Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **487,** 4393–4408 (2019) doi:10.1093/mnras/stz1494
Advance Access publication 2019 June 3

# The failure of stellar feedback, magnetic fields, conduction, and morphological quenching in maintaining red galaxies

Kung-Yi Su,[1★] Philip F. Hopkins,[1] Christopher C. Hayward,[2] Xiangcheng Ma,[1,3] Claude-André Faucher-Giguère,[4] Dušan Kereš,[5] Matthew E. Orr,[1] T. K. Chan[5] and Victor H. Robles[6]

[1]*TAPIR 350-17, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA*
[2]*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA*
[3]*Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA*
[4]*Department of Physics and Astronomy and CIERA, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[5]*Department of Physics, Center for Astrophysics and Space Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[6]*Center for Cosmology, Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*

Accepted 2019 May 24. Received 2019 May 7; in original form 2018 September 23

## ABSTRACT

The quenching 'maintenance' and related 'cooling flow' problems are important in galaxies from Milky Way mass through clusters. We investigate this in haloes with masses $\sim 10^{12}$–$10^{14}\,M_{\odot}$, using non-cosmological high-resolution hydrodynamic simulations with the FIRE-2 (Feedback In Realistic Environments) stellar feedback model. We specifically focus on physics present without AGN, and show that various proposed 'non-AGN' solution mechanisms in the literature, including Type Ia supernovae, shocked AGB winds, other forms of stellar feedback (e.g. cosmic rays), magnetic fields, Spitzer–Braginskii conduction, or 'morphological quenching' do not halt or substantially reduce cooling flows nor maintain 'quenched' galaxies in this mass range. We show that stellar feedback (including cosmic rays from SNe) alters the balance of cold/warm gas and the rate at which the cooled gas within the galaxy turns into stars, but not the net baryonic inflow. If anything, outflowing metals and dense gas promote additional cooling. Conduction is important only in the most massive haloes, as expected, but even at $\sim 10^{14}\,M_{\odot}$ reduces inflow only by a factor $\sim 2$ (owing to saturation effects and anisotropic suppression). Changing the morphology of the galaxies only slightly alters their Toomre-$Q$ parameter, and has no effect on cooling (as expected), so has essentially no effect on cooling flows or maintaining quenching. This all supports the idea that additional physics, e.g. AGN feedback, must be important in massive galaxies.

**Key words:** MHD – methods: numerical – cosmic rays – Galaxy: evolution – galaxies: clusters: intracluster medium – X-rays: galaxies: clusters.

## 1 INTRODUCTION

Perhaps the biggest unsolved question in galaxy formation is what explains the 'quenching'[1] of star formation and maintenance of 'red and dead' galaxy populations over a large fraction of cosmic time (Bell et al. 2003; Kauffmann et al. 2003; Madgwick et al. 2003; Baldry et al. 2004; Blanton et al. 2005; Dekel & Birnboim 2006; Kereš et al. 2009; Pozzetti et al. 2010; Wetzel, Tinker & Conroy 2012), at stellar masses $\gtrsim$3–5 $\times\,10^{10}\,M_{\odot}$ (above $\sim L_{*}$ in the galaxy luminosity function at $z \approx 0$). This is closely related to the classic 'cooling flow problem': X-ray observations show there exists significant radiative cooling of hot gas in massive ellipticals and clusters with cool-core, indicating cooling times much less than a Hubble time (Fabian et al. 1994; Peterson & Fabian 2006). Comparing with the inferred cooling flows (reaching up to $\sim 1000\,M_{\odot}\,yr^{-1}$ in clusters McDonald et al. 2018), there are neither sufficient amounts of cold gas (in observed H I, e.g. McDonald, Veilleux & Mushotzky 2011, or CO, Werner et al. 2013), or sufficient star formation rates (SFRs; Tamura et al. 2001; O'Dea et al. 2008; Rafferty, McNamara & Nulsen 2008), to account for the rapidly cooling gas (see also Silk 1976; Cowie & Binney 1977;

★ E-mail: ksu@caltech.edu

[1]Throughout this paper, when we refer to 'quenching' and red galaxies, we exclusively refer to central galaxies, as opposed to satellite galaxies which can be quenched by a variety of environmental processes (e.g. ram pressure and tidal stripping, starvation, strangulation, etc.).

Mathews & Bregman 1978; Ciotti et al. 1991; Fabian 1994 for the 'classical' cooling flow case). Simulations and semi-analytic models which do not suppress these cooling flows, and simply allow the material to cool into galaxies, typically overproduce the observed SFRs of massive galaxies by at least an order of magnitude (for recent examples, see e.g. the weak/no feedback runs in Sijacki et al. 2007; Booth & Schaye 2009; Choi et al. 2015; Li et al. 2015).

To compensate for the observed cooling, there must be some sort of heat source or pressure support. The presence of the shock heated hot-halo can help feedback mechanisms and quench galaxies (e.g. Kereš et al. 2005). However, the hot halo itself does not prevent later gas cooling from the cooling flows. The most popular, and perhaps promising, solution is 'feedback' from an active galactic nucleus (AGN) which can both expel gas from galaxies (shutting down star formation) and inject heat or stirring in the circum-galactic medium (CGM) or intra-cluster medium (ICM), preventing new gas accretion (for recent studies, see also e.g Faucher-Giguère & Quataert 2012; Anglés-Alcázar et al. 2017a; Eisenreich et al. 2017; Gaspari & Sądowski 2017; Li, Ruszkowski & Bryan 2017; Li et al. 2018; Martizzi et al. 2018; Pellegrini et al. 2018; Weinberger et al. 2018; Yoon et al. 2018; see e.g. Silk & Rees 1998; Fabian 1999; Ciotti & Ostriker 2001; Hopkins et al. 2005, 2006; Croton et al. 2006; McNamara & Nulsen 2007 for earlier works). However, despite its plausibility, the detailed physics of AGN feedback (e.g. what determines jet energetics and how they transfer energy into the ICM) remains uncertain, as do the relevant 'input parameters' (e.g. kinetic luminosities, duty cycles).

Perhaps as a result, a variety of 'non-AGN' mechanisms to quench galaxies and keep them red have been proposed in the literature. These generally invoked physics are un-ambiguously present, but play an uncertain role in quenching and the cooling flow problem, including: stellar feedback from shock-heated AGB winds, Type Ia supernovae (SNe), or SNe-injected cosmic rays (CRs); magnetic fields and thermal conduction in the CGM/IGM; or 'morphological quenching' via altering the galaxy morphology and gravitational stability properties. Our focus in this paper is therefore to attempt a systematic theoretical study of these possibilities, without considering AGN.

This is important for several reasons: if one (or more) of these mechanisms can, indeed, quench galaxies, this is critical to understand! Even if they do not quench galaxies, they could, in principle, 'help' by suppressing cooling or star formation (lessening 'requirements' for AGN). And although many previous studies have claimed AGN feedback is 'necessary' to explain quenching (see e.g. Ciotti et al. 1991; McNamara & Nulsen 2007; Conroy & Ostriker 2008; Khalatyan et al. 2008; Taylor & Kobayashi 2015, in addition to the references above), almost all studies of AGN feedback to date have neglected some or all of these additional processes (often treating e.g. stellar feedback in a highly simplified, sub-grid manner). Therefore, it is important to understand whether they alter the 'initial conditions' (e.g. typical CGM properties, cooling rates, etc.) for AGN feedback and 'how much more' AGN feedback need to do. We hope that by studying the 'over-cooling problem' in global simulations with higher resolution and more detailed physical treatments of the multiphase ISM and stellar feedback, we can better understand where and how AGN or other feedback, if indeed necessary, must act.

In Section 2, we summarize the physics considered here, and describe our numerical simulations. Results are presented in Section 3. We then discuss the effects of each of these physics in turn, in Section 4.

## 2 METHODOLOGY

We used simulations of isolated galaxies with halo masses ranging from $10^{12}$ to $10^{14}\,M_{\odot}$ to test how various physical processes affect galactic cooling flows. We set-up the initial conditions according to the observed profiles of cool-core clusters at low redshift, as detailed in Section 2.2. Without proper quenching mechanisms, although the galaxies have initial properties consistent with observations, their cooling flow rates and SFRs quickly runaway, surpassing the observational values by orders of magnitude. We ran the simulations with various proposed quenching mechanisms and tested to what extent (if any) they suppress the cooling flow and whether they can maintain stably quenched galaxies.

Our simulations use GIZMO (Hopkins 2015),[2] in its meshless finite mass (MFM) mode, which is a Lagrangian mesh-free Godunov method, capturing advantages of grid-based and smoothed-particle hydrodynamics (SPH) methods. Numerical implementation details and extensive tests are presented in a series of methods papers for e.g. the hydrodynamics and self-gravity (Hopkins 2015), magnetohydrodynamics (MHD; Hopkins & Raives 2016; Hopkins 2016), anisotropic conduction and viscosity (Hopkins 2017; Su et al. 2017), and cosmic rays (Chan et al. 2018).

Our default simulations use the FIRE-2 implementation of the FIRE physical treatments of the ISM and stellar feedback, details of which are given in Hopkins et al. (2018b,a) along with extensive numerical tests. This follows cooling from 10 to $10^{10}$ K, including the effects of photo-electric and photo-ionization heating by a UV background Faucher-Giguère et al. (2009) and local source, collisional, Compton, fine structure, recombination, atomic, and molecular cooling. Star formation is allowed only in gas that is molecular, self-shielding, locally self-gravitating (Hopkins, Narayanan & Murray 2013), and above a density $n > 100\,\mathrm{cm}^{-3}$. Star particles, once formed, are treated as a single stellar population with metallicity inherited from their parent gas at formation. All feedback rates (SNe and mass-loss rates, spectra, etc.) and strengths are IMF-averaged values calculated from STARBURST99 (Leitherer et al. 1999) with a Kroupa (2002) IMF. The feedback model includes: (1) Radiative feedback including photo-ionization and photo-electric heating, as well as single and multiple-scattering radiation pressure tracked in five bands (ionizing, FUV, NUV, optical-NIR, IR). (2) Stellar particles continuously lose mass and inject mass, metals, energy, and momentum in the form of OB and AGB winds. (3) Type II and Ia SNe happen stochastically according to the rate mentioned above. Once they occur, the stellar particles lose mass and inject the appropriate mass, metal, momentum, and energy to the surrounding gas.

### 2.1 Physics surveyed

#### *2.1.1 Stellar feedback: young/massive stars*

Feedback from massive stars is un-ambiguously crucial to galaxy evolution. In the last decade, with progress in modeling stellar feedback, simulations of $\lesssim L_{*}$ galaxies (see e.g. Governato et al. 2007; Ceverino & Klypin 2009; Hopkins, Quataert & Murray 2011, 2012a,b; Uhlig et al. 2012a; ; Kim, Ostriker & Kim 2014; Muratov et al. 2015; Agertz & Kravtsov 2016; Hu et al. 2016) are now able to produce reasonably realistic galaxy populations, without the runaway collapse and star formation that occurs absent feedback. However in these (mostly star forming) lower mass galaxies,

[2] A public version of this code is available at http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html.

feedback is dominated by young, massive stars (e.g. radiation and OB winds from massive stars, type-II SNe). Given the observed low specific SFRs (SSFRs) in quenched systems (e.g. $\lesssim 10^{-11}$ yr$^{-1}$ for $10^{14}$ M$_\odot$ haloes; Weinmann et al. 2006; Woo et al. 2013; Kravtsov, Vikhlinin & Meshcheryakov 2018), the number of massive stars is very low, so it seems unlikely this can maintain a quenched galaxy without (paradoxically) a much larger SFR. But these physics must be present whenever star formation does occur, so we include them with the methods described above.

### 2.1.2 Stellar feedback: SNe Ia

At the observed low SSFRs of massive (quenched) galaxies, the SNe Ia rate (including both prompt and delayed populations) is much larger than the core-collapse rate, giving a time-averaged energy-injection rate $\sim 10^{41.5}\,\mathrm{erg\,s^{-1}}\,(M_*/10^{11}\,_\odot)$, which can be comparable to the cooling rates in some systems. Since these come from old populations, and are distributed smoothly in space and time, it has been proposed that they could be an important CGM/ICM heating mechanism (e.g. Tang & Wang 2005; Tang et al. 2009; Sharma et al. 2012, and references therein). We include Ia's following the FIRE-2 method described above, using with the rates from Mannucci, Della Valle & Panagia (2006) (including both the prompt and delayed components), assuming $10^{51}$ erg per event. Note that although there has been considerable debate about Ia rates, it has focused on the prompt component, which is unimportant for our conclusions.

### 2.1.3 Stellar feedback: AGB winds

AGB winds from old stellar populations return a significant fraction of the stellar mass, but have low launch velocities $\sim 10\,\mathrm{km\,s^{-1}}$ and correspondingly negligible kinetic luminosities. However, Conroy, van Dokkum & Kravtsov (2015) note that if the AGB stars are moving through the ambient gas medium with large velocity dispersions $\gtrsim 300\,\mathrm{km\,s^{-1}}$, the kinetic luminosities and post-shock temperatures are greatly elevated in the wind bow shocks, and this can suppress cooling and inject energy well above the Ia rate. Crucially, our default FIRE-2 models account in detail for the relative star gas velocity when injecting stellar mass loss of any kind (AGB or OB winds or SNe), in an exactly conservative manner, as described and tested in Hopkins et al. (2018a).

### 2.1.4 Magnetic fields, conduction and viscosity

Magnetic fields can, in principle, directly suppress cooling flows via providing additional pressure support (Soker & Sarazin 1990; Beck et al. 1996, 2012), although they have limited effects on global star formation properties of sub-$L_*$ galaxies (Su et al. 2017). They can also non-linearly influence essentially all the gas dynamics.

Thermal conduction can carry heat from the outer parts of hot haloes into cool inner cores, and so might serve as an important heating mechanism (Binney & Cowie 1981; Tucker & Rosner 1983; Fabian, Voigt & Morris 2002; Voigt et al. 2002; Kim & Narayan 2003; Zakamska & Narayan 2003; Dolag et al. 2004; Voigt & Fabian 2004; Pope et al. 2006; Voit 2011; Wagh, Sharma & McCourt 2014). However, under the conditions in the CGM/ICM, this cannot be considered in the absence of MHD, as the conduction is highly anisotropic. Convective instabilities driven by anisotropic (Spitzer–Braginskii) conduction (Spitzer & Härm 1953; Sarazin 1988; Zakamska & Narayan 2003; ZuHone et al. 2015; Kannan et al. 2016) along magnetic field lines, including the heat-flux-driven buoyancy instability (HBI; Parrish & Quataert 2008; Quataert 2008) and the magnetothermal instability (MTI; Balbus 2000; Parrish, Stone & Lemaster 2008), may further change the magnetic configuration and conduction time-scale (Parrish, Quataert & Sharma 2009) or even drive turbulence and provide extra pressure support or mixing (Parrish et al. 2012). It is also been argued that conduction can help AGN feedback quench galaxies more effectively (e.g. Kannan et al. 2017).

We therefore consider a set of additional fluid 'microphysics' runs, with ideal MHD and physical (temperature-dependent, fully anisotropic) Spitzer–Braginskii conduction and viscosity (we assume the perpendicular transport coefficients are vanishingly small). The implementation is identical to Su et al. (2017).

### 2.1.5 Cosmic rays (not from AGN)

Cosmic rays (CRs) can provide additional pressure support to gas, drive galactic outflows, and heat the CGM/ICM directly via hadronic and streaming losses (Guo & Oh 2008; Sharma, Parrish & Quataert 2010; Enßlin et al. 2011; Fujita & Ohira 2011; Fujita, Kimura & Ohira 2013; Pfrommer 2013; Wiener, Oh & Guo 2013; Jacob & Pfrommer 2017a,b; Pfrommer et al. 2017a; Ruszkowski, Yang & Zweibel 2017a; Ruszkowski, Yang & Reynolds 2017b; Jacob et al. 2018). As a result several of the studies above suggest they can help quench star formation; however this is usually in the context of CRs from AGN. Here, we wish to explore non-AGN mechanisms, so we consider simulations adopting the CR physics and numerical implementation described in (Chan et al., in preparation). This CR treatment includes including streaming (at the local Alfvén speed or sound speed, whichever is larger, $v_{\rm st} \sim \sqrt{v_{\rm Alf}^2 + v_{\rm c}^2}$, with the appropriate streaming loss term, which thermalizes, following Uhlig et al. 2012b), diffusion (with a fixed diffusivity $\kappa_{\rm cr}$), adiabatic energy exchange with the gas and cosmic ray pressure in the gas equation of motion, and hadronic and Coulomb losses (following Guo & Oh 2008). We follow a single energy bin (i.e. GeV CRs, which dominate the pressure), treated in the ultra-relativistic limit. Streaming and diffusion are fully anisotropic along field lines. CRs are injected in SNe (a fixed fraction $\epsilon_{\rm cr} = 0.1$ of each SNe energy goes into CRs; see e.g. Pfrommer et al. 2017a,b). In Chan et al. (in preparation), we show that matching observed $\gamma$-ray luminosities, in simulations with the physics above requires $\kappa_{\rm cr} \sim 10^{29}\,\mathrm{cm^2\,s^{-1}}$ (in good agreement with detailed CR transport models that include an extended gaseous halo around the Galaxy, (see e.g. Strong & Moskalenko 1998; Strong et al. 2010; Trotta et al. 2011), so we adopt this as our fiducial value, but discuss variations below.

We note that in addition to SNe shocks, the other major non-AGN source of CRs of interest here is shocks from cosmological large-scale structure (LSS) formation/accretion. Since our simulations are not fully cosmological, this is not directly accounted for.[3]

### 2.1.6 Morphological quenching

Finally, Martig et al. (2009) and Dekel, Sari & Ceverino (2009) described a scenario they referred to as 'morphological quenching,' whereby quenching could be accomplished (SF suppressed) simply by altering a galaxy's morphology. Specifically, they argued that turning a stellar disc into a more gravitationally stable spheroid

[3]We implicitly effectively model this in the CR energy density in our initial conditions by assuming equipartition with magnetic energy.

**Table 1.** Simulation properties.

| Model | Resolution | | DM halo | | | | Stellar bulge | | Stellar disc | | Gas disc | | Gas halo | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\epsilon_g$ | $m_g$ | $M_{halo}$ | $c$ | $V_{Max}$ | $M_{bar}$ | $M_b$ | $a$ | $M_d$ | $r_d$ | $M_{gd}$ | $r_{gd}$ | $M_{gh}$ | $r_{gh}/r_{dh}$ |
| | (pc) | ($M_\odot$) | ($M_\odot$) | | (km s$^{-1}$) | ($M_\odot$) | ($M_\odot$) | (kpc) | ($M_\odot$) | (kpc) | ($M_\odot$) | (kpc) | ($M_\odot$) | |
| m12 | 1 | 8e3 | 1.5e12 | 12 | 174 | 2.2e11 | 1.5e10 | 1 | 5e10 | 3 | 5e9 | 6 | 1.5e11 | 1 |
| m12 Bulge | 1 | 8e3 | 1.5e12 | 12 | 174 | 2.2e11 | 6e10 | 1.6 | 5e9 | 1.4 | 5e9 | 6 | 1.5e11 | 1 |
| m12 Bulge-NoGD | 1 | 8e3 | 1.5e12 | 12 | 174 | 2.2e11 | 6e10 | 1.6 | 5e9 | 1.4 | 5e8 | 2.8 | 1.6e11 | 1 |
| m13 | 3 | 5e4 | 1.0e13 | 6 | 240 | 7.2e11 | 1e11 | 2.8 | 1.4e10 | 2.8 | 5e9 | 2.8 | 6e11 | 0.1 |
| m14 MR | 1 | 3e4* | 8.5e13 | 5.5 | 600 | 1.52e13 | 2e11 | 3.9 | 2e10 | 3.9 | 1e10 | 3.9 | 1.5e13 | 0.1 |
| m14 HR | 1 | 8e3* | 8.5e13 | 5.5 | 600 | 1.52e13 | 2e11 | 3.9 | 2e10 | 3.9 | 1e10 | 3.9 | 1.5e13 | 0.1 |

*Note*s. Parameters of the galaxy models studied here (Section 2.2): (1) Model name: The number following 'm' labels the approximate logarithmic halo mass. m12 is a disc dominant halo, while m12 Bulge, m13, and m14 are bulge-dominant. The run labelled NoGD have an order-of-magnitude smaller gas disc. (2) $\epsilon_g$: Gravitational force softening for gas (the softening for gas in all simulations is adaptive, and matched to the hydrodynamic resolution; here, we quote the minimum Plummer equivalent softening). (3) $m_g$: Gas mass (resolution element). There is a resolution gradient for m14, so its $m_g$ (with *) is the mass of the highest resolution elements. (4) $M_{halo}$: Halo mass. (5) $c$: NFW halo concentration. (6) $V_{max}$: Halo maximum circular velocity. (7) $M_{bar}$: Total baryonic mass. It is the sum of gas, disc, bulge, and stellar mass for isolated galaxy runs, and the sum of gas and stellar mass in the cosmological runs within 0.1 virial radius. (8) $M_b$: Bulge mass. (9) $a$: Bulge scale length (Hernquist profile). (10) $M_d$ : Stellar disc mass. For CosmoMW and CosmoDwarf runs, this is the total stellar mass within 0.1 virial radius. (11) $r_d$ : Stellar disc scale length (exponential disc). (12) $M_{gd}$: Gas disc mass. (13) $r_{gd}$: Gas disc scale length (exponential disc). (14) $M_{gh}$: Gas halo mass. (15) $r_{gh}/r_{dh}$: Gas halo scale length (beta profile) over dark matter scale length.

would raise the Toomre-$Q$ and stabilize the gas against fragmentation/star formation. This involves no new physics beyond those above (our simulations easily resolve $Q$ and the vertical scale heights and gravitational fragmentation of the cold gaseous discs), but rather different galaxy initial conditions given the same halo properties.

### 2.2 Initial conditions

It is important to note that the over-cooling problem exists over several orders of magnitude in halo mass, not just at $\sim L^*$ where most galaxies first quench, or in massive clusters where the classical 'cooling flow problem' is defined. We therefore consider three fiducial initial conditions (ICs), with halo masses of $1.5 \times 10^{12}$ (m12), $10^{13}$ (m13) and $8.5 \times 10^{13} M_\odot$ (m14), respectively. The DM halo, stellar bulge, stellar disc are set following Springel & White (1999). We assume a spherical, isotropic, NFW (Navarro, Frenk & White 1996) profile dark matter halo [scale lengths (20.4, 93, 218.5) kpc], and Hernquist (1990) profile stellar bulge [scale lengths (1, 2.8, 3.9) kpc]. We also assume exponential, rotation-supported gas and stellar discs (scale lengths (6, 2.8, 3.9) kpc and (3, 2.8, 3.9) kpc, respectively; scale height (0.3, 1, 1.4) kpc for both, gas temperatures initialized to pressure equilibrium Springel 2000), and an extended spherical, hydrostatic gas halo with a beta profile [scale-radius (20.4, 9.3, 21.85) kpc and $\beta = (0.5, 0.43, 0.5)$].[4] All the initial conditions except the m13 case are run adiabatically (no feedback, no cooling) for at least 50 Myr before putting into use, to ensure stability.[5] The properties are summarized in Table 1. In the m12 cases, we explicitly test the effect of different stellar morphology on the cooling, so besides the fiducial (disc-dominated) m12 IC, we also construct bulge-dominated and gas-disc-free ICs (m12 Bulge and m12 Bulge-NoGD).

The initial conditions are set-up to be similar to typical cooling-core systems observed, insofar as this is possible. For example, m12 is a Milky Way-mass galaxy, where the hot gas halo, roughly follows the observed Milky Way profile estimated in Miller & Bregman (2013, 2015) and Gupta, Mathur & Krongold (2017), except that we assume a universal baryonic fraction (0.16) inside twice the virial radius (we do not allow for missing baryons as suggested in the aforementioned papers). It is possible that the solution to over-cooling involves expulsion of a large fraction of the universal baryonic mass; however our intention here is to see if this does occur, not to put it in 'by hand' (moreover, direct observations indicate the full baryon content does exist within similar radii, in external systems, see Planck Collaboration XI 2013; Greco et al. 2015; Lim et al. 2017). Our m13 is the elliptical galaxy (Ell) in Su et al. (2017)[6] The mass and radial distribution of gas, stars, and dark mater are consistent with the observations of similar-mass haloes compiled in Humphrey et al. (2012b) and Anderson, Churazov & Bregman (2016). Our m14 is initialized as as cool core cluster, with a massive central elliptical galaxy, by design. The halo properties and profiles are consistent with typical observed systems of the same mass (Humphrey et al. 2012a; Humphrey & Buote 2013; Su, White & Miller 2013; Su et al. 2015).

The comparison of the X-ray luminosity of our halo to the observations is plotted in the X-ray luminosity–halo mass plane in Fig. 1.[7] The luminosity is calculated using the same method in Ressler, Quataert & Stone (2018), in which the cooling curve is calculated for the photospheric solar abundances (Lodders 2003), using the spectral analysis code SPEX (Kaastra, Mewe & Nieuwenhuijzen 1996) in the same way of Schure et al. (2009) and scaled according to the local hydrogen, helium, and metal mass fractions. All our initial conditions have cooling luminosity within the scatter of the observed X-ray luminosity–halo mass relation (Reiprich &

[4]The hot halo gas rotates at a fix fraction of the circular velocity, which is twice the DM halo streaming velocity in Springel & White (1999). It is then ∼10–15 per cent rotation-supported, and ∼85 − 90 per cent thermal-pressure-supported.

[5]This is necessary for the runs with radially dependent super-Lagrangian refinement scheme (m14), but less relevant for the others.

[6]There is minor typographical error in the tabulated Ell properties in Su et al. (2017). The values in Table 1 are correct.

[7]The X-ray luminosity in our simulation is measured over 0.5–7 keV. That from Reiprich & Böhringer (2002) and Reiprich & Böhringer (2002) is from 0.5–2.4 keV, while that from Kim & Fabbiano (2013) is measured from 0.3–8 keV, but these are corrected given their median estimated spectral slope to the same range we measure. Given that most of the X-ray emission halo in this mass range in below 2 keV, and we are not trying to reproduce a specific halo, the comparison here is not particularly sensitive to this.

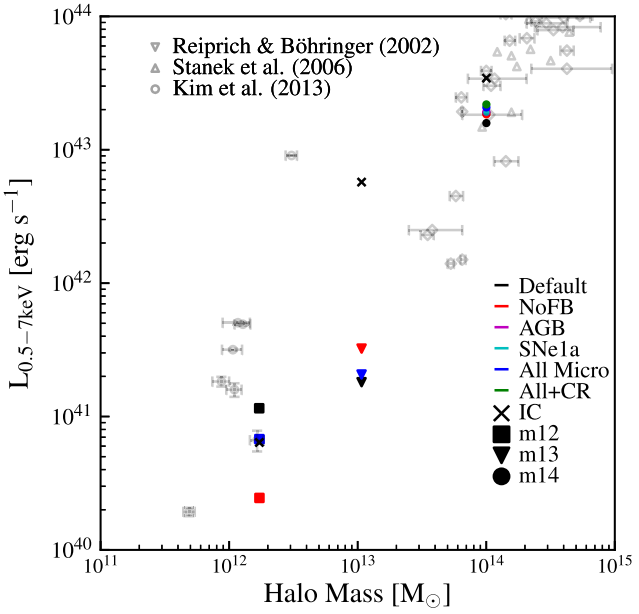


**Figure 1.** The X-ray luminosity (0.5–7 keV) of our initial conditions and the average luminosity of the last 100 Myr of each run are plotted on the X-ray luminosity–halo mass plane in comparison to the observations. All our runs lie reasonably within the scatter of the observed X-ray luminosity–halo mass relation. In m12, the low halo temperature means the X-ray luminosity is significantly influenced by SNe heating; in m13 and m14, the effects of stellar FB on the X-ray luminosity are small (most comes from the initial hot halo). Magnetic fields and CR feedback have little effect on the X-ray luminosity.

Böhringer 2002; Balogh et al. 2006; Stanek et al. 2006; Kim & Fabbiano 2013; Anderson et al. 2015).

In all runs, unless otherwise noted, the initial metallicity is set to solar ($Z = 0.02$) at the core, and drops to $Z = 0.001$ at larger radii.[8] The m14 'Low Metal' run is set to $Z = 0.001$ uniformly. For runs with magnetic fields, the initial magnetic field strength of the gas in the core is set to 0.03, 0.1, and 0.3 μG for m12, m13, and m14, respectively (roughly according to $m_{\rm vir}^{1/2}$). The initial field configuration is azimuthal, and decays as a function of radius.[9] For runs with cosmic rays, the initial CR energy density is set to be in local equipartition with the initial magnetic energy density, at all positions.

Given that m14 is very massive (with much of the gas mass in the extended, non-cooling halo at radii ∼Mpc), resolving it with a uniform gas mass resolution is computationally formidable and not necessary for the convergence of FIRE stellar feedback. We therefore adopt a radially dependent super-Lagrangian refinement scheme in this case: the target gas mass resolution is set to 8000 M$_\odot$ inside $r < 10$ kpc, and increases smoothly $\propto r$ outside this radius up to a maximum $2 \times 10^6$ M$_\odot$ at ∼300 kpc. Gas resolution elements are automatically merged or split appropriately if they move inward/outward, to maintain this mass resolution (to within a factor = 2 tolerance) at all times.

Most simulations have been rerun with different resolutions, with the initial mass resolution differing by at least 2 orders of magnitude. The conclusions are robust in this resolution range, with resolution studies provided in Appendix. The list of runs are summarized in Table 2. We note that the m14 runs with the 'extended fluid microphysics' set or cosmic rays are more expensive, and are therefore run with lower resolution, but within the range where our results appear robust (see Appendix).

[8]The metallicity scales as $Z_\odot(0.05 + 0.95/(1 + (r/R_c)^{1.5}))$, where $R_c$ is set to (20,10,20) kpc for (m12, m13, m14).

[9]Magnetic field strength scales as $(1/(1 + (r/R_c)^2))^\beta$, where $R_c$ = (20, 10, 20) kpc, and $\beta$ = (0.375, 1, 0.375) for (m12, m13, m14)

**Table 2.** List of runs.

| Model | Feedback | Microphysics | CR |
|---|---|---|---|
| m12 | | | |
| NoFB | None | Hydro | No |
| Default | FIRE 2 | Hydro | No |
| All Micro | FIRE 2 | MHD+viscosity+conduction | No |
| Default-Bulge | FIRE 2 | Hydro | No |
| NoFB-Bulge | FIRE 2 | Hydro | No |
| NoGD | | | |
| m13 | | | |
| NoFB | None | Hydro | No |
| Default | FIRE 2 | Hydro | No |
| All Micro | FIRE 2 | MHD+viscosity+conduction | No |
| m14 HR | | | |
| NoFB | None | Hydro | No |
| Default | FIRE 2 | Hydro | No |
| AGB | AGB winds only | Hydro | No |
| SNeIa | Type Ia SNe only | Hydro | No |
| Low Metal | FIRE 2 | Hydro | No |
| m14 MR | | | |
| Default | FIRE 2 | Hydro | No |
| All Micro | FIRE 2 | MHD+viscosity+conduction | No |
| All+CR | FIRE 2 | MHD+viscosity+conduction | Yes |

## 3 RESULTS

### 3.1 Gas masses and phases in cores

Fig. 2 plots the baryonic mass within < 30 kpc ($M_{\rm baryon}^{30\,\rm kpc}$). As gas cools, this increases, with rate $\dot{M}_{\rm baryon}^{30\,\rm kpc}$/M$_\odot$ yr$^{-1}$ ∼ (12, 6, 40) in (m12, m13, m14): there is a competition at increasing mass between higher temperatures (lower cooling rates per particle) and simply larger gas masses available to cool (so this decreases slightly from m12 to m13, then rises rapidly to m14). This is also partially because the viral temperature of 'm13' is roughly at the minimum of the cooling curve (a few times $10^6$ K). Gas with $T > 10^6$ K is mostly hot halo gas from the ICs and is replenished (from larger radius) at small radii as it cools (only a small fraction comes from stellar feedback) so the 'hot' gas mass evolves only weakly. In our default runs the cold and warm gas mass inside < 30 kpc grows rapidly as gas cools. In the 'NoFB' runs this does not appear only because that cold/warm gas turns into stars almost immediately (in ∼1 local free-fall time); stellar FB slows the cold gas consumption time to ∼1–2 Gyr.

The 'All Micro' and 'All+CR' runs are similar to our Default at lower masses, and produce modest effects at m14-mass, with conduction lowering cool inflow rates by a factor ∼2, and CRs suppressing SF in cool gas (and building up additional cold gas) by a similar factor at early times. 'AGB' and 'SNeIa' (runs using only these stellar FB mechanisms) are similar to NoFB, indicating most of the FB comes from massive stars in these runaway-cooling simulations. Our 'Low Metal' run suppresses the build-up of cold+warm gas by a factor ∼1.5–2, owing to the lack of metal-line cooling from the hot gas.

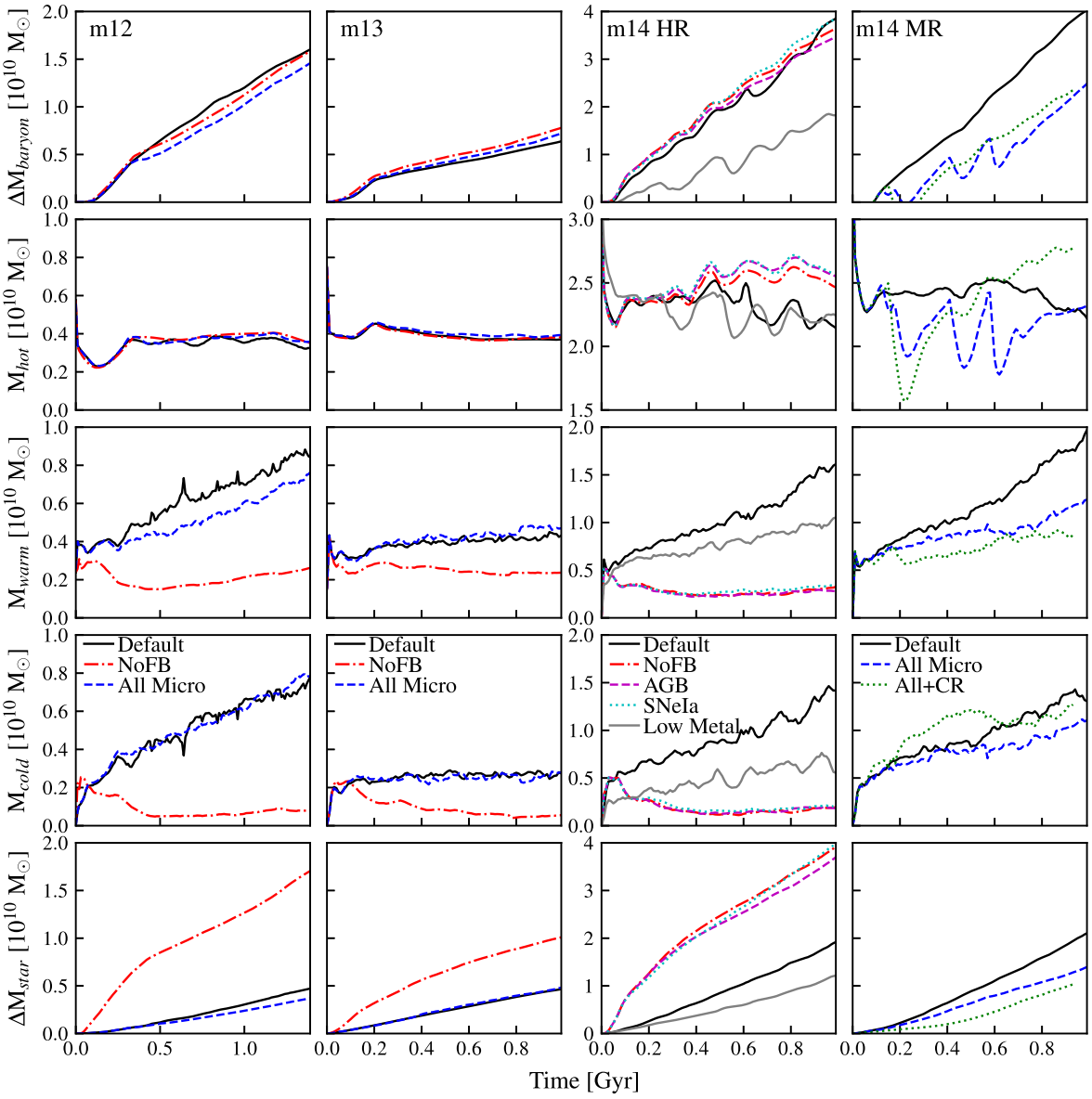


**Figure 2.** Cooling flows in different phases as a function of time. Top row: The baryonic mass variation within 30 kpc ($\Delta M_{\rm baryon}$). Second row: The total hot gas ($>10^6$ K) mass within 30 kpc ($M_{\rm hot}$). Third row: The total warm gas (8000–$10^6$ K) mass within 30 kpc ($M_{\rm warm}$). Fourth row: The total cold gas (<8000 K) mass within 30 kpc ($M_{\rm cold}$). Bottom row: The variation of stellar mass within 30 kpc ($\Delta M_{\rm star}$). In the 'Default' run, the cold and warm gas mass within 30 kpc grows rapidly as gas cools, but in the 'NoFB' runs, any cooled gas almost immediately forms star, so only stellar mass increases. Both the'AGB' and 'SNeIa' runs behave roughly similarly to the 'NoFB' runs, indicating most of the FB comes from massive stars in these runaway-cooling simulations. Without metal-line cooling ('Low Metal' m14 run), the built-up cold and warm gas are all suppressed by a factor of 1.5–2. Both the 'All Micro' and 'All+CR' runs are similar to the 'Default' run in lower mass cases, but have a modest difference in the m14 halo, where conduction suppresses cooling flow by a factor of 2, and CR feedback suppresses SF and allows the build-up of additional cold gas.

### 3.2 Star formation rates

Fig. 3 shows SFRs and sSFRs (averaged in rolling 10 Myr bins): typical $\dot{M}_* \sim 2$–$5\,{\rm M}_\odot\,{\rm yr}^{-1}$ in m12 and m13, and ∼20–40 $\rm M_\odot\,yr^{-1}$ in m14. In sSFR m12 and m14 have $\dot{M}_*/M_* \sim 10^{-10}\,{\rm yr}^{-1}$, m13 $\sim 3$–$5 \times 10^{-11}\,{\rm yr}^{-1}$: none of is 'quenched.' In fact, in m14, the SFRs and cooling flows are accelerating, indicating development of a stronger cooling flow with time. The 'NoFB' runs have an early, rapid burst, where gas in the initial core undergoes runaway collapse and SF, until ∼100 Myr when gas depletion in the core lowers the SFR and subsequent SF comes from gas initially at larger radii, tracking the cooling rate (again, the only-'AGB' and only-'SNeIa' runs resemble NoFB). The Default, 'All Micro,' and 'All+CR' runs initially turn less gas into stars, but this leads to their preserving a larger gas reservoir as cooling continues, until eventually the SFRs are similar to 'NoFB' (cooling-regulated). Again, effectively removing metal-line cooling in the 'Low Metal' run reduces cooling and late SFRs by a factor ∼2.

Note m13 and m14 are entirely bulge-dominated, but still feature high sSFRs. In m12, we explicitly test different initial stellar morphologies: the 'Default-Bulge' (bulge-dominated) and 'Default' (disc-dominated) runs give similar SFRs (with or without stellar FB). We also compare 'NoFB-Bulge-NoGD,' a run with no feedback and no gas disc initially, so the only SF can come from gas

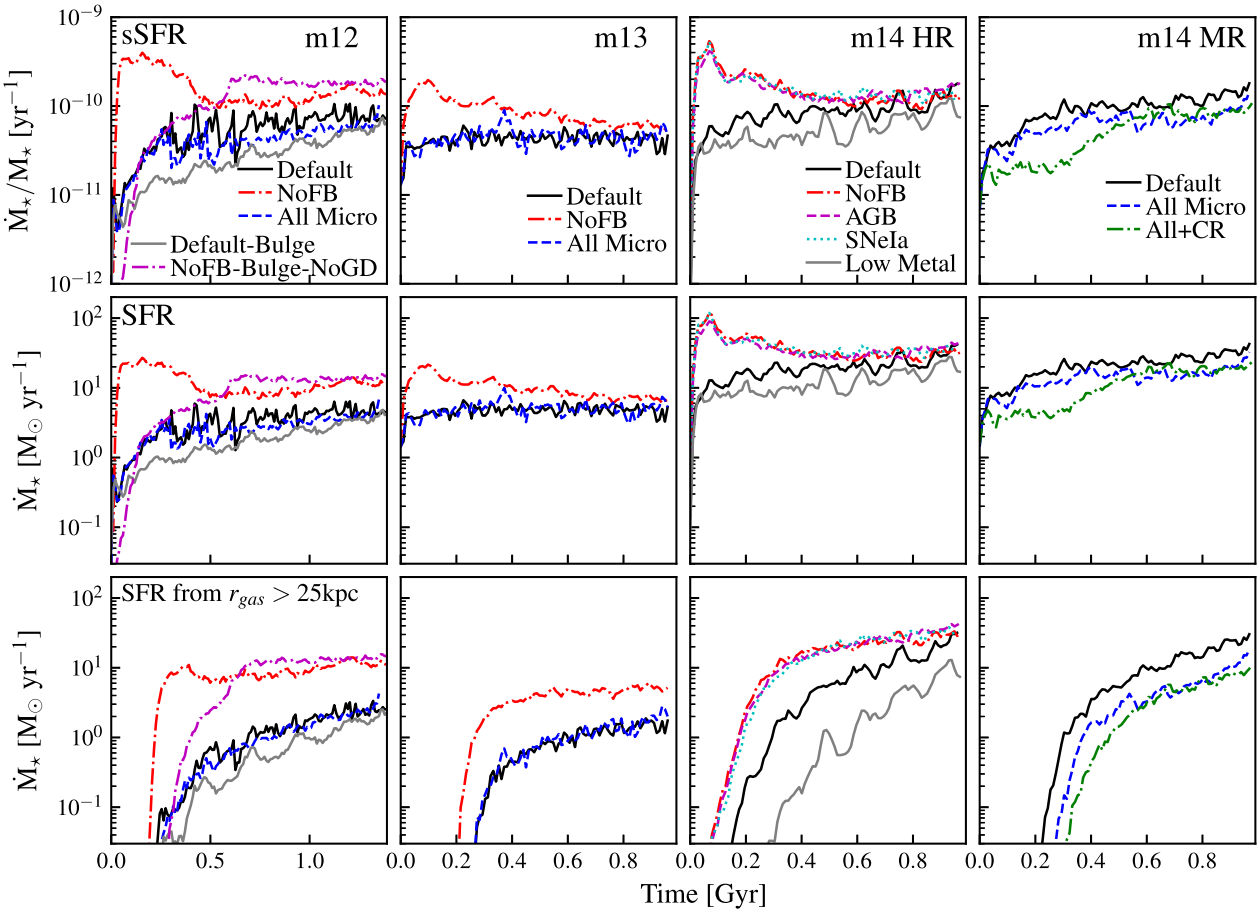


**Figure 3.** Top row: Specific star formation rate (sSFR) as a function of time. Second row: SFR as a function of time, averaged over 100 Myr. Third row: SFRs from gas initially at radii larger than 25 kpc (fuelled by cooling flow). None of the galaxies are quenched. The 'NoFB' runs (similarly also the 'AGB' and 'SNe1a' runs) have an initial rapid star burst, where the initial core gas cools rapidly and forms stars. After ∼100 Myr, gas depletion in the core lowers the initial core SF, and then the subsequent SF tracks the cooling flow gas. The other runs ('Default', 'All Micro', 'All+CR') which initially form fewer stars, and preserve larger gas reservoirs as gas cools, but they all eventually have cooling-regulated SFRs resembling the SFR of the 'NoFB' runs. Suppressing metal-line cooling ('Low Metal' run) lowers the SFR by a factor of 2. With or without feedback, the saturated SFRs of the disc-dominant m12 runs ('Default' and 'NoFB') are very similar to the corresponding bulge-dominant runs ('Default-Bulge' and 'NoFB-Bulge-NoGD'), indicating that the morphology of the stellar distribution has little effect on star formation.

cooling from large radius. Even in this case, while the initial SFR is lower owing to the lack of initial gas supply, the SFR saturates to the same value as the 'NoFB' (disc-dominated, with gas disc) run.

### 3.3 Cooling times and thermal stability

Fig. 4 shows the cooling time ($\tau_c$) of gas hotter than $10^5$ K as a function of radius averaged over the 80–90th Myr since the beginning of the simulations. Within ∼(200, 40, 100) kpc in (m12,m13,m14) cooling times are short compared to the Hubble time (at large radii temperatures are higher, metallicities and densities lower, so $\tau_c$ rise rapidly). The ratio of cooling time to dynamical time ($\tau_c/\tau_d$) is also shown.[10] In m12, the halo is not fully in the 'hot mode' given its relatively low virial temperature, so $\tau_c/\tau_d$ is steady at ∼10 at large radii (actually highest in the core, where $\tau_d$ becomes very short). Here, m13 is the 'most stable' case (consistent with its lower sSFR), with $\tau_c/\tau_d \sim 100$ from ∼5 to 100 kpc. The higher density of halo gas in m14's core gives $\tau_c/\tau_d \sim 20$ within ∼50 kpc. Note these all exhibit rapid cooling, despite $\tau_c/\tau_d \sim 20$–100 being the lowest values in m13 and m14, compared to the often-quoted critical value of ∼10 in the literature (Sharma et al. 2012; Voit et al. 2017). This partly owes to the structure being much more strongly multiphase here – the cooling gas has already cooled out of the thermally unstable temperature range (to $T < 10^5$ K; not included in Fig. 4), which makes the cooling rate of the remaining gas here appear longer.

Differences between physics variations are consistent with the SFR and cool gas mass plots above. Note the 'NoFB' runs actually feature the longest cooling times in the centre, as FB injects metals and dense gas into the hot phase, lowering its cooling time.

### 3.4 Cooling versus energy input

Fig. 5 compares cooling rates (all cooling channels added) and energy input rates (adding photo-ionization, photo-electric, cosmic ray, and the energy input from SNe and winds) within <30 kpc. Cooling always exceeds heating. As expected, energy input in our 'Default' run exceeds 'NoFB' owing to higher energy input from e.g. SNe, but 'Default' also maintains an even higher cooling rate.

[10] $\tau_c = E_{\rm thermal}/\dot{E}_{\rm cool}$, and $\tau_d = (r^3/GM_{\rm enc})^{1/2}$.

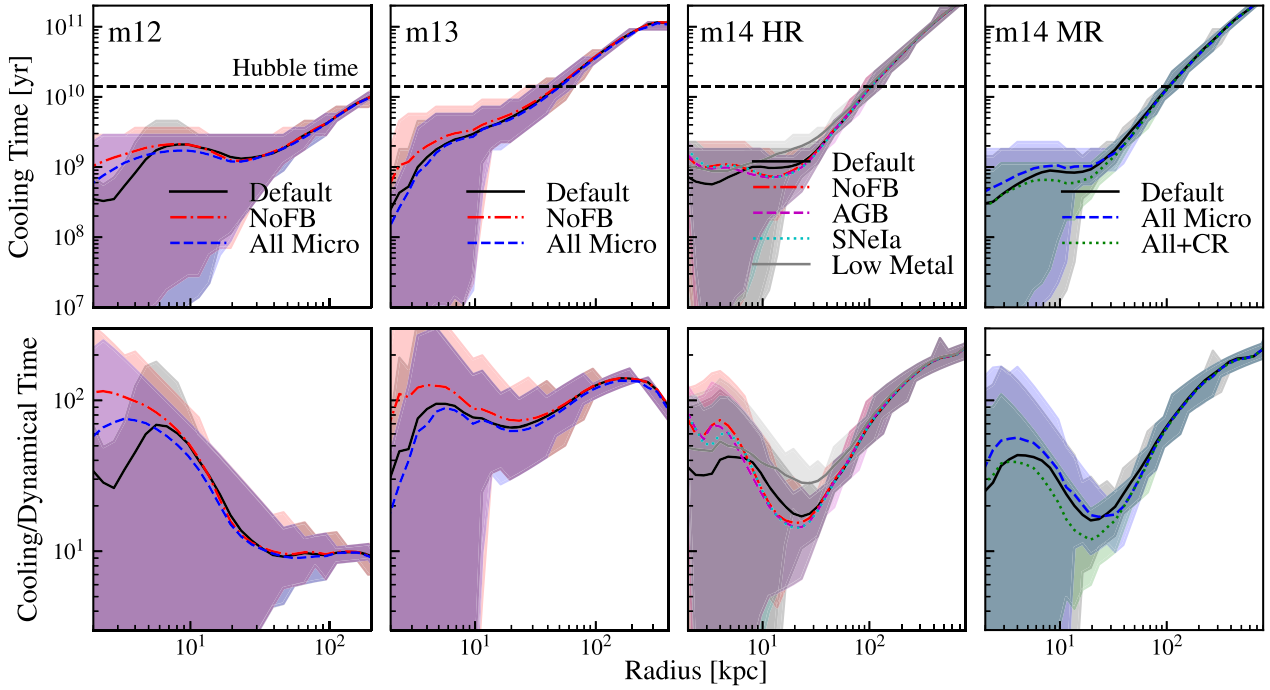


**Figure 4.** Top row: Cooling time of gas hotter than $10^5$K, as a function of radius. Black dashed lines indicate the Hubble time. Bottom row: Ratio of cooling time to dynamical time, as a function of radius. The lines are mass-weighted averages and the shaded region covers 10–90 per cent of the distribution. All the quantities are averaged over the 80–90th Myr since the beginning of the simulations. The 'Default' runs have lower $\tau_c/\tau_d$ than the 'NoFB' runs within ∼10 kpc since FB enriches the gas with metals. Both the 'AGB-only' and 'SNe1a-only' runs resemble the 'NoFB' run. Conduction and CR feedback do not have a significant effect. Suppressing metal-line cooling ('Low Metal' run) increases $\tau_c$ by a factor of 2.

'Low Metal' has a factor ∼2 lower cooling rate without metal-line cooling. Magnetic fields and viscosity produce negligible effects on their own. Conduction has weak effects here. We quantify the energy injection from each stellar feedback mechanism:[11] SNe Ia, O/B or AGB winds, and CRs contribute relatively little to total heating, with core-collapse SNe present (given that our galaxies are not quenched and have high SFRs). However, even the core-collapse input is only ∼1/4–1/3 of the cooling rate. SNe Ia input ∼20x lower energy compared to core-collapse, and AGB winds produce a factor ∼2–3x lower energy injection rate compared to Ia's.

The total X-ray luminosity of each run (in Fig. 1) scales with the cooling rates, as expected. In m12, the low halo temperature means the X-ray luminosity is significantly influenced by SNe heating; in m13 and m14, the effects of stellar FB on the X-ray luminosity are small (most comes from the initial hot halo) (van de Voort et al. 2016).

### 3.5 Energetic balance

Fig. 6 compares the specific energy in thermal, magnetic, CR, and turbulent forms averaged over 90–100th Myr.[12] Within a few kpc, turbulent energy dominates (the turbulence is supersonic and super-Alfvénic), consistent with studies of the ISM inside galaxies (Su et al. 2017, 2018b). At larger radii, thermal energy dominates (by at least ∼1 dex). Note at very large radii (≫10 kpc), the magnetic energies are simply dominated by the ICs, since the flow is approximately laminar so the classical global dynamo amplification time is many orbital times (≫Gyr). The same is true for the CR, as it does not have a chance to diffuse or stream to ≫10 kpc within the simulation time.

## 4 DISCUSSION: WHY DON'T WE QUENCH?

Here, we briefly discuss why none of the mechanisms in Section 2.1 produces quenching, at any mass we survey.

### 4.1 Stellar feedback

#### *4.1.1 Young/massive stars*

Feedback from massive stars clearly reduces the rate at which gas within the galaxy turns into stars, self-regulating to a gas consumption time ∼1–2 dex longer than a case without feedback (bringing it into agreement with the observed Schmidt–Kennicutt relation Schmidt 1959; Kennicutt 1998), and drives local outflows from the disc (e.g. Faucher-Giguère, Quataert & Hopkins 2013; Hayward & Hopkins 2017; Orr et al. 2018). However, cooling from the hot halo onto the galaxy eventually builds up the core gas mass and the SFR runs away.

Stellar feedback fails to suppress cooling in massive haloes on long time-scales for three reasons. (1) Even with the elevated (much higher than observed) runaway SFRs, the total energy input from SNe is ∼1/4–1/3 of the cooling rate (Fig. 5). (2) The energy is injected locally in the galaxy core, either as slow-moving (sub-$V_{esc}$) cool gas or fast but tenuous hot gas, so is rapidly decelerated

[11]The plotted SNe energy input rate includes $10^{51}$erg per event. The plotted stellar wind energy input rate is actually an upper bound, since a relative velocity between gas and stars is assumed to be 300 km s$^{-1}$ for purposes of this post-processing estimate (it is calculated self-consistently in the code).
[12]Turbulent energies are measured using the method from Su et al. (2017) which attempts to separate turbulent motion from outflows and non-circular but bulk orbital motion.

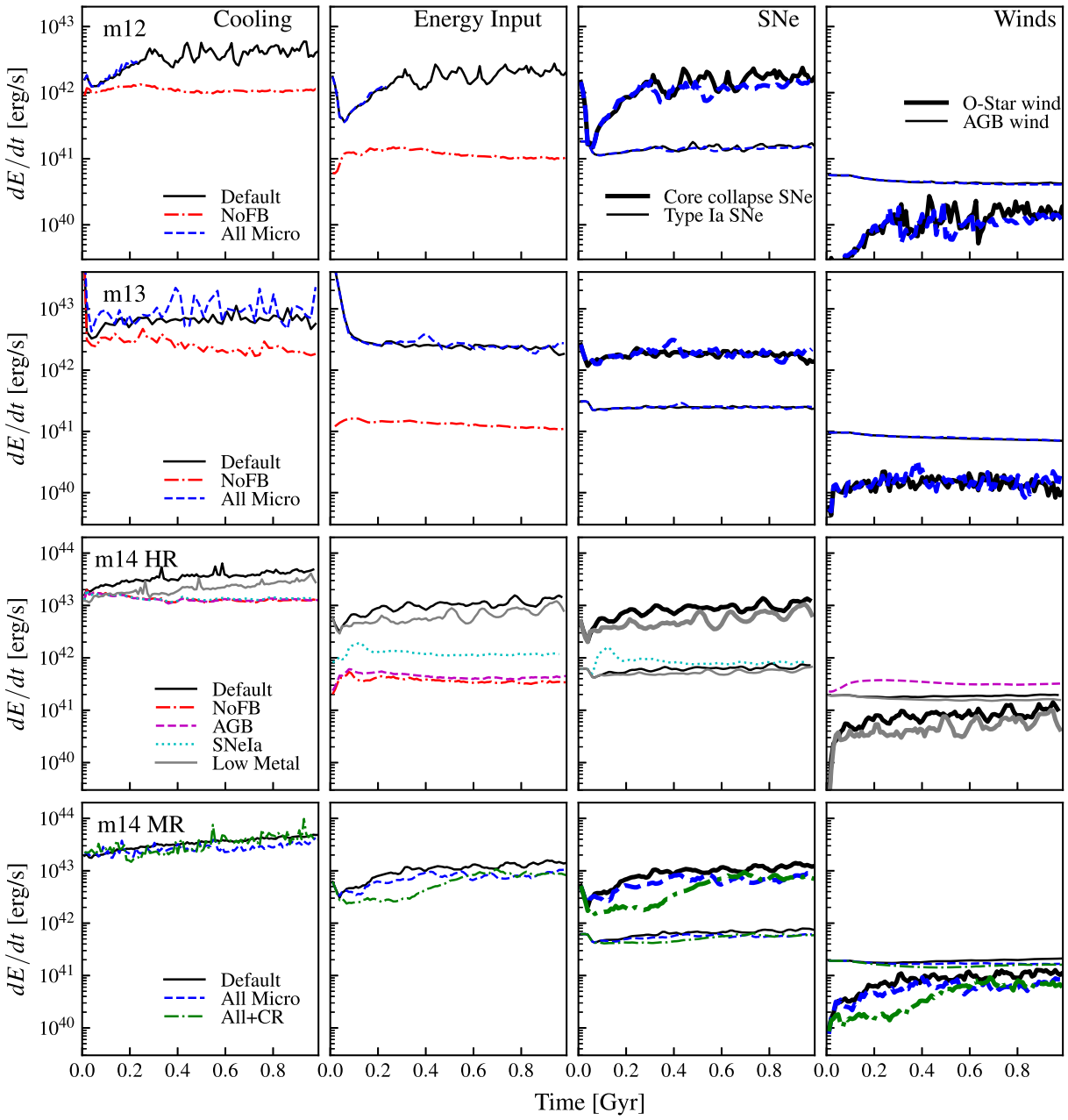


**Figure 5.** Energy input, cooling, and feedback, as a function of time in m12 (top row), m13 (2nd row), m14 HR (3rd row), and m14 MR (bottom row) runs. Left column: Total cooling rate within 30 kpc as a function of time. Middle left column: Energy input rate, including UV background, cosmic ray heating, dust collisional heating, photo-electric heating, and the energy input from SNe and winds, within 30 kpc. Middle right column: The energy input of core collapse SNe (thick lines) and type Ia SNe (thin lines) coloured by simulation. Right column: The energy input of O-Star winds (thick lines) and AGB winds (thin lines), coloured by simulation. The plotted energy input from SNe assumes $10^{51}$ erg per event, while winds energy input is calculated assuming the relative velocity between stars and gas is 300 km s$^{-1}$ (upper bound). Cooling rates are always higher than heating rates. Core collapse SNe input energy at ∼1/3–1/4 the cooling rate. Type 1a SNe input 20x lower energy (time-averaged) compared to core-collapse in these runs, and AGB winds produce 2–3x lower energy injection compared to Ia's. With suppressed metal-line cooling (Low Metal), the cooling rate is lowered by a factor of 2.

and does not couple outside the cooling radius (e.g. we verify that outflows in m14 rarely reach past ∼20–30 kpc; see also Muratov et al. 2015; Anglés-Alcázar et al. 2017b for more detail of the wind properties and the baryonic cycles in FIRE simulations). (3) As is commonly seen in galactic fountains (e.g. Silk 2009; Silk 2010, 2011; Fraternali et al. 2013), the outflows carry relatively dense, metal-rich gas into the halo, which increases the net cooling rate (Figs 3 and 5) as it mixes with a larger mass of less-dense and lower metallicity gas. The effect of metal enrichment on the cooling flows will be even clearer if the metal in the simulations were solely from the stellar feedback, instead of partially from the initial conditions. In fact, we do see a factor of 2 lower SFR and core baryonic mass in the run with negligible initial metallicity ('Low Metal' run).

Fig. 7 explicitly compares three variations of m12 that have initial central ($<10$ kpc) gas fractions $f_{\rm gas} =$

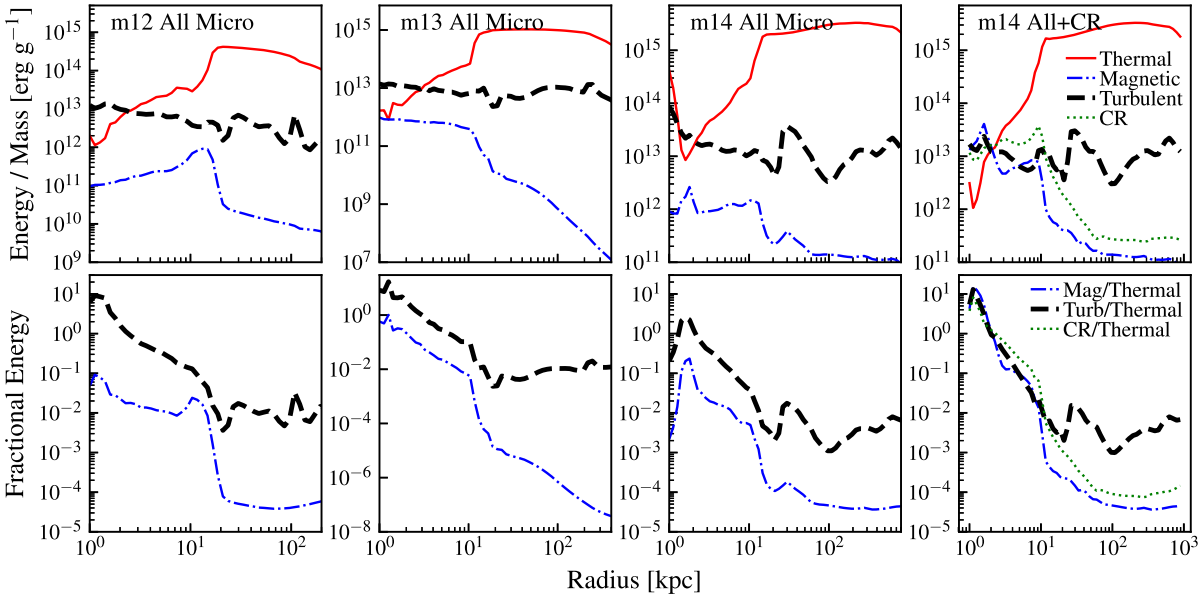


**Figure 6.** Top row: The comparisons of thermal, magnetic, CR, and turbulent energy per unit mass, averaged over the 90–100th Myr. Bottom row: The comparison of the ratio of magnetic, CR, and turbulent energy to thermal energy. For $r \lesssim 3$ kpc, magnetic energy is sub-dominant to turbulent energy, except in the 'All+CR' run, where CR and magnetic energy reach equipartition with turbulent energy, showing that the turbulence is mostly super-Alfvénic and supersonic. At larger radius, $E_{\rm thermal} > E_{\rm turb} > E_{\rm CR} \gtrsim E_{\rm mag}$, so magnetic and CR energy have little effect on cooling flow.

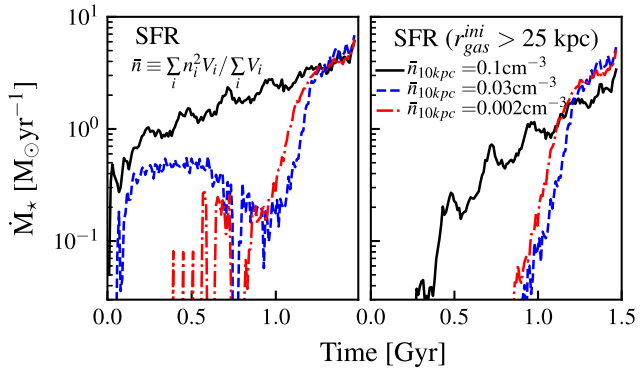


**Figure 7.** The SFR of m12 runs with different core gas density within 10 kpc. $\bar{n}$ is defined as $\sum_i n_i^2 V_i / \sum_i V_i$. The SFRs are significantly suppressed if the core gas density is low. However, in all the cases, the SFR eventually catches up as the gas supplied by cooling flows builds up.

$\bar{n}_{\rm gas}/n_{\rm star} \approx 0.05$, 0.01, 0.004 (mean central densities $\bar{n}/{\rm cm}^{-3} =$ 0.1, 0.03, 0.002). If there is little or no initial gas inside $<$ 10 kpc, then (as expected) the initial SFR is suppressed strongly. More interestingly, we also see the SFR from gas with initial $r > 25$ kpc is suppressed for $\approx$1 Gyr: this partially demonstrates how winds from SF in the disc (now absent) enhance cooling/inflow through enriching the halo gas. After $\approx$1 Gyr, however, cooling runs away and SF is dominated by the newly cooled gas.

### *4.1.2 SNe Ia and AGB winds*

Considering just Ia's or shock-heated AGB winds, we show in Fig. 5 that the feedback energy injection rate is even lower compared to cooling (by more than an order of magnitude), exacerbating problem (1) above, and problems (2) and (3) remain. The explosions decelerate rapidly, and mixing the highly metal-rich gas[13] promotes cooling.

Some previous studies (see Section 2.1.1) appeared to reach different conclusions. However, these were largely based on simple analytic energetics arguments, so could not follow the non-linear effects of (2) and (3) above. Moreover, even for (1), although we find energy injection rates 'per star' from SNe Ia and shocked AGB ejecta similar to these previous estimates, we find that clumping in the gas, mixing of winds, and cooling from larger radii enhances the central densities (and corresponding cooling rates) beyond the relatively low values assumed in those papers, rendering the heating insufficient.[14] Even if the initial gas density within 10 kpc is lowered to 0.002 cm$^{-2}$ (as shown in Fig. 7) , the stellar feedback (mostly SNe Ia given the SFR $\sim$ 0 within $\sim$1Gyr) can at most suppress the SFR for 1 Gyr, after which the core gas density builds up and the cooling runs away.

We note that the supernovae implementation deal with the unresolved Sedov–Taylor phases following Hopkins et al. (2018a),

[13]It is important to note that while AGB ejecta are much less metal-rich than pure Ia ejecta, it is still approximately solar (the mass-weighted mean stellar metallicity in massive galaxies) or somewhat more enriched in C and O (the primary coolants), and carries much larger mass, so it is less rapidly diluted. In fact, for a $\sim$10 Gyr old stellar population, the total metal return rate is higher by a factor of $\sim$4 in AGB winds, compared to Ia SNe.

[14]Consider m12: Type Ia SNe and shocked AGB winds input energy at $\sim 1.5 \times 10^{41}$ erg s$^{-1}$ and $\sim 5 \times 10^{40}$erg s$^{-1}$, respectively, roughly consistent with the value quoted in Conroy et al. (2015) for a similar-mass galaxy. However, in m12, the average effective gas density within $<$10 kpc can be $\sim$1 dex higher than the 0.01 cm$^{-3}$ assumed in Conroy et al. (2015). Note that clumping matters here: cooling rates scale $\propto n^2$, so the density that matters is $\bar{n} \equiv \langle n^2 \rangle^{1/2}$, which is a factor of several higher in our runs than $\langle n \rangle$ inside $R_{\rm cool}$. Even assuming primordial gas (ignoring metal-cooling, (3) above), and ignoring whether heating can reach large radii (2), the analytic scaling from Conroy et al. (2015) then predicts $\dot{e}_{\rm cool}/\dot{e}_{\rm heat} \sim 40\,(f_{\rm gas}/0.01)\,(\bar{n}/0.1\,{\rm cm}^{-3})$ within the galaxy.

which assumes negligible surrounding pressure (Cioffi, McKee & Bertschinger 1988). However, if a SN happens in CGM where the surrounding pressure is potentially high, it has to do extra pdV work as the blast wave expands, which lowers the coupling momentum. Therefore, assuming negligible surrounding pressure only means overestimating the effect from SNe, but, even so, the effect of SNe is still limited.

### 4.2 Magnetic fields, conduction, and viscosity

It is expected that magnetic fields alone cannot quench or suppress cooling flows: since they do not (directly) alter cooling, magnetic pressure would have to 'hold up' the cooling gas in the halo, requiring un-physically strong fields (plasma $\beta \ll 1$, while $\beta \gg 1$ is observed and expected).[15] Moreover even in that case, in 3D all field orientations are Rayleigh–Taylor unstable (see e.g. Stone & Gardiner 2007). And extensive previous work has shown the fields have a small effect on galaxy-scale SF (Su et al. 2017, and references therein). Viscosity has equally small effect: it is weak and, if anything, slows and mixes outflows, slightly enhancing cooling.

Because of its strong temperature dependence (diffusivity $\propto T^{5/2}/\rho$), Spitzer–Braginskii conduction is expected to play a role only in the most massive haloes, and we confirm this. The diffusion time for bulk heat transport across a distance $\sim R$ is $\sim 0.3\,\mathrm{Gyr}\,(R/10\,\mathrm{kpc})^2\,(n/0.01\,\mathrm{cm^{-3}})\,(T/10^7\,\mathrm{K})^{-5/2}$ – only comparable to cooling times (inside the cooling radius $R_{\rm cool}$) in our most massive halo (m14). Moreover in a turbulent medium, eddies mix with approximate diffusivity $\sim v_{\rm eddy}(\lambda_{\rm eddy})\,\lambda_{\rm eddy}$; if we assume transsonic, Kolmogorov turbulence then micro-physical conduction dominates over turbulent only at scales $\lesssim 50\,\mathrm{pc}\,(\lambda_{\rm driving}/10\,\mathrm{kpc})^{1/4}\,(T/10^6\,\mathrm{K})^{3/2}\,(n/0.01\,\mathrm{cm^{-3}})^{-3/4}$. This is only larger than our resolution (and correspondingly, microphysical conduction dominates over both turbulent and numerical conduction with $\kappa \sim c_{\rm s}\,\Delta x$ only) if $T \gtrsim 10^7\,\mathrm{K}\,(m_{\rm g}/10^4\,\mathrm{M_\odot})^{0.2}\,(n/0.01\,\mathrm{cm^{-3}})^{0.3}$. So except in the outer regions of our most massive halo, the dominant heat transport/mixing is not Braginskii conduction.

In m14, conduction is not negligible, but it only lowers the inflow rates and SFR from halo gas by a factor $\sim$2. The effect is modest because (1) Lower temperatures and higher densities inside the core make conduction globally less efficient. (2) Once something (e.g. dense outflows, turbulence) triggers thermal instability, cool 'clouds' radiate efficiently and develop sharp density contrasts with the ambient medium so the conduction becomes saturated and can only out-compete cooling in the very smallest clouds (McKee & Cowie 1977) – in fact, recent work (Brüggen & Scannapieco 2016; Armillotta et al. 2017; Hopkins et al. 2019) has shown that conduction often actually increases cold cloud lifetimes via cloud compression suppressing surface-mixing instabilities, in this limit. (3) Magnetic fields modestly suppress perpendicular transport (quantified in Fig. 8, where in cores the conductivity is suppressed by a factor $\sim$2, and in outskirts a factor $\sim$3–10, levels (Voit 2011 and Wagh et al. 2014 argue will suppress many effects of conduction).

Conduction may still be important to un-resolved small-scale thermal instabilities in hot haloes (as noted above), but we note first that at finite resolution our numerical diffusion dominates (so if anything we overestimate true conductivities), and secondly, most of the discussion in the literature on this question has focused on the regime where there is some global heat source injecting energy sufficient to offset cooling losses (where conduction plays the role of local heat transport into un-resolved clouds). When there is no heat input from feedback (or the heat input is less than the cooling rate, as in our default case here), then conduction does not significantly modify the consequences of the small-scale thermal instability (cold clouds 'absorb' the conducted heat, condense and rain out efficiently; see Voit & Donahue 2015).

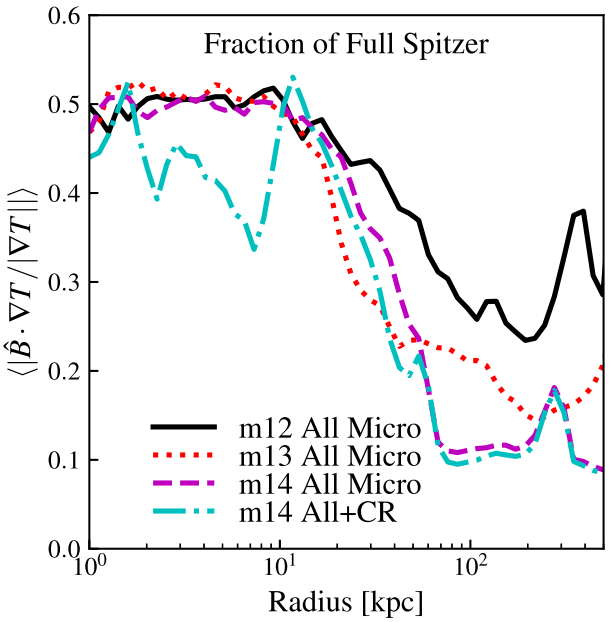


**Figure 8.** The effective fraction of Spitzer conductivity, as a function of radius, estimated as the mass weighted average of $|\hat{B}\cdot\nabla T/|\nabla T||$. This fraction drops as a function of radius. In the core region, the conductivity is effectively half of the full Spitzer value, while it drops to 0.1–0.3 of it at large radii.

### 4.3 Cosmic rays from stars and LSS (not AGN)

CRs from SNe have limited effect here: (1) Direct heating (from e.g. streaming/hadronic interactions) cannot compete with cooling – the total CR energy injection rate is $\sim$10 per cent of SNe, so order-of-magnitude less than cooling rates (Figs 2 and 5). (2) CR pressure is approximately in equipartition with magnetic energy (Fig. 6), so does not have a dramatic dynamical effect 'holding up' the halo (though it can help accelerate some winds from massive stellar feedback at the disc). (3) Like other feedback from massive stars, CRs do help suppress the collapse/SFR from cold gas locally (see Fig. 3, 'All+CR' m14), but this leads to a pile-up of that gas (Fig. 2) from large-scale cooling, which runs away. (4) It requires some fine-tuning to make CRs do all their 'work' around cooling-flow radii, rather than diffusing out.[16]

[15] We confirm that we can, in principle, 'quench' if we initialize enormously strong fields, but this requires magnetic field strengths exceeding self-gravity which simply 'explode' the halo gas in a dynamical time.

[16] While not shown here, we have experimented with e.g. variations in the CR diffusivity. If it is much lower ($\kappa_{\rm cr} \lesssim 10^{28}\,\mathrm{cm^2\,s^{-1}}$), CRs are trapped in the inner regions ($<10$ kpc) and suppress SF from cold gas in the short term more efficiently. But precisely because of this trapping and build-up of cold gas the CR energy is then mostly lost to catastrophic hadronic interactions, so the CRs have weaker long-term effects. For much higher diffusivity

Although we do not directly model it, CRs from structure formation will have all these limitations as well. We find, for example, that we can 'quench' via initializing an enormous CR density, but only if this is so large it overcomes gravity and blows out most of the halo gas. But this (a) does not resemble observed haloes, and (b) is not possible from structure formation, since (by definition) only a fraction of the gravitational energy (in e.g. shocks) goes into CRs.

### 4.4 Morphological quenching

Although galaxy colours and morphologies are correlated, that does not mean morphology causes quenching; we find that changing the stellar morphology of the galaxies here has very weak effects on their cooling/SF properties. (1) Changing the morphology of stars has no direct effect on cooling rates, so even if it somehow quenched SF within the galaxy, cool gas would still pile up. For m14, this would give $> 3 \times 10^{11}\,M_{\odot}$ of molecular gas by the end of our simulation (Fig. 2), orders of magnitude higher than usually observed (e.g. Salomé et al. 2006; Popping, Behroozi & Peeples 2015) (reference added). (2) Moreover, if gas is self-shielding (i.e. reaches surface densities $\gtrsim 10\,M_{\odot}\,pc^{-2}$, requiring just $\sim 3 \times 10^{9}\,M_{\odot}$ of gas within $< 10$ kpc, vastly less than that produced by the cooling flow), then it can cool to $T \ll 10^{4}$ K, at which point Toomre $Q \ll 1$ will always be true in the cold gas for any mass distribution with a physically plausible rotational velocity. Indeed, the few known BCGs with large gas reservoirs $> 10^{10}\,M_{\odot}$, as predicted by our simulations here, all have observed SFRS $\sim$10–100 $M_{\odot}\,yr^{-1}$, like our m14, and obviously not quenched (see O'Dea et al. 2008). (3) The effect of the stellar morphology on $Q$ is quite weak: somehow converting the entire MW stellar disc and all DM within $<10$ kpc to a compact bulge or point mass would only increase the $Q$ at the solar circle by $\sim$50 per cent (see e.g. Romeo & Wiegert 2011). In fact the stronger, but still weak, effect of changing an initial stellar disc to a bulge in Fig. 3 comes from slightly reducing the impact of stellar feedback on the inner halo (by making the stars older and less extended, so they have weaker feedback that reaches less far into the halo). (4) All our m13 and m14 runs, and of course observed massive galaxies, are completely bulge-dominated, yet still feature a 'cooling flow problem.'

## 5 CONCLUSIONS

In this paper we used high-resolution, idealized, isolated galaxy simulations including detailed physical treatments of star formation, stellar feedback, and ISM/CGM/ICM microphysics (cooling, magnetic fields, conduction, cosmic rays, etc.) to explore and quantify the quenching and cooling flow problems – in the absence of AGN feedback – across a range of halo masses from $\sim 10^{12}$ to $10^{14}\,M_{\odot}$. We started the simulations from initial conditions resembling observed cool-core clusters at low redshift and tested whether we could maintain stably quenched galaxies consistent with observations. We specifically explored several 'non-AGN' quenching or cooling-flow 'solution' mechanisms, which have been previously proposed in the literature (e.g. feedback from old stellar populations in Type Ia SNe or shocked AGB ejecta, heat transport from the outer halo via conduction, cosmic rays from Ia's or structure formation, or 'morphological quenching'). None of these mechanisms resolve the fundamental problem of overcooling and excessive star formation in massive galaxies, at any mass scale, that we simulate. The main effects of these physics are as follow:

(i) Stellar feedback alters the balance of cold/warm gas and suppresses SFRs for a given cold gas mass (i.e. controls the location of galaxies on the Kennicutt–Schmidt relation). However, it has only weak effects on cooling from the outer halo, and in fact tends to enhance cooling in the inner halo, as denser, more metal-rich ejecta mix and promote cooling in halo gas. This applies to all stellar feedback mechanisms (Ia's, AGB ejecta, feedback from young stars where present).

(ii) Magnetic fields and Braginskii viscosity have minor effects on the global cooling and inflow rates.

(iii) Conduction is, as expected, only important to bulk cooling/inflow in the most massive haloes ($\geq 10^{14}\,M_{\odot}$). Even there, the effects are modest, reducing inflow rates by a factor $\sim$2, owing to a combination of saturation effects, suppression by magnetic fields, and inefficient conduction in the cores once runaway cooling begins.

(iv) Cosmic rays from SNe (and shocked stellar winds) alone can, like other stellar feedback mechanisms, modestly reduce the SFR in cool gas already in/near the central galaxy, but their bulk energetics are insufficient to suppress cooling flows. We expect the same for CRs from structure formation.

(v) Stellar morphology has essentially no effect on cooling rates and only weakly alters star formation either in pre-existing gas discs or in gas discs formed via runaway cooling. Making our galaxies entirely bulge-dominated does not make them quenched.

Precisely because the effects of the above physics are weak, our summary of the quenching 'maintenance' and/or 'cooling flow' problems is consistent with many previous studies that treated some of the physics above in a more simplified manner:

(i) At all halo masses $\gtrsim 10^{12}\,M_{\odot}$, we find efficient cooling of halo gas in cores, with cooling luminosities similar to observations (where available), but excessive cooling/cold gas masses and SFRs in central galaxies (from $\sim 5\,M_{\odot}\,yr^{-1}$ in $\sim 10^{12-13}\,M_{\odot}$ haloes, to $\sim 50\,M_{\odot}\,yr^{-1}$ by $\sim 10^{14}\,M_{\odot}$).

(ii) The excess gas comes, in an immediate sense, from an overcooling core where higher densities and metallicities (enhanced by earlier generations of recycled galactic winds) produce rapid cooling and multiphase CGM structure. Although the 'median' cooling times in this core can be large compared to dynamical times ($t_{\rm cool}/t_{\rm dyn} \sim 100$), the fastest cooling (denser, more metal-rich) material reaches $t_{\rm cool}/t_{\rm dyn} \ll 10$–30 – this is what rapidly forms stars.

(iii) The core providing immediate fuel for SF can have a relatively small extent $\lesssim$30 kpc, but gas within this radius at later times originates from larger radii (up to $\sim$100 kpc) and migrates slowly inwards before 'runaway,' so it may be possible to 'starve' the cooling flow on longer time-scales by suppressing just cooling/inflow from $\sim$30 to 100 kpc.

(iv) It is possible, in principle, to temporarily quench galaxies in this mass range if one can remove all their dense gas within $\lesssim$10 kpc. This will suppress star formation for $\approx$1 Gyr (surprisingly independent of halo mass), before runaway cooling from the extended halo restores the excessive SFRs.

Our simulations have several limitations upon which future work could improve. We wished to construct idealized, controlled experiments so did not evolve fully cosmological simulations – we do not expect this to alter the fundamental conclusions above, but it could introduce additional important effects (e.g. stirring turbulence in

($\kappa_{\rm cr} \gtrsim 10^{30}\,cm^{2}\,s^{-1}$), CRs free stream completely out of the haloes with negligible interaction with gas.

haloes via structure formation). We of course have finite resolution, so micro-physical phase structure in the CGM remains un-resolved and could alter the effective large-scale cooling rates. Our treatment of some physics (SNe, cosmic rays) is necessarily approximate (sub-grid) but it would require truly dramatic qualitative changes to our assumptions to reverse our conclusions. We also did not include rare events like hypernovae. However, as we have shown in Su et al. (2018c), the net effects of hypernovae resemble those of overlapping supernovae, which readily occur in our simulations due to clustered star formation.

Due to computational expense we could only run ∼20 high-resolution simulations, which means we could not explore the potential diversity of properties of different haloes at the same mass. We also did not explore various initial cosmic ray and magnetic field configurations. Regarding cosmic rays, we refer readers to Su et al. (2018a), in which we explored how cooling flows are affected by the cosmic-ray profiles caused by various CR injection rates. In the simulations presented in this work, the magnetic fields in the core regions are quickly amplified to the saturation values, independent of the seed values. The initial magnetic field strength and configuration at larger radii could be sensitive to the seed magnetic field, but we do not expect this to alter any of our conclusions.[17]

Most obviously, we neglect AGN feedback, in various forms (jets, bubbles, winds, radiation, etc.). This work furthers the argument that something – perhaps AGN – beyond the 'known' physics we include here, must be at work. We emphasize that many or all of the physics explored here (e.g. magnetic fields, cosmic rays, etc.) may indeed play a critical role in AGN feedback, even if they do not dramatically alter cooling flows absent an AGN. In future work, we will explore generic classes of AGN feedback models proposed in the literature, in simulations incorporating the additional physics here which must (in reality) be present as well, in order to better understand the non-linear interactions of different feedback and ISM/CGM/ICM physics.

## ACKNOWLEDGEMENTS

We thank Eliot Quataert for useful discussion. The Flatiron Institute is supported by the Simons Foundation. Support for PFH was provided by an Alfred P. Sloan Research Fellowship, NASA ATP Grant NNX14AH35G, and NSF Collaborative Research Grant #1411920 and CAREER grant #1455342. CAFG was supported by NSF through grants AST-1412836, AST-1517491, AST-1715216, and CAREER award AST-1652522, by NASA through grant NNX15AB22G, by CXO through grant TM7-18007, and by a Cottrell Scholar Award from the Research Corporation for Science Advancement. DK was supported by NSF grant AST-1715101 and the Cottrell Scholar Award from the Research Corporation for Science Advancement. VHR acknowledges support from UC-MEXUS and CONACyT through the postdoctoral fellowship. Numerical calculations were run on the Caltech compute cluster 'Wheeler,' allocations from XSEDE TG-AST130039 and PRAC NSF.1713353 supported by the NSF, and NASA HEC SMD-16-7592.

[17]We have an m12 run that was accidentally started a magnetic field amplitude 2 orders of magnitude greater than the fiducial value. With this much higher value, the star formation was burstier, but the galaxy was still not quenched.

## REFERENCES

Agertz O., Kravtsov A. V., 2016, ApJ, 824, 79
Anderson M. E., Gaspari M., White S. D. M., Wang W., Dai X., 2015, MNRAS, 449, 3806
Anderson M. E., Churazov E., Bregman J. N., 2016, MNRAS, 455, 227
Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017a, MNRAS, 464, 2840
Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017b, MNRAS, 470, 4698
Armillotta L., Fraternali F., Werk J. K., Prochaska J. X., Marinacci F., 2017, MNRAS, 470, 114
Balbus S. A., 2000, ApJ, 534, 420
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Balogh M. L., Babul A., Voit G. M., McCarthy I. G., Jones L. R., Lewis G. F., Ebeling H., 2006, MNRAS, 366, 624
Beck R., Brandenburg A., Moss D., Shukurov A., Sokoloff D., 1996, ARA&A, 34, 155
Beck A. M., Lesch H., Dolag K., Kotarba H., Geng A., Stasyszyn F. A., 2012, MNRAS, 422, 2152
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Binney J., Cowie L. L., 1981, ApJ, 247, 464
Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Brüggen M., Scannapieco E., 2016, ApJ, 822, 31
Ceverino D., Klypin A., 2009, ApJ, 695, 292
Chan T.-K., Keres D., Hopkins P.-F., Quataert E., Su K.-Y., Hayward C.-C., Faucher-Giguere C.-A., 2018, preprint (arXiv:1812.10496)
Choi E., Ostriker J. P., Naab T., Oser L., Moster B. P., 2015, MNRAS, 449, 4105
Cioffi D. F., McKee C. F., Bertschinger E., 1988, ApJ, 334, 252
Ciotti L., Ostriker J. P., 2001, ApJ, 551, 131
Ciotti L., D'Ercole A., Pellegrini S., Renzini A., 1991, ApJ, 376, 380
Conroy C., Ostriker J. P., 2008, ApJ, 681, 151
Conroy C., van Dokkum P. G., Kravtsov A., 2015, ApJ, 803, 77
Cowie L. L., Binney J., 1977, ApJ, 215, 723
Croton D. J. et al., 2006, MNRAS, 365, 11
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
Dolag K., Jubelgas M., Springel V., Borgani S., Rasia E., 2004, ApJ, 606, L97
Eisenreich M., Naab T., Choi E., Ostriker J. P., Emsellem E., 2017, MNRAS, 468, 751
Enßlin T., Pfrommer C., Miniati F., Subramanian K., 2011, A&A, 527, A99
Fabian A. C., 1994, ARA&A, 32, 277
Fabian A. C., 1999, MNRAS, 308, L39
Fabian A. C., Arnaud K. A., Bautz M. W., Tawara Y., 1994, ApJ, 436, L63
Fabian A. C., Voigt L. M., Morris R. G., 2002, MNRAS, 335, L71
Faucher-Giguère C.-A., Quataert E., 2012, MNRAS, 425, 605
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Faucher-Giguère C.-A., Quataert E., Hopkins P. F., 2013, MNRAS, 433, 1970
Fraternali F., Marasco A., Marinacci F., Binney J., 2013, ApJ, 764, L21
Fujita Y., Ohira Y., 2011, ApJ, 738, 182
Fujita Y., Kimura S., Ohira Y., 2013, MNRAS, 432, 1434
Gaspari M., Sądowski A., 2017, ApJ, 837, 149
Governato F., Willman B., Mayer L., Brooks A., Stinson G., Valenzuela O., Wadsley J., Quinn T., 2007, MNRAS, 374, 1479
Greco J. P., Hill J. C., Spergel D. N., Battaglia N., 2015, ApJ, 808, 151
Guo F., Oh S. P., 2008, MNRAS, 384, 251
Gupta A., Mathur S., Krongold Y., 2017, ApJ, 836, 243
Hayward C. C., Hopkins P. F., 2017, MNRAS, 465, 1682

Hernquist L., 1990, ApJ, 356, 359
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., 2016, MNRAS, 462, 576
Hopkins P. F., 2017, MNRAS, 466, 3387
Hopkins P. F., Raives M. J., 2016, MNRAS, 455, 51
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Martini P., Robertson B., Springel V., 2005, ApJ, 630, 705
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1
Hopkins P. F., Quataert E., Murray N., 2011, MNRAS, 417, 950
Hopkins P. F., Quataert E., Murray N., 2012a, MNRAS, 421, 3488
Hopkins P. F., Quataert E., Murray N., 2012b, MNRAS, 421, 3522
Hopkins P. F., Narayanan D., Murray N., 2013, MNRAS, 432, 2647
Hopkins P. F. et al., 2018a, MNRAS, 477, 1578
Hopkins P. F. et al., 2018b, MNRAS, 480, 800
Hopkins P. F., et al., 2019, preprint (arXiv:1905.04321)
Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, MNRAS, 458, 3528
Humphrey P. J., Buote D. A., 2013, MNRAS, 436, 2879
Humphrey P. J., Buote D. A., Brighenti F., Flohic H. M. L. G., Gastaldello F., Mathews W. G., 2012a, ApJ, 748, 11
Humphrey P. J., Buote D. A., O'Sullivan E., Ponman T. J., 2012b, ApJ, 755, 166
Jacob S., Pfrommer C., 2017a, MNRAS, 467, 1449
Jacob S., Pfrommer C., 2017b, MNRAS, 467, 1478
Jacob S., Pakmor R., Simpson C. M., Springel V., Pfrommer C., 2018, MNRAS, 475, 570
Kaastra J. S., Mewe R., Nieuwenhuijzen H., 1996, in Yamashita K., Watanabe T., eds, UV and X-ray Spectroscopy of Astrophysical and Laboratory Plasmas. p. 411
Kannan R., Springel V., Pakmor R., Marinacci F., Vogelsberger M., 2016, MNRAS, 458, 410
Kannan R., Vogelsberger M., Pfrommer C., Weinberger R., Springel V., Hernquist L., Puchwein E., Pakmor R., 2017, ApJ, 837, L18
Kauffmann G. et al., 2003, MNRAS, 341, 54
Kennicutt Jr. R. C., 1998, ApJ, 498, 541
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kereš D., Katz N., Davé R., Fardal M., Weinberg D. H., 2009, MNRAS, 396, 2332
Khalatyan A., Cattaneo A., Schramm M., Gottlöber S., Steinmetz M., Wisotzki L., 2008, MNRAS, 387, 13
Kim D.-W., Fabbiano G., 2013, ApJ, 776, 116
Kim W.-T., Narayan R., 2003, ApJ, 596, 889
Kim C.-G., Ostriker E. C., Kim W.-T., 2014, ApJ, 786, 64
Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, Astron. Lett., 44, 8
Kroupa P., 2002, Science, 295, 82
Leitherer C. et al., 1999, ApJS, 123, 3
Li Y., Bryan G. L., Ruszkowski M., Voit G. M., O'Shea B. W., Donahue M., 2015, ApJ, 811, 73
Li Y., Ruszkowski M., Bryan G. L., 2017, ApJ, 847, 106
Li Y.-P. et al., 2018, ApJ, 866, 21
Lim S., Mo H., Wang H., Yang X., 2017, preprint (arXiv:1712.08619)
Lodders K., 2003, ApJ, 591, 1220
Madgwick D. S., Somerville R., Lahav O., Ellis R., 2003, MNRAS, 343, 871
Mannucci F., Della Valle M., Panagia N., 2006, MNRAS, 370, 773
Martig M., Bournaud F., Teyssier R., Dekel A., 2009, ApJ, 707, 250
Martizzi D., Quataert E., Faucher-Giguere C.-A., Fielding D., 2018, MNRAS, 483, 2465
Mathews W. G., Bregman J. N., 1978, ApJ, 224, 308
McDonald M., Veilleux S., Mushotzky R., 2011, ApJ, 731, 33
McDonald M., Gaspari M., McNamara B. R., Tremblay G. R., 2018, ApJ, 858, 45
McKee C. F., Cowie L. L., 1977, ApJ, 215, 213
McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117
Miller M. J., Bregman J. N., 2013, ApJ, 770, 118
Miller M. J., Bregman J. N., 2015, ApJ, 800, 14
Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
O'Dea C. P. et al., 2008, ApJ, 681, 1035
Orr M. E. et al., 2018, MNRAS, 478, 3653
Parrish I. J., Quataert E., 2008, ApJ, 677, L9
Parrish I. J., Stone J. M., Lemaster N., 2008, ApJ, 688, 905
Parrish I. J., Quataert E., Sharma P., 2009, ApJ, 703, 96
Parrish I. J., McCourt M., Quataert E., Sharma P., 2012, MNRAS, 419, L29
Pellegrini S., Ciotti L., Negri A., Ostriker J. P., 2018, ApJ, 856, 115
Peterson J. R., Fabian A. C., 2006, Phys. Rep., 427, 1
Pfrommer C., 2013, ApJ, 779, 10
Pfrommer C., Pakmor R., Schaal K., Simpson C. M., Springel V., 2017a, MNRAS, 465, 4500
Pfrommer C., Pakmor R., Simpson C. M., Springel V., 2017b, ApJ, 847, L13
Planck Collaboration XI, 2013, A&A, 557, A52
Pope E. C. D., Pavlovski G., Kaiser C. R., Fangohr H., 2006, MNRAS, 367, 1121
Popping G., Behroozi P. S., Peeples M. S., 2015, MNRAS, 449, 477
Pozzetti L. et al., 2010, A&A, 523, A13
Quataert E., 2008, ApJ, 673, 758
Rafferty D. A., McNamara B. R., Nulsen P. E. J., 2008, ApJ, 687, 899
Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716
Ressler S. M., Quataert E., Stone J. M., 2018, MNRAS, 478, 3544
Romeo A. B., Wiegert J., 2011, MNRAS, 416, 1191
Ruszkowski M., Yang H.-Y. K., Zweibel E., 2017a, ApJ, 834, 208
Ruszkowski M., Yang H.-Y. K., Reynolds C. S., 2017b, ApJ, 844, 13
Salomé P. et al., 2006, A&A, 454, 437
Sarazin C. L., 1988, X-ray emission from clusters of galaxies, Cambridge Univ. Press, Cambridge
Schmidt M., 1959, ApJ, 129, 243
Schure K. M., Kosenko D., Kaastra J. S., Keppens R., Vink J., 2009, A&A, 508, 751
Sharma P., Parrish I. J., Quataert E., 2010, ApJ, 720, 652
Sharma P., McCourt M., Quataert E., Parrish I. J., 2012, MNRAS, 420, 3174
Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Silk J., 1976, ApJ, 208, 646
Silk J., 2009, in Andersen J., Nordströara m B., Bland-Hawthorn J., eds, IAU Symposium Vol. 254, The Galaxy Disk in Cosmological Context. p. 401
Silk J., 2010, in Block D. L., Freeman K. C., Puerari I., eds, Galaxies and their Masks. Springer, New York, NY, p. 399
Silk J., 2011, in Carignan C., Combes F., Freeman K. C., eds, Proc. IAU Symp. Vol. 277, Tracing the Ancestry of Galaxies, Ouagadougou, Burkina Faso, p. 273
Silk J., Rees M. J., 1998, A&A, 331, L1
Soker N., Sarazin C. L., 1990, ApJ, 348, 73
Spitzer L., Härm R., 1953, Phys. Rev., 89, 977
Springel V., 2000, MNRAS, 312, 859
Springel V., White S. D. M., 1999, MNRAS, 307, 162
Stanek R., Evrard A. E., Böhringer H., Schuecker P., Nord B., 2006, ApJ, 648, 956
Stone J. M., Gardiner T., 2007, ApJ, 671, 1726
Strong A. W., Moskalenko I. V., 1998, ApJ, 509, 212
Strong A. W., Porter T. A., Digel S. W., Jóhannesson G., Martin P., Moskalenko I. V., Murphy E. J., Orlando E., 2010, ApJ, 722, L58
Su Y., White R. E., III, Miller E. D., 2013, ApJ, 775, 89
Su Y., Buote D., Gastaldello F., Brighenti F., 2015, ApJ, 805, 104
Su K.-Y. et al., 2018a, preprint (arXiv:1812.03997)
Su K.-Y. et al., 2018c, MNRAS, 480, 1666
Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, MNRAS, 471, 144

Su K.-Y., Hayward C. C., Hopkins P. F., Quataert E., Faucher-Giguère C.-A., Kereš D., 2018b, MNRAS, 473, L111
Tamura T. et al., 2001, A&A, 365, L87
Tang S., Wang Q. D., 2005, ApJ, 628, 205
Tang S., Wang Q. D., Mac Low M.-M., Joung M. R., 2009, MNRAS, 398, 1468
Taylor P., Kobayashi C., 2015, MNRAS, 448, 1835
Trotta R., Jóhannesson G., Moskalenko I. V., Porter T. A., Ruiz de Austri R., Strong A. W., 2011, ApJ, 729, 106
Tucker W. H., Rosner R., 1983, ApJ, 267, 547
Uhlig M., Pfrommer C., Sharma M., Nath B. B., Enßlin T. A., Springel V., 2012a, MNRAS, 423, 2374
Uhlig M., Pfrommer C., Sharma M., Nath B. B., Enßlin T. A., Springel V., 2012b, MNRAS, 423, 2374
van de Voort F., Quataert E., Hopkins P. F., Faucher-Giguère C.-A., Feldmann R., Kereš D., Chan T. K., Hafen Z., 2016, MNRAS, 463, 4533
Voigt L. M., Fabian A. C., 2004, MNRAS, 347, 1130
Voigt L. M., Schmidt R. W., Fabian A. C., Allen S. W., Johnstone R. M., 2002, MNRAS, 335, L7
Voit G. M., 2011, ApJ, 740, 28
Voit G. M., Donahue M., 2015, ApJ, 799, L1
Voit G. M., Meece G., Li Y., O'Shea B. W., Bryan G. L., Donahue M., 2017, ApJ, 845, 80
Wagh B., Sharma P., McCourt M., 2014, MNRAS, 439, 2822
Weinberger R. et al., 2018, MNRAS, 479, 4056
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2
Werner N. et al., 2013, ApJ, 767, 153
Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
Wiener J., Oh S. P., Guo F., 2013, MNRAS, 434, 2209
Woo J. et al., 2013, MNRAS, 428, 3306
Yoon D., Yuan F., Gan Z.-M., Ostriker J. P., Li Y.-P., Ciotti L., 2018, ApJ, 864, 6
Zakamska N. L., Narayan R., 2003, ApJ, 582, 162
ZuHone J. A., Kunz M. W., Markevitch M., Stone J. M., Biffi V., 2015, ApJ, 798, 90

## APPENDIX: RESOLUTION STUDY

Extensive resolution studies of the FIRE-2 feedback and physics models used in this paper are presented in Hopkins et al. (2018b). However, since these did not address all of the specific questions in this paper, we performed a series resolution studies using our 'Default' m14 halo, varying the mass resolution by a factor ∼300 (Table A1). The 'HR' ('MR') run in Table A1 matches the m14-HR (m14-MR) resolution in the text.

**Table A1.** Mass resolutions used in our studies for the default m14 Run.

| Resolution | $m_g$ | $m_h$ | $m_d$ | $m_b$ |
|---|---|---|---|---|
| HR-HRS | 8e3-2e6 | 4e7 | 8e3 | 8e3 |
| HR | 8e3-2e6 | 4e7 | 2.5e6 | 2.5e6 |
| MR-MRS | 3e4-2e6 | 4e7 | 3e4 | 3e4 |
| MR | 3e4-2e6 | 4e7 | 2.5e6 | 2.5e6 |
| LR | 2e6 | 4e7 | 1e7 | 1e7 |

*Note*. (1) Resolution name. LR: Low resolution. MR: Medium resolution. HR: High resolution. MRS(HRS): Medium (high) resolution initial stellar particles. (2) $m_g$: Gas particle mass. (3) $m_h$: Dark matter halo particle mass. (4) $m_d$: Pre-existing stellar disc particle mass. (5) $m_b$: Pre-existing bulge particle mass. Note: All runs use the m14 halo with 'Default' FIRE physics.

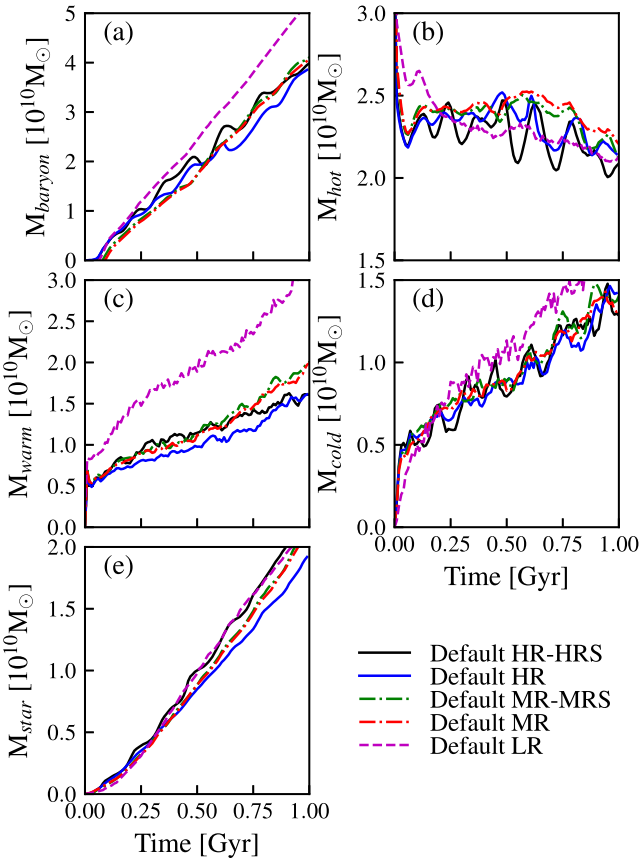


**Figure A1.** The comparison of (a) core (<30 kpc) baryonic mass, (b) hot gas (>$10^6$ K) mass, (c) warm gas (8000–$10^6$ K) mass, (d) cold gas (<8000 K) mass, and (e) stellar mass, for 'Default' m14 runs at different resolutions. 'MR' and 'HR' runs behave very similarly. 'LR' run, on the other hand has more gas accumulated in the warm phase.

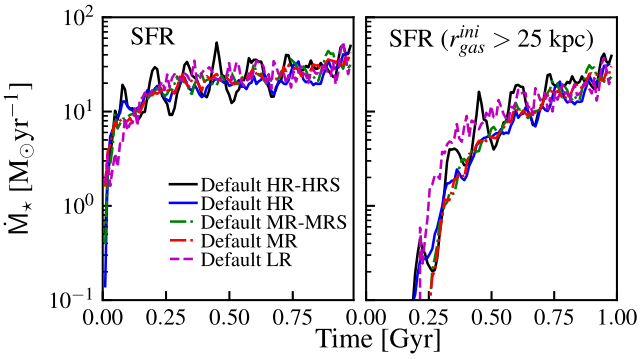


**Figure A2.** Comparison of total SFR, and SFR from gas initially outside 25 kpc, for 'Default' m14 runs with different resolutions. Runs at different resolutions have very similar SFRs.

Fig. A1 shows the evolution of the total core ($r$ < 30 kpc) baryonic, hot gas (>$10^6$ K), warm gas (8000–$10^6$ K), cold gas (<8000 K), and stellar masses at different resolutions. Fig. A2 shows the comparison of total SFR, and SFR from gas initially outside 25kpc. Fig. A3 shows the comparison of cooling time and cooling time over dynamical time for gas within 30 kpc.

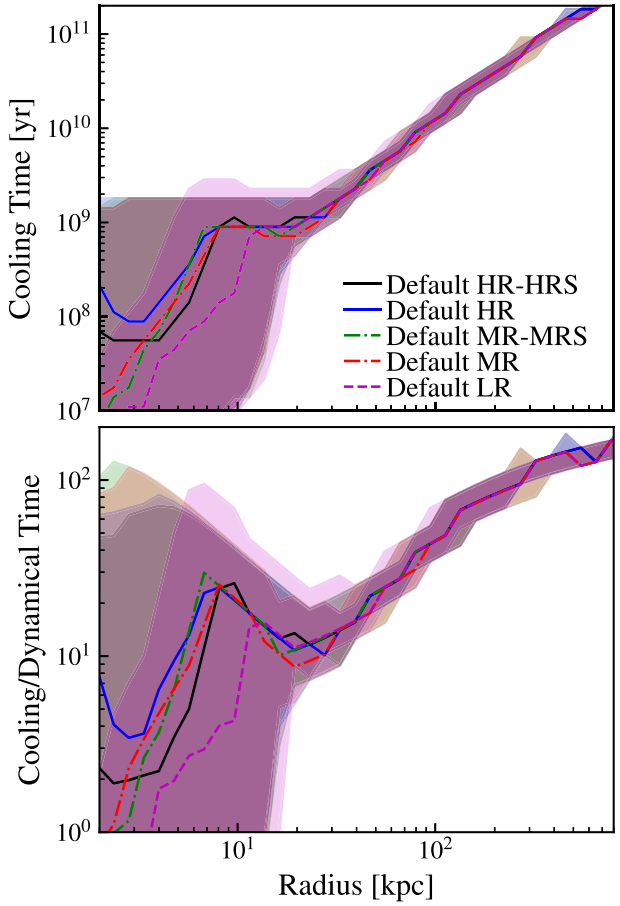


**Figure A3.** Cooling time, and cooling time over dynamical time, as a function of radius for gas hotter than $10^5$ K. 'MR' and 'HR' runs have very similar cooling properties. 'LR' run, has slightly shorter cooling time at small radius.

Fig. A4 shows the evolution of cooling and energy input rates within 30 kpc, and the total X-ray luminosity of the whole halo. In these calculated properties, runs with resolution equal to or higher than that of our 'MR' run do not differ significantly. The 'LR' run on the other hand, exhibits more gas build-up in the warm phase at small radius, which leads to the shorter cooling time there. This, and the more detailed resolution studies referenced above, suggest the results are here are robust to resolution, at least over the dynamical range, that we explore here.

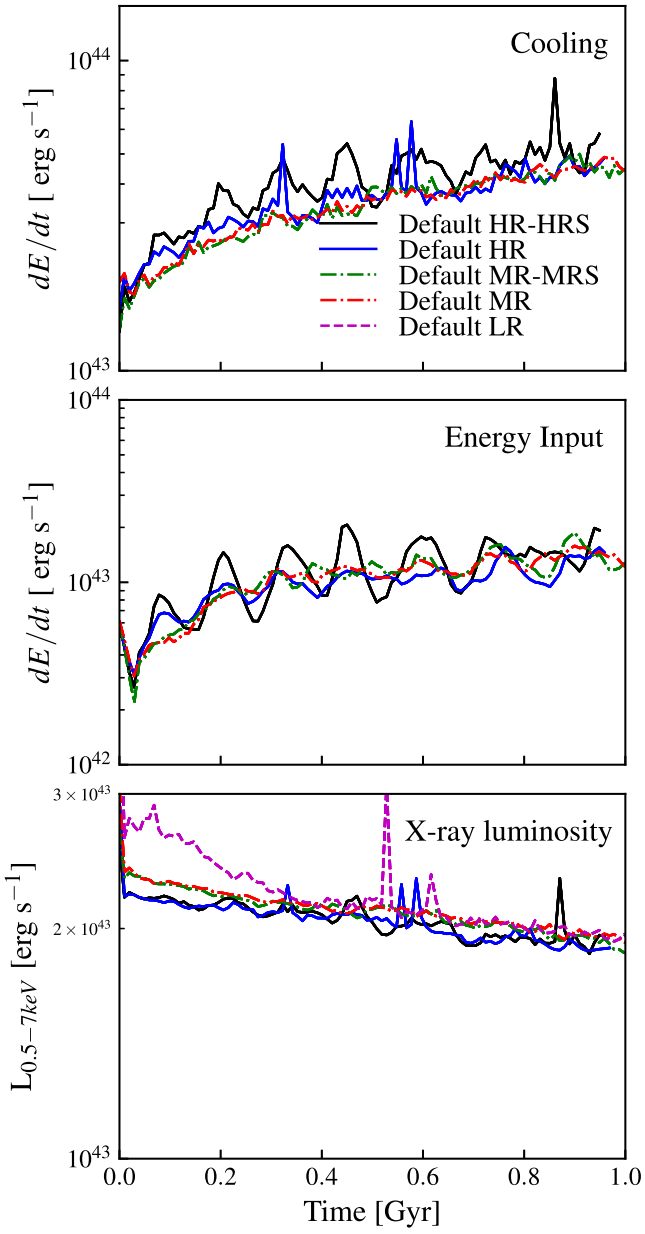


**Figure A4.** The upper row plots the cooling and energy input rates within 30 kpc, as a function of time. The bottom row shows the X-ray luminosity in the 0.5–7 kev band. Runs with different resolutions have very similar cooling and energy input rates.

This paper has been typeset from a TEX/LATEX file prepared by the author.