

Rapid formation of massive black holes in close proximity to embryonic protogalaxies

John A. Regan^{1,2*}, Eli Visbal^{3,4}, John H. Wise⁵, Zoltán Haiman^{3,6}, Peter H. Johansson⁷
and Greg L. Bryan^{3,4}

The appearance of supermassive black holes at very early times^{1–3} in the Universe is a challenge to our understanding of star and black hole formation. The direct-collapse^{4,5} black hole scenario provides a potential solution. A prerequisite for forming a direct-collapse black hole is that the formation of (much less massive) population III stars be avoided^{6,7}; this can be achieved by destroying H₂ by means of Lyman–Werner radiation (photons of energy around 12.6 eV). Here we show that two conditions must be met in the protogalaxy that will host the direct-collapse black hole. First, prior star formation must be delayed; this can be achieved with a background Lyman–Werner flux of $J_{\text{BG}} \gtrsim 100J_{21}$ (J_{21} is the intensity of background radiation in units of $10^{-21} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \text{ sr}^{-1}$). Second, an intense burst of Lyman–Werner radiation from a neighbouring star-bursting protogalaxy is required, just before the gas cloud undergoes gravitational collapse, to suppress star formation completely. Using high-resolution hydrodynamical simulations that include full radiative transfer, we find that these two conditions inevitably move the host protogalaxy onto the isothermal atomic cooling track, without the deleterious effects of either photo-evaporating the gas or polluting it with heavy elements. These atomically cooled, massive protogalaxies are expected ultimately to form a direct-collapse black hole of mass 10^4 – $10^5 M_{\odot}$.

To probe the unique combination of a background radiation field in tandem with an intense proximate burst, we perform a series of high-resolution radiation-hydrodynamic simulations using the adaptive mesh refinement code Enzo⁸ together with the Grackle multi-species library for solving the primordial (H+He) chemistry network and regulating the radiation backgrounds⁹. Radiative transfer is handled by Enzo's Moray ray-tracing package¹⁰.

We use relatively mild backgrounds (compared with the rather high values often cited in the literature, where background fields of $\gg 1,000J_{21}$ are typically invoked) from $J_{\text{BG}} = 100J_{21}$ up to $J_{\text{BG}} = 200J_{21}$ for an effective background temperature of 30,000 K (we normalize J_{21} at the hydrogen ionization edge throughout). This is the effective temperature expected from a population of metal-free and partially metal-enriched stars in the early Universe. This (relatively mild) global background radiation field is sufficient to delay the collapse but will not prevent the formation of H₂ as the halo mass increases (halo encompasses both the protogalaxy and the dark-matter structure surrounding the protogalaxy). To model the nearby source,

we use the 'synchronized pairs' scenario in which a protogalaxy (secondary halo) is exposed to the intense radiation from a star-burst (primary halo) that is sufficiently nearby^{11–14} and tightly synchronized in time¹⁵. The model is illustrated in Fig. 1. If the primary halo crosses the atomic cooling threshold (and begins forming stars through molecular collapse) shortly before a neighbouring, secondary halo, the primary halo can bombard the secondary halo with a critical flux of Lyman–Werner radiation. This bombardment destroys the H₂ in the secondary halo because of the extremely high flux of the primary halo. It provides the final push to the secondary halo, forcing the halo onto the atomic cooling track leading to direct-collapse black hole (DCBH) formation. Forming massive black holes in this way is a promising way of producing massive black hole seeds in the early Universe.

Further consideration must be given to preventing both metal contamination and photo-evaporation (from photons with energy $E > 13.6 \text{ eV}$) from neighbouring haloes. Metal cooling will rapidly reduce the temperature of the collapsing gas, resulting in strong fragmentation (the cooling time at a density of $n_{\text{HI}} \approx 100 \text{ cm}^{-3}$ is about 10 Myr for gas with a metallicity of $Z = 10^{-4}$ – $10^{-3} Z_{\odot}$). Therefore, for the Jeans mass to remain large, metal pollution from nearby galaxies must be avoided (although this can be mitigated by inefficient or slow mixing due to external metal enrichment¹⁶). In addition, photo-evaporation in haloes exposed to star-burst radiation for longer than $\sim 40 \text{ Myr}$ from a source with luminosity $L = 1.64 \times 10^{41} \text{ erg s}^{-1}$ and a separation of between 0.5 and 1 kpc was observed in previous simulations¹⁷, effectively limiting the time available for creating a pristine atomic cooling halo (ACH) using a nearby neighbour.

Our results show that to achieve an ACH, the separation, R_{sep} , between the primary and secondary haloes needs to be less than approximately 300 pc for the stellar luminosity of the primary halo adopted in our model ($1.2 \times 10^{52} \text{ photons s}^{-1}$, or a stellar mass of $M_{\star} \approx 10^5 M_{\odot}$). This critical distance will vary as $d_{\text{crit}} \propto \sqrt{M_{\star}}$ but for the luminosities examined in this paper will be approximately 300 pc. We define the synchronization time as the time between when the primary halo is turned on and when the gas in the secondary halo would have collapsed in the absence of the primary halo (see Fig. 1). The synchronization times that successfully produce an ACH can range from a few kiloyears to a few megayears (see column 6 in Supplementary Table 1). Short synchronization times reduce the probability of metal pollution and/or photo-evaporation from the primary halo.

¹Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK. ²Centre for Astrophysics and Relativity, School of Mathematical Sciences, Dublin City University, Glasnevin D09 Y5N0, Dublin, Ireland. ³Department of Astronomy, Columbia University, 550 West 120th Street, New York, New York, 10027, USA. ⁴Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, New York, 10003, USA. ⁵Center for Relativistic Astrophysics, Georgia Institute of Technology, 837 State Street, Atlanta, Georgia 30332, USA. ⁶Department of Physics, New York University, New York, New York 10003, USA. ⁷Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2a, FI-00014 Helsinki, Finland.

*e-mail: john.regan@dcu.ie

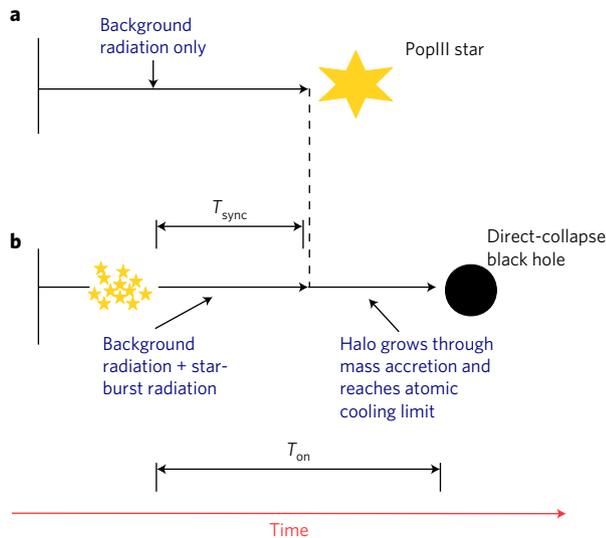


Figure 1 | Modelling synchronized haloes. a,b, The synchronized protogalaxy scenario. With only a background field in operation, a (delayed) PopIII star forms because of H₂ cooling (a). If a nearby star-burst galaxy, in conjunction with the background, provides the critical Lyman–Werner flux required, then a DCBH can form in an atomic cooling halo (b). T_{sync} is defined as the time between the star-burst turning on and the point at which a PopIII star would have formed. T_{on} is the time taken for an atomic cooling halo to collapse and form a DCBH (or the minimum time for which the source must shine).

In Fig. 2, we show a selection of ray profiles using our primary halo as the local galactic source (the stellar mass of the primary halo was determined from high-resolution simulations of the high-redshift Universe¹⁸). Ray profiles are generated by tracing 500 individual rays originating at the radiation source (centre of the primary halo) and terminating on a plane with a radius of approximately 2 pc around the centre of the secondary halo. For each simulation, we show the temperature, H₂ fraction, hydrogen number density and the intensity in the Lyman–Werner band as a function of distance from the centre of the secondary halo. Of the six cases examined in Fig. 2, four collapse nearly isothermally and form ACHs, whereas the two others form excessive H₂, cooling significantly. Three of the ‘successful’ simulations lie within the critical distance of 300 pc, with the fourth at 388 pc. The thermal profile of the fourth simulation is not as smooth as the closer-in simulations, and this case is therefore likely to be close to a tipping point (that is, a slightly larger separation would probably have resulted in a non-isothermal collapse). The virial radius of our secondary halo is approximately 300 pc, meaning that for the neighbouring source to have the greatest probability of completely suppressing H₂ cooling, the virial radii of the primary and secondary must overlap¹⁵. Primary sources that lie outside the virial radius of the secondary do not, in general, provide sufficient flux without unrealistically high star-formation efficiencies.

How spatially and temporally correlated do the halo pairs need to be? In Fig. 3, we show the average separation against the synchronization time, T_{sync} , for all of the simulations conducted in this study. As noted above, for separations of $R_{\text{sep}} \gtrsim 300$ pc, ACHs tend not to be formed, and instead a critical level of H₂ builds up, leading to a non-isothermal collapse. Owing to the possibility of metals from supernova explosions polluting the pristine environment of the secondary halo and the detrimental effects of long-term exposure of the gas to ionizing radiation, we disfavour sources that must be ‘on’ for more than 10 Myr (which is comparable to the lifetime of massive stars) before the initial collapse occurs. We therefore do not

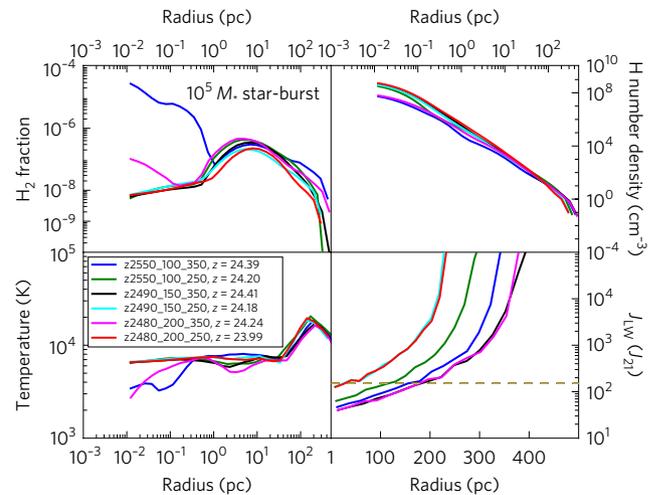


Figure 2 | Ray profiles for six selected haloes. A selection of ray profiles from simulations using a star-bursting galaxy source as the local radiation source. For an atomic cooling halo to form, the distance to the source must, in general, be less than approximately 300 pc. Four simulations make it to the atomic cooling track (green, black, cyan and red lines) while two simulations (blue and magenta) end up on the molecular cooling track. The simulations that ‘failed’ to achieve an isothermal collapse both show marked increases in H₂ towards the very centre of the halo. The golden dashed line in the bottom right panel shows the level of the uniform background radiation used (set at $J_{\text{BG}} = 150J_{21}$ in this illustrative case). Details for each simulation are given in Supplementary Table 1.

probe backgrounds with $J_{\text{BG}} < 100J_{21}$. As a result of these constraints, regions at the top and to the right in Fig. 3 are excluded. Ram pressure stripping will affect haloes that get too close to one another, and thus the bottom section of the graph is excluded for this reason^{15,19} (see Methods for further discussion). This leaves a (green) region in the left centre of the figure that allows for the formation of DCBHs. The crosses are results from the suite of simulations (see Supplementary Table 1). The simulations that formed an atomic cooling halo are labelled with green crosses and are in this region. Orange crosses indicate simulations that formed an atomic cooling halo but for which the background was set to $J_{\text{BG}} = 200J_{21}$ (the maximum level in our study). This is a strong background and may be beyond even the most clustered regions²⁰. Red crosses indicate simulations in which a non-isothermal collapse was observed because of H₂ cooling and in which a DCBH is therefore unlikely to form.

Three regions of particular interest are marked on the plot as A, B and C. Region A is where most of our simulation results cluster. Short ($T_{\text{sync}} \lesssim 2$ Myr) synchronization times combined with close separations almost always resulted in an ACH. Region B is outside our ‘synchronized halo zone’ (SHZ), and here most of the simulations show that the flux received by the secondary is too small and an atomic cooling halo does not form. Nonetheless, four atomic cooling haloes appear outside the SHZ. Two of the green crosses are from runs with a background of $J_{\text{BG}} = 150J_{21}$ while one is from a run with a background of $J_{\text{BG}} = 100J_{21}$. The orange cross had a background of $J_{\text{BG}} = 200J_{21}$. Similarly, two molecular collapses are seen inside the SHZ, showing that fluctuations limit the accuracy of the SHZ boundary at approximately the 10% level. Region C indicates the tip of the SHZ. As the synchronization time becomes longer, the risk of photoionization and/or metal pollution becomes larger. Synchronization limits will vary depending on the exact environmental conditions, the luminosity, the spectrum and the distance to the primary halo; we therefore only probed the limits of the separation and the limits of the synchronization (that is, how short or extended the synchronization time can be).

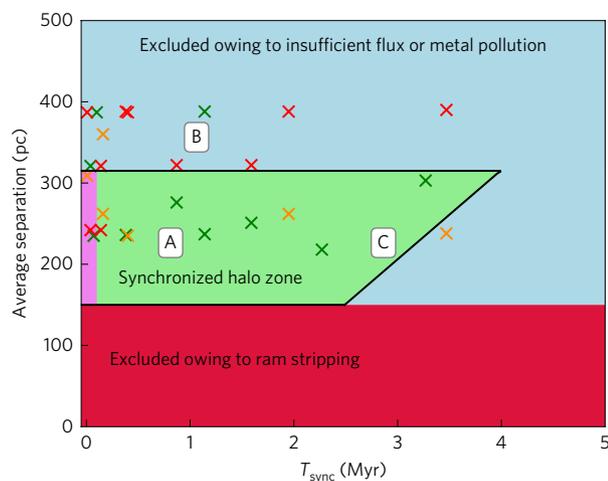


Figure 3 | The synchronized halo zone. This figure sums up the results of the suite of simulations conducted during this study. Plotted on the x-axis is the synchronization time, and on the y-axis is the average separation between the source and the target halo. As the synchronization time increases, the likelihood of attaining a DCBH diminishes as the risk of metal pollution and photo-evaporation increases. Green crosses indicate cases in which an atomic cooling halo was achieved; orange crosses indicate cases in which an atomic cooling halo was achieved but the background was at our highest value of $J_{\text{BG}} = 200J_{21}$. Red crosses indicate simulations in which a molecular core formed in the centre of the collapsing halo. The green region in the centre is the region in which we find atomic haloes most likely to form. We also tentatively indicate (purple region on the extreme left) a ‘point of no return’ at which we expect the formation of PopIII stars to be unstoppable, independent of the intensity of the nearby star-burst.

Our simulations do not include a star-formation prescription or the impact from subsequent supernova events. However, these omissions have no bearing on our results, as we include realistic stellar outputs in our radiative transfer calculation, while supernova feedback timescales are typically longer than the synchronization timescales found in this study.

A related issue concerning metal pollution is the possible impact of the background galaxies, which we have assumed provide the background radiation field. The metal pollution radius scales approximately as $r_s(M, t) \propto M_{\star}^{1/5} t^{2/5}$, where $r_s(M, t)$ is the metal spread radius, M_{\star} is the stellar mass of the galaxy and t is the time²¹. If we assume that the radiation field is created by a collection of five galaxies of similar mass to our primary galaxy, then the metallicity field spreads as $r_{\text{kpc},s} = 3 \times 10^{-2} (5 \times 10^5 M_{\odot})^{1/5} n^{-1/5} t_6^{2/5}$ (see Equation 5 in ref. 21). This leads to a metal spread of just over 2 kpc after 30 Myr under optimistic assumptions. This suggests, for background sources separated by a distance greater than 2 kpc from the primary and secondary haloes, that metal pollution is not a concern at the redshifts explored here. Our use of a massless radiation particle (see Methods) to act as a source of the nearby radiation flux also means that we neglect the impact of tidal disruption between the neighbouring galaxies. However, even in these circumstances, tidal disruption can be successfully avoided¹⁹.

Forming seeds for the supermassive black holes (SMBHs) that we observe at the centres of the most massive galaxies and as very-high-redshift quasars is an outstanding problem in modern astrophysics. We show here that the combination of a relatively mild ‘local’ radiation background field due to the clustering of early galaxies plus a nearby star-burst event is the perfect trigger for the creation of such an atomic cooling halo. The local background serves to delay population III (PopIII) star formation, allowing a sufficiently massive halo to develop. A nearby (synchronized) star-burst can then irradiate the now-massive halo with a flux greater than

the critical Lyman–Werner flux, pushing the collapsing halo onto the atomic cooling track while avoiding the deleterious effects of photo-evaporation or metal pollution. We find that a ‘synchronized halo zone’ exists where the separation between the neighbouring haloes is between approximately 200 pc and 300 pc and the synchronization time between the evolution of the neighbouring haloes is less than approximately 4 Myr (‘on’ time < 10 Myr). Furthermore, we find that the mass inflow rates onto the central object are greater than $0.1 M_{\odot} \text{ yr}^{-1}$ over several decades in radius.

Close halo pairs with tight synchronization times should easily fulfil the number density requirements of SMBHs at high redshift^{11,15,20} (see Methods for further details and a calculation of expected number densities) and may play an important role in the formation of all SMBHs. Upcoming observations with the James Webb Space Telescope will search for DCBH candidates up to redshifts of $z \approx 25$ and may be able to identify them from their unique spectral signatures^{22,23}. A detection of a DCBH candidate together with star-forming galaxies in close proximity would validate the synchronization mechanism.

Methods

Simulation set-up. We use the adaptive mesh refinement code Enzo in this study. Enzo uses an N -body adaptive particle-mesh solver to follow the dark matter dynamics. It solves the hydrodynamics equations using the second-order accurate piecewise parabolic method, while an HLLC Riemann solver ensures accurate shock capturing with minimal viscosity. Rather than using the internal chemistry solver, we use a modified version of the Grackle chemistry library that has been updated with the latest rates for modelling collapse in the face of radiation backgrounds^{24,25}. The chemical network includes 33 separate chemical reactions (see Table 1 in previous work¹⁷) from 10 species: H, H⁺, He, He⁺, He²⁺, e⁻, H₂, H₂⁺, H⁻ and HeH⁺.

The initial conditions were taken from previous work²⁶ and were generated with the MUSIC initial conditions generator²⁷. Haloes were initially located by running a large suite of dark-matter-only simulations in a cosmological context. Realizations were then selected by choosing simulations in which massive dark matter haloes formed relatively early and thus represented high ‘sigma’ peaks in the primordial density field. Simulations were then rerun with gas dynamics included. All simulations were run within a box of size $2 h^{-1}$ Mpc (comoving), the root grid size was 256^3 , and we used three levels of nested grids. Within the most refined nested grid, the dark matter resolution was $M_{\text{DM}} \approx 103 M_{\odot}$. To increase the dark matter resolution further, we split the dark matter particles²⁸ at a redshift of $z = 40$, well before the onset of the collapse. This has no adverse effect on the dynamics of the collapse but is a necessary step in high-resolution simulations²⁶. The baryon resolution is set by the size of the highest-resolution cells within the grid. Grids are refined in Enzo whenever user-defined criteria are breached. We allow refinement of grid cells based on three physical measurements: (1) the dark matter particle overdensity, (2) the baryon overdensity and (3) the Jeans length. The first two criteria introduce additional meshes when the overdensity ($\delta\rho/\rho_{\text{mean}}$) of a grid cell with respect to the mean density exceeds 8.0 for baryons and/or dark matter. Furthermore, we set the ‘MinimumMassForRefinementExponent’ parameter to -0.1 , making the simulation super-Lagrangian, and therefore reducing the threshold for refinement as higher densities are reached. For the final criteria, we set the number of cells per Jeans length to be 16 in these simulations. We set the maximum allowed refinement level to 18. This means that we reach a maximum grid resolution of $\Delta x \approx 2 \times 10^{-3}$ pc (physical distance at $z = 25$).

Choosing a background radiation field. The vast majority of studies undertaken to investigate the destruction of H₂ by an external radiation field have used a uniform background radiation field to demonstrate the viability of the mechanism. The general consensus has been that a critical Lyman–Werner intensity of $J \gtrsim 1,000 J_{21}$ is needed^{29–32} for a black-body background with an effective temperature of 10^5 K (a background intensity about an order of magnitude lower would be sufficient for low-mass stars with an effective temperature of $T = 10^4 \text{ K}^{29}$, but the stellar mass required is then significantly higher^{33,34}). However, achieving a background of $J_{\text{BG}} \gtrsim 1,000 J_{21}$ is very unlikely at the very high redshifts of interest ($z \gtrsim 25$)²⁰. Instead, what is much more likely to create the required flux is a mild background ($J_{\text{BG}} \approx 100–200 J_{21}$) augmented by a local burst of Lyman–Werner radiation from one or more nearby sources^{11,15,21}. The synchronized pairs scenario requires a mechanism to delay the formation of PopIII stars so that neighbouring haloes can build sufficient mass to cross the atomic cooling threshold in near-synchronization. In this study, we impose an external background field to facilitate this delay^{35,36}. For the halo studied in this work, we found, in agreement with other studies in the literature, that a background radiation field of $\sim 1,000 J_{21}$ is required to form an atomic halo with

a black-body effective temperature of 30,000 K (when only a global background is considered). In Supplementary Fig. 1, we have plotted the temperature as a function of radius for six different background fields. In this study, we examine background radiation fields, J_{BG} , with values of $100J_{21}$, $150J_{21}$, and $200J_{21}$. In all cases, we initiate our background radiation field at $z = 35$. In principle, the star-burst radiation from the primary galaxy could compensate for a lower background field. This, however, requires the star-burst to shine for longer, increasing the risk of metal pollution or photo-evaporation.

Radiation source. To study the effect that a local radiation source (the primary halo) can have on the ‘target’ halo (the secondary halo), we use a massless ‘radiation particle’ for which we can vary the emission intensity and source separation. The radiation particle acts as a source of radiation, and the radiation is propagated by Enzo’s radiative transfer scheme¹⁰, which traces individual photon packets through the AMR mesh by using 64-bit HEALPix algorithms^{37,38}. The ray tracing scheme allows us to follow photons in the infrared, Lyman–Werner and hydrogen ionization range. The exact energy levels and relative strength of each energy bin used are given in Supplementary Table 2. By using the ray tracing scheme, we are able to accurately track both the column density along each ray and the level of photodissociation of different ions along the ray path. The shielding effects of different species are accounted for by the exact knowledge of the column density (and therefore the optical depth). Self-shielding of H_2 is accounted for using the prescription given in previous work³⁹.

We control the physical distance between the primary and the secondary haloes by first running a test simulation where the radiation particle is placed at a distance of approximately 200 pc from the source. The simulation is then run, and we calculate the position of the point of maximum density averaged over multiple outputs. By doing this, we know in advance how the centre of mass of the system will change as the simulation proceeds (as the initial source position is changed, the centre of mass will change, but our numerical experiments showed that the change was only at the 10% level). We then choose a vector with the centre of mass as the origin. By placing the radiation particle at set distances along this given vector (we use the angular momentum vector for convenience), we find points at which the distance to the secondary halo remains approximately constant, although some variation is expected. This mimics a scenario in which the secondary halo and primary halo orbit one another.

The radiation spectrum emitted by the primary halo is modelled to be a partially metal-enriched galaxy. We use a luminosity of $\sim 1.2 \times 10^{52}$ photons per second (above the H^- photo-detachment energy of 0.76 eV; see Fig. 2 in ref. 17). We assume a stellar mass of $M_\star \approx 10^5 M_\odot$ at $z = 20$, consistent with the largest galaxies prior to reionization in the Renaissance Simulations⁴⁰. We then calculate its spectrum using the Bruzual–Charlot models⁴¹ with a metallicity of $10^{-2} Z_\odot$ and compute the photon luminosity from it. The spectrum does not include emission from the nebular component and is solely due to stellar emission. This spectrum is virtually identical to a PopIII source with a cluster of $9M_\odot$ stars⁴². However, a cluster of 10^4 PopIII stars is a challenging structure to produce even under extreme conditions⁴⁰ and hence we opt for the metal-enriched population. The details of the spectrum (and a comparable PopIII spectrum for comparison) are given in Supplementary Table 2.

Ram pressure stripping and tidal disruption. The small separation of the primary and secondary haloes implies that they are sub-haloes of a larger parent halo. As they move through the ‘intra-cluster medium’ of the parent halo, they are subject to ram pressure stripping. The ram pressure due to this movement is:

$$P_{\text{ram}} = \frac{1}{2} \rho_m v^2 \quad (1)$$

To ensure that this ram pressure does not unbind the core of the secondary galaxy, P_{ram} must obey^{43–45}

$$P_{\text{ram}} \leq \frac{\alpha G M_{\text{tot}}(R_{\text{core}}) \rho_{\text{gas}}(R_{\text{core}})}{R_{\text{core}}} \quad (2)$$

where ρ_m is the gas density that the core passes through, v is the relative orbital velocity of the protogalaxy, α is a variable of order unity which depends on the specifics of the dark matter and gas profiles, and $M_{\text{tot}}(R_{\text{core}})$ and $\rho_{\text{gas}}(R_{\text{core}})$ are the total mass and gas density within R_{core} , respectively. For the secondary halo studied here, we define the core of the halo as the radius at which the gas mass dominates over the dark matter mass, giving us a value of $R_{\text{core}} \approx 10$ pc. Using the virial mass of the secondary halo, we assume a relative orbital velocity of 20 km s^{-1} . We then use the gas density at an orbital distance of 100 pc ($\rho_m(R = 100 \text{ pc}) \approx 1 \times 10^{-23} \text{ g cm}^{-3}$) from the centre of the secondary halo to get

$$P_{\text{ram}} \approx 2 \times 10^{-8} \text{ g cm}^{-1} \text{ s}^{-2} \quad (3)$$

while the right-hand side of equation (2) gives a value of

$$\frac{\alpha G M_{\text{tot}}(R_{\text{core}}) \rho_{\text{gas}}(R_{\text{core}})}{R_{\text{core}}} \approx 2 \times 10^{-8} \text{ g cm}^{-1} \text{ s}^{-2} \quad (4)$$

at a radius of 10 pc. Inside this radius, the binding energy of the halo increases approximately as R^{-2} , and so, within 10 pc, ram pressure stripping will be unable to unbind the gas. To be conservative, we always constrain the radiation particle to be at a distance greater than 150 pc from the secondary halo centre. However, while the core will remain intact, some mass loss from the outer parts of the protogalaxy is inevitable. Given that the secondary and primary haloes are eventually expected to merge, the total mass available for the accreting DCBH is unaffected.

A recent study¹⁹ examined the formation of DCBHs in a large cosmological simulation. The focus of its simulations was to determine the impact of tidal disruption and ram pressure on the formation of DCBHs. In a study of 42 DC candidates, the authors found that two candidates successfully resulted in the rapid formation of stars within an ACH. One of the two successful candidates was due to a scenario similar to the synchronized scenario investigated here. However, their simulations were unable to accurately track the radiation coming from the primary halo and were not focused on this scenario. In their case, they found that the primary halo exceeded the atomic cooling threshold 20–30 Myrs before the secondary halo, with the Lyman–Werner intensity exceeding the critical value. This suggests that we are conservative with our 10-Myr limit on the primary halo ‘on’ time. However, similar to our study, the previous study¹⁹ does not include the effects of metal pollution (see below), which could limit the time for which the star-burst can be active before the secondary must collapse under atomic cooling. We also note that the authors did not include the impact of photoionization in their hydrodynamics simulations¹⁹, whereas we do so here, and are therefore able to show that photo-evaporation is avoided.

Metal pollution, photo-evaporation and the maximum irradiation time. Metal pollution, photo-evaporation or the natural end of the star-burst in the primary galaxy will ultimately limit the prospects for achieving a DCBH. A PopIII source model with individual stellar masses of $M_\star \approx 9M_\odot$ will have an expected lifetime⁴² of $T_\star \approx 20$ Myr. After this time, the stars would explode as supernovae and presumably pollute the secondary halo within a few million years²¹. Our galaxy model includes lower-mass stars with longer lifetimes but with a more distributed IMF, and so also stars with masses in excess of $9M_\odot$. To be conservative, we assume that in cases where the primary halo must be ‘on’ for greater than 10 Myr, a DCBH does not form. We note that we see no evidence of photo-evaporation of the secondary halo even when the secondary is irradiated for 22 Myr (the longest ‘on’ time encountered during our numerical experimentation), but since we do not include the impact of metal production and pollution, we limit the maximum irradiation time to 10 Myr.

Mass inflow rates. Large mass inflow rates are one of the prerequisites for forming a supermassive star. For the case of (super-)massive stars, the gravitational contraction timescale is much shorter than the Kelvin–Helmholtz timescale, meaning that the star begins nuclear burning before it has finished accreting, and in fact will continue to accrete (subject to the correct environmental conditions) throughout its lifetime. Its final mass then becomes the important quantity.

In Fig. 4, we have plotted the mass inflow rates as found for the same subset of simulations shown in Fig. 2. The mass inflow rates are calculated in spherical shells around the central density according to the equation

$$\dot{M}(t) = 4\pi R^2 \rho(R) V(R) \quad (5)$$

where $\dot{M}(t)$ is the mass inflow rate, R is the radius, ρ is the density and $V(R)$ is the radial velocity at R . We find that the mass inflow rates peak at approximately $0.75 M_\odot \text{ yr}^{-1}$ at $R \approx 5 \times 10^{-2}$ pc. Most importantly, mass inflow rates greater than $0.1 M_\odot \text{ yr}^{-1}$ are sustained over several decades in radius. Furthermore, we find that the more intense the flux (that is, the closer the separation), the higher the mass inflow rates. This feature is likely to be due to the higher Lyman–Werner flux experienced at this radius, which should in turn lead to high gas temperatures owing to the higher dissociation rates of H_2 .

The classical assumption regarding the formation of supermassive stars is that accretion rates greater than $0.03 M_\odot \text{ yr}^{-1}$ are required^{46,47}. At this mass accretion rate, the protostellar evolution changes completely. The stellar envelope swells greatly in radius, reaching up to 100 AU. The effective stellar temperature drops to close to 5,000 K, meaning that radiative feedback from the protostar becomes ineffective at preventing continued accretion. In a somewhat complementary scenario, a ‘quasi-star’ may be born when high accretion rates ($\dot{M}(t) \gg 0.14 M_\odot \text{ yr}^{-1}$)⁴⁸ onto an already existing supermassive star result in the collapse of the hydrogen core into a stellar-mass black hole. The highly optically thick gas that keeps falling on the black hole, bringing with it angular momentum, results in an accretion disk forming around the central black hole. The energy feedback inflates the innermost part of the inflow, resulting in a ‘quasi-star’. The mass-loss rates from these two types of objects are an area of active research, with possible negative consequences in the case of quasi-stars⁴⁹ (higher mass loss rates that may lead to less massive black holes) that may be absent in the case of more ‘normal’ supermassive stars⁵⁰. Either way, the accretion rates observed in our simulations satisfy the basic requirements that accretion rates of $\gtrsim 0.1 M_\odot \text{ yr}^{-1}$ are found over several decades in radius.

It should be noted that below a radius of approximately 3×10^{-2} pc, the inflow rates shown in Fig. 4 fall. This is a purely numerical effect created by both a diminishing resolution below that scale (the Jeans length of the gas at this

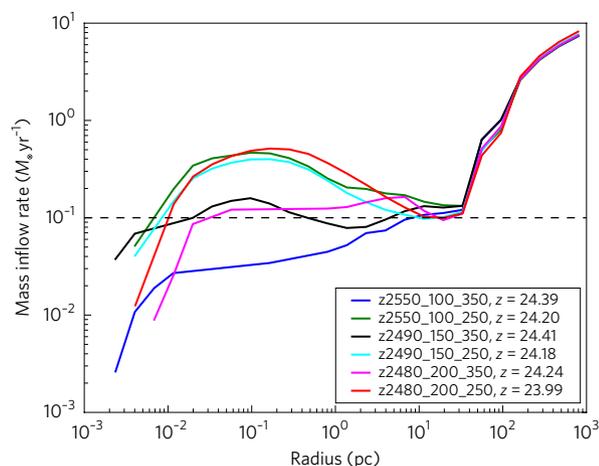


Figure 4 | Mass inflow rates. Mass inflow rates into the centre of collapsing halo as a function of distance from the central object for the same subset of simulations as shown in Fig. 2. The mass inflow rates are calculated by determining the flux in spherical shells surrounding the central object. Focusing on the core of the collapsing halo where $R < 10$ pc, the mass inflow rates are typically much greater than $0.1 M_{\odot} \text{ yr}^{-1}$, with values peaking at around $0.75 M_{\odot} \text{ yr}^{-1}$ at $R \approx 5 \times 10^{-2}$ pc. Interestingly, the highest rates of mass inflow are seen for the simulations in which the flux is largest (the smallest separations). Finally, accretion rates greater than $0.1 M_{\odot} \text{ yr}^{-1}$ are sustained over several decades in radius. Below a scale of $R \approx 3 \times 10^{-2}$ pc, the inflow rates fall because of the effects of limited resolution. Details for each simulation are given in Supplementary Table 1.

temperature and density is approximately 0.1 pc) and the fact that we stop the calculation once the gas reaches our maximum refinement level. We are therefore not evolving the gas at this scale over any significant fraction of its dynamical time. Further exploration of the inflow rate at this scale would require the adoption of sink particles.

Finally, it is not certain that if a monolithic inflow of $> 0.1 M_{\odot} \text{ yr}^{-1}$ is attained and then sustained over an extended period of up to 1 Myr (possibly up to 10 Myr as required for quasi-stars), a single central object will form. Fragmentation may, in this case, still occur within the collapsing object⁵¹ or else within a self-gravitating disk around the central object⁵². However, in either case the fragments are likely to form a dense cluster and ultimately a DCBH⁵³.

Expected and required Lyman–Werner background. Our simulations indicate that a background Lyman–Werner intensity of $J_{\text{BG}} \gtrsim 100 J_{21}$ is required for DCBH formation in the synchronized haloes scenario. Determining the precise abundance of DCBHs formed as a result of this requirement is beyond the scope of the present work; however, we argue here that such a high flux can plausibly be achieved and can potentially result in a DCBH number density large enough to explain observations of SMBHs at $z \gtrsim 6$.

The intensity of the Lyman–Werner background at $z \gtrsim 10$ is highly uncertain. Some theoretical estimates^{54–56} have found mean values in the range $(0.1–1) J_{21}$ at $z = 25$, and values roughly an order of magnitude higher by $z = 10$. Even though the $z = 25$ values are two to three orders of magnitudes below the requirement found in our simulations, such a high background may still be possible in overdense environments. Previous simulations⁵⁰ show that, even in a relatively small box ($35 \text{ Mpc } h^{-1}$), there are regions with Lyman–Werner flux several orders of magnitude higher than the minimum value (see Figure 11 of that paper²⁰). Given the extreme rarity of $z \approx 6$ SMBHs, with $M_{\text{BH}} > 10^9 M_{\odot}$ ($\sim 10^{-9} \text{ Mpc}^{-3}$), these objects are likely to have formed in highly biased regions with many more dark matter haloes than the mean number density. Thus, it seems plausible that nearby clustered sources could provide the necessary flux. This could be accomplished by, for example, about five sources similar to our primary halo within 2 kpc, or one source that is five times as bright.

We also note that a radiation background is not the only mechanism that could potentially delay PopIII star formation. The relative streaming velocities of baryons and dark matter⁵⁷ can significantly inhibit the collapse of gas in mini-haloes at very high redshift^{58–60} and could reduce the required intensity of the background field by two to three orders of magnitude^{5,61}. In particular, streaming velocities can suppress molecular cooling all the way up to the ACH limit⁶¹. Likewise, haloes assembling unusually rapidly through successive mergers on timescales shorter than the H_2 cooling timescale can avoid star formation at the mini-halo stage⁶². We do not include all of these effects in this study but note that they reduce the need

for a strong background. Nonetheless, an additional intense nearby source is always required to prevent H_2 forming in the centre of the collapsing halo⁶³.

Even if the abundance of DCBHs formed in the synchronized pair channel at $z \approx 25$ is very low, the number density at lower redshift could potentially explain the abundance of high-redshift supermassive black holes. Assuming separations similar to those in our simulations, previous work⁶³ estimated the abundance of synchronized pair DCBHs formed between $z = 10$ and $z = 11$ to be 0.0003 Mpc^{-3} . If the probability distribution function of the Lyman–Werner background for these pairs matched that from ref.²⁰ (see their Figure 11), combining this with the $J_{\text{BG}} \gtrsim 100 J_{21}$ background requirement would not reduce the number density below $\sim 10^{-9} \text{ Mpc}^{-3}$ (which is similar to the observed number density of high-redshift quasars). This rough estimate is only a lower limit, owing to the limited box sizes used in both calculations cited above and to the fact that other large-scale effects, as discussed, are neglected.

Data availability. The numerical experiments presented in this work were run with a fork of the Enzo code available from https://bitbucket.org/john_regan/enzo-3.0-rp, in particular change set d11330f. This altered version of Enzo also requires an altered version of the Grackle cooling library, available from https://bitbucket.org/john_regan/grackle-cverson, particularly change set d8df240.

Received 23 September 2016; accepted 7 February 2017;
published 13 March 2017

References

- Fan, X., Carilli, C. L. & Keating, B. Observational constraints on cosmic reionization. *Annu. Rev. Astron. Astrophys.* **44**, 415–462 (2006).
- Mortlock, D. J. *et al.* A luminous quasar at a redshift of $z = 7.085$. *Nature* **474**, 616–619 (2011).
- Wu, X.-B. *et al.* An ultraluminous quasar with a twelve-billion-solar-mass black hole at redshift 6.30. *Nature* **518**, 512–515 (2015).
- Loeb, A. & Rasio, F. A. Collapse of primordial gas clouds and the formation of quasar black holes. *Astrophys. J.* **432**, 52–61 (1994).
- Begelman, M. C., Volonteri, M. & Rees, M. J. Formation of supermassive black holes by direct collapse in pre-galactic haloes. *Mon. Not. R. Astron. Soc.* **370**, 289–298 (2006).
- Omukai, K. Primordial star formation under far-ultraviolet radiation. *Astrophys. J.* **546**, 635–651 (2001).
- Oh, S. P. & Haiman, Z. Second-generation objects in the universe: radiative cooling and collapse of halos with virial temperatures above 10^4 K . *Astrophys. J.* **569**, 558–572 (2002).
- Bryan, G. L. *et al.* Enzo: an adaptive mesh refinement code for astrophysics. *Astrophys. J. Suppl.* **211**, 19 (2014).
- Smith, B. D. *et al.* Grackle: a chemistry and cooling library for astrophysics. Preprint at <https://arxiv.org/abs/1610.09591> (2016).
- Wise, J. H. & Abel, T. Enzo+Moray: radiation hydrodynamics adaptive mesh refinement simulations with adaptive ray tracing. *Mon. Not. R. Astron. Soc.* **414**, 3458–3491 (2011).
- Dijkstra, M., Haiman, Z., Mesinger, A. & Wyithe, J. S. B. Fluctuations in the high-redshift Lyman–Werner background: close halo pairs as the origin of supermassive black holes. *Mon. Not. R. Astron. Soc.* **391**, 1961–1972 (2008).
- Yue, B., Ferrara, A., Salvaterra, R., Xu, Y. & Chen, X. The brief era of direct collapse black hole formation. *Mon. Not. R. Astron. Soc.* **440**, 1263–1273 (2014).
- Agarwal, B., Davis, A. J., Khochfar, S., Natarajan, P. & Dunlop, J. S. Unravelling obese black holes in the first galaxies. *Mon. Not. R. Astron. Soc.* **432**, 3438–3444 (2013).
- Agarwal, B., Dalla Vecchia, C., Johnson, J. L., Khochfar, S. & Paardekoooper, J.-P. The First Billion Years project: birthplaces of direct collapse black holes. *Mon. Not. R. Astron. Soc.* **443**, 648–657 (2014).
- Visbal, E., Haiman, Z. & Bryan, G. L. Direct collapse black hole formation from synchronized pairs of atomic cooling haloes. *Mon. Not. R. Astron. Soc.* **445**, 1056–1063 (2014).
- Smith, B. D., Wise, J. H., O’Shea, B. W., Norman, M. L. & Khochfar, S. The first Population II stars formed in externally enriched mini-haloes. *Mon. Not. R. Astron. Soc.* **452**, 2822–2836 (2015).
- Regan, J. A., Johansson, P. H. & Wise, J. H. Forming super-massive black hole seeds under the influence of a nearby anisotropic multi-frequency source. *Mon. Not. R. Astron. Soc.* **459**, 3377–3394 (2016).
- Xu, H., Wise, J. H. & Norman, M. L. Population III stars and remnants in high-redshift galaxies. *Astrophys. J.* **773**, 83 (2013).
- Chon, S., Hirano, S., Hosokawa, T. & Yoshida, N. Cosmological simulations of early black hole formation: halo mergers, tidal disruption, and the conditions for direct collapse. *Astrophys. J.* **832**, 134 (2016).
- Ahn, K., Shapiro, P. R., Iliev, I. T., Mellema, G. & Pen, U.-L. The inhomogeneous background of H_2 -dissociating radiation during cosmic reionization. *Astrophys. J.* **695**, 1430–1445 (2009).
- Dijkstra, M., Ferrara, A. & Mesinger, A. Feedback-regulated supermassive black hole seed formation. *Mon. Not. R. Astron. Soc.* **442**, 2036–2047 (2014).

22. Pacucci, F., Ferrara, A., Volonteri, M. & Dubus, G. Shining in the dark: the spectral evolution of the first black holes. *Mon. Not. R. Astron. Soc.* **454**, 3771–3777 (2015).
23. Natarajan, P. *et al.* Unveiling the first black holes with JWST: multi-wavelength spectral predictions. Preprint at <http://arxiv.org/abs/1610.05312> (2016).
24. Glover, S. C. O. Simulating the formation of massive seed black holes in the early Universe. I: An improved chemical model. *Mon. Not. R. Astron. Soc.* **451**, 2082–2096 (2015).
25. Glover, S. C. O. Simulating the formation of massive seed black holes in the early Universe. II: Impact of rate coefficient uncertainties. *Mon. Not. R. Astron. Soc.* **453**, 2901–2918 (2015).
26. Regan, J. A., Johansson, P. H. & Wise, J. H. The effect of dark matter resolution on the collapse of baryons in high-redshift numerical simulations. *Mon. Not. R. Astron. Soc.* **449**, 3766–3779 (2015).
27. Hahn, O. & Abel, T. Multi-scale initial conditions for cosmological simulations. *Mon. Not. R. Astron. Soc.* **415**, 2101–2121 (2011).
28. Kitsionas, S. & Whitworth, A. P. Smoothed particle hydrodynamics with particle splitting, applied to self-gravitating collapse. *Mon. Not. R. Astron. Soc.* **330**, 129–136 (2002).
29. Shang, C., Bryan, G. L. & Haiman, Z. Supermassive black hole formation by direct collapse: keeping protogalactic gas H_2 free in dark matter haloes with virial temperatures $T_{vir} \gtrsim 10^4 K$. *Mon. Not. R. Astron. Soc.* **402**, 1249–1262 (2010).
30. Latif, M. A. *et al.* A UV flux constraint on the formation of direct collapse black holes. *Mon. Not. R. Astron. Soc.* **443**, 1979–1987 (2014).
31. Latif, M. A., Schleicher, D. R. G., Bovino, S., Grassi, T. & Spaans, M. The formation of massive primordial stars in the presence of moderate UV backgrounds. *Astrophys. J.* **792**, 78 (2014).
32. Agarwal, B. & Khochfar, S. Revised rate coefficients for H_2 and H^- destruction by realistic stellar spectra. *Mon. Not. R. Astron. Soc.* **446**, 160–168 (2015).
33. Wolcott-Green, J. & Haiman, Z. Feedback from the infrared background in the early Universe. *Mon. Not. R. Astron. Soc.* **425**, L51–L55 (2012).
34. Wolcott-Green, J., Haiman, Z. & Bryan, G. L. Beyond T_{crit} : a critical curve for suppression of H_2 -cooling in protogalaxies. Preprint at <http://arxiv.org/abs/1609.02142> (2016).
35. Machacek, M. E., Bryan, G. L. & Abel, T. Simulations of pregalactic structure formation with radiative feedback. *Astrophys. J.* **548**, 509–521 (2001).
36. O’Shea, B. W. & Norman, M. L. Population III star formation in a Λ CDM universe. II: Effects of a photodissociating background. *Astrophys. J.* **673**, 14–33 (2008).
37. Abel, T. & Wandelt, B. D. Adaptive ray tracing for radiative transfer around point sources. *Mon. Not. R. Astron. Soc.* **330**, L53–L56 (2002).
38. Górski, K. M. *et al.* HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys. J.* **622**, 759–771 (2005).
39. Wolcott-Green, J., Haiman, Z. & Bryan, G. L. Photodissociation of H_2 in protogalaxies: modelling self-shielding in three-dimensional simulations. *Mon. Not. R. Astron. Soc.* **418**, 838–852 (2011).
40. Chen, P., Wise, J. H., Norman, M. L., Xu, H. & O’Shea, B. W. Scaling relations for galaxies prior to reionization. *Astrophys. J.* **795**, 144 (2014).
41. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
42. Schaerer, D. On the properties of massive Population III stars and metal-free stellar populations. *Astron. Astrophys.* **382**, 28–42 (2002).
43. Gunn, J. E. & Gott, J. R. III. On the infall of matter into clusters of galaxies and some effects on their evolution. *Astrophys. J.* **176**, 1–19 (1972).
44. Gisler, G. R. The fate of gas in elliptical galaxies and the density evolution of radio sources. *Astron. Astrophys.* **51**, 137–150 (1976).
45. McCarthy, I. G. *et al.* Ram pressure stripping the hot gaseous haloes of galaxies in groups and clusters. *Mon. Not. R. Astron. Soc.* **383**, 593–605 (2008).
46. Hosokawa, T., Omukai, K. & Yorke, H. W. Rapidly accreting supergiant protostars: embryos of supermassive black holes? *Astrophys. J.* **756**, 93 (2012).
47. Hosokawa, T., Yorke, H. W., Inayoshi, K., Omukai, K. & Yoshida, N. Formation of primordial supermassive stars by rapid mass accretion. *Astrophys. J.* **778**, 178 (2013).
48. Schleicher, D. R. G., Palla, F., Ferrara, A., Galli, D. & Latif, M. Massive black hole factories: Supermassive and quasi-star formation in primordial halos. *Astron. Astrophys.* **558**, A59 (2013).
49. Fiacconi, D. & Rossi, E. M. Bright vigorous winds as signposts of supermassive black hole birth. *Mon. Not. R. Astron. Soc.* **455**, 2–16 (2016).
50. Nakauchi, D., Hosokawa, T., Omukai, K., Saio, H. & Nomoto, K. Do stellar winds prevent the formation of supermassive stars by accretion? *Mon. Not. R. Astron. Soc.* **465**, 5016–5025 (2017).
51. Regan, J. A., Johansson, P. H. & Haehnelt, M. G. Numerical resolution effects on simulations of massive black hole seeds. *Mon. Not. R. Astron. Soc.* **439**, 1160–1175 (2014).
52. Inayoshi, K. & Haiman, Z. Does disc fragmentation prevent the formation of supermassive stars in protogalaxies? *Mon. Not. R. Astron. Soc.* **445**, 1549–1557 (2014).
53. Regan, J. A. & Haehnelt, M. G. Pathways to massive black holes and compact star clusters in pre-galactic dark matter haloes with virial temperatures 10000 K. *Mon. Not. R. Astron. Soc.* **396**, 343–353 (2009).
54. Fialkov, A., Barkana, R., Visbal, E., Tselikhovich, D. & Hirata, C. M. The 21-cm signature of the first stars during the Lyman–Werner feedback era. *Mon. Not. R. Astron. Soc.* **432**, 2909–2916 (2013).
55. Visbal, E., Haiman, Z., Terrazas, B., Bryan, G. L. & Barkana, R. High-redshift star formation in a time-dependent Lyman–Werner background. *Mon. Not. R. Astron. Soc.* **445**, 107–114 (2014).
56. Visbal, E., Haiman, Z. & Bryan, G. L. Limits on Population III star formation in minihaloes implied by Planck. *Mon. Not. R. Astron. Soc.* **453**, 4456–4466 (2015).
57. Tselikhovich, D. & Hirata, C. Relative velocity of dark matter and baryonic fluids and the formation of the first structures. *Phys. Rev. D* **82**, 083520 (2010).
58. Stacy, A., Bromm, V. & Loeb, A. Effect of streaming motion of baryons relative to dark matter on the formation of the first stars. *Astrophys. J. Lett.* **730**, L1 (2011).
59. Fialkov, A., Barkana, R., Tselikhovich, D. & Hirata, C. M. Impact of the relative motion between the dark matter and baryons on the first stars: semi-analytical modelling. *Mon. Not. R. Astron. Soc.* **424**, 1335–1345 (2012).
60. Naoz, S., Yoshida, N. & Gnedin, N. Y. Simulations of early baryonic structure formation with stream velocity. II: The gas fraction. *Astrophys. J.* **763**, 27 (2013).
61. Tanaka, T. L. & Li, M. The formation of massive black holes in $z \sim 30$ dark matter haloes with large baryonic streaming velocities. *Mon. Not. R. Astron. Soc.* **439**, 1092–1100 (2014).
62. Fernandez, R., Bryan, G. L., Haiman, Z. & Li, M. H_2 suppression with shocking inflows: testing a pathway for supermassive black hole formation. *Mon. Not. R. Astron. Soc.* **439**, 3798–3807 (2014).
63. Visbal, E., Haiman, Z. & Bryan, G. L. A no-go theorem for direct collapse black holes without a strong ultraviolet background. *Mon. Not. R. Astron. Soc.* **442**, L100–L104 (2014).

Acknowledgements

This work was supported by the Science and Technology Facilities Council (grant numbers ST/L00075X/1 and RF040365) and by NASA grant NNX15AB19G (to Z.H.). J.R. acknowledges support from the EU Commission via the Marie Skłodowska-Curie Grant SMARTSTARS, grant number 699941. J.W. is supported by National Science Foundation grants AST-1333360 and AST-1614333, and by Hubble theory grants HST-AR-13895 and HST-AR-14326. P.H.J. acknowledges the support of the Academy of Finland grant 1274931. G.B. acknowledges financial support from NASA grant NNX15AB20G and NSF grant AST-1312888. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1 and STFC DiRAC Operations grant ST/K003267/1, and by Durham University. DiRAC is part of the National E-Infrastructure. Some of the preliminary numerical simulations were also performed on facilities hosted by the CSC-IT Center for Science in Espoo, Finland, which are financed by the Finnish ministry of education. The Flatiron Institute is supported by the Simons Foundation. Z.H. acknowledges support from a Simons Fellowship for Theoretical Physics. The freely available astrophysical analysis code yt and plotting library matplotlib were used to construct numerous plots within this paper. Computations described in this work were performed using the publicly available Enzo code, which is the product of a collaborative effort of many independent scientists from institutions around the world.

Author contributions

J.A.R. modified the publicly available Enzo code and Grackle codes used in this work, ran and analysed the code results, and wrote the initial manuscript. J.A.R., Z.H., J.H.W. and E.V. determined the simulation set-up. The radiation particle model was conceived and designed by J.A.R., P.H.J. and J.H.W. All authors contributed to the interpretation of the results and to the text of the final manuscript.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.A.R.

How to cite this article: Regan, J. A. *et al.* Rapid formation of massive black holes in close proximity to embryonic protogalaxies. *Nat. Astron.* **1**, 0075 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Competing interests

The authors declare no competing financial interests.