The formation of hot gaseous haloes around galaxies

Camila A. Correa,[1,2]⋆ Joop Schaye,[1] J. Stuart B. Wyithe,[2] Alan R. Duffy,[2,3]
Tom Theuns,[4] Robert A. Crain[5] and Richard G. Bower[4]

[1]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[2]*School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia*
[3]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Melbourne, Victoria 3122, Australia*
[4]*Institute for Computational Cosmology, Physics Department, University of Durham, South Road, Durham DH1 3LE, UK*
[5]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*

ABSTRACT

We use a suite of hydrodynamical cosmological simulations from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) project to investigate the formation of hot hydrostatic haloes and their dependence on feedback mechanisms. We find that the appearance of a strong bimodality in the probability density function of the ratio of the radiative cooling and dynamical times for halo gas provides a clear signature of the formation of a hot corona. Haloes of total mass $10^{11.5}$–$10^{12}\,\mathrm{M}_\odot$ develop a hot corona independent of redshift, at least in the interval $z = 0$–4, where the simulation has sufficiently good statistics. We analyse the build-up of the hot gas mass in the halo, M_{hot}, as a function of halo mass and redshift and find that while more energetic galactic wind powered by SNe increases M_{hot}, active galactic nucleus feedback reduces it by ejecting gas from the halo. We also study the thermal properties of gas accreting on to haloes and measure the fraction of shock-heated gas as a function of redshift and halo mass. We develop analytic and semi-analytic approaches to estimate a 'critical halo mass', M_{crit}, for hot halo formation. We find that the mass for which the heating rate produced by accretion shocks equals the radiative cooling rate reproduces the mass above which haloes develop a significant hot atmosphere. This yields a mass estimate of $M_{\mathrm{crit}} \approx 10^{11.7}\,\mathrm{M}_\odot$ at $z = 0$, which agrees with the simulation results. The value of M_{crit} depends more strongly on the cooling rate than on any of the feedback parameters.

Key words: methods: analytical – methods: numerical – galaxies: evolution – galaxies: formation – galaxies: haloes.

1 INTRODUCTION

One of the major goals of modern galaxy formation theory is to understand the physical mechanisms that halt the star formation process, by removing, heating or preventing the infall of cold gas on to the galactic disc. X-ray observations suggest that for haloes hosting massive galaxies the majority of baryonic matter resides not only in the galaxies but also in the halo in the form of virialized hot gas (e.g. Lin, Mohr & Stanford 2003; Crain et al. 2010; Anderson & Bregman 2011). This work investigates the formation of the hot gaseous corona (also referred to as 'hot halo' or 'hot atmosphere') around galaxies, that may help reduce the rate of infall of gas on to galaxies, and has been suggested to explain the observed galaxy bimodality (Dekel & Birnboim 2006).

The hot gaseous corona is produced as a result of an important heating process, which was initially discussed by Rees & Ostriker (1977), Silk (1977), Binney (1977) and White & Rees (1978), and later in the context of the cold dark matter paradigm by White & Frenk (1991), in an attempt to explain the reduced efficiency of star formation within massive haloes. They proposed that while a dark matter halo relaxes to virial equilibrium, gas falling into it experiences a shock, and determined the cooling time of gas behind the shock. As long as the cooling time is shorter than the dynamical time, the infalling gas cools (inside the current 'cooling radius') and settles on to the galaxy. If, on the other hand, the cooling time exceeds the dynamical time, the gas is not able to radiate away the thermal energy that supports it. Therefore, it adjusts its density and temperature quasi-statically, forming a hot hydrostatic halo atmosphere, pressure supported against gravitational collapse. Over the past decade, the works of Birnboim & Dekel (2003) and Dekel & Birnboim (2006, hereafter DB06) investigated the stability of accretion shocks around galaxies, and concluded that a hot

⋆ E-mail: correa@strw.leidenuniv.nl

atmosphere forms when the compression time of shocked gas is larger than its cooling time, occurring when haloes reach a mass of about $10^{11.7}$ M$_\odot$.

Numerical simulations have shown, however, that cold gas accreting through filaments does not necessarily experience a shock when crossing the virial radius, even if the spherically averaged cooling radius is smaller than the virial radius. Many groups have concluded that there are two modes of gas accretion, named as hot and cold accretion, that are able to coexist in high-mass haloes at high redshift (e.g. Kereš et al. 2005; Dekel & Birnboim 2006; Ocvirk, Pichon & Teyssier 2008; Dekel et al. 2009; Faucher-Giguère, Kereš & Ma 2011; van de Voort et al. 2011; van de Voort & Schaye 2012; Nelson et al. 2013). The hot mode of accretion refers to the accreted gas that shock-heats to the halo virial temperature. The cold mode refers to gas that flows along dark matter filaments and is accreted on to the central galaxy without being shock-heated near the virial radius. It has been found that the cold streams end up being the dominant mode of accretion on to galaxies at high redshift (e.g. Dekel et al. 2009). However, it has also been found that most of the cold gas from filaments does experience significant heating when accreted by the galaxy at radii much smaller than the virial radius (Nelson et al. 2013).

Besides the rate of gas accretion, the hot halo can be influenced by feedback mechanisms and photoionization from local sources. Feedback mechanisms can suppress cooling from the hot halo, modify the distribution of hot gas in the halo (van de Voort & Schaye 2012) and (to a limited degree) reduce the accretion rates on to haloes (Faucher-Giguère et al. 2011; van de Voort et al. 2011; Nelson et al. 2015). In this work, we investigate the impact of feedback mechanisms on the hot halo in detail and analyse whether reasonable changes to the feedback implementation result in a change to the mass scale of hot halo formation. Increasing photoionizing flux (higher star formation rate or an active nucleus) from local sources can decrease the net cooling rate of gas in the proximity of the galaxy, potentially suppressing cold gas accretion in low-mass haloes ($<5 \times 10^{11}$ M$_\odot$) and decrease the mass scale for hot halo formation (Cantalupo 2010). However, the results are sensitive to the assumed escape fraction, and Vogelsberger et al. (2012) found only small effects when including local active galactic nucleus (AGN). For simplicity, we will assume the gas is only exposed to the metagalactic background radiation.

We use the suite of cosmological hydrodynamical simulations from the EAGLE project (Crain et al. 2015; Schaye et al. 2015) to investigate the physical properties of the hot gas in the halo, and their dependence on energy sources like stellar feedback and AGN feedback. Our main goal is to study the thermal properties of gas accreting on to haloes and the gas mass that remains hot in the halo ($M_{\rm hot}$). In addition, we develop analytic and semi-analytic approaches to calculate the heating rates of gas in the halo and the mass scale of hot halo formation, which we apply in a companion work (Correa et al. in preparation, hereafter Paper II). In Paper II, we derive a physically motivated model for gas accretion on to galaxies that accounts for the hot/cold modes of accretion on to haloes and for the rate of gas cooling from the hot halo. With this model, we aim to provide some insight into the physical mechanisms that drive the gas inflow rates on to galaxies.

The outline of this paper is as follows. We describe the EAGLE simulations series used in this study and the analysis methodology in Section 2. We present our main results concerning the physical properties of hot and cold gas in the halo in Section 3 and on the modes of gas accretion in Section 4. In Section 5, we develop an analytic approach to calculate a 'critical mass scale', $M_{\rm crit}$, for

hot halo formation, and compare it with our numerical results and previous works. Finally, in Section 6 we summarize our conclusions.

2 SIMULATIONS

To investigate the formation and evolution of hot haloes surrounding galaxies, we use cosmological, hydrodynamical simulations from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) project (Crain et al. 2015; Schaye et al. 2015). The EAGLE simulations were run using a modified version of GADGET 3 (Springel 2005), a *N*-Body Tree-PM smoothed particle hydrodynamic (SPH) code. The EAGLE version contains a new formulation of SPH, new time stepping and new subgrid physics. Below we present a summary of the EAGLE models. For a more complete description, see Schaye et al. (2015).

The EAGLE simulations assume a Λ cold dark matter (ΛCDM) cosmology with the parameters derived from *Planck-1* data (Planck Collaboration et al. 2014), $\Omega_{\rm m} = 1 - \Omega_\Lambda = 0.307$, $\Omega_{\rm b} = 0.04825$, $h = 0.6777$, $\sigma_8 = 0.8288$, $n_s = 0.9611$. The primordial mass fractions of hydrogen and helium are $X = 0.752$ and $Y = 0.248$, respectively.

Table 1 lists the box sizes and resolutions of the simulations used in this work. We use the notation *LxxxNyyyy*, where xxx indicates box size (ranging from 25 to 100 comoving Mpc) and yyyy indicates the cube root of the number of particles per species (ranging from 376^3 to 1504^3, with the number of baryonic particles initially equal to the number of dark mater particles). The gravitational softening was kept fixed in comoving units down to $z = 2.8$ and in proper units thereafter. We will refer to simulations with the mass and spatial resolution of L025N0376 as intermediate-resolution runs and to simulations with the resolution of L025N0752 as high-resolution runs.

2.1 Baryonic physics

Radiative cooling and photoheating are included as in Wiersma, Schaye & Smith (2009). The element-by-element radiative rates are computed in the presence of the cosmic microwave background (CMB), and the Haardt & Madau (2001) model for UV and X-ray background radiation from quasars and galaxies.

Star formation is modelled following the recipe of Schaye & Dalla Vecchia (2008). Star formation is stochastic above a density threshold, $n_{\rm H,0}$, that depends on metallicity (in the model of Schaye 2004, $n_{\rm H,0}$ is the density of the warm, atomic phase just before it becomes multiphase with a cold, molecular component), with the probability of forming stars depending on the gas pressure. The implementation of stellar evolution and mass loss follows the work of Wiersma et al. (2009). Star particles are treated as simple stellar populations with a Chabrier (2003) initial mass function, spanning in the range of 0.1–100 M$_\odot$. Feedback from star formation and supernovae events follows the stochastic thermal feedback scheme of Dalla Vecchia & Schaye (2012). Rather than heating all neighbouring gas particles within the SPH kernel, they are selected stochastically based on the available energy, and then heated by a fixed temperature increment of $\Delta T = 10^{7.5}$ K. The probability that a neighbouring SPH particle is heated is determined by the fraction of the energy budget that is available for feedback, $f_{\rm th}$. If ΔT is sufficiently high to ensure that radiative losses are initially small, the physical efficiency of feedback can be controlled by adjusting $f_{\rm th}$. The value $f_{\rm th} = 1$ corresponds to the expected value of energy injected by core collapse supernovae ($E_{\rm SN} = 1.736 \times 10^{49}$ erg M$_\odot^{-1}$

Table 1. List of simulations used in this work. From left to right, the columns show simulation identifier; comoving box size; number of dark matter particles (initially there are equally many baryonic particles); initial baryonic particle mass; dark matter particle mass; comoving (Plummer-equivalent) gravitational softening; maximum physical softening.

Simulation	L (comoving Mpc)	N	m_b ($M_\odot$)	m_{dm} ($M_\odot$)	ϵ_{com} (comoving kpc)	ϵ_{prop} (proper kpc)
L025N0376	25	376^3	1.81×10^6	9.70×10^6	2.66	0.70
L025N0752	25	752^3	2.26×10^5	1.21×10^6	1.33	0.35
L050N0752	50	752^3	1.81×10^6	9.70×10^6	2.66	0.70
L100N1504	100	1504^3	1.81×10^6	9.70×10^6	2.66	0.70

Table 2. List of feedback parameters that are varied in the simulations. From left to right, the columns show simulation identifier (prefix), asymptotic maximum and minimum values of the efficiency of star formation feedback (f_{th}), density term denominator ($n_{H,0}$) and exponents (n_n and n_Z) from equation (1), and temperature increment of stochastic AGN heating (ΔT_{AGN}).

Simulation	$f_{th,(max,min)}$	$n_{H,0}$ (cm^{-3})	$n_n (= n_Z)$ (cm^{-3})	ΔT_{AGN} (K)
Ref	3.0, 0.30	0.67	$2/\ln(10)$	$10^{8.5}$
Less energetic FB	1.5, 0.15	0.67	$2/\ln(10)$	$10^{8.5}$
More energetic FB	6.0, 0.60	0.67	$2/\ln(10)$	$10^{8.5}$
No AGN FB	3.0, 0.30	0.67	$2/\ln(10)$	—
More explosive AGN FB	3.0, 0.30	0.67	$2/\ln(10)$	$10^{9.5}$
Recal	3.0, 0.30	0.25	$1/\ln(10)$	10^9

per solar mass of stars formed). EAGLE takes f_{th} to be a function of the local physical conditions,

$$f_{th} = f_{th,min} + \frac{f_{th,max} - f_{th,min}}{1 + \left(\frac{Z}{0.1 Z_\odot}\right)^{n_Z} \left(\frac{n_{H,birth}}{n_{H,0}}\right)^{-n_n}}, \qquad (1)$$

which depends on maximum and minimum threshold values ($f_{th,max}$ and $f_{th,min}$, respectively), on density (n_H refers to hydrogen number density and $n_{H,birth}$ to the density inherited by the star particle) and metallicity (Z) of the gas particle. The reference simulations (hereafter Ref) use $f_{th,max} = 3$, $f_{th,min} = 0.3$ and $n_{H,0} = 0.67 \, \mathrm{cm}^{-3}$. These values were chosen to obtain good agreement with the observed present-day galaxy stellar mass function and disc galaxy sizes (as described by Crain et al. 2015).

Black hole seeds (of mass $\approx 1.4 \times 10^5 \, M_\odot$) are included in the gas particle with the highest density in haloes of mass greater than $\approx 1.4 \times 10^{10} \, M_\odot$ that do not contain black holes (Springel, Di Matteo & Hernquist 2005). Black holes can grow through mergers and gas accretion. The accretion events follow a modified Bondi–Hoyle formula that accounts for the angular momentum of the accreting gas (Rosas-Guevara et al. 2015; Schaye et al. 2015), and a free parameter that is related to the disc viscosity. AGN feedback follows the accretion of mass on to the black hole, where a fraction (0.015) of the accreted rest mass energy is released as thermal energy into the surrounding gas, and is implemented stochastically, as per the stellar feedback scheme, with a fixed free parameter heating temperature, ΔT_{AGN}, which is set to $10^{8.5}$ K in the reference simulations.

When the resolution is increased, the parameters may need to be (re-)calibrated to retain the agreement with observations. The high-resolution simulation with recalibrated parameters is called Recal. In addition to Ref and Recal, we also use simulations with different feedback implementations to test the impact of feedback on the formation of the hot halo. Table 2 lists the values of the feedback parameters adopted in each simulation. In the table, the

simulation identifier describes the differences in the feedback with respect to Ref. In the stellar feedback case, 'Less/More Energetic FB' means that in these simulations, the energy injected per mass of stars formed is lower/higher with respect to Ref. In the AGN case, 'More Explosive AGN FB' means that AGN feedback is more explosive and intermittent, but the energy injected per unit mass accreted by the BH does not change with respect to Ref. Additional information regarding the performance of the EAGLE simulations including an analysis of subgrid parameter variations, a study of the evolution of galaxy masses, star formation rates and sizes can be found in Crain et al. (2015), Furlong et al. (2015, 2017) and Schaye et al. (2015).

2.2 Hydrodynamics

There has been much debate regarding the systematic differences between SPH, grid codes and moving mesh grid codes when modelling fluid mixing and gas heating and cooling (e.g. Agertz et al. 2007; Vogelsberger et al. 2012; Nelson et al. 2013). It has been shown by Hutchings & Thomas (2000) and Creasey et al. (2011) that SPH simulations may not adequately resolve shocks of accreted gas. Since shocks are generally spread over several SPH kernel lengths, the heating rate is smoothed over time, potentially making it easier for radiative cooling to become important. In addition, if radiative cooling is able to limit the maximum temperature reached by the gas particle, numerical radiative losses may be enhanced.

In contrast, numerical simulations using grids do not smooth out the shocks, and are thus better at identifying shock temperatures spikes. Numerical simulations using moving mesh codes can also capture shocks accurately. However, in common with grid codes, they may suffer from numerical mixing of hot and cold gas as the fluid moves across cells. Recently, Nelson et al. (2013) compared the moving mesh code AREPO (Springel 2010) with the standard SPH version of GADGET, and calculated the rates of cold mode of gas accretion on to haloes and galaxies. They found that while the rates of gas accreted cold on to haloes are in very good agreement between the simulations run with GADGET and AREPO, the rates of gas accreted cold on to galaxies differ significantly, with galaxies in AREPO having a 20 per cent lower cold fraction in $10^{11} \, M_\odot$ haloes. Nelson et al. (2013) concluded that most of the cold gas from filaments experiences significant heating after crossing the virial radius, implying that the numerical deficiencies inherent in different simulation codes may modify the relative contributions of hot and cold modes of accretion on to galaxies.

Some differences in the contributions of hot and cold modes of accretion on to galaxies and haloes may, however, be due to the method employed to select shock-heated gas. Previous works (e.g. Kereš et al. 2005, 2009; Faucher-Giguère et al. 2011; van de Voort et al. 2011; Nelson et al. 2013, amongst others) followed the

thermal history of the gas and applied a fixed temperature cut to the distribution of the maximum past temperature (T_{max}), to separate hot from cold mode accretion. However, T_{max} is not suitable for identifying cold flows if the gas experiences a shock but cools immediately afterwards, as may happen for accretion on to galaxies. In this case, a filament that is mostly cold except at a point near the galaxy would be labelled as hot mode accretion by numerical studies using T_{max}, but observers would identify it as a cold flow. This practical problem may not be important for SPH simulations that suffer from 'in shock cooling' because they do not resolve the accretion shocks on to the galaxy, or, as in the case of van de Voort et al. (2011) and EAGLE, that impose a temperature–density relation on to high-density gas, but it may affect the conclusions inferred from moving mesh codes using the T_{max} statistic. To avoid this issue, we use an alternative method to identify shock-heated particles in Section 4, based on post-shock temperature values.

Hydrodynamic solvers may also produce differences in the hot/cold modes of accretion. The EAGLE version of GADGET uses the hydrodynamic solver 'Anarchy', which greatly improves the performance on standard hydrodynamical tests, when compared to the original SPH implementation in GADGET (Schaller et al. 2015, see Hu et al. 2014 for similar results). Anarchy makes use of the pressure-entropy formulation derived in Hopkins (2013), alleviating spurious jumps at contact discontinuities. It also uses an artificial viscosity switch advocated by Cullen & Dehnen (2010) that allows the viscosity limiter to be stronger when shocks and shear flows are present. In addition, Anarchy includes an artificial conduction switch (similar to that of Price 2008), the C^2 Wendland (1995) kernel and the time step limiters of Durier & Dalla Vecchia (2012). These changes ensure that ambient particles do not remain inactive when a shock is approaching.

Recently, Sembolini et al. (2016) compared cosmological simulations of clusters using SPH as well as mesh-based codes. They found that the modern SPH schemes (such as Anarchy) that allow entropy mixing produce gas entropy profiles that are indistinguishable from those obtained with grid-based schemes. In addition, Schaller et al. (2015) compared the EAGLE simulations with simulations run with the same subgrid physics, but using the standard GADGET rather than the Anarchy hydrodynamic solver. They found that while simulations with standard SPH contain haloes with a large number of dense clumps of gas at all radii, Anarchy's ability to mix phases allows dense clumps to dissolve into the hot halo. These substantial improvements of the SPH formulation in the EAGLE simulations motivate a detailed description of the resulting predictions for hot halo formation and of hot/cold mode accretion.

2.3 Identifying haloes and galaxies

Throughout this work, we select the largest subhalo in each Friends-of-Friends (FoF) group, and use the SUBFIND algorithm (Springel, White & Hernquist 2001; Dolag et al. 2009) to identify the substructures (subhaloes) within it. The FoF algorithm adopts a dimensionless linking length of 0.2, and the SUBFIND algorithm calculates halo virial masses and radii via a spherical overdensity routine that centres the main subhalo from the FoF group on the minimum of the gravitational potential. We define halo masses, M_{200}, as the mass of all matter within the radius, R_{200}, for which the mean internal density is 200 times the critical density of the Universe.

To select the gas associated with the central galaxies embedded in each resolved halo, we identify the gravitationally bound cold and dense gas within R_{200} that is star forming and/or has a hydrogen number density, $n_H > 0.01 \, cm^{-3}$, and temperature $T < 10^5 \, K$. We

also require all particles to be contained within a sphere of radius $0.15 \times R_{200}$ in order to avoid labelling infalling cold flows (that would be included by the $T - n_H$ cuts but are mostly at large radii) as part of the galaxy.

2.4 Measuring gas accretion

Once we have identified the haloes, we build merger trees across the simulation snapshots.[1] The standard procedure to build a halo merger tree is to link each progenitor halo with a unique 'descendant' halo in the subsequent output (see e.g. Fakhouri, Ma & Boylan-Kolchin 2010). To do so, we identify the main branches of the halo merger trees and compute the halo (and central galaxy) accretion rate between two consecutive snapshots. At each output redshift (snapshot), we select the most massive haloes within each FoF group and consider them to be 'resolved' if they contain more than 1000 dark matter particles, which corresponds to a minimum halo mass of $M_{200} = 10^{9.8} \, M_\odot$ ($10^{8.8} \, M_\odot$) in the intermediate-(high-) resolution simulations. This limit on the number of dark matter particles results from a convergence analysis that we present in Paper II, where we find that in smaller haloes the accretion on to galaxies does not converge, indicating that the inner galaxies are not well resolved. We refer to the resolved haloes as 'descendants', and then link each descendant with a unique 'progenitor' at the previous output redshift. This is non-trivial due to halo fragmentation, in which subhaloes of a progenitor halo may have descendants that reside in more than one halo. Such fragmentation can be either spurious or due to a physical unbinding event. To account for this, we link the descendant to the progenitor that contains the majority of the descendant's 25 most bound dark matter particles (see Correa et al. 2015b for an analysis of halo mass history convergence using these criteria to connect haloes between snapshots).

We distinguish between gas accreted on to a halo and gas accreted on to a galaxy. For each descendant halo at z_i and its linked progenitor at z_j ($z_j > z_i$), we identify the particles that are in the descendant but not in its progenitor by performing particle ID matching. We then select particles that are new in the halo and reside within the virial radius, as particles accreted on to the halo in the redshift range of $z_i \leq z < z_j$. The accretion rate on to galaxies is further explored in Paper II, where we follow the methodology described above for calculating accretion rates on to haloes, and we select the new particles within the radius $0.15 \times R_{200}$ as particles accreted on to the galaxies in the redshift range of $z_i \leq z < z_j$ (see Paper II, Section 2.1, for a discussion on methods for calculating gas accretion on to galaxies).

3 HOT HALO FORMATION

The simple models of galaxy formation (e.g. Rees & Ostriker 1977; White & Rees 1978) assume that as long as the cooling time, t_{cool}, is shorter than the dynamical time, t_{dyn}, the infalling gas cools (inside a 'cooling radius', White & Frenk 1991) and settles into the galaxy. Otherwise, the gas is unable to efficiently radiate its thermal energy and forms a hot hydrostatic atmosphere, which is pressure supported against gravitational collapse. More recent semi-analytic models of galaxy formation assume that the cooling radius expands outwards as a function of time, therefore the comparison is done between gas cooling time and a different time-scale representing the time

---

[1] The simulation data are saved in 10 discrete output redshifts between redshift 0 to 1, in 4 output redshifts between redshift 1 and 3, and in 8 output redshifts between redshift 3 and 8.
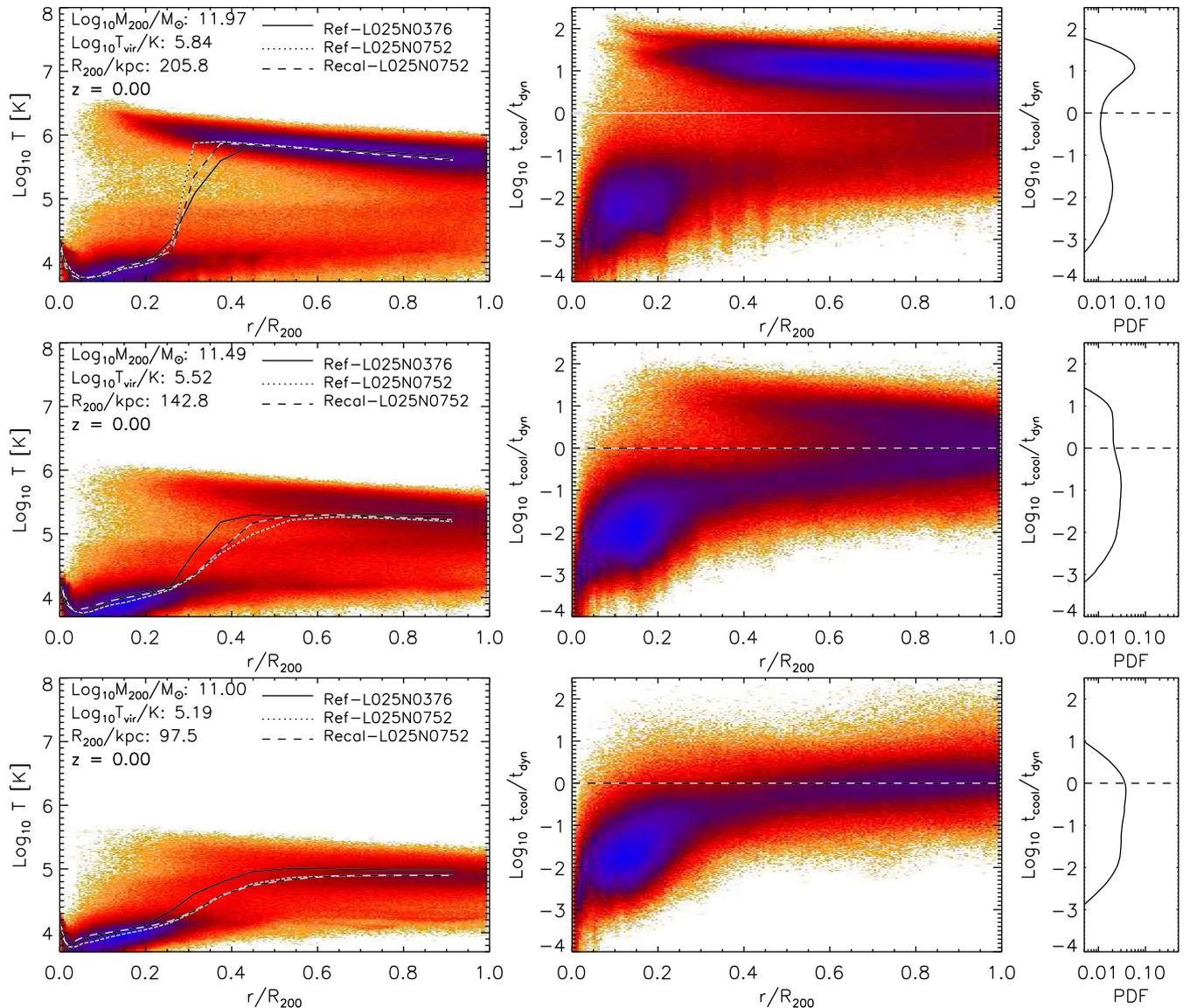
Figure 1. Temperature profile (left column), logarithm of the ratio between cooling times and local dynamical times (middle column) and the mass-weighted probability density function (PDF, right column) of $\log_{10} t_{\rm cool}/t_{\rm dyn}$ of gas from haloes in the mass range of $10^{11.9}$–$10^{12.1}$ M$_\odot$ (top row), $10^{11.4}$–$10^{11.6}$ M$_\odot$ (middle row) and $10^{10.9}$–$10^{11.1}$ M$_\odot$ (bottom row) at $z = 0$ taken from the Ref-L025N0752 simulation. The number of particles in a pixel is used for colour coding. The solid, dotted and dashed lines in the left-hand panels correspond to the median temperature per radial bin for different simulations.

available for cooling, like the time since the last major event or the time of halo formation (see e.g Lacey et al. 2016). In this section, we investigate when the hot hydrostatic halo forms in the EAGLE simulations, by analysing the interplay between the cooling and dynamical times of the gas particles in the halo. Throughout this work, we define hot halo gas as all gas particles that have $t_{\rm cool} > t_{\rm dyn}$ and that do not form part of the galaxy, i.e. $r > 0.15 R_{200}$.

We calculate $t_{\rm dyn}$ of the gas particle as

$$t_{\rm dyn} = r/V_{\rm c}(r), \qquad (2)$$

where $V_{\rm c}(r) = [GM(<r)/r]^{1/2}$ is the circular velocity and $M(<r)$ is the mass enclosed within r. We calculate $t_{\rm cool}$ as

$$t_{\rm cool} = \frac{3nk_{\rm B}T}{2\Lambda}, \qquad (3)$$

where n is the number density of the gas particle ($n = \rho_{\rm gas}/\mu m_{\rm p}$, μ is the molecular mass weight calculated from the cooling tables of

Wiersma et al. 2009, and $m_{\rm p}$ is the proton mass), $k_{\rm B}$ is the Boltzmann constant, T is the gas temperature and Λ is the cooling rate per unit volume with units of erg cm^{-3} s^{-1}. To calculate Λ, we use the tabulated cooling function for gas exposed to the evolving UV/X-ray background from Haardt & Madau (2001) given by Wiersma et al. (2009), which was also used by the EAGLE simulations. Note that the 'standard' definition for dynamical time of gas within a virialized system depends on R_{200} and $V_{\rm c}(R_{200})$, and not on the local radius and local circular velocity as defined here. We use local values rather than to investigate if shorter dynamical times in the inner dense regions give rise to a cooling flow. However, we find that changing equation (2) to $t_{\rm dyn} = R_{200}/V_{\rm c}(R_{\rm vir})$ does not change our conclusions.

Fig. 1 shows temperature profiles (left column), the logarithm of the ratio between cooling times and dynamical times (middle column) and the respective mass-weighted probability density function (PDF) of $\log_{10} t_{\rm cool}/t_{\rm dyn}$ (right column) for gas from haloes in the
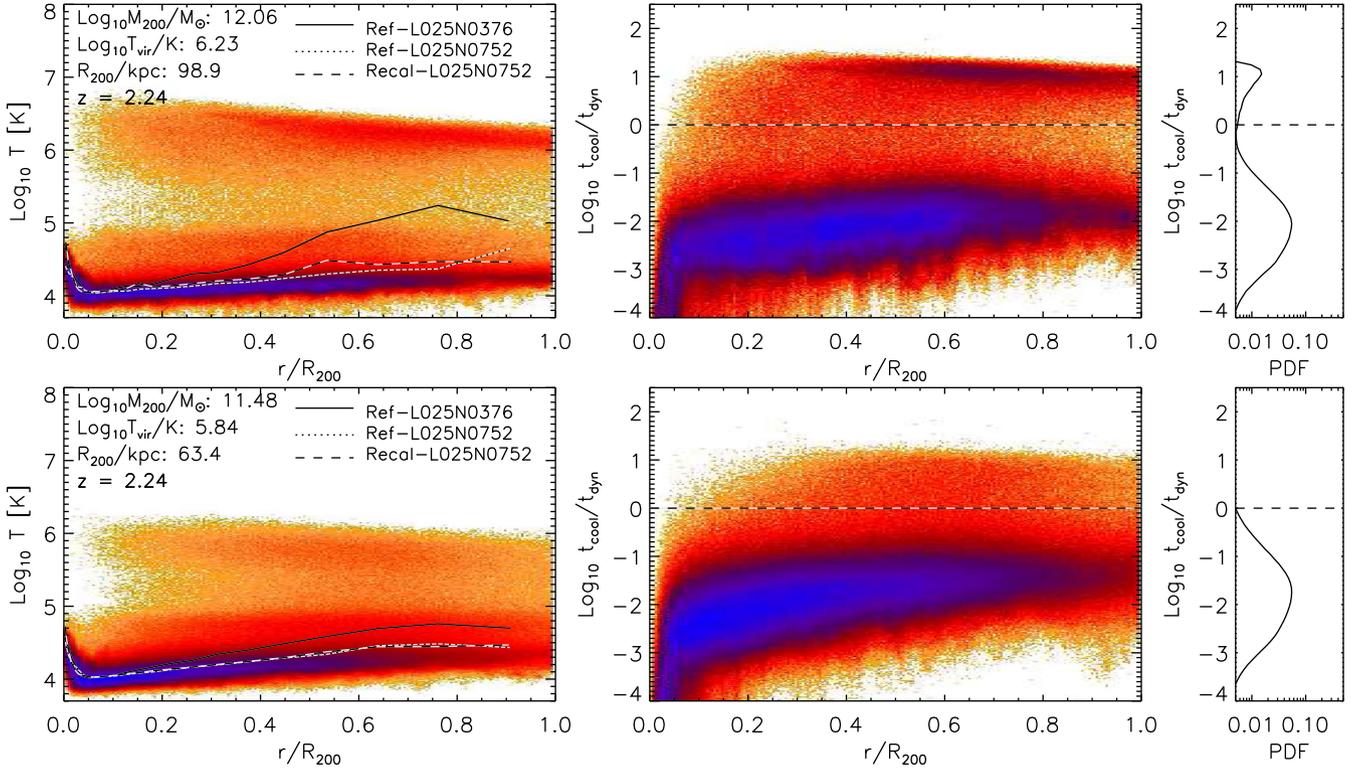
Figure 2. Same as Fig. 1 but for haloes at $z = 2.24$.

mass range of $10^{11.9}$–$10^{12.1}$ $M_\odot$ (top row), $10^{11.4}$–$10^{11.6}$ $M_\odot$ (middle row) and $10^{10.9}$–$10^{11.1}$ $M_\odot$ (bottom row) at $z = 0$, taken from the Ref-L025N0752 simulation. In the right-hand panels and throughout this work, the PDFs are calculated by stacking haloes in the selected mass range and distributing the gas particles in logarithmic bins of size 0.1 dex. We then sum the mass of the gas particles in each bin and normalize the distribution by the total gas mass. In the left-hand panels, the legend lists the median values of the mass, virial temperature (defined as $T_{\rm vir} = \frac{\mu m_{\rm p}}{2k_{\rm B}} V_{200}^2$, with $V_{200}^2 = GM_{200}/R_{200}$) and virial radius of haloes selected in each mass bin. The left panels also show in solid, dotted and dashed lines the median temperature per radial bin of gas from haloes taken from the simulations Ref-L025N0376, Ref-L025N0752 and Recal-L025N0752, respectively.

The left-hand panels of Fig. 1 show that while there is very good agreement in the median temperature profiles at small ($r/R_{200} < 0.2$) and large ($r/R_{200} > 0.4$) radii, at intermediate radii the median temperatures from the intermediate-resolution run (Ref-L025N0376) are larger by up to 0.3 dex than those from the high-resolution runs (L025N0752). This is in agreement with the convergence analysis of Nelson et al. (2016), who concluded that the physics (different models of stellar winds or AGN feedback) has a greater impact on $T(r)$ than resolution. We also find that in the radial range of $r = [0.2 - 0.4]R_{200}$, where the convergence with resolution is poorest, the median temperatures drop from $T_{\rm gas} \sim T_{\rm vir}$ to $T_{\rm gas} \sim 10^4$ K (in agreement with Nelson et al. 2016, $r_{\rm drop} \approx 0.25R_{200}$ and van de Voort & Schaye 2012, $r_{\rm drop} \approx 0.2 - 0.4R_{200}$), because of the high densities that rapidly decrease the gas cooling times, enabling it to radiate away its thermal energy and join the ISM.

The top and middle left-hand panels of Fig. 1 show that there is relatively little gas with $T \sim 10^5$ K at small and intermediate radii, reflecting the short cooling times at these temperatures. The cooling

flow in the hot halo is formed by $T = 10^6$ K gas that slowly decreases in temperature as it loses hydrostatic support due to cooling. The ISM consists of $T = 10^4$ K gas at $r/R_{200} < 0.15$. For a better understanding of the hot halo forming as a function of halo mass and its effect on the infall rate of gas on to the galaxy, we next analyse the middle and right columns. The bottom middle panel shows that in haloes with masses between $10^{10.9}$ and $10^{11.1}$ $M_\odot$, most of the gas has low temperature ($T_{\rm gas} < T_{\rm vir}$), short cooling times ($t_{\rm cool} < t_{\rm dyn}$) and infalls towards the central galaxy, although a substantial fraction of gas has $t_{\rm cool} > t_{\rm dyn}$ at $r \sim R_{200}$. At larger halo masses, a larger fraction of the gas is unable to cool and therefore forms a hot halo. The middle panel shows that haloes between $10^{11.4}$ and $10^{11.6}$ $M_\odot$ are in the intermediate stage between developing a hot stable atmosphere (gas with $T_{\rm gas} \sim T_{\rm vir}$ and $t_{\rm cool} > t_{\rm dyn}$) and continuing to fuel the galaxy. The top middle panel clearly shows a stable hot halo and a reduced amount of gas infalling towards the galaxies (gas at $r > 0.3R_{200}$ and $t_{\rm cool} < t_{\rm dyn}$).

Fig. 2 is similar to Fig. 1, but shows haloes in the mass range of $10^{11.9}$–$10^{12.1}$ $M_\odot$ (top panels), $10^{11.4}$–$10^{11.6}$ $M_\odot$ (bottom panels) at $z = 2.24$. The top middle panel shows that 10^{12} $M_\odot$ haloes develop a hot atmosphere, despite the significant fraction of cold gas at large radii that is accreted on to the halo. This cold gas forms part of the cold filamentary flows, which cross the virial radius, and are directly accreted on to the central galaxy. The cold accretion mode is best seen as the gas at $T = 10^4$ K and $r/R_{200} > 0.4$ in the bottom panel of Fig. 2, which remains cool (with $T = 10^4$ K), as it is accreted on to the galaxy.

The presence of cold flows produces gaseous haloes with an isothermal temperature profile of $T_{\rm gas} \sim 10^4$ K at all radii. Besides analysing the cooling time profiles, Nelson et al. (2016) and van de Voort & Schaye (2012) analysed the entropy profiles of haloes at $z = 2$, and concluded the that while the entropy of the

cold-mode gas decreases smoothly and strongly towards the centre, the entropy of the hot-mode gas decreases slightly down to $0.2R_{200}$, after which it drops steeply. We find that the cooling time profiles of the hot ($t_{\rm cool} > t_{\rm dyn}$) and cold ($t_{\rm cool} < t_{\rm dyn}$) modes follow the entropy profiles.

Figs 1 and 2 show that as the halo mass increases, so does the hot gas mass. In the case of 10^{11} M$_\odot$ haloes, the bottom left-hand panel of Fig. 1 shows that there is a large fraction of gas at $r > 0.4R_{200}$ with temperatures equal to or larger than the halo virial temperature. This seems to indicate that gas is shock-heating to the halo virial temperature when crossing R_{200} and forming a hot atmosphere. However, gas with $t_{\rm cool} \sim t_{\rm dyn}$ in the outer parts of $\sim 10^{11}$ M$_\odot$ haloes does not imply that the halo formed a stable hot atmosphere via gravitational accretion shocks, since the gas is affected by the extragalactic UV/X-ray background as we show in the next section.

The figures also show that as haloes are growing a hot atmosphere, the $t_{\rm cool}/t_{\rm dyn}$ PDF begins to present a strong bimodal shape, with a local maximum at $t_{\rm cool} > t_{\rm dyn}$ followed by a local minimum at $t_{\rm cool} < t_{\rm dyn}$ (see top right-hand panel). We then conclude that the bimodality in the cooling time PDF provides a clear signature of the formation of a hot halo, and the right-hand panels indicate that the hot hydrostatic halo forms in the halo mass range of $10^{11.5}$ M$_\odot$–10^{12} M$_\odot$ with a weak dependence on redshift.

We also analyse the radial velocity distributions of gas and find that in haloes of mass 10^{12} M$_\odot$ at $z = 0$, 92.2 per cent (61.3 per cent) of hot gas has an absolute radial velocity lower than $100\,{\rm km\,s}^{-1}$ ($50\,{\rm km\,s}^{-1}$), indicating that most of it is in hydrostatic equilibrium and not accreting on to the galaxy.

3.1 The impact of photoheating in low-mass haloes

In the previous section, we analysed the dependence of the gas cooling rates on halo mass and concluded that a halo with a hot hydrostatic atmosphere should present a strongly bimodal $t_{\rm cool}/t_{\rm dyn}$ PDF with a local maximum at $t_{\rm cool} > t_{\rm dyn}$. As the virial temperature decreases from $10^{5.5}$ K to $10^{5.2}$ K, we would naively expect the peak in the $t_{\rm cool}/t_{\rm dyn}$ PDF to shift towards shorter cooling times. This is, however, not the case: we find that at $z = 0$, the gas in haloes less massive than 10^{11} M$_\odot$ ($T_{\rm vir} \approx 10^{5.2}$ K) is affected by the extragalactic UV/X-ray background radiation, which strongly suppresses the net cooling rate of gas in the temperature range of $T \sim 10^4$–10^5 K (Efstathiou 1992; Wiersma et al. 2009). As a result, the peak in the $t_{\rm cool}/t_{\rm dyn}$ PDF remains at $t_{\rm cool} \sim t_{\rm dyn}$. This can be seen in Fig. 3, where we show the $t_{\rm cool}/t_{\rm dyn}$ PDF of gas from haloes in the mass range of $10^{10.4}$–$10^{10.6}$ M$_\odot$ (olive lines), $10^{10.8}$–$10^{11.0}$ M$_\odot$ (orange lines) and $10^{11.2}$–$10^{11.4}$ M$_\odot$ (red lines). The solid lines correspond to the case where the cooling rates are calculated for gas exposed to the evolving UV/X-ray background from Haardt & Madau (2001), while the dashed lines correspond to the case where the cooling rates are calculated for gas in collisional ionization equilibrium (CIE) and not exposed to the background radiation. Note, however, that we apply these CIE cooling rates to simulations that were run using cooling rates that did account for photoheating, limiting the gas temperature to $\sim 10^4$ K.

Fig. 3 shows that there is no large difference in the gas $t_{\rm cool}/t_{\rm dyn}$ PDF for haloes more massive than $10^{11.3}$ M$_\odot$, indicating that there is no strong impact of the background radiation on the cooling rates of gas from haloes with virial temperatures larger than 10^5 K. In smaller haloes, the peak in the $t_{\rm cool}/t_{\rm dyn}$ PDF curve is shifted to $t_{\rm cool} \sim 0.3 t_{\rm dyn}$ in the case of no background radiation.
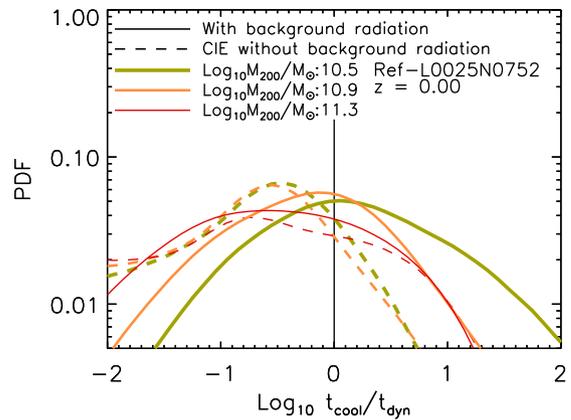
Figure 3. Mass-weighted PDF of the logarithm of the ratio between the net radiative cooling time and the local dynamical time for gas from haloes in the mass range of $10^{10.4}$–$10^{10.6}$ M$_\odot$ (green lines), $10^{10.8}$–10^{11} M$_\odot$ (orange lines), $10^{11.2}$–$10^{11.4}$ M$_\odot$ (red lines) at $z = 0$. The solid lines correspond to the case where the cooling rates are calculated for gas exposed to the evolving UV/X-ray background from Haardt & Madau (2001), while the dashed lines correspond to the case where the cooling rates are calculated for gas in collisional ionization equilibrium (CIE).

3.2 The impact of feedback

Feedback can affect the formation of the hot hydrostatic halo around galaxies. For example, very energetic SN activity generates large outflows and strong winds that shock against the gaseous halo. As a result, the winds can fill the halo with gas expelled from the galaxy, increasing the amount of hot gas at large radii. In this subsection, we compare the $t_{\rm cool}/t_{\rm dyn}$ mass-weighted PDF at fixed halo mass obtained from simulations with different feedback implementations (see Table 2 for reference and Section 2.1 for a brief description), and determine, by analysing whether the cooling time PDF is bimodal, the mass range where the hot halo is forming.

Fig. 4 shows the mass-weighted PDF of $t_{\rm cool}/t_{\rm dyn}$ of gas from haloes in the Ref-L100N1504 simulation in the mass range of $10^{11.4}$–$10^{11.6}$ M$_\odot$, $10^{11.9}$–$10^{12.1}$ M$_\odot$, $10^{12.4}$–$10^{12.6}$ M$_\odot$ and $10^{13.4}$–$10^{13.6}$ M$_\odot$ at $z = 0$ (left-hand panel) and $z = 2.24$ (right-hand panel). In the figures, the labels show the median halo mass for each mass bin. The region where $t_{\rm cool} > t_{\rm dyn}$ corresponds to hot gas in the halo.

As the halo mass increases so does the amount of hot gas (Section 3.3). The $t_{\rm cool}/t_{\rm dyn}$ PDF of gas in 10^{12} M$_\odot$ haloes at $z = 0$ shows a bimodal shape that becomes mostly unimodal in higher mass haloes. At $z = 2.24$, the bimodality persists up to the highest mass bin (10^{13} M$_\odot$) due to the contribution from cold flows that populate the peak at $t_{\rm cool} < t_{\rm dyn}$. We find that the presence of the bimodality in the $t_{\rm cool}/t_{\rm dyn}$ PDF indicates the increasing amount of hot gas at large radii and the eventual formation of the hot halo. Then, from visual inspection, we determine that the hot hydrostatic atmosphere is forming in haloes with masses between $10^{11.5}$ and 10^{12} M$_\odot$ at $z = 0$ and $z = 2.24$.

The panels in Fig. 5 repeat the analysis shown in the left-hand panel of Fig. 4, but instead show $t_{\rm cool}/t_{\rm dyn}$ mass-weighted PDFs for the L025N0376 simulations with different feedback prescriptions. In this case, the PDFs correspond to gas from haloes in the mass range of $10^{11.4}$–$10^{11.6}$ M$_\odot$, $10^{11.6}$–$10^{11.8}$ M$_\odot$, $10^{11.9}$–$10^{12.1}$ M$_\odot$ and $10^{12.4}$–$10^{12.6}$ M$_\odot$. In the panels, the top left legends indicate the total gas mass in the halo ($M_{\rm gas}$).

The simulation shown in the top left-hand panel of Fig. 5 does not have AGN feedback, while the one in the bottom left-hand panel uses a more explosive AGN feedback. Both include the same
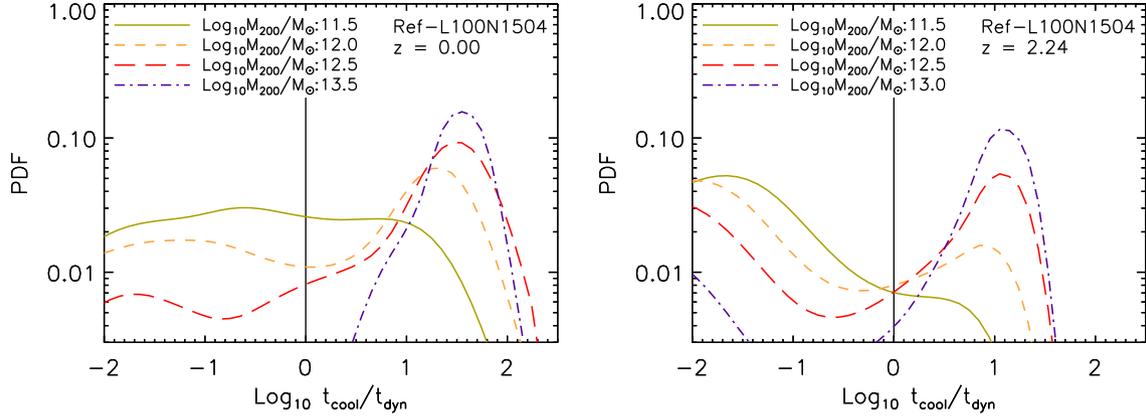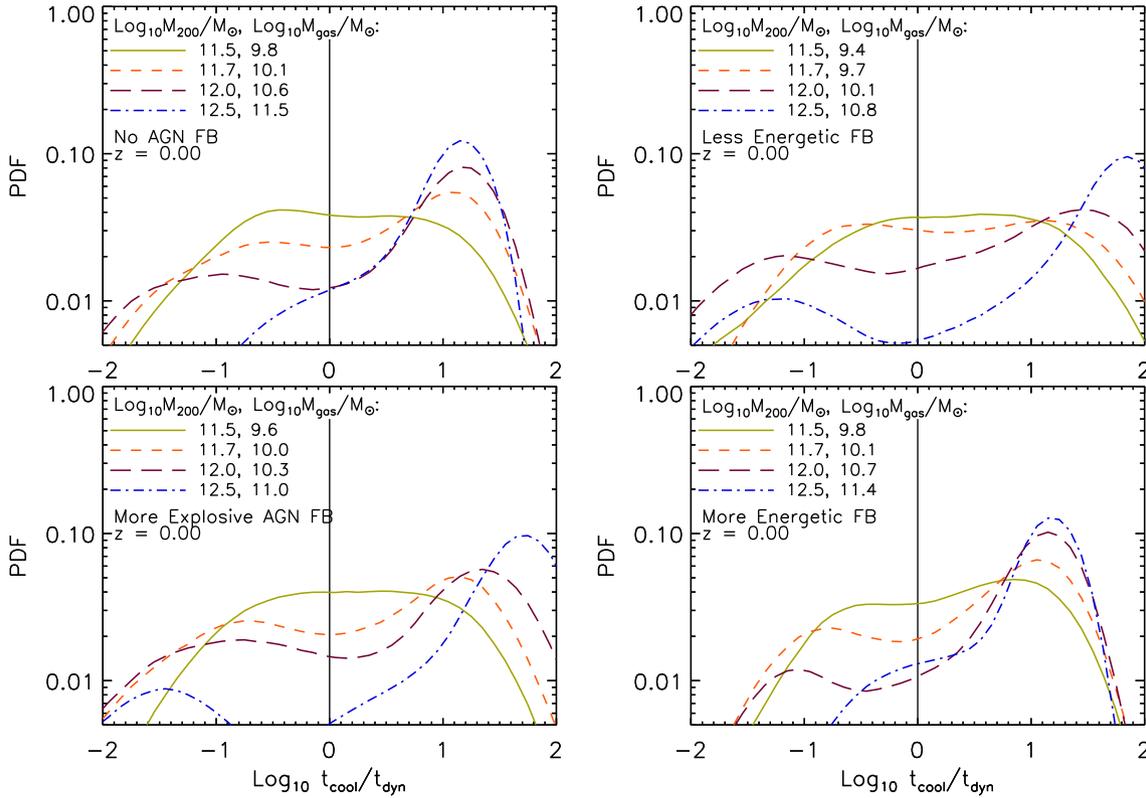
Figure 4. Mass-weighted PDF of the logarithm of the ratio between cooling times and local dynamical times of gas from haloes in the mass range of $10^{11.4}$–$10^{11.6}$ $M_\odot$, $10^{11.9}$–$10^{12.1}$ $M_\odot$, $10^{12.4}$–$10^{12.6}$ $M_\odot$ and $10^{13.4}$–$10^{13.6}$ $M_\odot$ at $z = 0$ (left-hand panel) and $z = 2.24$ (right-hand panel).

Figure 5. Mass-weighted PDF of the logarithm of the ratio between cooling times and local dynamical times of gas from haloes in the mass range of $10^{11.4}$–$10^{11.6}$ $M_\odot$, $10^{11.6}$–$10^{11.8}$ $M_\odot$, $10^{11.9}$–$10^{12.1}$ $M_\odot$ and $10^{12.4}$–$10^{12.6}$ $M_\odot$ at $z = 0$. The legends show the median mass of the haloes in each mass bin. The different panels show L025N0376 simulations with different feedback prescriptions: no AGN (top left), weak stellar feedback (top right), strong AGN (bottom left) and strong stellar feedback (bottom right).

feedback from star formation as in Ref. It can be seen that neither the bimodality of the t_{cool}/t_{dyn} PDF nor the amount of hot gas are strongly affected by AGN feedback in haloes with masses between $10^{11.5}$ and $10^{11.9}$ $M_\odot$. The right-hand panels in Fig. 5 show the t_{cool}/t_{dyn} PDFs in the less (top panel) and more energetic stellar feedback (bottom) scenarios (both including the same AGN feedback as in the Ref model). For these halo masses, stellar feedback has a strong impact on the t_{cool}/t_{dyn} PDFs. While a more energetic stellar feedback increases the fraction of hot gas, at least in the halo mass range probed by these simulations ($<10^{12}$ $M_\odot$) and thus limits the build-up of cold-mode gas in the halo centre (in agreement

with van de Voort & Schaye 2012), a less energetic stellar feedback maintains the bimodality in the t_{cool}/t_{dyn} PDF but shifts the peak in the t_{cool}/t_{dyn} PDF of hot gas towards larger cooling times. We find that the bimodality of the t_{cool}/t_{dyn} PDF is present in $10^{11.7}$ $M_\odot$ haloes with more energetic stellar feedback and in 10^{12} $M_\odot$ haloes with less energetic stellar feedback.

In the following section, we further analyse the dependence of the total hot gas mass on halo mass, redshift and feedback. In Section 5, we derive an analytic model for the formation of a stable hot hydrostatic atmosphere. In the model, we calculate a halo mass scale for which the gravitational heating rate of the hot halo gas

balances the gas cooling rate, thus keeping the gas hot and enabling the formation of a hot atmosphere. With the model, we show that the ability of a halo to develop a hot hydrostatic atmosphere depends on the amount of hot gas that the halo already contains, which we calculate in the next subsection, and on the fraction of accreted gas that shock-heats to the halo virial temperature (Section 4).

3.3 Hot gas mass

In order to better understand the build-up of the hot gas mass, M_{hot} (mass of gas with $t_{cool} > t_{dyn}$ and $r > 0.15R_{200}$) as haloes evolve, in this section we look for a correlation between M_{hot} and the total halo mass as a function of redshift. Fig. 6 shows the median ratio of $M_{hot}/(\Omega_b/\Omega_m)M_{200}$ (with $\Omega_b/\Omega_m = 0.146$ the universal baryon fraction) taken from a range of simulations (as indicated in the legends) at $z = 0$ (top panel) and at $z = 2.24$ (the second panel from the top). In these panels, the error bars show the 1σ scatter for the Ref-L025N0752 and Ref-L100N1504 simulations. The median ratio of $M_{hot}/(\Omega_b/\Omega_m)M_{200}$ is also shown in the third panel from the top, but in this case the values are taken from the Ref-L100N1504 simulation and at various output redshifts.

The top panel of Fig. 6 highlights the relatively poor agreement between the intermediate- and high-resolution simulations, with the latter predicting somewhat higher hot gas fractions. Good agreement is however achieved at $z = 2.24$ (middle panel). Although the Ref-L100N1504 simulation is not fully converged with respect to the numerical resolution at $z = 0$, the convergence with box size is excellent at all redshifts. The intermediate-resolution runs show that the hot gas represents <10 per cent of the total halo gas mass for $M_{200} < 10^{11.6}$ M$_\odot$ at $z = 0$. The hot gas mass fraction reaches 80–90 per cent in $10^{13.6}$ M$_\odot$ haloes and remains roughly constant for higher masses. In very low-mass haloes ($M_{200} < 10^{10.5}$ M$_\odot$), the hot gas mass fraction also remains roughly constant ($M_{hot}/(\Omega_m/\Omega_b)M_{200} \approx 0.02$–0.03). In these haloes, cold accretion dominates; therefore, the heating mechanism that maintains M_{hot} is the UV background as discussed in Section 3.1.

The third from the top panel of Fig. 6 shows the evolution of the hot gas fraction. In haloes larger than $10^{11.5}$ M$_\odot$, $M_{hot}/(\Omega_b/\Omega_m)M_{200}$ remains constant over the redshift range of 3–6 and at lower redshift it increases somewhat with time. In smaller haloes ($M_{200} < 10^{11.5}$ M$_\odot$), $M_{hot}/(\Omega_b/\Omega_m)M_{200}$ increases with time until $z = 1$ but decreases thereafter. We calculate the cooling rate of gas exposed to the UV background, and in the absence of it and compare the hot gas mass. We find that the hot gas mass in low-mass haloes increases due to the heating produced by the background radiation. In the case of gas not being exposed to the UV background, the total hot gas mass decreases with increasing redshift at fixed halo mass. We also find that the differences between M_{hot} occurs in haloes lower than $10^{11.3}$ M$_\odot$.

We next perform a least-square minimization to determine the best-fitting relation $M_{hot} - (\Omega_b/\Omega_m)M_{200}$ as a function of redshift. We apply equal weighting for each mass bin from the Ref-L100N1504 simulation (which we use to cover a large halo mass range) and minimize the quantity $\Delta_j = \frac{1}{N}\sum_{i=1}^{N} Y_i^2$, where

$$Y_i(z_j) = \log_{10}\left[\frac{M_{hot}}{(\frac{\Omega_b}{\Omega_m})M_{200}}\right]_i - F[M_{200,i}, \alpha(z_j), \beta(z_j), \gamma(z_j)], \quad (4)$$

N is the number of bins at each output redshift z_j, and F is

$$F = \alpha(z_j) + \beta(z_j)x_i + \gamma(z_j)x_i^2, \quad (5)$$

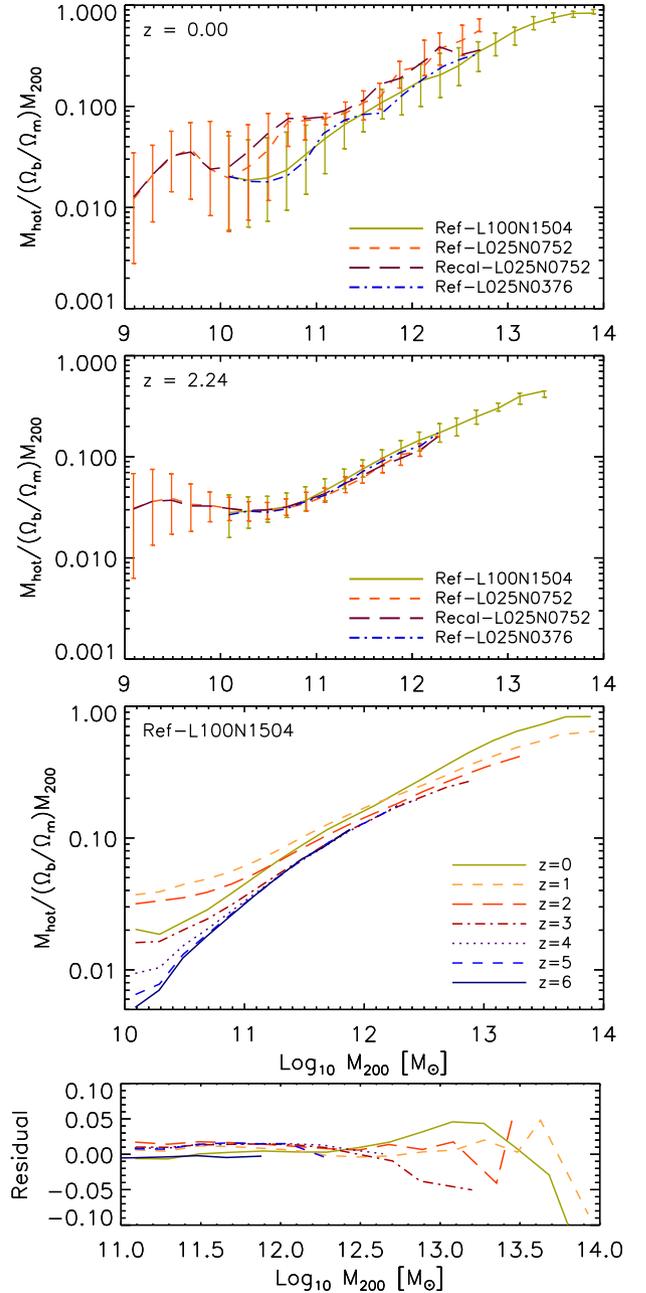$$x_i = \log_{10}(M_{200,i}/10^{12} \text{ M}_\odot). \quad (6)$$

Figure 6. Fraction of hot (with $t_{cool} > t_{dyn}$) gas mass with respect to the total halo mass, M_{200} (normalized by the universal baryon fraction), as a function of M_{200} for different simulations at $z = 0$ (top panel), at $z = 2.24$ (second panel from the top), and for Ref-L100N1504 at various redshifts (from $z = 0$ to $z = 6$, third panel from the top). The error bars in the top and middle panels show the 1σ scatter. Each bin contains at least five haloes. The bottom panel shows the residual of the data points with respect to the best-fitting expression (equations 7–10).

We obtain the best-fitting values for α, β and γ at each redshift z_j, and following the same methodology, we look for the best-fitting expression of these parameters as functions of redshift.

We find that the following expression best reproduces the relation in the halo mass range of $M_{200} = 10^{11}$–10^{14} M$_\odot$,

$$\log_{10}\left(\frac{M_{hot}}{(\frac{\Omega_b}{\Omega_m})M_{200}}\right) = \alpha(z) + \beta(z)x + \gamma(z)x^2, \quad (7)$$
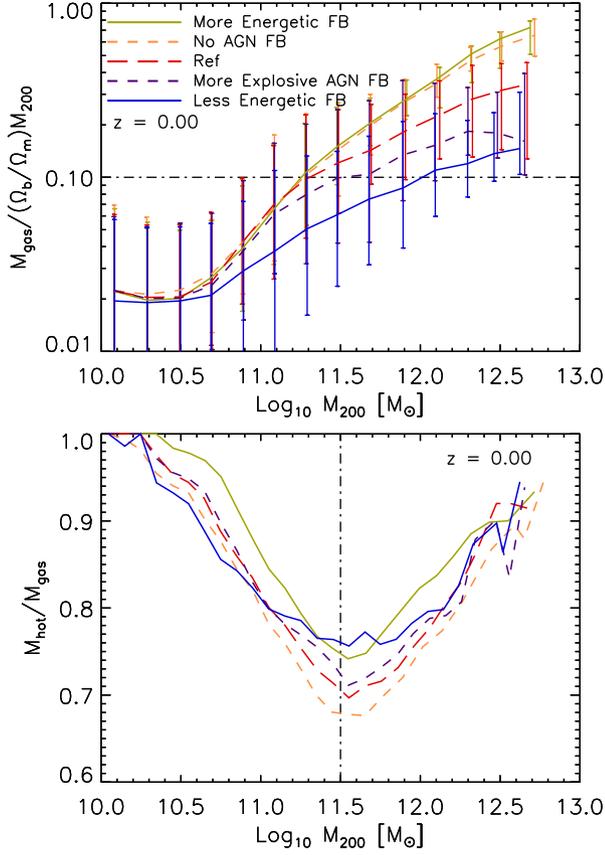
Figure 7. Top panel: Fraction of gas mass ($M_{\rm gas}(0.15R_{200} < r < R_{200})$ with respect to the total halo mass, M_{200} (normalized by the universal baryon fraction), as a function of M_{200} at $z = 0$. The different lines correspond to L025N0376 simulations with different feedback prescriptions (see Table 2 and/or Section 2). Bottom panel: Fraction of hot gas (gas with $t_{\rm cool} > t_{\rm dyn}$) with respect to $M_{\rm gas}$ as a function of M_{200}.

$$x = \log_{10}(M_{200}/10^{12}\,{\rm M}_\odot), \tag{8}$$

where α, β and γ are functions of z given by

$$
{\rm if}\, z \le 2 \begin{cases} \alpha(z) = -0.79 + 0.31\bar{z} - 0.96\bar{z}^2, \\ \beta(z) = 0.52 - 0.57\bar{z} + 0.85\bar{z}^2, \\ \gamma(z) = -0.05, \end{cases} \tag{9}
$$

$$
{\rm if}\, z > 2 \begin{cases} \alpha(z) = -0.38 - 1.56\bar{z} + 1.17\bar{z}^2, \\ \beta(z) = 0.12 + 0.94\bar{z} - 0.55\bar{z}^2, \\ \gamma(z) = -0.05, \end{cases} \tag{10}
$$

where $\bar{z} = \log_{10}(1 + z)$. The bottom panel of Fig. 6 shows the residual of the data points with respect to the best-fitting expression.

Next, we investigate how the presence of different feedback mechanisms affect the hot gas as well as the total gas mass ($M_{\rm gas}$) in the halo (all gas contained between 0.15 and $1R_{200}$). The top panel of Fig. 7 shows the $M_{\rm gas} - (\Omega_{\rm b}/\Omega_{\rm m})M_{200}$ relation for haloes in the mass range of 10^{10}–$10^{13}\,{\rm M}_\odot$ at $z = 0$ for the L025N0376 simulations. The different coloured lines correspond to simulations with different feedback prescriptions.

This panel shows that the impact of feedback increases with halo mass and that stellar feedback has a larger impact on the amount of gas in the halo than AGN feedback. We find that doubling the efficiency of the stellar feedback increases the gas mass fraction by

a factor of 1.3 in $10^{12}\,{\rm M}_\odot$ haloes relative to the Ref model, whereas halving the efficiency decreases the gas mass fraction by a factor of 2.5. No (Explosive) AGN feedback results in an increase (decrease) by a factor of 1.5 in the gas mass fraction.

While efficient stellar feedback increases the gas mass in the halo, more explosive AGN feedback decreases it. Overall, it seems that in haloes more massive than $10^{12}\,{\rm M}_\odot$ there is a greater difference in the gas mass fraction between Ref and More Energetic FB than between Ref and More Explosive AGN FB. This is due to two different reasons. Physically, AGN feedback mainly ejects gas mass from the halo, or prevents it from falling into the halo, whereas stellar feedback ejects gas out of the galaxy into the inner halo. Numerically, although stellar and AGN feedback use a similar thermal implementation (Dalla Vecchia & Schaye 2012), there is a difference in the actual energetics of the processes. The energy injected per mass of stars formed changes between Ref and More Energetic FB, whereas the energy injected per unit mass accreted by the BH does not change between Ref and More Explosive AGN FB. In the latter, it is only the intermittency and the explosiveness that changes as a consequence of the change in the temperature of the AGN. In the case of Less Energetic FB, we find that the gas mass in the halo decreases because more gas is accreted by the galaxy and locked up in stars.

The bottom panel of Fig. 7 shows the variation of the ratio $M_{\rm hot}/M_{\rm gas}$ with feedback. It can be seen that in haloes less massive than $10^{11.5}\,{\rm M}_\odot$, the $M_{\rm hot}/M_{\rm gas}$ ratio increases with decreasing halo mass, indicating that most of the halo gas is heated by the X-ray/UV background (see Section 3.1 for a discussion). In haloes more massive than $10^{11.5}\,{\rm M}_\odot$, $M_{\rm hot}/M_{\rm gas}$ increases with halo mass. While a more energetic stellar feedback increases the hot mass fraction by 10 per cent, no AGN feedback decreases it by 8 per cent in $10^{12}\,{\rm M}_\odot$ haloes. In the case of Less Energetic FB and More Explosive AGN, $M_{\rm hot}/M_{\rm gas}$ increases by 10 per cent and 3 per cent (on average) with respect to Ref, respectively, in the halo mass range of $10^{11.5}$–$10^{12}\,{\rm M}_\odot$.

In this section, gas particles with long cooling times ($t_{\rm cool} > t_{\rm dyn}$) are considered hot and counted in the calculation of $M_{\rm hot}$. Different from this work, van de Voort & Schaye (2012) separated hot and cold gas by performing a $T_{\rm max}$ cut and found that the hot fraction as a function of radius decreases not only when AGN feedback is switched on but also when stellar winds are enhanced. The reason for this is the way stellar feedback is implemented. In the more energetic stellar feedback simulation used by van de Voort & Schaye (2012), the wind velocity scales with the local sound speed, so it largely overcomes the pressure of the ISM, blowing the gas out of the galaxy and halo, thus decreasing the amount of hot gas. In our work, the efficiency of stellar feedback is regulated by the fraction of the energy budget available ($f_{\rm th}$), which makes it more/less energetic and controls the frequency of feedback events, but the temperature increase is kept fixed.

So far we have analysed the behaviour of the hot gas mass in the halo. In the next section, we investigate the fraction of gas that is accreted via hot and cold modes as a function of halo mass and redshift.

4 HOT AND COLD MODES OF ACCRETION

Over the last decade, numerical simulations have shown that gas accretion on to haloes occurs in two different modes, gas either shock-heats to the halo virial temperature near the virial radius (the hot accretion mode), or crosses the virial radius unperturbed (the
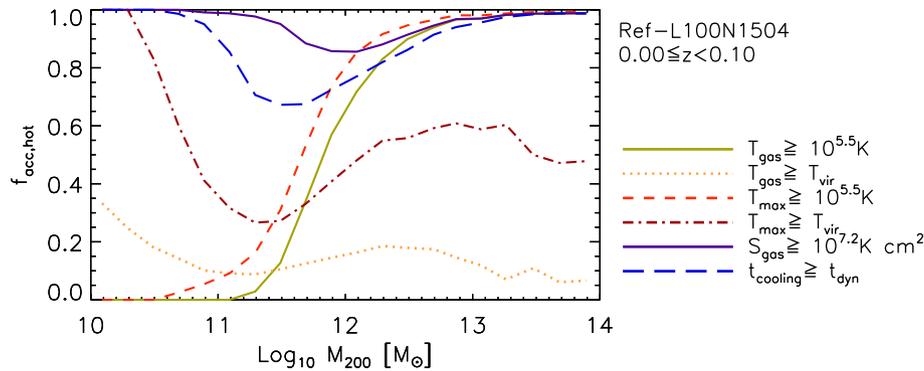
Figure 8. Fraction of gas accreted hot during the redshift range of $0 \leq z < 0.1$ as a function of halo mass. The various curves correspond to the hot fraction calculated using different methods as indicated in the legends.

cold accretion mode). Several works have found that cold accretion dominates in low-mass haloes ($M_{200} < 10^{12}\,\mathrm{M_\odot}$) and that the transition mass increases weakly with increasing redshift (e.g. Katz et al. 2003; Kereš et al. 2005, 2009; Ocvirk et al. 2008; van de Voort et al. 2011; van de Voort & Schaye 2012). The two modes coexist at high redshift in massive haloes, which develop a hot hydrostatic atmosphere despite experiencing significant cold accretion through filaments, generally referred to as 'cold flows' (Kereš et al. 2005; Dekel et al. 2009). Cold flows are important for galaxy formation, because even if they experience significant heating when crossing the hot atmosphere (Nelson et al. 2013), they are responsible for delivering cold, star-forming, gas deep within the halo (e.g. Dekel et al. 2009). In the following sections, we investigate different definitions that can be used to calculate the modes of accretion in the EAGLE simulations, analyse the impact of the hydrodynamic scheme and obtain best-fitting relations.

4.1 Definition of hot accretion

The contributions from the two different modes of accretion have generally been calculated from the temperature history of the accreted gas (e.g. Kereš et al. 2005; van de Voort et al. 2011; Nelson et al. 2013, amongst others), by following the maximum temperature, T_{max}, each gas particle has ever reached. Kereš et al. (2005) found a clear bimodality in the distribution of T_{max} of accreted particles, and they proposed a threshold value of $T_{max} = 2.5 \times 10^5$ K, given by the minimum in the distribution of T_{max} values, to determine whether gas is accreted hot ($T_{max} \geq 2.5 \times 10^5$ K) or cold ($T_{max} < 2.5 \times 10^5$ K). This is an often used method but other approaches have also been taken. For example, Brooks et al. (2009) identified hot gas accretion based on an entropy jump criterion and concluded that their method led to a distinction between hot/cold modes in very good agreement with the selection of hot/cold gas based on the use of a constant temperature threshold. In this section, we compare T_{max} with other variables that can also give us some insight into whether a particle experienced a shock when crossing the virial radius.

We follow the method described in Section 2.4, and look for gas particles that crossed the virial radius between two consecutive snapshots ($z_i \leq z < z_j$). Then, we calculate the mass-weighted PDFs of the gas particles T_{max}, temperature (T_{gas}) and entropy (S_{gas}) at redshift z_i (see Fig. A1). We find that the PDFs for the redshift interval $0.0 < z \leq 0.1$ and $2.0 < z \leq 2.2$ are bimodal, but only for haloes larger than $10^{12}\,\mathrm{M_\odot}$, with the location of the local minimum changing with M_{200}. A detailed analysis of the PDFs can be found

in Appendix A. We next use the minimum of the PDFs from the $10^{12}\,\mathrm{M_\odot}$ haloes as a threshold value to separate particles accreted hot and cold. For T_{gas} and T_{max} PDFs, we select the threshold to be $T_{min} = 10^{5.5}$ K, and for S_{gas}, we use $S_{min} = 10^{7.2}$ K cm^2.

Fig. 8 shows the fraction of gas particles accreted hot, $f_{acc,hot}$, in the redshift range of 0–0.1 as a function of halo mass. The different lines correspond to $f_{acc,hot}$ calculated using different definitions. We first calculated $f_{acc,hot}$ requiring that the gas particles at redshift 0 have temperatures higher than $10^{5.5}$ K (olive solid line), or that the gas particles have temperatures higher than the host halo virial temperature (orange dotted line). Then, we calculated $f_{acc,hot}$ requiring that the gas particle maximum past temperature is higher than $10^{5.5}$ K (red dashed line) or higher than the host halo virial temperature (dark red dot–dashed line). Finally, we calculated $f_{acc,hot}$ requiring that the entropy of the gas particles is larger than $10^{7.2}$ K cm^2 (purple dot-dot-dot–dashed line) or that the gas particles cooling time is larger than the local dynamical time (blue dashed line).

As previous works have shown (e.g. van de Voort et al. 2011; Nelson et al. 2013), the hot fraction depends very much on the definition. The fixed temperature cut, $T_{gas} \geq 10^{5.5}$ K ($T_{max} \geq 10^{5.5}$ K), gives a hot fraction that increases from 0.2 (0.5) in $10^{11.5}\,\mathrm{M_\odot}$ haloes to 0.7 (0.9) in $10^{12}\,\mathrm{M_\odot}$ haloes, thus showing a smooth transition from cold to hot accretion in the halo mass range of 10^{11} to $10^{13}\,\mathrm{M_\odot}$. On the other hand, the criterion $T_{gas} \geq T_{vir}$ ($T_{max} \geq T_{vir}$) indicates that for the most massive haloes the hot mode accounts for only 20 per cent (60 per cent) of the accreted gas particles. For the low-mass haloes, $T_{gas} \geq T_{vir}$ ($T_{max} \geq T_{vir}$) gives a hot fraction that increases from 0.1 (0.35) in $10^{11}\,\mathrm{M_\odot}$ haloes to 0.35 (1.0) in $10^{10}\,\mathrm{M_\odot}$ haloes. This seems to indicate that there is another mechanism, such as shocks driven by winds or heating by the extragalactic UV/X-ray background radiation, that makes gas reach temperatures higher than the halo's T_{vir}. In the case of T_{max}, stellar feedback events certainly increase the hot fraction, since we obtain a peak in the T_{max} PDFs at $10^{7.5}$ K for all halo mass, which disappears when stellar feedback is switched off.

Fig. 8 also shows that $f_{acc,\,hot}$ calculated using $S_{gas} \geq 10^{7.2}$ K cm^2 and $t_{cooling} \geq t_{dyn}$ decreases in the halo mass range of 10^{10}–$10^{12}\,\mathrm{M_\odot}$ and increases for higher halo masses, being in agreement with $f_{acc,hot}$ from $T_{gas} \geq 10^{5.5}$ K in haloes larger than $10^{12}\,\mathrm{M_\odot}$. From this upturn, we conclude that for haloes with $T_{vir} \gg 10^5$ K ($M_{200} > 10^{12}\,\mathrm{M_\odot}$), a large fraction of the accreted gas goes through a virial shock, and thus we can safely separate hot and cold accretion using $T_{gas} \geq 10^{5.5}$ K or $T_{max} \geq 10^{5.5}$ K. In lower mass haloes ($M_{200} < 10^{12}\,\mathrm{M_\odot}$), separating hot and cold accretion is not so easy. Although the hot halo is not expected to form (DB06) and the T_{gas}, T_{max} and S_{gas}
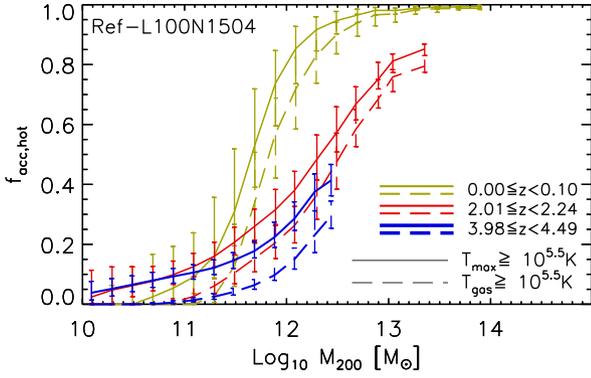
Figure 9. Fraction of gas accreted hot during the redshift ranges $0 \le z < 0.1$ (green lines), $2.0 \le z < 2.2$ (red lines) and $4.0 \le z < 4.49$ (thick blue lines), against halo mass. The solid curves correspond to the hot fraction calculated using $T_{\max} \ge 10^{5.5}$ K, whereas the dashed curves correspond to $T_{\mathrm{gas}} \ge 10^{5.5}$ K. The error bars show the 1σ scatter of the fractions.

PDFs are unimodal, UV background radiation (as discussed in Section 3.1) can significantly increase the cooling time and entropy of accreted gas.

Throughout this work, we have investigated how gas heated by either stellar or AGN feedback, UV/X-ray background radiation or accretion shocks, evolves with halo mass and redshift. We next aim to identify gas that is mostly heated by accretion shocks when crossing the virial radius. While stellar or AGN feedback does not strongly impact gas falling on to the halo (van de Voort et al. 2011, it does strongly affect gas falling on to the galaxy, van de Voort et al. 2011; Paper II), UV radiation does for low-mass haloes ($<10^{11.5}$ M$_\odot$). This can be seen in Fig. 8 through the unexpected increase in $f_{\mathrm{acc,hot}}$ towards very low halo masses using $S_{\mathrm{gas}} \ge 10^{7.2}$ K cm^2, $t_{\mathrm{cool}} \ge t_{\mathrm{dyn}}$, $T_{\mathrm{gas}} \ge T_{\mathrm{vir}}$ and $T_{\max} \ge T_{\mathrm{vir}}$. These methods clearly indicate that gas is hot after falling on to halo, but not necessarily due to shock-heating. For that reason, we decide to use a fixed temperature cut to calculate the hot mode of accretion. Note that $T_{\max}$ is updated whenever the gas particle reaches a higher temperature. However, if the gas particle is star-forming, $T_{\max}$ is not updated, because we impose a lower limit on the temperature of such gas. As in van de Voort et al. (2011), ignoring shocks in the ISM is appropriate because we are interested in the $T_{\max}$ the gas reached before accreting on to the galaxy.

Because the gas temperature after shock-heating can be slightly higher or lower than T_{vir}, we analyse the dependence of $f_{\mathrm{acc,hot}}$ on the fraction of T_{vir} used in the definition of hot accretion. As expected, at fixed halo mass $f_{\mathrm{acc,hot}}$ increases with decreasing fraction of T_{vir}, reaching $f_{\mathrm{acc,hot}} = 0.3$ and 0.6 in 10^{12} M$_\odot$ haloes for $T_{\mathrm{gas}} \ge 0.8 T_{\mathrm{vir}}$ and $T_{\mathrm{gas}} \ge 0.5 T_{\mathrm{vir}}$, respectively. In addition, the virial shock can be located slightly inwards or outwards of R_{200}. We therefore calculated $f_{\mathrm{acc,hot}}$ (using $T_{\mathrm{gas}} \ge 10^{5.5}$ K) for gas crossing $0.8 R_{200}$ and $1.2 R_{200}$. We found that for haloes more massive than 10^{12} M$_\odot$, $f_{\mathrm{acc,hot}}$ is insensitive to the precise value of the radius. In lower mass haloes, the difference between $f_{\mathrm{acc,hot}}$ for gas crossing $0.8 R_{200}$ and $1.2 R_{200}$ can be as high as 0.2dex in $10^{11.5}$ M$_\odot$ haloes.

Fig. 9 shows $f_{\mathrm{acc,hot}}$ calculated using $T_{\max} \ge 10^{5.5}$ K (solid lines) and $T_{\mathrm{gas}} \ge 10^{5.5}$ K (dashed lines) at $z = 0$ (green lines), $z = 2$ (red lines) and $z = 4$ (thick blue lines). For 10^{12} M$_\odot$ haloes, the fraction of shock-heated gas particles with $T_{\mathrm{gas}} \ge 10^{5.5}$ K is a factor of 1.25 lower than $f_{\mathrm{acc,hot}}$ given by $T_{\max} \ge 10^{5.5}$ K at $z = 0$, a factor of 1.6 at $z = 2$ and a factor of 2 at $z = 4$. Although changing the threshold value can bring the $f_{\mathrm{acc,hot}}$ curves into better agreement,
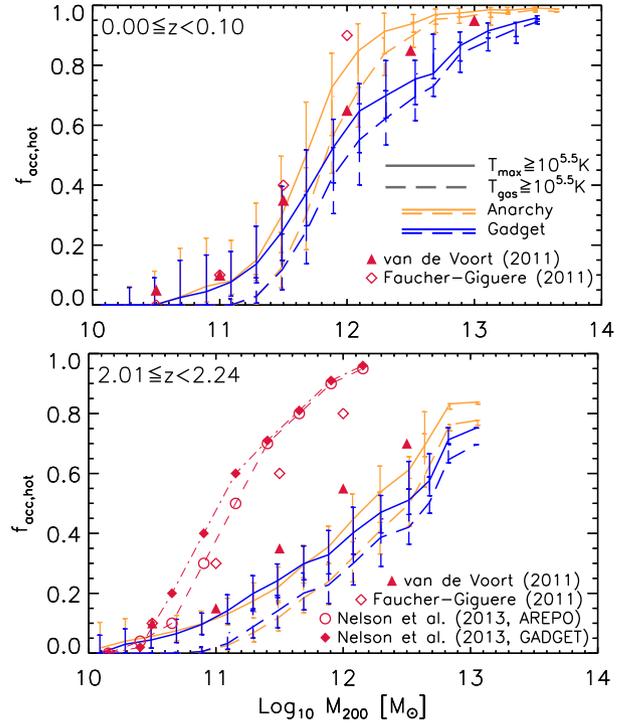

Figure 10. Fraction of gas accreted hot during the redshift ranges $0 \le z < 0.1$ (top) and $2.0 \le z < 2.2$ (bottom) against halo mass. The panels show the same as Fig. 9, but in this case, the curves correspond to the hot fraction calculated from the Ref-L050N0752 Anarchy (orange lines) and GADGET (blue lines) simulations. The error bars show the 1σ scatter of the fractions. The symbols correspond to the hot fraction estimates of van de Voort et al (2011, blue triangles), Faucher-Giguere et al. (2011, open diamonds) and Nelson et al. (2013, open circles).

some disagreement is expected because gas that was heated once may cool later.

T_{gas} and $T_{\max}$ are able to identify the gas particles that are shock-heated when crossing the virial radius; however since we have found that $T_{\max}$ is affected by stellar feedback, we decide to use the gas particle temperature, T_{gas}, after accretion. We find that a lower limit on T_{gas} is the most suitable method to calculate $f_{\mathrm{acc,hot}}$, since it also does not include gas that goes through a shock but cools immediately afterwards and therefore does not contribute to the hot halo formation process.

4.1.1 GADGET and Anarchy

In this section, we extend the discussion presented in Section 2.2 and analyse the differences in the hot/cold modes of accretion on to haloes when the formulation of the hydrodynamic scheme is changed. We compare two L050N0752 simulations that use the same subgrid models, one employs the standard SPH code GADGET, while the other employs the Anarchy hydrodynamic solver used in the fiducial EAGLE runs. Fig. 10 shows the same as Fig. 9 for $z = 0$ (top panel) and $z = 2$ (bottom panel), but the lines correspond to the standard GADGET (blue lines) and Anarchy (orange lines) simulations.

The top and bottom panels show excellent agreement between GADGET and Anarchy in haloes less massive than $10^{11.5}$ M$_\odot$ and 10^{12} M$_\odot$, respectively, and modest differences in larger haloes, irrespective of whether we use the $T_{\max}$ hot gas accretion fractions (solid lines) or the T_{gas} hot gas accretion fractions (dashed lines).

The Anarchy simulation exhibits a somewhat larger fraction of hot accretion on to massive haloes than its GADGET counterpart. This is expected, since the spurious surface tension appearing in the GADGET formulation of SPH prevents the cold dense clumps of gas from being disrupted, mixed and heated when crossing the virial shock (e.g. Schaller et al. 2015).

The top panel of Fig. 10 shows that the T_{max} hot gas accretion fractions taken from the GADGET simulation are in agreement with van de Voort et al. (2011) analysis of the OWLS simulations (Schaye et al. 2010). The bottom panel also shows broad agreement with van de Voort et al. (2011), but substantial differences with Nelson et al. (2013). While van de Voort et al. (2011) used the standard GADGET hydrodynamic solver in their simulations, Nelson et al. (2013) analysed two simulation series that employed either the moving mesh code AREPO (Springel 2010) or standard GADGET, both without stellar, AGN feedback or metal cooling. Nelson et al. (2013) traced the evolution of the gas properties using a Monte Carlo tracer particle technique that enable them to compute T_{max}. They did not find large differences between GADGET and AREPO in the cold mode of accretion on to haloes, and concluded that the cold fraction on to haloes mainly depends on the manner (either with T_{max} or other cut-off temperature) in which it is measured.

The large differences between our work and that of Nelson et al. (2013) is intriguing. Nelson et al. (2013) calculated the accretion rates and so the hot and cold fractions considering only smooth accretion (without including the merger contribution) and over an accretion time window of 1 Gyr. In this work, we did not separate gas accreted smoothly or through mergers and used smaller time windows; however, we find that considering only smooth accretion and/or larger time window does not significantly change the fraction of gas accreted hot (but increases the total rate of gas accretion). In addition, Nelson et al. (2013) used simulations without metal cooling, stellar feedback or AGN feedback. We compared the fraction of hot gas accretion between our reference model and a model without feedback and found that in the simulation without feedback the hot fraction increases from 10 per cent to 20 per cent in the mass range of $10^{12}\,M_{\odot}$ to $10^{12.5}\,M_{\odot}$, and agree in $10^{13}\,M_{\odot}$. However, this increase is not enough to explain the large differences we find with Nelson et al. (2013). Unfortunately, we do not have a model without metal cooling to test whether this plausible explanation is sufficient to account for the remaining differences.

We also compare our results with the work of Faucher-Giguère et al. (2011), who used a series of cosmological simulations run with the standard SPH code GADGET and including stellar feedback and metal cooling but no AGN feedback. They calculated the rates of gas crossing a virial shell, and similar to this work, they used an instantaneous temperature ($T_{gas} > 2.5 \times 10^{5}$ K rather than T_{max}) to separate hot from cold accretion. We find good agreement at $z = 0$, but at $z = 2$, we find large differences. The Faucher-Giguère et al. (2011) transition mass (i.e. the halo mass where the hot mode of accretion equals the cold mode) at $z = 2$ is between $10^{11.3}\,M_{\odot}$ and $10^{11.5}\,M_{\odot}$ (depending on the stellar feedback).

4.2 Hot/cold fraction

The final ingredient for our model of hot halo formation, which we will present in Section 5, is the fraction of gas accreted on to haloes in the hot and cold modes. Throughout this work, we calculate the fraction of hot mode gas accretion, $f_{acc,hot}(M, z)$, using $T_{gas} \geq 10^{5.5}$ K. $f_{acc,hot}$ can be considered as an indirect measure of the presence of hot gas in the halo, since large values of $f_{acc,hot}$ imply large values of M_{hot}/M_{200}. Fig. 11 shows $f_{acc,hot}(M, z)$ at $z = 0–0.1$

(top left-hand panel), $z = 1.0–1.26$ (top right-hand panel) and $z = 2.0–2.24$ (bottom left-hand panel). For each redshift range, we find excellent agreement between the $f_{acc,hot}(M, z)$ curves taken from simulations with different resolution and box size. We find that $f_{acc,hot}(M, z)$ increases smoothly with halo mass and decreasing redshift.

We look for the best-fitting expression for $f_{acc,hot}(M_{halo}, z)$ by performing a least-square minimization. We follow the method described in Section 3.2, we apply equal weighting for each mass bin from the Ref-L100N1504 simulation and minimize the quantity $\Delta_j = \frac{1}{N}\sum_i^N [f_{acc,hot}(M_i, z_j) - F(M_{200,i}, a(z_j), M_{1/2}(z_j))]^2$, where N is the number of mass bins at each output redshift z_j, and F is

$$F = 1/(1 + [M_{200,i}/M_{1/2}(z_j)]^{a(z_j)}). \tag{11}$$

We calculate the best-fitting values of a and $M_{1/2}$ at each redshift $0 \leq z_j < 6$ and for the halo mass range of $10^{10} \leq M_{200} < 10^{14}\,M_{\odot}$. We then look for the best-fitting expressions of a and $M_{1/2}$ as a function of redshift. We find that the relations

$$f_{acc,hot}(M_{200}, z) = 1/(1 + [M_{200}/M_{1/2}(z)]^{a(z)}), \tag{12}$$

$$a(z) = \begin{cases} -1.86 \times 10^{-1.26\bar{z}+1.29\bar{z}^2} & \text{if } 0 \leq z < 2, \\ -0.46 \times 10^{0.81\bar{z}-0.42\bar{z}^2} & \text{if } 2 \leq z < 4, \\ -1.07 & \text{if } z \geq 4, \end{cases} \tag{13}$$

$$\bar{z} = \log_{10}(1 + z), \tag{14}$$

$$M_{1/2}(z) = 10^{12}\,M_{\odot}$$
$$\times \begin{cases} -0.15 + 0.22z + 0.07z^2 & \text{if } 0 \leq z < 2, \\ -0.25 + 0.53z - 0.07z^2 & \text{if } 2 \leq z < 4, \\ 0.72 + 0.01z & \text{if } z \geq 4, \end{cases} \tag{15}$$

best reproduce the fraction of hot mode accretion as a function of halo mass and redshift. The bottom right-hand panel of Fig. 11 compares the fraction of hot accretion in various redshift ranges ($0 \leq z < 6$, symbols) with the best-fitting expression (lines). We find that $f_{acc,hot}$ evolves similarly to $M_{hot}/(\Omega_b/\Omega_m)M_{200}$ (shown in the bottom panel Fig. 6). For all halo masses, $f_{acc,hot}$ increases with time until $z = 1$. At $z < 1 f_{acc,hot}$ increases further in high-mass haloes ($M_{200} > 10^{11.5}\,M_{\odot}$) but decreases in low-mass haloes ($M_{200} < 10^{11.5}\,M_{\odot}$). This is in agreement with van de Voort et al. (2011), who calculated $f_{acc,hot}$ using the T_{max} criterion applied to the OWLS simulations.

We have also investigated how $f_{acc,hot}(M, z)$ is affected when we vary the feedback mechanisms. We find that although the total gas accretion rate on to the halo remains nearly unchanged under varying feedback mechanisms (in agreement with van de Voort et al. 2011), $f_{acc,hot}(M, z)$ increases somewhat at fixed halo mass for the strong stellar feedback and no AGN feedback scenarios. The impact of stellar feedback is largest. For example, strong stellar feedback increases $f_{acc,hot}(M, z)$ by a factor of 1.2 in $10^{12}\,M_{\odot}$ haloes, whereas weak stellar feedback decreases it by a factor of 1.26. Strong AGN feedback decreases $f_{acc,hot}(M, z)$ but only in high-mass haloes and by up to a factor of 1.1. As in van de Voort et al. (2011), we find that the impact of feedback mechanisms on the fraction of hot mode gas accretion is small.

In the next section, we derive an analytic model for hot halo formation. In the model, we assume that the halo develops a hot atmosphere depending on the fraction of hot mode gas accretion and on the amount of hot gas already in the halo.
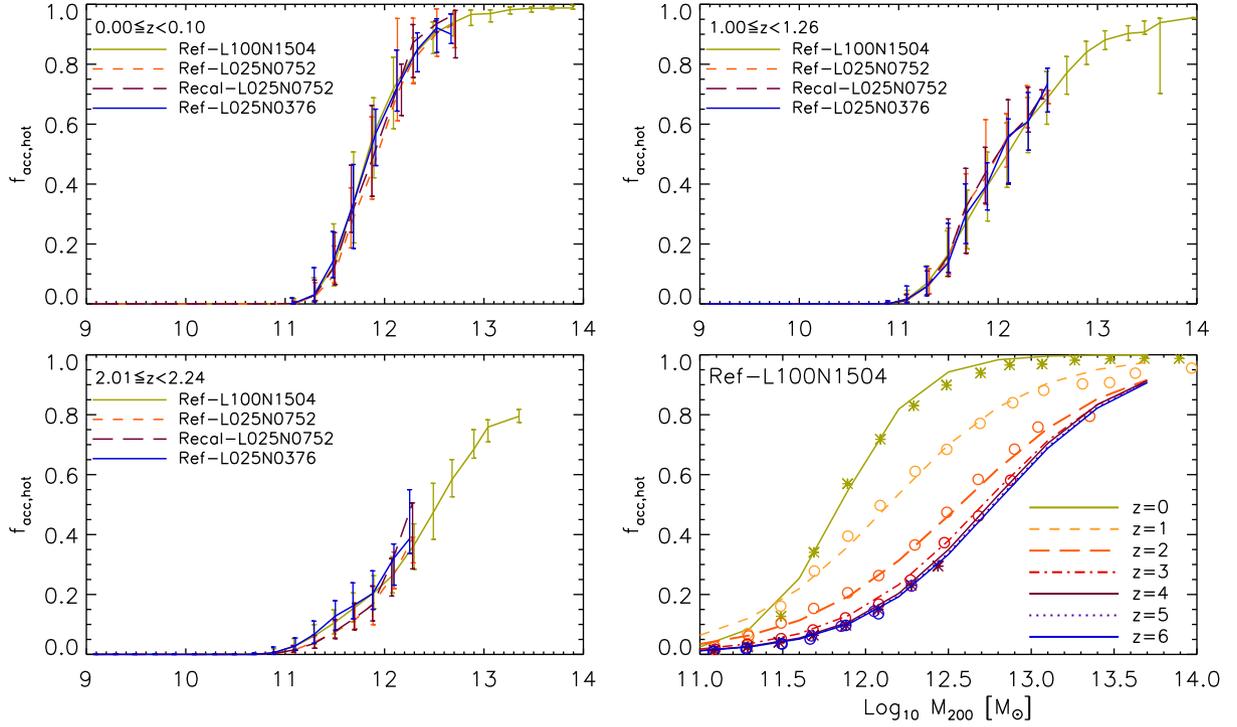
Figure 11. Fraction of hot mode gas accretion during $0 \leq z < 0.1$ (top left-hand panel), $1.0 \leq z < 1.26$ (top right-hand panel), $2.0 \leq z < 2.24$ (bottom left-hand panel) against halo mass. In these panels, the different lines correspond to simulations with different resolutions and box sizes and the error bars show the 1σ scatter. The bottom right-hand panel compares different redshifts (symbols) to the best-fitting expression (lines).

5 TOY MODEL

In Section 3, we investigated the formation of the hot halo in the EAGLE simulations. We found that the development of a strong bimodality in the cooling time PDF of the halo gas provides a clear signature of hot halo formation, which occurs in the halo mass range of $10^{11.5}$–$10^{12}\,\mathrm{M}_\odot$. We noticed however that, even when a stable hot atmosphere has not yet been formed, there is already some hot gas in haloes less massive than $10^{11.5}\,\mathrm{M}_\odot$. This is because gas can be heated by the extragalactic UV/X-ray background or by shocks with stellar or AGN outflows. In this section, we aim to determine the heating rate of gas produced by accretion shocks only, and the halo mass at which this heating overcomes the cooling. To do so, we present an analytic model for the shock-heating rate that takes into account the hot gas mass already in the halo, and the fraction of gas accretion occurring in the hot mode. With the model, we aim to assess the impact of feedback mechanisms (that change the hot gas mass), as well as filamentary cold accretion (that decreases the hot mode fraction of gas accretion), on the formation of a stable hot halo. The model assumes that the hot halo forms when the heating rate produced by accretion shocks is able to balance the radiative cooling rate.

We calculate the variation of the post-shock gas internal energy, $\mathscr{E}$ (in units of erg), due to the transformation of kinetic energy into thermal energy through accretion shocks and due to radiative losses as

$$\dot{\mathscr{E}} = \Gamma_{\mathrm{heat}} - \Gamma_{\mathrm{cool}}, \tag{16}$$

where $\Gamma_{\mathrm{heat}} = \frac{\mathrm{d}}{\mathrm{d}t}(\frac{3}{2}k_{\mathrm{B}}TN_{\mathrm{hot}})$ is the gas heating rate (in units of erg s^{-1}), defined as the variation in time of the thermal energy of the hot gas in the halo, and Γ_{cool} is the net radiative cooling rate (in units of erg s^{-1}). In the definition of Γ_{heat}, N_{hot} is number of hot

gas particles in the halo and T the mean gas temperature, which we assume to be $T = T_{\mathrm{vir}}$. Also, we assume the gas to be monatomic (i.e. the ratio of specific heat is 5/3). Therefore, when

$$\Gamma_{\mathrm{heat}} > \Gamma_{\mathrm{cool}}, \tag{17}$$

the accumulated shock-heated gas at the virial radius gains the necessary pressure through external shock-heating to overcome the energy loss from radiative cooling. We follow DB06 and define a critical mass, M_{crit}, above which haloes develop a hot atmosphere. We define M_{crit} as the halo mass at which the cooling rate, Γ_{cool}, of the hot gas in the halo equals the heating rate, Γ_{heat}, produced by the accretion shocks. In the following subsections, we present the calculations for the critical halo mass.

5.1 Virial heating rate and accretion history

In a ΛCDM cosmology, haloes grow through mergers and smooth accretion. Rapid mass accretion and mergers dynamically heat the gas when haloes form, transforming gravitational potential energy into kinetic energy of baryons and dark matter. For the gaseous component, kinetic energy associated with bulk and turbulent motions is transformed into thermal energy through shocks and viscous dissipation (e.g. Wang & Abel 2007). As a result, the heating rate defined above is driven by the transformation of the gravitational potential energy of baryons and dark matter into thermal energy through $\dot{T}_{\mathrm{vir}}$, and by the accretion rate of gas undergoing shocks through $\dot{N}_{\mathrm{hot}}$, as follows:

$$\Gamma_{\mathrm{heat}} = \frac{3}{2}k_{\mathrm{B}}\dot{T}_{\mathrm{vir}}N_{\mathrm{hot}} + \frac{3}{2}k_{\mathrm{B}}T_{\mathrm{vir}}\dot{N}_{\mathrm{hot}}. \tag{18}$$

We next rewrite equation (18), assuming that $N_{\mathrm{hot}} = M_{\mathrm{hot}}/\mu m_{\mathrm{p}}$ (with $\mu = 0.59$ and invariant), $f_{\mathrm{hot}} = M_{\mathrm{hot}}/[(\Omega_{\mathrm{b}}/\Omega_{\mathrm{m}})M_{200}]$,

$\dot{M}_{\rm hot} = f_{\rm acc,hot}(\Omega_{\rm b}/\Omega_{\rm m})\dot{M}_{200}$ and $\dot{T}_{\rm vir} = \frac{2}{3}\frac{\dot{M}_{200}}{M_{200}}T_{\rm vir}$. Equation (18) then yields

$$\Gamma_{\rm heat} = \frac{3}{2}\frac{k_{\rm B}T_{\rm vir}}{\mu m_{\rm p}}\frac{\Omega_{\rm b}}{\Omega_{\rm m}}\dot{M}_{200}\left[\frac{2}{3}f_{\rm hot} + f_{\rm acc,hot}\right]. \tag{19}$$

The virial temperature of a halo formed at redshift z is related to the total mass M_{200} as

$$T_{\rm vir} = 10^{5.3}\,{\rm K}\left(\frac{M_{200}}{10^{12}\,{\rm M_\odot}}\right)^{2/3}(1+z), \tag{20}$$

where we assumed that the halo encloses a characteristic virial overdensity $\Delta_c = 200$ relative to the critical density at redshift z, $\rho_{\rm crit}(z) = \left(\frac{3H_0^2}{8\pi G}\right)[\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda]$.

To calculate the halo accretion rate, we use the analytic derivation based on Press-Schechter theory from Correa et al. (2015c),

$$\dot{M}_{200}(z) = 71.6\,{\rm M_\odot yr^{-1}}\left(\frac{M_{200}(z)}{10^{12}{\rm M_\odot}}\right)\left(\frac{h}{0.7}\right)[-\tilde{\alpha} - \tilde{\beta}]$$

$$\times(1+z)[\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda]^{1/2}, \tag{21}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ depend on halo mass and the linear power spectrum. This formula gives the accretion rate at redshift z. See Correa et al. (2015a,b,c) for more details on the accretion rate model.

5.2 Cooling rate

To find haloes for which the infalling gas is shock-heated and prevented from cooling on to the inner halo, we compare the mechanical heating rate in equation (19) to the net radiative cooling rate, $\Gamma_{\rm cool}$ (erg s^{-1}),

$$\Gamma_{\rm cool} = M_{\rm hot}\frac{\Lambda(T_{\rm hot}, Z_{\rm hot}, \rho_{\rm hot})}{\rho_{\rm hot}},$$

$$= f_{\rm hot}\frac{\Omega_{\rm b}}{\Omega_{\rm m}}M_{200}\frac{\Lambda(T_{\rm hot}, Z_{\rm hot}, \rho_{\rm hot})}{\rho_{\rm hot}}, \tag{22}$$

where $\Lambda(T_{\rm hot}, Z_{\rm hot}, \rho_{\rm hot})$ (erg cm^{-3} s^{-1}) is the net cooling rate per unit volume and $M_{\rm hot}/\rho_{\rm hot}$ is the volume the hot gas occupies. In the calculation of $\Gamma_{\rm cool}$, we assume that $T_{\rm hot} = T_{\rm vir}$, that the density of the hot gas is $\rho_{\rm hot} = 10^{0.6}\rho_{\rm crit}$ and that the metallicity, $Z_{\rm hot}$, is $Z_{\rm hot} = 0.1\,{\rm Z_\odot}$ (both constant with redshift). The $\rho_{\rm hot}$ and $Z_{\rm hot}$ values were chosen after the analyses of the hot gas density and metallicity, as well as the dependence on the halo mass, which is included in Appendices B and C, respectively. In particular, Appendix C shows that at $z = 0$, the mean density of the hot gas in the halo is around $10^{0.6}\rho_{\rm crit}$ in both $10^{12}\,{\rm M_\odot}$ and $10^{11.5}\,{\rm M_\odot}$ haloes, is slightly higher at $z = 2.2$ than at $z = 0$ and does not change significantly with the host halo mass.

5.3 Critical halo mass for the formation of a hot halo

5.3.1 Analytic estimate

In this subsection, we use the energy condition of post-shock gas given by equation (16) and calculate the critical halo mass, $M_{\rm crit}$, for which the mechanical heating rate, $\Gamma_{\rm heat}$ (equation 19), equals the gas cooling rate, $\Gamma_{\rm cool}$ (equation 22). $M_{\rm crit}$ is the halo mass above which the heating rate exceeds the cooling rate, and as a result the halo develops a stable hot hydrostatic atmosphere. To calculate $M_{\rm crit}$, we assume values for the fraction of hot mode gas accretion ($f_{\rm acc,hot}$), as well as the fraction of hot gas mass in the halo ($f_{\rm hot}$).

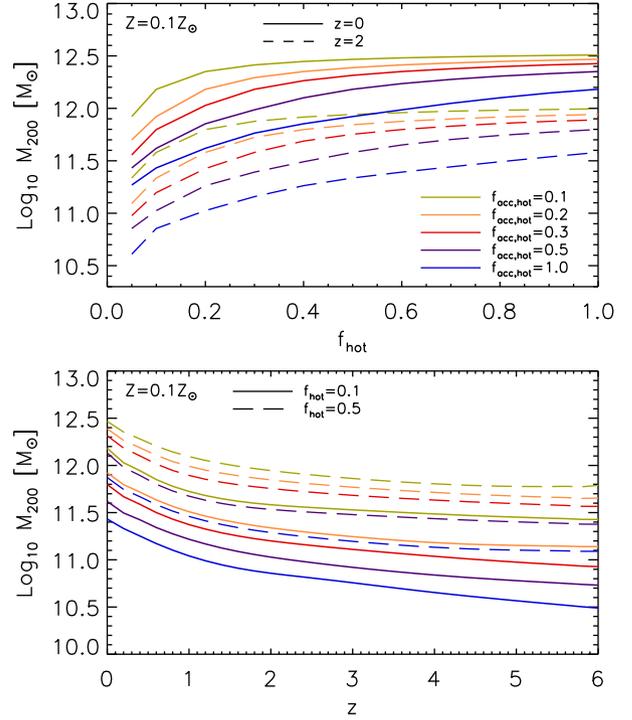The top panel of Fig. 12 shows $M_{\rm crit}$ as a function of $f_{\rm hot}$ and $f_{\rm acc,hot}$

Figure 12. Top panel: Halo mass obtained by equating $\Gamma_{\rm heat}$ and $\Gamma_{\rm cool}$ for redshift $z = 0$ (solid lines) and $z = 2$ (dashed lines) as a function of $f_{\rm hot} = M_{\rm hot}/(\Omega_{\rm b}/\Omega_{\rm m})M_{200}$. The different colour lines correspond to $M_{\rm crit}$ calculated assuming fixed values for the fraction of the hot mode gas accretion ($f_{\rm acc,hot}$, as indicated in the legends). Bottom panel: Same as the top panel, but in this case $M_{\rm crit}$ is calculated assuming $f_{\rm hot} = 0.1$ (solid lines) and $f_{\rm hot} = 0.5$ (dashed lines) as a function of redshift.

for $z = 0$ (solid lines) and $z = 2$ (dashed lines). From the panel, it can be seen that for fixed $f_{\rm acc,hot}$, $M_{\rm crit}$ increases with increasing $f_{\rm hot}$. This is because as $f_{\rm hot}$ increases, so does $\Gamma_{\rm cool}$, and therefore for $\Gamma_{\rm heat}(\propto M_{200}^{2/3}\dot{M}_{200})$ to be able to balance $\Gamma_{\rm cool}$, M_{200} needs to be larger.

For fixed $f_{\rm hot}$, $M_{\rm crit}$ increases with decreasing $f_{\rm acc,hot}$. In this case, when $f_{\rm acc,hot}$ decreases, the heating rate is able to balance the cooling rate only if the accretion rate is large ($\Gamma_{\rm heat} \propto \dot{M}_{200}T_{\rm vir}$), and since the accretion rate increases with halo mass, the halo needs to grow in mass in order to develop a heating rate large enough to keep the gas hot. The top panel of Fig. 12 also shows that for fixed $f_{\rm acc,hot}$ and $f_{\rm hot}$, $M_{\rm crit}$ at $z = 2$ is lower than $M_{\rm crit}$ at $z = 0$. This can also be explained in terms of the halo accretion rate. If $f_{\rm acc,hot}$ and $f_{\rm hot}$ do not change, the heating rate for fixed halo mass still increases because $\dot{M}_{200}$ and $T_{\rm vir}$ increase with increasing redshift. As a result, lower mass haloes are able to produce a heating rate that balances the gas cooling rate.

It is challenging to calculate analytically the hot gas mass in the halo, and the fraction of gas accreted hot, as a function of halo mass and redshift. For that reason, we follow our analysis from Section 3 and make the ansatz that a halo develops a hot atmosphere when the hot gas mass is 10 per cent (which is roughly the hot fraction in haloes with masses between $10^{11}-10^{12}\,{\rm M_\odot}$ in the redshift range of 0–6, see the bottom panel from Fig. 7). In the case of the fraction of hot gas accretion, it is known that for fixed halo mass, $f_{\rm acc,hot}$ is large at low redshift, and it decreases with increasing redshift due to the presence of cold gas accretion from filaments (as shown in Section 4.2). We assume that at $z = 0\,f_{\rm acc,hot} \sim 0.5$–1, and obtain that the mass scale of hot halo formation is between $10^{11.4}$and$10^{11.7}\,{\rm M_\odot}$.

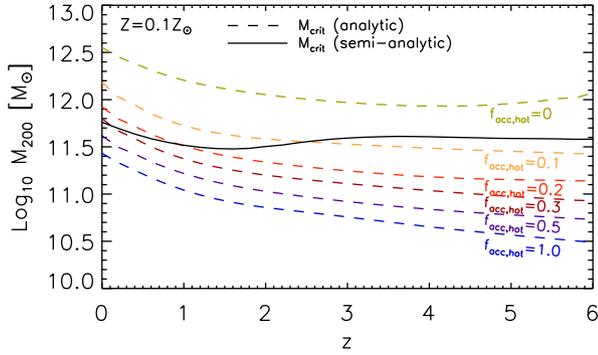**Figure 13.** Halo mass obtained by equating $\Gamma_{\rm heat}$ and $\Gamma_{\rm cool}$, for constant values of $f_{\rm hot}$ and $f_{\rm acc,hot}$ (analytic $M_{\rm crit}$, coloured dashed lines) and for $f_{\rm hot}$ and $f_{\rm acc,hot}$ as a function of halo mass and redshift (semi-analytic $M_{\rm crit}$ using equations 7–10 and 12–15, black solid line).

This is in agreement with the analysis from Section 3, where, by visual inspection of the gas cooling time PDF, we concluded that the hot halo forms between $10^{11.5}$ and $10^{12}\,{\rm M}_\odot$.

We next analyse how $M_{\rm crit}$ changes with redshift. The bottom panel of Fig. 12 shows $M_{\rm crit}$ as a function of redshift for $f_{\rm hot} = 0.1$ (solid lines) and $f_{\rm hot} = 0.5$ (dashed lines). The different colour lines correspond to $M_{\rm crit}$ calculated assuming fixed values for $f_{\rm acc,hot}$ (as indicated in the legend in the top panel). It can be seen from the panel that, for any redshift, the higher $f_{\rm hot}$ and lower $f_{\rm acc,hot}$, the higher $M_{\rm crit}$. It can also be seen that for any $f_{\rm hot}$ and $f_{\rm acc,hot}$ values, $M_{\rm crit}$ remains roughly constant in the redshift range of 6–2, and then increases. This is possibly due to the rapid drop of the accretion rate ($\dot{M}_{200}$, hence $\Gamma_{\rm heat}$) in the redshift range of 0–1 caused by the accelerated expansion of the Universe.

Ocvirk et al. (2008), along with DB06, argued that chemical enrichment has a crucial impact on shock stability, since metallicity, as well as gas density, determines the cooling rate. DB06 and Ocvirk et al. (2008) found that increasing the metallicity increases the critical halo mass for shock stability. We analyse the impact of metallicity on $M_{\rm crit}$ in Appendix A, where we show that increasing metallicity increases $M_{\rm crit}$, but if the hot gas metallicity is lower than $0.1{\rm Z}_\odot$, it does not strongly impact the normalization of $M_{\rm crit}$. This is expected, since metal cooling only becomes important for $Z \gtrsim 0.1{\rm Z}_\odot$ (e.g. Wiersma et al. 2009).

5.3.2 Semi-analytic estimate

In the previous subsection, we used the analytic model to analyse how the mass scale of hot halo formation changes with the fraction of hot gas mass in the halo, $f_{\rm hot}$, and the fraction of gas that shock-heats when crossing the virial radius, $f_{\rm acc,hot}$. Since it is challenging to derive analytically $f_{\rm hot}$ and $f_{\rm acc,hot}$ as a function of halo mass and redshift, we assumed typical values and concluded that the hot halo forms in the halo mass range of $10^{11.4}$–$10^{11.7}\,{\rm M}_\odot$ at $z = 0$ and remains roughly constant with redshift. In this section, we make a 'semi-analytic' estimate of $M_{\rm crit}$ (hereafter $M_{\rm crit,\,sa}$) by using the best-fitting relations from our simulations for $f_{\rm hot}(M_{200}, z)$ (equations 7–10), $f_{\rm acc,hot}(M_{200}, z)$ (equations 12–15). Note that $f_{\rm acc,hot}(M_{200}, z)$ relation may underestimate the fraction of hot gas accretion for haloes with virial temperatures lower than $10^{5.5}\,$K (see Section 4.1 for a discussion). In this section, however, we do not intend to predict a mass scale for hot halo formation, but compare the result with the analysis done in the previous subsection.

Fig. 13 shows the semi-analytic estimate of $M_{\rm crit}$ in black solid lines and the analytic estimates calculated in the previous subsection

in coloured dashed lines. By using the best-fitting relations, we obtain a critical mass scale of $10^{11.75}\,{\rm M}_\odot$ at $z = 0$ that remains roughly constant with redshift in agreement with the analysis done in the previous section.

We find that $f_{\rm acc,hot}(M_{\rm crit,\,sa}) \approx 0.3$ at $z = 0$ and $f_{\rm acc,hot}(M_{\rm crit,\,sa}) < 0.3$ at $z > 0$, from which we conclude that $M_{\rm crit}$ does not necessarily correspond to the mass scale where the hot and cold modes of accretion contribute equally ($f_{\rm acc,hot} = 0.5$), because equal hot/cold modes of accretion do not imply the existence of a stable hot atmosphere or lack thereof. In fact, $f_{\rm acc,hot}(M_{\rm crit,\,sa}) < 0.3$ at $z > 0$ implies that massive haloes are able to develop a hot atmosphere (and hence virial shocks), even when they are accreting the majority of gas in the cold mode.

We also find that $M_{\rm crit,\,sa}$ is somewhat lower than the analytic halo mass calculated by DB06 at $r = R_{200}$. In their work, DB06 found that the critical halo mass of shock-heating occurring in the inner halo at $r = 0.1R_{200}$ is $6 \times 10^{11}\,{\rm M}_\odot$ and $2 \times 10^{12}\,{\rm M}_\odot$ at $r = R_{200}$, in both cases the critical halo mass was nearly constant with redshift. In our case, we focus on shocks occurring at R_{200} and find that $M_{\rm crit,\,sa}$ is a factor of 3.5 lower and changes slightly with redshift. The small change of $M_{\rm crit,\,sa}$ with redshift is driven by the interplay between the accretion rate, $\dot{M}_{200}$, which increases with redshift, and the fraction of hot gas accretion, $f_{\rm acc,hot}$, which decreases with redshift, but it is also due to the fact that we assume a fixed value of $10^{0.6}\rho_{\rm crit}$ for the hot gas density. In Appendix D, we do a more detailed comparison with the work of DB06, and in Appendix C, we discuss how the density of the hot gas in the halo changes with redshift.

5.4 Comparison between $t_{\rm cool}/t_{\rm dyn}$ and $\Gamma_{\rm heat}/\Gamma_{\rm cool}$

In this subsection, we aim to show that the analytical model gives a better prediction for $M_{\rm crit}$ compared to the ratio of the halo dynamical and cooling time-scales. To do so, we compute $\Gamma_{\rm heat}$ and $\Gamma_{\rm cool}$, as well as $t_{\rm cool}$ and $t_{\rm dyn}$, for each halo in the simulations.

In the calculation of $\Gamma_{\rm heat}$, we compute the individual gas accretion rates for each halo, as well as the fraction of particles that shock-heat, and define hot gas as all particles with temperatures larger than the host halo's $T_{\rm vir}$ (to avoid using $t_{\rm cool}/t_{\rm dyn}$ as we do in section 3.3). In the case of $t_{\rm cool}$ and $t_{\rm dyn}$, we assume that the hot halo is formed when the cooling time at the virial radius is larger than the dynamical time, and to calculate them we only use gas between $(0.8 - 1) \times R_{200}$.

The top panel of Fig. 14 shows the ratio between $\Gamma_{\rm heat}$ and $\Gamma_{\rm cool}$ (olive lines) and $t_{\rm cool}$ and $t_{\rm dyn}$ (blue lines). We find that the halo mass where $\Gamma_{\rm heat} = \Gamma_{\rm cool}$ at $z = 0$ is in very good agreement with the semi-analytic prediction ($M_{\rm crit,\,sa} = 10^{11.8}$). At $z = 2$, the halo mass is larger than $M_{\rm crit,\,sa}$ due to the hot gas density being slightly lower than the fiducial value adopted in Section 5.2. We find that at $z = 0$, the median gas $t_{\rm cool}$ is always larger than $t_{\rm dyn}$ in the halo mass range of 10^{11}–$10^{12.5}\,{\rm M}_\odot$, indicating that the hot halo should form, in contradiction with the results from Section 3. Such long cooling times at all masses are due to the presence of additional heating mechanisms (like UV/X-ray background). At $z = 2$, the gas median cooling times at large radii are shorter, due to the cold, dense filamentary gas, and $t_{\rm cool}$ overcomes $t_{\rm dyn}$ only in haloes larger than $10^{12.2}\,{\rm M}_\odot$.

We believe that $\Gamma_{\rm heat} = \Gamma_{\rm cool}$ is a better method to determine when the hot halo forms because, unlike $t_{\rm cool} = t_{\rm dyn}$, $\Gamma_{\rm heat}$ by definition only considers the heating due to halo growth and accretion shocks.
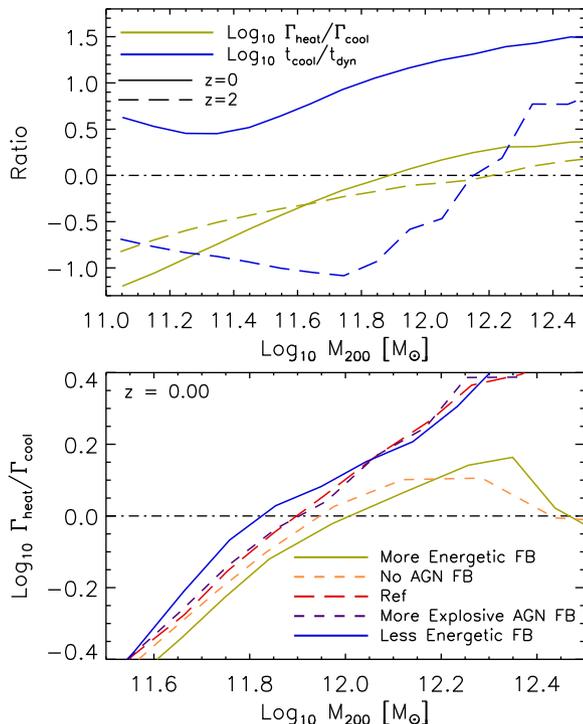
Figure 14. Top panel: median logarithmic ratio between Γ_{heat} and Γ_{cool} (green lines), and between t_{cool} and t_{dyn} (blue lines) as a function of halo mass (M_{200}) for gas in haloes at $z = 0$ (solid lines) and $z = 2$ (dashed lines). Bottom panel: median logarithmic ratio of Γ_{heat} and Γ_{cool} calculated from simulations with different feedback prescriptions as indicated in the legend.

5.4.1 Feedback variations

The model presented in this section assumes that the formation of the hot halo is only driven by the heating from gravitational accretion shocks. However, in the presence of other energy sources, like stars or AGN, the heating rate should increase, and therefore extra terms (like Γ_{stellar} or Γ_{AGN}) should be added to equation (16). Although we do not include extra heating sources accounting for the presence of feedback, we still find good agreement between the analytical results and the numerical analysis. Also, varying feedback may change the metallicity and the hot gas density (e.g. Crain et al. 2013) and hence Γ_{cool}. The sign of the effect is however difficult to predict. On the one hand, a more efficient feedback will reduce the stellar and hence the total metal mass. On the other hand, a greater fraction of the metals may reside in the hot halo gas.

We next investigate how the critical halo mass is affected by different feedback implementations. To do so, we calculate Γ_{heat} and Γ_{cool} from simulations with feedback variations and show these in the bottom panel of Fig. 14. We find that at $z = 0$, in the less energetic stellar feedback scenario M'_{crit} (at which $\Gamma_{\text{heat}} = \Gamma_{\text{cool}}$) is $\sim 10^{11.8} \, \text{M}_{\odot}$, in the more explosive AGN FB $M'_{\text{crit}} \sim 10^{11.9} \, \text{M}_{\odot}$, in the no AGN FB (but moderate stellar feedback) $M'_{\text{crit}} \sim 10^{11.9} \, \text{M}_{\odot}$, and in the more energetic stellar feedback case $M'_{\text{crit}} \sim 10^{12} \, \text{M}_{\odot}$. This is in excellent agreement with the analytic model, which predicts that M_{crit} increases when f_{hot} increases, which occurs when stellar feedback is more energetic and/or there is no AGN feedback. The result is also in good agreement with Section 3.2, where we concluded that the hot halo formation is not as strongly affected by AGN feedback as it is by stellar feedback.

5.5 Comparison with simulated $t_{\text{cool}}/t_{\text{dyn}}$

In this subsection, we compare the semi-analytic predictions for $M_{\text{crit, sa}}$ presented in the previous section with results from our simulations. To do so, we investigate the halo mass range for which haloes develop a hot atmosphere. The model predicts that haloes with masses $M_{200} > M_{\text{crit, sa}}$ form a hot atmosphere. We test this by analysing the PDF of $t_{\text{cool}}/t_{\text{dyn}}$ for haloes with masses $M_{200} \ll M_{\text{crit, sa}}$ (the left-hand panel of Fig. 15), $M_{200} \sim M_{\text{crit, sa}}$ (middle panel) and $M_{200} \gg M_{\text{crit, sa}}$ (right-hand panel) at $z = 0$–4. We select gas particles in the halo (located between $r = [0.15 - 1]R_{200}$) that are not star forming, and use the Ref-L100N1504 simulation.

The left-hand panel of Fig. 15 shows that in haloes with masses $M_{200} \sim 0.1 M_{\text{crit, sa}}$ most of the gas has $t_{\text{cool}} < t_{\text{dyn}}$ and is thus able to cool effectively and accrete on to the galactic disc. In larger haloes ($M_{200} \sim M_{\text{crit, sa}}$, middle panel), the PDF is bimodal with a peak in either side of $t_{\text{cool}}/t_{\text{dyn}} = 1$, indicating that a hot atmosphere has been formed in these haloes at each redshift, despite the increasing contribution of cold gas from filaments at higher redshifts. In haloes with masses $M_{200} \sim 10 M_{\text{crit, sa}}$ (right-hand panel), the increase in shock-heated hot gas enhances the peak at $t_{\text{cool}} \gg t_{\text{dyn}}$ at the expense of the peak at $t_{\text{cool}} \ll t_{\text{dyn}}$. We conclude that the semi-analytic model for hot halo formation captures the mass scale of hot halo formation in the simulations.

6 SUMMARY AND CONCLUSIONS

We have studied the formation of the hot hydrostatic halo, its dependence on feedback mechanisms, on the hot gas mass that is already in the halo, and on the fraction of gas accreted hot using the EAGLE suite of hydrodynamical simulations, as well as analytic calculations.

We began by analysing the PDF of the ratio of the radiative cooling time and the dynamical time for gas in the halo and found that when the hot halo is formed, it produces a strong bimodality in the PDF (Figs 1 and 2, top right-hand panels). By inspection of cooling time PDFs, we found that the mass scale for hot halo formation is $10^{11.5}$–$10^{12} \, \text{M}_{\odot}$ at $z = 0$–4 (Fig. 4).

We found, however, that the cooling time PDF is strongly affected by the extragalactic UV/X-ray background radiation and stellar feedback. The UV/X-ray background radiation suppresses the net cooling rate of gas in the temperature range of $T \sim 10^4$–10^5 K; therefore, the peak of the cooling time PDF shifts towards larger cooling times as the halo's virial temperature decreases to these values (Fig. 3).

In the case of stellar feedback, galactic winds expel gas from the galaxy into the halo, changing the distribution of the hot gas. As a result, more energetic stellar feedback increases the fraction of hot gas. We also analysed the build-up of the total gas mass, M_{gas}, as well as the hot gas mass, M_{hot}, as haloes evolve, and concluded that stellar feedback has a large impact on the amount of gas in the halo. For example, doubling the strength of the stellar feedback increases the gas mass fraction by a factor of 1.3 in $10^{12} \, \text{M}_{\odot}$ haloes relative to the Ref model, whereas halving the strength of the stellar feedback decreases the gas mass fraction by a factor of 2.5 (Fig. 7).

In the case of AGN feedback, neither the bimodality of the cooling time PDF nor the hot gas mass are strongly affected in haloes smaller than $10^{12} \, \text{M}_{\odot}$ since they do not form massive black holes. However, the PDFs and hot gas mass do change at higher halo masses and in a manner opposite from that of stellar feedback. While efficient stellar feedback increases the gas mass in the halo, more explosive AGN feedback decreases it. For example, with strong
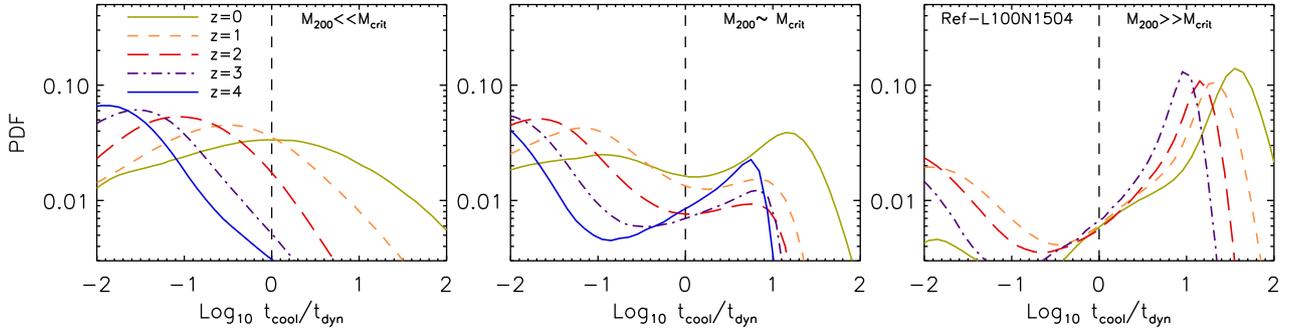
Figure 15. Mass-weighted PDF of the ratio of the cooling and local dynamical time for gas in haloes with masses $M_{200} \ll M_{crit}$ (left-hand panel), $M_{200} \sim M_{crit}$ (middle panel) and $M_{200} \gg M_{crit}$ (right-hand panel) at $z = 0, 1, 2, 3, 4$ and 5, taken from the Ref-L100N1504 simulation.

(without) AGN feedback the total gas mass in the halo decreases (increases) by a factor of 1.5 (Fig. 7). The effect on M_{hot} is slightly different, efficient stellar feedback increases the hot mass fraction by 10 per cent relative to Ref, but no AGN feedback decreases it by 8 per cent in $10^{12}\,M_{\odot}$ haloes. In the case of less energetic stellar feedback and more explosive AGN feedback, the ratio M_{hot}/M_{gas} increases by 10 per cent and 3 per cent (on average) with respect to Ref, respectively, in the halo mass range of $10^{11.5}$–$10^{12}\,M_{\odot}$.

In addition to the hot gas in the halo, we calculated the fraction of gas accretion occurring in the hot mode (Fig. 11), $f_{acc,hot}$. Rather than using a lower limit on the maximum past temperature to select shock-heated gas particles, as done in most previous works, we used the gas temperature after accretion, T_{gas}, which yields lower $f_{acc,hot}$ values than the maximum past temperature. We believe that a lower limit on T_{gas} is a better method to select hot gas accretion, because it excludes the gas that goes through a shock but cools immediately afterwards or that has not passed through an accretion shock but was heated in the past by stellar feedback and has since cooled, and therefore does not contribute to the formation of a hot halo.

We derived an analytic model for hot halo formation that depends on the fraction of hot gas mass that is already in the halo, f_{hot}, as well as on the accretion rates and the fraction of gas accretion occurring in the hot mode, $f_{acc,hot}$. We assumed that a hot halo develops when the heating rate from accretion shocks balances the radiative cooling rate. We computed M_{crit}, the critical mass scale above which the hot halo forms, as a function of f_{hot} and $f_{acc,hot}$. We found that M_{crit} increases with increasing f_{hot} and decreasing $f_{acc,hot}$ (Fig. 12). The analytic model yields a mass estimate of $M_{crit} \approx 10^{12}$–$10^{11.5}\,M_{\odot}$ at $z = 0$, which agrees with the simulation results.

Because estimating f_{hot} and $f_{acc,hot}$ analytically is very challenging, we combined the analytic model with fits to the hot gas mass, and hot mode accretion rates as a function of mass and redshift. We computed a semi-analytic critical mass, $M_{crit, sa}$, and found that $M_{crit, sa} = 10^{11.75}\,M_{\odot}$ at $z = 0$. At higher redshift, $M_{crit, sa}$ remains roughly constant (Fig. 13). We tested the $M_{crit, sa}$ values by inspecting the cooling time PDF of hot gas in haloes with mass $M_{crit, sa}$ and confirmed that at all redshifts, the PDF has a clear bimodal shape (Fig. 15). Note that because the semi-analytic model uses our simulation results as input, its prediction for the mass scale for hot halo formation cannot be tested using the same simulations. We can, however, use it to compare with the analytic analysis.

We compared the ratio of the heating due to accretion (Γ_{heat}, derived in the analytic model) and the radiative cooling (Γ_{cool}) rates of hot gas, with the ratio of the cooling (t_{cool}) and dynamical (t_{dyn}) times of gas at the virial radius, and found that unlike $\Gamma_{heat}/\Gamma_{cool}$, the median t_{cool}/t_{dyn} of gas is always greater than unity in the halo mass

range of 10^{11}–$10^{12.5}\,M_{\odot}$ (Fig. 14). On the contrary, $\Gamma_{heat}/\Gamma_{cool}$ is only greater than unity for haloes more massive than $10^{11.8}\,M_{\odot}$, indicating that this ratio better captures the heating due to halo growth and accretion shocks. We believe that compared with the ratio of cooling and dynamical times, the analytic model of hot halo formation is better at indicating when a hot halo forms.

Finally, we investigated how feedback impacts the hot halo formation mass scale. We calculated $\Gamma_{heat}/\Gamma_{cool}$ using simulations with different feedback prescriptions, and concluded that while a hot hydrostatic atmosphere forms in more (less) massive haloes in scenarios with more (less) energetic stellar feedback, the mass scale of hot halo formation is not strongly affected by AGN feedback. This result is driven by the dependence of Γ_{heat} and Γ_{cool} on the hot gas mass fraction, f_{hot}. When f_{hot} increases (i.e. in the strong stellar feedback scenario), so does the rate of cooling and therefore the halo needs to grow in mass in order to develop a heating rate that overcomes the cooling rate.

Cosmological hydrodynamical simulations have shown that the manner in which galaxies accrete gas depends on the complex interaction between the hot halo, AGN and stellar feedback (see e.g. van de Voort et al. 2011; Faucher-Giguère et al. 2011; Nelson et al. 2015). In Paper II, we will make use of the semi-analytic calculations presented in this work and derive a semi-analytic model for gas accretion on to galaxies that accounts for the hot/cold modes of gas accretion on to haloes and for the rate of gas cooling from the hot halo. By doing so, we aim to provide some insight into the physical mechanisms that drive the gas inflow rates on to galaxies.

REFERENCES

Agertz O. et al., 2007, MNRAS, 380, 963

Anderson M. E., Bregman J. N., 2011, ApJ, 737, 22

Andrews B. H., Martini P., 2013, ApJ, 765, 140

Binney J., 1977, ApJ, 215, 483

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, ApJ, 694, 396

Cantalupo S., 2010, MNRAS, 403, L16

Chabrier G., 2003, ApJL, 586, L133

Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015a, MNRAS, 450, 1514

Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015b, MNRAS, 450, 1521

Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015c, MNRAS, 452, 1217

Crain R. A., McCarthy I. G., Frenk C. S., Theuns T., Schaye J., 2010, MNRAS, 407, 1403

Crain R. A., McCarthy I. G., Schaye J., Theuns T., Frenk C. S., 2013, MNRAS, 432, 3005

Crain R. A. et al., 2015, MNRAS, 450, 1937

Creasey P., Theuns T., Bower R. G., Lacey C. G., 2011, MNRAS, 415, 3706

Cullen L., Dehnen W., 2010, MNRAS, 408, 669

Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2 (DB06)

Dekel A. et al., 2009, Nature, 457, 451

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 465

Efstathiou G., 1992, MNRAS, 256, 43P

Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267

Faucher-Giguère C.-A., Kereš D., Ma C.-P., 2011, MNRAS, 417, 2982

Furlong M. et al., 2015, MNRAS, 450, 4486

Furlong M. et al., 2017, MNRAS, 465, 722

Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays Modelling the UV/X-ray Cosmic Background with CUBA. p. 64

Hopkins P. F., 2013, MNRAS, 428, 2840

Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014, MNRAS, 443, 1173

Hutchings R. M., Thomas P. A., 2000, MNRAS, 319, 721

Katz N., Keres D., Dave R., Weinberg D. H., 2003, in Rosenberg J. L., Putman M. E., eds, The IGM/Galaxy Connection. The Distribution of Baryons at z = 0. Vol. 281 of Astrophysics and Space Science Library, How Do Galaxies Get Their Gas? Kluwer, Dordrecht, p. 18

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Kulkarni V. P., Khare P., Péroux C., York D. G., Lauroesch J. T., Meiring J. D., 2007, ApJ, 661, 88

Lacey C. G. et al., 2016, MNRAS, 462, 3854

Lin Y.-T., Mohr J. J., Stanford S. A., 2003, ApJ, 591, 749

Nagamine K., Springel V., Hernquist L., 2004, MNRAS, 348, 435

Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353

Nelson D., Genel S., Vogelsberger M., Springel V., Sijacki D., Torrey P., Hernquist L., 2015, MNRAS, 448, 59

Nelson D., Genel S., Pillepich A., Vogelsberger M., Springel V., Hernquist L., 2016, MNRAS, 460, 2881

Ocvirk P., Pichon C., Teyssier R., 2008, MNRAS, 390, 1326

Péroux C., Dessauges-Zavadsky M., D'Odorico S., Kim T.-S., McMahon R. G., 2007, MNRAS, 382, 177

Planck Collaboration et al., 2014, A&A, 571, A16

Price D. J., 2008, J. Comput. Phys., 227, 10040

Prochaska J. X., Gawiser E., Wolfe A. M., Castro S., Djorgovski S. G., 2003, ApJl, 595, L9

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Rosas-Guevara Y. M. et al., 2015, MNRAS, 454, 1038

Savaglio S., 2006, New J. Phys., 8, 195

Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015, MNRAS, 454, 2277

Schaye J., 2004, ApJ, 609, 667

Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

Schaye J. et al., 2010, MNRAS, 402, 1536

Schaye J. et al., 2015, MNRAS, 446, 521

Sembolini F. et al., 2016, MNRAS, 457, 4063

Silk J., 1977, ApJ, 211, 638

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., White M., Hernquist L., 2001, ApJ, 549, 681

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

van de Voort F., Schaye J., 2012, MNRAS, 423, 2991

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458

Vogelsberger M., Sijacki D., Kereš D., Springel V., Hernquist L., 2012, MNRAS, 425, 3024

Wang P., Abel T., 2007, e-prints (arXiv:astro-ph/0701363)

Wendland H., 1995, Advances Comput. Math., 4, 389

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574

Wiersma R. P. C., Schaye J., Smith B. D., 2009, MNRAS, 393, 99

Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, ApJ, 771, L19

APPENDIX A: SHOCK ANALYSIS

In this section, we analyse the mass-weighted PDF of T_{max}, T_{gas}, S_{gas}. Fig. A1 shows the PDFs for the redshift interval $0.0 < z \leq 0.1$ (top left-hand panel) and $2.0 < z \leq 2.2$ (top right-hand panel). The curves are coloured according to the colour bars at the top of the figure, which indicate the halo mass (and virial temperature) of the halo that gas is accreted on to. It can be seen that the T_{max} PDF varies with M_{200}, being unimodal in low-mass haloes and bimodal in high-mass haloes. The location of the local minimum of the bimodal distribution also changes with M_{200}, going from $T_{max,min} \sim 10^{5.5}$ K in 10^{12} $M_{\odot}$ haloes to $T_{max,min} \sim 10^6$ K in 10^{14} $M_{\odot}$ haloes. Besides this local minimum, there is a local maximum at $10^{7.5}$ K for all halo masses. This peak is produced by stellar feedback instead of accretion shocks. Some of the gas that is ejected out of the halo due to stellar feedback is eventually re-accreted. However, if it does not reach a temperature larger than $10^{7.5}$ K when crossing R_{200}, T_{max} is not updated, and the gas will be considered hot mode accretion by the maximum temperature criterion. When applying a T_{max} criterion to separate hot from cold accretion, rather than calculating a $T_{max,\,min}$ threshold value that changes with M_{200}, we follow previous works from the literature (e.g. van de Voort et al. 2011; Nelson et al. 2013) and use $T_{max} = 10^{5.5}$ K.

As per T_{max}, the middle and bottom panels of Fig. A1 show that the T_{gas} and S_{gas} PDFs have a bimodal shape. To identify the accreted gas that does not cool immediately after the shock, we analyse the post-shock gas temperature and entropy. We select the gas particles that were accreted during the redshift interval $z_i - z_j$ ($z_i < z_j$), and are hot using a temperature and entropy threshold value. We calculate the gas temperature and entropy mass-weighted PDFs at redshift z_i. We use the local minima of the bimodal distribution
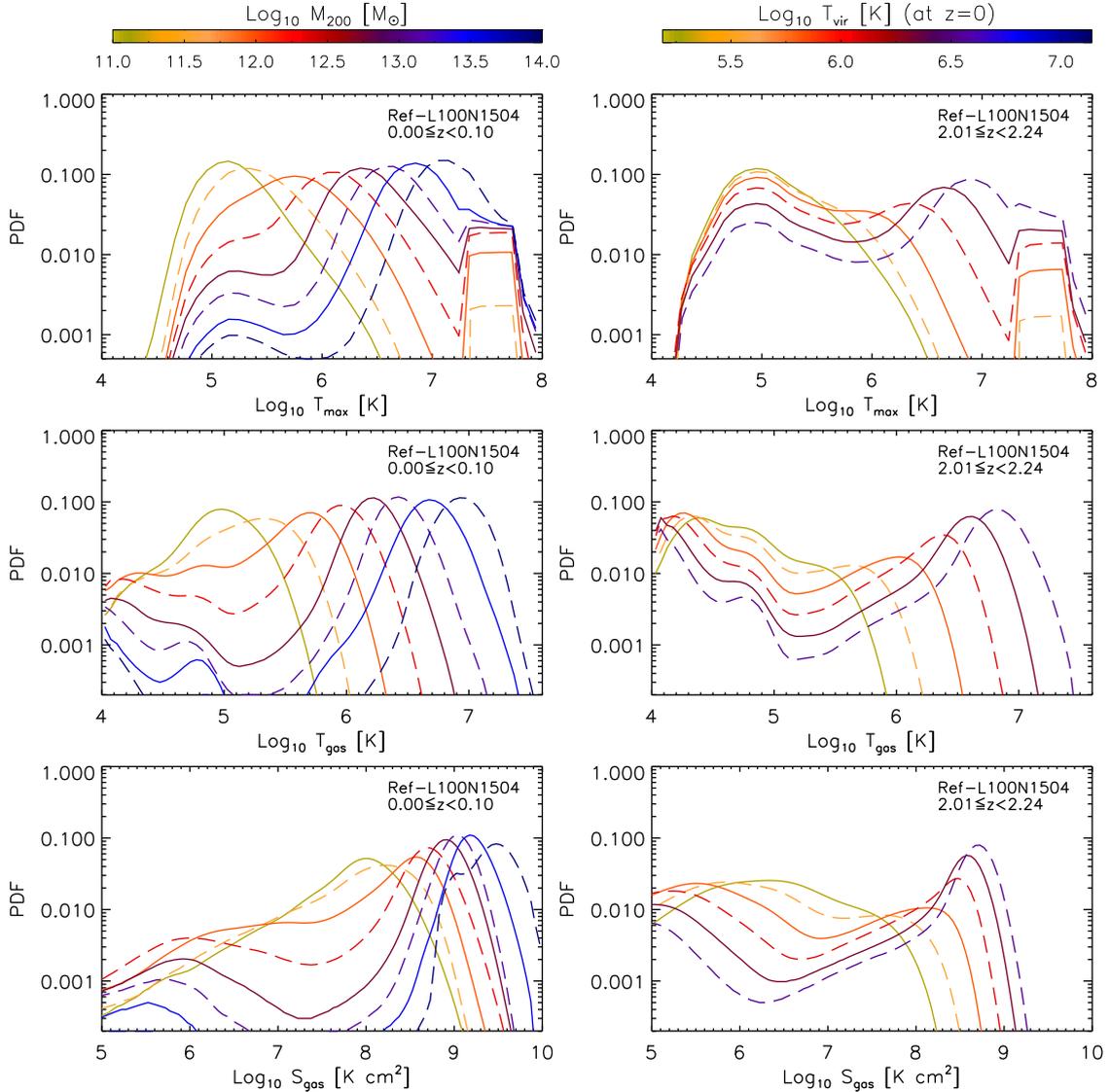
Figure A1. PDF of maximum past temperature (top panels), temperature (middle panels) and entropy (bottom panels) of gas accreted on to haloes in the redshift ranges $z = 0$–0.1 (left-hand panels) and $z = 2.0$–2.2 (right-hand panels). The curves are coloured according to the colour bars at the top of the figure, which indicate the halo mass and virial temperature.

as the threshold values ($T_{min} = 10^{5.5}$ K and $S_{min} = 10^{7.2}$ K cm^2) to calculate the fraction of the gas accreted hot. Our motivation for using the gas entropy (besides temperature) to identify shocked gas is based on the fact that gas generally undergoes a large entropy increase when it encounters a shock (Brooks et al. 2009).

APPENDIX B: METALLICITY

In Section 5, we derived an analytic model for hot halo formation. The model predicts that a hot halo develops when the heating rate from accretion shocks balances the radiative cooling rate. To calculate the gas cooling rate, we assumed a constant hot gas metallicity, $Z_{hot\,gas}$, of $0.1\,Z_\odot$. However, it has been argued that chemical enrichment has a crucial impact on shock stability (e.g. DB06; Ocvirk et al. 2008). Therefore, in this section, we investigate how the hot gas metallicity in haloes changes with halo mass and redshift, and analyse how the halo mass scale, M_{crit}, at which the heating rate balances the gas cooling rate, depends on metallicity.

Using the Ref-L100N1504 simulation, we define hot gas as all gas that is within R_{200} and that has a cooling time longer than the local dynamical time, and calculate the mass-weighted median metallicity per halo mass bin. The left-hand panel of Fig. B1 shows the $Z_{hot\,gas}$–M_{200} relation for various output redshifts (here $Z_{hot\,gas}$ is normalized by solar metallicity, which we assume to be $Z_\odot = 0.0129$). The error bars show the 16–84th percentiles and each bin contains at least five haloes. Interestingly, at fixed halo mass the mass-weighted median metallicity of the hot gas slightly increases with redshift (i.e. by up to a factor of 1.6 in $10^{12}\,M_\odot$ haloes). Many studies of metal abundance have shown that galaxies tend to have lower metallicities at higher redshift (see e.g. Prochaska et al. 2003; Nagamine et al. 2004; Savaglio 2006; Kulkarni et al. 2007; Péroux et al. 2007). It is important to note that the left-hand panel does not show the median metallicity of the ISM (Z_{ISM}), but of the hot mostly ionized gas in the halo. In the case of the ISM, we obtain Z_{ISM}–M_* relations (with M_* stellar mass) in agreement with the galaxy mass–metallicity relation from Andrews & Martini (2013)
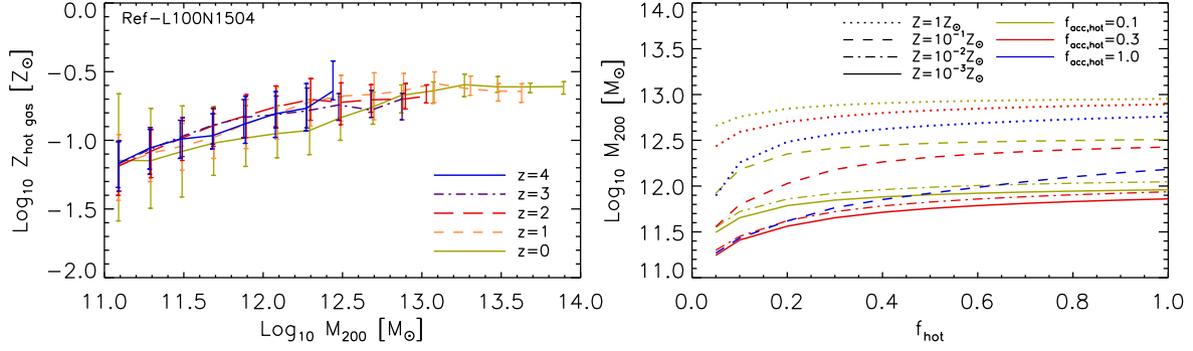
Figure B1. Left-hand panel: Mass-weighted median metallicity of the hot gas in the halo as a function of halo mass for various output redshifts, as indicated in the legends. Hot gas is defined as gas within R_{200} that has a cooling time longer than the local dynamical time. The error bars in the panel correspond to the 16–84th percentiles (i.e. 1σ scatter) and each bin contains at least five haloes. Right-hand panel: Halo mass obtained by equating Γ_{heat} and Γ_{cool} for redshift $z = 0$ as a function of $f_{\mathrm{hot}} = M_{\mathrm{hot}}/(\Omega_{\mathrm{b}}/\Omega_{\mathrm{m}})M_{200}$. The different colour lines and types correspond to M_{crit} calculated assuming fixed values for the fraction of the hot mode gas accretion ($f_{\mathrm{acc,hot}}$) and metallicity, respectively, as indicated in the legends.

and Zahid et al. (2013), as recently shown by Somerville & Davé (2015). As expected, Z_{ISM} at fixed stellar mass decreases with increasing redshift.

We find that for all halo masses, the median mass-weighted metallicity of the gas in the halo increases towards the halo centre (in agreement with Ocvirk et al. 2008; van de Voort & Schaye 2012). Interestingly, we find that the diffuse hot gas in the halo has lower median mass-weighted metallicity (by up to 0.5 dex in haloes larger than 10^{11} M$_{\odot}$) than the cold gas. However, if we define hot gas as all gas within R_{200} that has a maximum past temperature lower than $10^{5.5}$ K, we obtain that the median mass-weighted metallicity of the hot gas is higher (also by to 0.5 dex in haloes larger than 10^{11} M$_{\odot}$), as in van de Voort & Schaye (2012).

We next analyse how M_{crit} changes with metallicity. The right-hand panel of Fig. B1 shows M_{crit} at $z = 0$, calculated assuming constant values of the fraction of hot mode gas accretion, $f_{\mathrm{acc,hot}}$, and metallicity (as indicated in the legends), as a function of the fraction of hot gas mass in the halo, f_{hot}. The panel shows that increasing metallicity increases M_{crit}, but if the hot gas metallicity is lower than 10^{-1} Z$_{\odot}$, it does not strongly impact on the normalization of M_{crit}. This is expected, since metal cooling only becomes important for $Z \gtrsim 0.1$ Z$_{\odot}$ (e.g. Wiersma et al. 2009). The left-hand panel of Fig. B1 shows that the typical metallicity reached by hot gas in the redshift range of 0–4 is ~ 0.1 Z$_{\odot}$ in $\sim 10^{12}$ M$_{\odot}$ haloes, we then find that assuming $Z_{\mathrm{hot\,gas}} \sim 0.1$ Z$_{\odot}$ in the analysis of Section 5.3 is a good approximation.

APPENDIX C: DENSITY

In Section 5, we derived an analytic model for hot halo formation, which considers the heating and cooling rates of gas in the halo. When calculating the cooling rates, we assumed that the hot gas density is about $10^{0.6}\rho_{\mathrm{crit}}$. In this section, we analyse the density and temperature of the hot gas in $10^{11.5}$–10^{12} M$_{\odot}$ haloes.

Fig. C1 shows the temperature versus density of gas in 10^{12} M$_{\odot}$ (top left-hand panel) and $10^{11.5}$ M$_{\odot}$ haloes (top right-hand panel) at $z = 0$, and in 10^{12} M$_{\odot}$ (bottom left-hand panel) and $10^{11.5}$ M$_{\odot}$ haloes (bottom right-hand panel) at $z = 2.2$. The red contours indicate the distribution of the hot gas (i.e. with $t_{\mathrm{cool}} \geq t_{\mathrm{dyn}}$), while the blue contours indicate the cold gas (i.e. $t_{\mathrm{cool}} < t_{\mathrm{dyn}}$). It is interesting to see that there is gas with temperatures below 10^5 K (and a large spread in density) that has a net cooling time longer than the local dynamical time. This gas is close to the equilibrium temperature

obtained when photoheating balances radiative cooling rate. We define the hot halo gas as all gas with densities below $10^2\rho_{\mathrm{crit}}$ and temperature above 10^5 K. At $z = 0$, we find that the mean density of the hot gas in the halo is around $10^{0.6}\rho_{\mathrm{crit}}$ in both 10^{12} M$_{\odot}$ and $10^{11.5}$ M$_{\odot}$ haloes. At $z = 2.2$, the mean density increases slightly to $10^{0.8}\rho_{\mathrm{crit}}$ for both halo masses. We find that the hot gas density does not change significantly with the host halo mass, nor with redshift (we also did the analysis for $z = 3$ and 4), we then assume that the mean density of the hot gas is $10^{0.6}\rho_{\mathrm{crit}}$ and use this value to calculate of the mass scale for the hot halo formation.

APPENDIX D: COMPARISON WITH THE DEKEL AND BIRNBOIM MODEL

DB06 derived a post-shock stability criterion based on the interplay between the cooling time and the compression time. In their derivation, DB06 began by defining the adiabatic index

$$\gamma_{\mathrm{eff}} \equiv \gamma - \rho q/(\dot{\rho}\mathscr{E}), \tag{D1}$$

which they rewrote in terms of the compression time, defined as $t_{\mathrm{comp}} \equiv \Gamma\rho/\dot{\rho}$, with $\Gamma = (3\gamma + 2)/[\gamma(3\gamma - 4)]$ and $\rho = N/V$, and the cooling time, $t_{\mathrm{cool}} = \mathscr{E}/q$ with q the cooling rate, as follows:

$$\gamma_{\mathrm{eff}} = \gamma - \Gamma^{-1}t_{\mathrm{comp}}/t_{\mathrm{cool}}. \tag{D2}$$

They found that the shock is stable if $\gamma_{\mathrm{eff}} > \gamma_{\mathrm{crit}} = 2\gamma/(\gamma + 2/3)$, which is equivalent to $t_{\mathrm{cool}} > t_{\mathrm{comp}}$. Once the cooling time is larger, the pressure gained by compression can balance the loss by radiative cooling, and thus support the shock. In their calculation, $t_{\mathrm{comp}} \propto \frac{r_s}{u} \propto \frac{R_{200}}{V_{\mathrm{vir}}}$, with $r_s \approx R_{200}$, the radius where the spherical shock occurs, and $u \propto V_{\mathrm{vir}}$, the post-shock radial velocity. Thus, obtaining that t_{comp} is comparable to the Hubble time at the corresponding epoch (but at inner radii becomes significantly shorter).

We compare our condition for hot halo formation with that of DB06. We begin by writing equation (16) in terms of t_{cool} and t_{heat}:

$$\frac{\dot{\mathscr{E}}}{\mathscr{E}} = \frac{\Gamma_{\mathrm{heat}}}{\mathscr{E}} - \frac{\Gamma_{\mathrm{cool}}}{\mathscr{E}}, \tag{D3}$$

$$= \frac{\frac{\mathrm{d}}{\mathrm{d}t}(\frac{3}{2}k_{\mathrm{B}}TN_{\mathrm{hot}})}{\frac{3}{2}k_{\mathrm{B}}TN_{\mathrm{hot}}} - \frac{M_{\mathrm{hot}}\Lambda/\rho_{\mathrm{hot}}}{\frac{3}{2}k_{\mathrm{B}}TN_{\mathrm{hot}}}, \tag{D4}$$

$$= \frac{\dot{M}_{200}}{M_{200}}(2/3 + f_{\mathrm{acc,hot}}/f_{\mathrm{hot}}) - \frac{\Lambda}{\frac{3}{2}n_{\mathrm{hot}}k_{\mathrm{B}}T}, \tag{D5}$$

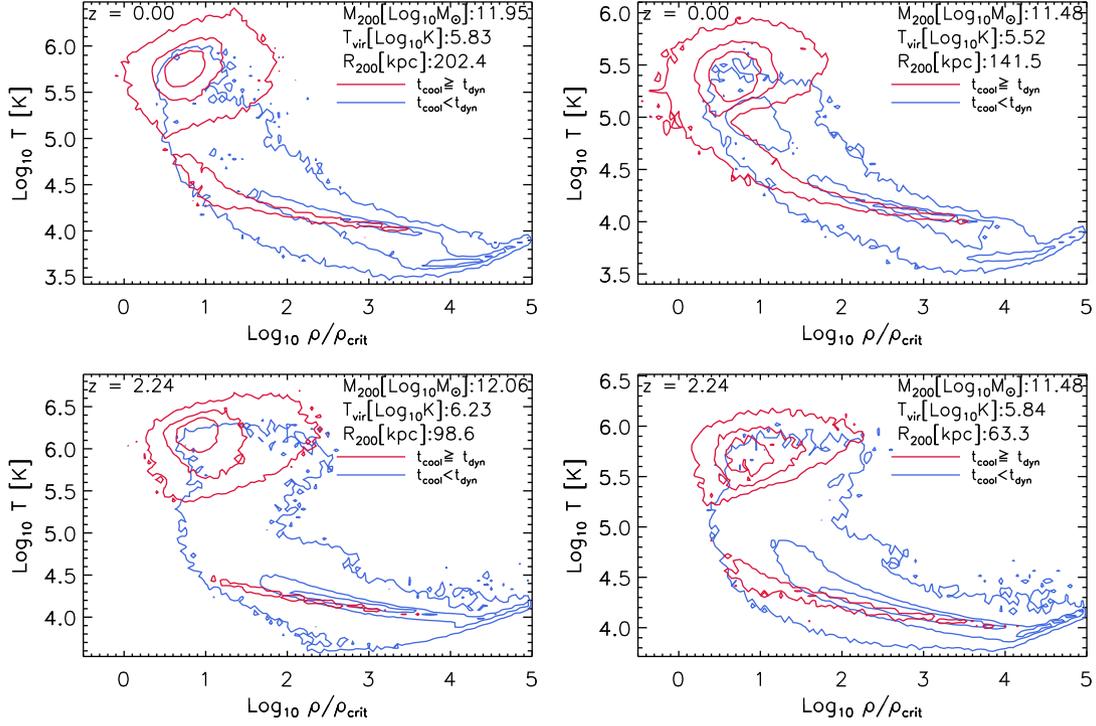$$= t_{\mathrm{heat}}^{-1} - t_{\mathrm{cool}}^{-1}, \tag{D6}$$

Figure C1. Temperature versus density of gas in 10^{12} M$_\odot$ (top left-hand panel) and $10^{11.5}$ M$_\odot$ haloes (top right-hand panel) at $z = 0$, and in 10^{12} M$_\odot$ (bottom left-hand panel) and $10^{11.5}$ M$_\odot$ haloes (bottom right-hand panel) at $z = 2.2$. The contours in the panels enclose 25, 50 and 75 per cent of the distribution and the different colours correspond to the distribution of hot gas (i.e. with $t_{cool} \geq t_{dyn}$, red contours) and cold gas (blue contours). Note that the cooling time is computed from the net cooling rate, i.e. the difference between the photoheating and radiative cooling rate, and depends on metallicity.

where in equation (D3) we divided by $\mathscr{E} = \frac{3}{2}kT_{vir}N_{hot}$.

We find that our model is not equivalent to that of DB06 due to the different redshift dependence of t_{comp} and t_{heat}. While $t_{comp}^{-1} \propto [\Omega_m(1 + z)^3 + \Omega_\Lambda]^{1/2}$, $t_{heat}^{-1} \propto (1 + z)[\Omega_m(1 + z)^3 + \Omega_\Lambda]^{1/2}$. However, we obtain similar results at all redshifts when comparing to DB06's critical mass for the shock at $0.1R_{200}$ (6×10^{11} M$_\odot$). We believe that we have improved upon the DB06 model by introducing a dependence on the amount of shock-heated gas, which we find to decrease with increasing redshift at fixed halo mass due to the presence of cold filaments (see panels in Fig. 9). To include the impact of cold filaments in their calculations, DB06 had to modify the gas

density and assume $\rho_{stream}/\rho_{vir} \sim (3M_*/M)^{-2/3}$ (with ρ_{stream} the filamentary gas density, ρ_{vir} the gas density at the virial radius and M_* the non-linear clustering mass scale). By doing so they obtained an upper limit for cold streams that increases with increasing redshift, in agreement with our results, but reaches 10^{14} M$_\odot$ at $z = 3.5$.

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