Jet-regulated cooling catastrophe

Yohan Dubois,¹ ⋆ Julien Devriendt,¹,2 Adrianne Slyz¹ and Romain Teyssier³,⁴

¹Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH
²Centre de Recherche Astronomique de Lyon, 9 Avenue Charles André, 69561, St-Genis-Laval Cedex, France
³Institute for Theoretical Physics, University of Zurich, CH-8057 Zurich, Switzerland
⁴CEA Saclay, DSM/IRFU/SAP , Bâtiment 709, F-91191 Gif-sur-Yvette, Cedex, France

Accepted 2010 July 9. Received 2010 June 4; in original form 2010 April 11

ABSTRACT
We present the first implementation of active galactic nuclei (AGN) feedback in the form of momentum-driven jets in an adaptive mesh refinement (AMR) cosmological resimulation of a galaxy cluster. The jets are powered by gas accretion on to supermassive black holes (SMBHs) which also grow by mergers. Throughout its formation, the cluster experiences different dynamical states: both a morphologically perturbed epoch at early times and a relaxed state at late times allowing us to study the different modes of black hole (BH) growth and associated AGN jet feedback. BHs accrete gas efficiently at high redshift (z > 2), significantly pre-heating proto-cluster haloes. Gas-rich mergers at high redshift also fuel strong, episodic jet activity, which transports gas from the proto-cluster core to its outer regions. At later times, while the cluster relaxes, the supply of cold gas on to the BHs is reduced leading to lower jet activity. Although the cluster is still heated by this activity as sound waves propagate from the core to the virial radius, the jets inefficiently redistribute gas outwards and a small cooling flow develops, along with low-pressure cavities similar to those detected in X-ray observations. Overall, our jet implementation of AGN feedback quenches star formation quite efficiently, reducing the stellar content of the central cluster galaxy by a factor of 3 compared to the no-AGN case. It also dramatically alters the shape of the gas density profile, bringing it in close agreement with the β model favoured by observations, producing quite an isothermal galaxy cluster for gigayears in the process. However, it still falls short in matching the lower than universal baryon fractions which seem to be commonplace in observed galaxy clusters.

Key words: methods: numerical – galaxies: active – galaxies: clusters: general – galaxies: jets.

1 INTRODUCTION
The well-known overcooling problem in galaxy formation is encountered in numerical simulations for a range of galaxy sizes, spanning dwarves to red ellipticals. Stellar feedback mechanisms, such as winds from young stars and supernovae explosions, are potentially good candidates for expelling large amounts of cold gas from galaxies. Unfortunately, these mechanisms are inadequate for the most massive galaxies, where the energy liberated by star formation activity is too low to unbind material from their gravitational potential wells.

Different theories find ways to partially or totally prevent the cooling catastrophe in the most massive structures: galaxy groups and clusters. Propagating heat with a Spitzer conductivity from the outskirts to the central parts of a cluster, thermal conduction could solve the cooling problem for the more massive clusters (Voigt & Fabian 2004). However, due to the propagation of electrons along magnetic field lines, the conduction is essentially anisotropic leading to instabilities (such as the heat flux buoyancy instability (HBI); see Quataert 2008; Parrish & Quataert 2008) in the cluster core. These can reorient the magnetic field lines and stop the net inflow of heat towards the centre (Bogdanović et al. 2009; Parrish, Quataert & Sharma 2009). Pre-heating of the gas destined to fall into a cluster potential well during proto-galactic stages has also been proposed as a means to empty gas on galactic scales in these massive haloes (Babul et al. 2002).

However, in most models of massive galaxy formation, active galactic nuclei (AGN) play a crucial role in regulating their gas content. Red and dead galaxies commonly exhibit the signatures of supermassive black holes (SMBHs) which are thought to power high-velocity jets into the hot surroundings of the galaxies. Observational evidence for strong AGN activity in groups and clusters is plenty (Arnaud et al. 1984; Carilli, Perley & Harris 1994; McNamara et al. 2001, 2005; Fabian et al. 2002; Birzan et al. 2004; Forman et al. 2007). When spatially resolved, this activity often
takes the form of radio lobes or cavities (Boehringer et al. 1993; Owen, Eilek & Kassim 2000; Birzan et al. 2004; McNamara et al. 2005; Fabian et al. 2006; Taylor et al. 2006; Dong, Rasmussen & Mulchaey 2010; Dunn et al. 2010), or thin extended jets (Brüggen et al. 1994). The radio lobes or cavities are associated with low AGN activity, i.e. the radio mode, and jets with the quasar modes.

AGN feedback is invoked to efficiently suppress star formation in the most massive galaxies either by ejecting gas from their interstellar medium (ISM) into the intra-cluster medium (ICM) or by preventing the ICM gas from collapsing into galactic discs (Binney & Tabor 1995; Rephaeli & Silk 1995). The powerful ejection of gas by AGN is also supposed to suppress the formation of cool cores in some fraction of galaxy clusters, and in turn to make the ICM turbulent (Dubois et al. 2009).

Many of the previous numerical studies that attempt to simulate the growth of black holes (BHs) and their associated AGN feedback have been performed using the Smoothed Particle Hydrodynamics (SPH) technique (Sijacki et al. 2007; Di Matteo et al. 2008; Booth & Schaye 2009). As shown by Mitchell et al. (2009), SPH codes suffer from underestimating the true entropy profile in cluster cores because of their trouble resolving Kelvin–Helmholtz instabilities in regions of strong density contrast (Agertz et al. 2007). As it is generally assumed that BHs accrete gas at a Bondi rate which is related to the local entropy $M_{\text{BH}} \propto K^{-3/2}$, a poor estimate of this entropy leads to an incorrect calculation of the accretion rate on to a BH and, thus, its energy release. To circumvent this issue, we use a sink particle approach to follow the growth and AGN feedback of BHs within an adaptive mesh refinement (AMR) code.

Moreover, in previous cosmological simulations, AGN feedback is modelled with a thermal input of energy. Meanwhile, theoretical work and observations suggest that AGN feedback is mostly mechanical, not thermal. Numerous numerical simulations have implemented and tested the formation and propagation of AGN jets on cluster scales and their impact on the ICM using either idealized simulations (Churazov et al. 2001; Quilis, Bower & Balogh 2001; Reynolds, Heinz & Begelman 2001; Basson & Alexander 2003; Omma et al. 2004; Ruszkowski, Brüggen & Begelman 2004b; Vernaleo & Reynolds 2006; Cattaneo & Teyssier 2007; Simionescu et al. 2009; Gaibler, Krause & Cenamondz 2009; O’Neill & Jones 2010) or cosmological simulations (Heinz et al. 2006; Morsony et al. 2010), but none so far has followed the BH growth self-consistently (i.e. resolving both the BH growth and jet-AGN feedback in cosmological simulation over a Hubble time). As the AGN feedback is tightly linked to its BH growth history, it is of crucial importance to model both the BH evolution through time and its jet-energy release. In this paper, we propose to bridge this gap by performing the first cosmological simulation including a self-consistent treatment of BH evolution and its associated AGN jet-energy release.

This paper is organized as follows. In Section 2, we describe the physical ingredients of our simulation. We start by presenting our cooling and star formation prescriptions and then we introduce the scheme for following the formation and mergers of BHs, gas accretion on to them, and their energy release in the form of jets. In Section 3, we describe our initial condition set-up and our simulation run. In Section 4, we show how the BH growth is linked to the galaxy cluster formation history, and what drives the different modes of AGN feedback (radio versus quasar). In Section 5, we demonstrate that this type of anisotropic mechanical AGN feedback is able to suppress the cooling catastrophe occurring in the galaxy cluster. Finally, in Section 6, we discuss our results.

2 MODELLING THE PHYSICS OF GALAXY FORMATION

2.1 Modelling star formation

Gas in our simulation is allowed to radiate energy by atomic collisions in an H/He primordial gas (Sutherland & Dopita 1993) so that it can collapse into dark matter (DM) potential wells to form galaxies (Silk 1977). To model reionization from $z = 8.5$, heating from an UV background is followed with the prescriptions from Haardt & Madau (1996). Star formation occurs in high-density regions $\rho \geq \rho_0$ ($\rho_0 = 0.1 \, \text{H cm}^{-3}$). When the density threshold is surpassed, a random Poisson process spawns star cluster particles according to a Schmidt law $\dot{\rho}_* = \epsilon \rho/\tau_{\text{ch}}$, where $\tau_{\text{ch}}$ is the gas free-fall time and $\epsilon$ is the star formation efficiency, taken to be $\epsilon = 0.02$ in order to reproduce the observational surface density laws (Kimmcutt 1998). The reader can consult Rasera & Teyssier (2006) and Dubois & Teyssier (2008b) for more information on the star formation implementation.

In this work, we do not model supernova feedback. Several authors have argued that supernovae can only have a dynamical impact on low-mass galaxies (Springel & Hernquist 2003; Dubois & Teyssier 2008b). Thus, as a first-order approximation, we assume that supernova feedback has very little effect on the growth of BHs in massive galaxies as they alone appear incapable of removing a substantial fraction of gas from the ISM. Whilst this simplification allows us to properly isolate the effect of AGN feedback on the surrounding gas from any other galactic feedback mechanism, it does not allow us to study the role of metal enrichment on cooling in the centre of massive haloes. However, we will see that the main features of a cooling catastrophe are already captured with zero metallicity cooling.

Finally, it is possible to view the modification of the temperature at high density $\rho \geq \rho_0$ by a polytropic equation of state (EoS) that we introduce for numerical reasons in Section 2.3 as a way to take into account the thermal effect of the heating of the ISM by supernovae. As a matter of fact, a similar EoS approach is used by other authors (e.g. Springel & Hernquist 2003) as a simple model for the unresolved multi-phase structure of the ISM in cosmological simulations. More specifically, the minimum temperature in dense regions becomes $T_{\text{min}} = T_0 (\rho/\rho_0)^{-1}$, with $T_0 = 10^5 \, \text{K}$, and $n = 4/3$ which leads to a constant Jeans mass $M_J = 1.3 \times 10^6 \, \text{M}_\odot$. Such a value of the polytropic index $n$ roughly compares with the complex functional form of the EoS obtained by analytical considerations on the multi-phase structure of the ISM in Springel & Hernquist (2003).

2.2 SMBHs as sink particles

Sink particles were first introduced by Bate, Bonnell & Price (1995) in a SPH code. Sinks are massive particles that capture gas particles in their surroundings. They mimic the formation of unresolved compact objects, e.g. proto-stellar cores in the ISM, black holes in the ISM, central SMBHs in galaxies, etc. Due to the very Lagrangian nature of the sink particle technique, it had been extensively and exclusively used in SPH codes until Krumholz, McKee & Klein (2004) extended its use to grid codes. The version in RAMSES

1 Here and later, the entropy is defined as $K = T \rho^{-2/3}$, where $T$ is the gas temperature and $\rho$ is the gas mass density.
Silk particles are created in regions where the Jeans criterion is violated, i.e. in regions where the maximum level of refinement is reached and where the gas density is large enough to potentially produce a numerical instability, in other words where

$$\frac{\Delta x}{4} > \lambda_J = \frac{\pi c_s^2}{G\rho}.$$  

(1)

Here $\Delta x$ is the size of the smallest cell, $\lambda_J$ the Jeans length, $c_s$ the sound speed and $\rho$ the gas density. According to Truelove et al. (1997), the numerical stability of a gravitationally bound object is ensured if it is resolved with at least four cells. With a mixed composition of matter (DM, gas, stars), Jeans stability is not trivial anymore, but we can reasonably assume that gas is the dominant source of gravitational potential inside dense collapsed objects, like galaxies, in our case.

For numerical stability, each time that the Jeans criterion is violated we should spawn a sink particle with a mass corresponding to the depleted mass. However, in cosmological simulations this leads to excessively large sink masses. The reason is that the gas is concentrated in structures (galaxies) that are poorly resolved (kpc scale). As a result an entire galactic disc can be defined by only a few Jeans-violating cells leading to massive sinks. To form sufficiently small seed BHs in the centres of the galaxies, we prefer to choose their initial mass, $M_{\text{seed}}$, thereby introducing a free parameter. We set $M_{\text{seed}} = 10^7 M_\odot$ in agreement with previous cosmological simulations (e.g. Booth & Schaye 2009). However, BHs are still spawned only in cells belonging to the maximum level of refinement and that verify equation (1). One consequence of this self-controlled formation of the sink BHs is that they are not allowed to accrete gas when the Jeans criterion is violated. The only way for them to accrete gas is to do so by a reasonable physical process such as the Bondi accretion. With this prescription for initializing the mass of the seed black hole, it is conceivable that gas could be numerically violently Jeans unstable, but this issue is partially solved by the consumption of gas in the star-forming process that temporarily restores gravitational stability.

To get only one BH per massive galaxy, a halo finder is usually run on-the-fly during the simulation to check if candidate galaxies already host a BH (Di Matteo, Springel & Hernquist 2005; Booth & Schaye 2009). We prefer a simpler, more direct and computationally cheaper approach. To avoid creating multiple sinks inside the same galaxy, we ensure that each time a cell could potentially produce a sink (i.e. it verifies equation 1), it is farther than a minimum radius $r_{\text{min}}$ from all other pre-existing sinks. This distance has to be larger than the typical size of galactic discs and smaller than the typical average inter-galactic distance. Test runs suggest that the choice $r_{\text{min}} = 100$ kpc produces very satisfactory results.

To avoid the formation of sink particles in low-density regions that are Jeans unstable, we set a minimum threshold for the density $\rho > \rho_0$ of gas that can create a new sink, where $\rho_0$ is the same density threshold that we use for star formation. In order to avoid the creation of sink black holes before the formation of the very first stars, we check that the star density $\rho_*$ verifies

$$f_* = \frac{\rho_0}{\rho_* + \rho} > 0.25,$$

(2)

before a sink is spawned, where $\rho$ is the gas density.

When a sink particle is finally created it is split into several cloud particles with equal mass. Cloud particles are spread over a $4\Delta x$ radius sphere and positioned every $0.5\Delta x$ in $(x,y,z)$. The exact number of cloud particles in this configuration is therefore $n_{\text{cloud}} = 2109$ per sink. This splitting process is useful in many ways. First, it keeps a heavy sink particle from becoming the dominant gravitational contribution in one single cell. The latter could catapult particles far from their host galaxy by two body encounters. Secondly, it provides a simple canvas over which to compute averaged quantities around the sink, for example we use it to determine the Bondi accretion rate.

Sinks are also allowed to merge together if they lie at a distance closer than $4\Delta x$ from each other. Mass is conserved in this process and momentum vectors of the old sinks are simply added to compute the momentum of the new sink.

Finally, we insist on the fact that sink positions and velocities are updated in the classical way used to update standard particles such as DM particles. No correction on their positions and velocities is done to force them to stay near their host galaxy (as could be done with the halo finder approach). Thus, weakly bound BHs, such as BHs in satellite galaxies of large groups and clusters, may easily be stripped from their host galaxy. These BHs behave like star particles that tidal forces compel to populate the stellar halo of massive galaxies.

### 2.3 Accretion rate on to SMBHs

Since we fail to resolve the accretion disc around our SMBHs, whose size is sub-parsec even for the most massive ones ($\sim 10^{-3}$ pc according to Morgan et al. 2010 from micro-lensing estimates), we use the common prescription that these BHs accrete gas at a Bondi–Hoyle–Lyttleton rate (Bondi 1952)

$$M_{\text{BH}} = \frac{4\pi c_s G^2 M_{\text{BH}}^2 \rho}{(c_s^2 + v^2)^{3/2}},$$

(3)

where $\alpha$ is a dimensionless boost factor ($\alpha \geq 1$), $M_{\text{BH}}$ is the black hole mass, $\rho$ is the average gas density, $c_s$ is the average sound speed and $v$ is the gas velocity relative to the BH velocity. One of the major difficulties encountered with the computation of the relative gas velocity is that in cosmological runs, the ISM is poorly resolved and leads to a very thin scaleheight for galaxies (compared to the resolution). Moreover, due to poor sampling of the gravitational force in the galactic disc, BHs can slightly oscillate in their host galaxy. For this reason a BH close to the centre of a galaxy can feel the infalling material coming from the halo at a relative velocity much higher than the typical velocity inside the bulge. As $v$ is not a reliable quantity, we do not compute $v$ as the gas velocity relative to the sink velocity but we prefer to set it to the average gas velocity dispersion in the ISM which is assumed constant and equal to $\sigma = 10$ km s$^{-1}$ (Dib, Bell & Burkert 2006; Agertz et al. 2009).

The average density $\bar{\rho}$ and sound speed $c_s$ are computed around the BH using the cloud particles for this operation, as mentioned in Section 2.2. To compute the averages, the cell in which each particle sits is assigned a weight given by a kernel function $w$, similar to the one used in Krumholz et al. (2004):

$$w \propto \exp \left(-r^2/r_k^2\right),$$

(4)

where $r$ is the distance from the cloud particle to the sink particle and $r_k$ is the radius defined as

$$r_k = \begin{cases} \frac{\Delta x}{4} & r_{\text{BH}} < \frac{\Delta x}{4}, \\ r_{\text{BH}} & \frac{\Delta x}{4} \leq r_{\text{BH}} \leq 2\Delta x, \\ 2\Delta x & r_{\text{BH}} > 2\Delta x. \end{cases}$$

(5)
The Bondi–Hoyle radius $r_{\text{BH}}$ is given by

$$r_{\text{BH}} = \frac{GM_{\text{BH}}}{c^2},$$

(6)

where $c$ is the exact sound speed in the cell where the sink lies.

The true accretion rate on to the sink is finally limited by its Eddington rate

$$\dot{M}_{\text{Edd}} = \frac{4\pi GM_{\text{BH}} m_p}{\epsilon_f \eta_c},$$

(7)

where $\sigma_T$ is the Thompson cross-section, $c$ is the speed of light, $m_p$ is the proton mass and $\epsilon_f$ is the radiative efficiency, assumed to be equal to 0.1 for the Shakura & Sunyaev (1973) accretion on to a Schwarzschild BH.

The accretion rate is computed at each time-step and a fraction $\dot{M}_{\text{BH}} \Delta t / n_{\text{seed}}$ of gas mass is depleted from the cell where the cloud particle lies and is added to that cloud particle, and its sink mass is updated accordingly. At each coarse time-step cloud particles are re-scattered with equal-mass $\dot{M}_{\text{BH}} / n_{\text{seed}}$. As the time-step does not depend on the accretion speed on to BHs and as low-density cells can be close to high-density cells, a BH might remove more mass than is acceptable. To avoid negative densities and numerical instabilities arising from this, we do not allow a cloud particle to deplete more than 25 per cent of the gas content in a cell.

In such large-scale cluster simulations, it is impossible to resolve the scale and the clumpiness of the ISM. To prevent the collapse of the gas from numerical instabilities and to take into account the mixing of the different phases in the ISM (cold and warm components), we use the polytropic EoS described in Section 2.1. Applying this EoS means that it is impossible to know the ‘true’ density and the ‘true’ sound speed in the ISM; thus the accretion rate on to the BHs must be modified. Previous work modelling the accretion rate on to BHs with such a polytropic EoS set the $\alpha$ parameter to a constant 100 (Springel, Di Matteo & Hernquist 2005; Sijacki et al. 2007; Di Matteo et al. 2008). Here, we follow the prescription from Booth & Schaye (2009) who show that $\alpha = (\rho/\rho_0)^2$ is the best parametric choice to match observational laws.

We stress that this polytropic EoS has important consequences on the accretion rate on to BHs in gas-rich systems: equation (3) turns into $\dot{M}_{\text{BH}} \propto M_{\text{BH}}^{3/2}$, and the temperature dependence is removed. On the other hand, as soon as the cold gas component gets evaporated by star formation or feedback mechanisms from massive galaxies, the accretion rate of the black hole is, by definition, the proper Bondi accretion rate. This $\alpha$ boost of the accretion rate is an artificial way of modelling the very fast accretion of gas within cold and gas-rich galaxies at early epochs, where the clumpiness of the ISM is unresolved in large-scale cosmological simulations.

2.4 AGN feedback modelling

Following Omma et al. (2004), we assume that the primary source of AGN feedback is a sub-relativistic, momentum-imparting bipolar outflow. As discussed in detail by these authors, there are many considerations which support this hypothesis, the most blatant one being that bipolar outflows are observed around virtually all accreting objects in the Universe: stars, black holes and galaxies. Along with these authors, we further assume that the advection-dominated inflow–outflow solution (ADIOS) developed by Blandford & Begelman (1999) to explain the low luminosities of AGNs compared to their estimated Bondi–Hoyle accretion rates is correct. The most important feature of the ADIOS model, as far as we are concerned, is that the bulk of the accretion energy (released as plasma falls into the SMBH) drives a (sub-relativistic) wind from the surface of the accretion disc. We emphasize that, as pointed out in Omma et al. (2004), this bipolar wind is distinct from observed relativistic synchrotron jets which are probably powered by the spin of the SMBH itself; although both jets can simultaneously be present. However the synchrotron jet is very likely irrelevant in terms of AGN feedback since the mechanical luminosity of the sub-relativistic outflow is much higher than the synchrotron luminosity. At present, it is beyond the reach of cosmological simulations to resolve the scales of the accretion disc and to follow the formation process of both the relativistic jet and the sub-relativistic bipolar outflow. For the sake of simplicity we denote the sub-relativistic collimated outflow that we are modelling here as a jet. Finally, we note that in very dense environments it is probable that much of the accretion energy is radiated away, driving outflows with velocities $\approx 0.1c$ in objects with photon luminosities also on the order of the Eddington luminosity (King & Pounds 2003). We argue that as more and more shock-heated rarefied gas fills the central parts of the growing host halo, more and more energy will come out as mechanical energy. More specifically, we assume that a fraction $\epsilon_f$ of the radiated energy is imparted to the ambient gas

$$E_{\text{AGN}} = \epsilon_f L_\nu = \epsilon_f \epsilon_r \dot{M}_{\text{BH}} c^2,$$

(8)

in the form of a kinetic jet.

Such an implementation has the advantage of continuously releasing energy without radiating it away by cooling on the scale of a hydro time-step. Indeed, this problem plagues supernovae feedback modelling where energy is generally injected in thermal form (Navarro & White 1993) and leads to the feedback having no dynamical impact on the surrounding gas. To bypass this issue, some authors release AGN thermal energy only after a sufficient amount of gas has been accreted on to the BH so as to more severely impact the ambient medium (Sijacki et al. 2007; Booth & Schaye 2009; Teyssier et al. 2010). In contrast, we model the AGN feedback as a jet-like structure with the same momentum profile defined in Omma et al. (2004). This model has already been used to follow the self-consistent BH growth and its energy release in an isolated galaxy cluster (Cattaneo & Teyssier 2007; Dubois et al. 2009). Mass, momentum and energy are spread over a small cylinder of radius $r$ and height $2h$ ($h$ for one side of the jet) multiplied with a kernel window function

$$\psi(r_{\text{cyl}}) = \frac{1}{2\pi r_f^2} \exp\left(\frac{-r_{\text{cyl}}^2}{2r_f^2}\right),$$

(9)

where $r_{\text{cyl}}$ is the distance to the axis of the cylinder. The mass deposition follows

$$M_{f}(r_{\text{cyl}}) = \frac{\psi(r_{\text{cyl}})}{\|\psi\|} \eta M_{\text{BH}},$$

(10)

where $\|\psi\|$ is the integrated value of $\psi$ over the whole cylinder, and $\eta = 100$ is an arbitrary value that represents the mass loading factor of the jet on unresolved scales. Mass is transferred from the central cell (where the sink lies) to all the enclosed cells within the jet. Momentum $q_j$ is deposited in opposite directions from the centre along the jet axis, according to

$$\|q_j\| = M_j(r_{\text{cyl}}) \|u_j\| = \frac{\psi(r_{\text{cyl}})}{\|\psi\|} M_{\text{BH}} \sqrt{2\epsilon_f \epsilon_r c} J \cdot dr,$$

(11)

where $\|u_j\| = (2\epsilon_f \epsilon_r)^{1/2} c$ is the velocity of the jet ($\|u_j\| \approx 10^8 \text{ km s}^{-1}$ for $\epsilon_f = 1$), $J$ is the unit spin vector of the BH which defines the jet axis and $dr$ is the distance vector from the centre of
the black hole, $j$ is computed by adding the different contributions from the neighbouring cells (sampled with the cloud particles) to the total angular momentum
\[ J = \sum_{i \in \text{clouds}} m_i \mathbf{r}_i \times \mathbf{u}_i, \]
where $m_i$ and $\mathbf{u}_i$ are the mass and velocity of the gas in the cell harbouring the cloud particle $i$, so that $j = J/|J|$. Finally the kinetic energy deposited within a cell is
\[ \dot{E}_j(r_{\text{cyl}}) = \frac{\dot{\mathbf{q}}_j^2(r_{\text{cyl}})}{2M_j(r_{\text{cyl}})} = \frac{\psi(r_{\text{cyl}})}{\|\dot{\psi}\|} \dot{E}_{\text{AGN}}. \]
Integrating this energy deposition over all the cells within the jet, we recover $\dot{E}_{\text{AGN}}$.

We point out that our jet has no opening angle and should therefore propagate along a straight line as it is perfectly collimated. Omma et al. (2004) have shown that this kind of jet stays collimated over quite long distances (100 kpc) compared to its initial broadness and length (1 kpc). It broadens as it reaches equilibrium with the surrounding hot ambient medium. The same behaviour is also pointed out by Cattaneo & Teyssier (2007) and Dubois et al. (2009), but with the difference that when strong turbulent motions begin to develop in the cluster core due to the formation of a cooling flow, the jet is more quickly mixed with the ICM. The choice of the jet velocity input $|\mathbf{u}_i| \sim 10^4$ km s$^{-1}$ (or equivalently the mass loading factor $\eta = 100$) is particularly arbitrary but based on earlier work by Omma et al. (2004), Cattaneo & Teyssier (2007) and Dubois et al. (2009). The same simulation performed with $|\mathbf{u}_i| \sim 3 \times 10^4$ km s$^{-1}$ produces results in very strong agreement with the ones presented here (mass of the most massive BH and stellar mass of the central galaxy agrees within a few per cent) suggesting that even strong variations of $|\mathbf{u}_i|$ keep our results unchanged.

We set $r_j$ and $\rho_j$ equal to $\Delta x$, and the energy efficiency $\epsilon_j = 1$ so as to reproduce the $M_{\text{BH}}-M_*$ and $M_{\text{BH}}-\sigma_*$ observational relations. Larger values of $r_j$ and $\rho_j$ have been tested at the same resolution $\Delta x \approx 1$ kpc in a cosmological simulation (as opposed to a resimulation like the one presented in this paper) and they produce BHs which are too massive with respect to their host galaxy. Note that our total efficiency $\epsilon_j, \epsilon_i = 0.1$ is also in good agreement with the average value obtained by general relativistic magnetohydrodynamic numerical simulations of the accretion-ejection mechanism in accretion discs around spinning BHs (e.g. De Villiers et al. 2005, Hawley & Krolik 2006 or Benson & Babul 2009 and references therein). Lower $\epsilon_i$ values again cause black holes to become more massive, overshooting the $M_{\text{BH}}-M_*$ observational relation.

3 SIMULATION SET-UP

The simulations are run with the AMR code RAMSES (Teyssier 2002). The evolution of the gas is followed using a second-order unsplit Godunov scheme for the Euler equations. The Riemann solver used to compute the flux at a cell interface is the acoustic solver using a first-order MinMod Total variation diminishing scheme to reconstruct the interpolated variables from their cell-centred values. Collisionless particles (DM, stars and sink particles) are evolved using a particle-mesh solver with cloud-in-cell (CIC) interpolation.

We assume a flat Λ cold dark matter ($\Lambda$CDM) cosmology with total matter density $\Omega_m = 0.3$, baryon density $\Omega_b = 0.045$, dark energy density $\Omega_\Lambda = 0.7$, fluctuation amplitude at $8 h^{-1}$ Mpc $\sigma_8 = 0.90$ and Hubble constant $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$ that corresponds to the Wilkinson Microwave Anisotropy Probe (WMAP) 1-year best-fitting cosmology (Spergel et al. 2003). The simulations are performed using a resimulation (zoom) technique: the coarse region is a $128^3$ grid with $M_{\text{DM}} = 2.9 \times 10^{10}$ M$_\odot$ DM resolution in a $80 h^{-1}$ Mpc simulation box. This region contains a smaller 256$^3$ equivalent grid in a sphere of radius 20 $h^{-1}$ Mpc with $M_{\text{DM}} = 3.6 \times 10^8$ M$_\odot$ DM resolution, which in turn encloses the final high-resolution sphere with radius 6 $h^{-1}$ Mpc, 512$^3$ equivalent grid and $M_{\text{DM}} = 4.5 \times 10^8$ M$_\odot$ DM resolution. Fig. 1 shows the distribution of DM and the distribution of stars in the zoom region and in the galaxy cluster at $z = 0$.

The smallest region is the resimulation zone where cells may be refined up to $\ell_{\text{max}} = 16$ levels of refinement, reaching 1.19 $h^{-1}$ kpc, following a quasi-Lagrangian criterion: if more than eight DM particles lie in a cell, or if the baryon mass exceeds eight times the initial DM mass resolution, the cell is refined. This strategy allows AMR codes, such as RAMSES, which use CIC interpolation in their gravity solver, to avoid propagating discreetness noise from small scales (Romeo et al. 2008). A Jeans length criterion is also added to ensure the numerical stability of the scheme on all levels $\ell < \ell_{\text{max}}$ (Truelove et al. 1997), and where $\delta \rho = \rho/\bar{\rho} > 10^2$: the cells fulfilling these conditions must sample the local Jeans length with more than four cells. We point out that the $\ell_{\text{max}} = 16$ level of refinement is only reached at $a_{\text{exp}} = (1 + z)^{-1} = 0.8$, and that the actual maximum level of refinement for a given redshift is increased as the expansion factor grows with time, i.e. $\ell_{\text{max}} = 15$ at $a_{\text{exp}} = 0.4$, $\ell_{\text{max}} = 14$ at $a_{\text{exp}} = 0.2$, etc. This allows us to resolve the smallest scales with a roughly constant physical size ($0.95 < \Delta x < 1.9 h^{-1}$ kpc), rather than a constant comoving size.

The resimulated region tracks the formation of a galaxy cluster with a 1:1 major merger occurring at $z = 0.8$. Indeed throughout its formation, the cluster chosen for resimulation passes through different dynamical stages: both a morphologically perturbed epoch occurring at half the age of the Universe, and a relaxed state at late times, which permits us to study the different associated states of the BH self-regulated growth. Fig. 2 shows the DM halo merger tree history for this cluster. Haloes and sub-haloes are identified and followed using the Most massive Sub-node Method (MSM) algorithm described in Tweed et al. (2009), which isolates bound substructures. The cluster experiences a major halo merger at $z \approx 1.7$ and two proto-cluster progenitors merge together earlier at $z \approx 3.1$. These mergers end at $z \approx 0.8$ and $z \approx 1.7$, respectively, when the central galaxies hosted in the (sub)haloes merge. Central BHs hosted by these galaxies merge later, as shown in Fig. 3. At later times, most of the mass growth of the cluster occurs through diffuse accretion or minor mergers. In the following, we discuss how such events might trigger or halt the AGN activity of the central (most massive) BH.

4 GROWTH OF SMBHs AND THEIR ACTIVITY

The growth of a BH is tightly linked to the accretion history of its host halo (cf. coeval growth scenario advocated by Miller et al. 2006, Hopkins et al. 2007). In principle, cold gas which flows directly into the central nucleus in a free-fall time will very efficiently grow BHs. However, this rapid growth might be substantially reduced by AGN activity that could expel both energy and material in the vicinity of the BH. In this case, self-regulation of the BH growth possibly drives the relations observed between BH mass and their host galaxy properties (Magorrian et al. 1998; Tremaine et al. 2002; Haring & Rix 2004).

At high redshift ($z > 2$), most galaxies seem to harbour a massive cold gas disc component, both in cosmological hydrodynamic simulations (Ocvirk, Pichon & Teyssier 2008) and in the
Figure 1. Projected DM density (upper panels) and stellar density (bottom panels) in the resimulated 14 Mpc on a side region (left-hand panels) and in the cluster, 1.8 Mpc on a side region (right-hand panels). The red circle in the upper-right panel shows the $r_{500} = 940$ kpc radius of the DM halo at $z = 0$, and green circles in the lower-right panel show the virial radii of the stellar structures as detected by the halo finder.

observations (Shapiro et al. 2008), which can be tapped to fuel rapid BH growth. Accordingly, in the cosmological re-simulation of a cluster presented here, the central BH reaches a few tenths of its final mass when the Universe is less than 2 Gyr old, accreting at a rate above $\sim 1$ per cent of its Eddington limit (Fig. 3). As the initial seed mass of the black hole is $10^5 M_\odot$, this means that its mass increases by a factor $10^4$ in a tenth of the Hubble time. In Fig. 3, the growth of two of its most massive BH progenitors (indicated by red dashed and blue dot-dashed lines in the figure) is also displayed until they merge with the final BH ($M_{\text{BH}} = 1.7 \times 10^{10} M_\odot$ at $z = 0$). These mergers (vertical dotted lines in upper panel of Fig. 3) coincide with two important halo mergers in the history of the cluster (branches labelled 2 and 3 in Fig. 2). At high redshift ($z > 2$), these BH progenitors behave like the main one: they accrete gas at a fraction of their Eddington rate and this fraction steadily decreases with time from $z = 4$ to $z = 2$. We know (Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Ocvirk et al. 2008) that for the most massive haloes [those with masses $M_{\text{stream}} \geq 6 \times 10^7 (1 + z)^3 M_\odot$ at $z \geq 2$], cold accretion of gas from the IGM is efficiently thermalized at a few virial radii by an accretion shock, and as a result we expect the accretion rate in the centre of the halo to drop. From Fig. 2 one can see that the DM haloes of branches 2
Jet-regulated cooling catastrophe

Figure 2. DM halo merger history of the resimulated galaxy cluster. Only the eight most massive branches and sub-branches are shown. Haloes are designated with light blue circles and sub-haloes with dark blue squares. The final merger between a sub-halo and its host halo is represented by a solid line connecting both objects. Note that the cluster main branch 1 experiences two major mergers, one with branch 2 at \( z \sim 3.1 \) and one with branch 3 at \( z \sim 1.7 \). However, these mergers are completed at \( z \sim 1.7 \) and \( z \sim 0.8 \), respectively, epochs which coincide with the merger of the central galaxies hosted in these haloes. Central BHs merge later (see Fig. 3).

and 3 have masses \( \approx 10^{13} M_\odot \) at redshift \( z = 3 \) and therefore satisfy the \( M_{\text{stream}} \) criterion. On that account we claim that this explains, in part, the decrease of the BH accretion rate relative to its Eddington rate.

Mergers are a non-negligible growing mode at intermediate and lower redshift, as there is less cold gas to feed the BH in the massive cluster. Indeed, Fig. 3 shows that the BH doubles its mass at redshift \( \approx 1.6 \) when two BHs of comparable mass (\( M_{\text{BH}} = 3.1 \times 10^9 M_\odot \) and \( M_{\text{BH}} = 8.2 \times 10^8 M_\odot \)) coalesce. The extra amount of mass (\( \approx 2 \times 10^9 M_\odot \)) comes from the fast accretion of material brought in by the galaxy major merger (\( M_* = 1.3 \times 10^{12} M_\odot \) and \( M_* = 6.1 \times 10^{11} M_\odot \)). This merger appears in the DM merger tree (Fig. 2) when branches 1 and 2 join at redshift 1.6. It is interesting to note that the first BH that forms is not necessarily the most massive one at late times (in our case the BH hosted by branch 2 halo forms first), as already pointed out by Di Matteo et al. (2008).

Figs 4, 5 and 6 show three different episodes of the formation and evolution of the cluster, respectively a high-redshift major merger between two gas-rich galaxies (Fig. 4), the major merger of the two clusters (Fig. 5), and the relaxation of the cluster at late times (Fig. 6).

The halo merger between branches 1 and 2 at \( z = 3.1 \) results in a cataclysmic episode for its host galaxies at \( z \approx 1.6 \): a large amount of gas is expelled far from the core of the halo, reaching the virial radius and the resulting disc of gas from the two merging galaxies is almost completely disrupted. This sequence of events is illustrated in Fig. 4. On the left-hand panel, we observe the encounter of the two gas-rich galaxies before they merge. Shortly after they merge (Fig. 4, middle panel), their respective BHs do so as well which results in a strong jet that disrupts most of the cold baryon content in the galaxy and shock heats the ambient medium to high temperature. The jet propagates supersonically at Mach 3 (\( u_{\text{jet}} \approx 3000 \text{ km s}^{-1} \) and \( c_s \approx 1000 \text{ km s}^{-1} \)) before being stopped by the intergalactic medium at \( r \approx 1.2 \text{ Mpc} \), which corresponds to about \( 3r_{\text{vir}} \) at this redshift (Fig. 4, right-hand panel).

The disruption of cold material by AGN feedback has already been noted by Di Matteo et al. (2005) in idealized simulations of a gas-rich merger. It is comforting to confirm their results within a cosmological setting. Khalatyan et al. (2008) have also pointed out that mergers could trigger a high level of AGN activity during the formation of a small galaxy group. Finally, we see two hotspots during the jet propagation that look like radio lobes (Fig. 4,

Figure 3. Mass evolution (upper panel) and accretion rate relative to the Eddington limit (bottom panel) of the most massive BH (black solid lines), and two of its most massive progenitors involved in the episodes illustrated in Figs 4 and 5, i.e. central BH of branch 2 (blue dot-dashed lines) and branch 3 (red dashed lines) of the merger tree (Fig. 2). Black vertical dotted lines in the upper panel mark the mergers of the progenitors with the most massive BH.
Figure 4. Projected temperature in linear scale along the $z$ direction at $z = 1.76$, $z = 1.56$ and $z = 1.49$ from left to right of the most massive cluster progenitors. Branch 1 of the merger tree corresponds to the red circle and branch 2 to the blue circle (see Fig. 2). This redshift sequence corresponds to pre-central galaxy merger, central galaxy and BH mergers and post-central galaxy merger, respectively. The red saturated region corresponds to a temperature $T \approx 2.5-3$ keV. The size of the images is 3.6 Mpc in comoving units.

right-hand panel). Such events (strong jets following a merger) become rarer as time goes on since the combined action of star formation and early AGN activity strongly diminishes the cold and dense gas content in massive haloes.

To check how this striking result is affected by our limited spatial resolution, we performed the same simulation with one more level of refinement ($\Delta x = 0.6 h^{-1}$ kpc). The same burst appears at the same redshift but its power is slightly lower, because more gas has been pre-heated by a strong AGN activity in a previous merger taking place at higher redshift $z \approx 4$. This is not a very surprising effect: with more resolution, the density contrast is more pronounced especially in poorly resolved high-redshift galaxies, which, in turn, leads to a faster accretion rate at early times. As in our standard run, the cold gas in the core of the halo is strongly disrupted by the AGN activity triggered by the wet merger occurring at $z = 1.6$, so that BH growth and AGN luminosity are suppressed for 3 Gyr (see Fig. 7; from $z = 1.6$ to $z = 0.6$). Therefore, we can reasonably claim that most of the important features describing the BH growth are already captured by our standard resolution run.

Fig. 7 also shows the energy that would be released by supernova feedback if it were implemented in our simulation. To estimate this energy release, we assume that stars are distributed according to a Salpeter (1955) initial mass function, for which each massive star ($M_\star > 8 M_\odot$) deposits $10^{51}$ erg per $10 M_\odot$ into the ISM. We see that the energy from this form of feedback is always lower than that from the AGN at all times, except during the post-merger phase from $z = 1.6$ to $z = 0.6$. However, in this post-merger phase the high level of supernova feedback is an artefact of the way star formation is computed: we trace back the star formation history (SFH) of the central galaxy using all the stars which belong to it at $z = 0$, so that the star formation and hence the supernova rate we derive from it include that of its accreted satellites. In this redshift range ($0.6 < z < 1.6$), the star formation rate is dominated by the galaxy progenitor that has not undergone the cataclysmic quasar phase (hosted by the branch 3 halo in Fig. 2), whilst the star formation activity in the quasar galaxy progenitor (hosted by the branch 1 halo) is completely suppressed. In light of this, it is a fair approximation to neglect the feedback from supernovae on the evolution of this galaxy cluster.

At $z \approx 1.7$, another halo major merger (1:1) occurs (branches 1 and 3 in the merger tree; Fig. 2). However, the most massive ($M_{BH} = 5.9 \times 10^9 M_\odot$) BH only merges with the ($M_{BH} = 4.3 \times 10^9 M_\odot$) BH of its cluster companion at a much later epoch ($z = 0.58$; right-hand panel of Fig. 5). Note that, in this case, the BH merger also takes place quite a long time after the central galaxies hosting the BHs merge together. As a matter of fact, the $z = 0.8$ major galaxy merger drives the cluster gas to temperatures twice the virial

Figure 5. Projected temperature in linear scale along the $x$ direction at $z = 0.88$, $z = 0.74$ and $z = 0.58$ from left to right of the cluster during its merging phase. Branch 1 of the merger tree corresponds to the red circle and branch 3 to the blue circle (see Fig. 2). This redshift sequence corresponds to pre-central galaxy merger, slightly post-central galaxy merger and BH merger, respectively. The red saturated region corresponds to a temperature $T \sim 4-6$ keV. The size of the images is 3.6 Mpc in comoving units.
temperature, thanks to a violent shock wave (middle panel of Fig. 5). However, during this galaxy merger phase, the accretion rate on to the most massive BH drops to negligible values (≈10^{-5} M_{\text{Edd}}; Fig. 3, solid black line), and only its companion continues to accrete at a moderate rate (a few ≈10^{-3} M_{\text{Edd}}; Fig. 3, red dashed line). The resulting AGN activity, even when boosted by the final BH merger at z = 0.58, does not seriously impact the highly pressurized ICM gas which confines the jet energy to the very central parts of the cluster (right-hand panel of Fig. 5). After the major BH merger, the accretion rate on to the central BH becomes extremely small (10^{-4} M_{\text{Edd}}) due to the complete evaporation of leftover cold material by the final outburst of AGN activity. Subsequently, the BH stays in this almost-dead phase for 2 Gyr.

It is striking that the less massive BH progenitor of the latter BH merger (at z = 0.58) accretes gas at 0.01 M_{\text{Edd}} before the merger, whereas accretion on to the most massive BH is negligible. The reason for this different behaviour can be understood by looking at the temperature maps of both cluster progenitors (left-hand panel of Fig. 5): the most massive cluster (branch 1, red circle) is slightly warmer than its companion (branch 3, blue circle), as it has been pre-heated by important quasar activity at earlier redshifts (z ∼ 1.56).

On the other hand, the less massive progenitor did not experience such strong pre-heating, and as a result has a lower gas temperature, and therefore a higher accretion rate during the pre-merger phase.

Finally, the cluster relaxes and the inner halo cold gas reservoir gets replenished, fuelling a faint accretion on to the BH. This translates into low AGN activity. Episodically, stronger jets are produced by the AGN in this phase which yields small perturbations of the ICM temperature in the form of sound waves (Fig. 6). These jets are roughly sonic (u_{\text{jet}} ≈ 1000 km s^{-1} and c_s ≈ 1300 km s^{-1}), and are efficiently thermalized by the ICM. As a result they do not propagate farther than a few 10 kpc.

Nevertheless, the part of the jet energy which is carried by these sound waves limits the cooling flow in the cluster core. However, the cooling time in the core (<100 kpc) is extremely short (smaller than a Gyr; see Fig. 8), so the cooling flow eventually develops again and feeds the BH afresh. As a result, the BH accretion rate increases from 10^{-4} M_{\text{Edd}} to a few 10^{-2} M_{\text{Edd}} at z ~ 0, giving rise to late-time AGN activity.

Fig. 9 shows different physical properties in a slice of gas cut through the AGN jet at z = 0. They show that the jet is at low density and high temperature in good agreement with high-resolution simulations of jet formation (see Heinz et al. 2006; Simionescu et al. 2009 for example). A remarkable feature is the formation of two underdense cavities filled with sonic-jet material. We have computed a simulated X-ray map of these cavities in three different temperature bands (see Fig. 10). These cavities are reminiscent of the ones observed in Perseus A by Fabian et al. (2006) in which a strong cooling core is also visible. In our simulation, an extended cooling core (a few 10 kpc across) is absent, but the cooling flow which gives rise to late-time AGN activity is clearly present. As in the Fabian et al. (2006) observations, we interpret the ripples induced by our jet modelling as sound waves. This can be seen in Fig. 11 where radial velocities are always sub-sonic. These sound waves, provided one can dissipate them viscously (Fabian et al. 2003; Ruszkowski, Britzen & Begelman 2004a), can reheat the ICM at distances larger than the scalelength of the jet. In our simulation, no explicit viscosity is included, but these spherical sound waves do not appear at radii larger than r_{300}, suggesting that they have been dissipated by numerical viscosity on these scales.

---

**Figure 6.** Projected temperature in linear scale along the z direction at z = 0.09, z = 0.04 and z = 0 from left to right of the cluster during its relaxed phase. The size of the images is 1.8 Mpc in comoving units.

**Figure 7.** AGN luminosity of the most massive BH at z = 0 as a function of time (red curve). For comparison, a simple estimate of the contribution from supernova activity based on star formation rate is given (black curve). Note, however, that this contribution is negligible compared to that of the AGN and is not included in the simulation.
Figure 8. Volume-weighted cooling time for the run without AGN (black) at $z = 0$ and with AGN (red) at $z = 0$ (solid), $z = 0.04$ (dashed) and $z = 0.09$ (dot–dashed). The horizontal dotted lines are the Hubble time and the time since the major merger of the two central galaxies at $z \approx 0.8$.

5 REGULATION OF THE COOLING CATASTROPHE

In the absence of strong feedback processes to offset the cooling of gas in the potential wells of massive DM haloes, too many massive galaxies are formed both in CDM cosmological hydrodynamical simulations and semi-analytic models of galaxy formation and evolution. In our cluster zoom simulation, when no AGN feedback is considered, the final mass of stars in the central cD galaxy is very high, $M_* \simeq 1.7 \times 10^{13} \, M_\odot$, for a $M_{500} = 2.4 \times 10^{14} \, M_\odot$ (at $M_{200} = 2.9 \times 10^{14} \, M_\odot$) DM halo with radius $r_{500} = 940$ kpc (respectively $r_{200} = 1370$ kpc; see Fig. 1).\(^2\) In the presence of stirring from AGN feedback, the total stellar mass is reduced to $M_* \simeq 5.6 \times 10^{12} \, M_\odot$, i.e. by more than a factor of 3. To compute the stellar mass content, we use the same MSM algorithm (Tweed et al. 2009) as for the DM but with different parameters, since stars are more clustered than DM particles. This tool efficiently separates one galaxy from another, especially the central galaxy from its satellite galaxies (see Fig. 1). However, the algorithm used in this method attributes all the stars present in the DM halo and not part of satellite galaxies to the central one. As a result, a non-negligible part of the stellar mass of the central galaxy resides in the intra-cluster stellar halo (composed of all the stars stripped from satellite galaxies of the cluster), which has a very large extent (up to ~400 kpc). This caveat must be borne in mind when comparing the stellar mass of the central object with observations: our estimate only provides an upper limit of the stellar mass content of the central galaxy.

Fig. 12 compares the stellar mass evolution as a function of redshift for the most massive galaxy at $z = 0$ (solid line) and its two most massive progenitors (dashed and dot–dashed lines, central galaxies of DM haloes identified as branches 1 and 3 in Fig. 2) in the AGN and no-AGN runs (red and black sets of curves, respectively). We observe that the reduction of the stellar mass is a continuous

\(^2\) All quantities with subscripts 200 or 500 refer to regions with overdensities 200 or 500 times larger than critical ($\rho_c = 3HF^2/(8\pi G)$).
process which begins at an early stage (around $z \sim 5$) but gets amplified as time goes on to reach a factor of 3 at $z \sim 1$. Both progenitors seem to follow the same reduction of their stellar content which suggests that even the branch 3 cluster, which does not exhibit any strong quasar activity at high redshift, is able to prevent some gas from falling on to the central galaxy. The process is more quiescent as can be seen in its BH accretion rate (red dashed curve in Fig. 3), but even this continuous and moderate AGN activity can efficiently reduce the star formation.

We evaluate the bulge to disc mass ratio of the central galaxy by using the kinematics of its star particles. First, we identify the rotation axis of the galaxy to define the correct cylindrical reference frame in which we project the velocity components of each star particle. A particle belongs to the bulge if its circular velocity is lower than half its total velocity. With that definition, the bulge-to-total mass ratio of the central galaxy is 0.75 for the simulation with AGN, and 0.80 for the simulation without AGN. Thus it appears that although AGN dramatically change the SFH, they have a much less significant impact on the morphology of a galaxy.

Fig. 13 shows the star formation rate for the central galaxy as a function of time for the two runs. To compute the time evolution of the SFR, we simply have identified the stars belonging to the central galaxy at $z = 0$, and traced them back in time. Thus it is the SFH summed over all the stars of all the satellite galaxies that have been accreted on to the central galaxy throughout its evolution. The SFR continuously decreases with time due to early AGN activity ($z \sim 4$), but the dramatic decline in SFR occurs around $z = 0.6$ when the BHs hosted by the central galaxy merger remnant of branches 1 and 3 haloes finally merge. The vast majority of the cold gas is heated up by the AGN during this violent merger. In contrast, the cold gas in the no-AGN case is simply compressed in the galaxy merger which results in a double small star formation peak around the same redshift ($z \sim 0.6$). The latter effect is a well-known property of merging galaxies without AGN (Mihos & Hernquist 1996). However, mergers of galaxies containing BHs boost the accretion of gas on to the BH fuelling strong AGN activity and produce a dip in the SFR by reducing the cold gas content. These features are clearly seen at redshift $z = 1.6$ and $z = 0.6$ in Fig. 13. Such a behaviour has already been analysed in detail in idealized (as opposed to cosmological) simulations of galaxy merger (Di Matteo et al. 2005; Springel et al. 2005). Our work confirms that it also occurs in more realistic cosmological configurations.

We have measured the cumulative mass profiles of the different components (gas, stars, DM) as a function of radius for the cluster at $z = 0$ (Fig. 14). In the absence of feedback, most of the cold baryons are concentrated within the galaxy (in the inner 10 kpc):
i.e. the cooling catastrophe has occurred. There is a strong difference in the cumulative profiles between the runs with and without AGN activity, especially in the central region of the cluster. They differ by a factor of 5 in total mass at $r = 10$ kpc and by a factor of 1.5 at $r = 100$ kpc. Without AGN, the gravitational potential is steeper in the centre of the cluster, baryons accumulate and gravitationally pull DM along with them. This is predicted to have severe consequences when simulating the gravitational lensing effect of such structures (Peirani et al. 2008; Meneghetti et al. 2010).

Baryon fraction as a function of radius provides us with a good benchmark to quantify the influence of feedback processes. We have computed the cumulative fractions of stars, gas and DM for both runs (Fig. 15). We observe a very small difference in the total baryon fraction at $r_{200}$ and $r_{500}$ which means that this quantity is relatively immune to the presence of feedback. In contrast, the baryon fraction has decreased by a factor of 2 at $r = 0.1r_{500}$. This means that half of the baryons that were concentrated in the inner parts of the halo have been redistributed by the AGN feedback at larger radii. However, as the baryon fraction seems relatively independent of the presence of the AGN at very large radii ($> r_{500}$), we can conclude that only
Table 1. Comparison of the stellar, gas and baryon fractions at $r_{500}$ in our simulations with the observational data.

|            | no AGN | AGN | L03$^a$ | G07$^b$ | G09$^c$ |
|------------|--------|-----|---------|---------|---------|
| $f_{500}^{\text{star}}$ | 0.090  | 0.036 | 0.019   | 0.023   | 0.032   |
| $f_{500}^{\text{gas}}$  | 0.073  | 0.116 | 0.117   | 0.101   | 0.090   |
| $f_{500}^{\text{baryon}}$ | 0.163  | 0.152 | 0.136   | 0.124   | 0.122   |

$^a$Observational data from Lin et al. (2003).
$^b$Observational data from Gonzalez et al. (2007). Their gas fraction is the best fit to data from Vikhlinin et al. (2006) and Gastaldello et al. (2007).
$^c$Observational data from Giodini et al. (2009). Their stellar fraction is the best fit of their data combined with data from Lin et al. (2003). Their gas fraction is the best fit to data from Vikhlinin et al. (2006), Arnaud et al. (2007) and Sun et al. (2009).

a modest fraction of the gas is expelled by the AGN outside of the cluster.

Stellar and gas fractions yield more clues as to the impact of AGN feedback on the cluster history. There is a strong decrease in the stellar fraction, even at large radii ($r > r_{500}$), meaning that star formation has been efficiently suppressed by the AGN activity. We see that the stellar fraction at $r_{500}$ has been lowered by more than a factor of 2, which is comparable with the results obtained by Duffy et al. (2010). Table 1 shows the stellar fractions $f_{500}^{\text{star}}$, gas fractions $f_{500}^{\text{gas}}$ and baryon fractions (gas and stars) $f_{500}^{\text{baryon}}$ measured at $r_{500}$ in our simulations and compared to X-ray or near-infrared measurements made by various groups (Lin, Mohr & Stanford 2003; Vikhlinin et al. 2006; Arnaud, Pointecouteau & Pratt 2007; Gonzalez, Zaritsky & Zabludoff 2007; Gastaldello et al. 2007; Giodini et al. 2009; Sun et al. 2009). Observational data values are obtained using the best fit these authors provide for $f_{500}^{\text{star}}$, $f_{500}^{\text{gas}}$ and $f_{500}^{\text{baryon}}$ as a function of $M_{500}$.

From this comparison with observations, the simulation which does not include AGN feedback is clearly ruled out, as the stellar fraction is too high by a factor of $\sim 3$--4 and the gas fraction too low by 25 per cent in the most favourable case. This clearly indicates that the ICM has undergone a cooling catastrophe: too much gas has been depleted and transformed into stars. On the other hand, the simulation with AGN feedback shows reasonable agreement with the stellar fraction estimated from X-ray measurements by Gonzalez et al. (2007) and Giodini et al. (2009), but still overestimates the stellar fraction from near-infrared data by Lin et al. (2003) by about a factor of 2. Gas fraction in the simulation with AGN is also within the range of values inferred from X-ray gas emission, albeit on the high side. The major flaw of our simulations is their inability to match the lower than universal baryon fraction observed in small galaxy clusters ($M_{500} < 10^{15} M_\odot$). We note that the discrepancy would be even more blatant had we used WMAP 5 year parameters (Dunkley et al. 2009) since the universal baryon fraction goes up to 18 per cent for this cosmology. In the case of our AGN simulation, the cluster baryon fraction is close to the universal baryon fraction $\Omega_b/\Omega_m = 0.15$, but we fail to push gas far enough out of the cluster potential well to match lower observational values sitting around $\sim 0.13$.

The gas fraction behaviour is somewhat counterintuitive: it is larger in the run with AGN feedback, whatever the radius is. This is explained by the fact that AGN removes gas from the central regions of the cluster to replenish its outer parts. AGN feedback thus transforms cold gas contained in the central galactic disc into hot and diffuse halo gas. Moreover, the gas fraction has a remarkable feature in the form of a pronounced dip at intermediate radius (15 kpc for the no-AGN and 100 kpc for the AGN case) which marks the transition between the cold/dense phase ($n > 0.1$ cm$^{-3}$), and the hot/diffuse component. Such dips in the gas fraction also are commonplace in X-ray cluster surveys (Vikhlinin et al. 2006). Another interesting result from Fig. 15 is that the DM to total mass fraction in the cluster core ($r < 10$ kpc) is higher in the case of AGN feedback, even though the total amount of DM is lower in this case. We attribute this to the domination of the mass budget by the stellar component which pulls DM along with it through adiabatic contraction (Blumenthal et al. 1986).

Finally, we compare the thermodynamical properties of the gas in the two runs in Fig. 16. We have fitted the gas density profile in the AGN run with a $\beta$ profile of the form

$$\rho = \rho_s \left(1 + (r/r_c)^2\right)^{-3\beta/2},$$

where $\rho_s = 0.5$ cm$^{-3}$, $r_c = 10$ kpc and $\beta = 0.6$. This profile matches the density profile of the relaxed cluster at different times ($z = 0$, $z = 0.04$ and $z = 0.09$, which are separated by 500 Myr) in the intermediate-radius range 0.05--1$r_{500}$. When the cluster is relaxed, the same analytic profile extends to the core of the cluster, but as soon as a cooling flow develops it fails to describe the numerical gas density accurately. Indeed, the core of the cluster shrinks as gas flows in, so $r_c$ drops, but the $\beta$ index stays identical as the profile remains unaffected by the cooling flow on large scales. By contrast, fitting the gas density profile with a $\beta$ profile for the simulation where no AGN is present turns out to be an impossible task, since the core radius becomes smaller than the spatial resolution in that case.

Surprisingly, the simulation without the AGN, which has endured a cooling catastrophe for Gyr, exhibits a very hot gas core (temperature in excess of 9 keV) with a very steep profile (see second panel from the top in Fig. 16). Actually a massive central cold disc component is also present in this run and would appear in Fig. 16 if we were measuring mass-weighted instead of volume-weighted quantities, but the properties of the gas at the centre of the cluster would then reflect the properties of the ISM instead of the diffuse ICM. This very hot thermal part, in the no-AGN run, arises from the cluster need for more thermal energy to support the gas against the extra gravitational compression generated by adiabatic contraction. Due to the very high temperature and a lack of diffuse gas around the post-merger galaxy this energy is not easily radiated away (the cooling time is greater than 2 Gyr as shown in Fig. 8). However, we can clearly see in Fig. 17 that the gas, close to the galactic disc ($r < 50$ kpc), still collapses on to that galaxy due to the lack of pressure support (Fig. 16), which explains the depletion of the gas component at $r \approx 100$ kpc.

On the other hand, when the AGN is active the temperature profile is stabilized and looks quasi-isothermal in the range 0.05--1$r_{500}$ (second panel of Fig. 16). Before $z = 0$, the temperature is a factor of 2 to 4 higher in the inner 10 kpc, due to heating from the jet, which remains confined in that region. As the gas radiates away the jet energy, its temperature drops and a cooling flow develops because of a lack of pressure support in the core (second panel from the bottom in Fig. 16). Small variations of temperature with radius in the form of wiggles can be observed in Fig. 16 (second panel from the top, solid red curve). These correspond to the propagation of sound waves into the ICM. These waves contribute to reheating the cooling plasma in the cluster as a whole by propagating and isotropizing the energy injected by the AGN jet. They manage to offset the extremely short cooling time within $r < 0.1r_{500}$ which is at least one order of magnitude shorter than the time elapsed since the last major merger (see Fig. 8), and thus prevents most of the gas from collapsing on to the central galaxy.
Figure 16. From top to bottom: volume-weighted average number density, temperature, pressure and entropy as a function of radius for the run without AGN (black solid) at $z = 0$, and the run with AGN at $z = 0$ (red solid), $z = 0.04$ (red dashed) and $z = 0.09$ (red dot-dashed). The blue line in the number density plot is a $\beta$ profile fit with $\beta = 0.6$. The dashed black line in the entropy plot describes a $K \propto r^{1.2}$ power law.

The pressure cavities seen in Fig. 9 and in the X-ray map (Fig. 10) are visible in the pressure profile of Fig. 16 (second panel from the bottom): at $z = 0$, there is a small depression in the pressure profile around $r \approx 15$–$30$ kpc, which does not appear at earlier times when the AGN is not active enough to form these cavities. This feature is also detectable in the radial velocity profile of the gas in Fig. 17 (bottom panel): there is a net radial gas outflow at $r \approx 15$–$30$ kpc whose maximum corresponds to the maximum extent of the jet and whose outwardly decreasing profile reflects the pre-shocked cocoon region. The volume-averaged velocity which we plot in this figure is, however, underestimated, because the sonic outflowing component of the jet is mixed with the quasi-steady flow or inflowing regions. The velocity inside the jet is much faster, around $1000$ km s$^{-1}$.

Finally, entropy profiles (bottom panel of Fig. 16) shows a plateau in the cluster core (at $r < 20$ kpc for the AGN run, and $r < 300$ kpc for the no-AGN run) with a strong departure from the scaling power law $K \propto r^{1.2}$. This indicates the level of turbulent mixing in the gas as explained in detail by Mitchell et al. (2009) and is nicely illustrated by comparing entropy profiles in Fig. 16 (bottom panel) with radial velocity dispersion profiles in Fig. 17 (top panel). Such a comparison clearly shows that the stronger the turbulence level (or equivalently the radial velocity dispersion), the higher the entropy.

6 DISCUSSION AND CONCLUSION

We have numerically studied how a self-regulating model of SMBH growth and AGN feedback impacts the formation history of a large cluster of galaxies. Using a resimulation technique to explicitly account for the cosmological context ($\Lambda$CDM universe) which drives the cluster growth and an AMR technique to solve the equations of hydrodynamics without running into entropy issues, we find the following.

(i) BHs accrete gas efficiently at high redshift ($z > 2$), significantly pre-heating proto-cluster haloes in the process.

(ii) Some, but not all, wet (gas-rich) mergers fuel strong episodic jet activity which transport gas from the cluster core to its intermediate/outer regions.

© 2010 The Authors. Journal compilation © 2010 RAS, MNRAS 409, 985–1001
(iii) Reduced infall of cold gas during the more secular phase evolution of clusters produces smaller outbursts from the central AGN which contribute to heat the whole cluster via sound waves but are inefficient in redistributing the gas outwards.

(iv) Late-time AGN activity forms two large cavities correlated with the emergence of a small cooling flow.

Whilst our model for accretion on to the black hole is commonly used in simulations in the literature (Di Matteo et al. 2005; Springel et al. 2005; Sijacki et al. 2007; Di Matteo et al. 2008; Booth & Schaye 2009; Teyssier et al. 2010), this is the first time, to the best of our knowledge, that a momentum-driven jet is implemented as AGN feedback in cosmological simulations and followed in a self-consistent way. We argue that this is one step in the correct direction since powerful jets are observed in the Universe on scales well resolved by any of these simulations (e.g. Bridle et al. 1994). Other authors have adopted a more phenomenological approach where energy is either accumulated by the BH before being released as a thermal pulse (Booth & Schaye 2009; Teyssier et al. 2010) or simply used as a continuous heat source (Springel et al. 2005; Di Matteo et al. 2005, 2008), or used both heating modes (Sijacki et al. 2007). Injection of non-thermal relativistic protons in rising buoyant AGN bubbles has also been explored as an alternative feedback mechanism (Sijacki et al. 2008). It is worth noting the recrudescence of efforts done to improve models of AGN feedback in idealized (non-cosmological) simulations of galaxy evolution, where kinetic energy is deposited either isotropically (DeBuhr et al. 2010; Power, Nayakshin & King 2010) or – as in the present paper – as a collimated jet (Nayakshin & Power 2010). These models probably capture more of the relevant jet physics than we capture here, and we believe that as the spatial resolution in cosmological simulations increases, more physical insight into the impact of AGN feedback on the ICM will be gained by coupling them to such models. While we believe that most of the results we get are similar to those obtained with the continuous heating model and that our feedback seems less efficient at stopping the cooling catastrophe than the accumulated heating prescription, the devil certainly is in the details and we defer a more detailed comparison to a future paper (Dubois et al. in prep). It is however interesting to note that all these models, including ours, are calibrated to provide an acceptable $M_{\text{BH}}$ versus $M_*$ (or $M_{\text{BH}}$ versus $\sigma_*$) relation so that it is very unlikely that this relation can be used to constrain feedback mechanisms.

From both semi-analytic prescriptions (Bower et al. 2006; Cattaneo et al. 2006) and recent numerical simulations (Ruszkowski & Oh 2010; Duffy et al. 2010), it has become clear that AGN feedback must play a major role in reducing the central galaxy stellar mass in groups and clusters by a large amount. Our implementation of feedback succeeds at least partially in reaching this goal: the stellar mass of the central object is reduced by more than a factor of 3 in the run where AGN feedback stirs the gas whose properties (density profile, temperature, radial velocity) also are in better agreement with observations. Another success of our AGN modelling is its capacity to reproduce double cavities separated by a cold component as seen in the X-ray emission of observed clusters (e.g. in Perseus, Fabian et al. 2006). Our experiments with a simple isotropic thermal input (Teyssier et al. 2010) suggest that these cannot be reproduced.

One drawback of our simulation is that no supernova feedback is included. However, we do not expect this feedback to be energetically relevant as a simple estimate based on the SFH of our galaxies shows that it is always an order of magnitude lower than the AGN energy input (Fig. 7). On the other hand, supernovae feedback releases metals into the ICM. As the cluster gas is heated to very high temperatures ($\sim$3 keV), cooling in the ICM is primarily due to free–free collisions, thus we do not expect metals to strongly alter the cooling rate of the plasma. Indeed, at $T = 3.5$ keV, the relative difference in the net cooling rate between a zero and a one-third of solar metallicity plasma is $0.2$ according to Tozzi & Norman (2001). This means that, assuming a $Z = 0.3Z_{\odot}$ metallicity in the ICM, our simulation underestimates the gas cooling rate by 20 per cent.

As always when analysing numerical simulations, one has to worry about spatial and mass resolution related issues. Indeed, accretion on to a SMBH happens on (sub)parsec scales, well below the kpc size of the smallest cell in our cluster re-simulation. Even though we use the empirically well-motivated sub-grid model of Booth & Schaye (2009) to account for this lack of resolution, one would like to test its validity by performing direct simulations. Whilst this is beyond the reach of the current generation of supercomputers, we have tried to assess both the robustness of our sub-grid and our jet implementations by performing a sub-kpc run and conclude that our results are by-and-large unchanged at this increased spatial resolution. This is also true for a modest increase in mass resolution. We are therefore quite confident that the conclusions we draw in this paper are robust vis–vis resolution issues and defer a more thorough resolution study to future work.

Finally, other physical mechanisms, which we do not model here, could potentially play a significant role in preventing the development of massive cooling flows. These mechanisms involve tapping into the heat reservoir provided by the outer regions of galaxy clusters to raise the gas temperature in their core. In particular, recent efforts have been made to investigate the importance of anisotropic thermal conduction in idealized galaxy clusters (Parrish & Quataert 2008; Parrish et al. 2009; Bogdanović et al. 2009). These studies have shown that the HBI can reorient the magnetic field lines in the cluster core in an azimuthal configuration that stops the inward heat flux. The most recent of these simulations (Parrish, Quataert & Sharma 2010) demonstrated that some small turbulence is brought to break this magnetic field topology, heating is able to proceed. One has to wonder, in the context of anisotropic thermal conduction, whether the turbulence induced by small-scale and large-scale motions is able to reorder magnetic fields in cosmological simulations of galaxy clusters (Dolag et al. 2005; Asai, Fukuda & Matsumoto 2007; Dubois & Teyssier 2008a; Pfrommer & Dursi 2010), and thus, break the shield-shielding of the heat-flux. In particular, the presence of AGN stirring can help to disturb this magnetic equilibrium (Dubois et al. 2009), so we believe that this problem needs to be addressed with MHD cosmological simulations including both AGN stirring and anisotropic thermal conduction.

ACKNOWLEDGMENTS

We thank Taysun Kimm and Stas Shabala for useful comments and discussion. YD is supported by an STFC Postdoctoral Fellowship. The simulations presented here were run on the TITANE cluster at the Centre de Calcul Recherche et Technologie in CEA Saclay on allocated resources from the GENCI grant c2009046197.

REFERENCES

Agertz O. et al., 2007, MNRAS, 380, 963
Agertz O., Lake G., Teyssier R., Moore B., Mayer L., Romeo A. B., 2009, MNRAS, 392, 294

© 2010 The Authors. Journal compilation © 2010 RAS, MNRAS 409, 985–1001
Jet-regulated cooling catastrophe

Truelove J. K., Klein R. I., McKee C. F., Holliman II J. H., Howell L. H., Greenough J. A., 1997, ApJ, 489, L179
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
Vernaleo J. C., Reynolds C. S., 2006, ApJ, 645, 83

Vikhlinin A., Kravtsov A., Forman W., Jones C., Markevitch M., Murray S. S., Van Speybroeck L., 2006, ApJ, 640, 691
Voigt L. M., Fabian A. C., 2004, MNRAS, 347, 1130

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