



# A Global Model for Circumgalactic and Cluster-core Precipitation

G. Mark Voit<sup>1</sup>, Greg Meece<sup>1</sup>, Yuan Li<sup>2</sup>, Brian W. O’Shea<sup>1,3</sup>, Greg L. Bryan<sup>4</sup>, and Megan Donahue<sup>1</sup>

<sup>1</sup>Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA; [voit@pa.msu.edu](mailto:voit@pa.msu.edu)

<sup>2</sup>Department of Astronomy, University of Michigan, Ann Arbor, MI, USA

<sup>3</sup>Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA

<sup>4</sup>Department of Astronomy, Columbia University, New York, NY, USA

Received 2016 July 7; revised 2017 June 1; accepted 2017 June 27; published 2017 August 14

## Abstract

We provide an analytic framework for interpreting observations of multiphase circumgalactic gas that is heavily informed by recent numerical simulations of thermal instability and precipitation in cool-core galaxy clusters. We start by considering the local conditions required for the formation of multiphase gas via two different modes: (1) uplift of ambient gas by galactic outflows, and (2) condensation in a stratified stationary medium in which thermal balance is explicitly maintained. Analytic exploration of these two modes provides insights into the relationships between the local ratio of the cooling and freefall timescales (i.e.,  $t_{\text{cool}}/t_{\text{ff}}$ ), the large-scale gradient of specific entropy, and the development of precipitation and multiphase media in circumgalactic gas. We then use these analytic findings to interpret recent simulations of circumgalactic gas in which global thermal balance is maintained. We show that long-lasting configurations of gas with  $5 \lesssim \min(t_{\text{cool}}/t_{\text{ff}}) \lesssim 20$  and radial entropy profiles similar to observations of cool cores in galaxy clusters are a natural outcome of precipitation-regulated feedback. We conclude with some observational predictions that follow from these models. This work focuses primarily on precipitation and AGN feedback in galaxy-cluster cores, because that is where the observations of multiphase gas around galaxies are most complete. However, many of the physical principles that govern condensation in those environments apply to circumgalactic gas around galaxies of all masses.

**Key words:** galaxies: active – galaxies: clusters: intracluster medium – galaxies: elliptical and lenticular, cD – galaxies: evolution – galaxies: halos

## 1. Introduction

The relationship between thermal instability and galaxy formation is a classic topic in theoretical astrophysics that has recently come back into fashion. Its reemergence has been driven by the need to understand how accretion onto supermassive black holes regulates cooling and star formation in galaxy-cluster cores. Both observational and theoretical evidence are accumulating in support of the idea that development of a multiphase medium through condensation in the vicinity of a supermassive black hole triggers strong feedback that limits further condensation (e.g., Pizzolato & Soker 2005, 2010; Soker 2006; Cavagnolo et al. 2008; Gaspari et al. 2012, 2013, 2015; McCourt et al. 2012; Sharma et al. 2012b; Li et al. 2015; Voit et al. 2015b; Tremblay et al. 2016). Perhaps most intriguingly, if there is a similar link between feedback heating and condensation of circumgalactic gas around smaller galaxies, then this regulation mechanism has much broader implications for galaxy evolution (e.g., Soker 2010a; Sharma et al. 2012a; Voit et al. 2015a).

### 1.1. Heritage of the Topic

All modern discussions of thermal instability in the context of galaxy formation are rooted in the classic works of Rees & Ostriker (1977), Binney (1977), and Silk (1977), which themselves owe a debt to Hoyle (1953). These landmark papers derived the maximum stellar mass of a galaxy ( $\sim 10^{12} M_{\odot}$ ) by comparing the time for gas to fall through a galaxy’s potential well to the time required to cool from the potential’s virial temperature. If the cooling time is less than the freefall time, then infalling gas can potentially condense and fragment into star-forming clouds via thermal instability. That

can happen relatively easily in galaxy-scale objects with virial temperatures  $\lesssim 10^7$  K but is more difficult to arrange in hotter, more massive systems, leading to a natural division between the mass scales of individual galaxies and those of galaxy groups and clusters.

Many subsequent papers have made interesting use of the cooling time to freefall time ratio to analyze how galaxies form and evolve (e.g., Blumenthal et al. 1984; Fall & Rees 1985; Maller & Bullock 2004). However, some form of negative feedback is necessary to explain the inefficient transformation of a galaxy’s gas supply into stars (e.g., Larson 1974; White & Rees 1978; Dekel & Silk 1986; White & Frenk 1991; Baugh et al. 1998; Somerville & Primack 1999; Kauffmann & Haehnelt 2000). A complete understanding of the relationship between thermal instability and galaxy formation must therefore account for the interplay between radiative cooling and the energetic feedback that opposes cooling.

Field (1965) worked out many of the fundamental features of astrophysical thermal instability but did not consider the complications that arise when thermal instability couples with buoyancy. The most comprehensive analytical treatment of that coupling is by Balbus & Soker (1989), who astutely summarized much of the preceding work. Balbus & Soker were primarily concerned with the development of inhomogeneity in galaxy-cluster cores. At the time, hot gas in the cores of many galaxy clusters was suspected to condense at rates  $\sim 10^{-3} M_{\odot} \text{ yr}^{-1}$  (e.g., Fabian 1994), but models of homogeneous cooling flows into the cluster’s central galaxy produced X-ray surface-brightness profiles with central peaks far greater than what were observed. This conundrum led to speculation that the mass inflow rate could decline inward because of thermal instability and spatially distributed condensation, which would reduce the

radiative losses required to maintain a steady state at the center of the flow (e.g., Thomas et al. 1987; White & Sarazin 1987a, 1987b, 1987c). Upon closer examination, this speculation was found to be problematic, because buoyancy generally tends to suppress the development of thermal instability (e.g., Cowie et al. 1980; Nulsen 1986; Balbus & Soker 1989). The attention of the field therefore gradually shifted away from steady-state cooling-flow models in favor of models in which feedback from a central active galactic nucleus compensates for cooling (e.g., Tabor & Binney 1993; Binney & Tabor 1995; Soker et al. 2001; McNamara & Nulsen 2007, 2012).

### 1.2. Renaissance of the Topic

More than two decades later, the coupling between buoyancy and thermal instability is being reexamined, because the presence of inhomogeneous gas in galaxy-cluster cores consisting of multiple phases that are orders of magnitude cooler and denser than the ambient medium now appears closely linked with the ratio of cooling time to freefall time (e.g., Gaspari et al. 2012; McCourt et al. 2012; Voit & Donahue 2015; Voit et al. 2015b). In those studies, the cooling time is typically defined with respect to the specific heat at constant volume, so that  $t_{\text{cool}} \equiv [3kT/n_e\Lambda(T)](n/2n_i)$ , where  $\Lambda(T)$  is the usual cooling function at temperature  $T$ , and the number densities of electrons, ions, and gas particles are  $n_e$ ,  $n_i$ , and  $n$ , respectively. The freefall time  $t_{\text{ff}} \equiv (2r/g)^{1/2}$  is defined with respect to the local gravitational potential  $g$  at radius  $r$ . Given these definitions, a floor appears to be present near  $t_{\text{cool}}/t_{\text{ff}} \approx 10$  in the radial cooling-time profiles of the ambient hot gas in galaxy clusters. A large majority of the cluster cores known to contain multiphase gas have minimum values of  $t_{\text{cool}}/t_{\text{ff}}$  within a factor of 2 of this floor. Conversely, almost all of the clusters without multiphase gas have  $\min(t_{\text{cool}}/t_{\text{ff}}) > 20$ .

McCourt et al. (2012) and Sharma et al. (2012b) interpreted this relationship between multiphase gas and  $t_{\text{cool}}/t_{\text{ff}}$  as resulting from the amplification of initially small perturbations by thermal instability. Under conditions of global thermal balance, numerical simulations of thermal instability in a plane-parallel potential by McCourt et al. (2012) showed that  $t_{\text{cool}}/t_{\text{ff}} \lesssim 1$  was required for thermal instability resulting in condensation, but the simulations of Sharma et al. (2012b) in a spherically symmetric potential indicated that condensation could happen in spherical systems with  $t_{\text{cool}}/t_{\text{ff}} \lesssim 10$ . The threshold value of  $t_{\text{cool}}/t_{\text{ff}}$  was therefore assumed to be geometry dependent. Subsequent simulations implementing more sophisticated treatments of feedback appeared to corroborate that assumption because they showed that condensation-fueled accretion of cold gas onto a central black hole can lead to long-lasting self-regulation with  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  (e.g., Gaspari et al. 2012; Li & Bryan 2014a; Li et al. 2015; Prasad et al. 2015).

In the meantime, it has become clear that the critical ratio  $t_{\text{cool}}/t_{\text{ff}}$  is not a geometry-dependent manifestation of local thermal instability. For example, numerical simulations by Meece et al. (2015) of thermal instability in a plane-parallel potential under conditions seemingly quite similar to those adopted by McCourt et al. (2012) showed that condensation could occur at the midplane of systems with any value of  $t_{\text{cool}}/t_{\text{ff}}$ . Also, Choudhury & Sharma (2016) presented a detailed thermal stability analysis of systems in global thermal balance showing that the growth rates of linear perturbations are largely independent of the gravitational potential's geometry. So why,

then, do both real and simulated galaxy-cluster cores appear to self-regulate at  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ ?

### 1.3. Precipitation-regulated Feedback

The most general answer seems to involve a phenomenon that we have come to call precipitation. As feedback acts on a galaxy-scale system, the outflows it drives can promote condensation of the hot ambient medium by raising some of it to greater altitudes (e.g., Revaz et al. 2008; Li & Bryan 2014b; McNamara et al. 2016). Adiabatic uplift promotes condensation by lowering the  $t_{\text{cool}}/t_{\text{ff}}$  ratio of the uplifted gas (see Section 2). This process is loosely analogous to the production of raindrops during adiabatic cooling of uplifted humid gas in a thunderstorm—hence, the name “precipitation.” As in a thunderstorm, the condensates rain down toward the bottom of the potential well after they form. In simulations, this rain of cold gas into the galaxy at first provides additional fuel for feedback and temporarily boosts the strength of the outflows, but eventually those strengthening outflows add enough heat to the ambient medium to raise  $t_{\text{cool}}/t_{\text{ff}}$  high enough to stop the condensation. Precipitation is therefore naturally self-regulating.

A prescient series of “cold feedback” papers by Soker & Pizzolato anticipated many features of the precipitation mechanism now seen in simulations (Pizzolato & Soker 2005, 2010; Soker 2006, 2008). They proposed a feedback cycle in cluster cores in which energetic AGN outbursts produce a wealth of nonlinear density perturbations, the densest of which cool, condense, fall back toward the black hole, and provide more fuel for accretion. In this scenario, the cooling times of the blobs must be short enough that  $t_{\text{cool}} \lesssim t_{\text{ff}}$  and also shorter than the time interval between AGN heating outbursts. They argued that this source of accretion fuel could potentially provide much more fuel than Bondi accretion from the hot medium alone while responding to changes in the ambient cooling time far more quickly. They also noted that a shallow central entropy gradient would promote condensation.

Shortly thereafter, Gaspari and collaborators produced sets of numerical simulations in which such a cold feedback loop was realized (Gaspari et al. 2012, 2013, 2015; but see also Sharma et al. 2012b). In these simulations, cold clouds form through thermal instability and accrete toward the center of the simulation volume while the heating mechanism needed to maintain approximate global thermal balance stimulates turbulence. This turbulence is critical, because it ensures a steady supply of cold gas blobs with low specific angular momentum, which can plunge to the center through a process the authors call “chaotic cold accretion.”

The proliferation of terminology has a way of making these mechanisms seem more different than they really are. Precipitation-regulated feedback, as described in this paper, is a “cold feedback” mechanism that fuels a central black hole through “chaotic cold accretion.” Here we are proceeding with the term “precipitation,” despite the prior existence of these other terms, because the processes that promote thermal instability and condensation in circumgalactic gas may be responsible for more than just the feeding of black holes. Therefore, they merit a more general term.

#### 1.4. Implications for Galaxy Evolution

Precipitation is potentially of broader interest because it may link a galaxy's time-averaged star formation rate with the multiphase structure of its circumgalactic gas. Observations of circumgalactic absorption lines are showing that the masses of gas and metals within a few hundred kiloparsecs of a galaxy are at least as great as the galaxy's mass in the form of stars (e.g., Tumlinson et al. 2011). Also, the amount of circumgalactic gas at intermediate temperatures ( $10^5$ – $10^6$  K) appears closely linked with a galaxy's star formation rate. Observations of the circumgalactic medium at other wavelengths likewise show rich multiphase structure (e.g., Putman et al. 2012), which has been challenging for simulations to reproduce (e.g., Hummels et al. 2013; Ford et al. 2016).

Sharma et al. (2012a) proposed that thermal instability, through precipitation-regulated feedback, places a lower limit of  $t_{\text{cool}}/t_{\text{ff}} \approx 10$  on the ambient density of circumgalactic gas (but see also Meece et al. 2015). Voit et al. (2015a) built on that idea to show how precipitation-regulated feedback could be responsible for governing not only the relationships between stellar mass, metallicity, and the stellar baryon fraction observed among galaxies but also the relationship between a galaxy's stellar velocity dispersion and the mass of its central black hole. However, they did so without having a satisfactory explanation for the crucial assumption that  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  or a complete model for the global structure of precipitation-regulated systems.

#### 1.5. Purpose of the Paper

This paper's purpose is to propose a global context for interpreting observations of multiphase gas around galaxies, as well as the numerical simulations that strive to reproduce those observations, in terms of the long legacy of theoretical papers on astrophysical thermal instability. Many of the theoretical results derived here were published decades ago by others, most notably by Defouw (1970), Cowie et al. (1980), Nulsen (1986), Malagoli et al. (1987), Loewenstein (1989); and Balbus & Soker (1989). Our rederivations of them are intended to provide a common conceptual framework for a meta-analysis of the simulations in which precipitation occurs. We are focusing on simulations of precipitation in galaxy-cluster cores, because that is where the observations of multiphase gas around galaxies and interactions of outflows with the circumgalactic medium are most complete. However, many of the physical principles that govern condensation in those environments apply to circumgalactic gas around galaxies of all masses.

#### 1.6. A Readers' Guide

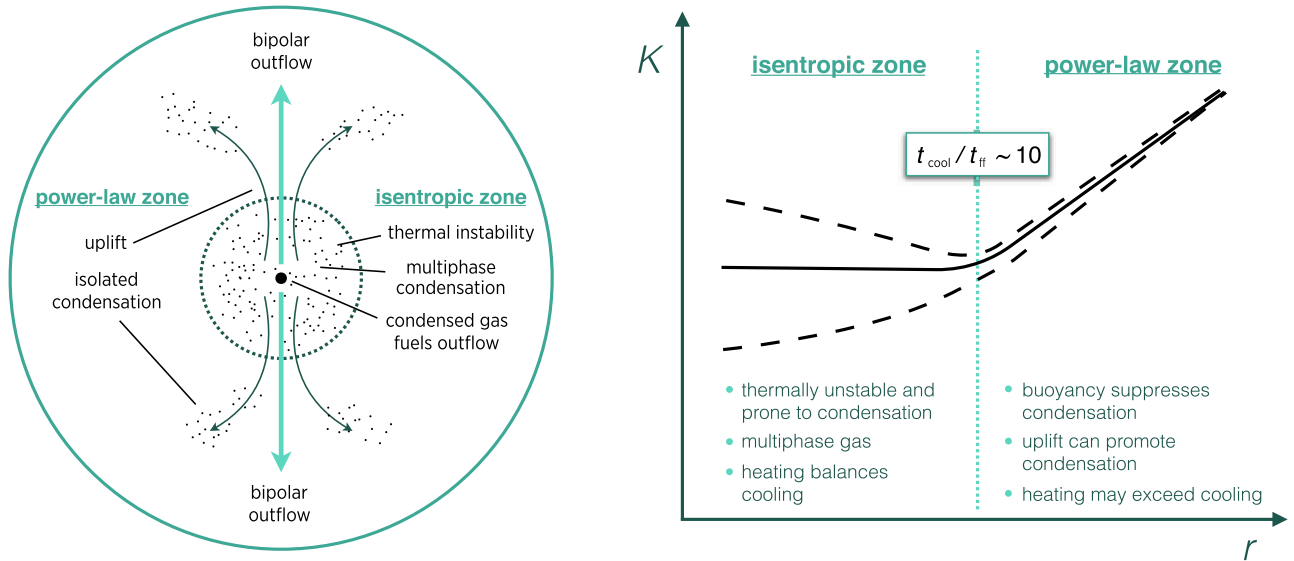
Busy readers may wish to be selective in deciding which sections of this long paper will reward their close attention. For them, we have prepared this guide, along with a cartoon (Figure 1) that sketches out the main ideas.

1. The next three sections (Sections 2–4) consider the local conditions required for condensation and precipitation. Two different modes of precipitation emerge from those considerations. One is analogous to rain that is stimulated by the uplift of humid gas in Earth's atmosphere, because of the role that adiabatic cooling plays in bringing on

condensation. The other is analogous to drizzle or fog, in that condensation happens without uplift, when the conditions are right.

- (a) Section 2 initiates the discussion with a brief reminder about the deep connections between adiabatic uplift and condensation. It also presents a short calculation showing that ambient gas uplifted at speeds comparable to a halo's circular velocity is likely to condense if it initially has  $t_{\text{cool}}/t_{\text{ff}} \lesssim 10$ . This finding suggests that the ambient medium around a galaxy cannot persist in a state with  $t_{\text{cool}}/t_{\text{ff}} \ll 10$  if there is significant vertical circulation. Galactic outflows in which the energy source is fueled by condensation therefore tend to drive the ambient medium toward  $t_{\text{cool}}/t_{\text{ff}} \approx 10$ .
  - (b) Section 3 outlines the general conditions necessary for thermal instability without uplift to progress to condensation in a hydrostatic medium that is thermally balanced within each equipotential layer. The main result is that the condition for condensation to occur depends not only on  $t_{\text{cool}}/t_{\text{ff}}$  but also on the slope of the entropy gradient: thermal instability leads to condensation only if a low-entropy perturbation can cool faster than it sinks to a layer of equivalent entropy. This is not a new result, but its significance is often not fully appreciated.
  - (c) Section 4 uses the results of Section 3 to interpret recent simulations of thermal instability in circumgalactic gas. In particular, it calls attention to the critical role of the global entropy gradient in determining where condensation can occur and where it is suppressed by buoyancy. The main result is that media in thermal balance are prone to condensation in regions where the large-scale entropy gradient is flat. However, buoyancy tends to delay the onset of condensation if  $t_{\text{cool}}/t_{\text{ff}} \gg 1$ .
2. The following two sections (Sections 5 and 6) apply the findings from the first part of the paper to interpret the global evolution of simulated galactic systems in which condensation fuels feedback. Our objective is to understand why those systems end up in long-lasting configurations with  $5 \lesssim \min(t_{\text{cool}}/t_{\text{ff}}) \lesssim 20$  and with radial entropy profiles in agreement with observations of multiphase galaxy-cluster cores.
    - (a) Section 5 examines recent simulations of condensation in globally balanced but locally unstable galactic systems in light of the findings summarized in Section 4. It points out that feedback is *required* for the development of a multiphase medium, because phase separation cannot happen in a globally balanced medium without the flow of free energy through the system. Condensation also cannot happen through linear thermal instability in regions with a significant entropy gradient and  $t_{\text{cool}} \gg t_{\text{ff}}$ . Steady self-regulation of a precipitating system therefore favors a global configuration with a shallow inner entropy gradient and a steeper outer entropy gradient (Figure 1). The shallow inner gradient promotes the precipitation needed for fuel, while the steeper outer entropy gradient prevents condensation from running away into a cooling catastrophe. The boundary between these regions tends to be where  $t_{\text{cool}}/t_{\text{ff}}$  reaches a





**Figure 1.** This schematic cartoon outlines the main ideas presented in the paper. On the left is a diagram of a galactic environment in which feedback is active. Accretion of condensed gas onto the central black hole releases feedback energy, and a bipolar outflow distributes that energy over a large volume. In order for the system to develop a well-regulated feedback loop, it must separate into two zones, an inner “isentropic zone” and an outer “power-law zone” in which the specific entropy ( $K \equiv kTn_e^{-2/3}$ ) follows  $d \ln K / d \ln r \approx 1$ , as observed in both real and simulated galaxy-cluster cores (Sections 5 and 6). The power-law entropy gradient allows buoyancy to limit the growth of thermal instability (Sections 3 and 4), implying that condensation in the power-law zone requires uplift of lower-entropy gas (Section 2). In contrast, buoyancy cannot suppress thermal instability in the isentropic zone, which proceeds to develop a multiphase structure, as indicated by the dashed lines showing the dispersion in  $K$  at each  $r$  (Sections 4 and 6). Phenomenologically, the ratio of cooling time to freefall time in the ambient medium is observed to reach a minimum value in the range  $5 \lesssim t_{\text{cool}}/t_{\text{ff}} \lesssim 20$  at the boundary between these zones in cluster cores with multiphase gas (Section 6). Such a system cannot remain in a steady feedback-regulated state with a central cooling time  $\lesssim 1$  Gyr unless a large proportion of the feedback energy is thermalized *outside* of the central isentropic region (Section 5).

minimum value  $\sim 10$ , for reasons outlined in Section 2. In order to ensure long-term global stability with a central cooling time  $\lesssim 1$  Gyr, much of the feedback energy must propagate beyond the isentropic zone before thermalizing. This finding has deep implications for implementations of black-hole feedback in numerical simulations.

- (b) Section 6 puts all of these pieces together to interpret the time-dependent behavior of precipitation-regulated feedback in a numerical simulation from Li et al. (2015). It shows that unopposed cooling leads to a power-law entropy gradient at the center of the system, which focuses condensation onto the central black hole. The feedback response disrupts that central gradient out to where  $t_{\text{cool}}/t_{\text{ff}} \approx 10$  in the ambient medium. Much of the gas uplifted from that central region can be induced to condense, particularly if it is inhomogeneous. The system then settles into a long-lasting steady state in which condensed gas fuels the outflow and feedback maintains the isentropic central region at a level corresponding to  $5 \lesssim \min(t_{\text{cool}}/t_{\text{ff}}) \lesssim 20$ . Catastrophic cooling is prevented because much of the outflow’s energy is thermalized outside of the isentropic zone. When the condensed gas is depleted, the outflow shuts down. Cooling then proceeds almost homogeneously, because there is no source of free energy to promote phase separation. However, thermal instability eventually initiates condensation near the outer edge of the isentropic region, at the minimum of  $t_{\text{cool}}/t_{\text{ff}}$  in the ambient medium. Newly condensed gas subsequently reignites feedback, and the cycle repeats.

The paper concludes with two sections that acknowledge the loose ends (Section 7) and present some concluding thoughts about how the overall model can be tested (Section 8), followed by an Appendix that constructs a useful toy model for global configuration changes of precipitation-regulated systems.

## 2. Uplift and Condensation

In the precipitation framework for self-regulating feedback, adiabatic uplift is one of the main mechanisms for causing condensation of circumgalactic gas. Transporting gas upward without changing its entropy increases its propensity to condense by lowering its  $t_{\text{cool}}/t_{\text{ff}}$  ratio. Most of the change in this critical ratio comes from the increase in  $t_{\text{ff}}$  owing to uplift, because the change in cooling time produced by adiabatic uplift is relatively modest.

### 2.1. Adiabatic Uplift

To illustrate this point, consider a blob of gas that begins at radius  $r_1$  with entropy  $K_1$  in an isothermal potential with constant circular velocity  $v_c$ . The freefall time in this potential is linearly proportional to radius, and the cooling time of the blob depends on its pressure. When written in terms of the entropy  $K$  and gas pressure  $P$ , this cooling time is

$$t_{\text{cool}} = \frac{3K^{6/5}}{P^{1/5} \Lambda[T(P, K)]} \left( \frac{n}{n_e} \right)^{1/5} \left( \frac{n}{2n_i} \right) \quad (1)$$

$$\approx 235 \text{ Myr} \left( \frac{K}{10 \text{ keV cm}^2} \right)^{6/5} \left( \frac{P}{k \cdot 10^6 \text{ K cm}^{-3}} \right)^{-1/5} \times \left( \frac{\Lambda}{2 \times 10^{-23} \text{ erg cm}^3 \text{ s}^{-1}} \right)^{-1}. \quad (2)$$

For gas with the temperatures and abundances typical of galaxy-cluster cores and large elliptical galaxies ( $10^7 \text{ K} \lesssim T \lesssim 3 \times 10^7 \text{ K}$ ), the cooling function  $\Lambda(T)$  is nearly independent of temperature, meaning that  $t_{\text{cool}}$  changes very little as gas adiabatically expands or contracts. One can account for the temperature dependence of  $\Lambda$  by defining  $\lambda \equiv d \ln \Lambda / d \ln T$  and noting that  $T \propto K^{3/5} P^{2/5}$ , to obtain  $t_{\text{cool}} \propto K^{(6-3\lambda)/5} P^{-(1+2\lambda)/5}$ . Adiabatic uplift of a gas blob in a stratified atmosphere therefore changes the ratio of cooling time to freefall time within the blob according to

$$\left( \frac{t_{\text{cool}}}{t_{\text{ff}}} \right)_r = \left( \frac{t_{\text{cool}}}{t_{\text{ff}}} \right)_{r_1} \left( \frac{r_1}{r} \right)^{1-2\beta(1+2\lambda)/5}, \quad (3)$$

where  $d \ln P / d \ln r = -2\beta$  with  $\beta \equiv \mu m_p v_c^2 / 2kT$  in a hydrostatic ambient medium. Uplift by an order of magnitude in radius consequently lowers the  $t_{\text{cool}}/t_{\text{ff}}$  ratio local to the blob by roughly an order of magnitude, given the typical  $\beta$  and  $\lambda$  values of circumgalactic gas ( $0.5 \lesssim \beta \lesssim 1$  and  $-1 \lesssim \lambda \lesssim 0.5$ ), as long as magnetic fields sufficiently suppress thermal conduction.

This relationship helps explain why precipitation-driven feedback tends to push a galaxy's volume-filling ambient medium into a state with  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ . If the central gas is so dense that  $t_{\text{cool}}/t_{\text{ff}} \ll 10$ , then outflows propelled by feedback can promote additional condensation of the ambient medium with relatively modest amounts of adiabatic uplift. Newly condensed clouds that have not achieved escape velocity can then rain back down into the galaxy to fuel more star formation and AGN feedback, and the rain will continue until the dissipation of feedback energy raises the ambient value of  $\min(t_{\text{cool}}/t_{\text{ff}})$  enough to inhibit condensation of uplifted gas. Then, feedback finally succeeds in stopping the accumulation of cold clouds within the galaxy. The critical value of  $\min(t_{\text{cool}}/t_{\text{ff}})$  at which cessation of condensation should happen is hard to calculate from first principles, but simulations of the process indicate that it happens around  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  (e.g., Gaspari et al. 2012; Li et al. 2015; Prasad et al. 2015; Meece et al. 2016).

## 2.2. Ballistic Condensation

A more quantitative relationship between uplift and condensation can be derived from the cooling history of an uplifted gas blob that follows a purely radial ballistic trajectory without incurring any heat input. For simplicity, consider a gravitational potential in which  $v_c$  is constant with radius, so that  $\phi(r) = v_c^2 \ln(r/r_{\text{max}})$ , where  $r_{\text{max}}$  is the radius at which the blob reaches its apex. This implies

$$r = r_{\text{max}} \exp\left(-\frac{v^2}{2v_c^2}\right), \quad (4)$$

where  $v$  is the radial velocity, which leads to the equation of motion

$$\ddot{r} = \dot{v} = -\frac{v_c^2}{r_{\text{max}}} \exp\left(\frac{v^2}{2v_c^2}\right). \quad (5)$$

We are interested in knowing the blob's time of flight  $\Delta t$  from an initial velocity  $v_1$  to a final velocity  $v_2$ , which can be determined by integrating the equation of motion:

$$\begin{aligned} \Delta t &= \frac{r_{\text{max}}}{v_c^2} \int_{v_2}^{v_1} \exp\left(\frac{v^2}{2v_c^2}\right) dv \\ &= \frac{\pi^{1/2}}{2} \left[ \text{erf}\left(\frac{v_1}{\sqrt{2} v_c}\right) - \text{erf}\left(\frac{v_2}{\sqrt{2} v_c}\right) \right] t_{\text{ff}}(r_{\text{max}}). \end{aligned} \quad (6)$$

The maximum time of flight is then  $\max(\Delta t) = \pi^{1/2} t_{\text{ff}}(r_{\text{max}})$ , and setting  $v_1 = -v_2 = v_c$  gives  $\Delta t \approx 1.2 t_{\text{ff}}(r_{\text{max}})$ .

Assuming for the time being that the blob's cooling time remains nearly independent of gas pressure during its rise and fall, we can estimate the time of flight necessary for condensation by integrating  $d \ln K / dt = -t_{\text{cool}}^{-1}$  to obtain

$$K \approx K_1 \left[ 1 - \frac{6-3\lambda}{5} \frac{\Delta t}{t_{\text{cool}}(r_1)} \right]^{5/(6-3\lambda)}. \quad (7)$$

The condition for producing condensation is therefore

$$\Delta t \gtrsim \frac{5}{6-3\lambda} t_{\text{cool}}(r_1), \quad (8)$$

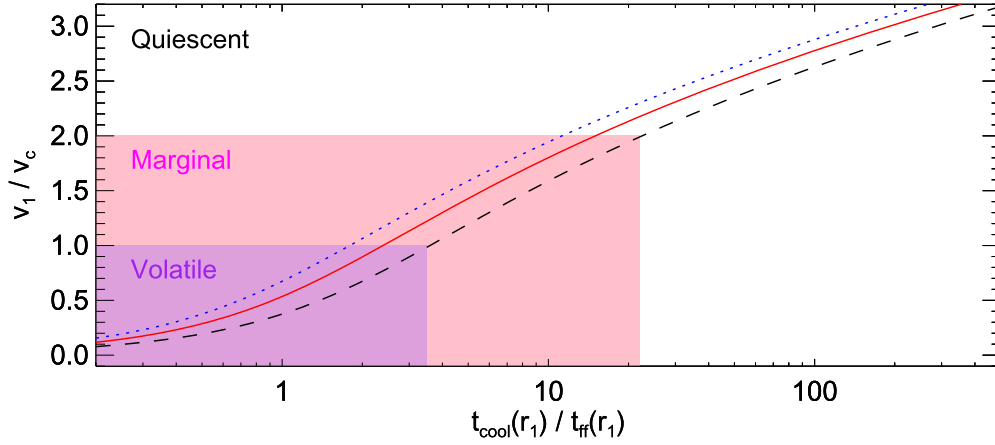
or, equivalently,

$$\frac{v_1}{v_c} \gtrsim \sqrt{2 \ln \left[ \frac{5}{6-3\lambda} \frac{t_{\text{ff}}(r_{\text{max}})}{\Delta t} \left( \frac{t_{\text{cool}}}{t_{\text{ff}}} \right)_{r_1} \right]}. \quad (9)$$

The right-hand side of this second expression is generally close to unity, as long as  $t_{\text{cool}}/t_{\text{ff}}$  is not orders of magnitude greater than unity. Thus, the question of condensation depends primarily on the initial velocity of the blob's ballistic trajectory. Figure 2 shows the critical values of  $v_1/v_c$  required for condensation, according to Equation (9), as functions of  $t_{\text{cool}}/t_{\text{ff}}$  for  $-1 \leq \lambda \leq 0.5$ . Notice that uplift velocities  $v_1 \approx v_c$  can induce condensation of blobs that initially have  $t_{\text{cool}}/t_{\text{ff}} \approx 3$  but that  $v_1 \gtrsim 1.5v_c$  is needed for uplift to induce condensation of blobs that initially have  $t_{\text{cool}}/t_{\text{ff}} \gtrsim 10$ . This simple relationship is likely to be the biggest driver of condensation in more complex treatments of multiphase outflows (e.g., Thompson et al. 2016).

## 3. Thermal Instability: General Considerations

Uplift is not the only route to condensation. Under certain conditions, a medium in thermal balance can be unstable to condensation when small isobaric entropy perturbations are introduced. This topic has considerable heritage in the literature, and some of those findings can appear contradictory. Therefore, in order to clarify the conditions under which condensation can occur, this section goes back to basics to outline what happens to an entropy perturbation in a background medium in which the net cooling rate per unit mass  $\mathcal{L}$  depends on both local thermodynamic conditions and the



**Figure 2.** Values of uplift velocity  $v_1$  required for a gas blob to condense on a ballistic trajectory starting at radius  $r_1$  as a function of its value of  $t_{\text{cool}}/t_{\text{ff}}$  at  $r_1$ . Lines correspond to different values of the cooling-function slope  $\lambda \equiv d \ln \Lambda / d \ln T$ , with a dotted blue line showing  $\lambda = 1/2$  (characteristic of free-free cooling at  $T \gtrsim 10^7$  K), a dashed black line showing  $\lambda = -1$  (characteristic of collisional emission-line cooling at  $10^5 \text{ K} < T < 10^7 \text{ K}$ ), and a solid red line showing  $\lambda = 0$  (characteristic of the crossover at  $T \approx 10^6 \text{ K}$ ). In each case, the line corresponds to trajectories along which the cooling blob formally reaches  $K = 0$  when the blob returns to  $r_1$  in a potential with constant circular velocity  $v_c$ , and cooling time is assumed to be independent of pressure. Rectangles schematically indicate regimes in which a precipitating system is volatile to feedback fueled by condensation (purple;  $v_1/v_c < 1$ ), marginally susceptible to such feedback (pink;  $1 < v_1/v_c < 2$ ), or quiescent (white;  $2 < v_1/v_c$ ).

location within the system. Here we will use the entropy equation  $d \ln K / dt = -(2\mu_p/3kT)\mathcal{L}$ , which is appropriate for a monatomic ideal gas, but the results are similar for more complicated equations of state.

### 3.1. Condensation in a Uniform Background

Without gravity, the pressure  $P$  and entropy  $K$  of the background medium can be uniform. The contrast in entropy between the background and a perturbation with entropy  $\tilde{K}$  is  $\delta \ln K = \ln(\tilde{K}/K)$ , and it evolves with time according to

$$\begin{aligned} \frac{d}{dt}(\delta \ln K) &= -\left(\frac{2\mu_p}{3k}\right)\delta\left(\frac{\mathcal{L}}{T}\right) \\ &= -\left(\frac{2\mu_p}{3k}\right)\frac{\partial(\mathcal{L}/T)}{\partial \ln K} \bigg|_A \delta \ln K, \end{aligned} \quad (10)$$

where  $A$  is an arbitrary thermodynamic quantity complementary to  $K$  that remains constant during the process of condensation. Spatial gradients of  $\mathcal{L}$  do not appear in this equation because they must vanish in order for the background medium to remain uniform. The perturbation's amplitude therefore grows monotonically with time if

$$\frac{\partial(\mathcal{L}/T)}{\partial \ln K} \bigg|_A < 0, \quad (11)$$

which is equivalent to the thermal instability condition originally derived by Balbus (1986). If the background medium is thermally balanced ( $\mathcal{L} = 0$ ), then this condition reduces to

$$\frac{\partial \mathcal{L}}{\partial \ln K} \bigg|_A < 0, \quad (12)$$

which is equivalent to the thermal instability condition originally derived by Field (1965).

To obtain the timescale on which isobaric gas condenses, we set  $A = P$ , so that

$$\begin{aligned} \frac{d}{dt}(\delta \ln K) &= \omega_{\text{ti}}(\delta \ln K), \\ \omega_{\text{ti}} &\equiv -\left(\frac{2\mu_p}{3k}\right)\frac{\partial(\mathcal{L}/T)}{\partial \ln K} \bigg|_P. \end{aligned} \quad (13)$$

Then we separate  $\mathcal{L}$  into a cooling rate per unit mass  $C$  and a heating rate per unit mass  $H$ , so that  $\mathcal{L} = C - H$ . If all of the cooling is radiative, then  $C/T \propto K^{-(6-3\lambda)/5}P^{(1+2\lambda)/5}$  and

$$\omega_{\text{ti}} = \frac{6-3\lambda}{5t_{\text{cool}}} \left[ 1 + \left(\frac{5}{6-3\lambda}\right)\frac{H}{C}\frac{\partial \ln(H/T)}{\partial \ln K} \bigg|_P \right]. \quad (14)$$

The second term inside the square brackets vanishes if  $H = 0$ . It also vanishes if heating per unit volume is constant, because in that case  $H/T$  is constant in an isobaric medium. Condensation therefore happens as long as  $\lambda < 2$  and the factor in square brackets is positive, and it progresses exponentially on a timescale  $\omega_{\text{ti}}^{-1}$ . Temporal variations in  $H$  have no effect on a perturbation's growth rate if both the perturbation and the ambient medium share the same value of  $H/T$  at each moment in time. However, small entropy perturbations in a medium with no heating (i.e.,  $H = 0$ ) do not get much of an opportunity to develop into nonlinear condensates, because the timescale for perturbation growth is then quite similar to the timescale for cooling of the background medium (e.g., Malagoli et al. 1987; Balbus & Soker 1989).

### 3.2. Condensation in a Gravitationally Stratified Background

In the presence of gravity, the pressure of a static background medium cannot be uniform, and condensation couples interestingly with buoyancy (e.g., Defouw 1970; Cowie et al. 1980; Nulsen 1986; Malagoli et al. 1987; Balbus & Soker 1989; Loewenstein 1989). The strength of this coupling depends on the entropy gradient  $\nabla \ln K$  in the

medium, because

$$\frac{\partial}{\partial t}(\delta \ln K) = \omega_{\text{ti}}(\delta \ln K) + \omega_{\text{sw}}(\delta \ln P) - \dot{\xi} \cdot \nabla \ln K, \quad (15)$$

where  $\xi(\mathbf{r}, t) \equiv \mathbf{r} - \mathbf{r}_0(\mathbf{r}, t)$  represents the displacement of gas at location  $\mathbf{r}$  and time  $t$  from its initial location  $\mathbf{r}_0$  at time  $t = 0$ . In this expression, the quantity

$$\omega_{\text{sw}} \equiv -\left(\frac{2}{3} \frac{\mu m_p}{k}\right) \frac{\partial(\mathcal{L}/T)}{\partial \ln P} \Big|_{K,r} = -\frac{6 + 2\lambda}{5t_{\text{cool}}} \quad (16)$$

is the inverse of the timescale on which sound waves either grow or decay. The right-hand side of Equation (16) applies to a thermally balanced medium in which heating per unit volume is constant at each radius and implies that circumgalactic sound waves tend to be thermally stable, because  $\omega_{\text{sw}} < 0$  for essentially all environments of interest. Thermal instability consequently does not develop unless  $\omega_{\text{ti}} > 0$ . However, subsequent motions resulting from changes in buoyancy can either increase or diminish the perturbation's (Eulerian) entropy contrast  $\delta \ln K$ , depending on the sign of  $\dot{\xi} \cdot \nabla \ln K$ .

We are primarily interested in systems in which  $\mathcal{L}$  depends explicitly on position, but it is illuminating to consider first the case in which  $\mathcal{L}$  depends on only the local thermodynamic state, as quantified by  $K$  and  $P$ . In that case, the thermal instability condition  $\omega_{\text{ti}} > 0$  cannot be satisfied in a hydrostatic medium that is convectively stable and in which sound waves are thermally stable, as emphasized by Balbus & Soker (1989). In such a medium,  $K$  increases with altitude while  $P$  decreases with altitude. Therefore, the condition  $\mathcal{L}(K, P) = 0$  requires  $\omega_{\text{ti}}$  and  $\omega_{\text{sw}}$  to have the same sign. Otherwise, there cannot be a configuration that is simultaneously hydrostatic, convectively stable, and thermally balanced. Hence, a background configuration with all of three of these desirable stability characteristics *does not exist* unless  $\omega_{\text{ti}} \leq 0$ .

An explicit dependence of  $\mathcal{L}$  on position  $\mathbf{r}$  allows hydrostatic, convectively stable configurations to exist for a medium with  $\omega_{\text{ti}} > 0$  even if sound waves are thermally stable. In a spherically symmetric environment, a thermally balanced medium in which  $\mathcal{L}$  depends explicitly on position has

$$\alpha_K \frac{\partial \mathcal{L}}{\partial \ln K} \Big|_{P,r} + \alpha_P \frac{\partial \mathcal{L}}{\partial \ln P} \Big|_{K,r} + \frac{\partial \mathcal{L}}{\partial \ln r} \Big|_{K,P} = 0, \quad (17)$$

where  $\alpha_K \equiv \mathbf{r} \cdot \nabla \ln K$  and  $\alpha_P \equiv \mathbf{r} \cdot \nabla \ln P$ . The physical implications of this relationship are easiest to see if  $\mathcal{L}$  depends only on  $K$  and  $P$  while  $H/T$  depends only on  $r$ , giving

$$\frac{d \ln(H/T)}{d \ln r} = -(\alpha_K \omega_{\text{ti}} + \alpha_P \omega_{\text{sw}}) t_{\text{cool}}. \quad (18)$$

According to this equation, thermal instability can proceed in a thermally balanced medium satisfying the other three stability conditions as long as the decline in  $H/T$  with radius is sufficiently rapid.

### 3.3. Circumgalactoseismology

From here on, we will focus on media that have  $\omega_{\text{ti}} > 0$  but in which sound waves are stable ( $\omega_{\text{sw}} < 0$ ). Entropy perturbations can then be thermally unstable and undergo buoyancy-driven

motions governed by the momentum equation

$$(1 + \delta \ln \rho) \ddot{\xi} = \left[ \frac{3}{5}(\delta \ln K) + \frac{2}{5}(\delta \ln P) \right] \nabla \phi - \frac{P}{\rho} \nabla(\delta \ln P). \quad (19)$$

Here,  $\nabla \phi$  represents the gradient of the gravitational potential, and we have assumed  $\nabla P = -\rho \nabla \phi$ , so that the background state is in hydrostatic equilibrium. Differentiating Equation (15) with respect to time and applying Equation (19) after dropping the nonlinear  $(\delta \ln \rho) \ddot{\xi}$  term then gives

$$\left[ \frac{\partial^2}{\partial t^2} - \omega_{\text{ti}} \frac{\partial}{\partial t} + \omega_{\text{buoy}}^2 \right] (\delta \ln K) = \left[ \omega_{\text{sw}} \frac{\partial}{\partial t} - \frac{2}{3} \omega_{\text{buoy}}^2 + \omega_{\text{buoy}}^2 \frac{c_s^2}{g} \frac{\partial}{\partial r} \right] (\delta \ln P), \quad (20)$$

where  $\omega_{\text{buoy}} \equiv [(3/5) \nabla \phi \cdot \nabla \ln K]^{1/2} = (6\alpha_K/5)^{1/2} t_{\text{ff}}^{-1}$  is the usual Brunt-Väisälä frequency for buoyant oscillations and  $c_s = (5P/3\rho)^{1/2}$  is the adiabatic sound speed. Note that this equation has been obtained without using either the continuity equation or the Boussinesq approximation and relates internal gravity waves (left-hand side) to the corresponding pressure disturbances (right-hand side).

If the right-hand side of the equation is negligible, then an entropy perturbation can move through the background medium without dissipative resistance, and the qualitative behavior of the perturbation's (Eulerian) entropy contrast  $\delta \ln K$  is obvious. Entropy perturbations in a medium with  $\omega_{\text{ti}} \gg \omega_{\text{buoy}} > 0$  grow monotonically with time, while those in a medium with  $\omega_{\text{buoy}} \gg \omega_{\text{ti}} > 0$  oscillate with continually increasing amplitude. The oscillations grow because of a phase difference between  $\xi_r \equiv (\mathbf{r} \cdot \xi)/r$ , which is the perturbation's displacement in the radial direction and the radial restoring force, which is proportional to  $\delta \ln K$ . During most of each oscillation cycle, the product of  $\xi_r$  and  $\delta \ln K$  is negative, meaning that buoyancy forces are usually pushing the perturbation back toward the midpoint of the oscillations. However,  $\delta \ln K > 0$  at the beginning of each upward excursion above the midpoint, and  $\delta \ln K < 0$  at the beginning of each downward excursion below the midpoint. The resulting buoyancy impulses cause the amplitudes of both  $\xi$  and  $\delta \ln K$  to increase with time, as originally found by Defouw (1970).

Substituting  $\delta \ln K \propto e^{-i\omega t}$  into Equation (20) with the right-hand side set to zero gives the frequency solutions

$$\omega = \omega_{\pm} \equiv \frac{i\omega_{\text{ti}}}{2} \pm \left( \omega_{\text{buoy}}^2 - \frac{\omega_{\text{ti}}^2}{4} \right)^{1/2}. \quad (21)$$

Low-entropy perturbations in such a medium progress steadily toward nonlinear condensation if  $\omega_{\text{buoy}} < \omega_{\text{ti}}/2$ , which corresponds to

$$\alpha_K < \frac{3(2 - \lambda)^2}{40} \left( \frac{t_{\text{ff}}}{t_{\text{cool}}} \right)^2, \quad (22)$$

when expressed in terms of  $t_{\text{cool}}$ ,  $t_{\text{ff}}$ , and  $\alpha_K$ . However, the qualitative behavior of the oscillatory entropy perturbations that arise when  $\omega_{\text{buoy}} \gg \omega_{\text{ti}}$  depends critically on how one treats the



$\delta \ln P$  side of the equation. We will now proceed with a careful look at those  $\delta \ln P$  terms, in order to clarify some seemingly contradictory claims in the astrophysical literature on local thermal instability.

### 3.3.1. The Bobbing-blob Approximation

One intuitively appealing approach is to idealize an entropy perturbation as a discrete, isolated blob in which all of the gas shares the same displacement vector (e.g., Cowie et al. 1980; Nulsen 1986; Loewenstein 1989). In a medium with  $\omega_{\text{buoy}} \gg \omega_{\text{ti}}$ , the vertical bobbing of the blob necessarily excites internal gravity waves that propagate away from the blob. This transfer of kinetic energy from the blob to its environment implies that the blob's motion must experience some resistance in addition to the background pressure gradient responsible for maintaining the oscillations. Adding the necessary resistance by replacing the pressure and density perturbation terms in the equation of motion with a heuristic dissipation term, so that

$$\ddot{\xi} = \frac{3}{5}(\delta \ln K) \nabla \phi - \omega_{\text{D}} \dot{\xi}, \quad (23)$$

qualitatively accounts for those kinetic energy losses by dissipating the blob's kinetic energy on the timescale  $\omega_{\text{D}}^{-1}$ . This replacement results in the frequency solutions

$$\omega = \omega_{\pm} \equiv \frac{i(\omega_{\text{ti}} - \omega_{\text{D}})}{2} \pm \left[ \omega_{\text{buoy}}^2 - \frac{(\omega_{\text{ti}} + \omega_{\text{D}})^2}{4} \right]^{1/2}, \quad (24)$$

which show that the blob's oscillations decay instead of growing if the timescale for dissipating kinetic energy is shorter than the thermal timescale for pumping the oscillations.

Associating the dissipation term with hydrodynamical drag gives  $\omega_{\text{D}} \sim |\dot{\xi}|/\Delta r$ , where  $\Delta r$  is the blob's thickness in the vertical direction. Given this scaling for  $\omega_{\text{D}}$ , thermal pumping will increase the amplitudes of small oscillations until they saturate at  $\delta \ln K \sim \alpha_K (\omega_{\text{ti}}/\omega_{\text{buoy}})(\Delta r/r)$ . The largest (Eulerian) density fluctuations,  $\delta \ln \rho \sim \alpha_K^{1/2} (t_{\text{ff}}/t_{\text{cool}})$ , are therefore obtained for  $\Delta r \sim r$ . This analytical estimate of the saturation threshold agrees well with numerical simulations of oscillatory thermal instability in thermally balanced media (e.g., McCourt et al. 2012; Meece et al. 2015). However, the route through which it was obtained indicates that the saturation process is inherently nonlinear. Furthermore, the bobbing-blob approximation ignores what happens to the propagating internal gravity waves stimulated by the blob's oscillations.

### 3.3.2. Plane-wave Perturbations

Alternatively, one can treat the perturbations as Eulerian plane waves (e.g., Malagoli et al. 1987; Balbus 1988; Balbus & Soker 1989), but then some care needs to be taken with the physical interpretation. Making the usual plane-wave assumption that  $\delta \ln K$ ,  $\delta \ln P$ , and  $\xi$  are all  $\propto \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$  and defining  $\mathcal{R} \equiv \delta \ln P / \delta \ln K$  allows us to convert Equation (20) to the dispersion relation

$$\omega^2 - i(\omega_{\text{ti}} + \mathcal{R}\omega_{\text{sw}})\omega - \left(1 + \frac{2}{3}\mathcal{R} - i\mathcal{R}\frac{c_s^2 k_r}{g}\right)\omega_{\text{buoy}}^2 = 0. \quad (25)$$

In order to account for the influence of pressure perturbations on this dispersion relation through the ratio  $\mathcal{R}$ , we need to bring in the continuity equation for the perturbation, which can be expressed as

$$\frac{5}{3}(\nabla \cdot \xi) = \Delta \ln K - \Delta \ln P \quad (26)$$

when written in terms of the Lagrangian perturbation amplitudes  $\Delta \ln P = \delta \ln P + \alpha_P(\xi_r/r)$  and  $\Delta \ln K \equiv \delta \ln K + \alpha_K(\xi_r/r)$ . Using the plane-wave assumption to simplify the tangential momentum equation, one can write the divergence of the displacement field as

$$\nabla \cdot \xi = (ik_r r + \nabla \cdot \mathbf{r}) \frac{\xi_r}{r} - \frac{3c_s^2 k_{\perp}^2}{5\omega^2} \delta \ln P, \quad (27)$$

where  $k_r$  and  $k_{\perp}$  are the radial and tangential wavenumbers, respectively. The coordinate divergence term is  $\nabla \cdot \mathbf{r} = 2$  in a spherical environment but can be ignored in a plane-parallel approximation. Combining Equations (26) and (27) and recognizing that  $\alpha_P = -5gr/3c_s^2$  then leads to

$$\mathcal{R} = \frac{\omega^2}{\omega^2 - c_s^2 k_{\perp}^2} \left[ \frac{5}{3\alpha_K} \left( ik_r r + \nabla \cdot \mathbf{r} - \frac{gr}{c_s^2} \right) \times \left( 1 - \frac{\Delta \ln K}{\delta \ln K} \right) + \frac{\Delta \ln K}{\delta \ln K} \right]. \quad (28)$$

Applying the plane-wave assumption to Equation (15) gives the ratio of Lagrangian to Eulerian entropy perturbation amplitudes:

$$\frac{\Delta \ln K}{\delta \ln K} = i \left( \frac{\omega_{\text{ti}} + \mathcal{R}\omega_{\text{sw}}}{\omega} \right). \quad (29)$$

The dispersion relation for perturbations with short radial wavelengths ( $|k_r r| \gg 1$ ) therefore reduces to

$$\omega^2 - i(\omega_{\text{ti}} + \mathcal{R}\omega_{\text{sw}})\omega - \left(1 + \frac{2}{3}\mathcal{R}\right) \left( \frac{c_s^2 k_{\perp}^2 - \omega^2}{c_s^2 k^2 - \omega^2} \right) \omega_{\text{buoy}}^2 = 0, \quad (30)$$

where  $k^2 \equiv k_r^2 + k_{\perp}^2$ .

The literature on thermal instability in circumgalactic media has generally focused on short-wavelength modes with  $\omega^2 \ll c_s^2 k_{\perp}^2 / |k_r r|$ , which are nearly pure internal gravity waves. In that case, the amplitude ratio  $\mathcal{R} = \delta \ln P / \delta \ln K$  is small, and the dispersion relation reduces further to

$$\omega^2 - i\omega_{\text{ti}}\omega - \frac{k_{\perp}^2}{k^2} \omega_{\text{buoy}}^2 = 0 \quad (31)$$

(e.g., Malagoli et al. 1987; Binney et al. 2009). For these modes, Eulerian pressure perturbations influence the gravity-wave oscillator only through the pressure-gradient term in Equation (20). The resulting  $k_{\perp}^2/k^2$  factor slows the oscillations because the gravitational force correcting a given radial displacement meets with greater inertial resistance from transverse displacements as  $k_{\perp}^2/k^2$  declines (e.g., Binney et al. 2009). As a consequence, the critical value of  $\alpha_K$  separating monotonic perturbation growth from oscillatory



growth for a given  $t_{\text{cool}}/t_{\text{ff}}$  ratio becomes larger than one would obtain by setting  $\delta \ln P = 0$ .

Some treatments in the literature (e.g., Balbus & Soker 1989; McCourt et al. 2012) arrive at a similar dispersion relation by setting  $\nabla \cdot \xi = 0$ . With that restriction, we find

$$\mathcal{R} = \frac{\alpha_P}{\alpha_K} \left( 1 - \frac{\Delta \ln K}{\delta \ln K} \right) + \frac{\Delta \ln K}{\delta \ln K}, \quad (32)$$

and obtain the dispersion relation

$$\omega^2 - i(\omega_{\text{ti}} + \mathcal{R}\omega_{\text{sw}})\omega - \left( 1 + \frac{2}{3}\mathcal{R} \right) \times \left[ \frac{k_{\perp}^2 r^2}{k^2 r^2 + ik_r r (\nabla \cdot \mathbf{r})} \right] \omega_{\text{buoy}}^2 = 0 \quad (33)$$

without having to assume  $|k_r r| \gg 1$ . If the  $ik_r r (\nabla \cdot \mathbf{r})$  term can be ignored, then this relation would reduce to Equation (31) if  $\mathcal{R}$  were small. However,  $|\mathcal{R}|$  is typically of order unity for divergence-free modes, indicating that they are qualitatively different from the internal gravity waves discussed in the previous paragraph. In fact, the restriction  $\nabla \cdot \xi = 0$  implies that the radial wavenumber for such modes is imaginary, with

$$k_r r = i \left[ k_{\perp}^2 r^2 \left( \frac{c_s^2}{gr} \right) \frac{\omega_{\text{buoy}}^2}{\omega^2} + (\nabla \cdot \mathbf{r}) \right] \quad (34)$$

in the adiabatic limit. These divergence-free modes are therefore evanescent in the radial direction. Furthermore, the associated growth rate of thermal instability is  $\omega_{\text{ti}} + (\alpha_P/\alpha_K)\omega_{\text{sw}}$ , which exceeds  $\omega_{\text{ti}}$  in media with stable sound waves (i.e.,  $\omega_{\text{sw}} < 0$ ), because  $\alpha_P/\alpha_K < 0$ .

### 3.3.3. Onset of Nonlinearity

Prior analyses of circumgalactic thermal instability did not focus on the consequences of modes in which  $\omega_{\text{sw}}$  is important, but here we would like to understand why the growth of oscillatory thermal instability saturates with  $\delta \ln \rho \sim \alpha_K^{1/2}(t_{\text{ff}}/t_{\text{cool}})$ , as predicted by the heuristic bobbing-blob model and as observed in numerical simulations (McCourt et al. 2012; Meece et al. 2015). If only low- $\mathcal{R}$  or divergence-free modes are permitted, then there is no channel (other than viscosity) for the dissipation of the kinetic energy that oscillatory thermal instability introduces into the system. Nonlinear mode coupling can exchange energy among those gravity-wave modes but does not by itself limit the overall growth of a set of modes that are all thermally unstable. In order to achieve a steady state in which  $\delta \ln \rho$  depends on  $t_{\text{cool}}$ , nonlinear mode coupling must instead channel the accumulating gravity-wave energy into acoustic modes, which damp through radiative losses on a timescale  $\sim |\omega_{\text{sw}}^{-1}| \sim t_{\text{cool}}$ .

In lieu of presenting a complete solution to that mode-coupling problem in this paper, we would instead like to briefly illustrate how the gravity-wave saturation scale  $\delta \ln \rho \sim \alpha_K^{1/2}(t_{\text{ff}}/t_{\text{cool}})$  arises from nonlinearity. We are interested in knowing how mode coupling changes the amplitudes  $A_i$  of the radial-displacement eigenmodes  $\xi_i = A_i \exp[i(\mathbf{k}_i \cdot \mathbf{r} - \omega_i t)]$ , for which  $\omega_i(\mathbf{k}_i)$  satisfies Equations (25), (28), and (29). The full radial momentum equation for a superposition of those modes

can be expressed as

$$(1 + \delta \ln \rho) \left( \sum_i \ddot{\xi}_i \right) = - \sum_j \omega_j^2 \xi_j, \quad (35)$$

and the amplitudes  $A_i$  remain constant if the nonlinear terms are ignored. After canceling the linear terms, we can rewrite this equation in a form that relates the rates of amplitude change to those nonlinear terms,

$$\begin{aligned} \sum_i \left[ \left( \frac{\ddot{A}_i - i2\omega_i \dot{A}_i}{A_i} \right) \xi_i \right] &= - \left( \sum_j \omega_j^2 \xi_j \right) \left[ \sum_{n=1}^{\infty} (-\delta \ln \rho)^n \right] \\ &\approx - \left( \sum_j \omega_j^2 \xi_j \right) \left[ \sum_m (\delta \ln \rho)_m \right], \end{aligned} \quad (36)$$

where  $(\delta \ln \rho)_m$  represents the oscillating density perturbation corresponding to eigenmode  $\xi_m$ , and we have kept only the lowest-order nonlinear terms on the right-hand side. Resonant energy transfer can therefore happen among triads of modes that satisfy  $\mathbf{k}_i = \mathbf{k}_j + \mathbf{k}_m$  and  $\omega_i = \omega_j + \omega_m$ .

Here, we are most interested in coupling between a typical internal gravity wave of wavenumber  $k$  and frequency  $\omega_g \approx (k_{\perp}/k)\omega_{\text{buoy}}$ , and a pair of sound waves with frequencies  $\sim c_s k$  that differ by  $\omega_g$  so that their beat pattern resonates with the gravity wave. In a purely adiabatic system, this triad of waves would eventually come into energy equipartition with  $A_g \omega_g \sim A_s c_s k$ , where  $A_g$  and  $A_s$  are the gravity-wave and sound-wave displacement amplitudes, respectively. On the right-hand side of Equation (36), terms corresponding to pairs of sound waves of appropriate wavelength have magnitudes  $\sim (A_s c_s k)^2 k$  because  $(\delta \ln \rho)_m \sim A_s k_m$ . In a system in equipartition, these terms are  $\sim (A_g \omega_g)^2 k$ , implying that mode coupling transfers energy from sound waves into gravity waves and back again at a rate  $\sim A_g \omega_g k$ .

The system we are considering is not adiabatic. Instead, gravity waves grow in amplitude on a timescale  $\sim \omega_{\text{ti}}^{-1}$  until their amplitudes are large enough for mode coupling to channel their energy into dissipative sound waves on the same timescale. This happens when

$$\begin{aligned} \frac{A_g}{r} &\sim \frac{\omega_{\text{ti}}}{\omega_{\text{buoy}}} \frac{1}{kr}, \\ (\delta \ln \rho)_g &\sim \delta \ln K \sim \alpha_K \frac{A_g}{r} \\ &\sim \alpha_K \frac{\omega_{\text{ti}}}{\omega_{\text{buoy}}} \frac{1}{kr} \sim \alpha_K^{1/2} \frac{t_{\text{ff}}}{t_{\text{cool}}} \frac{1}{kr}, \end{aligned} \quad (37)$$

where  $(\delta \ln \rho)_g$  is the density perturbation scale associated with gravity-wave oscillations.

### 3.4. Buoyancy Damping

The saturation amplitude derived from nonlinear mode coupling considerations in Section 3.3.3 closely resembles the one derived from the heuristic bobbing-blob model in Section 3.3.1. In the bobbing-blob case, the timescale for the energy transfer from the blob to the surrounding medium is

$\sim (A_g \omega_{\text{buoy}})^{-1} \Delta r$ , and energy is permanently lost if it propagates away from the blob without contributing to the restoring force on it. Analogously, the timescale for energy transfer from internal gravity waves to damped acoustic modes is  $\sim (A_g \omega_{\text{buoy}})^{-1} k^{-1}$ , regardless of the mechanism that dissipates sound-wave energy. In both cases, the saturation amplitudes are largest for inhomogeneities with  $kr \sim 1$ .

Saturation of thermally unstable gravity-wave oscillations therefore has important consequences for condensation in the circumgalactic medium. In an otherwise static medium with  $\alpha_K \gg (t_{\text{ff}}/t_{\text{cool}})^2$ , energy transfer from buoyant oscillations into dissipative modes prevents a gas blob with a slightly lower entropy than its surroundings from condensing. At first, the blob descends faster than it can radiate away its thermal energy and accelerates until it passes through a layer of equivalent entropy. After it overshoots that layer, it begins to bob, and its subsequent behavior depends on the amplitude  $|\delta \ln K|$  of its Eulerian entropy contrast. If that amplitude is less than the saturation amplitude, then thermal pumping will cause it to grow to the saturation amplitude. But if  $|\delta \ln K|$  is greater than the saturation amplitude, then coupling of buoyant oscillations to acoustic modes causes  $|\delta \ln K|$  to decay until it reaches the saturation amplitude. Hereafter, we will refer to such a decline in perturbation amplitude as *buoyancy damping*, because it results from the damping of buoyant oscillations in a gravitationally stratified medium.

In contrast, buoyancy does not inhibit condensation in a medium with  $\alpha_K \ll (t_{\text{ff}}/t_{\text{cool}})^2$ , because there are no oscillations. Instead, low-entropy blobs proceed steadily toward becoming nonlinear condensates. In fact, the onset of nonlinearity can actually *assist* perturbation growth, because it inhibits the descent of low-entropy gas toward layers of equivalent entropy (e.g., Nulsen 1986). Therefore, the condition

$$\alpha_K \sim \left( \frac{t_{\text{ff}}}{t_{\text{cool}}} \right)^2 \quad (38)$$

corresponds to a critical entropy-profile slope separating media in which linear thermal instability leads to condensation from media in which linear thermal instability saturates at relatively small amplitudes.<sup>5</sup>

At the margin between these regimes, where  $\omega_{\text{buoy}} \sim \omega_{\text{ti}}$ , thermal instability itself starts to reduce the frequency of gravity-wave oscillations, which slows the rate of energy transfer into sound waves. In a system with  $\alpha_K \gtrsim (t_{\text{ff}}/t_{\text{cool}})^2$ , the saturation scale for thermally unstable gravity-wave oscillations with  $kr \sim 1$  becomes

$$\frac{A_g}{r} \sim \frac{\omega_{\text{ti}}}{\omega_{\text{buoy}}} \left( 1 - \frac{\omega_{\text{ti}}^2}{4\omega_{\text{buoy}}^2} \right)^{-1/2},$$

$$(\delta \ln \rho)_g \sim \alpha_K \frac{A_g}{r} \gtrsim \frac{t_{\text{ff}}^2}{t_{\text{cool}}^2} \left( 1 - \frac{\omega_{\text{ti}}^2}{4\omega_{\text{buoy}}^2} \right)^{-1/2}. \quad (39)$$

<sup>5</sup> Binney et al. (2009) arrive at a similar-looking criterion, their Equation (10), via different reasoning. Instead of arguing that oscillatory perturbations saturate at a relatively small amplitude when  $t_{\text{cool}} \gg t_{\text{ff}}$ , they instead state without proof that radiative cooling will not have a big impact on gravity-wave oscillations that spend half of their time overdense and cool and the other half of it underdense and warm.

In other words, there is a lower limit of  $(\delta \ln \rho)_g \sim (t_{\text{ff}}/t_{\text{cool}})^2$  on the saturation amplitude of gravity-wave perturbations, and the limiting amplitude approaches unity as  $\omega_{\text{buoy}}$  approaches  $\omega_{\text{ti}}/2$  from above. A reduction of the entropy gradient to  $\alpha_K \sim (t_{\text{ff}}/t_{\text{cool}})^2$  consequently eliminates buoyancy damping and allows thermal instability to proceed to condensation.<sup>6</sup>

#### 4. Thermal Instability in Idealized Numerical Simulations

The analysis in Section 3 provides a conceptual foundation for interpreting the conditions governing thermal instability and condensation in numerical simulations of circumgalactic media. McCourt et al. (2012) showed that thermal instability in the oscillatory regime does indeed saturate at relatively small perturbation amplitudes, while the simulations of Meece et al. (2015) show something more: condensation and feedback in one part of a globally balanced system can alter the behavior of thermal instability elsewhere in the system by changing the large-scale slope of its entropy profile. This section applies the results of Section 3 to show how that happens.

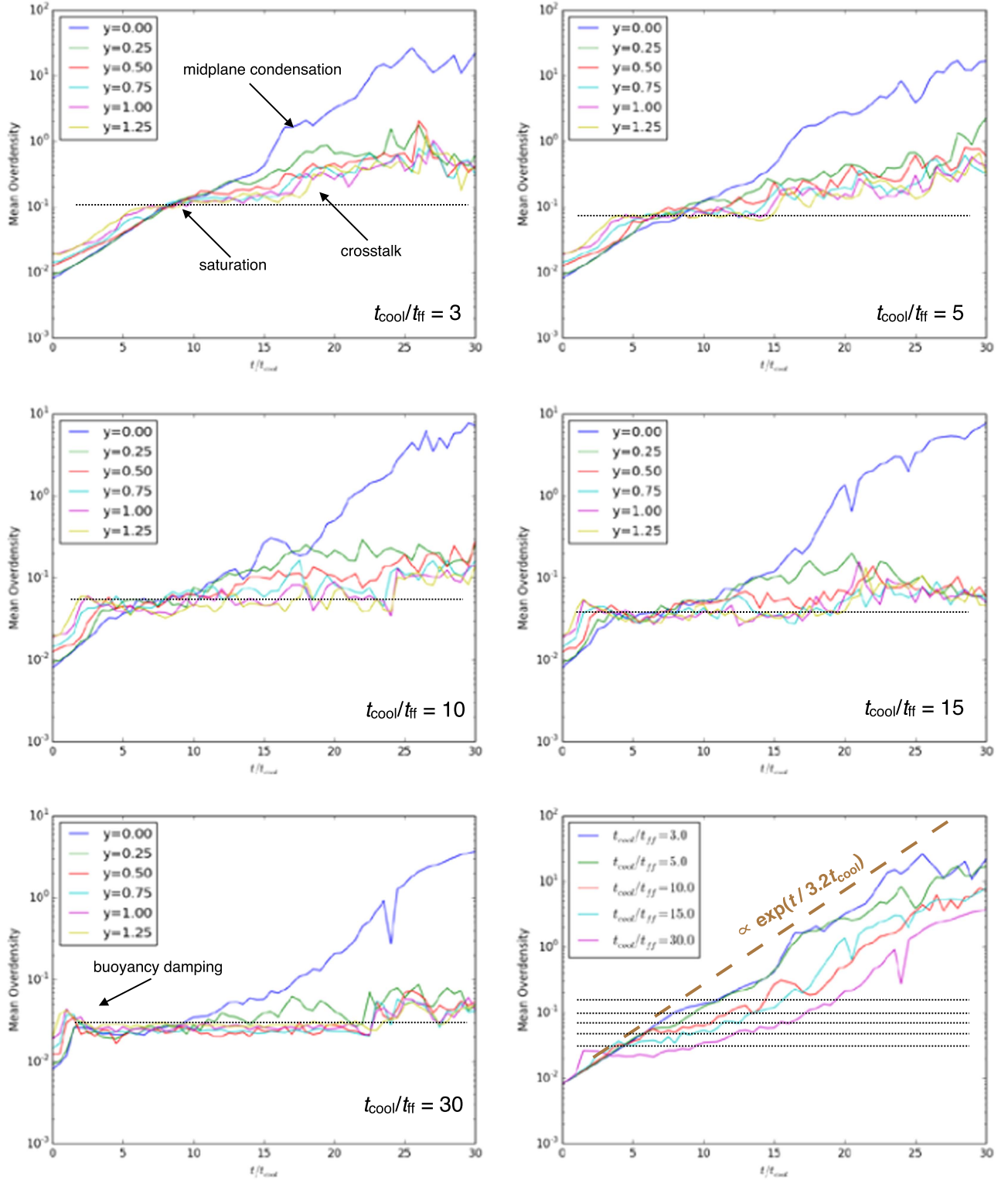
Before proceeding, we want to emphasize that this connection between thermal instability and the large-scale entropy gradient turns out to be critical for understanding how feedback regulates condensation, star formation, and black-hole accretion. Most critically, it implies that the central injection of thermal energy inevitably promotes condensation because it reduces the central entropy gradient. Therefore, if feedback energy is to suppress condensation, it most effectively does so when distributed in ways that *increase* the system's entropy gradient, as we will show in Sections 5 and 6.

##### 4.1. Buoyancy Damping in Simulations

McCourt et al. (2012) drew attention to the dependence of saturation amplitude on  $t_{\text{cool}}/t_{\text{ff}}$  using simulations of a plane-parallel medium in which heating balances cooling within each equipotential layer. They showed that small entropy perturbations introduced into an initially isothermal medium grow to an Eulerian entropy contrast  $\sim \alpha_K (\omega_{\text{ti}}/\omega_{\text{buoy}})$  before saturating. The interpretation they offered for this saturation was that nonlinear mode coupling leads to dissipation when  $\xi_r/r \sim \omega_{\text{ti}}/\omega_{\text{buoy}}$ , but they were not specific about the nature of the dissipation or the mechanism through which mode coupling disrupts the growth of gravity-wave oscillations. They concluded that the criterion  $t_{\text{cool}}/t_{\text{ff}} \lesssim 1$  was a necessary condition for condensation in a plane-parallel medium.

Meece et al. (2015) performed similar simulations showing that the condensation criterion is more complex, in that condensation is always possible near the midplane, even in systems with  $t_{\text{cool}}/t_{\text{ff}} \gg 1$ , if enough time is allowed (see also Choudhury & Sharma 2016). Figure 3 shows some previously unpublished results from the Meece et al. effort. Perturbation growth far from the midplane saturates as in McCourt et al. (2012) but is not fully suppressed near the midplane. Instead, it is significantly delayed for large values of  $t_{\text{cool}}/t_{\text{ff}} \gg 1$  and

<sup>6</sup> After submitting this paper, we recognized that turbulence can alter the condensation criterion by interfering with buoyancy damping. Turbulence enters the picture when the turbulent velocity dispersion is similar in magnitude to the buoyancy-driven motions that result in damping of thermal instability. It can enable condensation by levitating low-entropy blobs that would otherwise descend in a medium with  $\alpha_K > (t_{\text{ff}}/t_{\text{cool}})^2$ . We will explore more deeply how turbulence alters the condensation criterion in a separate paper. For a complementary viewpoint on the role of turbulence in promoting condensation, see Gaspari et al. (2017).



**Figure 3.** Growth of the rms amplitudes of density perturbations through thermal instability in thermally balanced plane-parallel simulations. Each panel shows perturbation growth as a function of  $t/t_{\text{cool}}$  in the simulations of Meece et al. (2015), in which average heating equals average cooling in each equipotential layer. The simulations all begin with nearly isentropic initial conditions within a pressure scale height of the midplane, and labels in five of the panels show the initial value of  $t_{\text{cool}}/t_{\text{ff}}$  in the simulation domain. Lines of different colors in those panels show  $\langle \Delta\rho/\rho \rangle_{\text{rms}}$  at different heights above the midplane, where  $y$  is the height in units of the initial pressure scale height; the blue line shows condensation at the midplane ( $y = 0$ ). Black dotted lines in those five panels show the level at which perturbations saturate because of buoyancy damping. The sixth panel (lower right) shows the time required for each simulation to progress toward midplane condensation (i.e., mean overdensity  $\approx 1$ ). The five black dotted lines in that panel are the same as those in the other five panels, and a brown dashed line extrapolates the exponential growth rate that precedes saturation. When midplane condensation happens, the compensating release of heat produces buoyant bubbles (convective crosstalk) that amplify the perturbation amplitudes in higher-lying layers.

eventually generates crosstalk that affects perturbation growth in higher-lying layers, for reasons we will now proceed to analyze.

#### 4.2. Midplane Condensation

Condensation is always possible near the midplane of a macrophysically balanced system in a plane-parallel potential well because buoyancy damping is ineffective there as long as the gas density does not become infinite at the midplane. In the bobbing-blob approximation of Section 3.3.1, the boundary between monotonic growth and oscillatory thermal instability is at  $\omega_D = \omega_{ti}$ , implying that the condition for condensation is

$$\alpha_K < \frac{3(2 - \lambda)^2}{10} \left( \frac{t_{ff}}{t_{cool}} \right)^2. \quad (40)$$

In order to determine the zone around the midplane in which condensation can proceed, one can adopt a linear approximation to the potential, in which the freefall time  $t_{ff,0}$  is constant with height. The pressure and entropy at the midplane are  $P_0$  and  $K_0$ , respectively, giving a midplane density  $\rho_0 = \mu m_p (n/n_e)^{2/5} (P_0/K_0)^{3/5}$ . To lowest order in height  $z$  above the midplane, the pressure profile is  $P(z) = P_0 [1 - (z/z_P)^2]$ , where  $z_P \equiv t_{ff,0} (P_0/\rho_0)^{1/2}$ , and the linearized entropy profile can be expressed as  $K(z) = K_0 [1 + (z/z_K)]$ . The condensation condition in this zone is therefore

$$\frac{z}{z_K} \lesssim \frac{3(2 - \lambda)^2}{10} \left[ \frac{t_{ff,0}}{t_{cool,0}} \right]^2 \left[ 1 + \left( \frac{z}{z_K} \right) \right]^{6(2-\lambda)/5} \times \left[ 1 - \left( \frac{z}{z_P} \right)^2 \right]^{2(1+2\lambda)/5}, \quad (41)$$

where  $t_{cool,0} \equiv t_{cool}(K_0, P_0)$ . In other words, the thickness of the midplane region that is unstable to condensation is approximately  $(t_{ff}/t_{cool})^2$  times the entropy scale height in media with  $t_{cool}/t_{ff} \gg 1$ .

#### 4.3. Isentropic Initial Conditions

At first glance, the susceptibility to condensation of a system in which  $\alpha_K \lesssim (t_{ff}/t_{cool})^2$  would appear to be in conflict with the plane-parallel simulations of McCourt et al. (2012) that start with isentropic initial conditions ( $\alpha_K = 0$ ) and fail to produce nonlinear density contrasts. In those simulations, the fractional fluctuations in entropy and density at one pressure scale height above the midplane saturate at  $\sim (t_{ff}/t_{cool})^2$  and remain at this amplitude for many cooling times. The reason for saturation of perturbation growth in what would seem to be an isentropic medium is that buoyant migration of the initial perturbations causes an entropy gradient to develop. As lower-entropy fluctuations sink and higher-entropy fluctuations rise, the resulting radial sorting of entropy perturbations increases  $\alpha_K$ . Eventually, the entropy gradient exceeds  $\alpha_K \sim (t_{ff}/t_{cool})^2$ , and the condensation condition is no longer satisfied. The system then transitions to the bobbing-blobs regime, in which (Eulerian) entropy fluctuations saturate at a fractional amplitude  $\sim \alpha_K (\omega_{ti}/\omega_{buoy})$ , which corresponds to  $\sim (t_{ff}/t_{cool})^2$  for  $\alpha_K \sim (t_{ff}/t_{cool})^2$ , as shown in Section 3.4.

McCourt et al. (2012) recognized that the development of an entropy gradient through convective sorting of the initial

perturbations was responsible for suppressing perturbation growth away from the midplane but did not fully explore the consequences for stability close to the midplane. Instead, they suppressed both heating and cooling near the midplane. In contrast, Meece et al. (2015) chose to allow heating and cooling to proceed near the midplane, and those simulations ended up displaying qualitative behavior distinctly different from the initially isentropic McCourt et al. (2012) simulations, even though they started with nearly isentropic initial conditions within a pressure scale height of the midplane.

Midplane condensation always occurs in the Meece et al. (2015) simulations, no matter what the initial value of  $t_{cool}/t_{ff}$ , because there is always a zone around the midplane in which  $\alpha_K \lesssim (t_{cool}/t_{ff})^2$ . However, buoyancy damping delays condensation by a time interval depending on  $t_{cool}/t_{ff}$ . Without gravity and buoyancy damping, the initial entropy perturbations of Meece et al. (2015) would experience monotonic exponential growth. Nearly exponential growth is indeed seen in simulations with  $t_{cool}/t_{ff} \lesssim 3$ , but larger values of  $t_{cool}/t_{ff}$  lead to the saturation of perturbation growth at progressively smaller amplitudes (see Figure 3). Buoyancy damping suppresses perturbation growth away from the midplane but succeeds only in delaying it near the midplane. Delays occur because buoyancy damping limits the initial amplitudes of the seed perturbations that enter the isentropic zone and ultimately condense. As a result, the time required for condensation at the midplane is a larger multiple of  $t_{cool}$  in systems with larger values of  $t_{cool}/t_{ff}$ .

#### 4.4. Convective Crosstalk

Once condensation begins at the midplane in the Meece et al. (2015) simulations, the idealized algorithm for enforcing thermal balance within each layer ends up stimulating perturbation growth at greater altitudes. As condensates form near the midplane, the requirement for heating to balance cooling in each layer causes compensating hot bubbles to form. Those hot bubbles then rise out of the condensation zone. When they reach higher layers, the bubbles catalyze additional condensation, because regions that were formerly at the mean entropy for their layer, and therefore close to thermal balance, suddenly find themselves significantly below the new mean entropy level. The net cooling rate in lower-entropy regions of the layer must therefore increase in order to satisfy the thermal balance constraint. As a result, convective crosstalk between layers causes nonlinear perturbation growth to propagate beyond the thermally unstable midplane zone. The same effect was also found in simulations by Choudhury & Sharma (2016).

Does this crosstalk phenomenon, which arises in Meece et al. (2015) and Choudhury & Sharma (2016) from a highly idealized representation of global thermal balance, have a qualitative counterpart in more realistic treatments of feedback? We suggest that the answer to this question may be yes, because the outflow of feedback energy from a central source in a real galactic system is likely to propagate primarily through the lowest-density, highest-entropy ambient gas as it follows the path of least resistance to greater altitudes. Unless some mechanism (such as thermal conduction) exists for preferentially channeling feedback energy into denser, cooler gas clumps, then lower-entropy regions will inevitably cool and condense while feedback energy flows outward around them, even if heating and cooling are in global balance.



This phenomenon illustrates how centrally injected thermal feedback renders the surrounding region vulnerable to condensation: convection eliminates the central entropy gradient and therefore shuts off buoyancy damping of perturbation growth. Consequently, centrally injected thermal feedback actually *promotes* precipitation rather than suppressing it. This is why centrally injected thermal feedback fails so miserably to prevent condensation in the cluster-core simulation of Meece et al. (2017) that relies on pure thermal feedback, as well as many other such simulations that rely on central thermal feedback alone. With that thought in mind, we now turn our attention from local thermal instability to its implications for global thermal balance.

## 5. Evolution of Systems in Global Thermal Balance

The analyses of Sections 3 and 4 show that  $t_{\text{cool}}/t_{\text{ff}} \approx 10$  is not a general threshold for condensation. So why is the lower limit  $t_{\text{cool}}/t_{\text{ff}} \gtrsim 10$  so prevalent among multiphase systems in both numerical simulations (e.g., Gaspari et al. 2012, 2013; Sharma et al. 2012b; Li et al. 2015) and in nature (e.g., McCourt et al. 2012; Voit & Donahue 2015; Voit et al. 2015b, 2015c)? This section examines how  $\min(t_{\text{cool}}/t_{\text{ff}})$  evolves in systems constrained to be in global thermal balance. It shows that if each individual layer is restricted to remain in thermal balance, then buoyancy damping substantially slows the process of condensation in simulated galactic systems as  $\min(t_{\text{cool}}/t_{\text{ff}})$  rises, causing it to stall near  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  in simulated cluster cores. However, if the thermal balance restriction is relaxed in individual layers, then a globally balanced system evolves toward a value of  $\min(t_{\text{cool}}/t_{\text{ff}})$  that depends on the spatial distribution of heat input. In that case, global stability near  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  can be realized if heat input exceeds cooling in the inner and outer parts of the system but falls short of matching cooling at intermediate altitudes.

### 5.1. Unmixing at Constant Pressure

From a thermodynamic perspective, development of a multiphase medium under conditions of global thermal balance is essentially the opposite of thermal mixing in a system with constant total energy. For example, consider a system in which a small fraction of the gas condenses by radiating an amount of heat energy  $\delta Q$ . To maintain strict thermal balance, the rest of the gas in the system must gain the same amount of heat energy through feedback. If this condensation and feedback process proceeds at constant pressure, then the volume of the overall system does not change, but its overall entropy must *decrease*, because hotter gas gains heat energy while colder gas radiates it away. This inhomogeneous condensation process is prohibited in a closed system by the second law of thermodynamics. However, the supply of free energy that is added to balance cooling in a globally balanced system enables it to “unmix” via condensation. In this context, feedback is not only necessary to keep the system in overall thermal balance, it is also *required* to transform the original single-phase medium into a multiphase medium via thermal instability.

### 5.2. Instability in Thermally Balanced Equipotential Layers

Gravity can suppress “unmixing” by causing buoyancy damping. The simulation results discussed in Section 4 show that buoyancy damping in a gravitationally stratified medium causes

thermal instability to saturate in layers with  $\alpha_K \gg (t_{\text{ff}}/t_{\text{cool}})^2$  and can substantially delay condensation even if  $\alpha_K \ll (t_{\text{ff}}/t_{\text{cool}})^2$ . Extrapolating these findings to a cluster-core environment suggests that a thermally balanced cluster core with  $t_{\text{cool}} \lesssim 1$  Gyr and  $\min(t_{\text{cool}}/t_{\text{ff}}) \sim 10$  may be able to resist condensation for  $\gtrsim 5$  Gyr.

The idealized cluster-core simulations of Sharma et al. (2012b), which strictly maintain thermal balance in each equipotential layer, appear to corroborate this suggestion. In their fiducial simulation (run C10), the core configuration begins with central entropy  $\approx 10$  keV cm<sup>2</sup> (corresponding to  $t_{\text{cool}} \approx 200$  Myr for the adopted cooling function) and  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ . Even though the central cooling time is  $\ll 1$  Gyr, the system remains almost static for  $\approx 2$  Gyr. Then a large condensation event raises the central entropy to  $\sim 20$  keV cm<sup>2</sup>, which corresponds to  $\min(t_{\text{cool}}) \approx 600$  Myr and  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 15$ . Afterward, the system remains relatively stable for  $\sim 3$  Gyr. The initial delay time before condensation ( $\sim 10 t_{\text{cool}}$ ) is consistent with a period of buoyancy damping followed by steady exponential growth, as illustrated in Figure 3.

A similar interpretation suits run C1 from Sharma et al. (2012b), which starts with central entropy  $\approx 1$  keV cm<sup>2</sup> and  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 1$ . Condensation begins promptly in that simulation and raises the ambient specific entropy at small radii to  $\sim 10$  keV cm<sup>2</sup> in  $\lesssim 2$  Gyr, at which time  $\min(t_{\text{cool}}) \sim 200$  Myr and  $\min(t_{\text{cool}}/t_{\text{ff}}) \sim 10$ . Condensation then subsides, and the entropy and cooling-time profiles remain relatively stable for the next  $\sim 2$  Gyr, presumably for the same reason that run C10 resisted condensation for the first 2 Gyr.

In contrast, the transition to  $\min(t_{\text{cool}}/t_{\text{ff}}) \gtrsim 10$  requires only  $\sim 40$  Myr for the Gaspari et al. (2013) simulations, which also enforce thermal balance within equipotential layers. There are three reasons for the action to proceed on a shorter timescale than in Sharma et al. (2012b). First, the greater metallicity (i.e., solar values) and lower temperature compared with Sharma et al. (2012b) result in a shorter central cooling time at a given central entropy level. Second, the gravitational potential is deeper because the gravity of the central galaxy’s stars and a central supermassive black hole are both included, meaning that the freefall time at each radius is shorter. Third, and most importantly, Gaspari et al. (2013) continually drive turbulence, which counteracts the effects of buoyancy damping and raises the amplitude of the seed perturbations that thermal instability amplifies.

Taken together, the simulations of Sharma et al. (2012b) and Gaspari et al. (2013) have been viewed as support for the suggestion that systems in global thermal balance naturally stabilize near  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ , as observed in more realistic simulations of cluster cores regulated by precipitation-fed bipolar outflows (Gaspari et al. 2012; Li et al. 2015; Prasad et al. 2015; Meece et al. 2016; Yang & Reynolds 2016). However, both the analysis of Section 4 and the analytical toy model for configuration changes presented in the Appendix show that there is nothing unique about  $t_{\text{cool}}/t_{\text{ff}} \approx 10$  from the point of view of local thermal stability. Instead, the Sharma et al. (2012b) simulations remain near  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  for long time periods because increases in  $t_{\text{cool}}/t_{\text{ff}}$  stretch the time interval between condensation events to multiple gigayears, even though  $\min(t_{\text{cool}})$  remains  $\lesssim 1$  Gyr.

### 5.3. Global Thermal Balance with Distributed Heating

Relaxing the requirement that each equipotential layer remain in thermal balance while retaining the global thermal balance constraint leads to a somewhat different interpretation for the minimum value of  $t_{\text{cool}}/t_{\text{ff}}$ . When this restriction is relaxed, the entropy profile of a globally balanced system can change with time in response to the spatial distribution of heat input. In systems that can be divided into an inner isentropic zone and an outer power-law zone (see Figure 1 and the Appendix), the ambient  $t_{\text{cool}}/t_{\text{ff}}$  ratio is lowest at the boundary between those zones. Therefore, the minimum value of  $\min(t_{\text{cool}}/t_{\text{ff}})$  can depend on how heat input is distributed with radius.

To see how the structure of a globally balanced system responds to the radial distribution of feedback heating, consider how heating affects the evolution of the entropy-profile slope. The Lagrangian derivative of  $\alpha_K(r, t)$  with respect to time is

$$\begin{aligned} \frac{d\alpha_K}{dt} &= \frac{\partial}{\partial t} \left( \frac{\partial \ln K}{\partial \ln r} \right) + v \cdot \frac{\partial}{\partial r} \left( \frac{\partial \ln K}{\partial \ln r} \right) \\ &= \frac{\partial}{\partial \ln r} \left( \frac{d \ln K}{dt} \right) + \left( \frac{\partial \ln K}{\partial \ln r} \right) \frac{\partial}{\partial \ln r} \left( \frac{v}{r} \right), \end{aligned} \quad (42)$$

where  $v$  is the radial velocity of the Lagrangian shell at  $r$ . That radial velocity depends on how quickly the volume bounded by the shell is expanding or contracting due to heating or cooling. It can be expressed in terms of the volume-weighted average of the change in specific volume ( $1/\rho$ ) within  $r$ :

$$\begin{aligned} \frac{v}{r} &= \frac{d \ln r}{dt} = -\frac{1}{3} \left\langle \frac{d \ln \rho}{dt} \right\rangle, \\ \left\langle \frac{d \ln \rho}{dt} \right\rangle &\equiv \frac{3}{4\pi r^3} \int_0^r \left( \frac{d \ln \rho}{dt} \right) 4\pi r'^2 dr'. \end{aligned} \quad (43)$$

The Lagrangian derivative of  $\alpha_K$  can therefore be written as

$$\begin{aligned} \frac{d\alpha_K}{dt} &= \frac{\partial}{\partial \ln r} \left( \frac{d \ln K}{dt} \right) \\ &\quad + \alpha_K \frac{d \ln \rho}{dt} - \alpha_K \left\langle \frac{d \ln \rho}{dt} \right\rangle. \end{aligned} \quad (44)$$

In this equation, the first term represents the changes in  $\alpha_K$  that arise from the radial gradient of the net heating rate. The second term represents changes that arise from the compression or rarefaction of the local entropy gradient. The third term represents changes that result from the expansion or contraction in radius of the Lagrangian shell under consideration.

We are interested in knowing whether  $\alpha_K$  increases or decreases with time in response to the distribution of heating and cooling with radius. If  $\alpha_K = 0$ , then the answer is relatively simple. The entropy slope steepens in regions where the net heating rate increases with radius and declines in regions where it decreases with radius. However, the answer becomes more complicated when  $\alpha_K \neq 0$  and the other two terms come into play. To better understand how the additional terms affect the evolution of the system as a whole, we explore three specific radial distributions of heating and cooling in the following three subsections, and in a fourth subsection (Section 5.3.4) we bring those explorations to bear on interpreting  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ .

#### 5.3.1. Two Regions: Central Heating, Distributed Cooling

In the simplest case to consider, heating dominates cooling at the center and cooling dominates heating in the outer regions. Central heating is the mode of AGN feedback most often implemented in simulations of galaxy evolution. Examples in the literature are too numerous to mention here, but their generic behavior is captured by the simulation from Meece et al. (2017) in which the kinetic energy of the feedback response is set to zero: condensation proceeds until feedback raises the central cooling time of the ambient medium to several gigayears by creating a large isentropic zone. In such simulations, the amounts of condensed gas and star formation become unacceptably large unless the rise in the central cooling time is rapid, which requires a high feedback efficiency.

This outcome occurs because central heating in a globally balanced system corresponds to a radial gradient of  $d \ln K / dt$  that is negative, which reduces  $\alpha_K$  with time. The other two terms in Equation (44) consequently decline in importance. The result is a central region that is effectively isentropic and also has access to a flow of free energy. These two features are exactly what is needed for condensation to produce a multiphase structure in a medium with  $t_{\text{cool}} \gg t_{\text{ff}}$ . In such a globally balanced system, central heating continues to increase the central entropy and can promote condensation until the minimum cooling time in the ambient medium becomes comparable to the age of the universe. As this happens, the system's ambient medium inevitably evolves toward  $\min(t_{\text{cool}}/t_{\text{ff}}) \gg 10$ .

#### 5.3.2. Two Regions: Distributed Heating, Central Cooling

The opposite case, in which heating exceeds cooling at large radii and cooling exceeds heating at small radii, evolves in the opposite sense. In such a system, the radial gradient of  $d \ln K / dt$  is positive and tends to increase the entropy slope  $\alpha_K$ . However, increases in  $\alpha_K$  make the second and third terms in Equation (44) more important. The effects of those other two terms are clearest if heating is insignificant. Then, cooling causes  $\rho$  to increase with time in the central cooling-dominated zone on a timescale  $\sim t_{\text{cool}}$ . It also causes the third term to increase more than the second term, because the average cooling time within a sphere of radius  $r$  is generally shorter than the average cooling time at  $r$  in a system with a positive entropy gradient. As a result, the third term limits the entropy gradient to be no greater than the value at which the third term offsets the other two.

One can estimate the limiting value of  $\alpha_K$  by assuming that the system remains approximately isothermal as it loses entropy through radiative cooling, which is usually valid for gas confined by a nearly isothermal gravitational potential, as long as  $t_{\text{cool}} \gg t_{\text{ff}}$ . With that assumption, we obtain

$$\frac{d\alpha_K}{dt} \approx \frac{3\alpha_K}{t_{\text{cool}}} \left( \frac{1 - \alpha_K}{2 - \alpha_K} \right) \quad (45)$$

if the volume average in Equation (43) is performed assuming  $\alpha_K$  is constant with radius. This estimate shows that cooling produces an inflow that approaches a limiting entropy slope of  $\alpha_K \approx 1$ , corresponding approximately to  $\rho \propto r^{-3/2}$  in gas that remains nearly isothermal, and is in alignment with the self-similar cooling-flow solution derived by Bertschinger (1989). In a more realistic cluster-core potential, deviations from isothermality can slightly steepen the entropy slope of a

steadily cooling system to  $\alpha_K \approx 1.2$ , but it always remains close to unity (e.g., Voit 2011). Formally, Equation (45) also suggests that a cooling region initialized without an entropy gradient remains isentropic, because the cooling time in the isothermal approximation we have adopted is then independent of radius. In that limit, the effects of temperature gradients and gradients in heat input can no longer be ignored. However, even if there is some growth of the entropy gradient, it proceeds on a timescale comparable to or longer than the cooling time.

We conclude that if cooling dominates heating at small radii, then entropy losses and the resulting compression cause the isentropic region to shrink in radius as the center of the system evolves toward a state with  $\alpha_K \approx 1$ . Condensation then happens only at small radii, where  $t_{\text{cool}} \lesssim t_{\text{ff}}$ . The result is to focus condensation and mass deposition onto the neighborhood of the black hole (e.g., Li & Bryan 2012). This outcome is highly favorable for the coupling of black-hole accretion to the global evolution of the surrounding medium. At larger radii, where heating exceeds cooling, the response of  $\alpha_K$  depends on the radial gradient of heat input, which depends in turn on the feedback delivery mechanism. However, the resulting changes in  $\alpha_K$  typically proceed on a much longer timescale than the cooling-induced changes to the inner profile, because the heat input required for global balance is spread over a much larger volume, in gas with much lower density and a much longer cooling time.

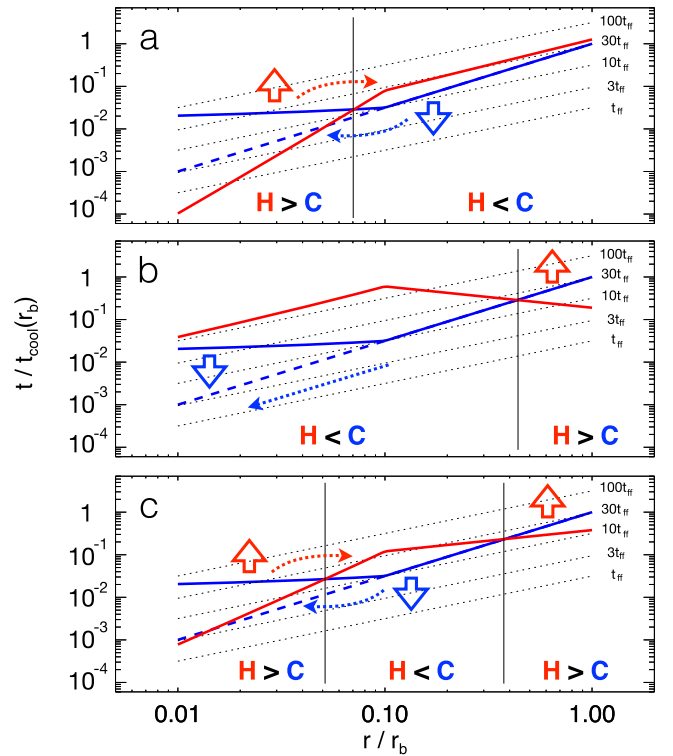
### 5.3.3. Three Regions: Heating–Cooling–Heating

Neither of the simple two-region cases yields a satisfactory explanation for how feedback in a cluster core can maintain a central cooling time  $\lesssim 1$  Gyr along with  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  for periods of several gigayears. We must therefore take one more small step toward complexity and consider what happens if there are *three* nested regions in which the sign of  $d \ln K/dt$  alternates, with heating exceeding cooling at the smallest and largest distances from the black hole, while cooling exceeds heating at intermediate radii.

Figure 4 schematically shows how the response to feedback of this three-region system contrasts with the two-region cases we have just discussed. The radial decline in  $d \ln K/dt$  across the boundary between the innermost two regions causes convection that drives the inner part of the system toward an isentropic configuration that is prone to condensation. Meanwhile, the rise in  $d \ln K/dt$  across across the boundary between the outermost two regions maintains a large-scale entropy gradient that suppresses thermal instability through buoyancy damping. The natural outcome of such a feedback response is a long-lasting configuration with an inner isentropic zone and an outer power-law zone (see Figure 1 again). In this configuration, the central production of cold clouds can fuel a feedback response that prevents condensation in the outer zone. The Appendix presents a toy model that makes this conceptual argument more quantitative.

### 5.3.4. Interpretation of $t_{\text{cool}}/t_{\text{ff}} \approx 10$

This chain of reasoning leads to an interpretation of  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  with several facets. Section 5.3.3 outlines the radial distribution of feedback heat input needed to keep  $\min(t_{\text{cool}}/t_{\text{ff}})$  steady but does not explain how this balance remains stable. If the system is to remain near a particular balance point for time periods much longer than the cooling



**Figure 4.** Schematic illustrations of configuration changes in globally balanced systems. In each panel, cooling equals heating within a boundary radius  $r_b$ , and solid blue lines show identical cooling-time profiles derived from the toy model in the Appendix with  $r_0 = 0.1r_b$  and  $\alpha_K = 1.0$ . The blue dashed line shows the cooling-time profile that would result if the power-law entropy profile extended to small radii. Dotted black lines show the slope of  $t_{\text{ff}}(r)$ . The solid red line in each panel shows the local heating timescale  $t_{\text{heat}}$ , corresponding to heat input that has a power-law dependence on radius. It is normalized so that the total heat input  $\dot{E}(r_b)$  equals the total luminosity generated by radiative cooling within  $r_b$ . Panel (a) shows  $t_{\text{heat}}$  for a heat-input distribution ( $\dot{E} \propto r^{0.3}$ ) for which the heating per unit volume exceeds cooling at small radii, causing the central entropy to rise with time, and falls short of cooling at large radii (see Section 5.3.1). Panel (b) shows a heat-input distribution ( $\dot{E} \propto r^2$ ) that fails to match cooling at small radii but exceeds it at large radii, causing the center to approach a steady cooling flow with  $\alpha_K = 1$  (see Section 5.3.2). Panel (c) shows a heat-input distribution ( $\dot{E} \propto r$ ) in which heating exceeds cooling at small radii, falls short at intermediate radii, and exceeds it again at large radii (see Section 5.3.3).

time, then fluctuations in  $\min(t_{\text{cool}}/t_{\text{ff}})$  need to produce compensating responses. Here we suggest that volatility places a lower bound on  $\min(t_{\text{cool}}/t_{\text{ff}})$  and that inefficient central heating places an upper limit on  $\min(t_{\text{cool}}/t_{\text{ff}})$ .

First, consider how such a system responds to a decline in  $\min(t_{\text{cool}}/t_{\text{ff}})$ . As the value of this ratio drops below  $\approx 10$ , the system becomes increasingly vulnerable to feedback outbursts fueled by condensation. Outbursts that uplift gas at speeds close to the circular velocity of the gravitational potential promote condensation of gas that initially has  $t_{\text{cool}}/t_{\text{ff}} \lesssim 10$  (see Section 2). Condensation then removes the lowest entropy gas from the ambient medium, leaving behind higher entropy gas with a longer cooling time. Promoting condensation with uplift becomes progressively more difficult as this process continues, because  $\min(t_{\text{cool}}/t_{\text{ff}})$  rises and the uplift of large amounts of gas at speeds substantially exceeding the circular velocity is unlikely. Therefore, this part of the feedback loop restores the system to  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ .

Next, consider the response to a rise in  $\min(t_{\text{cool}}/t_{\text{ff}})$  driven by central heating. Condensation can continue in the isentropic



zone but is spread over a larger region, making it more difficult for condensed gas clouds to shed enough angular momentum to accrete onto the black hole. In nature, this reduction in the fuel supply might be all that is needed to diminish central heating and restore  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ . In simulations, however, spatial resolution is more limited and subgrid feedback algorithms have much less of an angular momentum problem to overcome than in nature. Nevertheless, simulations with precipitation-fueled bipolar outflows also manage to self-regulate.

We suggest that in these simulations, and maybe also in real galactic systems, the restoring response to a rise in  $\min(t_{\text{cool}}/t_{\text{ff}})$  comes about because the central region less efficiently taps the feedback energy that is transported through it. The [Appendix](#) presents this suggestion more quantitatively in the context of our toy model. Global regulation requires the total amount of feedback heating to be at least as great as losses to radiative cooling, and we have shown that much of that feedback energy must pass through the isentropic zone without thermalizing there. A rise in  $\min(t_{\text{cool}}/t_{\text{ff}})$  corresponds to a decrease in the density of the isentropic zone, and if  $\alpha_K > 2/3$ , a decrease also in its column density. It is therefore plausible that reducing the density of the isentropic zone reduces the fraction  $f_0$  of feedback energy deposited there (see, for example, the recent simulations by Tchekhovskoy & Bromberg 2016).

To appreciate the consequences of an inverse relationship between  $f_0$  and  $K_0$ , consider first what happens if  $f_0$  remains constant. Suppose the system begins in a volatile state, with  $\min(t_{\text{cool}}/t_{\text{ff}}) < 10$ , so that condensation feeds the black-hole engine and boosts the output rate  $\dot{E}$  of feedback power. Increasing condensation then boosts  $\dot{E}$  until  $f_0 \dot{E}$  balances the radiative luminosity of an isentropic zone with  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$ . As  $\min(t_{\text{cool}}/t_{\text{ff}})$  rises beyond that point, condensation should diminish, but a large cold gas reservoir may already have accumulated, which can continue to fuel the outflow. Upward fluctuations in  $\dot{E}$  fueled by that cold gas reservoir can in principle cause  $K_0$  and  $\min(t_{\text{cool}}/t_{\text{ff}})$  to grow further. However, if  $df_0/dK_0 < 0$ , then increases in  $K_0$  diminish the heat supply to the isentropic zone and allow cooling to return the system to its original value of  $K_0$ . Simultaneous long-term regulation of both the isentropic zone and the power-law zone can therefore be achieved if radiative cooling of the power-law zone is  $\lesssim (1 - f_0)\dot{E}$ . A precise balance between heating and cooling is not necessary in the power-law zone because the timescale for configuration changes at large radii is long, and feedback regulation of the isentropic zone will quickly respond to the slow feedback-driven expansion of the power-law zone.

## 6. Illustrations from Simulations

Our objective until this point in the paper has been to understand the physical principles that govern precipitation-regulated feedback, both in nature and in simulations of galaxy evolution. Now it is time to apply those principles to interpret what happens in such simulations, using an idealized cluster-core simulation from Li et al. (2015) as a paradigmatic example. In that simulation, condensation produces cold gas clouds that feed a central black hole, which responds by producing a bipolar outflow. This mode of feedback regulates the system so that  $\min(t_{\text{cool}}/t_{\text{ff}})$  remains in the range  $5 \lesssim \min(t_{\text{cool}}/t_{\text{ff}}) \lesssim 20$  for long time periods while the outflows are active. However, star formation eventually consumes the cold gas that the black hole needs for fuel. When that cold gas reservoir becomes depleted,

the outflow shuts down, cooling is now unopposed, and  $\min(t_{\text{cool}}/t_{\text{ff}})$  declines until condensation renews the supply of cold gas. The outflow then resumes, restoring regulation at  $5 \lesssim \min(t_{\text{cool}}/t_{\text{ff}}) \lesssim 20$ .

The Li et al. (2015) simulation is far from a complete treatment of all of the relevant physics. Perhaps most importantly, it is not subject to cosmological structure formation, and is therefore artificially symmetric and undisturbed. It is quite likely that the cold gas reservoirs in real systems of this type are never fully depleted, in which case there would be no time periods during which the central cooling time would need to decline below  $\min(t_{\text{cool}}/t_{\text{ff}}) \sim 5$  in order to re-initiate condensation. However, those episodes of cold gas depletion in the Li et al. (2015) simulation turn out to be extremely useful for illustrating the respective roles of unopposed cooling, uplift-driven condensation, buoyancy damping, and feedback regulation, as this section will demonstrate.

### 6.1. Simulated Feedback Cycles

Figure 5 shows the feedback cycles in the Li et al. (2015) simulation. At the beginning, the gas is homogeneous at each radius and remains homogeneous as it cools, until condensation triggers strong feedback at  $t = 0.27$  Gyr. The amount of gas that is present within the central 20 kpc prior to the feedback outburst can be estimated by noting that  $\alpha_K \approx 1$  and  $K \approx 10 \text{ keV cm}^2$  at 20 kpc, giving a gas-mass estimate  $\sim 3 \times 10^{11} M_\odot$  for a gas temperature  $\sim 3 \text{ keV}$ . Approximately  $\sim 4 \times 10^9 M_\odot$  of this gas condenses within  $\sim 200 \text{ Myr}$  and triggers feedback that leaves the system in a steady feedback-regulated state with  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  for the next  $\sim 1 \text{ Gyr}$ , until star formation uses up all of the cold gas.

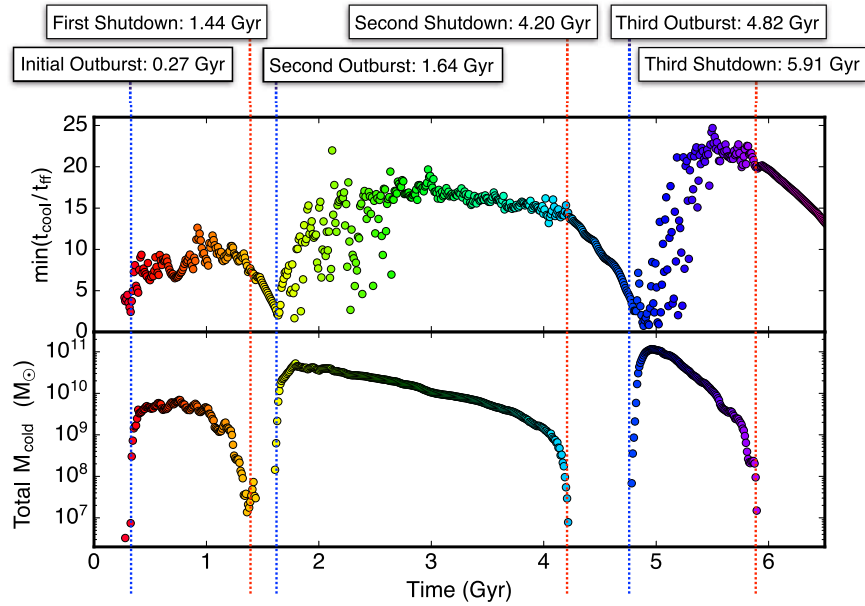
Two more outburst and shutdown cycles follow the first one, and each cycle proceeds similarly. Without heat input, the core cools homogeneously and develops a power-law central entropy gradient that focuses condensation on the black hole. When the black hole erupts, uplift stimulates rapid condensation and spreads condensing gas over a much larger volume. After each eruption, the system is left in a long-lasting, well-regulated state with a flat central entropy gradient.

### 6.2. Anatomy of the Specific Entropy Distribution

In order to understand why these cycles unfold as they do, it is important to look closely at the distribution of specific entropy  $K$  at each radius. Figure 6 shows the distribution of  $K$  as a function of  $r$  at  $t = 1 \text{ Gyr}$ , which is in the midst of the first feedback cycle. Broadly speaking, the entropy profile is approximately isentropic in the inner regions and approximately a power law with  $\alpha_K \approx 1$  in the outer regions. The transition from  $\alpha_K \ll 1$  to  $\alpha_K \approx 1$  is at the radius where  $t_{\text{cool}}/t_{\text{ff}} \approx 10$ . However, gas at any given radius can span a wide range in  $K$ .

The actual dispersion in specific entropy is not quite as broad as the figure's logarithmic heat map makes it seem. Lines in the inset figure show the median entropy at each radius (solid line) and the 20%–80% percentile range (dashed lines). Despite the wide range of extremes in entropy, the dispersion is actually quite narrow at large radii. The extremes are associated with the high-entropy outflow and the lower-entropy gas uplifted by it (Li & Bryan 2014b), but buoyancy damping keeps most of the





**Figure 5.** Outburst history of precipitation-regulated AGN feedback in a simulation from Li et al. (2015). The top panel shows the minimum value of  $t_{\text{cool}}/t_{\text{ff}}$  in the ambient medium at each moment in time, and the bottom panel shows the amount of cold gas available for accretion. Cold gas rapidly accumulates at the onset of each outburst, but the feedback response regulates the condensation rate as  $\min(t_{\text{cool}}/t_{\text{ff}})$  rises. Feedback continues to regulate condensation while cold gas is plentiful, but shuts down as star formation uses up the cold gas supply. The value of  $\min(t_{\text{cool}}/t_{\text{ff}})$  subsequently declines until cooling triggers another feedback outburst.

gas nearly homogeneous at radii where the entropy gradient is significant.

In the center, where the entropy gradient vanishes, buoyancy damping is less successful at suppressing entropy perturbations, and thermal instability is able to produce condensation (Section 3). Gas blobs below the median entropy in this zone are declining in specific entropy and will soon join the condensed component. Outside of this thermally unstable zone, condensation is scarcer. At the epoch shown, there are two isolated patches of multiphase gas at larger radii, roughly 30 and 50 kpc from the center. Buoyancy damping cannot suppress condensation in these patches because they have been uplifted by the outflow (Section 2) and are on nearly ballistic trajectories.

### 6.3. Initial Outburst, Uplift, and Condensation

The six panels of Figure 7 show how the entropy distribution evolves during the first outburst of feedback. Before the outburst erupts at  $t = 0.27$  Gyr, the  $K$  distribution at each  $r$  is still nearly homogeneous for two reasons: (1) pure cooling without compensating heating does not lead to phase separation (Section 5.1), and (2) even if there were some entropy perturbations, they would be damped by buoyancy (Section 3.4). The next panel ( $t = 0.37$  Gyr) shows the system 100 Myr after the beginning of the outburst, and the leading edge of the high-entropy outflow (at  $\sim 40$  kpc) is well defined. Gas at larger radii is still undisturbed, and the median entropy still follows an approximate power law to  $\ll 10$  kpc. However, there are significant amounts of condensing low-entropy gas at  $\sim 20$  kpc. That cannot be gas condensing out of the ambient medium at 20 kpc, because the median cooling time there was  $\gtrsim 200$  Myr at the time the outburst began. Instead, it is low-entropy gas that was transported outward from smaller radii by the outflow, and uplift is allowing it to condense.

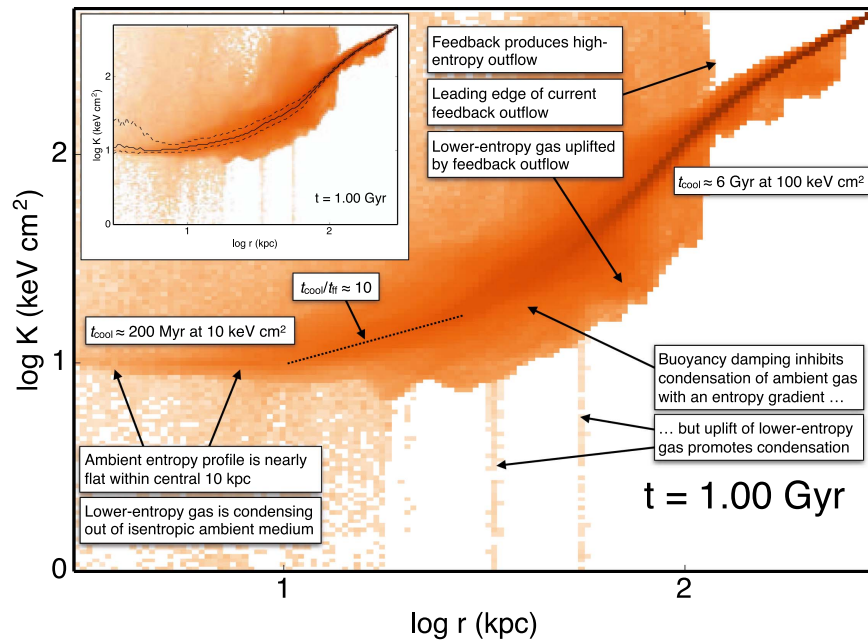
After another 100 Myr (at  $t = 0.47$  Gyr), an isentropic central region has started to develop at  $K_0 \approx 10 \text{ keV cm}^2$ ,

where the cooling time is  $\approx 200$  Myr. In order for an isentropic central region to form, the bulk of the lower-entropy gas must be eliminated through either condensation, heating, or mixing. Figure 5 shows that less than  $\sim 10^{10} M_{\odot}$  condenses during this event, implying that the rest of the  $\sim 3 \times 10^{11} M_{\odot}$  initially within the central  $\lesssim 20$  kpc is either heated or mixed into gas of greater entropy.

The remaining three frames, spanning the next 300 Myr, show that the system becomes well-regulated after arriving at a state with an isentropic core and a power-law outer region (see Section 5.3.3). Condensation continues in the isentropic core at  $\lesssim 20$  kpc but proceeds only in isolated patches beyond where the entropy gradient steepens, and those patches are moving on approximately ballistic trajectories. The outflow continues to populate the extremes of the entropy distribution. However, buoyancy damping causes those extremes to evolve toward the median, as shown by the lines in the lower-right panel, which mark the median and 20%–80% percentile range, as in the Figure 6 inset.

### 6.4. Second Outburst

Figure 8 shows six more frames, this time spanning the period from the end of the first feedback episode to a time approximately 400 Myr after the beginning of the second. At  $t = 1.44$  Gyr, the cold fuel has just run out, causing the AGN to shut down, but the residual outflow has not finished propagating. Some of the gas is far from the median entropy level, and the central cooling time is  $\approx 200$  Myr. During the next  $\sim 100$  Myr, buoyancy damping reduces the extremes in entropy at each radius in the power-law zone. Meanwhile, the central cooling time begins to decline, but it does so almost uniformly, because there is no source of free energy to cause phase separation (Section 5.1). Condensation finally starts to happen as the simulation approaches  $t = 1.59$  Gyr, and it develops most quickly at the boundary between the isentropic



**Figure 6.** Anatomy of the specific entropy distribution at  $t = 1$  Gyr in the simulation from Figure 5. Shades of orange show the relative amounts of gas in the uncondensed component with specific entropy  $K$  at radius  $r$ . The median entropy value (shown with a solid black line in the inset figure) is nearly constant with radius inside of 10 kpc and steadily rises at larger radii. The dispersion in entropy (dashed lines in the inset figure show the 20%–80% percentile range) declines with increasing radius. Buoyancy damping largely suppresses condensation where the entropy gradient is significant but cannot suppress condensation where the entropy gradient is shallow. The transition between these two regions is where  $t_{\text{cool}}/t_{\text{ff}} \approx 10$  (shown with a dotted line).

and power-law zones, because that is where  $t_{\text{cool}}/t_{\text{ff}}$  is smallest (Section 5.3).

Once condensation begins, cooling gas quickly converges toward the center and triggers a second feedback outburst at  $t = 1.64$  Gyr. At this moment, the median entropy distribution is nearly a power law with  $\alpha_K \approx 1.3$  from 100 kpc to within 10 kpc. This median slope is almost identical to the triggering slope at  $t = 0.37$  Gyr in Figure 7, but the pre-triggering entropy distribution at  $\lesssim 10$  kpc is much less homogeneous because of the condensation that started happening 50 Myr earlier.

A major consequence of this pre-existing inhomogeneity is that uplift induces condensation far more effectively when the outflow begins, yielding  $\sim 5 \times 10^{10} M_{\odot}$  of cold gas during the next 200 Myr (see Figure 5). The gas mass within  $\sim 20$  kpc of the center as condensation begins is similar to the  $\sim 3 \times 10^{11} M_{\odot}$  that was there during the first outburst, but 10 times more of it condenses (see Figure 5). The rest must be either heated or mixed to  $\gtrsim 20$  keV  $\text{cm}^2$  because the central region becomes nearly isentropic at this level by  $t = 2.05$  Gyr, as shown in the final panel of Figure 8.

After the second outburst produces another reservoir of cold gas fuel, the system enters a long period of steady self-regulation. The median entropy profile within  $\sim 20$  kpc stays flat at  $K_0 \approx 20$  keV  $\text{cm}^2$ , so that  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 15$ , apart from some stochastic excursions as the black-hole energy output fluctuates. However, the cold fuel eventually runs out after sustaining  $\approx 2.5$  Gyr of activity, setting the stage for the third outburst.

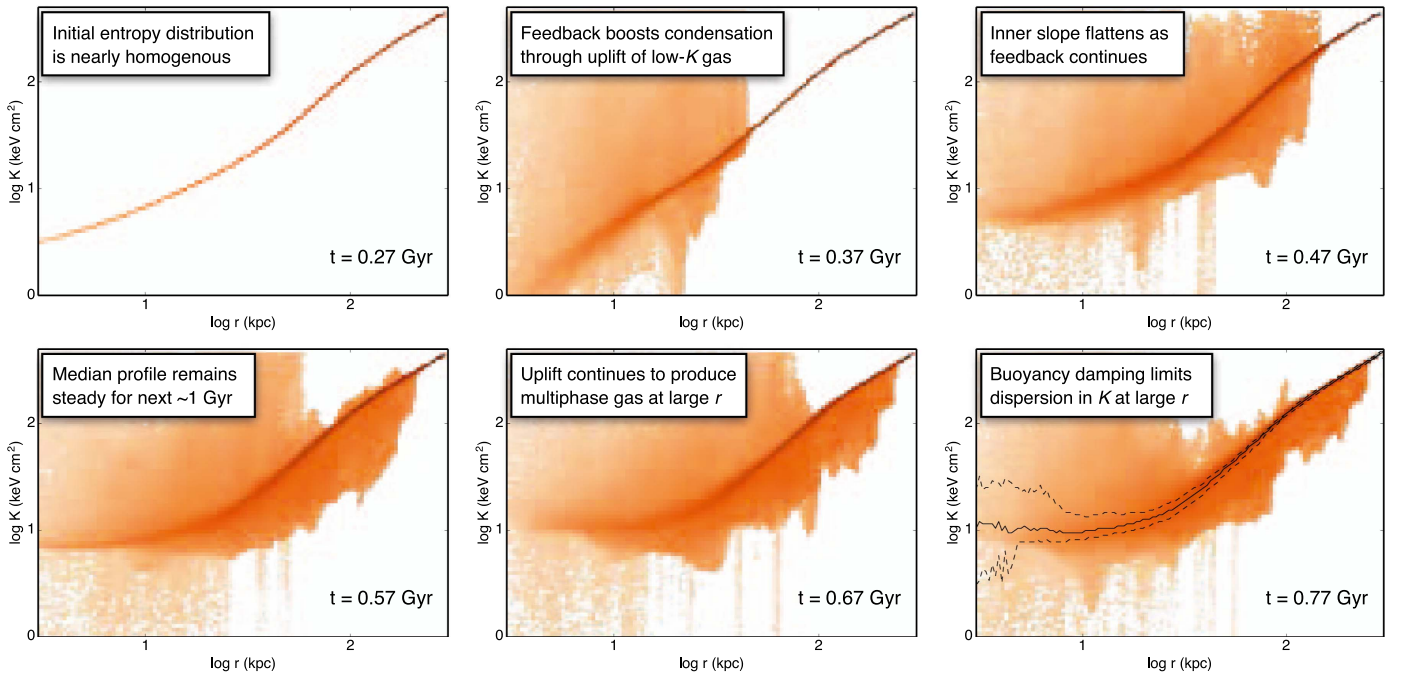
### 6.5. Third Outburst

Figure 9 concentrates on the events leading up to the third outburst, which require more time to unfold because the central

cooling time of the system is  $\approx 600$  Myr when the cold fuel runs out. Complete depletion of cold fuel could be a consequence of the idealizations in this simulation that never happen in real cluster cores. We mention this possibility because very few real clusters have central cooling times  $\lesssim 1$  Gyr and no multiphase gas (Voit & Donahue 2015). However, the period of quiescence before the third outburst in this simulation beautifully illustrates the physics of AGN triggering via condensation, and similar processes are likely to operate in less idealized systems.

The events that lead to triggering are almost identical to those in Figure 8, but here the progression of thermal instability is easier to see. Buoyancy damping again reduces entropy perturbations in the power-law zone, and the central regions again cool uniformly because there is no source of free energy to promote phase separation. However, buoyancy damping is least effective at the junction between the isentropic and power-law zones, where  $t_{\text{cool}}/t_{\text{ff}}$  is smallest and is also declining with time. That junction moves inward as the isentropic zone loses thermal energy and becomes more highly compressed. As a result, the median entropy profile approaches a pure power law in radius, while thermal instability progresses toward condensation at the outer edge of the isentropic zone. Condensation then occurs between  $t = 4.77$  Gyr and  $t = 4.82$  Gyr. It happens first at  $r \approx 20$  kpc, but soon the entire region within 20 kpc contains multiphase gas.

The feedback outflow then sweeps through the condensing region, and 200 Myr later ( $t = 5.00$  Gyr) a new isentropic zone has begun to form. This triggering period produces even more condensation than the previous two outbursts, yielding  $\approx 10^{11} M_{\odot}$  of cold gas. With this fuel source, the feedback outflow stabilizes the system's global configuration at  $K_0 \approx 30$  keV  $\text{cm}^2$  and  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 20$  for another  $\sim 1$  Gyr.



**Figure 7.** Evolution of the specific entropy distribution in the simulation from Figure 5 during triggering of the initial episode of feedback. The outburst of kinetic energy from the center begins when the central entropy becomes low enough for condensation. At that moment, the median entropy distribution is nearly a power law with  $K \propto r$ , and  $\min(t_{\text{cool}}/t_{\text{ff}}) \sim 2$ . Shortly thereafter, the feedback outburst flattens the median entropy profile within  $\sim 10$  kpc and stabilizes at  $\min(t_{\text{cool}}/t_{\text{ff}}) \sim 10$  for the next  $\sim 1$  Gyr.

### 6.6. Not-quite-ballistic Multiphase Regions

During the long periods of relatively steady feedback regulation between these triggering events, condensation happens mostly within the isentropic zone, where it is not inhibited by buoyancy damping. However, multiphase regions are sometimes observed far outside the isentropic zone, at radii where  $\alpha_K \approx 1$  and  $t_{\text{cool}}/t_{\text{ff}} > 10$  in the ambient medium, implying that buoyancy damping should be suppressing condensation. In plots of the  $K(r)$  distribution, those multiphase regions show up as isolated vertical stripes extending far below the median entropy level.

The origin of this multiphase gas in the power-law zone is not the growth of linear thermal instability but rather the uplift. These gas patches originate in the isentropic zone, are uplifted by the outflow, and then follow trajectories that are approximately ballistic. Black arrows in Figure 10 show how one such region rises and falls. It can be seen rising out of the isentropic zone, reaching  $\approx 30$  kpc at  $t = 5.10$  Gyr. Between  $t = 5.20$  Gyr and  $t = 5.30$  Gyr, it reaches an apex at  $\approx 80$ – $90$  kpc and falls back through  $\approx 30$  kpc at  $t \approx 5.42$  Gyr. If the motion of the gas blob were purely ballistic, this excursion in radius would take  $\approx 240$  Myr and would require the projectile to rise and fall through  $r = 30$  kpc, where the circular velocity is  $\approx 550$  km s<sup>−1</sup>, at a speed  $\gtrsim 900$  km s<sup>−1</sup>. Instead, the peak speed during both the upward and downward portions of the trajectory is  $\sim 500$  km s<sup>−1</sup>, while most of the time is spent at smaller speeds near the trajectory’s turning point. These deviations from ballistic motion imply that the outflow is still providing impetus to the multiphase gas blob during the upward stage, and that drag is limiting the infall speed to  $\approx 550$  km s<sup>−1</sup> during the downward stage.

It can be argued that these multiphase gas clumps would be deviating more strongly from ballistic trajectories in a more

realistic and highly resolved simulation because of hydrodynamic shredding and drag (e.g., Nulsen 1986; Pizzolato & Soker 2005). In this simulation, however, drag is clearly not necessary to promote the development of multiphase gas at large altitudes, because the gas is multiphase on both the upward and downward portions of the trajectory. High-resolution simulations specifically designed to track the hydrodynamical evolution (and perhaps even magnetohydrodynamical evolution) of uplifted gas will be needed to draw more certain conclusions about the roles of uplift and drag in promoting condensation and multiphase structure within the power-law zone.

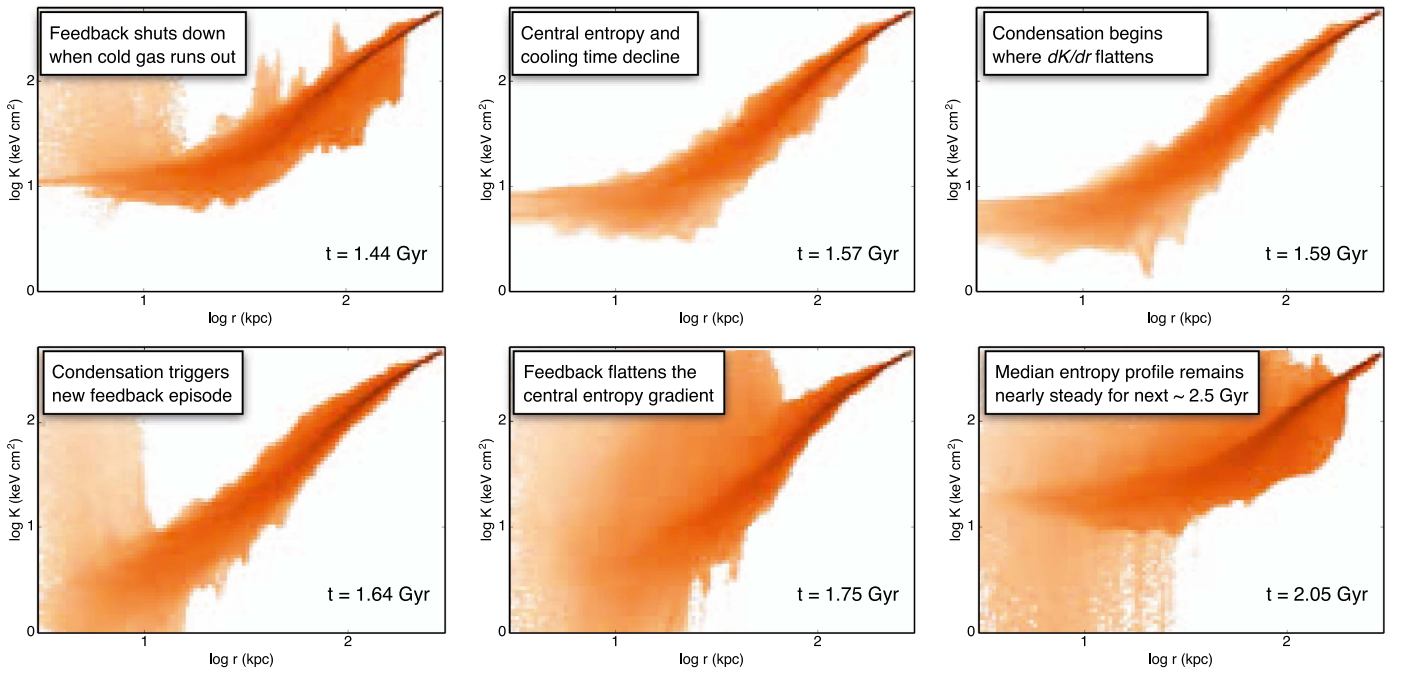
## 7. Loose Ends

Knowledgeable readers have surely noticed by now that some important physical effects certain to affect the thermal stability of circumgalactic gas are absent from our analysis. We hope that those readers are also wise enough to recognize when a paper has already gotten a little too long. Very briefly, we will point out some of the missing physics and speculate about its consequences.

### 7.1. Angular Momentum

The rotation of the system, which we have ignored, has at least two major consequences: (1) locally it can reduce the effective gravitational potential and therefore the stratification on which buoyancy damping depends, and (2) globally it can inhibit the accretion of condensed gas onto the central black hole. Consequence (1) can promote condensation in regions where it otherwise would not happen. Consequence (2) has the potential to decouple condensation from black-hole fueling. However, Gaspari et al. (2015) have explored how rotation affects condensation and chaotic cold accretion using high-resolution simulations of thermally balanced media and show





**Figure 8.** Evolution of the specific entropy distribution in the simulation from Figure 5 during the triggering of the second episode of feedback. When the supply of cold gas for accretion runs out, the central cooling time is  $\approx 200$  Myr. The central gas then cools uniformly, because feedback is not supplying the flow of free energy needed to promote phase separation. Meanwhile, buoyancy damping is reducing the entropy perturbations at large radii. Condensation therefore begins at the outer edge of the isentropic region, where  $t_{\text{cool}}/t_{\text{ff}}$  is smallest and buoyancy damping cannot operate.

that precipitation can still deliver plenty of cold fuel to the vicinity of the black hole. The key is for turbulence and cloud–cloud collisions to continually populate the low end of the distribution of specific angular momentum among cold clouds. This is more easily done when the velocity dispersion of turbulence exceeds the mean rotational speed.

It is tempting to speculate about the implications of consequence (1) for the replenishment of condensed gas in galactic disks. In a rotating frame of reference, the effective gravitational potential of a spherically symmetric halo with a stellar disk becomes more like that in the plane-parallel simulations of Meece et al. (2015). If the halo contains hydrostatic ambient gas with  $t_{\text{cool}}/t_{\text{ff}} \gg 1$ , then buoyancy damping can inhibit condensation everywhere except the midplane. Condensation of hot ambient halo gas may then replenish the disk’s reservoir of cold gas without any obvious evidence for infall in the form of cool clouds at larger altitudes. For a more formal analysis of thermal instability in a rotating circumgalactic medium, see Nipoti & Posti (2014).

### 7.2. Inhomogeneous Enrichment

If cooling of the circumgalactic medium depends strongly on metallicity, as it certainly will in systems with virial temperatures  $\lesssim 1$  keV, then inhomogeneities in the distribution of metals will promote thermal instability and condensation of the more highly enriched gas (e.g., Fraternali & Binney 2008; Marinacci et al. 2010; Oppenheimer et al. 2010). An over-enriched gas blob with the same  $K$  and  $P$  as the medium surrounding it will cool faster, causing it to sink, and it will not respond as strongly to buoyancy damping. This coupling among enrichment, entropy, and buoyancy has complex consequences for condensation, meaning that metal transport and mixing in circumgalactic gas will be important to simulate

properly in order to clarify the role of inhomogeneous enrichment.

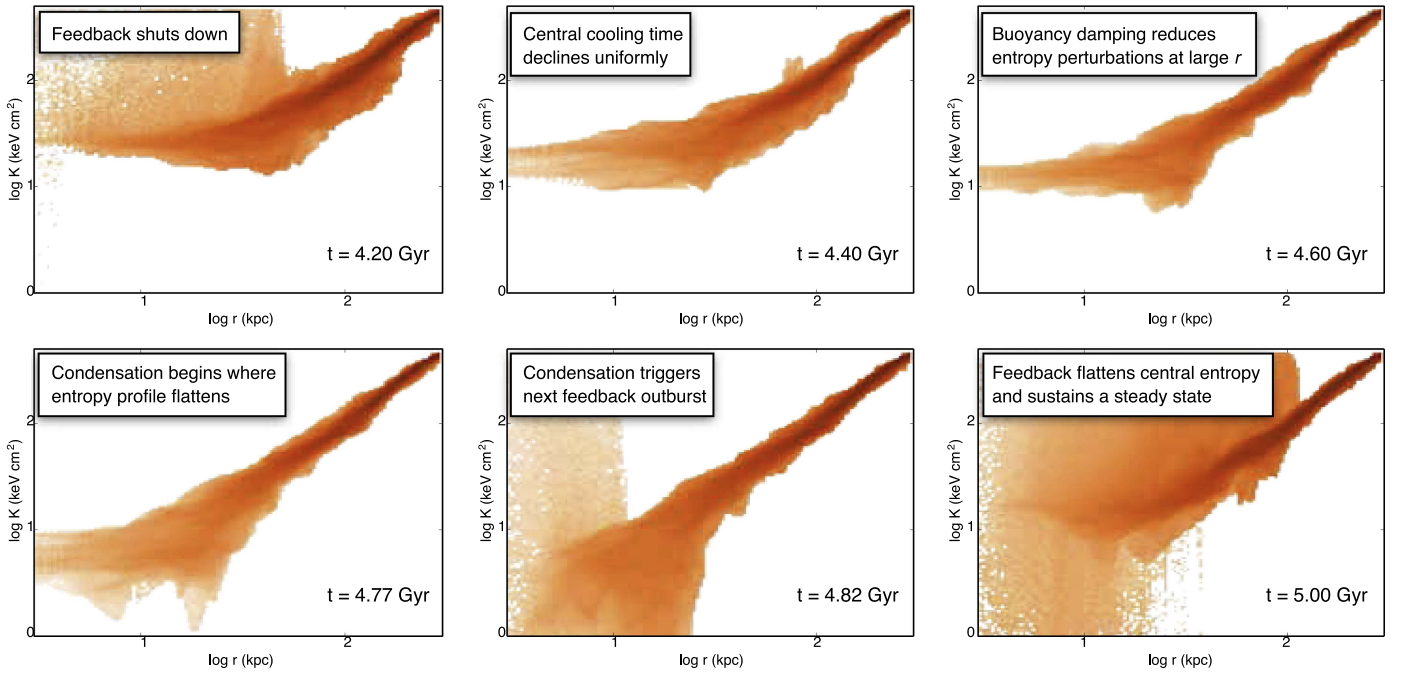
Perhaps counterintuitively, inhomogeneities in enrichment could make it easier for lower-metallicity gas to condense within the galactic disk itself. This may happen if high-metallicity gas more easily condenses into cold, ballistic clouds at greater altitudes, while lower metallicity gas settles more homogeneously into the disk and condenses there instead. Again, high-resolution targeted simulations including metal transport will be necessary to resolve this question.

### 7.3. Magnetic Fields

Our entire analysis has implicitly assumed the existence of tangled magnetic fields because we have ignored electron thermal conduction (see Soker 2010b for a discussion of the role of magnetic fields in cooling flows). In reality, conduction is likely to be somewhat anisotropic without being completely suppressed, which will naturally cause condensation to be filamentary (e.g., Sharma et al. 2010; McCourt et al. 2012; Wagh et al. 2014). Also, larger-scale conduction may help lift cluster cores out of precipitation-regulated states, if the central entropy rises enough for radial heat transport to exceed radiative cooling in the core (Guo et al. 2008; Voit et al. 2008, 2015b; Parrish et al. 2010). If that happens, then feedback is no longer necessary to prevent catastrophic cooling.

Another potential consequence of magnetic fields is enhanced drag, which can promote condensation (Section 3.3.1). This has long been recognized as a possibility (e.g., Nulsen 1986), and recent ALMA observations of remarkably small, sub-Keplerian velocity dispersions among the bulk of the molecular gas clouds in galaxy-cluster cores (McNamara et al. 2014; Russell et al. 2016) have rekindled interest in it (McNamara et al. 2016). Obviously, magnetohydrodynamical simulations are required to





**Figure 9.** Evolution of the specific entropy distribution in the simulation from Figure 5 during the triggering of the third episode of feedback. This triggering episode proceeds similarly to the second one in Figure 8, except that the quiescent period is longer because the central cooling time is  $\approx 0.6$  Gyr when the supply of cold gas runs out.

clarify whether magnetic tension is strong enough to suspend molecular clouds within a medium that is many orders of magnitude less dense.

Cosmological MHD simulations will also be needed to clarify whether the magnetothermal instability (Balbus 1991) or heat-flux driven buoyancy instability (Quataert 2008; Parrish & Quataert 2008) alter the general picture of susceptibility to condensation that we have outlined here. When anisotropic conduction is important, buoyancy is more closely related to temperature gradients than to entropy gradients (e.g., Balbus 2000; Sharma et al. 2009), which is likely to reduce the dependence of buoyancy damping on  $\alpha_K$ . Recent simulations have shown that maximally efficient anisotropic conduction does indeed alter the effects of AGN feedback on the intracluster medium (Kannan et al. 2017).

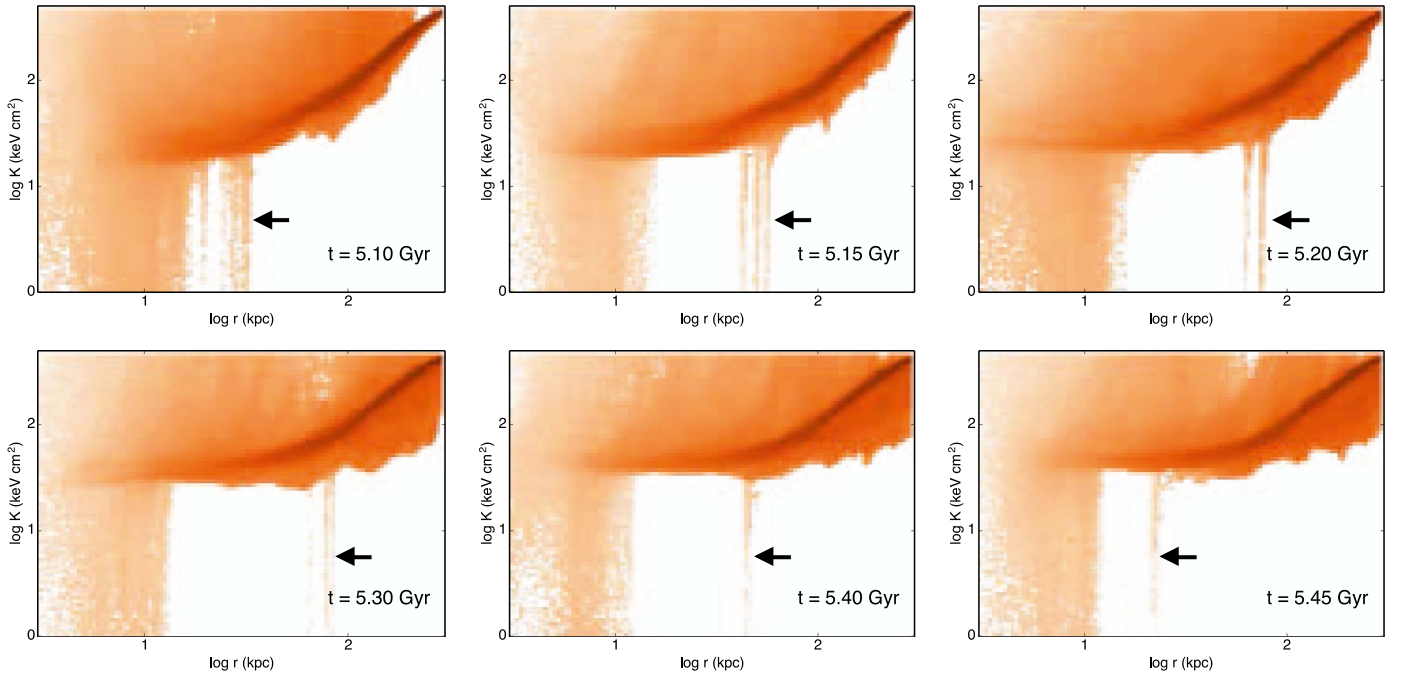
#### 7.4. Heat Transport and Dissipation

We have emphasized the importance of heat transport in stabilizing the circumgalactic medium of massive galaxies but have not paid any attention to how the transported energy is transformed into heat. The dissipation of feedback energy in galaxy-cluster cores is currently an active area of research. Many possibilities are being investigated, including turbulence, shock heating, and mixing of hot gas with cooler gas (e.g., Banerjee & Sharma 2014; Zhuravleva et al. 2014; Hillel & Soker 2016; Yang & Reynolds 2016). These dissipation processes will need to be included in a more complete model.

### 8. Concluding Thoughts

A summary of the paper’s main findings can be found in Section 1.6. Instead of repeating that summary here, we will conclude with some suggestions about how our global model for the ambient circumgalactic medium may be tested with observations and numerical simulations:

1. *Buoyancy Damping.* The model proposes that buoyancy damping can suppress condensation in circumgalactic media with a significant entropy gradient but allows it to proceed if the median entropy profile is nearly isentropic. This proposal can be tested with observations of correlations between homogeneity of the ambient medium and the slope of its entropy gradient. Cluster cores with short central cooling times are ideal for such tests, because both their entropy gradients and their levels of inhomogeneity are observable with X-ray telescopes. One expects to detect an increasing dispersion in gas entropy and temperature as the large-scale entropy gradient flattens.
2. *Two Precipitation Modes.* Multiphase gas can precipitate out of the circumgalactic medium in two different ways: (1) through growth of thermal instability in the isentropic zone, or (2) through uplift of low-entropy ambient gas. Isolated clumps of multiphase gas outside of the isentropic zone must therefore originate at lower altitudes and can result from condensation if they are uplifted from regions with  $t_{\text{ff}}/t_{\text{cool}} \lesssim 10$ . Condensation in the power-law zone of gas with greater initial values of  $t_{\text{ff}}/t_{\text{cool}}$  requires either an uplift velocity exceeding  $\approx 1.5v_c$  or some combination of drag and turbulence that slows the infall speed as the uplifted gas falls back toward its original altitude. These should be the two most prevalent modes of condensation in more complex simulations of galaxy evolution with sufficiently high spatial resolution.
3. *Lower Limit on  $t_{\text{cool}}/t_{\text{ff}}$  in the Ambient Medium.* Feedback outbursts tend to produce a lower limit of  $\min(t_{\text{cool}}/t_{\text{ff}}) \approx 10$  in the ambient medium, because ambient gas with lower values of  $t_{\text{cool}}/t_{\text{ff}}$  is vulnerable to uplift-driven condensation. X-ray observations show that galactic systems with  $v_c > 300$  km s<sup>-1</sup> generally adhere to this limit. It should also apply to ambient



**Figure 10.** Isolated region of multiphase gas rising and falling through the power-law zone. At large radii, isolated multiphase regions appear as conspicuous stripes extending directly downward in  $K$  far below the median entropy. They remain multiphase as they move vertically within the simulation environment and sometimes develop spatial substructure, which manifests as multiple components in the radial domain. Because they are long lasting, their vertical trajectories can be tracked as the simulation progresses. The black arrows in these panels show how a particular multiphase region is moving with time. Its trajectory rises and falls at a peak speed of  $\approx 500 \text{ km s}^{-1}$ . This speed is comparable to the circular velocity of the potential, which is  $\approx 550 \text{ km s}^{-1}$  at  $r = 30 \text{ kpc}$  and  $\approx 730 \text{ km s}^{-1}$  at  $r = 60 \text{ kpc}$ .

circumgalactic gas near the virial temperature in lower-mass systems, in which the conditions can be probed by *Hubble*-COS observations (e.g., Stocke et al. 2013; Werk et al. 2014).

4. *Consequences of Central Thermal Feedback.* Central thermal feedback destabilizes the circumgalactic medium because it expands the isentropic zone and suppresses buoyancy damping. Condensation can proceed there until feedback raises its cooling time to approximately the current age of the universe. Simulations implementing this mode of feedback cannot produce realistic galaxy-cluster cores, many of which have central cooling times  $\lesssim 1 \text{ Gyr}$ . Instead, heat deposition by feedback must extend beyond the isentropic zone. Feedback energy is probably transported there by bipolar outflows, and this is likely to be why kinetic feedback reproduces the characteristics of galaxy-cluster cores with much greater success than pure thermal feedback. Our analysis predicts that any numerical simulation of massive-galaxy evolution in which feedback is centrally injected and purely thermal requires heat input to raise the central cooling time to several Gyr before it succeeds in stopping star formation.
5. *Quenching and the Circumgalactic Medium Entropy Gradient.* Buoyancy damping, when coupled with kinetic feedback, can explain how star formation remains quenched in massive elliptical galaxies with short central cooling times. If feedback from a bipolar outflow can maintain a significant entropy gradient in the ambient medium, then buoyancy damping will suppress thermal instability and condensation. Episodic feedback outbursts may occasionally induce condensation through uplift, but the total amount of condensed gas should remain small,

as long as the mass of ambient gas in the isentropic zone remains small. More generally, one expects the time-averaged supply rate of condensed gas from the circumgalactic medium to be similar to the gas mass of the isentropic zone divided by the cooling time of the isentropic zone, as long as that cooling time is significantly less than the age of the universe. Quenched galaxies with short central cooling times should therefore be observed to have strong entropy gradients and small isentropic cores (as in Werner et al. 2012, 2014).

6. *Midplane Condensation in Galactic Disks.* Buoyancy damping may allow the ambient circumgalactic medium to supply a galactic disk with condensing gas via midplane condensation without any obvious “rainfall.” This can happen if there is an entropy gradient outside of the midplane and sufficient rotation to inhibit buoyancy damping in the radial direction. Given enough time, condensation will happen near the midplane (see Section 4.2), in a region with a thickness that depends on  $(t_{\text{ff}}/t_{\text{cool}})^2$ . A steady supply of hot ambient gas then settles into the midplane to replenish what is lost to condensation, without any need for infall of cold clouds at large distances from the galactic disk. This mode of gas supply should be present in numerical simulations of sufficiently high spatial resolution.
7. *Episodic Line Emission from Cooling Gas.* Condensation and star formation in a precipitation-regulated system do not necessarily happen simultaneously, in a nearly steady state (see Figure 5). Line emission from intermediate-temperature ( $10^5$ – $10^6 \text{ K}$ ) condensing gas is expected to be most luminous during the uplift events that trigger rapid condensation. In between these events, the star formation rate may exceed the condensation rate for much of the

duty cycle, causing a steady decline in the amount of cold gas.

We are indebted to A. Babul, J. Bregman, A. Crocker, A. Evrard, M. Fall, M. Gaspari, A. Kravtsov, M. McDonald, B. McNamara, P. Nulsen, P. Oh, M. Peebles, M. Ruszkowski, P. Sharma, D. Silvia, N. Soker, M. Sun, G. Tremblay, J. Tumlinson, S. White, and an insightful but anonymous referee for discussions that have shaped our thinking. Some of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1066293. G.M.V., B.W.O., and M.D. acknowledge NSF for support through grant AST-0908819 and NASA for support through grants NNX12AC98G, NNX15AP39G, HST-AR-13261.01-A, and HST-AR-14315.001-A. G.L.B. acknowledges financial support from NSF grants AST-1312888, and NASA grants NNX12AH41G and NNX15AB20G, as well as computational resources from NSF XSEDE, and Columbia University's Yeti cluster. Simulations were run on the NASA Pleiades supercomputer through allocation SMD-15-6514 and were performed using the publicly available Enzo code (<http://enzo-project.org>) and analyzed with the `yt` package (Turk et al. 2011). Enzo is the product of a collaborative effort of many independent scientists from numerous institutions around the world. Their commitment to open science has helped make this work possible.

## Appendix

### A Toy Model for the Reconfiguration of Thermally Balanced Hydrostatic Systems

In order to gain insight into how cooling and feedback regulate the global configuration of the ambient gas in a galactic gravitational potential, we consider here the behavior of an idealized but representative hydrostatic system that is isentropic within an inflection radius  $r_0$  and isothermal outside of that radius. For a monatomic ideal gas, the equation of hydrostatic equilibrium in spherical symmetry can be written as

$$\frac{d \ln P}{d \ln r} = \frac{5}{2} \frac{d \ln T}{d \ln r} - \frac{3}{2} \frac{d \ln K}{d \ln r} = -2 \frac{T_\phi}{T}, \quad (46)$$

where  $T_\phi$  is defined so that  $2kT_\phi/\mu m_p = d\phi/d \ln r$ . The temperature gradient is therefore determined by

$$\frac{d \ln T}{d \ln r} = \frac{3}{5} \alpha_K - \frac{4}{5} \frac{T_\phi}{T}, \quad (47)$$

where  $\alpha_K \equiv d \ln K / d \ln r$ . Isothermality at temperature  $T_0$  can always be enforced in a general potential well by setting  $\alpha_K(r) = 4T_\phi(r)/3T_0$ . Instead, we will simplify the mathematics by assuming a potential well with a constant value of  $T_\phi$ , in which case the temperature of the isothermal region,  $T_0 = 4T_\phi/3\alpha_K$ , is set by the constant power-law index of the entropy profile. The system's configuration is then completely specified by the entropy level  $K_0$  of the isentropic region, the gas mass  $M_0$  within that region, the power-law slope  $\alpha_K$  of the entropy profile outside that region, the depth  $T_\phi$  of the potential well, and a pressure boundary condition that ensures  $T = T_0$  at the outer edge of the isentropic region.

#### A.1. The Isentropic Region

The temperature and electron density profiles in the isentropic region are given by

$$T(r) = T_0 \left[ 1 - \frac{3\alpha_K}{5} \ln \left( \frac{r}{r_0} \right) \right],$$

$$n_e(r) = \left( \frac{kT_0}{K_0} \right)^{3/2} \left[ 1 - \frac{3\alpha_K}{5} \ln \left( \frac{r}{r_0} \right) \right]^{3/2}, \quad (48)$$

and integrating to obtain the ambient gas mass inside  $r_0$  yields

$$M_0 = \frac{4\pi}{3} r_0^3 \mu_e m_p \left( \frac{4kT_\phi}{3K_0} \right)^{3/2} \frac{I_{3/2}(\alpha_K)}{\alpha_K^{3/2}},$$

$$I_n(\alpha_K) \equiv \int_0^1 \left( 1 - \frac{3\alpha_K}{5} \ln x \right)^n 3x^2 dx, \quad (49)$$

with  $I_{3/2}(\alpha_K) \approx (1 + 0.33 \alpha_K^{1.1})$  to within 0.2% in the range  $2/3 < \alpha_K < 2$ . Reorganizing the equation for  $M_0$  then gives  $r_0$  in terms of the three adjustable parameters ( $K_0$ ,  $M_0$ ,  $\alpha_K$ ) that control the system's configuration.

In most astrophysical systems of interest, the  $t_{\text{cool}}/t_{\text{ff}}$  ratio declines with radius in isentropic gas. In this particular system, the dependence of cooling time on radius in the isentropic region follows

$$\frac{d \ln t_{\text{cool}}}{d \ln r} = (1 + 2\lambda) \frac{3\alpha_K}{10} \left[ 1 - \frac{3\alpha_K}{5} \ln \left( \frac{r}{r_0} \right) \right]^{-1}, \quad (50)$$

where  $\lambda \equiv d \ln \Lambda / d \ln T$ . The condition for  $t_{\text{cool}}/t_{\text{ff}}$  to decline with radius in a system with constant  $T_\phi$  is  $d \ln t_{\text{cool}} / d \ln r < 1$ . A decline with radius is therefore guaranteed for  $\alpha_K < 5/3$ , since astrophysical cooling functions have  $\lambda \leq 1/2$  at the temperatures characteristic of ambient circumgalactic gas.

Buoyancy damping does not suppress thermal instability in this part of the system. It is therefore prone to producing multiphase gas via condensation on a timescale  $\sim \ln(1/\delta_K) \times t_{\text{cool}}$ , where  $\delta_K$  is the fractional amplitude of the initial entropy fluctuations.

#### A.2. The Isothermal Region

The structure of the system's isothermal region outside of  $r_0$  is given by

$$K(r) = K_0 \left( \frac{r}{r_0} \right)^{\alpha_K},$$

$$n_e(r) = \left( \frac{kT_0}{K_0} \right)^{3/2} \left( \frac{r}{r_0} \right)^{-3\alpha_K/2},$$

$$\frac{t_{\text{cool}}}{t_{\text{ff}}} \propto r^{(3\alpha_K - 2)/2}. \quad (51)$$

Buoyancy damping suppresses condensation in the isothermal region as long as  $\alpha_K \gtrsim (t_{\text{ff}}/t_{\text{cool}})^2$ . Suppression of condensation therefore applies to the entire isothermal region if  $\alpha_K > 2/3$  and  $t_{\text{cool}}/t_{\text{ff}} \gg 1$  at  $r_0$ . In such systems, the minimum value of  $t_{\text{cool}}/t_{\text{ff}}$  is at  $r_0$ .

#### A.3. Redistribution of Heat

In order for the system to remain in overall thermal balance as it evolves, the heat energy lost by one of its subsystems must



be gained by another subsystem. Let subsystem 1 with gas mass  $\delta M_g$  be the one that loses an amount of heat energy  $\delta Q$  and experiences a corresponding change in specific entropy from  $K_1$  to  $K_1 - \delta K_1$ , such that

$$\delta Q = \frac{3k}{2} \frac{\delta M_g}{\mu_e m_p} \int_{K_1 - \delta K_1}^{K_1} T_1(K) \frac{dK}{K}. \quad (52)$$

In general, heat transfer changes the temperature  $T_1$  of the subsystem, and the relationship between  $\delta Q$  and  $\delta K$  depends on  $T_1(K)$  during the heat-transfer episode. For example, the particular case of condensation at constant pressure corresponds to  $T_1 \propto K^\eta$  with  $\eta = 3/5$ , and condensation at constant volume corresponds to  $\eta = 1$ . In either case, a condensation process in which  $K_1 - \delta K_1 \ll K_1$  leads to

$$\delta Q = \frac{3}{2} \frac{k T_1(K)}{\eta} \frac{\delta M_g}{\mu_e m_p}. \quad (53)$$

A more complex condensation process can be characterized by the effective value of  $\eta$  for which this equation holds.

Let subsystem 2 be that one that gains heat energy  $\delta Q$ . If subsystem 2 has a gas mass  $M_g \gg \delta M_g$ , then the gain in heat energy raises its specific entropy by

$$\delta K_2 = \frac{K_2}{\eta} \frac{T_1(K_1)}{T_2(K_2)} \frac{\delta M_g}{M_g}. \quad (54)$$

Note that this equation can also be used to derive the relation  $M_g \propto K^{-\eta}$  between the specific entropy  $K$  and gas mass  $M_g$  of the ambient medium in a system that is kept in global thermal balance as condensation gradually removes mass from the ambient phase and adds it to the condensed phase.

#### A.4. Reconfiguration of Thermally Balanced Equipotential Layers

Heat redistribution algorithms that keep each equipotential layer in independent thermal balance (e.g., McCourt et al. 2012; Sharma et al. 2012b; Gaspari et al. 2013; Meece et al. 2015) allow gas to condense out of the ambient phase in the isentropic region and produce an accompanying rise in  $K_0$ . Meanwhile, the specific entropy  $K$  of each Lagrangian layer outside of the isentropic region remains unchanged because buoyancy damping causes entropy fluctuations there to saturate at a fractional amplitude  $\delta K \sim (t_{\text{ff}}/t_{\text{cool}})^2$  instead of condensing. Consequently, a rise in  $K_0$  drives convective mixing in the neighborhood of  $r_0$  as gas with  $K < K_0$  sinks into the isentropic zone and adds to  $M_0$ .

Those changes in system configuration are best analyzed in the gas-mass domain. Let  $M_g(K)$  represent the mass of ambient gas with specific entropy less than  $K$ , and let  $M_c$  be the mass of gas that has already condensed. Inverting  $M_g(K)$  gives a two-segment entropy distribution that is isentropic at  $K_0$  for  $M_g < M_0$  and follows a power law with  $d \ln K / d \ln M_g = 2\alpha_K / (6 - 3\alpha_K)$  for  $M_g > M_0$ . A condensation event that raises  $K_0$  by  $\delta K_0$  adds gas mass  $\delta M_c = \eta M_0 \delta K_0 / K_0$  to the condensed phase and changes  $M_0$  by

$$\delta M_0 = M_0 \left( \frac{6 - 3\alpha_K}{2\alpha_K} - \eta \right) \frac{\delta K_0}{K_0}. \quad (55)$$

Condensation therefore drives the ratio of condensed to isentropic gas toward  $M_c/M_0 = 2\alpha_K \eta / (6 - 3\alpha_K)$ , implying

that the condensed gas mass is comparable to  $M_0$  for representative values of  $\alpha_K \sim 1$  and  $\eta \approx 3/5$ .

As  $K_0$  rises, there is a particular value of  $\eta$  that ensures that the  $K(r)$  profile remains isothermal,

$$\begin{aligned} \eta_0 &= \frac{3}{2} - \frac{3}{\alpha_K} \left[ 1 - \frac{1}{I_{3/2}(\alpha_K)} \right] \\ &\approx \frac{3}{2} \left[ \frac{3 + \alpha_K^{1.1} - 2\alpha_K^{0.1}}{3 + \alpha_K^{1.1}} \right], \end{aligned} \quad (56)$$

because it ensures that  $M_0 + M_c = M_0 [2/(2 - \alpha_K)] I_{3/2}^{-1}(\alpha_K)$ , which is the gas mass that would be contained within  $r$  if the isothermal region extended all the way to  $r = 0$ . For  $\alpha_K \approx 1$ , this equation gives  $\eta_0 \approx 3/4$ , which is in between the characteristic values for constant-pressure condensation ( $\eta = 3/5$ ) and constant-volume condensation ( $\eta = 1$ ). Evolution of  $K(r)$  in a system with thermally balanced equipotential layers therefore remains close to the idealized isentropic/isothermal configuration, with  $K(r)$  staying nearly fixed at  $r > r_0$  as condensation in lower-lying layers raises the value of  $K_0$  (see, e.g., Sharma et al. 2012b; Gaspari et al. 2013). Deviations of the actual value of  $\eta$  from  $\eta_0$  cause  $M_0$  to deviate somewhat from the required value for maintaining the idealized configuration and result in small deviations from isothermality just outside of  $r_0$ .

The rate at which such a system's  $K(r)$  profile evolves depends primarily on  $t_{\text{cool}}(K_0)$  but also on the initial sizes of the entropy fluctuations that lead to condensation. If the fractional amplitudes of these fluctuations are of order unity, then  $d \ln K_0 / dt \approx t_{\text{cool}}^{-1}(K_0)$ . If instead the seed fluctuations are smaller, then more time is required for condensation. The results of Section 4 show that buoyancy damping leads to saturation of perturbation growth at smaller amplitudes for larger values of  $t_{\text{cool}}/t_{\text{ff}}$ . Condensation at the center of the system, where the entropy gradient vanishes, is still possible but requires a time period that grows progressively longer compared to the cooling time as  $t_{\text{cool}}/t_{\text{ff}}$  increases.

#### A.5. Reconfiguration of a Globally Balanced System

A globally balanced system in which one layer gains heat energy while another layer loses the same amount of heat energy has more freedom to change configuration. Qualitatively, the two simplest kinds of configuration changes are (1) if cooling exceeds heating inside of  $r_0$ , causing  $K_0$  to decrease, then heating must exceed cooling in the outer region, or (2) if heating exceeds cooling inside of  $r_0$ , causing  $K_0$  to increase, then cooling must exceed heating in the outer region.

In our toy model, a small change in entropy  $\delta K(M_g)$  as a function of the gas mass  $M_g$  enclosed within a given radius corresponds to a change in heat energy given by

$$\begin{aligned} \delta Q &= \frac{3}{2} \frac{k T_0}{\mu m_p} \left[ \delta \ln K_0 \cdot \int_0^{M_0} \frac{T}{T_0} dM_g \right. \\ &\quad \left. + \int_{M_0}^{M_b} \delta \ln K \cdot dM_g \right], \end{aligned} \quad (57)$$

where  $M_b$  is the ambient gas mass within the boundary radius  $r_b$ . The configuration changes permitted by global thermal



balance within the radius  $r_b$  have the property

$$\int_{M_0}^{M_b} \delta \ln K(M_g) \cdot dM_g = -M_0 \frac{I_{5/2}(\alpha_K)}{I_{3/2}(\alpha_K)} \delta \ln K_0, \quad (58)$$

where  $I_{5/2}(\alpha_K)$  and  $I_{3/2}(\alpha_K)$  are both given by the definition in Equation (49). From Equations (49) and (51), we obtain an expression for the entropy-profile change in terms of changes in the three model parameters:

$$\delta \ln K = \left(1 - \frac{\alpha_K}{2}\right) \delta \ln K_0 - \frac{\alpha_K}{3} \delta \ln M_0 - \left[ \ln\left(\frac{K}{K_0}\right) - \frac{\alpha_K}{2} + \frac{\alpha_K}{3} \frac{d \ln I_{3/2}}{d \ln \alpha_K} \right] \delta \ln \alpha_K. \quad (59)$$

Integrating this function over  $M_g$  then gives

$$\left(\frac{M_b - M_0}{M_0}\right) \delta \ln K_b - \left(\frac{2\alpha_K}{6 - 3\alpha_K}\right) \left[\left(\frac{M_b - M_0}{M_0}\right) + \frac{3}{\alpha_K I_{3/2}}\right] \delta \ln \alpha_K = -\frac{I_{5/2}}{I_{3/2}} \delta \ln K_0, \quad (60)$$

where  $\delta \ln K_b$  is the change in specific entropy at the boundary, obtained by substituting  $K_b \equiv K(r_b)$  into Equation (59). Parameter adjustments satisfying this equation change the system's configuration without changing the amount of heat energy it contains.

Some configuration changes require a flow of free energy through the system, while others do not. The change in the system's global entropy  $S$  when  $K_0$  changes under conditions of global thermal balance is

$$\delta S = \frac{3k}{2\mu m_p} \int_0^{M_b} \frac{\delta K}{K} dM_g = -\frac{3kM_0}{2\mu m_p} \left[ \frac{I_{5/2}}{I_{3/2}} - 1 \right] \delta \ln K_0. \quad (61)$$

A reduction of  $K_0$  can proceed without a source of free energy (since  $I_{5/2} > I_{3/2}$ ) because it corresponds to a transfer of heat from the higher temperature isentropic zone to the lower temperature isothermal zone, thereby boosting the global entropy. However, a growth of  $K_0$  requires a source of free energy, because the lower-temperature outer regions then lose heat while the higher-temperature inner regions gain it, resulting in  $\delta S < 0$ .

With the global balance constraint in place, the system has only two degrees of freedom. Applying another constraint at the boundary leaves only one degree of freedom. One option is to constrain the temperature at the outer boundary to remain at  $T_0$ , which has the consequence of holding  $\alpha_K$  fixed. In that case, Equation (58) reduces to

$$\delta \ln M_0 = \left[ \left( \frac{6 - 3\alpha_K}{2\alpha_K} \right) + \frac{3}{\alpha_K} \frac{I_{5/2}}{I_{3/2}} \times \left( \frac{M_0}{M_b - M_0} \right) \right] \delta \ln K_0, \quad (62)$$

and central heating with  $\alpha_K < 2$  causes both  $K_0$  and  $M_0$  to rise while  $\alpha_K$  stays fixed. Alternatively, holding the entropy constant at the boundary gives

$$\delta \ln K_0 = \left( \frac{2}{2 - \alpha_K} \right) \left[ \frac{I_{3/2}}{I_{5/2}} \times \left( \frac{M_b - M_0}{M_0} \right) + \frac{3}{I_{5/2} \alpha_K} \right] \delta \ln \alpha_K, \quad (63)$$

$$\delta \ln M_0 = \frac{3}{\alpha_K} \left[ \frac{I_{3/2}}{I_{5/2}} \left( \frac{M_b - M_0}{M_0} \right) + \frac{3}{I_{5/2} \alpha_K} - \ln \frac{K_b}{K_0} + \frac{\alpha_K}{2} - \frac{\alpha_K}{3} \frac{d \ln I_{3/2}}{d \ln \alpha_K} \right] \delta \ln \alpha_K. \quad (64)$$

For this boundary condition,  $M_0$  must again increase as  $K_0$  grows in order for the system to remain in global thermal balance. The entropy slope  $\alpha_K$  must increase as well, but that change is small if  $M_b \gg M_0$ .

This result has important qualitative implications for feedback-regulated systems. If they are thermally balanced by a central heat source, then central heating makes ambient gas there isentropic, and the values of both  $K_0$  and  $M_0$  increase with time. Condensation in that region is not suppressed by buoyancy damping and can add to the mass  $M_c$  of condensed gas at a rate  $\dot{M}_c \sim M_0/t_{\text{cool}}(K_0)$  until the rise in  $K_0$  is sufficient for  $t_{\text{cool}}(K_0)$  to approach a Hubble time. This is not a finely regulated feedback loop but rather a monotonically evolving system in which feedback gradually shuts off condensation while a potentially large amount of condensed gas accumulates.

In contrast, globally balanced systems in which heating exceeds cooling outside of the isentropic zone generally have values of  $K_0$  and  $M_0$  that *decrease* with time. Buoyancy damping limits condensation to a relatively small central region, and the associated cooling time also decreases with time, with  $t_{\text{cool}} \sim K_0^{3/2}$ . Ultimately, something must prevent the central cooling time in the ambient medium from becoming singular. One lower limit on the cooling time of ambient gas at small radii is  $t_{\text{cool}} \gtrsim t_{\text{ff}}$ , because otherwise thermal instability is unavoidable. Another is stochastic heating that occasionally exceeds cooling, even if cooling exceeds heating during most of the duty cycle. In a less idealized system, the ambient cooling time that prevails after a heating outburst will be the timescale on which coupling and self-regulation happen, and this timescale can be much less than a Hubble time.

#### A.6. Global Regulation

Stable global regulation at a particular  $K_0$  and  $r_0$  requires the heat input balancing radiative cooling to be divided between the isentropic and isothermal zones so that each zone is individually balanced. The radiative luminosity  $L$  of each zone can be calculated by integrating

$$\frac{dL}{dr} = 4\pi \left( \frac{n_i}{n_e} \right) \left( \frac{r}{K} \right)^3 (kT)^3 \Lambda(T) \quad (65)$$

over radius. This results in a luminosity

$$L_0 = \frac{4\pi}{3} \left( \frac{n_i}{n_e} \right) \left( \frac{r_0}{K_0} \right)^3 (kT_0)^3 \Lambda(T_0) I_{3+\lambda}(\alpha_K) \quad (66)$$

for the isentropic zone and a total luminosity

$$L = L_0 \left[ 1 + \frac{3 \ln(K_b/K_0)}{I_{3+\lambda}(1)} \right] \quad (67)$$

for the particular case of  $\alpha_K = 1$ . In order to remain in a configuration with a particular value of  $K_0$ , the fraction  $f_0$  of the globally balancing heat input that is deposited into the isentropic zone must be equal to  $L_0/L$ .

Specifying a function  $f_0(K_0)$  to represent how the feedback required for global balance distributes heat as a function of radius places one more constraint on the system and leads to a potentially unique solution for the global configuration. However, not all functions  $f_0(K_0)$  permit a stable solution. Global stability requires a heat-distribution response with the property

$$\frac{d \ln f_0}{d \ln K_0} < \frac{3f_0}{I_{3+\lambda}(1)} \quad (68)$$

so that upward perturbations of  $K_0$  lead to net cooling of the isentropic zone and downward perturbations lead to net heating. If that happens, then the system's global configuration remains stable, with a value of  $K_0$  satisfying

$$f_0(K_0) \approx \left[ 1 + \frac{3 \ln(K_b/K_0)}{I_{3+\lambda}(1)} \right]^{-1}. \quad (69)$$

Qualitatively, the requirement for global stability corresponds to a heat distribution mechanism in which heat deposition onto the isentropic zone does not rise very much, and perhaps even declines, as  $K_0$  rises.

In the context of a configuration with  $\alpha_K \approx 1$ , this feature is a physically plausible consequence of feedback energy transport with a bipolar outflow, because the column density of the isentropic zone is  $N_0 \propto r_0 K_0^{-3/2} \propto K_0^{(2-3\alpha_K)/(2\alpha_K)}$ . For systems with  $\alpha_K > 2/3$ , a rise in  $K_0$  therefore causes  $N_0$  to decrease. The isentropic zone then becomes progressively less able to capture and thermalize the kinetic energy of the outflow as  $K_0$  rises, which enables global regulation.

## ORCID iDs

G. Mark Voit  <https://orcid.org/0000-0002-3514-0383>  
 Yuan Li  <https://orcid.org/0000-0001-5262-6150>  
 Brian W. O'Shea  <https://orcid.org/0000-0002-2786-0348>  
 Greg L. Bryan  <https://orcid.org/0000-0003-2630-9228>  
 Megan Donahue  <https://orcid.org/0000-0002-2808-0853>

## References

- Balbus, S. A. 1986, *ApJL*, 303, L79  
 Balbus, S. A. 1988, *ApJ*, 328, 395  
 Balbus, S. A. 1991, *ApJ*, 372, 25  
 Balbus, S. A. 2000, *ApJ*, 534, 420  
 Balbus, S. A., & Soker, N. 1989, *ApJ*, 341, 611  
 Banerjee, N., & Sharma, P. 2014, *MNRAS*, 443, 687  
 Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 1998, *ApJ*, 498, 504  
 Bertschinger, E. 1989, *ApJ*, 340, 666  
 Binney, J. 1977, *ApJ*, 215, 483  
 Binney, J., Nipoti, C., & Fraternali, F. 2009, *MNRAS*, 397, 1804  
 Blumenthal, G. R., Tabor, G. 1995, *MNRAS*, 276, 663  
 Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, *Natur*, 311, 517  
 Cavagnolo, K. W., Donahue, M., Voit, G. M., & Sun, M. 2008, *ApJL*, 683, L107  
 Choudhury, P. P., & Sharma, P. 2016, *MNRAS*, 457, 2554  
 Cowie, L. L., Fabian, A. C., & Nulsen, P. E. J. 1980, *MNRAS*, 191, 399  
 Defouw, R. J. 1970, *ApJ*, 160, 659  
 Dekel, A., & Silk, J. 1986, *ApJ*, 303, 39  
 Fabian, A. C. 1994, *ARA&A*, 32, 277  
 Fall, S. M., & Rees, M. J. 1985, *ApJ*, 298, 18  
 Field, G. B. 1965, *ApJ*, 142, 531  
 Ford, A. B., Werk, J. K., Davé, R., et al. 2016, *MNRAS*, 459, 1745  
 Fraternali, F., & Binney, J. J. 2008, *MNRAS*, 386, 935  
 Gaspari, M., Brighenti, F., & Temi, P. 2015, *A&A*, 579, A62  
 Gaspari, M., Ruszkowski, M., & Oh, S. P. 2013, *MNRAS*, 432, 3401  
 Gaspari, M., Ruszkowski, M., & Sharma, P. 2012, *ApJ*, 746, 94  
 Gaspari, M., Temi, P., & Brighenti, F. 2017, *MNRAS*, 466, 677  
 Guo, F., Oh, S. P., & Ruszkowski, M. 2008, *ApJ*, 688, 859  
 Hillel, S., & Soker, N. 2016, *MNRAS*, 455, 2139  
 Hoyle, F. 1953, *ApJ*, 118, 513  
 Hummels, C. B., Bryan, G. L., Smith, B. D., & Turk, M. J. 2013, *MNRAS*, 430, 1548  
 Kannan, R., Vogelsberger, M., Pfrommer, C., et al. 2017, *ApJL*, 837, L18  
 Kauffmann, G., & Haehnelt, M. 2000, *MNRAS*, 311, 576  
 Larson, R. B. 1974, *MNRAS*, 169, 229  
 Li, Y., & Bryan, G. L. 2012, *ApJ*, 747, 26  
 Li, Y., & Bryan, G. L. 2014a, *ApJ*, 789, 54  
 Li, Y., & Bryan, G. L. 2014b, *ApJ*, 789, 153  
 Li, Y., Bryan, G. L., Ruszkowski, M., et al. 2015, *ApJ*, 811, 73  
 Loewenstein, M. 1989, *MNRAS*, 238, 15  
 Malagoli, A., Rosner, R., & Bodo, G. 1987, *ApJ*, 319, 632  
 Maller, A. H., & Bullock, J. S. 2004, *MNRAS*, 355, 694  
 Marinacci, F., Binney, J., Fraternali, F., et al. 2010, *MNRAS*, 404, 1464  
 McCourt, M., Sharma, P., Quataert, E., & Parrish, I. J. 2012, *MNRAS*, 419, 3319  
 McNamara, B. R., & Nulsen, P. E. J. 2007, *ARA&A*, 45, 117  
 McNamara, B. R., & Nulsen, P. E. J. 2012, *NJPh*, 14, 055023  
 McNamara, B. R., Russell, H. R., Nulsen, P. E. J., et al. 2014, *ApJ*, 785, 44  
 McNamara, B. R., Russell, H. R., Nulsen, P. E. J., et al. 2016, *ApJ*, 830, 79  
 Meece, G. R., O'Shea, B. W., & Voit, G. M. 2015, *ApJ*, 808, 43  
 Meece, G. R., Voit, G. M., & O'Shea, B. W. 2017, *ApJ*, 841, 133  
 Nipoti, C., & Posti, L. 2014, *ApJ*, 792, 21  
 Nulsen, P. E. J. 1986, *MNRAS*, 221, 377  
 Oppenheimer, B. D., Davé, R., Kereš, D., et al. 2010, *MNRAS*, 406, 2325  
 Parrish, I. J., & Quataert, E. 2008, *ApJL*, 677, L9  
 Parrish, I. J., Quataert, E., & Sharma, P. 2010, *ApJL*, 712, L194  
 Pizzolato, F., & Soker, N. 2005, *ApJ*, 632, 821  
 Pizzolato, F., & Soker, N. 2010, *MNRAS*, 408, 961  
 Prasad, D., Sharma, P., & Babul, A. 2015, *ApJ*, 811, 108  
 Putman, M. E., Peek, J. E. G., & Joing, M. R. 2012, *ARA&A*, 50, 491  
 Quataert, E. 2008, *ApJ*, 673, 758  
 Rees, M. J., & Ostriker, J. P. 1977, *MNRAS*, 179, 541  
 Revaz, Y., Combes, F., & Salomé, P. 2008, *A&A*, 477, L33  
 Russell, H. R., McNamara, B. R., Fabian, A. C., et al. 2016, *MNRAS*, 458, 3134  
 Sharma, P., Chandran, B. D. G., Quataert, E., & Parrish, I. J. 2009, *ApJ*, 699, 348  
 Sharma, P., McCourt, M., Parrish, I. J., & Quataert, E. 2012a, *MNRAS*, 427, 1219  
 Sharma, P., McCourt, M., Quataert, E., & Parrish, I. J. 2012b, *MNRAS*, 420, 3174  
 Sharma, P., Parrish, I. J., & Quataert, E. 2010, *ApJ*, 720, 652  
 Silk, J. 1977, *ApJ*, 211, 638  
 Soker, N. 2006, *NewA*, 12, 38  
 Soker, N. 2008, *ApJL*, 684, L5  
 Soker, N. 2010a, *MNRAS*, 407, 2355  
 Soker, N. 2010b, arXiv:1007.2249  
 Soker, N., White, R. E., III, David, L. P., & McNamara, B. R. 2001, *ApJ*, 549, 832  
 Somerville, R. S., & Primack, J. R. 1999, *MNRAS*, 310, 1087  
 Stocke, J. T., Keeney, B. A., Danforth, C. W., et al. 2013, *ApJ*, 763, 148  
 Tabor, G., & Binney, J. 1993, *MNRAS*, 263, 323  
 Tchekhovskoy, A., & Bromberg, O. 2016, *MNRAS*, 461, L46  
 Thomas, P. A., Fabian, A. C., & Nulsen, P. E. J. 1987, *MNRAS*, 228, 973  
 Thompson, T. A., Quataert, E., Zhang, D., & Weinberg, D. H. 2016, *MNRAS*, 455, 1830  
 Tremblay, G. R., Oonk, J. B. R., Combes, F., et al. 2016, *Natur*, 534, 218  
 Tumlinson, J., Thom, C., Werk, J. K., et al. 2011, *Sci*, 334, 948  
 Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011, *ApJS*, 192, 9  
 Voit, G. M. 2011, *ApJ*, 740, 28

- Voit, G. M., Bryan, G. L., O'Shea, B. W., & Donahue, M. 2015a, [ApJL](#), **808**, L30
- Voit, G. M., Cavagnolo, K. W., Donahue, M., et al. 2008, [ApJL](#), **681**, L5
- Voit, G. M., & Donahue, M. 2015, [ApJL](#), **799**, L1
- Voit, G. M., Donahue, M., Bryan, G. L., & McDonald, M. 2015b, [Natur](#), **519**, 203
- Voit, G. M., Donahue, M., O'Shea, B. W., et al. 2015c, [ApJL](#), **803**, L21
- Wagh, B., Sharma, P., & McCourt, M. 2014, [MNRAS](#), **439**, 2822
- Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, [ApJ](#), **792**, 8
- Werner, N., Allen, S. W., & Simionescu, A. 2012, [MNRAS](#), **425**, 2731
- Werner, N., Oonk, J. B. R., Sun, M., et al. 2014, [MNRAS](#), **439**, 2291
- White, S. D. M., & Frenk, C. S. 1991, [ApJ](#), **379**, 52
- White, S. D. M., & Rees, M. J. 1978, [MNRAS](#), **183**, 341
- White, R. E., III, & Sarazin, C. L. 1987a, [ApJ](#), **318**, 621
- White, R. E., III, & Sarazin, C. L. 1987b, [ApJ](#), **318**, 629
- White, R. E., III, & Sarazin, C. L. 1987c, [ApJ](#), **318**, 612
- Yang, H.-Y. K., & Reynolds, C. S. 2016, arXiv:1605.01725
- Zhuravleva, I., Churazov, E., Schekochihin, A. A., et al. 2014, [Natur](#), **515**, 85