

# Are cosmological gas accretion streams multiphase and turbulent?

Nicolas Cornuault<sup>1</sup>, Matthew D. Lehnert<sup>1</sup>, François Boulanger<sup>2,3</sup>, and Pierre Guillard<sup>1</sup>

<sup>1</sup> Sorbonne Universités, UPMC Paris VI, CNRS UMR 7095, Institut d’Astrophysique de Paris, 98bis bd Arago, 75014 Paris, France  
e-mail: cornuault@iap.fr; lehnert@iap.fr

<sup>2</sup> Institut d’Astrophysique Spatiale, CNRS UMR 8617, Université Paris-Sud 11, Bâtiment 121, Orsay, France

<sup>3</sup> Research associate at the Institut d’Astrophysique de Paris, 98bis bd Arago, 75014 Paris, France

Received 30 June 2016 / Accepted 20 October 2017

## ABSTRACT

Simulations of cosmological filamentary accretion reveal flows (“streams”) of warm gas,  $T \sim 10^4$  K, which bring gas into galaxies efficiently. We present a phenomenological scenario in which gas in such flows, if it is shocked as it enters the halo as we assume and depending on the post-shock temperature, stream radius, its relative overdensity, and other factors, becomes biphasic and turbulent. We consider a collimated stream of warm gas that flows into a halo from an overdense 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, expansion of the post-shock gas, amount of turbulence in the clouds, and 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 mainly occurs in the mass range,  $M_{\text{halo}} \sim 10^{11}$  to  $10^{13} M_{\odot}$ , where the bulk of stars have formed in galaxies. Because of the expansion of the stream and turbulence, gas accreting along cosmic web filaments may eventually lose 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. Decollimating flows make the direct interaction between galaxy feedback and accretion streams more likely, thereby further reducing the overall accretion efficiency. As we discuss in this work, moderating the gas accretion efficiency through these mechanisms may help to alleviate a number of significant challenges in theoretical galaxy formation.

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

## 1. Introduction

The realization that we live in a dark matter-dominated universe 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 of these problems 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 overcooling problem (Benson et al. 2003).

As numerical simulations allowed for galaxy growth to be coupled to the development of large-scale structure, these studies showed 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” (Dekel & Birnboim 2006).

This cold mode accretion in simulations occurs as streams of warm ( $10^4$  K) gas entering the halo are smoothed 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 overcooling 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, thereby lowering their total gas content.

The circumgalactic 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-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 motivation to analyze gas accretion flows phenomenologically developed from the notion that overcooling remains a problem in simulations, there is scant observational evidence for the streams of the type currently simulated, and 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).

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 leads 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  $z \sim 2$  of  $\sim 50\%$  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 determine whether or not microphysical gas 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, making a quantitative investigation of the impact of expansion in the post-shock gas on an accretion flow, discussing the case in which the post-shock gas has small fluctuations in density and temperature, and developing a criterion for when the gas fragments (Sect. 2). We also discuss the thermodynamic evolution of the streaming gas and how it becomes multiphased in Sects. 2.3 and 2.4. In Sect. 3, we discuss the consequences of the formation of a multiphase flow on the dynamical evolution of the stream after it penetrates the halo. To gauge the astrophysical pertinence of our model and to illustrate specific phenomena within the streaming gas, we

analyze the physical characteristics of an idealized gas accretion stream penetrating a halo of  $10^{13} M_{\odot}$  at  $z = 2$  throughout Sects. 2 and 3. In Sect. 4, 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.

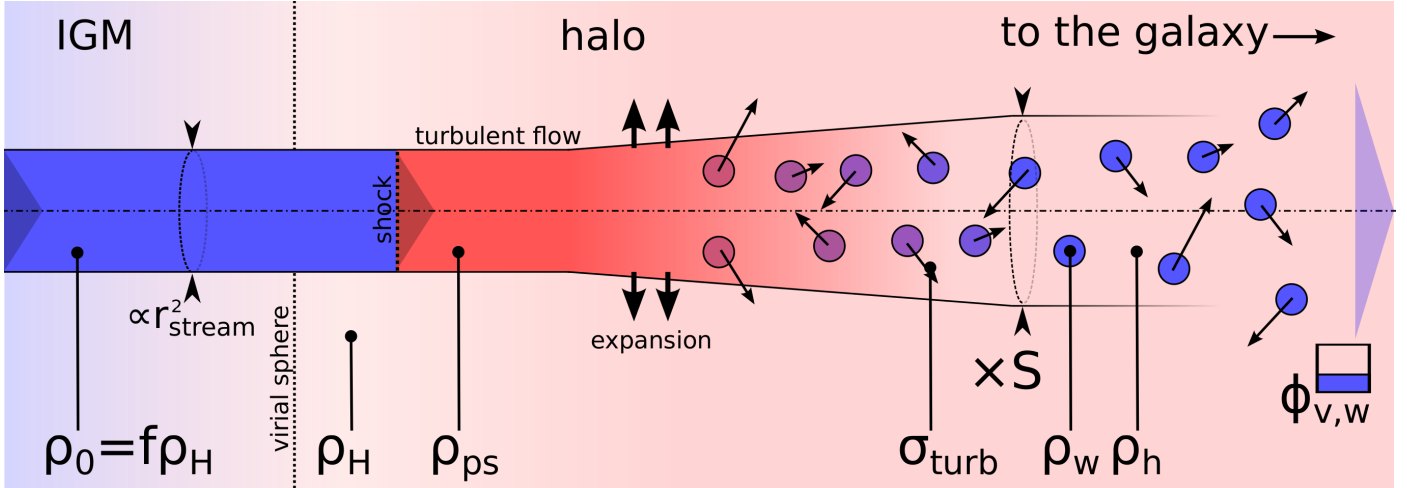
## 2. Our framework for multiphased streams

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). Gas instabilities are 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). Ambient gas in halos and outflows from galaxies must often meet the necessary requirements for instability 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 a balance between 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  $10^4$  K that penetrates into a dark matter halo filled with hot gas at the halo virial temperature. The density distribution of the hot gas smoothly 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 (Sect. 2.2). The post-shock gas is overpressurized 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 develops 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 cools and fragments, forming a two phase medium – a hot phase with an embedded warm cloudy phase (Sect. 2.2). Density and velocity inhomogeneities in the post-shock gas may be amplified through gas cooling, leading to the formation of a multiphased flow (see Sect. 2.3). The thermodynamic evolution of the hot and warm gas follow distinct paths in the density-temperature plane (Sect. 2.4).



**Fig. 1.** Sketch of our phenomenological picture of flows of gas passing through a virial shock. The initial inflowing gas (in blue) is overdense relative to the halo at the boundary by a factor  $f$  ( $=\rho_0/\rho_H$  or the stream density divided by the density of the ambient halo gas at the virial radius), shocks at the boundary between the intergalactic medium (labeled “IGM”) and the hot halo gas (labeled “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 “expansion”). If certain conditions are met, the post-shock gas becomes 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,  $\sigma_{\text{turb}}$ , and a volume-filling factor,  $\phi_{v,w}$ . If the level of turbulence is high enough, the clouds move beyond the initial radius of the stream, decollimating 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.

We expect that part of the kinetic energy of the infalling gas is converted to turbulence within the warm clouds and random cloud-cloud motions (e.g., Hennebelle & Péroul 1999; Kritsuk & Norman 2002; Heigl et al. 2018). 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 Sect. 3). If these conditions are met, the stream decollimates. The thermodynamic evolution of the post-shock gas in the stream is largely determined by the timescales of several relevant processes – gas 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. 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 determine whether 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 (DB06; Mo et al. 2010). In our study, we reconsidered the sustainability of the accretion shock that occurs as the infalling filament collides with the ambient halo gas. To conduct this analysis, we compared the cooling time of the post-shock gas to various other measures of its dynamical evolution. For simplicity, we only compared 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 quickly cools down and main-

tain 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 mixes 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). In this paper, we describe this more complex regime phenomenologically.

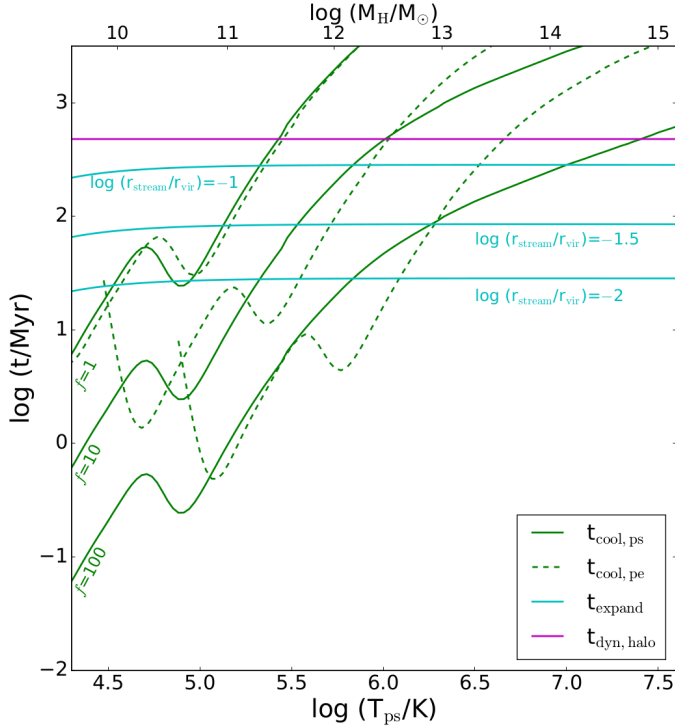
We now consider the impact of expansion on the properties of the post-shock gas. To estimate the cooling, expansion, and dynamical timescales, we introduce two parameters specific to the flow, its overdensity relative to that of mean density of gas at the virial radius,  $f$ , and the radius of the stream,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

$$P_H = \left( \frac{k_B}{\mu m_p} \right) \rho_H T_H = \frac{(\gamma - 1)}{2} \rho_H v_{\text{vir}}^2 = \frac{\gamma \mathcal{M}_1^2}{3f} P_0, \quad (1)$$

where  $T_H$  is the temperature of the ambient halo gas, which we assume to be the virial temperature. The pressure ratio just after the shock is

$$\frac{P_{\text{ps}}}{P_H} = \frac{2f}{\gamma(\gamma + 1) \cdot g(\mathcal{M}_1)}, \quad (2)$$

where  $g : \mathcal{M}_1 \mapsto (2\gamma/(\gamma-1) - \mathcal{M}_1^{-2})^{-1}$ . For  $\gamma = 5/3$  and within the domain  $[1, +\infty]$ , this function rapidly decreases from 1:4 to 1:5.



**Fig. 2.** Timescale comparison of the competing relevant processes of cooling, expansion, and halo dynamics, for various post-shock temperatures (i.e., various halo masses), at  $z = 2$ . The solid green curves indicate the instantaneous post-shock cooling time,  $t_{\text{cool,ps}}$ , and the green dashed curves indicate the post-expansion cooling time,  $t_{\text{cool,pe}}$  – the cooling time of the post-shock gas 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,  $10^4$  K, of the cooling efficiency used in our analysis (Gnat & Sternberg 2007). The cooling time curves were computed for three values of the initial overdensity  $f$ , as indicated by inclined green labels on the left side of the panel. Expansion times (solid cyan lines) are shown for three different relative filament radii,  $\log r_{\text{stream}}/r_{\text{vir}} = -1, -1.5,$  and  $-2$ . The dynamical time,  $t_{\text{dyn,halo}}$ , which is independent of halo mass, is indicated by the horizontal magenta line.

Hence the pressure ratio,  $P_{\text{ps}}/P_{\text{H}}$  is between 9:5 and 9:4 times the initial stream overdensity,  $f$ . Since  $f \geq 1$ , the pressure of the post-shock gas is higher than the pressure of the ambient halo gas. Thus, since the pressure ratio is always greater than one, the flow expands into the 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 written as

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

where  $r_{\text{stream}}$  is the radius of the stream before it expands and  $c_{\text{ps}}$  is the sound speed of the gas immediately after it has passed through the shock front. As a first-order approximation, for 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

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

where  $k_{\text{B}}$  is the Boltzmann constant,  $\Lambda(T)$  is the cooling efficiency as a function of temperature,  $T$  (Sutherland & Dopita 1993; Gnat & Sternberg 2007),  $\mu$  is the mean molecular weight<sup>1</sup>,  $m_{\text{p}}$  is the mass of the proton,  $\rho$  is the mass density of the cooling gas,  $n_{\text{e}}$  is the electron density, and  $n_{\text{H}}$  the hydrogen particle density<sup>2</sup>. Hereafter, the cooling time of the post-shock is denoted by,  $t_{\text{cool,ps}}$ . We also refer to the cooling time after expansion as  $t_{\text{cool,pe}}$ .

The other timescale with which to compare the cooling time is the dynamical time of the halo. 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

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

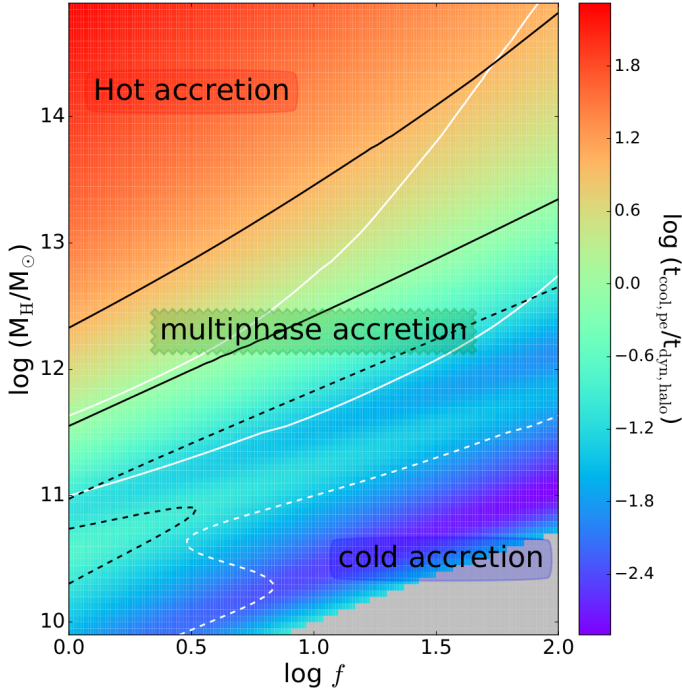
For this estimate, we assume a NFW dark matter potential (Navarro et al. 1997) and  $\alpha = 1$ .  $\alpha$  is a factor of order unity to account for the integrated gravitational acceleration during the fall. Assuming that the characteristic velocity and radius are the virial values,  $v_{\text{vir}}$  and  $r_{\text{vir}}$ , the dynamical time is independent of halo mass (Mo et al. 2010).

Figure 2 illustrates how these timescales depend on the post-shock temperature that are directly related to the halo mass (abscissa at the top) and for a range of relative stream overdensities,  $f$ , 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 ionization by the meta-galactic flux or photons from the galaxy at the center of the halo potential. From this analysis, we see that the expansion times are almost constant above  $M_{\text{H}} \gtrsim \text{few } 10^{10} M_{\odot}$  since the virial shock is sufficiently strong such that  $t_{\text{expand}} \propto t_{\text{dyn,halo}}$ . As  $f$  increases, the 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 overdense 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 cools globally neither purely adiabatically or isobarically.

We can understand the competition between these timescales,  $t_{\text{cool,ps}}$ ,  $t_{\text{cool,pe}}$ ,  $t_{\text{dyn,halo}}$ , and  $t_{\text{expand}}$ , more clearly by considering them as a function of the stream overdensity,  $f$ , and post-shock temperature or halo mass (Fig. 3). For the case where  $t_{\text{cool,ps}} \gg t_{\text{expand}}$  and  $t_{\text{cool,pe}} > t_{\text{dyn,halo}}$ , the post-shock gas expands rapidly remaining hot over at least a dynamical time of the halo. In this regime the hot post-shock gas likely mixes 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  $t_{\text{cool,ps}} \ll t_{\text{dyn,halo}}$ , the gas cools before there is significant expansion and before the stream has penetrated deeply into the halo. Accretion in this regime corresponds to their cold mode accretion and we labeled this as such in Fig. 3. Between these two regimes,  $t_{\text{cool,pe}}$ ,  $t_{\text{cool,ps}}$ ,  $t_{\text{dyn,halo}}$ , and  $t_{\text{expand}}$

<sup>1</sup>  $\mu = 0.6 m_{\text{p}}$  for a fully ionized gas.

<sup>2</sup> The factor 4.3 in Eq. (4) comes from  $n_{\text{e}} = 1.2 n_{\text{H}}$  and  $n_{\text{H}} \approx \rho/(1.4 m_{\text{p}})$ .

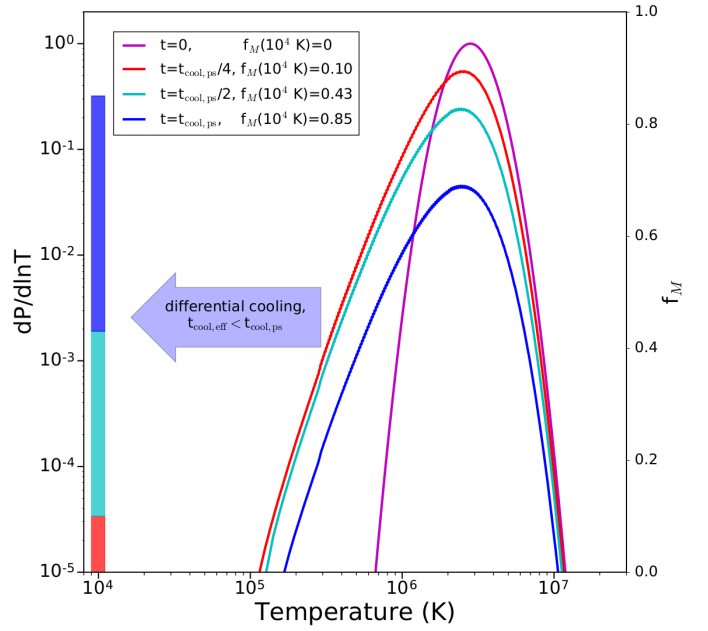


**Fig. 3.** Ratio of the post-expansion cooling and halo dynamical timescales as a function of halo mass,  $M_H$ , and stream overdensity,  $f$ , at  $z = 2$ . The colors represent  $\log(t_{\text{cool,pe}}/t_{\text{dyn,halo}})$  (Eqs. (4) and (5)) with corresponding values indicated in the color bar. Black contours indicate  $\log(t_{\text{cool,pe}}/t_{\text{dyn,halo}}) = 0$  and 1 (solid) and  $\log(t_{\text{cool,pe}}/t_{\text{dyn,halo}}) = -1$  (dashed). The white contours indicate values of  $\log(t_{\text{cool,pe}}/t_{\text{expand}})$  (Eqs. (3) and (4)) equal to 0 and 1 (solid), and  $-1$  (dashed). For this analysis, we assumed  $\log(r_{\text{stream}}/r_{\text{vir}}) = -1.5$ . We highlight regions in which the cooling times are significantly longer than the halo dynamical times as hot accretion, regions in which the cooling time is significantly shorter than the dynamical times as cold accretion, and regions in which 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 separation of accretion modes. The gray region in the bottom right corner has  $T_{\text{pe}} < 10^4$  K, which is outside of the range of the cooling efficiency used in our analysis (Gnat & Sternberg 2007), and thus the values of  $\log(t_{\text{cool,pe}}/t_{\text{dyn,halo}})$  are not defined.

are all about the same order of 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. Over this region, the thermodynamic evolution of the post-shock gas is more complicated. We labeled this regime “multiphase accretion”.

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

We now discuss the physical mechanisms behind the development of multiphase 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 streams are not accreted homogeneously but have substructure in both density and velocity (Nelson et al. 2016). Inhomogeneities may arise owing 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



**Fig. 4.** Temperature probability distribution function,  $dP/d\ln T$ , as a function of the gas temperature at various times during isobaric cooling. The initial distribution (magenta line) is lognormal with a dispersion,  $\sigma_T = 0.3$ , initial mean gas temperature,  $T_{\text{ps}} = 2.7 \times 10^6$  K, and initial pressure,  $P_{\text{ps}}/k_B = 5.8 \times 10^4$  K cm $^{-3}$ . We show the distributions of gas temperature at 4 different times relative to  $t_{\text{cool,ps}}$  as indicated in the legend. In the last column of the legend, we provide the fraction of the gas that has cooled to  $10^4$  K,  $f_M$ , at each fraction of the post-shock cooling time. For example, in this illustration, after one post-shock cooling time, even though the peak gas temperature of the PDF has changed very little,  $\approx 85\%$  of the gas has cooled to  $10^4$  K. We represent these fractions in the colored bar in which each color corresponds to the PDF of the same color (the fractions,  $f_M$  are relative to the ordinate on the right side of the panel). Once the gas starts to cool differentially, it does so very rapidly, in less than  $t_{\text{cool,ps}}$  (large blue arrow).

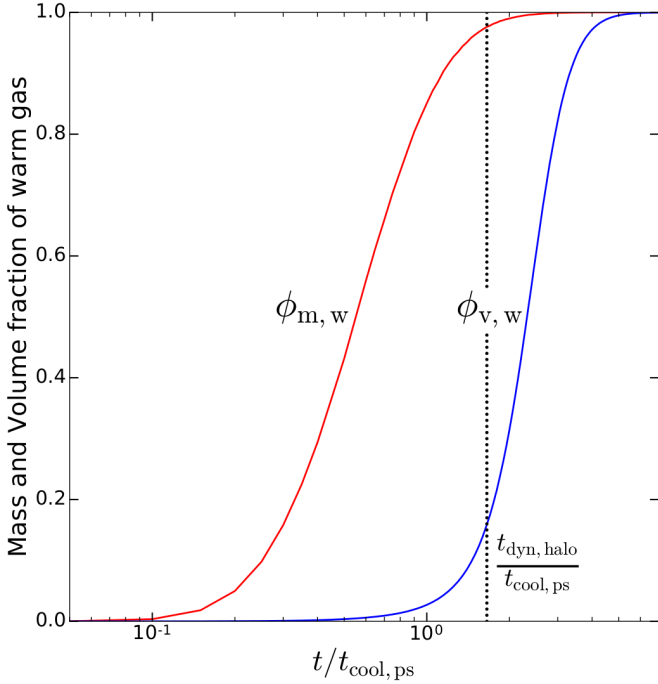
velocity also lead to inhomogeneities in the post-shock gas in both temperature and density resulting in a range of cooling times (Guillard et al. 2009). As we now discuss, once such “differential cooling” sets in, it acts like the thermal instability, leading to phase separation in the flow (Sharma et al. 2012b), but over a finite range of time in the absence of heating processes that balance the 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 presented in Sharma et al. (2012b), we begin with a parcel of gas that is overdense relative to an ambient medium. The magnitude of the overdensity is,  $\delta \equiv |\delta\rho/\rho| \sim |\delta T/T|$  for isobaric conditions. The inverse of the effective cooling time of the overdense gas parcel relative to the ambient medium is, hence,

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

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

$$\frac{1}{t_{\text{cool,eff}}} \sim -\delta \frac{\partial}{\partial \ln T} \left( \frac{1}{t_{\text{cool}}}_p \right) = \frac{\delta}{t_{\text{cool}}} \left[ 2 - \frac{d \ln \Lambda(T)}{d \ln T} \right]. \quad (7)$$



**Fig. 5.** 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 (Eqs. (9) and (10)). The timescale is expressed in units of the isobaric cooling time,  $t_{cool,ps}$ , as defined in the text. The initial values of the pressure and temperature are the same as those used in Fig. 4. We indicate the ratio of the halo dynamical and post-shock cooling times as a vertical black dotted line. This suggests that with our initial conditions, the stream has sufficient time to fragment significantly, but not reach a high volume-filling factor of warm gas. This suggests that even highly fragmented streams with high  $\phi_{m,w}$  are challenging to identify and study via QSO absorption lines.

Marginally denser gas cools faster than the gas in which it is embedded and the cooling becomes a runaway process for temperatures  $\geq 10^5$  K and around few  $10^4$  K at low metallicities (right panel in Fig. 6). 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 cool rapidly and form clouds, leaving a high volume-filling inter-cloud medium that is hotter and rarer than the clouds. Thus, differential cooling leads to multiphase 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 heating processes of the hot gas to balance its 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 multiphase medium.

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,

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

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 lognormal PDF in temperature with a dispersion of  $\sigma_T = 0.3$  ( $1\sigma$  dispersion in  $\log \frac{\rho}{\rho_{ps}} = \log \frac{T}{T_{ps}}$ ). We chose a gas pressure and mean temperature appropriate for the post-shock gas of a flow into a massive,  $10^{13} M_\odot$ , halo at  $z = 2$ . For this illustration, the gas pressure is  $P_{ps}/k_B = 5.8 \times 10^4 \text{ K cm}^{-3}$  and the initial post-shock temperature  $T_{ps} = 2.7 \times 10^6 \text{ K}$ . We find that a substantial fraction of the accreting gas cools to  $10^4$  K in a fraction of  $t_{cool,ps}$ , while the peak of the hot gas temperature PDF does not shift with time (Fig. 4).

We show in Fig. 5 both the mass fraction and volume-filling factor of the gas for the warm,  $10^4$  K gas as a function of time. We calculated the mass fraction by integrating the amount of gas that cools to  $10^4$  K as a function of time,

$$\phi_{m,w} = \int_0^{10^5 \text{ K}} \frac{dm}{dT} dT / \int_0^{T_{max}} \frac{dm}{dT} dT, \quad (9)$$

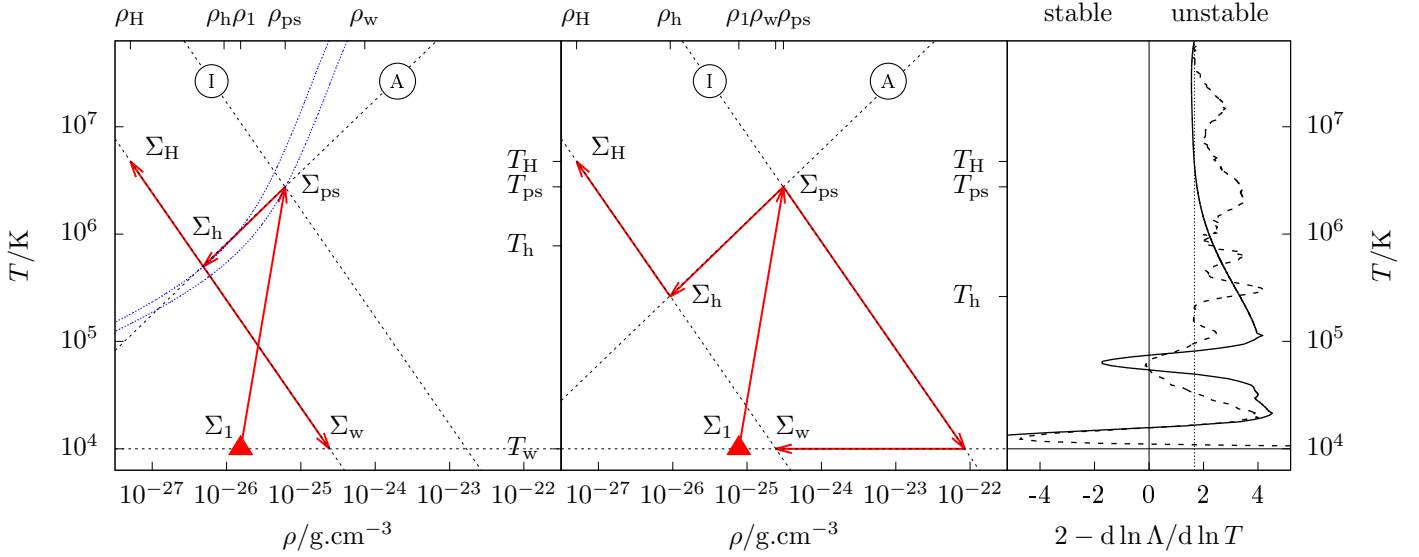
where  $dm$  is the gas mass between  $T$  and  $T + dT$ . The volume-filling factor was calculated using,

$$\phi_{v,w} = \int_0^{10^5 \text{ K}} \frac{1}{\rho} \frac{dm}{dT} dT / \int_0^{T_{max}} \frac{1}{\rho} \frac{dm}{dT} dT. \quad (10)$$

We note that the results shown on Figs. 4 and 5 are plotted as a function of the post-shock cooling time  $t_{cool,ps}$ . The timescale normalized in this way is, to first order, independent of the specific values of  $\rho_{ps}$  and  $T_{ps}$ . These results even apply to the gas cooling after expansion when  $t_{expand} < t_{cool,ps}$ . Figure 5 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 higher value of the temperature dispersion  $\sigma_T$ ,  $\phi_{m,w}$  increases over a broader range of fractional times distributed around the same time required for half the mass to cooling in our specific illustration.

In Fig. 5, the gas is multiphase for only a few  $t_{cool,ps}$ . 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 eventually mixes 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, cloud-cloud collisions, 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).

We ask the poignant question whether the flowing post-shock gas can 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 occurs if the post-shock conditions are sufficiently inhomogeneous. The gas becomes multiphased around when  $t_{cool,pe}$  is the same order of magnitude as  $t_{dyn,halo}$ . This justifies where we placed the “multiphase accretion” label in Fig. 3. This figure shows that this occurs in the important mass range of



**Fig. 6.** *Left and middle:* sketch of the thermodynamic path for cases (1) and (2) (see text).  $\Sigma_1$  (red triangle) indicates the density and temperature of the pre-shock gas. The gas is shocked reaching the point  $\Sigma_{ps}$ . Subsequently, the gas cools adiabatically due to the expansion of the stream until the halo pressure is reached at  $\Sigma_h$ . Two phases separate. For case (1), phase separation occurs at  $\Sigma_h$ , while for case (2), it occurs at  $\Sigma_{ps}$ . In both cases, the hot component mixes with the surrounding halo gas, reaching  $\Sigma_H$ . For case (1), the warm component cools radiatively and isobarically to the point  $\Sigma_w$ . 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* indicate cooling 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. (8)) 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 differential cooling if  $\delta$  is sufficiently large. The post-shock temperature of our illustration lies within the region where the stream can become unstable.

$M_{halo} \sim 10^{11}$  to  $10^{13} M_{\odot}$  at  $z = 2$ . The values of course depend on the relative overdensity of the streams ( $f$ ) and, through the characteristics of the halos, on redshift. The bulk of stars have formed in galaxies in halos within this mass range and  $z \approx 2$  is the epoch in which the peak of the comoving density of star formation occurs (Madau & Dickinson 2014, and references therein). Moreover, because the volume filling-factor of the warm gas is likely to always be small, even in the cases where the mass fraction is large, multiphase streams are likely to be difficult to identify observationally using QSO absorption lines. We discuss this further in Sect. 4.

#### 2.4. Thermodynamic evolution of streams: why are some streams cloudy?

There are two limiting cases to consider specifically when attempting to understand the thermodynamic development of a biphasic stream. The two cases are (1)  $t_{expand} \ll t_{cool,ps}$  and (2)  $t_{expand} \gg t_{cool,ps}$ . 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 (“paths”) of a stream in Fig. 6. To make these illustrations, we adopted  $f = 30$  and  $r_{stream}/r_{vir} = 0.01$  for case (1); and  $f = 150$  and  $r_{stream}/r_{vir} = 0.1$  for case (2) for a halo with a mass of  $10^{13} M_{\odot}$  at  $z = 2$ . For a halo of this mass, the post-shock gas temperature is such that, if the stream has a sufficiently high  $\delta$ , it becomes unstable and fragments into warm and hot phases (right panel in Fig. 6).

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 phases 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 limiting (simplest) cases, in reality, the gas has  $t_{expand}$  within an order of magnitude of  $t_{cool,ps}$  (Fig. 3). For these “intermediate” cases, the thermodynamic evolution is more complex, but, with our assumption that the hot gas cools adiabatically during expansion and the warm gas evolves isobarically, the clouds reach the same final thermodynamical state as illustrated in the two limiting cases. However, the gas in these intermediate cases does not necessarily end up with the same distribution of volume-filling factors and mass fractions.

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

The thermodynamic paths we sketched in Fig. 6 exhibit phenomena that are generic to the evolution of fragmenting streams. We only use the initial conditions of the stream to illustrate the quantitative evolution of the gas (i.e., the temperatures and densities and their evolution). Streams fragment for a wide range

of initial conditions and have a generally similar, but more complex, thermodynamic evolution.

### 3. 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,  $\eta$ , of the turbulent energy density of the clouds and the initial bulk kinetic energy density of the stream. We define this ratio as

$$\eta = \frac{\langle \rho_w \rangle_v \sigma_{\text{turb}}^2}{\rho_1 v_1^2}, \quad (11)$$

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  $\rho_w$  is the density of the warm clouds, respectively),  $\sigma_{\text{turb}}$  is the cloud-cloud velocity dispersion,  $\rho_1$  and  $v_1$  are the pre-shock gas density and velocity. We assume that  $v_1$  is equal to the virial velocity,  $v_{\text{vir}}$ . The initial density of the stream is related to the hot halo density,  $\rho_H$ , 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  $\eta$  which is free.

As we briefly alluded to in Sect. 2.3, as the turbulent energy dissipates, it reheats the warm clouds, thereby moderating the cooling. The cooling of the clouds is also 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 and the radiation from the high-speed accretion shock (Allen et al. 2008). The parameter  $\eta$  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,  $\eta$  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 is not 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

$$S \rho_2 v_2 = \rho_1 v_1, \quad (12)$$

where  $\rho_2$  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

ultimately reaches pressure equilibrium with the halo, which implies  $\rho_w T_w = \rho_h T_h = \rho_H T_H$ , where  $\rho_w$  and  $\rho_h$  are the densities of the warm and hot components and  $T_H$  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

$$S (\rho_2 v_2^2 + \eta \rho_1 v_1^2 + P_H) = \rho_1 v_1^2 + P_1, \quad (13)$$

where  $P_1$  is the initial pressure of the stream. We assume the fragmenting gas radiates away its heat until reaching a floor temperature,  $T_w = 10^4$  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,  $\rho_h^\gamma P_{\text{ps}} = \rho_{\text{ps}}^\gamma P_H$ , where  $\gamma$  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), (12), and (13), as

$$S = \frac{1}{2} \left( \eta + \frac{\gamma - 1}{2f} \right)^{-1} \left( 1 - \sqrt{1 - 4 \frac{\rho_H}{\rho_2} \left( \eta f + \frac{\gamma - 1}{2} \right)} \right). \quad (14)$$

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 Sect. 2, it is the relative ratios of the thermal cooling time and expansion time that influences how the stream evolves.

#### 3.3. Disruption of cloudy streams

The relative cloud-cloud motions may lead to warm clouds spreading beyond the initial 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 decollimate. 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,

$$t_{\text{disrupt}} = \frac{r_{\text{stream}}}{\sigma_{\text{turb}}}, \quad (15)$$

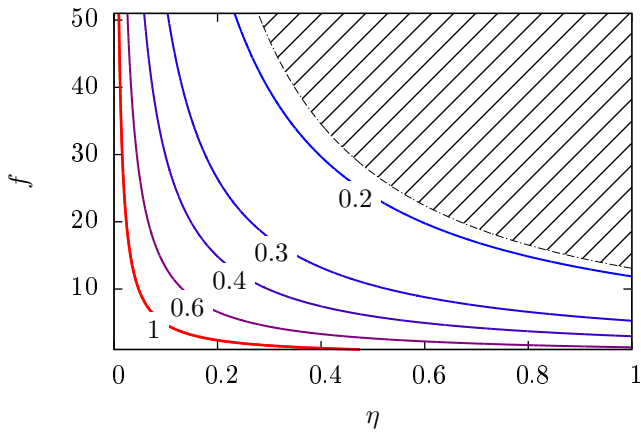
where  $\sigma_{\text{turb}}$  may be computed from  $\eta$ ,  $f$ , and the volume-filling factor of the warm clouds,  $\phi_{v,w}$ . From Eqs. (1) and (11), assuming pressure equilibrium between gas phases, yields

$$\begin{aligned} \sigma_{\text{turb}}^2 &= v_1^2 (\eta f / \phi_{v,w}) (T_w / T_H) \\ &= \frac{2}{(\gamma - 1)} \left( \frac{k_B}{\mu m_p} \right) (\eta f / \phi_{v,w}) T_w. \end{aligned} \quad (16)$$

Characterized this way,  $\sigma_{\text{turb}}^2 / v_1^2 > \eta$ . For a wide range of relative amounts of turbulent energy, stream overdensity, 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 remain collimated as they flow, otherwise, the streams decollimate.

This is the only process we consider in determining whether or not the warm clouds travel coherently toward the galaxy





**Fig. 7.** Disruption of the flow as a function of the filament overdensity,  $f$ , and level of turbulence,  $\eta$ . The contours represent constant ratios of  $t_{\text{disrupt}}/t_{\text{dyn,halo}}$  as labeled (cf. Eqs. (5) and (15)). We assume a volume-filling factor of 0.1 (see Fig. 5) for the warm clouds and  $r_{\text{stream}} = r_{\text{vir}}/10$ . In regions with values less than 1, the streams are “disrupted” (Sect. 3.3). 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  $t_{\text{disrupt}} \propto \sigma_{\text{turb}}^{-1}$ , the contours of constant  $t_{\text{disrupt}}/t_{\text{dyn,halo}}$  are shaped like contours of constant  $\sigma_{\text{turb}}$  in the same plane. For example, the contour,  $t_{\text{disrupt}}/t_{\text{dyn,halo}} = 0.2$ , is close to the contour for  $\sigma_{\text{turb}} = 200 \text{ km s}^{-1}$ , which is almost half the initial velocity of the flow.

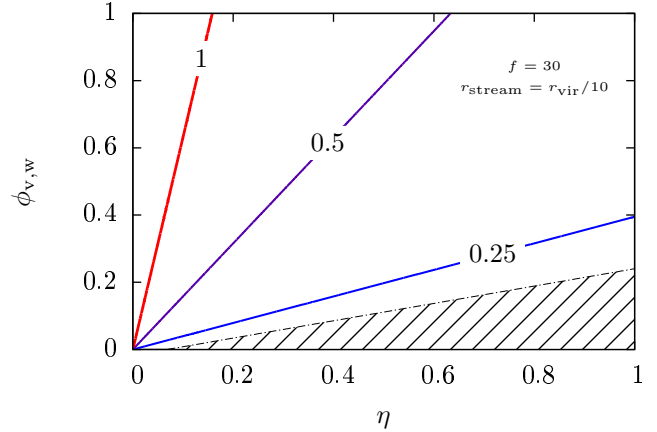
proper as observed in numerical simulations (e.g., Brooks et al. 2009; Danovich et al. 2015). To illustrate this concept, we estimate  $t_{\text{disrupt}}/t_{\text{dyn,halo}}$  as functions of  $f$  and  $\phi_{\text{v,w}}$  for a halo with a mass of  $10^{13} M_{\odot}$  at  $z = 2$ . (Figs. 7 and 8; Eqs. (5) and (15)). We find that the disruption time is shorter than or approximately equal to the halo dynamical time for this halo mass, redshift, and our assumed  $r_{\text{stream}}$ . Thus it appears that for a wide range of relative turbulent energy densities, stream overdensities, and volume-filling factors of the warm gas, the flows do 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).

## 4. Discussion

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

### 4.1. 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  $\text{km s}^{-1}$  (van de Voort & Schaye 2012; Goerdt & Ceverino 2015) or much greater than the sound speed of the stream,  $c_s \sim 10\text{--}20 \text{ km s}^{-1}$ . 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) showed that atomic shocks with velocities of



**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  $t_{\text{disrupt}}/t_{\text{dyn,halo}}$  as labeled (cf. Eqs. (5) and (15)). We assume  $f = 30$  and  $r_{\text{stream}} = r_{\text{vir}}/10$ . In regions with values less than 1, the streams are disrupted. The shaded region has the same meaning as in Fig. 7.

$\sim 100 \text{ km s}^{-1}$  reach their post-shock temperatures within a distance of  $\approx 1\text{--}10 \times 10^{15} \text{ cm}$  in less than  $\sim 30 \text{ yr}$ . For higher shock velocities, the spatial and temporal scales are 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, 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  $10^4 \text{ K}$  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 have taken advantage of numerical viscosity or 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). Viscosity spreads the shock over several resolution elements, thereby 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 were unchanged but the viscosity increased, the Reynolds number of the flow would be artificially low. Simulations with artificial or numerical viscosity have flows with low Reynolds numbers. Simulated low Re flows,  $\text{Re} \lesssim 1000$ , 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).

### 4.2. Nature of flows into galaxies: observational tests

As a consequence of our assumption that all energetic quantities scale as  $v_{\text{vir}}$ , the cloud-cloud dispersion is a simple linear

function of  $\eta$ ,  $f$ , and  $\phi_{v,w}^{-1}$ . 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 understood well, the two parameters,  $\eta$  and  $\phi_{v,w}$ , 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: clumpy, turbulent streams and 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, namely, that a large fraction of the bulk kinetic energy of the accretion flow is transferred to turbulent motions among cold clouds, is observed in large-scale galaxy colliding flows, such as the situation in the Taffy Galaxies or 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 post-shock gas, most of the 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 colliding flows 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, which may explain why streams have not been conspicuously identified so far. This is most obviously seen in Fig. 5, where even though a large fraction of the gas is warm, its volume-filling factor is, for a significant fraction of the dynamical time, minuscule. Along most lines of sight, absorption-line 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, a process that depends on the square of the density, since the density increases during differential cooling, the emission is enhanced. Their emission can be powered by the meta-galactic flux, UV radiation from the central galaxy of the halo, and loss of their gravitational potential energy as they fall into the halo. These processes have already been modeled in simulations (e.g., [Goerdt et al. 2010](#); [Rosdahl & Blaizot 2012](#); [Cen & Zheng 2013](#)). In our scenario, streams can be additionally powered by the localized dissipation of turbulent energy, shock heating through cloud-cloud collisions, and radiation from the shock front. This may have already been observed in Ly $\alpha$  (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](#)).

#### 4.3. 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: disruption and fragmentation of the flow and 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 decollimation. Any decollimation 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  $z = 2$ , depending on the level of inhomogeneities, post-shock temperature, stream radius, and stream overdensity. 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, thereby 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 AGNs 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 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 ([Hayes et al. 2016](#)). Thus, the generation of turbulent cloudy media in both accretion flows and starburst-driven outflows allows for efficient interaction between these two types of flows. This dynamical interaction likely sustains turbulence in the halo, which compensates for 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. Other mechanisms, such as 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).

## 5. 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 throughout the text 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 the following results:

- The post-shock gas expands into the halo gas and can fragment due to differential cooling and hydrodynamic instabilities if it satisfies a number of conditions. Instabilities lead to the formation of a biphasic flow. The formation of a hot post-shock phase mixes with the ambient hot halo gas, ultimately limiting how much of the gas can cool. As a result of the phase separation and 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 manuscript.
- The development of a biphasic medium converts some of the bulk kinetic energy into random turbulent motions in the gas (e.g., Hennebelle & Péroul 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. Owing to this transfer of momentum and energy, after passing through the shock, the fragmented and turbulent stream has a lower bulk velocity.
- For a wide range of turbulent energy densities, our model for a fiducial halo shows that the stream loses 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 multiwavelength observations useful for constraining the properties of turbulent astrophysical flows are necessary.

The post-virial shock gas is not isothermal, and accretion streams are both hot and cold. Whether or not a stream fragments after the virial shock depends on how inhomogeneous it is, its average overdensity relative to the halo gas, its radius, and whether or not the post-shock gas has a temperature that makes it potentially differentially unstable. We found that at  $z = 2$ , fragmentation is most important over the halo mass range,  $10^{11}$  to

$10^{13} M_{\odot}$ . This redshift and halo mass range is where the ensemble of galaxies is most rapidly increasing their stellar masses (e.g., Madau & Dickinson 2014, and references therein). Thus “hot-cold dichotomy” (see DB06) is no longer a simple function of whether or not the shock is stable, but now relies on under what circumstances the post-shock gas becomes multiphase and turbulent. However, what 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. This last notion needs more theoretical investigation.

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 inflow 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. Thus, 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 Île-de-France. N.C. 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 significantly improve our manuscript.

## 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., Lehnert, M. D., & Péroul, C. 2006, *MNRAS*, **367**, L16
- Bouché, N., Finley, H., Schroetter, I., et al. 2016, *ApJ*, **820**, 121
- 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
- Cen, R., & Zheng, Z. 2013, *ApJ*, **775**, 112
- 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
- Goerdt, T., Moore, B., Read, J. I., & Stadel, J. 2010, *ApJ*, **725**, 1707
- Gressel, O. 2009, *A&A*, **498**, 661
- Guillard, P., Boulanger, F., Pineau Des Forêts, G., & Appleton, P. N. 2009, *A&A*, **502**, 515
- Guillard, P., Boulanger, F., Cluver, M. E., et al. 2010, *A&A*, **518**, A59
- 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., & Thompson, T. A. 2017, ArXiv e-prints [[arXiv:1701.09062](https://arxiv.org/abs/1701.09062)]
- Heckman, T. M., Armus, L., & Miley, G. K. 1990, *ApJS*, **74**, 833
- Heigl, S., Burkert, A., & Gritschneider, M. 2018, *MNRAS*, **474**, 4881
- 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
- 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., & Norman, M. L. 2002, *ApJ*, **569**, L127
- Kritsuk, A. G., Nordlund, Å., Collins, D., et al. 2011, *ApJ*, **737**, 13
- 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
- Lu, Y., Mo, H. J., & Wechsler, R. H. 2015, *MNRAS*, **446**, 1907
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- 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**, 1025
- Mo, H., van den Bosch, F. C., & White, S. 2010, *Galaxy Formation and Evolution* (Cambridge, UK: Cambridge University Press)
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493
- Nelson, D., Vogelsberger, M., Genel, S., et al. 2013, *MNRAS*, **429**, 3353
- Nelson, D., Genel, S., Vogelsberger, M., et al. 2015, *MNRAS*, **448**, 59
- Nelson, D., Genel, S., Pillepich, A., et al. 2016, *MNRAS*, **460**, 2881
- Ocvirk, P., Pichon, C., & Teyssier, R. 2008, *MNRAS*, **390**, 1326
- 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, J. R., Kahn, S. M., Paerels, F. B. S., et al. 2003, *ApJ*, **590**, 207
- Peterson, B. W., Appleton, P. N., Helou, G., et al. 2012, *ApJ*, **751**, 11
- 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
- Rosdahl, J., & Blaizot, J. 2012, *MNRAS*, **423**, 344
- Salomé, P., Combes, F., Revaz, Y., et al. 2011, *A&A*, **531**, A85
- Sharma, P., Parrish, I. J., & Quataert, E. 2010, *ApJ*, **720**, 652
- 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
- Singh, A., & Sharma, P. 2015, *MNRAS*, **446**, 1895
- Suresh, J., Bird, S., Vogelsberger, M., et al. 2015, *MNRAS*, **448**, 895
- Sutherland, R. S., & Dopita, M. A. 1993, *ApJS*, **88**, 253
- Sutherland, R. S., Bicknell, G. V., & Dopita, M. A. 2003, *ApJ*, **591**, 238
- Thompson, T. A., Quataert, E., Zhang, D., & Weinberg, D. H. 2016, *MNRAS*, **455**, 1830
- Tillson, H., Devriendt, J., Slyz, A., Miller, L., & Pichon, C. 2015, *MNRAS*, **449**, 4363
- 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

**Table A.1.** Halo and gas characteristics for  $M_{\text{H}} = 10^{13} M_{\odot}$  at  $z = 2$ .

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

## 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  $M_{\text{H}}$ ,  $v_{\text{vir}}$ , and  $r_{\text{vir}}$ . 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  $T_{\text{H}}$ , which we assume to be equal to the virial temperature of the halo,  $T_{\text{vir}}$ . The density of the hot halo,  $\rho_{\text{H}}$ , 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,  $f_{\text{B}} = 0.18$ . This is  $\approx 37 f_{\text{B}} \rho_{\text{crit}}$ , where  $\rho_{\text{crit}}$  is the critical density of the Universe. The halo pressure,  $P_{\text{H}}$ , is related to  $T_{\text{H}}$  and  $\rho_{\text{H}}$ . 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).

The gas accretes through a stream of infalling gas with radius,  $r_{\text{stream}}$ , 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.,  $\rho_{\text{H}}$ ). We further assume that there is a temperature floor in the post-shock gas of  $10^4$  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  $10^4$  K because this gas lacks heavy metals and is likely heated by the meta-galactic flux and 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 has one of the longest possible radiative cooling times (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, such 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.

Variable name	Symbol	Equation
Temperature floor	$T_0$	$= 10^4 \text{ K}$
Initial speed of sound	$c_1$	$= \sqrt{\gamma k_B T_0 / \mu m_p}$
Incoming Mach number	$\mathcal{M}_1$	$= v_{\text{vir}} / c_1$
Density of the halo gas at $r_{\text{vir}}$	$\rho_H$	$= \rho_{\text{NFW}}(r_{\text{vir}}) \approx 37.0 \times f_B \rho_{\text{crit}}(z)$
Temperature of the halo gas	$T_H$	$= T_{\text{vir}} = \mu m_p v_{\text{vir}}^2 (\gamma - 1) / 2 k_B$
Pressure of the halo gas at $r_{\text{vir}}$	$P_H$	$= k_B T_H \rho_H / \mu m_p$
Density of the post-shock gas <sup>a</sup>	$\rho_{\text{ps}}$	$= (\gamma + 1) / (\gamma - 1) \mathcal{M}_1^2 / [\mathcal{M}_1^2 + 2 / (\gamma - 1)] \rho_0$
Pressure of the post-shock gas <sup>a</sup>	$P_{\text{ps}}$	$= (\gamma - 1) / (\gamma + 1) [2\gamma / (\gamma - 1) \mathcal{M}_1^2 - 1] P_0$
Post-shock speed of sound	$c_{\text{ps}}$	$= \gamma P_{\text{ps}} / \rho_{\text{ps}}$
Temperature of the warm phase	$T_w$	$= T_0$
Density of the warm phase	$\rho_w$	$= \rho_H T_H / T_w$
Density of the hot phase (post-expansion)	$\rho_h$	$= \rho_{\text{ps}} (P_{\text{ps}} / P_H)^{1/\gamma}$
Temperature of the hot phase (post-expansion)	$T_h$	$= \rho_H T_H / \rho_h$
Volume-averaged density of the warm phase	$\langle \rho_w \rangle_v$	$= \phi_{v,w} \rho_w$
Volume-averaged density	$\rho_2$	$= \phi_{v,w} \rho_w + (1 - \phi_{v,w}) \rho_h$
Expansion factor of the post-shock stream <sup>b</sup>	$S$	$= (1 - \sqrt{\Delta}) / [(\gamma - 1) / f + 2\eta]$
Velocity dispersion of the warm clouds	$\sigma_{\text{turb}}$	$= \sqrt{2 k_B T_w \eta f / (\gamma - 1) \phi_{v,w} \mu m_p}$
Halo dynamical time	$t_{\text{dyn,halo}}$	$= r_{\text{vir}} / v_{\text{vir}}$
Cooling time of the phase $\Phi \in \{\text{ps}, \text{w}, \text{h}\}$	$t_{\text{cool},\Phi}$	$= k_B \mu m_p T_\Phi / \rho_\Phi \Lambda(T_\Phi)$
Expansion time of the post-shock gas	$t_{\text{expand}}$	$= 2\gamma r_{\text{stream}} / c_{\text{ps}}$
Disruption time of the turbulent warm phase	$t_{\text{disrupt}}$	$= r_{\text{stream}} / \sigma_{\text{turb}}$
Isobaric cooling length $\Phi \in \{\text{ps}, \text{w}, \text{h}\}$	$\lambda_{\text{cooling}}$	$= c_\Phi t_{\text{cool},\Phi}$

**Notes.** A graphical representation of many of these variables is shown in Fig. 1. <sup>(a)</sup> Standard normal shock equation from the Rankine-Hugoniot jump conditions. <sup>(b)</sup> The equation assumes  $\Delta = 1 - 4[\eta f + (\gamma - 1)/2]\rho_H/\rho_2$ .