H$_2$ suppression with shocking inflows: testing a pathway for supermassive black hole formation

Ricardo Fernandez, Greg L. Bryan,* Zoltan Haiman and Miao Li

Department of Astronomy, Columbia University, 550 West 120th Street, New York, NY 10027, USA

Accepted 2014 January 31. Received 2014 January 28; in original form 2013 August 25

ABSTRACT

The presence of quasars at redshifts $z > 6$ indicates the existence of supermassive black holes (SMBHs) as massive as a few times $10^6$ M$_\odot$, challenging models for SMBH formation. One pathway is through the direct collapse of gas in $T_{\mathrm{vir}} \gtrsim 10^4$ K haloes; however, this requires the suppression of H$_2$ cooling to prevent fragmentation. In this paper, we examine a proposed new mechanism for this suppression which relies on cold-mode accretion flows leading to shocks at high densities ($n > 10^4$ cm$^{-3}$) and temperatures ($T > 10^4$ K). In such gas, H$_2$ is efficiently collisionally dissociated. We use high-resolution numerical simulations to test this idea, demonstrating that such haloes typically have lower temperature progenitors, in which cooling is efficient. Those haloes do show filamentary flows; however, the gas shocks at or near the virial radius (at low densities), thus preventing the proposed collisional mechanism from operating. We do find that if we artificially suppress H$_2$ formation with a high-UV background, so as to allow gas in the halo centre to enter the high-temperature, high-density ‘zone of no return’, it will remain there even if the UV flux is turned off, collapsing to high density at high temperature. Due to computational limitations, we simulated only three haloes. However, we demonstrate, using Monte Carlo calculations of $10^6$ halo merger histories, that a few rare haloes could assemble rapidly enough to avoid efficient H$_2$ cooling in all of their progenitor haloes, provided that the UV background exceeds $J_{\nu} >$ few at redshifts as high as $z \sim 20$.

Key words: black hole physics – methods: numerical – cosmology: theory.

1 INTRODUCTION

Dynamical evidence indicates that most nearby galaxies harbour a central supermassive black hole (SMBH; e.g. Ferrarese & Ford 2005), including our own Milky Way, which hosts a central SMBH with mass $\lesssim 10^6$ M$_\odot$ (Ghez et al. 2005). Furthermore, the discovery of quasars at redshifts greater than six signals the existence of SMBHs as massive as a few times $10^6$ M$_\odot$ at an epoch when the Universe was less than a billion years old (e.g. Fan 2006; Mortlock et al. 2011). Such massive and early SMBHs pose a challenge to current models of their formation.

One possible formation scenario is the growth of a remnant black hole (BH) seed, generated from a Population III (Pop III) star ($\sim 100$ M$_\odot$), by mergers and gas accretion (e.g. Haiman & Loeb 2001; Volonteri, Haardt & Madau 2003; Li et al. 2007). However, this formation scenario poses certain difficulties. The time to assemble a $10^6$ M$_\odot$ SMBH by standard Eddington accretion is comparable to the age of the universe at $z \sim 6$ and it is unlikely that the seed BH will have continual accretion due to negative feedback and merger-induced gravitational recoils (Alvarez, Wise & Abel 2009; Milosavljević, Couch & Bromm 2009; Tanaka & Haiman 2009; Jeon et al. 2012; Tanaka, Perna & Haiman 2012).

An alternative pathway is the direct collapse of metal-free primordial gas with virial temperature $\gtrsim 10^4$ K into a BH seed of mass $10^3$–$10^6$ M$_\odot$ (Oh & Haiman 2002; Bromm & Loeb 2003; Regan & Haehnelt 2009; Shang, Bryan & Haiman 2010). Such a large seed requires fewer Salpeter times to grow to quasar size and so bypasses many of the difficulties of Eddington growth of stellar BH seeds. The exact mechanism by which the collapse occurs is not entirely clear (e.g. Bromm & Loeb 2003; Begelman, Rossi & Armitage 2008; Begelman 2010); however, a vital condition for this scenario is that the collapsing gas avoids fragmentation into stars. A natural way to avoid fragmentation is to have a long Jeans length due to a high gas temperature. The temperature of the gas depends on the interplay of atomic and molecular cooling. In the absence of H$_2$, a halo with $T_{\mathrm{vir}} \gtrsim 10^4$ K cools to $\sim 8000$ K by atomic hydrogen. Including H$_2$, the halo cools further to $\sim 200$ K. The corresponding Jeans mass at characteristic central densities $M_s \approx 10^6$ M$_\odot$ ($T/10^4$ K)$^{3/2}$ is $10^3$ M$_\odot$ for the latter, suggesting the formation of a Pop III star, and $10^6$ M$_\odot$ for the former, suggesting direct collapse into a massive BH. Therefore, a necessary condition for direct collapse is to prevent cooling by H$_2$.

The suppression of H$_2$ can be accomplished by a strong far-ultraviolet (UV) radiation flux in the Lyman–Werner (LW) bands,
suppression with shocking inflows

$J_{21} \geq 10^2 - 10^3$ (in units of $10^{-21}$ erg s$^{-1}$ cm$^{-2}$ Hz$^{-1}$ sr$^{-1}$; Omukai 2001; Bromm & Loeb 2003; Shang et al. 2010). This photodissociates the molecular hydrogen. However, only a small subset $\lesssim 10^{-6}$ of all atomic cooling haloes are estimated to be exposed to such levels, due to the presence of close luminous neighbour (Dijkstra et al. 2008; Agarwal et al. 2012). This makes it difficult to explain the production of the observed quasar population at $z > 6$. Note that the threshold only increases in the presence of a cosmic ray/X-ray flux (Inayoshi & Omukai 2011).

Inayoshi & Omukai (2012) have proposed an alternative mechanism for the suppression of H$_2$ cooling that does not depend on having a high-UV flux. In this scenario, cold-accretion flows penetrate to the centre of the halo, colliding with each other and shocking to produce hot and dense gas. The post-shock layer cools efficiently due to atomic hydrogen cooling and contracts isobarically until the gas reaches $\sim$8000 K. If the shocked gas at this high temperature is already at a high enough number density $n \gtrsim 10^{13}$ cm$^{-3}$, then H$_2$ rotational–vibrational levels reach local thermodynamic equilibrium and collisional dissociation can destroy the molecular hydrogen. Crucially, once the gas is shocked to this high-temperature, high-density regime, it will no longer be able to cool via H$_2$, even in the absence of any LW radiation. Throughout this paper, we will thus refer to this regime (defined more precisely below) as the ‘zone of no return’. Inayoshi & Omukai (2012) have argued that this mechanism may be able to produce a massive BH seed without strong radiative feedback; however, their numerical experiments focused on one-zone models and so questions still remain about the applicability of this pathway in cosmological simulations. In particular, only fluid elements that have a sufficiently high temperature and density have their fragmentation suppressed, and it is not clear if: (1) the gas, which ends up in haloes with such large virial temperatures, is not first processed through lower mass haloes in which fragmentation and star formation can occur$^1$ and (2) a sufficient amount of gas enters the ‘zone of no return’ for this mechanism to be important in realistic haloes.

To address these questions, we explore the possibility of H$_2$ suppression via cold-accretion shocks by conducting numerical simulations. This paper is organized in the following manner. In Section 2, we describe the ingredients and initial setup of the code. In Section 3, we describe the results of our numerical simulations followed by a discussion in Section 4. Finally, in Section 5, we summarize our conclusions.

2 NUMERICAL METHOD

The simulations were performed with the publicly available Eulerian adaptive mesh refinement (AMR) ENZO code. (Bryan 1999; Norman & Bryan 1999; O’Shea et al. 2004; Bryan et al. 2013). The code implements an N-body particle mesh technique (Efstathiou et al. 1985; Hockney & Eastwood 1988) to follow the dynamics of the dark matter (DM) particles and an Eulerian AMR method (Berger & Colella 1989) for the gas. In addition, ENZO provides modules which compute the radiative cooling of the gas as well as solve the chemical reaction network of a primordial mixture of H and He. Our simulations use the H$_2$ cooling function of Galli & Palla (1998) and solve the non-equilibrium evolution of the following nine species: H, H$^+$, He, He$^+$, He$^{++}$, H$_2$, H$_3^+$, and e$^-$ (Abel et al. 1997; Abel, Bryan & Norman 2000). Density-dependent collisional dissociation (Martin, Schwarz & Mandy 1996) is important – we include this with a rate as described in Shang et al. (2010).

We use a set of zoom simulations to focus on a number of haloes selected from a 1 h$^{-1}$ Mpc comoving box, with a root grid resolution of 128$^3$, using standard Λ cold DM model parameters: $\Omega_{\Lambda,0} = 0.721, \Omega_{m,0} = 0.233, \Omega_{b,0} = 0.233, \sigma_8 = 0.817, n_s = 0.96$, and $h = 0.701$ (Komatsu et al. 2009). Initially, we performed a low-resolution run with the maximum refinement level set to 4 and radiative cooling turned off to inhibit the gas from collapsing to high densities. We evolved this simulation from $z = 99$ to 10. Then, we applied the HOP halo finder (Eisenstein & Hut 1998) to the resulting data files at various redshifts to identify haloes with masses corresponding to virial temperatures $\gtrsim 10^4$ K. Throughout this paper, we adopt the relation $T_{\text{vir}} = 0.75 \times 1800(M/10^9$ M$_\odot)^{1/2}$ (1 + $z$)/21 K between halo mass and virial temperature. This is consistent with the commonly adopted version for neutral primordial gas with mean molecular weight $\mu = 1.2$ (Bryan & Norman 1998), except that we reduced the normalization by a factor of 0.75. We have found that this correction agrees better with our simulations – that is, it yields $T_{\text{vir}} = 10^4$ K for haloes at the redshift and mass when they begin to cool efficiently via atomic H.

Three haloes where selected at random (see Table 1) to be re-run at high resolution. We re-generated the initial conditions for the volumes, adding three nested grids that enclosed the Lagrangian volume of each halo. Since each additional grid doubles the spatial resolution, this resulted in an innermost grid with an effective resolution of 1024 and a DM particle mass of $\sim$85 M$_\odot$. Radiative cooling and multispecies were turned on to self-consistently follow the build-up of molecular hydrogen. During the course of the simulation, each cell was adaptively refined using the following three criteria: baryon mass, DM mass, and Jeans length. For the first two criteria, refinement is added when the baryon or DM mass exceeds four times the mass of the initial most refined cell, corresponding to mass resolutions of 68 and 340 M$_\odot$ for the baryons and DM, respectively. The third criterion enforces the Truelove et al. (1997) condition which states that at least four cells should resolve the Jeans length to avoid artificial fragmentation. In our simulations, the Jeans length was resolved by 16 cells to be sure that we adequately followed the collapse. Based on these criteria, the simulations were allowed to refine to a maximum level of 18, which corresponds to a comoving scale of 0.0298 h$^{-1}$ pc. The DM distribution was smoothed at refinement level 13 (about 0.065 proper pc at $z = 20$) to suppress numerical effects from the discreteness of DM particles.

We carried out two sets of runs for each of the haloes, which differed only in the background LW flux that we adopted. In the first set of runs, we used $J_{21} = 10$, where $J_{21}$ is the specific intensity in the LW bands (11.2–13.6 eV) in units of $10^{-24}$ erg cm$^{-2}$ sr$^{-1}$ Hz$^{-1}$. This corresponds to a typical (but slightly high) value in the late pre-ionization period (Dijkstra et al. 2008) and is well below that required to suppress H$_2$ radiatively (e.g. Shang, Bryan &...
Haiman 2010; Wolcott-Green, Haiman & Bryan 2011). We use these simulations to determine if the H$_2$ suppression mechanism suggested by Inayoshi & Omukai (2012) can be responsible for halting fragmentation in these haloes. We carry out these runs until they collapse to high densities and then examine the resulting gas distribution. As we will show, these haloes do form an abundant supply of H$_2$, and so the Inayoshi & Omukai (2012) mechanism by itself does not appear to be sufficient to allow direct collapse. In fact, cooling and collapse set in well before the virial temperature reaches $10^4$ K.

In a second set of simulations, we adopt a much higher value of the LW background, in particular, we take $J_{21} = 10^7$, which is well above the critical flux required to suppress H$_2$ formation and cooling. We evolve these simulations until their virial temperatures are above $10^4$ K, which allows us to (artificially) run the halo until it has a virial temperature sufficiently high for the Inayoshi & Omukai (2012) mechanism to operate. We then turn the LW background down to $J_{21} = 10$ to see if and how the purely collisional H$_2$ suppression acts. These runs are intended as an academic exercise to see if gas that is in or near the ‘zone of no return’ will indeed stay there.

3 RESULTS

In the following sections, we present the results of our numerical simulations, first focusing on runs with low background flux, and then follow with high-UV background simulations.

3.1 $J_{21} = 10$ runs

The addition of radiative cooling causes the haloes to cool and collapse before the redshifts at which they were identified in the low-resolution run. This occurs because the haloes build up in a hierarchical fashion and so have progenitor haloes with virial temperatures that allow them to cool via H$_2$ cooling. We stop the runs when they reach the maximum refinement level allowed in the simulation, as at this point they will start to form stars. The collapse redshifts, virial masses, and virial temperatures of the haloes at the point at which they collapse are shown in Table 2. Only the third halo reaches a virial temperature close to $10^4$ K. In this context, we define collapse as the point when we would need additional refinement in order to continue evolving the halo, which corresponds to a density of approximately $10^6$ cm$^{-3}$.

To investigate if the density-dependent H$_2$ suppression mechanism is operating, we analyse the simulation results at the point at which they collapse. In Fig. 1, we plot the temperature and density distribution of all cells within a radius of a few times the virial radius for each simulation run. Additionally, we overlay the ‘zone of no return’ in this figure, using the approximate curves for the boundary of the zone from Inayoshi & Omukai (2012). From the plots, we see directly that no cells have been shock-heated into this ‘no return zone’. This indicates that the H$_2$ suppression mechanism is not operating, at least for these haloes at this point.

Inayoshi & Omukai (2012) argued that cold, filamentary flows would penetrate to the centre of these haloes, shock-heating the gas at high densities. To better understand why this is not happening, in Fig. 2, we plot slices through the centre of halo C (with $T_{\text{vir}} = 8.1 \times 10^5$ K), showing the density, temperature, entropy, negative of the divergence of the velocity field, and H$_2$ fraction. From the density and temperature slices, we clearly see the cold filamentary structure feeding the halo. However, it is evident that no cold flows ($T \sim 10^3$–$10^4$ K) penetrate unperturbed into the centre. Instead, the gas quickly heats to the average temperature around the virial radius. This is consistent with the phase plots which show continual heating of the cells as number densities approach $\sim 1$ cm$^{-3}$. Most of the cells at this number density have roughly $T \sim 10^4$ K. At higher densities $\gtrsim 1$ cm$^{-3}$, H$_2$ cooling becomes dominant, leaving the gas roughly at $\sim 700$ K.

One way to identify shocks is to look for sources of entropy production (although shocks are not the only source of entropy). Fig. 2 shows that the largest entropy production happens around the virial radius. This is largely due to the spherical accretion of the surrounding cold gas. This is more apparent in the Mach number slice, where the flow outside the virial radius is supersonic and the flow interior to the virial radius is mostly transonic. The transition from supersonic to subsonic flow happens abruptly. We also see that the cold inflows are not immune to this shock, and gas in the filaments begins to exhibit higher entropy at the virial radius (although the entropy is considerably lower than most of the gas at the virial temperature). The entropy generation is at the expense of its kinetic energy as evident in the Mach number slice where the inflow also becomes transonic as it passes the virial radius. The negative divergence of the velocity shown in Fig. 2 reinforces our previous statements. Shocks produce a large value of $-\nabla \cdot V$. It is evident there is a strong shock at the virial radius and the filamentary structure is disrupted as it passes through this radius.

This shows clearly that the cold, filamentary flows in these haloes shock at or around the virial radius, where their densities are low – too low for the H$_2$ suppression mechanism suggested in Inayoshi & Omukai (2012) to operate. Instead, the shocked filaments can efficiently form H$_2$ and cool.

We argue that this conclusion is largely consistent with (comparable) previous work. For example, Wise & Abel (2007) carried out similar simulations, finding that although filamentary gas can, in some cases, penetrate through the virial radius without shocking, it does not get past one-third of the virial radius (as stated in the abstract of that paper). Moreover, this occurs only in the absence of H$_2$ cooling (i.e. in their H+He only runs A6 and B6), which is consistent with our own findings. In any case, this penetration is not nearly far enough – for example, fig. 2 of that paper demonstrates that the densities do not get to $10^6$ cm$^{-3}$ until about 0.005 $r_{\text{vir}}$. The Smoothed Particle Hydrodynamics (SPH) simulations of Greif et al. (2008) also examine the inflow of filaments in haloes of similar size. They appear to find that cold filaments may not shock-heat as the flow in; however, two points are relevant here. The first is that later work (Nelson et al. 2013) has shown that SPH simulations incorrectly predict substantial amount of cold accretion compared to Eulerian or moving-mesh codes. The second point is that even in the simulations of Greif et al. (2008), gas does not get close to the H$_2$ zone of no return (e.g. their fig. 7). Therefore, we conclude that our key results are in agreement with previous work.

We have seen that – for the three haloes simulated – the haloes collapse and fragment without any gas entering the ‘zone of no

| Halo | $z_{\text{col}}$ | $T_{\text{vir, col}}$ (K) | $M_{\text{vir, col}}$ ($M_\odot$) | $r_{\text{vir, col}}$ (pc) |
|------|-----------------|--------------------------|-----------------------------|-------------------|
| A    | 21.81           | $2.86 \times 10^3$      | $2.72 \times 10^6$        | 199.14            |
| B    | 16.12           | $6.43 \times 10^3$      | $1.41 \times 10^7$        | 421.10            |
| C    | 12.89           | $0.81 \times 10^4$      | $2.75 \times 10^7$        | 642.13            |
Figure 1. Phase plots of number density and temperature for each of the $J_{21} = 10$ simulations (from left to right: haloes A, B, and C). The colour indicates how much mass is at each point in the phase diagram. In all three simulations, no cells were shock-heated to the ‘zone of no return’, shown by the dashed line and defined as in Inayoshi & Omukai (2012).

Figure 2. Slices of the $y$–$z$ plane for density, temperature, entropy, Mach number, $-\nabla \cdot \mathbf{V}$ and $H_2$ fraction for halo C (with $T_{\text{vir}} = 8.1 \times 10^3$ K) with a low LW background. The field of view of each plot is 5 kpc (physical) and the virial radius is 0.64 kpc.

return’. They do this in part because the $10^5$ K haloes do not collapse monolithically – instead, the hierarchical formation naturally involves lower temperature haloes which can cool efficiently via $H_2$.

One possible objection is that we have simply not simulated enough haloes, and that a small fraction of haloes might collapse quickly enough for the suppression mechanism to be efficient. We examine this possibility in more detail in Section 4; however, first (in the next section), we look at what would happen if we could somehow suppress cooling in the lower mass haloes until high-$T_{\text{vir}}$ haloes build up.

Before examining that issue, we make a brief remark about resolution – a concern with any result from a numerical simulation. Although we have not explicitly carried out a resolution study as part of this work, this point has been examined regularly in the past. For example, Machacek, Bryan & Abel (2001) explicitly ran multiple simulations (also with the enzo code) to examine how mass resolution impacted the cooling of primordial haloes like the ones simulated here. They used DM particles masses of 306 and 38 M$_\odot$, finding only minor changes in the statistics of cooling haloes, implying that, with a DM particle mass of 85 M$_\odot$, our results are robust to mass resolution. A related but slightly different concern is the resolution of the Jeans length – recent work has suggested that the Jeans length may, in some circumstances, need to be resolved by more than the 16 cells adopted here. For example, Turk et al. (2012) find that the magnetic properties require ratios greater than 32, with possible changes to the temperature profile as the resolution is changed. However, the temperature changes are not found in Greif, Springel & Bromm (2013), and moreover, the impact is only significant at higher densities than explored here, $n \gtrsim 10^{12}$ cm$^{-3}$. At lower densities, the profiles are well converged. Therefore, we conclude that this issue is unlikely to impact the results found here. We can also compare our profiles to previous work. For example, O'Shea & Norman (2008) carried out simulations of similar haloes with similar LW backgrounds – the temperature profiles they find (e.g. their fig. 9) are very similar to ours. As that paper demonstrates, the relatively high temperatures ($\sim 700$ K) we find in the $H_2$ cooling region are an indirect consequence of the relatively high LW background used here.

3.2 $J_{21} = 10^5$ run

To investigate if the collisional $H_2$ suppression mechanism can keep a halo from fragmenting if we could place its central region in or near the ‘zone of no return’, we re-simulate halo A using a much higher value of the LW background to artificially suppress $H_2$ cooling. We adopt $J_{21} = 10^5$ – with such a large (and unrealistic) background, most of the $H_2$ will be dissociated, as shown by...
Shang et al. (2010) and Wolcott-Green et al. (2011), where the critical flux was determined to be $J_{21} = 10^3$. We picked the halo with the smallest virial temperature at collapse in the previous runs ($T_{\text{vir}} = 2.86 \times 10^3$ K) and re-ran it with the large $J_{21}$. The halo collapses at a later redshift, $z = 16.74$, with a larger virial temperature $T_{\text{vir}} = 0.97 \times 10^4$ K, mass $2.48 \times 10^7 M_\odot$, and radius 534 pc.

In Fig. 3, we show the number density and temperature phase plot for this run. We see the cells for $n \gtrsim 1$ cm$^{-3}$ have now increased in temperature. The large $J_{21}$ has dissociated most of the H$_2$, leaving atomic cooling as the dominant mechanism and keeping the gas at $T \sim 8000$ K. For $n \lesssim 1$ cm$^{-3}$, the same structure is apparent as the previous simulations, where cells are heated to $T \sim 10^6$ K at the virial radius.

Fig. 4 shows slices with the same quantities as in the previous section, but now for the large $J_{21}$ run. Although the general structure is quite similar, there are now some signs of cold flows penetrating past the virial radius. This is more apparent in the Mach number slice where we see high-velocity flows passing the virial radius where the majority of the fluid is transonic. It can also be seen in the temperature distribution, where low-temperature gas flows into the halo before heating. However, we note that the ‘zone of no return’ occurs at densities above $\sim 10^3$ cm$^{-3}$, which corresponds to the very central region (coloured red) of the density slice, and so apparently, the cold flows are not penetrating this zone.

### 3.2.1 Stability of the zone of no return

Next, we investigate if gas in the ‘zone of no return’ will indeed stay there without any external flux. We do this by turning down the high-UV background flux in the previous simulation. We run two identical simulations as in the last section, but now reduce $J_{21}$ at different evolutionary points. The two points are distinguished by whether or not gas cells have reached the critical H$_2$ number density for local thermodynamic equilibrium, $n_{cr} \sim 10^4$ cm$^{-3}$. Thus, in our first simulation, we allow the gas to evolve with a large $J_{21}$ until

the maximum density just reaches $n_{cr}$, and at this point the UV background is reduced to $J_{21} = 10$ and the gas is allowed to evolve. Similarly, our second simulation is identical with the exception that the reduction is applied when a small fraction of cells have entered into the ‘zone of no return’.

Fig. 5 shows the phase plot evolution for the first simulation (the instant of flux reduction corresponds to the upper-left panel, at $z = 16.831$). As the simulation evolves with the reduced $J_{21}$, H$_2$ begins to form and cool efficiently. The cooling occurs first at densities $n \sim 10^3$ cm$^{-3}$, leaving the higher density regions still hot ($T \sim 10^4$ K). As the evolution progresses, this gas eventually cools and by the final time shown, when the halo collapses to the highest refinement level we allow, the phase distribution looks very much like the low-$J$ runs.

It is interesting to analyse this evolution in more detail. In Fig. 6, we show temperature (left) and density (right) slices that go through the centre of this halo for the same times presented in Fig. 5. We see that the cooling process creates a cold envelope surrounding the hot core. This core is overpressurised, which generates an outflow, driving clumps of dense material out of the central region, as evident in the density slices of Fig. 6. The outflow reduces the density (and temperature) of the gas in the core, which is then effectively evaporated. The final remnant is a dense shell surrounding a core that has cooled on average to $T \sim 700$ K.

In the second simulation of this section, we allow the run with $J_{21} = 10^3$ to continue a bit further, to $z = 16.754$, when a significant amount of gas has entered the ‘zone of no return’ (see the top-left panel of Fig. 7) before turning the UV background down to $J_{21} = 10$. The resulting evolution of the phase diagram is shown in Fig. 7. In this case, the evolution is quite different: although some gas does cool, the dense gas evolves more quickly, and by the time we stop the simulation (when it reaches the highest refinement level at
Figure 5. Phase plots showing the distribution of number density and temperature in halo A at a variety of redshifts, as indicated. The simulation ran with $J_{21} = 10^5$ until the maximum density reached the edge of the ‘zone of no return’ (as shown in the top-left panel). Thereafter, the gas is evolved with a reduced $J_{21} = 10$. Gas quickly forms $H_2$ and cools to a temperature well below the zone.

Figure 6. The left set of four panels shows temperature slices through the centre of halo A corresponding to the same simulation and output times as in Fig. 5, while the right set of four panels shows number density slices for the same halo at the same sets of times. After the background flux is reduced, the lower density region outside the core cools first, driving an outflow which evaporates the centre.

$z = 16.749$), the amount of mass inside the ‘zone of no return’ has not changed by more than a few per cent. There is some cooling – in particular, at $n \sim 10^3$ cm$^{-3}$ and $\sim 10^5$ cm$^{-3}$; however, this does not significantly affect the evolution of the clump.

The resulting baryonic mass in the ‘zone of no return’ is $\sim 10^5$ M$_\odot$, roughly 10 per cent of the halo mass within the virial radius.

4 DISCUSSION

Using a set of numerical simulations, we have investigated the viability of the mechanism for collisional suppression of $H_2$ suggested by Inayoshi & Omukai (2012). We found that, for the three haloes examined, $H_2$-mediated cooling sets in before the conditions required for suppression could be established – in particular, before...
a $T_{\text{vir}} \gtrsim 10^4$ K halo forms, which would be a pre-requisite for the generation of shocks strong enough to put gas into the ‘zone of no return’.

However, we did confirm (using a set of simulations that employed a high-UV background to temporary suppress cooling) some aspects of the mechanism: if the gas in the core of an atomic cooling halo can avoid H$_2$ cooling until it enters the ‘zone of no return’, then subsequent H$_2$ cooling can be naturally averted. The gas would then remain at $\sim 10^4$ K, possibly leading to conditions suitable for SMBH formation, as envisioned in the many previous works mentioned in the Introduction (see, e.g. Haiman 2013 for a recent review).

Unfortunately, H$_2$ cooling at any time in the past history of the same gas would likely invalidate this possibility, if this H$_2$ cooling allowed the gas to reach temperatures of a few 100 K and to evolve to high density at such low temperatures. This cold and dense gas, at the centre of the progenitor minihalo, would form one, or possibly a few, Pop III stars. The subsequent evolution of this progenitor into an atomic cooling halo is uncertain; however, there are many obstacles to rapid SMBH formation.

First, even a handful of SNe (or a single pair instability SNe) can enrich the entire atomic cooling halo to a metallicity of $Z \gtrsim 10^{-3} Z_\odot$ (e.g. Bromm & Yoshida 2011, see p. 396), the critical value above which direct SMBH formation is replaced by fragmentation (Inayoshi & Omukai 2012). It is possible that none of these Pop III stars produce any metals (such as non-rotating metal-free stars between 40 and 140 $M_\odot$; Heger et al. 2003). Such massive Pop III stars would leave behind stellar-mass BHs, which, in principle, could grow into SMBHs, provided that they are surrounded by very dense gas, and accrete sufficiently rapidly to trap their own radiation (e.g. Volonteri & Rees 2005). However, the parent Pop III stars of these seed BHs will create a large ionized bubble and therefore the seed BHs will likely begin their life in a low-density medium. As they later begin to accrete, their radiation likely self-limits the time-averaged accretion rate to a fraction of the Eddington rate (Alvarez et al. 2009; Milosavljević et al. 2009; Park & Ricotti 2012). Although the above scenarios are worth investigating further in detail, we will hereafter assume that these obstacles prevent rapid SMBH formation in any halo that experienced H$_2$ cooling at any time in its history.

### 4.1 Avoiding H$_2$ cooling by rapid halo assembly

However, since we only simulated three haloes, a natural question is whether H$_2$ cooling may be avoided in a few, highly atypical, atomic cooling haloes, even if the background flux $J_{BLW}$ was much lower. One possibility is that – in rare cases – the progenitor haloes could experience unusually rapid mergers, continuously shock-heating the gas on a time-scale that always remains shorter than the H$_2$-cooling time. Here, we evaluate the likelihood of this scenario, using Monte Carlo realizations of the merger histories of $10^6$ atomic cooling haloes.

In particular, we start by creating DM halo merger trees, using the Monte Carlo algorithm in Zhang, Fakhouri & Ma (2008b). This paper presents three different numerical algorithms, which are based on the conditional halo mass functions in the ellipsoidal collapse model (Sheth & Tormen 2002; Zhang, Ma & Fakhouri 2008a). We adopted their ‘method B’, which was found in subsequent work to provide the best match to the statistics of merger trees in $N$-body simulations (Zhang, private communication). This method represents an improvement over previous merger-tree methods based on

---

**Figure 7.** Phase plots showing the distribution of number density and temperature in halo A at a variety of redshifts, as indicated. The simulation ran with $J_{21} = 10^5$ until the maximum density reached well into the ‘zone of no return’ (top-left panel). Thereafter, the gas is evolved with a reduced $J_{21} = 10$. Although some gas at moderate densities begins to cool, the gas in the ‘zone of no return’ collapses more quickly than the outer region can cool.
spherical collapse – in particular, the ellipsoidal collapse models tend to predict a larger number of more massive progenitors, or a ‘flatter’ assembly history (Tanaka, Li & Haiman 2013).

We have created $10^6$ merger trees of a DM halo with $T_{\text{vir}} = 10^4$ K ($M_0 = 6.3 \times 10^7 M_\odot$) at redshift $z_0 = 10$, extending back to redshift $z = 20$. We then follow the mass of the most massive progenitor back in time, and at each redshift $z$, we compute the H$_2$-cooling time

$$t_{\text{H}_2} = \frac{1.5 n_e k_B T}{\Delta m_\text{H}_2}$$

in this progenitor, using the procedures and rates in Machacek et al. (2001). Specifically, we compute the maximum density

$$n_{\text{max}} = 187 \Omega_m h^2 (T_{\text{vir}}/1000 \text{ K})^{1.5} \text{ cm}^{-3}$$

that could be reached by the gas in the nucleus of this progenitor, in the absence of H$_2$-cooling (i.e. by adiabatic compression). We then specify a constant background flux $J_{\text{LW}}$ and compute the equilibrium H$_2$ fraction and the corresponding $t_{\text{H}_2}$ at this density, given $J_{\text{LW}}$ and $n_{\text{max}}$. Finally, we require that the longer of the dynamical time $t_{\text{dyn}} = \sqrt{3M/16G \rho_{\text{max}}}$ and $t_{\text{H}_2}$ remains longer than the cosmic time $\Delta t(z) = t(z) - t(z_0)$ elapsed between the $z$ and $z_0$. This last requirement ensures that the most massive progenitor of our atomic cooling halo at redshift $z > z_0$ cannot cool via H$_2$ and evolve dynamically, before it is incorporated by mergers into the $M_0 = 6.3 \times 10^7 M_\odot$ atomic cooling halo at redshift $z_0 = 10$. Note that this requirement must be satisfied at all redshifts $z > z_0$ – in other words, if our atomic cooling halo had a progenitor, at any redshift $z$, that was massive enough to cool and collapse, the ‘SMBH formation by direct collapse’ scenario is no longer feasible.

In practice, as we run our merger trees backwards, we simply look, at each redshift step, whether the most massive progenitor violates the criterion $\Delta t(z) > n_{\text{max}}[t_{\text{H}_2}, t_{\text{dyn}}]$. If it does, we record the redshift $z_{\text{max}}$ where this happened, and we discard the rest of this merger tree.

In Fig. 8, we show the probability distribution of the ‘terminal lookback redshift’ $z_{\text{max}}$ we have found among the $10^6$ merger trees, for four different values of a constant $J_{\text{LW}} = 1, 3, 10$, and 30. For an LW background $J_{\text{LW}} = 1$ (shown in the figure as the solid black histogram), we found that all of the $10^6$ merger trees contain an H$_2$-cooling progenitor by $z < 20$. However, for the higher background fluxes of $J_{\text{LW}} = 3$ (dot–dashed green), $J_{\text{LW}} = 10$ (dotted red), or $J_{\text{LW}} = 30$ (dashed blue), we find approximately 2, 200, and 4000 cases, respectively, where none of the progenitors, out to $z = 20$, could cool via H$_2$ – this is shown by the pile-up of the probability distribution in the last bin at $z_{\text{max}} = 10$.

**5 SUMMARY**

One way to explain the presence of SMBHs at $z > 6$ with masses greater than $10^9 M_\odot$ is through the formation of a massive ($\sim 10^5 M_\odot$) BH seed from gas in $T_{\text{vir}} \gtrsim 10^4$ K haloes, either via direct collapse to a BH or through the formation of a supermassive star or a quasi-star (Begelman et al. 2008; Hosokawa et al. 2013). As noted in the Introduction, one difficulty with these modes is the need to prevent fragmentation of the cooling halo into lower mass stars, and, in particular, to prevent cooling due to H$_2$. Inayoshi & Omukai (2012) have suggested that this can be accomplished through the action of cold flows, which results in gas shocking to high densities ($n > 10^4$ cm$^{-3}$) and temperatures ($T > 10^4$ K). This shocked gas cools by Ly$\alpha$ emission to about 8000 K; however, if the density is high enough, enhanced H$_2$ collisional dissociation suppresses the gas from cooling further. This was demonstrated using one-zone models calculated by Inayoshi & Omukai (2012), who identified a ‘zone of no return’. This mechanism is appealing as it does not require a high LW background to destroy the H$_2$; however, it is unclear if this mechanism operates in nature.

To test this idea, we carried out cosmological hydrodynamic simulations with the AMR code ENZO. We first identified three haloes from a low-resolution simulation which all had $T_{\text{vir}} \gtrsim 10^4$ K at redshifts ranging from 12 to 17. We then re-simulated these haloes at high resolution with a relatively modest UV flux in the LW band, $J_{\text{LW}} = 10$. We found that in all three cases, cooling from H$_2$ was able to efficiently lower the gas temperature below that of the ‘zone of no return’, indicating that, at least for these three haloes, the mechanism was not operating.

To determine why this occurred, we examined the structure of the simulated haloes in detail, and found that, while cold flows do occur, they generally shock at or near the virial radius and do not penetrate into the halo centre where the densities are high. We note that although these small haloes have low virial temperatures and
so might be naively classified as ‘cold-mode’ haloes (Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006), they actually have low cooling rates because of the inefficiency of H₂ cooling. Therefore, the characteristic cooling times of these haloes is longer (or comparable) to their dynamical times, making them more akin to hot-mode haloes. This is consistent with the clear virial shocks that we see in these haloes (e.g. Fig. 2).

To determine if the suppression mechanism could function if we artificially suppressed H₂ cooling, we re-ran one of the simulations with a high LW background (J₂₁ = 10¹⁰), well above the critical flux required to suppress H₂ cooling radiatively (Shang et al. 2010; Wolcott-Green et al. 2011). We showed that, in this case, the gas did indeed enter the ‘zone of no return’. To see if this situation was stable, we ran two variations of this simulation, in which we turned the flux off either just before or just after the gas in the halo centre entered the ‘zone of no return’. In the first case, cooling eventually won out, with the region just outside the core cooling first and driving an evaporative wind which led to cooling by H₂ throughout the halo. However, in the second case, the gas inside the ‘zone of no return’ stayed there, and eventually collapsed to high densities while remaining at T ~ 8000 K, despite the lack of any LW flux.

This demonstrates that the mechanism would work if a halo could collapse quickly enough, but that typical Tvir > 10⁶ K haloes have progenitors which can cool efficiently. To investigate if any haloes can collapse quickly enough to escape this fate, we ran Monte Carlo merger-tree calculations of 10⁴ haloes with an ellipsoidal collapse model. We used a simple analytic prescription to determine if any haloes could assemble quickly enough to prevent cooling and collapse prior to building up to a Tvir = 10⁶ K halo. We found that for an LW background of J₂₁ = 1, none of the 10⁶ merger-tree histories collapsed quickly enough. For larger LW backgrounds, a small fraction (up to ≲ 10⁻³ for J₂₁ as high as 30) of the haloes did form quickly enough to evade cooling. We note that at higher redshift, J₂₁ would certainly be expected to be lower.

In summary, we confirm the essential physics proposed by Inayoshi & Omukai (2012), but we find that the supersonic filamentary flows are unlikely to shock gas into the ‘zone of no return’, as they had originally envisioned. Rather, we propose here a modification of their scenario. In the core of a few rare haloes, which assembled unusually rapidly, the gas may have been kept continuously shock-heated throughout their history, and eventually shocked into the ‘zone of no return’. This scenario reduces the required LW flux by a factor of ≲ 300, from J₂₁ ≳ 10⁷ (Shang et al. 2010; Wolcott-Green et al. 2011) to J₂₁ ≳ 1, but it does require that the latter LW background is in place at a redshift as high as z ~ 20. Future, self-consistent global models of the build-up of the LW background, together with the assembly of a large number of atomic-cooling haloes, are required to assess the viability of this scenario.

ACKNOWLEDGEMENTS

We thank Jun Zhang for useful discussions about ellipsoidal collapse-based merger trees and for sharing unpublished results on comparisons between Monte Carlo methods and N-body simulations. ZH acknowledges financial support from NASA through grant NNX11AE05G. GB acknowledge financial support from NSF grants AST-0908390 and AST-1008134, and NASA grant NNX12AH41G, as well as computational resources from NSF XSEDE and Columbia University’s Hotfoot cluster. Analysis was done with the yt package (Turk et al. 2011) and we thank Matthew Turk for useful advice and discussion.

REFERENCES

Abel T., Anninos P., Zhang Y., Norman M. L., 1997, New Astron., 2, 181
Abel T., Bryan G. L., Norman M. L., 2000, ApJ, 540, 39
Agarwal B., Khochfar S., Johnson J. L., Neistein E., Dalla Vecchia C., Livio M., 2012, MNRAS, 425, 2854
Alvarez M. A., Wise J. H., Abel T., 2009, ApJ, 701, L133
Begelman M. C., 2010, MNRAS, 402, 673
Begelman M. C., Rossi E. M., Armitage P. J., 2008, MNRAS, 387, 1649
Berger M. J., Colella P., 1989, J. Comput. Phys., 82, 64
Birnboim Y., Dekel A., 2003, MNRAS, 345, 349
Bromm V., Loeb A., 2003, ApJ, 596, 34
Bromm V., Yoshida N., 2011, ARA&A, 49, 373
Bryan G., 1999, Comput. Sci. Eng., 1, 46
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bryan G. L. et al., 2013, preprint (arXiv:1307.2265)
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dijkstra M., Haiman Z., Mesinger A., Wyithe J. S. B., 2008, MNRAS, 391, 1961
Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241
Eisenstein D. J., Hut P., 1998, ApJ, 498, 137
Fan X., 2006, New Astron., 50, 665
Ferrarese L., Ford H., 2005, Space Sci. Rev., 116, 523
Galler D., Palla F., 1998, A&A, 335, 403
Ghez A. M., Salim S., Hornstein S. D., Tanner A., Lu J. R., Morris M., Becklin E. E., Duchêne G., 2005, ApJ, 620, 744
Greif F. H., Johnson J. L., Klessen R. S., Bromm V., 2008, MNRAS, 387, 1021
Greif F. H., Springel V., Bromm V., 2013, MNRAS, 434, 3408
Haiman Z., 2013, in Wiklind T., Mobasher B., Bromm V., eds, Astrophysics and Space Science Library, Vol. 396, The First Galaxies. Springer-Verlag, Berlin, p. 293
Haiman Z., Loeb A., 2001, ApJ, 552, 459
Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, ApJ, 591, 288
Hockney R. W., Eastwood J. W., 1988, Computer Simulation Using Particles. IoP Publishing, Bristol
Hosokawa T., Yorke H. W., Inayoshi K., Omukai K., Yoshida N., 2013, ApJ, 778, 178
Inayoshi K., Omukai K., 2011, MNRAS, 416, 2748
Inayoshi K., Omukai K., 2012, MNRAS, 422, 2539
Jeon M., Pawlik A. H., Greif T. H., Glover S. C. O., Bromm V., Milosavljević M., Klessen R. S., 2012, ApJ, 754, 34
Kereš D., Katz N., Weinberg D. H., Dave R., 2005, MNRAS, 363, 2
Komatsu E. et al., 2009, ApJS, 180, 330
Li Y. et al., 2007, ApJ, 665, 187
Machacek M. E., Bryan G. L., Abel T., 2001, ApJ, 548, 509
Martin P. G., Schwarz D. H., Mandy M. E., 1996, ApJ, 461, 265
Milosavljević M., Couch S. M., Bromm V., 2009, ApJ, 696, L146
Mortlock A., Conselice C. J., Bluck A. F. L., Bauer A. E., Gützbauch R., Buitrago F., Ownsworth J., 2011, MNRAS, 413, 2845
Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353
Norman M. L., Bryan G. L., 1999, in Miyama S. M., Tomisaka K., Hanawa T., eds, Astrophysics and Space Science Library, Vol. 240, Numerical Astrophysics. Kluwer, Dordrecht, p. 19
O’Shea B. W., Norman M. L., 2008, ApJ, 673, 14
O’Shea B. W., Bryan G., Bordner J., Norman M. L., Abel T., Harkness R., Krissuk A., 2004, preprint (astro-ph/0403044)
Oh S. F., Haiman Z., 2002, ApJ, 569, 559
Omukai K., 2001, ApJ, 546, 635
Park K., Ricotti M., 2012, ApJ, 747, 9
Regan J. A., Haehnelt M. G., 2009, MNRAS, 393, 858
Shang C., Bryan G. L., Haiman Z., 2010, MNRAS, 402, 1249
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Tanaka T., Haiman Z., 2009, ApJ, 696, 1798
Tanaka T., Perna R., Haiman Z., 2012, MNRAS, 425, 2974
Tanaka T., Li M., Haiman Z., 2013, MNRAS, 435, 3559

MNRS 439, 3798–3807 (2014)
Truelove J. K., Klein R. I., McKee C. F., Holliman J. H., II, Howell L. H., Greenough J. A., 1997, ApJ, 489, L179
Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, ApJS, 192, 9
Turk M. J., Oishi J. S., Abel T., Bryan G. L., 2012, ApJ, 745, 154
Volonteri M., Rees M. J., 2005, ApJ, 633, 624
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
Wise J. H., Abel T., 2007, ApJ, 665, 899
Wolcott-Green J., Haiman Z., Bryan G. L., 2011, MNRAS, 418, 838
Zhang J., Ma C.-P., Fakhouri O., 2008a, MNRAS, 387, L13
Zhang J., Fakhouri O., Ma C.-P., 2008b, MNRAS, 389, 1521

This paper has been typeset from a \LaTeX\ file prepared by the author.