Are Cosmological Gas Accretion Streams Multiphase and **Turbulent?**

Nicolas Cornuault^{1*}, Matthew D. Lehnert¹, François Boulanger^{2**}, and Pierre Guillard¹

- ¹ Sorbonne Universités, UMPC Paris 6 et CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
- ² Institut d'Astrophysique Spatiale, CNRS, UMR8617, Université Paris-Sud 11, Bâtiment 121, Orsay, France

September 5, 2017

ABSTRACT

Simulations of cosmological filamentary accretion reveal flows ("s into galaxies. We present a phenomenological scenario where gas – become biphasic and, as a result, turbulent. We consider a collim filament of the cosmic web. The post-shock streaming gas expand fragments as it cools. The fragmented stream forms a two phase m argue that the hot phase sustains the accretion shock. During frag gas is converted into turbulence among and within the warm cloudetermined by the relative timescales of several processes. These copost-shock gas, the amount of turbulence in the clouds, and the dy when the gas cooling and dynamical times are of the same orderimportant mass range of Mhalo~10¹¹ to 10¹³ M_☉, where the bulk of gas accreting along cosmic web filaments may eventually loose cof separation and "disruption" of the stream, the accretion efficiency flows make the direct interaction between galaxy feedback and a accretion efficiency. As we discuss, moderating the gas accretion efficiency. As we discuss, moderating the gas accretion efficiency accretion efficiency. As we discuss, moderating the gas accretion efficiency of significant challenges in theoretical galaxy formation.

Key words. galaxies: evolution – galaxies: halos – galaxies: form verse led to the development of the first comprehensive theory of galaxy formation (e.g. White & Rees 1978; Fall & Efstathiou 1980). These analytic models embedded simple gas physics into the hierarchical growth of structure whereby smaller halos merged over time forming successively more massive halos (White & Rees 1978). Despite the successes of this model in understanding the scale of observed galaxy masses, it was soon realized that there were a number of problems. The most significant is that modeled galaxies form with a higher fraction of baryons than is observed (e.g., Ferrara et al. 2005; Bouché et al. 2006; Anderson & Bregman 2010; Werk et al. 2014). This failure was dubbed the "over-cooling problem" (Benson et al. 2003).

As numerical simulations allowed for galaxy Simulations of cosmological filamentary accretion reveal flows ("streams") of warm gas, T~10⁴K, which are efficient in bringing gas into galaxies. We present a phenomenological scenario where gas in such flows – if it is shocked as it enters the halo as we assume - become biphasic and, as a result, turbulent. We consider a collimated stream of warm gas that flows into a halo from an over dense filament of the cosmic web. The post-shock streaming gas expands because it has a higher pressure than the ambient halo gas, and fragments as it cools. The fragmented stream forms a two phase medium: a warm cloudy phase embedded in hot post-shock gas. We argue that the hot phase sustains the accretion shock. During fragmentation, a fraction of the initial kinetic energy of the infalling gas is converted into turbulence among and within the warm clouds. The thermodynamic evolution of the post-shock gas is largely determined by the relative timescales of several processes. These competing timescales characterize the cooling, the expansion of the post-shock gas, the amount of turbulence in the clouds, and the dynamical time of the halo. We expect the gas to become multiphase when the gas cooling and dynamical times are of the same order-of-magnitude. In this framework, we show that this occurs in the important mass range of $M_{halo} \sim 10^{11}$ to 10^{13} M_{\odot} , where the bulk of stars have formed in galaxies. Due to expansion and turbulence, gas accreting along cosmic web filaments may eventually loose coherence and mix with the ambient halo gas. Through both the phase separation and "disruption" of the stream, the accretion efficiency onto a galaxy in a halo dynamical time is lowered. De-collimating flows make the direct interaction between galaxy feedback and accretion streams more likely, thereby further reducing the overall accretion efficiency. As we discuss, moderating the gas accretion efficiency through these mechanisms may help to alleviate a number

Key words. galaxies: evolution – galaxies: halos – galaxies: formation – methods: analytical – turbulence – instabilities

that much of the accreting mass may penetrate into the halo as filaments of gas and dark matter (Kereš et al. 2005; Ocvirk et al. 2008). Whether or not these streams pass through a stable accretion shock as they penetrate the halo depends on the mass and redshift of the halo (Birnboim & Dekel 2003; Dekel & Birnboim 2006, hereafter BD03 and DB06 respectively). If the shock is not stable, the accretion flow is "cold". This "cold mode" accretion in simulations occurs as streams of warm (10⁴ K) gas

entering the halo, smooth at kpc-scale and weakly coupled to the infalling dark matter filaments (Danovich et al. 2015; Wetzel & Nagai 2015). Cold mode accretion is efficient in reaching down to a few tenths of a virial radius (Dekel & Birnboim 2006; Behroozi et al. 2013). The high efficiency of gas accretion in some simulations leads to model galaxies with unrealistically high baryon fractions emphasizing the over-cooling problem. To alleviate the problem of excess baryons in simulated galaxies, efficient outflows and feedback were introduced (e.g., Hopkins et al. 2012, 2016). Feedback both heats the gas in the halo, preventing it from cooling, and also ejects gas from both the galaxy and halo lowering their total gas content.

The circum-galactic media of galaxies are certainly not devoid of gas, perhaps containing up to approximately half of the total baryon content of the halo (e.g., Werk et al. 2014; Peek et al. 2015). This gas is known empirically to be multiphase. The multiphase nature of halo gas is most evident in local high mass halos, those with masses on the scales of cluster or groups. In clusters, for example, even at constant pressure, a very wide range of gas phases are observed, from hot X-ray emitting gas to cold, dense molecular gas (e.g., Jaffe et al. 2005; Edge et al. 2010; Salomé et al. 2011; Tremblay et al. 2012; Hamer et al. 2016; Emonts et al. 2016). In galaxy halos, the detection of multiphase gas is mostly through the absorption lines from warm neutral and ionized gas ($\lesssim 10^4$ to $\sim 10^6$ K) and dust via the reddening of background galaxies and quasars (Ménard et al. 2010; Peek et al. 2015, but see Pinto et al. 2014 for the detection of hot gas in X-

email: cornuault@iap.fr

^{**} Research associate at the Institut d'Astrophysique de Paris

ray emission lines). Outflows from galaxies are also multiphase (e.g., Beirão et al. 2015; Heckman & Thompson 2017) and are likely crucial for creating and maintaining the multiphase gas in halos (e.g., Gaspari et al. 2012; Sharma et al. 2012b; Borthakur et al. 2013; Voit et al. 2015a; Hayes et al. 2016).

There is only circumstantial evidence for smooth, collimated accretion streams penetrating into galaxy halos (e.g., Martin et al. 2015; Bouché et al. 2016; Vernet et al. 2017, and references therein). In analogy with analyses of gas in halos and outflows, a phenomenological approach may provide additional insights into the nature of flows and halos that galaxy simulations are perhaps not yet achieving (e.g., Sharma et al. 2010, 2012b,a; Singh & Sharma 2015; Voit et al. 2015a,b; Thompson et al. 2016). The notion that over-cooling remains a problem in simulations, that there is scant observational evidence for the streams of the type currently simulated, and because high speed collisions of gas can lead to multiphase turbulent media (Guillard et al. 2009, 2010; Ogle et al. 2010; Peterson et al. 2012; Appleton et al. 2013; Alatalo et al. 2015), all motivated us to analyze gas accretion flows phenomenologically.

Just as with the explanation for the lack of cooling flows in clusters (e.g., Peterson et al. 2003; Rafferty et al. 2008), our current understanding of accretion flows in galaxy halos may also suffer from an overly simplistic view of gas thermodynamics. In clusters, it is now understood that heating and cooling are in approximate global balance, preventing the gas from cooling catastrophically (e.g., Rafferty et al. 2008; Sharma et al. 2012b,a; McCourt et al. 2012; Zhuravleva et al. 2014; Voit et al. 2015b). In analogy with gas in cool core clusters, and in contrast to what a number of cosmological simulations currently show, the gas in streams may not cool globally. Instead, if the gas in streams is inhomogeneous and subject to thermal and hydrodynamic instabilities (e.g., Sharma et al. 2010), the differences in cooling times between the gas phases will lead to fragmentation of the gas. If streams are unstable, their gas will not remain monophasic or laminar. Thus, our goal in this paper is to investigate the question posed in the title: "Are Cosmological Gas Accretion Streams Multiphase and Turbulent?". If yes, the gas energetics may regulate the gas accretion efficiency. Dekel et al. (2013) estimated the penetration efficiency over large halo scales at $z\sim2$ of $\sim50\%$ but this estimate only considered macroscopic processes that may influence the accretion efficiency. Because heating and cooling are controlled by the mass, energy and momentum exchanges between the gas phases, a careful investigation of the gas physics on microscopic scales is required to investigate whether those microphysical processes may further reduce the efficiency of gas accretion onto galaxies.

To investigate the question posed in the title, we begin by presenting a qualitative sketch of our scenario, make a quantitative investigation of the impact of expansion in the post-shock gas on an accretion flow, and then discuss the case where the post-shock gas has small fluctuations in density and temperature, developing a criterion for when the gas will fragment (§ 2). In § 3, we discuss the consequences of the formation of a multiphase flow on the thermodynamic evolution of the stream after it penetrates the halo. To gauge the astrophysical pertinence of our model, we analyze the evolution of an idealized gas accretion stream into a halo of 10^{13} M $_{\odot}$ at z=2 (§ 4). In Section 5, we discuss why simulations may be missing some ingredients necessary for modeling accretion shocks robustly and outline a few simple consequences of our proposed scenario.

2. Our framework for multiphased streams

The idea that gas in halos is multiphase has been suggested for decades (e.g., Binney 1977; Maller & Bullock 2004). More recent studies of halo gas attribute the development of multiphase gas to the growth of local thermal instabilities (e.g., Sharma et al. 2010) or galaxy outflows (e.g., Thompson et al. 2016; Hayes et al. 2016). Thermal instability is only relevant when the cooling time of the unstable gas is of the same order-of- magnitude or smaller than the dynamical time of the halo (e.g., Sharma et al. 2012b; McCourt et al. 2012). In ambient halo gas and outflows from galaxies, the gas must often meet this requirement given that they are observed to be multiphase. However, in an accreting stream of gas, it is difficult to understand how the gas might achieve such a balance in heating and cooling. We may have to consider other processes to determine if it is possible for streams themselves to become multiphased as they flow into the halo. In the following sections, we examine the physics of gas flowing into halos from cosmic web filaments.

2.1. Qualitative sketch of our specific framework

We briefly qualitatively outline our scenario of gas accretion through streams, sketched in Fig. 1, introducing the concepts developed later in the paper. We consider a collimated stream of warm gas with a temperature of 10⁴ K that penetrates into a dark matter halo filled with hot gas at the halo virial temperature. The smooth density distribution of the hot gas follows the density distribution of the underlying dark matter halo. The ambient halo gas has a long cooling time and has constant density and temperature during the flow. The speed of the flow is set to the virial velocity of the halo and is highly supersonic relative to the sound speed of the gas within the stream. In the following we introduce the three basic physical ingredients of our modeling of streams.

As the stream penetrates at the virial radius, we assume it is shock-heated whenever the hot halo gas provides the necessary pressure support for sustaining the shock (§ 2.2). The postshock gas is over-pressurized relative to the ambient halo gas and expands, invalidating the classical one-dimensional analysis of streams (DB06, Mo et al. 2010). The hot post-shock gas mixes with the ambient halo gas, which prevents much of the gas initially in the stream from cooling completely to form a monophasic post-shock stream. We further posit that the post-shock gas will develop inhomogeneities due to, for example, non-planarity or obliqueness of the shock-front or through inhomogeneities in velocity and/or density of the stream before it is shocked. The fragmentation of the gas into hot and warm cloudy phases is central to our scenario. If certain physical conditions are met, the expanding inhomogeneous gas will cool and fragment, forming a two phase medium – a hot phase with an embedded warm cloudy phase (§ 2.2). Density and velocity inhomogeneities in the postshock gas may be amplified through gas cooling, leading to the formation of a multiphased flow (see § 2.3).

We expect that part of the kinetic energy of the infalling gas will be converted to turbulence within the warm clouds and random cloud-cloud motions (e.g., Hennebelle & Pérault 1999; Kritsuk & Norman 2002; Heigl et al. 2017). If the level of turbulence is high or if the clouds are formed while the gas is expanding, the warm clouds may spread beyond the initial boundary of the collimated inflowing stream (Fig. 1 and see § 3). If these conditions are met, the stream will de-collimate. The thermodynamic evolution of the post-shock gas in the stream is largely determined by the timescales of several relevant processes – gas

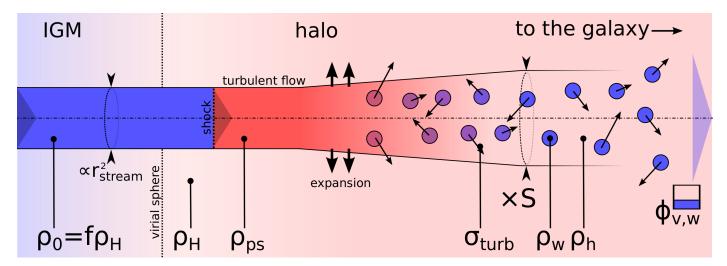


Fig. 1. Sketch of our phenomenological picture of flows of gas passing through a virial shock. The initial inflowing gas (in blue) is over-dense relative to the halo gas at the boundary by a factor f (= $\rho_0/\rho_{\rm H}$ or the stream density divided by the density of the ambient halo gas at the virial radius), shocks at the boundary between the inter-galactic medium (labeled as "IGM") and the hot halo gas (labeled as "halo"). The persistent virial shock and the higher pressure of the post-shock gas compared to the ambient halo gas, allows the flow to expand after being shocked (labeled as "expansion"). The post-shock gas may become unstable, fragmenting to form a biphasic medium. The fragmentation enables a fraction of the initial momentum and energy of the stream to be captured as turbulent clouds of warm gas with a dispersion, $\sigma_{\rm turb}$, and a volume-filling factor, $\phi_{\rm v,w}$. The clouds may move beyond the initial radius of the stream, de-collimating the flow. Eventually, the hot post-shock gas mixes with the ambient halo gas which prevents it from cooling further. See text and the Appendix for definitions of the variables.

cooling, expansion of the hot phase, dynamical time of the halo, and the stream disruption due to turbulent motions – which we quantify in the remainder of this and the next section.

2.2. The Impact of Expansion on the Phenomenology of Accretion Flows

Since the pioneering work of White & Rees (1978) decades ago, the accretion of gas onto galaxies or into halos has been analyzed in one, radial dimension, assuming homogeneity of the flow (see Mo et al. 2010). The one-dimensional approximation is also used for streams, where it is valid only if one can ignore lateral expansion. To know if this approximation is valid, one has to compare the expansion to cooling time of the gas within the flow. Historically, even the analysis of the stability of the accretion shock neglected the possibility that the post-shock gas may expand into the ambient halo gas gas (DB06; Mo et al. 2010). In our study, we reconsider the sustainability of the accretion shock that occurs as the infalling filament collides with the ambient halo gas. To conduct this analysis, we compare the cooling time of the post-shock gas to various other measures of its dynamical evolution. For simplicity, we only compare the cooling time to the expansion time of the post-shock gas into the surrounding ambient halo gas and the dynamical time of the halo. Intuitively, one can understand that if the cooling time is significantly shorter than either the expansion or dynamical times, the stream after the shock will quickly cool down and maintain much of its integrity. Such a shock can be unstable and this is essentially the textbook situation that has been considered already (Mo et al. 2010). If the cooling time of the post-shock gas is significantly longer than either the expansion or dynamical times, then the post-shock gas will mix the ambient halo gas further supplying it with hot gas and entropy. This is another textbook example that has been analyzed one-dimensionally (DB06; Mo et al. 2010). However, if the cooling time is of the same order-of-magnitude of the expansion and halo dynamical times, a wider range of outcomes of the post-shock gas are possible (e.g., Sharma et al. 2010, 2012a). This is the point of this paper – to describe phenomenologically this more complex regime.

We now consider the impact of expansion on the properties of the post-shock gas. To estimate the cooling, expansion, and dynamical timescales, we introduce two parameters specific to the flow, its over-density relative to that of mean density of gas at the virial radius, f, and the radius of the stream, $r_{\rm stream}$. All variables used in this and subsequent sections are summarized in Tables A.1 and A.2 of the Appendix. As we now show, the gas in the stream immediately after being heated to a high post-shock temperature, $T_{\rm ps}$, has a pressure, $P_{\rm ps}$, higher than the ambient halo gas, $P_{\rm H}$, and thus expands into the surrounding ambient halo gas of density, $\rho_{\rm H}$. All of the post-shock quantities are given by the normal Rankine-Hugoniot shock relations (see Appendix for appropriate formulae). The halo pressure at the virial radius is,

$$P_{\rm H} = \left(\frac{k_{\rm B}}{\mu m_{\rm P}}\right) \rho_{\rm H} T_{\rm H} = \frac{(\gamma - 1)}{2} \rho_{\rm H} v_{\rm vir}^2 = \frac{\gamma \mathcal{M}_1^2}{3 f} P_0 \tag{1}$$

where $T_{\rm H}$ is the temperature of the ambient halo gas, which we have assumed to the be virial temperature. The pressure ratio just after the shock is,

$$\frac{P_{\rm ps}}{P_{\rm H}} = \frac{2f}{\gamma(\gamma+1) \cdot g(\mathcal{M}_1)} \tag{2}$$

where $g: M_1 \mapsto (2\gamma/(\gamma-1) - M_1^{-2})^{-1}$. For $\gamma = 5/3$ and within the domain $[1,+\infty]$, this function rapidly decreases from 1/4 to 1/5. Hence the pressure ratio, $P_{\rm ps}/P_{\rm H}$ is between 9/5 and 9/4 times the initial stream over-density, f. Since $f \gtrsim 1$, the pressure of the post-shock gas is higher than the pressure of the ambient halo. Thus, since the pressure ratio is always greater than one, the flow will expand into the ambient halo gas.

To define the expansion time of the flow, we approximate this as the inverse of the relative rate of change of pressure in the post-shock gas as it expands. Quantitatively, this is,

$$t_{\rm expand} = P_{\rm ps}/\dot{P}_{\rm ps} = -2\gamma r_{\rm stream}/\dot{r}_{\rm stream} \sim 2\gamma r_{\rm stream}/c_{\rm ps}$$
 (3)

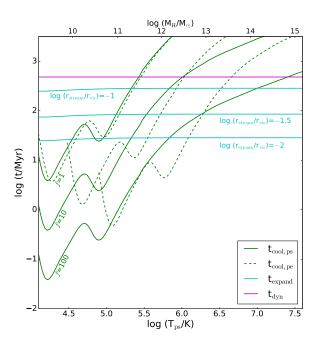


Fig. 2. Timescale comparison of the competing relevant processes of cooling, expansion, and halo dynamics, for different post-shock temperatures (i.e., different halo masses), at redshift z = 2. Cooling times (green) are times of isobaric cooling from a given temperature-density couple. The solid green curves indicate the instantaneous post-shock cooling time, $t_{\text{cool,ps}}$, and the green dashed curves indicate the postexpansion cooling time, $t_{\text{cool},pe}$ – the cooling time of the post-shock after it has expanded sufficiently to reach pressure equilibrium with the ambient halo gas. The cooling time curves for the post-expansion gas are truncated at the minimum temperature spanned by the curves, 10⁴ K. The cooling time curves were computed for three values of the initial over-density f, as indicated by green labels on the left side of the panel. Expansion times (solid cyan lines) are shown for three different relative filament radii, log_{10} $r_{stream}/r_{vir}=-1$, -1.5, and -2. The dynamical time, $t_{\rm dyn,halo}$, which is independent of halo mass, is indicated by the horizontal magenta line.

where $r_{\rm stream}$ is the initial radius of the stream before it expands and $c_{\rm ps}$ is the sound speed of the gas immediately after passing through the shock front. As a first order approximation, in the case of a homogeneous post-shock medium, one can easily compare the cooling and expansion times. Our definition of the expansion time means it is a differential measure of the expansion and so the most appropriate comparison is with the instantaneous cooling time of the gas immediately after the shock. The isobaric cooling time is defined as,

$$t_{\rm cool}(T) = T \left| \frac{dT}{dt} \right|_P^{-1} = \frac{5k_{\rm B}\rho T}{2\mu n_e n_{\rm H}\Lambda(T)} \approx 4.3 \times \frac{5k_{\rm B}\mu m_{\rm p}T}{2\rho\Lambda(T)} \tag{4}$$

where $k_{\rm B}$ is the Boltzmann constant, $\Lambda(T)$ is the electronic cooling efficiency as a function of temperature, T (Sutherland & Dopita 1993; Gnat & Sternberg 2007), μ is the mean molecular weight¹, $m_{\rm P}$ is the mass of the proton, ρ is the mass density of the cooling gas, n_e is the electronic density and $n_{\rm H}$ the Hydrogen particle density². Hereafter, the cooling time of the post-shock is denoted by, $t_{\rm cool,ps}$. We also refer to the cooling time after expansion, noted $t_{\rm cool,pe}$.

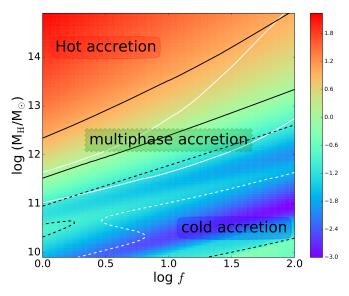


Fig. 3. Ratio of the post-expansion cooling and halo dynamical timescales as a function of halo mass, $M_{\rm H}$, and stream over-density, f, at redshift, z=2. The colors represent the log ($t_{\rm cool,pe}/t_{\rm dyn,halo}$) (Eqs. 4 and 5) whose corresponding values are indicated by the color bar. The solid black lines indicate $\log{(t_{\rm cool,pe}/t_{\rm dyn,halo})}$ of increasing values of 0 and 1 from the middle to the top of the diagram. Dashed black lines indicate $\log{(t_{\rm cool,pe}/t_{\rm dyn,halo})}$ =-1. The white contours indicate values of $\log{(t_{\rm cool,pe}/t_{\rm expand})}$ (Eqs. 3 and 4) equal to 0 and 1 (solid), and -1 (dashed). For this analysis, we have assumed $\log{(r_{\rm stream}/r_{\rm vir})}$ =-1.5. We highlight regions where the cooling times are significantly longer than the halo dynamical times as "hot accretion", regions where the cooling time is significantly shorter than the dynamical times as "cold accretion", and regions where the cooling times are of the same order-of-magnitude as the dynamical times as "multiphase accretion" (see text for details). These labels are illustrative and are not intended to precisely delineate a clean separation in the timescales.

The other timescale with which to compare the cooling time is the dynamical time of the halo. It is simple to estimate the dynamical time. We estimate the dynamical time for matter falling from the virial radius with a radial velocity of v_2 (the post-shock velocity) directed at the center of the potential as,

$$t_{\rm ff} = \alpha \frac{r_{\rm vir}}{v_2} \approx t_{\rm dyn,halo}$$
 (5)

For this estimate, we assume a NFW dark matter potential (Navarro et al. 1997) and α , is a factor of-order unity to account for the integrated gravitational acceleration during the fall. Assuming that the characteristic velocity and radius are the virial values, $v_{\rm vir}$ and $r_{\rm vir}$, the dynamical time is independent of halo mass (Mo et al. 2010).

Fig. 2 illustrates how these time scales depend on the post-shock temperature, directly related to the halo mass (top x-axis) and for a range of relative stream over-densities, f, and relative stream radii, $r_{\text{stream}}/r_{\text{vir}}$ at z=2. The initial stream penetrating the halo gas and the warm post-shock clouds are assumed to have a temperature of 10^4 K, which is maintained through external heating processes, such as the ionization by the meta-galactic flux or ionizing photons from the galaxy in the halo. Since the post-shock temperatures are proportional to velocity, the post-shock temperature translates into dark matter halo mass. From this analysis, we see that the expansion times are almost constant above $M_{\rm H} \gtrsim {\rm few} 10^{10} {\rm M}_{\odot}$ since the virial shock is sufficiently strong such that $t_{\rm expand} \propto t_{\rm dyn,halo}$. As f increases, the

 $^{^{1}}$ $\mu = 0.6 m_{\rm p}$ for a fully ionized gas.

The factor 4.3 in Eq. 4 comes from $n_e = 1.2n_{\rm H}$ and $n_{\rm H} \approx \rho/(1.4 \, m_{\rm p})$.

cooling times decrease systematically, which means for low temperatures (low mass halos at z=2), the cooling time is always less than the expansion time. For very high mass halos, even for wide streams with relatively high densities, the cooling time is much longer than the expansion time and can be longer than the halo dynamical time for low to moderately over-dense filaments. A wide range of values in f and $r_{\text{stream}}/r_{\text{vir}}$ lead to cooling times that are less than the halo dynamical time and within an order-of-magnitude of the expansion time. Whenever the expansion time is comparable to the cooling time, the gas flow will globally cool neither purely adiabatically or isobarically.

We can understand the competition between these time scales, t_{cool,ps}, t_{cool,pe}, t_{dyn,halo}, and t_{expand}, more clearly by considering them as a function of the stream over-density, f, and post-shock temperature or halo mass (Fig. 3). For the case where $t_{\text{cool,ps}} >> t_{\text{expand}}$ and $t_{\text{cool,pe}} > t_{\text{dyn,halo}}$, the post-shock gas will expand rapidly remaining hot over at least a dynamical time of the halo. In this regime the hot post-shock gas will likely mix with the ambient halo gas preventing significant cooling. This is akin the "hot mode accretion" discussed by Dekel and collaborators (e.g., DB06). In Fig. 3, we have labeled this regime, "hot accretion". At the other extreme in Fig. 3, when $t_{\text{cool,ps}} \ll t_{\text{expand}}$ and $t_{\rm cool,ps} << t_{\rm dyn,halo}$, the gas cools before significant expansion and before the stream has penetrated deeply into the halo. Accretion in this regime corresponds to their "cold mode accretion" and have labeled it as such in Fig. 3. In between these two regimes, $t_{\rm cool,pe}, t_{\rm cool,ps}, t_{\rm dyn,halo},$ and $t_{\rm expand}$ are all about the same orderof-magnitude. In this regime, there is not a simple dichotomy between the types of accretion streams – they can be both hot and warm if the post-shock gas has a range of temperature and/or densities and the thermodynamic evolution of the post-shock gas will be more complicated. We labeled this regime "multiphase accretion" in Fig. 3.

2.3. Inhomogeneous infalling streams and post-shock gas: differential cooling

We now discuss the physical mechanisms behind the development of multiphasic accretion streams. Although our previous discussion of the relevant timescales considered homogeneous streams, density and velocity inhomogeneities may arise through the dynamics of the shock itself (Kornreich & Scalo 2000; Sutherland et al. 2003). Simulations show that stream are not accreted homogeneously but have substructure in both density and velocity (Nelson et al. 2016). Inhomogeneities may arise due to a range of curvature in accretion shock-fronts, translating into a range of Mach numbers and post-shock temperatures and densities. Density fluctuations at constant shock velocity will also lead to inhomogeneities in the post-shock gas in both temperature and density resulting in a range of cooling times (Guillard et al. 2009). As we now discuss, once such "differential cooling" sets in, it acts like the thermal instability, leading to phase separation in the flow (Sharma et al. 2012b), but over a finite range of time in the absence of a heating process balancing cooling of the hot phase.

To understand how the inhomogeneous post-shock gas evolves, we investigate how small fluctuations in the density and/or temperature are amplified. Neglecting the influence of any heating process, the radiative cooling is not balanced and there are no fixed equilibrium points. Following closely the development in Sharma et al. (2012b), we begin with a parcel of gas that is over-dense relative to an ambient medium. The magnitude of the over-density is, $\delta \equiv |\delta \rho/\rho| \sim |\delta \, T/T|$ for isobaric conditions. The inverse of the effective cooling time of the over-

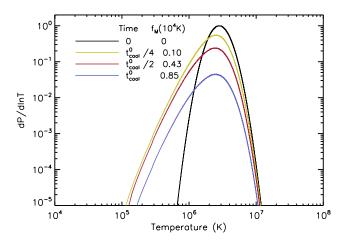


Fig. 4. Probability distribution function of the gas temperature at different times for isobaric cooling. The initial distribution (black line) is log-normal with a dispersion, $\sigma_T = 0.3$, and an initial mean gas temperature, $T_0 = 2.7 \times 10^6$ K. The distribution at a time equal to the initial isobaric gas cooling time is plotted in blue, that of half and a quarter of this time in red and green, respectively. In the last column of the legend, we indicate the fraction of the gas that has cooled to 10^4 K within the indicated fraction of the cooling time. For example, in this illustration, after one cooling time, approximate 85% of the gas has cooled to 10^4 K.

dense gas parcel relative to the ambient medium is then

$$\frac{1}{t_{\text{cool,eff}}} = -\frac{1}{t_{\text{cool,parcel}}} + \frac{1}{t_{\text{cool,ambient}}}.$$
 (6)

As already shown by Sharma et al. (2012b), for small overdensities, $\delta \le 1$, the inverse of the effective cooling time under isobaric conditions is

$$\frac{1}{t_{\rm cool,eff}} \sim -\delta \frac{\partial}{\partial \ln T} \left(\frac{1}{t_{\rm cool}} \right)_{\rm P} = \frac{\delta}{t_{\rm cool}} \left[2 - \frac{\mathrm{d} \ln \Lambda(T)}{\mathrm{d} \ln T} \right]. \tag{7}$$

Marginally denser gas cools faster than the gas in which it is embedded and the cooling becomes a runaway process for temperatures below a few 10⁵ K. We refer to this process as "differential cooling". As soon as heterogeneities form in approximate pressure equilibrium with their surroundings, denser cooler regions of the gas will cool rapidly, forming clouds, leaving a hotter, rarer, high volume-filling inter-cloud medium. Thus, differential cooling leads to multiphasic post-shock accretion flows. This is akin to what occurs for thermally unstable gas (Field 1965) but the phase separation is transient in the absence of a heating process of the hot gas balancing cooling.

Of course, a natural consequence of differential cooling is that even for halo masses where the post-shock gas has on average a long cooling time, gas with fluctuations in temperature or density may locally cool sufficiently rapidly to form a multiphased stream.

Analytically, inhomogeneities can be accounted for as is done in phenomenological analyses of the formation of stars through the use of probability distribution functions (PDFs) in gas density and/or temperature (e.g., Hennebelle & Chabrier 2009). We follow the thermodynamical evolution of the temperature and density PDFs using the energy equation, namely,

$$\dot{e} = -\rho \frac{\Lambda(T)}{4.3\mu m_p} + P \frac{\dot{\rho}}{\rho^2} \,, \tag{8}$$

where e is the specific internal energy. We illustrate the isobaric evolution of inhomogeneities in a flow using such an approach in Fig. 4. In this illustration, we consider a log-normal PDF in temperature with a dispersion of $\sigma_T = 0.3$ (1σ dispersion of $\log \frac{\rho}{\rho_0} = \log \frac{T}{T_0}$, where ρ_0 and T_0 are the initial mean values of the density and temperature, with corresponding cooling time $t_{\rm cool}^0$). We have chosen a gas pressure and mean temperature appropriate for the post-shock gas of a flow into a massive, 10^{13} ${\rm M}_{\odot}$, halo at z=2 (see § 4). For this illustration, the gas pressure is ${\rm P/k} = 5.8 \times 10^4$ K cm⁻³, and the initial mean temperature T_0 =2.7 × 10^6 K. We find that a substantial fraction of the accreting gas cools to 10^4 K in a fraction of $t_{\rm cool}^0$, while the peak of the hot gas temperature PDF does not shift with time (see Fig. 4).

We show in Fig. 5 both the mass fraction and volume-filling factor of the gas that cools to 10^4 K as a function of time. Note that since the results shown on Figs. 4 and 3 are plotted as a function of the initial cooling time $t_{\rm cool}^0$, to a first approximation they are not much dependent on the specific values of ρ_0 and T_0 . Those results also apply to the gas cooling after expansion when $t_{\rm expand} << t_{\rm cool}^0$. Figure 3 shows that the warm gas becomes the dominant mass phase, within about half of the post-shock cooling time, and that the warm gas only fills a small fraction of the volume. For a larger value of the temperature dispersion $\sigma_{\rm T}$, $\Phi_{\rm m,w}$ increases over a broader range of fractional times distributed around the same mean value.

In Fig. 5, the gas is multiphase for only a few t_{cool}^0 . The hot gas eventually cools because we have not considered any heating processes. However, the process of fragmentation of the gas could be sustained if there is heating of the hot gas. Even in absence of heating, after expansion the hot gas will eventually mix with the ambient halo gas thus sustaining the hot phase. Possible local sources of heating for the post-shock gas are the radiative precursor of high velocity shocks, thermal conduction, radiation from the surrounding hot medium, and dissipation of turbulence. The mechanical energy input from active galactic nuclei (AGN), winds generated by intense star formation, and other processes can plausibly balance the cooling of the hot gas globally (e.g., Best et al. 2007; Rafferty et al. 2008). Overall, there is more than enough energy, but what is unknown is how and with what efficiency this energy is transferred to the streaming gas. Even though simulations do not capture this process, it has been suggested that the energy from the galaxy is transferred efficiently through turbulent energy cascade and dissipation (Zhuravleva et al. 2014; Banerjee & Sharma 2014).

A poignant question to ask is can the flowing post-shock gas actually become multiphase while cooling in a halo dynamical time? It is a difficult question to answer in the particular case of accretion streams because it depends on how inhomogeneous the gas is. It will occur if the post-shock conditions are sufficiently inhomogeneous. The gas will become multiphased around when $t_{\text{cool,pe}}$ is the same order-of-magnitude as $t_{\text{dyn,halo}}$. This justifies where we placed the "multiphase accretion" label in Fig. 3. This figure shows that this occurs in the important mass range of $M_{halo} \sim 10^{11}$ to 10^{13} M_{\odot}. The values of course depend on the relative over-density of the streams (f) and through the characteristics of the halos, on redshift. Halos within this mass range is where the bulk of stars have formed in galaxies. Moreover, because the volume filling factor of the warm gas is likely to be always small, even in the cases where the mass fraction is large, multiphase streams are likely to be difficult to identify observationally. We discuss this further in § 5.

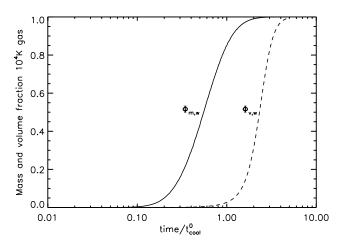


Fig. 5. The change in the mass fraction, $\phi_{m,w}$, and the volume filling-factor, $\phi_{v,w}$, of the warm, 10^4 K gas as a function of the post-shock cooling time. The time scale is expressed in units of the isobaric cooling time, t_{cool}^0 as defined in the text. The values of the pressure and temperature are the same as those used in Fig. 4.

Consequences of the formation of multiphased, cloudy accretion flows

3.1. Turbulence in warm clouds

As the gas fragments, perhaps due to differential cooling and/or thermal instabilities, we assume that some of the initial kinetic energy of the stream is converted into turbulence within the warm component. We parameterize the turbulence after phase separation as the ratio, η , of the turbulent energy density of the clouds and the initial bulk kinetic energy density of the stream. We define this ratio as,

$$\eta = \frac{\langle \rho_{\rm w} \rangle_{\rm v} \, \sigma_{\rm turb}^2}{\rho_1 v_1^2} \tag{9}$$

where $\langle \rho_w \rangle_v$ is the volume-averaged density of the warm clouds $(\langle \rho_w \rangle_v = \phi_{v,w} \ \rho_w)$, where $\phi_{v,w}$ is volume-filling factor and ρ_w is the density of the warm clouds respectively), σ_{turb} is the cloud-cloud velocity dispersion, ρ_1 and v_1 are the pre-shock gas density and velocity. We assume that v_1 is equal to the virial velocity, v_{vir} . The initial density of the stream is related to the hot halo density, ρ_H , as $\rho_1 = f\rho_H$. The amount of turbulence generated in the post-shock gas is likely determined by as yet poorly understood and undoubtedly complex gas physics. In our model we do not attempt to investigate the complexity involved in this transformation of energy, we simply parameterize the amount of turbulence by η which we allow to be free (see § 3.3).

As we briefly alluded to in § 2.3, as the turbulent energy dissipates, it reheats the warm clouds, moderating the cooling. The cooling of the clouds will also be moderated by mixing with the ambient halo gas as the stream penetrates deeper into the halo potential (Fragile et al. 2004). Dissipation of turbulent kinetic energy in the hot phase may also contribute to heating the post-shock gas as well as the radiation from the high-speed accretion shock (Allen et al. 2008). The parameter η may be high enough to provide the required heating of the hot gas in the stream through turbulent dissipation and mixing with the halo gas, perhaps instigating thermal instabilities. On the other hand, η must be low enough such that turbulent dissipation and mixing does not prevent instabilities, both thermal and cooling, from

growing in the post-shock gas (Banerjee & Sharma 2014; Zhuravleva et al. 2014). These constraints and considerations may ultimately provide limits on how much or how little turbulence is sustainable in the post-shock gas, but determining these exact values requires further detailed study and will not be considered here.

3.2. Mass and momentum budget of the multiphase medium

We assume that as the stream flows into the halo, its mass flow rate is conserved during the post-shock expansion and does not mix immediately with the ambient halo gas. This leads to the relation,

$$S\rho_2 v_2 = \rho_1 v_1 \tag{10}$$

where ρ_2 is the density after the hot gas has expanded by a factor, S, and similarly, v_2 is its velocity (Fig. 1). The expansion factor, S, is defined as the ratio of the initial over the final mass fluxes (per unit surface perpendicular to the flow). The post-shock gas will ultimately reach pressure equilibrium with the halo which implies, $\rho_{\rm w}T_{\rm w}=\rho_{\rm h}T_{\rm h}=\rho_{\rm H}T_{\rm H}$, where $\rho_{\rm w}$ and $\rho_{\rm h}$ are the densities of the warm and hot components and $T_{\rm H}$ is the temperature of the hot components after phase separation in the post-shock gas.

Before there is any momentum exchange with the halo gas, the momentum of the streaming gas is conserved, implying,

$$S\left(\rho_2 v_2^2 + \eta \rho_1 v_1^2 + P_H\right) = \rho_1 v_1^2 + P_1 \tag{11}$$

where P_1 is the initial pressure of the stream. We assume the fragmenting gas radiates away its heat until reaching a floor temperature, $T_{\rm w}=10^4$ K. The hot component only cools adiabatically and reversibly (no heat transfer and no entropy increase) expanding after it passed through the shock front, namely, $\rho_{\rm h}^{\gamma}P_{\rm ps}=\rho_{\rm ps}^{\gamma}P_{\rm H}$, where γ is the ratio of specific heats. This approximation holds until the expanding gas reaches pressure equilibrium with the ambient halo gas.

The expansion factor, S, is derived from Eqs. 1, 10, and 11, as

$$S = \frac{1}{2} \left(\eta + \frac{\gamma - 1}{2f} \right)^{-1} \left(1 - \sqrt{1 - 4\frac{\rho_{H}}{\rho_{2}} \left(\eta f + \frac{\gamma - 1}{2} \right)} \right)$$
 (12)

The expansion of the stream is important in our formulation. It leads to the mixing of the expanding post-shock gas with the ambient halo gas. This mixing couples the hot post-shock gas to the larger energy reservoir of the ambient halo gas which acts as a thermostat preventing the gas from cooling, thereby possibly maintaining the pressure necessary to support a sustained shock. As we discussed in § 2, it is the relative ratios of the thermal cooling time and the expansion time that will influence how the stream evolves.

3.3. Cloudy stream disruption

The relative cloud-cloud motions may lead to the warm clouds spreading beyond the original radius of the stream. So instead of the streams being highly collimated as we assumed they are initially (before the virial shock), the flows may de-collimate. This can be thought of as disruption since the warm clouds expand

away from the original trajectory of the stream, thus "disrupting" the flow. We define the timescale for disruption as the cloud crossing time of the stream, namely,

$$t_{\text{disrupt}} = \frac{r_{\text{stream}}}{\sigma_{\text{turb}}} \tag{13}$$

 σ_{turb} may be computed from η , f, and the volume-filling factor of the warm clouds, $\phi_{\text{v,w}}$. From Eqs. 1 and 9, assuming pressure equilibrium between gas phases, yields,

$$\sigma_{\text{turb}}^{2} = v_{1}^{2} (\eta f / \phi_{\text{v,w}}) (T_{\text{w}} / T_{\text{H}})$$

$$= \frac{2}{(\gamma - 1)} \left(\frac{k_{\text{B}}}{\mu m_{p}} \right) (\eta f / \phi_{\text{v,w}}) T_{\text{w}}$$
(14)

Characterized this way, $\sigma_{\text{turb}}^2/v_1^2 > \eta$. For a wide range of relative amounts of turbulent energy, stream over-density, and volume-filling factor of the warm gas, the cloud-cloud dispersion be up to $v_{\text{vir}}/2$. The dynamical evolution of the stream is determined by the ratio of the disruption and halo dynamical time. If $t_{\text{disrupt}} \gg t_{\text{dyn,halo}}$, then the warm clouds within the stream will remain collimated as they flow, otherwise, the streams will decollimate.

4. A specific case: $10^{13} \, \mathrm{M}_{\odot}$ halo at z=2

To gauge whether any of the phenomenology we have discussed is pertinent astrophysically, we calculate the stream characteristics for a single dark matter halo of mass, $10^{13}\,M_\odot$, at redshift 2. We chose this halo mass and redshift because in DB06, halos at this mass and redshift were determined to have substantial accretion rates in the "hot mode". Our analysis in § 2 indicated that this halo mass and redshift would be a revealing illustration of how the impact of expansion and accelerated differential cooling might change the physical characteristics of gas accretion flows (i.e., it would no longer simply be "hot mode accretion"). Both the halo mass and the redshift set the initial stream density and the characteristics of the halo gas. We provide all the characteristics of the halos and initial gas conditions in the Appendix.

4.1. Why are streams cloudy?

There are two limiting cases to specifically consider when attempting to understand the development of a biphasic stream. The two cases are: (1) $t_{\rm expand} \ll t_{\rm cool,ps}$ and (2) $t_{\rm expand} \gg t_{\rm cool,ps}$. In case (1), the warm phase develops only after the expansion has occurred, while in case (2) the clouds form before the expansion. We sketch the thermodynamic evolution ("path") of a stream in Fig. 6. To make these illustrations, we adopted f=30 and $r_{\rm stream}/r_{\rm vir}=0.01$ for case (1), and f=150 and $r_{\rm stream}/r_{\rm vir}=0.1$ for case (2).

In case (1), clouds fill the entire expanse of the expanded flow. Adiabatic expansion occurs before radiative cooling becomes important, the stream cools and the density of the hot phase declines without a change in entropy. In case (2), the two phase separate before expansion. In both cases, the hot post-shock gas reaches pressure equilibrium and mixes with the ambient halo gas.

These two conditions are of course for the extreme cases, in reality, the gas will have $t_{\rm expand} \sim t_{\rm cool,ps}$ (Fig. 3). For these intermediate cases, the thermodynamic evolution is more complex, but the clouds reach the same final thermodynamical state.

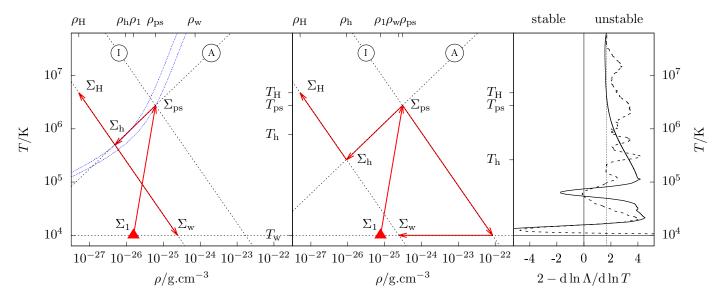


Fig. 6. (*left and middle*) Sketch of the thermodynamic path for cases (1) and (2) (see text). Σ_1 (red triangle) indicates the density and temperature of the pre-shock gas. The gas is shocked reaching the point Σ_{ps} . Subsequently, the gas cools adiabatically due to the expansion of the stream until the halo pressure is reached at Σ_h . Two phases separate. For case (1), phase separation occurs at Σ_h , while for case (2), it occurs at Σ_{ps} . In both cases, the hot component mixes with the surrounding halo gas, reaching Σ_H . For case (1), the warm component cools radiatively and isobarically to the point Σ_w . For case (2), the same point is reached but the gas takes a different thermodynamic path. Dashed black lines represent adiabats and isobars labeled A and I respectively. The two blue-dashed curves in the left panel are contours of constant cooling time, $t_{cool} \sim 0.2t_{dyn,halo}$ (upper curve) and $t_{cool} \sim 0.12t_{dyn,halo}$ (lower curve). These curves indicate that the cooling time during the expansion remains approximately constant. (*right*) Analysis of the isobaric differential cooling "instability" (Eq. ??) of low-, 10^{-3} solar (solid line) and solar-metallicity (dashed line) gas as a function of temperature. When $2 - d \ln \Lambda/d \ln T > 0$ the gas can become heterogeneous through accelerated differential cooling.

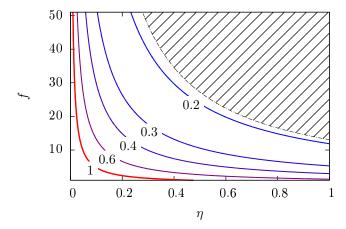


Fig. 7. Disruption of the flow as a function of the filament over-density, f, and level of turbulence, η. The contours represent constant ratios of $t_{\rm disrupt}/t_{\rm dyn,halo}$ as labeled (cf. Eqs. 13 and 5). We assume a volume-filling factor of 0.1 (see Fig. 5) for the warm clouds and $r_{\rm stream} = r_{\rm vir}/10$. In regions with values less than 1, the streams are "disrupted" (§ 4.2). The shaded region indicates regions that are forbidden because for these values of the parameters, the post-shock pressure is less than the halo pressure. We note that because $t_{\rm disrupt} \propto \sigma_{\rm turb}^{-1}$, the contours of constant $t_{\rm disrupt}/t_{\rm dyn,halo}$ are shaped like contours of constant $\sigma_{\rm turb}$ in the same plane. For example, the contour, $t_{\rm disrupt}/t_{\rm dyn,halo}$ =0.2, is close to the contour for $\sigma_{\rm turb}$ =200 km s⁻¹, which is almost half the initial velocity of the flow.

The cooling length at the post-shock temperature is the size over which structures can cool isobarically (i.e., cooling length, $\lambda_{\text{cooling,ps}} = c_{\text{ps}} t_{\text{cool,ps}}$). In case (1), the cooling length is much larger than the stream radius, $\lambda_{\text{cooling,ps}} >> r_{\text{stream}}$, clouds may form over all scales within the stream. In case (2), it is much

smaller, $\lambda_{\text{cooling,ps}} << r_{\text{stream}}$, and clouds may form over a range of sizes smaller than the stream radius. The expansion of the hot gas does not inhibit the growth of thermal and differential cooling instabilities because the decrease in the pressure is roughly compensated by the decrease of temperature along an adiabat in the expression of the cooling time. In other words, for the post-shock temperature, i.e., for the halo mass we have adopted, the gas cooling time remains roughly constant as the gas expands (Fig. 6; see also Fig. 2). Eventually, the warm phase equilibrates at approximately the halo pressure and the hot phase mixes with the halo gas. We are obviously considering fragmentation on scales much smaller than the scale height of the gravitational potential well and thus we can safely ignore thermal stabilization by convection (Balbus & Soker 1989; Sharma et al. 2010).

4.2. Does the accretion flow disrupt?

The only process we consider in determining whether or not the warm clouds will travel coherently towards the galaxy proper as observed in numerical simulations (e.g. Brooks et al. 2009; Danovich et al. 2015) is the cloud-cloud velocity dispersion. The cloud-cloud dispersion will broaden the stream as it penetrates into the halo. Fig. 7 and Fig. 8 show contours of $t_{\text{disrupt}}/t_{\text{dyn,halo}}$ for a constant r_{stream} (Eqs. 13 and 5). We find that the disruption time is shorter than or approximately equal to the halo dynamical time. Thus it appears that for a wide range of relative turbulent energy densities, stream over-densities, and volume-filling factors of the warm gas, the flows will not simply fall directly into the potential as a highly collimated, coherent streams. In reality, the clouds are dynamical entities, we expect clouds to keep forming through cooling as other clouds are destroyed by hydrodynamic instabilities and heated by dissipation (Cooper et al. 2009).

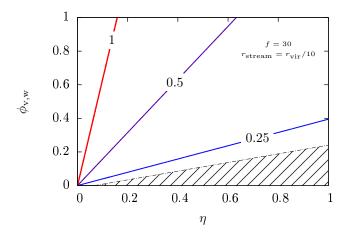


Fig. 8. Disruption of the flow as a function of the volume-filling factor of the warm gas and level of turbulence. The contours represent constant ratios of $t_{\rm disrupt}/t_{\rm dyn,halo}$ as labelled (cf. Eqs. 13 and 5). We assume f=30 and $r_{\rm stream} = r_{\rm vir}/10$. In regions with values less than 1, the streams are disrupted. The shaded region has the same meaning as in Fig. 7.

5. Discussion

We now discuss broadly how our findings relate to aspects of galaxy formation and evolution.

5.1. Are virial shocks persistent?

In our scenario, the existence of a hot phase in hydrostatic equilibrium supports a persistent shock (Binney 1977; Maller & Bullock 2004). Cosmological simulations appear to show a similar phenomenology as that described in DB06. Depending on the mass of the halo and redshift, streams penetrate at about one to many 100s of km s⁻¹ (van de Voort & Schaye 2012; Goerdt & Ceverino 2015) or much greater than the sound speed of the stream, $c_s \sim 10-20 \text{ km s}^{-1}$. Simulated accretion shocks are "isothermal" at high Mach numbers and not stable (e.g., Nelson et al. 2016, 2015). Perhaps this uniform isothermality is due to the spatial and temporal resolutions adopted in the cosmological simulations. The scales that simulations must probe are roughly delineated in one-dimensional shock calculations. Raymond (1979) show that atomic shocks with velocities of ~ 100 km s⁻¹ reach their post-shock temperatures within a distance of \approx 1-10 \times 10¹⁵ cm in less than \sim 30 yrs. For higher shock velocities, the spatial and temporal scales will be even shorter (Allen et al. 2008). The gas cools after the shock on timescales that are at most only a couple of orders-of-magnitude longer. In addition, in order to capture the differential cooling and thermal instabilities in the post-shock gas, resolutions much finer than the Field length are required (Koyama & Inutsuka 2004; Gressel 2009). Simulations should be specifically designed to capture the multiphase nature of streams penetrating halos to test our scenario by resolving the Field length (Koyama & Inutsuka 2004). Since the Field length decreases strongly with decreasing temperature, this is most easily done with ad hoc floor temperature higher than 10⁴ K but lower than the virial temperature as done for simulating thermal instabilities in cool core clusters (McCourt et al. 2012).

The resolution and temporal scales necessary to resolve high Mach number shocks are not achievable in galaxy- or cosmological-scale simulations. To overcome this limitation, numericists use artificial viscosity in the form of a dissipative term either in dynamical equations or dispersion relations, depending on the properties of the gas or the flow (e.g., Kritsuk et al. 2011; Price 2012; Hu et al. 2014; Beck et al. 2016). Artificial viscosity spreads the shock over several resolution elements enabling simulations to resolve heating and cooling across the shock front. The Reynolds number is inversely proportional to the kinetic viscosity of the fluid. If the flow properties are unchanged but the viscosity increased, the Reynolds number of the flow will be artificially low. Simulations with artificial viscosity have flows with low Reynolds numbers. Simulated low Re flows, Re $\lesssim 1000$, tend to be laminar. Those with low spatial and temporal resolutions, due to not resolving the Field length and having unrealistically low Reynolds numbers likely fail to produce biphasic turbulent flows (see, e.g., Kritsuk & Norman 2002; Sutherland et al. 2003; Koyama & Inutsuka 2004; Kritsuk et al. 2011; Nelson et al. 2016, for discussion).

5.2. Nature of Flows into Galaxies: Observational tests

As a consequence of our assumption that all energetic quantities scale as $v_{\rm vir}$, the cloud-cloud dispersion is a simple linear function of η , f, and $\phi_{\rm v,w}^{-1}$. This relation implies that turbulent velocities of the warm clouds in the post-shock gas are independent of both halo mass and redshift. In principle this means that post-shock streams may be turbulent in any halo at any redshift. The reality is probably much more complex, through both macro-and microscopic gas physics which is not yet well-understood, the 2 parameters, η and $\phi_{\rm v,w}$, likely depend on the accretion velocity and the physical state of the ambient halo gas – both of which undoubtedly depend on redshift and halo mass.

Our scenario has observationally identifiable consequences. If our scenario is realistic, then observations should reveal: (1) clumpy, turbulent streams; (2) strong signs of the dissipation of turbulent mechanical energy in the warm medium (e.g., Guillard et al. 2009; Ogle et al. 2010; Tumlinson et al. 2011). The situation described in our model where a large fraction of the bulk kinetic energy of the accretion flow is transfered to turbulent motions amongst cold clouds is observed in large-scale galaxy colliding flows, like the situation in the Taffy Galaxies or the Stephan's Quintet, where we see evidence for this energy cascade (Peterson et al. 2012; Cluver et al. 2010). In Stephan's Quintet, two atomic gas filaments are colliding at $\sim 1000 \text{ km s}^{-1}$ and yet instead of finding intense X-ray emission from the postshock gas, most of the bulk kinetic energy is contained in the turbulent energy of the warm molecular gas (Guillard et al. 2009, 2010). Remarkably, roughly 90% of the bulk kinetic energy has not been dissipated in the large-scale shock and is available to drive turbulence. If gas in halos is multiphase and turbulent, then it may be the case as well for accretion streams.

Obviously, a clumpy stream is difficult to identify as such through absorption line spectroscopy and this may explain why streams have not been conspicuously identified so far. This is most obviously seen in Fig. 5 where despite a large fraction of the gas is warm, its volume-filling factor is minuscule. Along most lines-of-sight, absorption spectroscopy is expected to sample only the hot, high volume-filling factor halo gas or probe a population of warm ambient halo clouds (Maller & Bullock 2004; Tumlinson et al. 2011; Werk et al. 2014). The clouds should be looked for in emission. Their emission can be powered by the UV radiation from the galaxy but also through the localized dissipation of turbulent energy and through losses of their gravitational potential energy as they fall into the halo. This may have already been observed in Ly α (e.g., Cantalupo et al. 2014; Martin et al. 2015). In particular, it would be promising to interpret spectral-imaging observations such as those provided by

MUSE on the ESO/VLT within the context of our scenario (see Borisova et al. 2016; Fumagalli et al. 2016; Vernet et al. 2017).

5.3. Moderating the accretion rate: Biphasic streams and increased coupling between "feedback" and accretion

In our phenomenological model, two mechanisms moderate the accretion efficiency on to galaxies: (1) disruption and fragmentation of the flow; (2) interaction between streams and outflows of mass, energy, and momentum due to processes occurring within galaxies (e.g., AGN, intense star-formation).

First, streams potentially become multiphase and turbulent leading to short disruption times resulting in de-collimation. Any de-collimation undoubtedly leads to longer accretion times and thus lower overall accretion efficiencies compared to smooth isothermal streams (Danovich et al. 2015; Nelson et al. 2016). The post shock gas becomes multiphase over a wide range of halo masses at z=2. A fraction of the initial stream mass flow becomes hot gas and ultimately mixes with the surrounding ambient hot halo gas. Thus even accretion streams potentially feed gas into the hot halo which may have long cooling times compared to the halo free fall time (White & Rees 1978; Maller & Bullock 2004).

Second, simulations indicate that the mass and energy outflows from galaxies can interact with streams, regulating or even stopping the flow of gas (e.g., Ferrara et al. 2005; Dubois et al. 2013; Nelson et al. 2015; Lu et al. 2015). Simulated streams are relatively narrow (e.g., Ocvirk et al. 2008; Nelson et al. 2013, 2016) and generally penetrate the halo perpendicular to the spin axes of disk galaxies and their directions are relatively stable for long periods (e.g., Pichon et al. 2011; Dubois et al. 2014; Welker et al. 2014; Laigle et al. 2015; Codis et al. 2015; Tillson et al. 2015). Feedback due to mechanical and radiative output of intense galactic star formation and active galactic nuclei is observed to be highly collimated in inner regions of disk galaxies (opening angle, $\Omega \sim \pi$ sr, e.g., Heckman et al. 1990; Lehnert & Heckman 1995, 1996; Beirão et al. 2015). In the case of dwarf galaxies, their outflows are generally more weakly collimated (e.g., Marlowe et al. 1995, 1997; Martin 1998, 2005). The geometry of the accretion flows and the significant collimation of outflows from galaxies in simulations, result in only weak direct stream-outflow interactions. However, accretion flows in simulations can be moderated or stopped when the halo gas is pumped with mass and energy via feedback to sufficiently high thermal pressures and low halo-stream density contrasts to induce instabilities in the stream and disrupt it; or when the halo gas develops a sufficiently high ram pressure in the inner halo due to angular momentum exchange between the gas, dark matter, and galaxy disk to disrupt accretion flows (van de Voort et al. 2011; Dubois et al. 2013; Nelson et al. 2016).

The processes we have described are generic to flows, whether they are inflows or outflows. It is only the context and timescales that change (Thompson et al. 2016). Just as with the accretion flows modeled here, we also expect the galaxy winds to be highly uncollimated as they flow from the galaxy due to the formation (and destruction) of turbulent clouds. Thus, the generation of turbulent cloudy media in both accretion flows and starburst-driven outflows will allow for efficient interaction between these two types of flows. This dynamical interaction likely sustains turbulence in the halo compensating for the dissipation. It is perhaps through this interaction that galaxies become "self-regulating" on a halo scale (e.g., Fraternali et al. 2013), and not only on a galaxy scale (e.g., Lehnert et al. 2013, 2015).

While the fate of the turbulent clouds is beyond the scope this paper, the qualitative implication is that the gas accretion efficiency may be moderated through the generation of turbulence in biphasic flows. The fragmented, turbulent nature of the gas in streams and outflows likely makes their dynamical and thermal interaction and coupling efficient. Note that other mechanisms, like the growth of Rayleigh-Taylor and Kelvin-Helmholtz instabilities, associated with gas cooling, can also trigger the formation of cold clouds in the surrounding halo (e.g. Kereš & Hernquist 2009). Moderating the overall gas accretion efficiency onto galaxies may help to alleviate two significant challenges in contemporary astrophysics: the distribution of the ratio of the baryonic to total halo mass as a function of halo mass (e.g. Behroozi et al. 2013), where low mass galaxies have especially low baryon fractions; and the requirement for models to drive extremely massive and efficient outflows to reduce the baryon content of galaxies (e.g., Hopkins et al. 2012, 2016).

6. Conclusions

We developed a phenomenological model of filamentary gas accretion, "streams", into dark matter halos. We assume both that streams penetrate ambient hot halo gas as homogeneous flows of $10^4~K$ gas and that they undergo a shock at the virial radius of the halo. The ingredients of the model, those which sets it apart from other phenomenological models of gas accretion, are that we assume the "virial shock" is sustained, the post-shock gas expands into a ambient hot halo gas, and through several mechanisms or characteristics of the shock front, the post-shock gas is inhomogeneous. To gauge whether this model is astrophysically pertinent, we discuss the thermodynamic evolution of a single stream penetrating a dark matter halo of mass $10^{13}~M_{\odot}$ at z=2. From this analysis, we find that:

- The post-shock gas expands into the halo gas and it can fragment due to differential cooling and hydrodynamic instabilities. Instabilities lead to the formation of a biphasic flow. It is the formation of a hot post-shock phase which mixes with the ambient hot halo gas, ultimately limiting how much of the gas can cool. As a result of the phase separation and the pressure provided by the hot post-shock gas, we argue that the virial shock may be sustained. However, we have not analyzed the sustainability of the shock in detail in this paper.
- The development of a biphasic medium converts some of the bulk kinetic energy into random turbulent motions in the gas (e.g., Hennebelle & Pérault 1999; Kritsuk & Norman 2002). The turbulent energy cascades from large to small scales and across gas phases. The flows, while retaining significant bulk momentum as they penetrate into the halo, are turbulent with cloud-cloud dispersion velocities that can be up to 1/2 of the initial velocity of the stream.
- For a wide range of turbulent energy densities, our model shows that the stream will lose coherence in less than a halo dynamical time. We emphasize that the turbulent energy density is not in reality a free parameter but is determined by macro- and microscopic multiphase gas physics about which we have only a rudimentary understanding. To understand what processes regulate the amount of turbulence in streams, high resolution simulations of accreting gas need to be made and additional multi-wavelength observations useful for constraining the properties of turbulent astrophysical flows are necessary.

The post-virial shock gas is not isothermal, accretion streams are both hot and cold. The "hot-cold dichotomy" (see DB06) is

no longer a simple function of whether or not the shock is stable, but now relies both on the shock occurring and under what circumstances the post-shock gas becomes multiphase and turbulent. However, we have discussed may apply if there is no virial shock provided that inflowing gas is hot and already inhomogeneous (Kang et al. 2005; Cen & Ostriker 2006). Thus, even in absence of a virial shock, the gas may become multiphased by compression as it falls deeper into the halo potential.

Moderating the gas accretion efficiencies on to galaxies through this and other mechanisms may help to alleviate some significant challenges in theoretical astrophysics. If gas accretion is actually not highly efficient, then perhaps models will no longer have to rely on highly mass-loaded outflows to regulate the gas content of galaxies. It is likely that the underlying physical mechanisms for regulating the mass flow rates and evolution of outflows are very similar to those that regulate gas accretion (Thompson et al. 2016). If so, then observing outflows in detail can provide additional constraints on the physics of astrophysical flows generally. We do not only have to rely on apparently challenging detections of direct accretion onto galaxies.

Acknowledgements. This work is supported by a grant from the Région Ile-de France. NC wishes to thank the DIM ACAV for its generous support of his thesis work. We thank Yohan Dubois for insightful discussions and an anonymous referee for their skepticism, poignant questions, and significant challenges which helped to improve our paper significantly.

```
References
Alatalo, K., Appleton, P. N., Lisenfeld, U., et al. 2015, ApJ, 812, 117
Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J.
   2008, ApJS, 178, 20
Anderson, M. E. & Bregman, J. N. 2010, ApJ, 714, 320
Appleton, P. N., Guillard, P., Boulanger, F., et al. 2013, ApJ, 777, 66
Balbus, S. A. & Soker, N. 1989, ApJ, 341, 611
Banerjee, N. & Sharma, P. 2014, MNRAS, 443, 687
Beck, A. M., Murante, G., Arth, A., et al. 2016, MNRAS, 455, 2110
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
Beirão, P., Armus, L., Lehnert, M. D., et al. 2015, MNRAS, 451, 2640
Benson, A. J., Bower, R. G., Frenk, C. S., et al. 2003, ApJ, 599, 38
Best, P. N., von der Linden, A., Kauffmann, G., Heckman, T. M., & Kaiser, C. R.
   2007, MNRAS, 379, 894
Binney, J. 1977, ApJ, 215, 483
Birnboim, Y. & Dekel, A. 2003, MNRAS, 345, 349
Borisova, E., Cantalupo, S., Lilly, S. J., et al. 2016, ApJ, 831, 39
Borthakur, S., Heckman, T., Strickland, D., Wild, V., & Schiminovich, D. 2013,
   ApJ, 768, 18
Bouché, N., Finley, H., Schroetter, I., et al. 2016, ApJ, 820, 121
Bouché, N., Lehnert, M. D., & Péroux, C. 2006, MNRAS, 367, L16
Brooks, A. M., Governato, F., Quinn, T., Brook, C. B., & Wadsley, J. 2009, ApJ,
```

694, 396 Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, Nature, 506, 63

Cen, R. & Ostriker, J. P. 2006, ApJ, 650, 560

Cluver, M. E., Appleton, P. N., Boulanger, F., et al. 2010, ApJ, 710, 248

Codis, S., Pichon, C., & Pogosyan, D. 2015, MNRAS, 452, 3369

Cooper, J. L., Bicknell, G. V., Sutherland, R. S., & Bland-Hawthorn, J. 2009, ApJ, 703, 330

Danovich, M., Dekel, A., Hahn, O., Ceverino, D., & Primack, J. 2015, MNRAS, 449, 2087

Dekel, A. & Birnboim, Y. 2006, MNRAS, 368, 2

Dekel, A., Zolotov, A., Tweed, D., et al. 2013, MNRAS, 435, 999 Dubois, Y., Pichon, C., Devriendt, J., et al. 2013, MNRAS, 428, 2885

Dubois, Y., Pichon, C., Welker, C., et al. 2014, MNRAS, 444, 1453

Edge, A. C., Oonk, J. B. R., Mittal, R., et al. 2010, A&A, 518, L46

Emonts, B. H. C., Lehnert, M. D., Villar-Martín, M., et al. 2016, Science, 354, 1128

Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189

Ferrara, A., Scannapieco, E., & Bergeron, J. 2005, ApJ, 634, L37

Field, G. B. 1965, ApJ, 142, 531

Fragile, P. C., Murray, S. D., Anninos, P., & van Breugel, W. 2004, ApJ, 604, 74 Fraternali, F., Marasco, A., Marinacci, F., & Binney, J. 2013, ApJ, 764, L21 Fumagalli, M., Cantalupo, S., Dekel, A., et al. 2016, MNRAS, 462, 1978

```
Gaspari, M., Ruszkowski, M., & Sharma, P. 2012, ApJ, 746, 94
Gnat, O. & Sternberg, A. 2007, ApJS, 168, 213
Goerdt, T. & Ceverino, D. 2015, MNRAS, 450, 3359
 Gressel, O. 2009, A&A, 498, 661
Guillard, P., Boulanger, F., Cluver, M. E., et al. 2010, A&A, 518, A59
Guillard, P., Boulanger, F., Pineau Des Forêts, G., & Appleton, P. N. 2009, A&A,
         502, 515
Hamer, S. L., Edge, A. C., Swinbank, A. M., et al. 2016, MNRAS, 460, 1758 Hayes, M., Melinder, J., Östlin, G., et al. 2016, ApJ, 828, 49 Heckman, T. M., Armus, L., & Miley, G. K. 1990, ApJS, 74, 833 Heckman, T. M. & Thompson, T. A. 2017, ArXiv e-prints Heigl, S., Burkert, A., & Gritschneder, M. 2017, arXiv.org, arXiv:1705.03894
Hennebelle, P. & Chabrier, G. 2009, ApJ, 702, 1428
Hennebelle, P. & Pérault, M. 1999, A&A, 351, 309
Hopkins, P. F., Quataert, E., & Murray, N. 2012, MNRAS, 421, 3522
Hopkins, P. F., Torrey, P., Faucher-Giguère, C.-A., Quataert, E., & Murray, N.
         2016, MNRAS, 458, 816
 Hu, C.-Y., Naab, T., Walch, S., Moster, B. P., & Oser, L. 2014, MNRAS, 443,
         1173
 Jaffe, W., Bremer, M. N., & Baker, K. 2005, MNRAS, 360, 748
Jaffe, W., Bremer, M. N., & Baker, K. 2005, MNRAS, 360, 748
Kang, H., Ryu, D., Cen, R., & Song, D. 2005, ApJ, 620, 21
Kereš, D. & Hernquist, L. 2009, ApJ, 700, L1
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Kornreich, P. & Scalo, J. 2000, ApJ, 531, 366
Koyama, H. & Inutsuka, S.-i. 2004, ApJ, 602, L25
Kritsuk, A. G., Nordlund, Å., Collins, D., et al. 2011, ApJ, 737, 13
Kritsuk, A. G. & Norman, M. L. 2002, ApJ, 569, L127
Laigle, C., Pichon, C., Codis, S., et al. 2015, MNRAS, 446, 2744
Lehnert, M. D. & Heckman, T. M. 1995, ApJS, 97, 89
Lehnert, M. D. & Heckman, T. M. 1996, ApJ, 462, 651
Lehnert, M. D., Le Tiran, L., Nesvadba, N. P. H., et al. 2013, A&A, 555, A72
Lehnert, M. D., van Driel, W., Le Tiran, L., Di Matteo, P., & Haywood, M. 2015, A&A, 577, A112
         A&A, 577, A112
 Lu, Y., Mo, H. J., & Wechsler, R. H. 2015, MNRAS, 446, 1907
Maller, A. H. & Bullock, J. S. 2004, MNRAS, 355, 694
 Marlowe, A. T., Heckman, T. M., Wyse, R. F. G., & Schommer, R. 1995, ApJ,
```

438, 563 Marlowe, A. T., Meurer, G. R., Heckman, T. M., & Schommer, R. 1997, ApJS,

112, 285
Martin, C. L. 1998, ApJ, 506, 222
Martin, C. L. 2005, ApJ, 621, 227
Martin, D. C., Matuszewski, M., Morrissey, P., et al. 2015, Nature, 524, 192
McCourt, M., Sharma, P., Quataert, E., & Parrish, I. J. 2012, MNRAS, 419, 3319
Ménard, B., Scranton, R., Fukugita, M., & Richards, G. 2010, MNRAS, 405,

No, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493 Nelson, D., Genel, S., Pillepich, A., et al. 2016, MNRAS, 460, 2881 Nelson, D., Genel, S., Vogelsberger, M., et al. 2015, MNRAS, 448, 59 Nelson, D., Vogelsberger, M., Genel, S., et al. 2013, MNRAS, 429, 3353 Ocvirk, P., Pichon, C., & Teyssier, R. 2008, MNRAS, 390, 1326 Ogle, P., Boulanger, F., Guillard, P., et al. 2010, ApJ, 724, 1193 Peek, L. E. G. Ménard, B. & Corrales, L. 2015, ApJ, 813, 7 Ogle, P., Boulanger, F., Guillard, P., et al. 2010, ApJ, 724, 1193
Peek, J. E. G., Ménard, B., & Corrales, L. 2015, ApJ, 813, 7
Peterson, B. W., Appleton, P. N., Helou, G., et al. 2012, ApJ, 751, 11
Peterson, J. R., Kahn, S. M., Paerels, F. B. S., et al. 2003, ApJ, 590, 207
Pichon, C., Pogosyan, D., Kimm, T., et al. 2011, MNRAS, 418, 2493
Pinto, C., Fabian, A. C., Werner, N., et al. 2014, A&A, 572, L8
Price, D. J. 2012, MNRAS, 420, L33
Rafferty, D. A., McNamara, B. R., & Nulsen, P. E. J. 2008, ApJ, 687, 899
Raymond, J. C. 1979, ApJS, 39, 1
Salomé, P., Combes, F., Revaz, Y., et al. 2011, A&A, 531, A85
Sharma, P., McCourt, M., Parrish, I. J., & Ouataert, E. 2012a, MNRAS

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

Singh, A. & Sharma, P. 2015, MNRAS, 446, 1895 Suresh, J., Bird, S., Vogelsberger, M., et al. 2015, MNRAS, 448, 895 Sutherland, R. S., Bicknell, G. V., & Dopita, M. A. 2003, ApJ, 591, 238 Sutherland, R. S. & Dopita, M. A. 1993, ApJS, 88, 253 Thompson, T. A., Quataert, E., Zhang, D., & Weinberg, D. H. 2016, MNRAS,

Tillson, H., Devriendt, J., Slyz, A., Miller, L., & Pichon, C. 2015, MNRAS, 449,

Tremblay, G. R., O'Dea, C. P., Baum, S. A., et al. 2012, MNRAS, 424, 1026 Tumlinson, J., Thom, C., Werk, J. K., et al. 2011, Science, 334, 948 van de Voort, F. & Schaye, J. 2012, MNRAS, 423, 2991

van de Voort, F., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, MNRAS,

415, 2782

Vernet, J., Lehnert, M. D., De Breuck, C., et al. 2017, A&A, 602, L6 Voit, G. M., Bryan, G. L., O'Shea, B. W., & Donahue, M. 2015a, ApJ, 808, L30 Voit, G. M., Donahue, M., Bryan, G. L., & McDonald, M. 2015b, Nature, 519, 203

Welker, C., Devriendt, J., Dubois, Y., Pichon, C., & Peirani, S. 2014, MNRAS, 445, L46
Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, ApJ, 792, 8
Wetzel, A. R. & Nagai, D. 2015, ApJ, 808, 40
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
Zhuravleva, I., Churazov, E., Schekochihin, A. A., et al. 2014, Nature, 515, 85

Appendix A: Parameters in the model

The quantities that are important in setting the initial conditions of the stream-ambient halo gas interaction are the mass, virial velocity, and virial radius of the dark matter halo which we denote as $M_{\rm H}$, $v_{\rm vir}$, and $r_{\rm vir}$. The dark matter distribution is given by a NFW profile with a concentration parameter, c, of 10 (Navarro et al. 1997). The halo is filled by a hot gas of temperature $T_{\rm H}$, which we assume to be equal to the virial temperature of the halo, $T_{\rm vir}$. The density of the hot halo, $\rho_{\rm H}$, is assumed to follow that of the dark matter density with radius, but is multiplied by the cosmological baryon density relative to the dark matter density, $f_{\rm B}=0.18$. This is $\approx 37 f_{\rm B} \rho_{\rm crit}$, where $\rho_{\rm crit}$ is the critical density of the Universe. The halo pressure, $P_{\rm H}$, is related to $T_{\rm H}$ and $\rho_{\rm H}$. The filling factor of this gas is assumed to be one. We are agnostic about how this hot, high volume-filling factor halo at the virial temperature formed but note that it likely forms by a combination of accretion of gas from the intergalactic medium and heating through the radiative and mechanical energy output of the galaxy embedded in the halo (e.g., Suresh et al. 2015; Lu et al. 2015).

Table A.1. Halo and gas parameters for example in § 4

Parameter Name	Symbol	Value
Halo mass	$M_{ m H}$	$10^{13}\mathrm{M}_\odot$
Baryonic fraction	$f_{ m B}$	0.18
Redshift	Z	2
Virial radius	$r_{ m vir}$	220 kpc
Virial velocity	$v_{ m vir}$	440 km s^{-1}
Critical density	$ ho_{ m crit}/\mu m_{ m p}$	$7.6 \times 10^{-5} \text{ cm}^{-3}$
Number density at $r_{\rm vir}$	n_0	$5.1 \times 10^{-4} \text{ cm}^{-3}$
Adiabatic index	γ	5/3
Mean particle mass	$\mu m_{ m p}$	$0.6 \times m_{\rm p}$

The gas accretes through a stream of infalling gas with radius, r_{stream} , we assume that it passes through a shock and that the properties of the post-shock gas is given by the standard set of shock equations. We simply scale the density of the accreting stream by a factor, f, which is its density contrast of the background dark matter density at the virial radius multiplied by the cosmological baryon fraction (i.e., $\rho_{\rm H}$). We further assume that there is a temperature floor in the post-shock gas of 10⁴ K. We assumed this temperature mainly because we also assume that the metallicity of the accreting stream is 10^{-3} of the solar value. The gas cannot cool much beyond 10⁴ K due to it lacking heavy metals (and is likely heated by the meta-galactic flux and the ionizing field of the galaxy embedded in the halo). This assumption, although naive, is also extremely conservative in that this implies the post-shock gas will have one of the longest possible radiative cooling time (see Sutherland & Dopita 1993; Gnat & Sternberg 2007). We use the cooling curve, $\Lambda(T)$, from Gnat & Sternberg (2007) to compute the cooling times in the post-shock gas. We assume that the temperature of the gas in the stream before passing through the shock is also 10^4 K (T_1) . At those temperatures and very low metallicity, we assume that no molecules form, so that the adiabatic index of the gases is always that of a monatomic gas, namely $\gamma = 5/3$.

The parameters we use in the model, given our assumed mass and redshift are given in Table A.1. We enumerate for completeness and clarity all variables used in our analysis in Table A.2.

Table A.2. Model variables and their relationships † .

X7	C11	To die
Variable Name	Symbol	Equation
Temperature floor	T_0	
Initial speed of sound	c_1	$=\sqrt{\gamma k_{\mathrm{B}}T_{\mathrm{0}}/\mu m_{\mathrm{p}}}$
Incoming Mach number	\mathcal{M}_1	$=v_{\rm vir}/c_1$
Density of the halo gas at r_{vir}	$ ho_{ m H}$	$= \rho_{\text{NFW}}(r_{\text{vir}}) \approx 37.0 \times f_{\text{B}}\rho_{\text{crit}}(z)$
Temperature of the halo gas	$T_{ m H}$	$= T_{\rm vir} = \mu m_{\rm p} v_{\rm vir}^2 (\gamma - 1)/2k_{\rm B}$
Pressure of the halo gas at r_{vir}	$P_{ m H}$	$= k_{\rm B} T_{\rm H} \rho_{\rm H} / \mu m_{\rm p}$
Density of the post-sock gas ^a	$ ho_{ m ps}$	$= (\gamma + 1) / (\gamma - 1) \mathcal{M}_1^2 / \left[\mathcal{M}_1^2 + 2/(\gamma - 1) \right] \rho_0$
Pressure of the post-shock gas ^a	$P_{ m ps}$	$= (\gamma - 1) / (\gamma + 1) \left[2\gamma / (\gamma - 1) \mathcal{M}_1^2 - 1 \right] P_0$
Post-shock speed of sound	$c_{ m ps}$	$= \gamma P_{\rm ps}/\rho_{\rm ps}$ $= T_0$
Temperature of the warm phase	$T_{ m w}$	$=T_0$
Density of the warm phase	$ ho_{ m w}$	$= ho_{ m H} T_{ m H} / T_{ m w}$
Density of the hot phase (post-expansion)	$ ho_{ m h}$	$= \rho_{\rm ps} \left(P_{\rm ps} / P_{\rm H} \right)^{1/\gamma}$
Temperature of the hot phase (post-expansion)	$T_{ m h}$	$= \rho_{\rm H} T_{\rm H} / \rho_{\rm h}$ $= \phi_{\rm v,w} \rho_{\rm w}$
Volume-averaged density of the warm phase	$\langle \rho_{\rm w} \rangle_{\rm v}$	$=\phi_{\mathrm{v,w}}\rho_{\mathrm{w}}$
Volume-averaged density	$ ho_2$	$= \phi_{v,w} \rho_w + (1 - \phi_{v,w} \rho_h)$
Expansion factor of the post-shock stream ^b	S	$= \left(1 - \sqrt{\Delta}\right) / \left[(\gamma - 1) / f + 2\eta \right]$
Velocity dispersion of the warm clouds	$\sigma_{ m turb}$	$= \sqrt{2k_{\rm B}T_{\rm w}\eta f/(\gamma-1)\phi_{\rm v,w}\mu m_{\rm p}}$
Halo dynamical time	t _{dyn,halo}	$= r_{\rm vir}/v_{\rm vir}$
Cooling time of the phase $\Phi \in \{ps, w, h\}$	$t_{\mathrm{cool},\Phi}$	$=k_{\rm B}\mu m_{\rm p}T_{\rm \Phi}/\rho_{\rm \Phi}\Lambda\left(T_{\rm \Phi}\right)$
Expansion time of the post-shock gas	$t_{\rm expand}$	$=2\gamma r_{\rm stream}/c_{\rm ps}$
Disruption time of the turbulent warm phase	$t_{ m disrupt}$	$= r_{\rm stream}/\sigma_{ m turb}$
Isobaric cooling length $\Phi \in \{ps, w, h\}$	$\lambda_{ m cooling}$	$=c_{\Phi}t_{\mathrm{cool},\Phi}$

Notes.^(†) A graphical representation of many of these variables is shown in Fig. 1.
^(a) Standard normal shock equation from the Rankine-Hugoniot jump conditions.
^(b) The equation assumes $\Delta = 1 - 4 \left[\eta f + (\gamma - 1)/2 \right] \rho_{\rm H}/\rho_2$