OPEN ACCESS



Thermal Instability in Radiation Hydrodynamics: Instability Mechanisms, Positiondependent S-curves, and Attenuation Curves

Daniel Proga^{1,2}, Tim Waters^{1,2,3}, Sergei Dyda⁴, and Zhaohuan Zhu^{1,2}

Department of Physics & Astronomy University of Nevada, Las Vegas 4505 S. Maryland Pkwy Las Vegas, NV, 89154-4002, USA; daniel.proga@unlv.edu

Nevada Center for Astrophysics (NCfA), University of Nevada, Las Vegas, NV, USA

Theoretical Division, Los Alamos National Laboratory, NM 87545, USA

Department of Astronomy University of Virginia 530 McCormick Rd. Charlottesville, VA 22904, USA

Received 2022 July 7; revised 2022 August 4; accepted 2022 August 6; published 2022 August 23

Abstract

Local thermal instability can plausibly explain the formation of multiphase gas in many different astrophysical environments, but the theory of local TI is only well-understood in the optically thin limit of the equations of radiation hydrodynamics (RHD). Here, we lay groundwork for transitioning from this limit to a full RHD treatment assuming a gray opacity formalism. We consider a situation where the gas becomes thermally unstable due to the hardening of the radiation field when the main radiative processes are free–free cooling and Compton heating. We identify two ways in which this can happen: (i) when the Compton temperature increases with time, through a rise in either the intensity or energy of a hard X-ray component; and (ii) when attenuation reduces the flux of the thermal component such that the Compton temperature increases with depth through the slab. Both ways likely occur in the broad-line region of active galactic nuclei where columns of gas can be ionization-bounded. In such instances where attenuation is significant, thermal equilibrium solution curves become position-dependent and it no longer suffices to assess the stability of an irradiated column of gas at all depths using a single equilibrium curve. We demonstrate how to analyze a new equilibrium curve—the attenuation curve —for this purpose, and we show that, by Field's instability criterion, a negative slope along this curve indicates that constant-density slabs are thermally unstable whenever the gas temperature increases with depth.

Unified Astronomy Thesaurus concepts: Active galactic nuclei (16)

1. Introduction

Several outstanding problems in the theory of active galactic nuclei (AGN) are finally becoming tractable due to the rapid progress that has been made in recent years in solving the equations of radiation hydrodynamics (RHD; e.g., Davis et al. 2012; Jiang et al. 2012, 2014; Ryan et al. 2015; Ryan & Dolence 2020; Jiang 2021). Prime among them is the structure of the broad-line region (BLR), as this is a strongly radiating flow by definition. In the class of models popularized by the AGN unification model, only a small volume of the BLR—the gas within BLR clouds-contributes to the line emission (for a review, see Netzer 2015). While there is still debate as to the relative contribution of matter-bounded clouds (i.e., clouds optically thin to the ionizing continuum, such that hydrogen is fully ionized) to the line luminosity (e.g., Collin-Souffrin & Lasota 1988; Shields et al. 1995; Snedden & Gaskell 2007), photoionization modeling generally requires BLR clouds to be ionization-bounded (i.e., moderately optically thick) to account for the observed properties of the broad lines (for textbook accounts, see Krolik 1999; Osterbrock & Ferland 2006). Testing the viability of BLR cloud models from first principles therefore requires solving the equations of RHD at a minimum.

Proga et al. (2014) and Dyda et al. (2020) reported on results from a preliminary investigation of BLR clouds that are ablated by an intense radiation field in the gray limit of RHD. The main focus of these studies was on assessing the role of the type and magnitude of the opacity on the evolution of a pre-existing cloud. Although these simulations did not include Compton processes,

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

they confirmed the early analytic results of Krolik (1988) and Mathews & Doane (1990): BLR clouds are highly prone to disruption and various hydrodynamical instabilities. Proga et al. (2014) noted that optically thin clouds survived the longest. Therefore, in subsequent studies, we focused on the optically thin regime by instead solving the equations of nonadiabatic gas dynamics, which allowed including other important radiative processes. In Proga & Waters (2015), for example, we studied the formation dynamics of matter-bounded BLR clouds in a radiative environment where line cooling and Compton heating resulted in an equilibrium curve (S-curve hereafter) that is thermally unstable. We demonstrated that, in nonisotropic radiation fields, the nonlinear regime of thermal instability (TI; Field 1965) is associated with cloud acceleration, because the line opacity increases as unstable gas condenses (see also Waters & Proga 2016).

Here, we return to the equations of RHD, but now with Compton heating included to allow for the possibility that cloud formation can be triggered due to thermal instability, thus alleviating the need to assume a pre-existing cloud as a part of the initial conditions in RHD calculations. In Section 2, we derive a net cooling function that allows us to bridge our previous BLR cloud studies, as it permits studying TI in RHD. By then focusing our analysis on an irradiated column of gas with constant density, in Section 3 we show how the effects of attenuation modify the basic stability analysis of TI, leading to a new type of equilibrium curve and a physical interpretation of its associated instability criterion.

2. Theory

Local TI refers to the exponential growth of linear wave modes in a homogeneous plasma. The entropy mode is sometimes referred to as a trivial mode, because it is locally nonpropagating, meaning that for a given flow field, it will simply be advected along streamlines. Only this mode has its stability determined by Balbus's criterion for TI (Balbus 1986),

$$\left[\frac{\partial(\mathcal{L}/T)}{\partial T}\right]_{p} < 0, \tag{1}$$

where the subscript notation indicates that the temperature derivative is taken at constant pressure. Here, \mathcal{L} is the radiative loss function that enters the first law of thermodynamics, commonly written for gas dynamics as

$$\rho \frac{D\mathcal{E}}{Dt} = -p\nabla \cdot \boldsymbol{v} - \rho \mathcal{L} - \nabla \cdot \boldsymbol{q}. \tag{2}$$

The variables ρ , v, and p are the gas density, velocity, and pressure, respectively, $\mathcal{E} = p\rho^{-1}/(\gamma-1)$ is the gas internal energy (with adiabatic index γ , taken to be 5/3 here), \boldsymbol{q} is the heat flux due to thermal conduction, and D/Dt is the comoving frame derivative. We note that Equation (1) will no longer determine the stability of an irradiated gas column at high optical depths, as condensation growth will be suppressed in a radiation diffusion regime; we return to this point in Section 4. As argued in Appendix A, Equation (1) will remain valid, provided the mean free path for radiation is much greater than both the cooling length of the gas and the gradient length scales of ρ , v, and p. Notice that this instability criterion reduces to Field's criterion, $[\partial \mathcal{L}/\partial T]_p < 0$, for gas in thermal equilibrium with $\mathcal{L} = 0$.

The goal of this section is to arrive at a function \mathcal{L} from first principles, beginning with the source term S(I, n) describing the interactions between gas and radiation in the radiation transfer equation,

$$\frac{\partial I}{\partial t} + c \, \boldsymbol{n} \cdot \nabla I = S(I, \boldsymbol{n}), \tag{3}$$

where I is the frequency-integrated lab frame specific intensity. In the mixed-frame approach, S(I, n) is specified in the comoving frame, where it is given by $S(I_0, n_0)$. To leading order, S(I, n) and $S(I_0, n_0)$ differ by O(v/c) terms, and these terms will be negligible for the dynamics associated with TI whenever the sound speed satisfies $c_s \ll c$. This is because, in the isobaric regime of TI, the condensation velocity is always highly subsonic, while in the nonisobaric regime, it is at most $v \sim c_s$ (Waters & Proga 2019). Thus, for the purpose of stability analysis, there is no need to distinguish between these frames.

Assuming local thermodynamic equilibrium (LTE) and using a gray treatment of radiation, the (frequency-integrated) source term is (see, e.g., Jiang 2021)

$$S(I, \mathbf{n}) = c\rho\kappa_{aR} \left(\frac{aT^4}{4\pi} - I\right) + c\rho\kappa_e (J - I)$$

$$+ c\rho(\kappa_{aP} - \kappa_{aR}) \left(\frac{aT^4}{4\pi} - J\right)$$

$$+ c\rho\kappa_e \frac{4(T - T_{C,eff})}{T_e} J, \tag{4}$$

where κ_{aR} and κ_{aP} are the Rosseland and Planck mean absorption opacities, κ_e is the electron scattering opacity, and

J is the mean intensity, defined as

$$J = \frac{1}{4\pi} \int I \, d\Omega. \tag{5}$$

The last term in Equation (4) is an approximate way to account for energy exchange through Compton scattering. In previous works that use this term (e.g., Hirose et al. 2009; Jiang et al. 2013), $T_{\text{C,eff}}$ corresponds to the temperature of the thermal radiation alone, i.e., $T_{\text{C,eff}} = T_r \equiv (E_r/a)^{1/4}$, with E_r being the energy density of that radiation, $4\pi J/c$. As will made clear in Section 3, the heating due to this term is not efficient enough to lead to TI. What is needed is a nonthermal component of higher-energy radiation.

2.1. Gray Two-band Treatment of Compton Heating

We first define the thermal radiation component as the frequencies extending through the soft X-ray band,

$$J_{\rm S} = \int_0^{\nu_{\rm h}} J_{\nu} d\nu,\tag{6}$$

i.e., up to some frequency ν_h that fully covers the blackbody tail of this distribution. The presence of high-energy photons can be accounted for in a gray formalism in various simplified ways. For example, the mean intensity of hard X-rays can be assumed to be time-dependent by taking $J_h = \int_{\nu_h}^{\infty} J_{\nu} d\nu$ to vary in proportion to J_s . Even simpler, the hard X-ray mean intensity can be parameterized in terms of the initial value of J_s , $J_0 \equiv J_s(0) = \int_0^{\nu_h} J_{\nu} d\nu|_{t=0}$, allowing the time dependence of $T_{C,\text{eff}}$ to be specified using two additional free parameters: $T_{C,h}$, the Compton temperature corresponding to the hard X-ray component (see Equation (B3)), and $f_h \equiv J_h/J_0$. This choice is quite physical and should be a good approximation that applies to the BLR whenever a column of gas is optically thin to this high-energy component. The radiation source term then becomes

$$S(I, \mathbf{n}) = c\rho\kappa_{aR} \left(\frac{aT^4}{4\pi} - I_{\rm s}\right) + c\rho\kappa_{e}(J_{\rm s} - I_{\rm s})$$

$$+ c\rho(\kappa_{aP} - \kappa_{aR}) \left(\frac{aT^4}{4\pi} - J_{\rm s}\right)$$

$$+ c\rho\kappa_{e} \frac{4(T - T_{\rm C,eff})}{T_{e}} (J_{\rm s} + J_{\rm h}), \tag{7}$$

where, as shown in Appendix B,

$$T_{\text{C,eff}} = \frac{T_r E_r / E_{r,0} + f_{\text{h}} T_{\text{C,h}}}{E_r / E_{r,0} + f_{\text{h}}}.$$
 (8)

Here, $E_{r,0} = 4\pi J_0/c$ is the initial value of the radiation energy density.

2.2. Net Cooling Function in RHD

By Equation (5), we see that the two κ_{aR} terms in Equation (7) cancel upon integrating over solid angle, Ω . The source term in the energy equation, $S_E \equiv \int S(I, \mathbf{n}) d\Omega$, can therefore be written

$$S_E = \rho \kappa_{aP} (aT^4 - E_r) + A_C \rho (E_r + f_h E_{r,0}) (T - T_{C,eff}),$$
 (9)

where E_r corresponds only to the thermal radiation,

$$E_r = \frac{4\pi J_{\rm s}}{c},\tag{10}$$

and $A_C = 4 k \kappa_e/(m_e c^2)$ is a constant. Note that, by the definition of f_h , $f_h E_{r,0} = 4\pi J_h/c$. With the source term for the momentum equation given by

$$S_F = -\frac{\rho(\kappa_e + \kappa_{aR})}{c} F_r, \tag{11}$$

we can now write down the equations of RHD:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \tag{12}$$

$$\frac{\partial(\rho v)}{\partial t} + \nabla \cdot (\rho vv + pI) = -\mathbf{S}_F,\tag{13}$$

$$\frac{\partial E}{\partial t} + \nabla \cdot [(E+p)v] = -cS_E - \nabla \cdot \boldsymbol{q}. \tag{14}$$

Comparing Equation (14) with Equation (2), we have the correspondence $\rho \mathcal{L} = cS_E$. Upon adopting a Kramers opacity appropriate for free–free absorption to use for the Planck mean opacity,

$$\kappa_{aP} = A_{\rm ff} \rho \ T^{-3.5},\tag{15}$$

we finally arrive at the net cooling function in our RHD framework (noting again that $T_{\text{C.eff}}$ is given by Equation (8)):

$$\mathcal{L} = c A_{\rm ff} \rho T^{-3.5} (aT^4 - E_r) + A_C (E_r + f_{\rm h} E_{r,0}) (T - T_{\rm C,eff}).$$
(16)

In Appendix A, we justify using Equation (1), which is formally derived under the optically thin approximation, to analyze the stability of this net cooling function.

3. Results

It has been known for over four decades that the thermal state of irradiated gas is determined by the shape of the incident spectrum together with the value of the ionization parameter, Ξ (Krolik et al. 1981; London et al. 1981). To the best of our knowledge, all previous dynamical calculations of TI have relied upon the assumption that the shape of the incident spectrum is fixed in both time and space. We now analyze Equation (16) to reveal how, by relaxing this assumption, an irradiated slab can transition from being thermally stable to unstable and how instability first sets in deep within the slab.

An S-curve can develop an unstable branch whenever the transition from one dominant heating and cooling process to another is too abrupt. From Equation (16), the ratio of the free-free cooling and Compton heating timescales is

$$\frac{t_{\rm cool}}{t_{\rm heat}} = \frac{A_C(E_r + f_{\rm h}E_{r,0})T_{\rm C,eff}}{acA_{\rm ff}\,\rho\sqrt{T}} \propto \xi \, T^{-\frac{1}{2}}T_{\rm C,eff},\tag{17}$$

where the proportionality relation comes from introducing the density ionization parameter, ξ (defined in Section 3.1 and Equation (21)). At low enough ξ where free–free processes dominate to give $t_{\rm cool} \ll t_{\rm heat}$, the gas can remain stable. An unstable branch arises where Compton heating overtakes free–free cooling in Equation (16). For a thermally stable region within a column of gas, the temperature is fixed by the local value of ξ in that region, so the only way to increase the ratio

 $t_{\rm cool}/t_{\rm heat}$ is by raising $T_{\rm C,eff}$. Examining Equation (8) in the optically thin limit $(E_r/E_{r,0}=1)$ and assuming $T_{r,0}/T_{\rm C,h} \ll f_{\rm h} \lesssim 1$, we find

$$T_{\text{C,eff}} = \frac{f_{\text{h}} T_{\text{C,h}}}{1 + f_{\text{h}}}.$$
 (18)

Hence, $T_{\rm C,eff}$ will increase due to either an increase in the temperature $(T_{\rm C,h})$ or mean intensity $(f_{\rm h})$ of high-energy photons. If the higher value of $T_{\rm C,eff}$ raises $\xi \, T^{-1/2}$ enough that $t_{\rm cool}/t_{\rm heat}$ continues to increase, the gas has transitioned into a thermally unstable state.

When the optical depth through the slab approaches unity, $T_{\text{C,eff}}$ will increase without any changes to $T_{\text{C,h}}$ or f_{h} due to attenuation. That this can lead to TI has been noted already by Gonçalves et al. (2007) in the context of photoionization modeling, as it is purely a radiative transfer effect. This cause of TI is readily understood from Equation (8). In the limit $f_{\text{h}}T_{\text{C,h}} \gg T_r E_r/E_{r,0}$ and neglecting emission through the slab so that $E_r/E_{r,0} \approx e^{-\tau}$, where $\tau = \int \kappa \, \rho \, dx$ is the optical depth (with $\kappa = \kappa_e + \kappa_{aP}$), one finds

$$T_{\text{C,eff}} = \frac{f_{\text{h}} T_{\text{C,h}}}{e^{-\tau} + f_{\text{h}}}.$$
 (19)

We see that Equation (18) is recovered for $\tau \ll 1$. As τ increases beyond 1, however, $T_{\text{C,eff}}$ exponentially approaches $T_{\text{C,h}}$, its maximum value.

To summarize, Equation (17) shows that the basic instability mechanism under investigation is an increase in $T_{\rm C,eff}$, and this increase can occur in two different ways: (i) through temporal changes in the incoming radiation field, i.e., from flux or spectral variability; and (ii) by a nontemporal route, namely the spectral hardening of the radiation field with depth. As demonstrated in Section 3.3, the parameter space for route (ii) to lead to TI is naturally encountered as an evolutionary phase of route (i). Put differently, when $T_{\rm C,eff}$ is too close to T_r , neither route (i) nor (ii) can lead to TI, but as $T_{\rm C,h}$ or $f_{\rm h}$ increase with time, the additional spectral hardening with depth makes route (ii) a possibility before route (i).

3.1. Position-dependent S-curves

In X-ray astronomy, it is customary to analyze TI in the phase space spanned by the temperature and either the density or pressure ionization parameter, ξ or Ξ , respectively. To express these in terms of E_r , first recall that only the soft X-ray flux is treated as thermal radiation, i.e., $E_r \equiv (4\pi/c)J_s$. The total ionizing flux is $F_{\rm ion} = F_s + F_h$, which for a point source is related to the total mean intensity J as $F_{\rm ion} = 4\pi J$, or $F_{\rm ion} = 4\pi (J_s + J_h) = 4\pi J_0 (E_r/E_{r,0} + f_h)$. From the definition $\xi \equiv 4\pi F_{\rm ion}/n$ where n the gas number density, we find that $\xi = (4\pi)^2 (J_0/n)(E_r/E_{r,0} + f_h)$. Taking n_0 to be the number density at the face of a slab where $E_r = E_{r,0}$, the incident mean intensity can be parameterized by $\xi_0 = (4\pi)^2 (J_0/n_0)(1+f_h)$. Combining the above relations gives, by the definition $\Xi \equiv (F_{\rm ion}/c)/p$,

$$\Xi = \frac{1}{4\pi c k} \frac{\xi}{T},\tag{20}$$

where

$$\xi = \frac{4\pi c}{n} (E_r + f_h E_{r,0}), \tag{21}$$

and

$$E_{r,0} = \frac{\xi_0 \, n_0}{4\pi \, c(1 + f_{\rm h})}.\tag{22}$$

From Equation (21), we see that, for a gas column exposed to a given mean intensity, a rise in density is not the only way the ionization can decrease; as noted above, attenuation will also reduce ξ by resulting in $E_r < E_{r,0}$. This represents a major departure from the textbook approach to analyzing TI, where a single S-curve suffices to delineate the unstable parameter space. That S-curves become position-dependent in attenuated gas was also pointed out by Gonçalves et al. (2007).

Conveniently, there is an analytic expression for these S-curves. Setting $\mathcal{L}=0$ in Equation (16), solving for ρ , and equating this with $\rho=\bar{m}(k\ T\ \Xi)^{-1}(E_r+f_{\rm h}E_{r,0})$ from Equation (21) gives

$$\Xi_{S}(T) = \frac{\bar{m} \ c \ A_{\rm ff}}{A_{C}} \frac{a \ T^{4} - E_{r}}{k(T_{C \ \rm eff} - T)} T^{-4.5}. \tag{23}$$

Because $T_{C,eff} = T_{C,eff}(E_r)$, the position dependence is solely due to the local value of $E_r = E_r(x)$. For gas columns with $\tau \ll 1$, a negligible amount of attenuation implies $E_r = E_{r,0}$, so S-curves in the optically thin limit have E_r given by Equation (22) and $\xi = \xi_0 n_0/n$ given by Equation (21). The S-curve is so named because, on the (T, Ξ) -plane, it has a negative-sloping unstable branch connecting low- and high-temperature positive-sloping stable branches. At small and large Ξ , these stable branches correspond to the asymptotic limits obvious from Equation (23): as $\Xi_S \to \infty$, $T \to T_{C,eff}$, while as $\Xi_S \to 0$, $T \to T_r \equiv (E_r/a)^{1/4}$.

3.2. Attenuation Curves

When TI is analyzed using S-curves, smaller values of $\xi = 4\pi F_{\rm ion}/n$ are associated with either smaller ionizing fluxes or higher densities, which is to say that $n\,\xi$ is treated as a proxy for $F_{\rm ion}\,\propto J$. Degeneracy will arise when attenuation becomes important, because J is approximately $J_0e^{-\tau}$, i.e., $n\,\xi\,\propto\,J_0e^{-\tau}$ becomes a proxy for both J_0 and the density. To break the degeneracy, it becomes necessary to supplement the S-curve analysis by focusing on gas at some fixed density. Specifically, taking $\rho=constant$ in Equation (21) leads to an alternative stability analysis of an attenuated column of gas, with smaller ξ values corresponding to greater depths through the column.

A second equilibrium curve, unique to RHD, that we term the *attenuation curve*, or A-curve for short, applies to all gas at a given density. It corresponds to the $\mathcal{L}=0$ contour of Equation (16) after replacing E_r with ξ using Equation (21). In contrast with S-curves, A-curves represent the phase space the column will occupy as the flux drops with depth due to significant absorption or scattering. An A-curve is shown in the left panel of Figure 1 for a constant-density slab, where it is compared with three S-curves, one computed at the face of the slab (solid line) and two at significant depths into the slab (solid lines with dots overplotted), computed by taking $E_r = 0.5E_{r,0}$ and $E_r = 0.25E_{r,0}$ in Equation (23).

We emphasize again that the Ξ -axis for an S-curve is a proxy for pressure, while the Ξ -axis for an A-curve is a proxy for E_r ,

i.e., these axes are separately built by taking $E_r = constant$ and n = constant in Equation (21), respectively. With Ξ_0 specified at the face of the slab (marked by the solid black dot), points along the A-curve with $\Xi < \Xi_0$ correspond to greater depths into the slab. The S-curves illustrate the spectral hardening with depth because, as $E_r/E_{r,0}$ decreases, the Compton branch of each one tends to higher values of $T_{\rm C,eff}$. The A-curve reveals that the temperature tends to decrease with greater depth through the column, the behavior expected for an attenuated slab.

We can now state a key result of this study: an irradiated constant-density slab is thermally unstable if, due to attenuation, the temperature *increases* with depth. This result follows from the mathematical identity $(\partial T/\partial p)_{\mathcal{L}}(\partial p/\partial \mathcal{L})_T(\partial \mathcal{L}/\partial T)_p = -1$. As shown in Appendix C, in the optically thin limit, the following version of this identity applies to both S-curves and A-curves:

$$\left(\frac{\partial \log T}{\partial \log \Xi}\right)_{\mathcal{L}} = \frac{p}{T} \frac{(\partial \mathcal{L}/\partial p)_{T}}{(\partial \mathcal{L}/\partial T)_{p}}.$$
(24)

The left-hand side derivative is taken along surfaces of constant \mathcal{L} ; this equation therefore relates Field's instability criterion, $(\partial \mathcal{L}/\partial T)_p < 0$, to the slopes of S-curves and A-curves (values of $\partial \log T/\partial \log \Xi$ along the contours $\mathcal{L}=0$). This correspondence between instability according to Field's criterion and the slope having $(\partial \log T/\partial \log \Xi)_{\mathcal{L}} < 0$ holds so long as $(\partial \mathcal{L}/\partial p)_T > 0$, which is readily seen to be the case for S-curves but not for A-curves (see Appendix C). In other words, in the case of A-curves, it is possible for a slab to be thermally unstable when $(\partial \log T/\partial \log \Xi)_{\mathcal{L}} > 0$, because $(\partial \mathcal{L}/\partial p)_T$ can become negative. We will show an example of this below. Focusing on the case $(\partial \mathcal{L}/\partial p)_T > 0$ shows that, unlike a negative slope for an S-curve, a negative slope for an A-curve physically corresponds to a slab with a temperature that increases with depth.

3.3. Stability Analysis Using S-curves and A-curves

For a stratified or otherwise inhomogeneous column of gas, S-curves are position-dependent unless attenuation within the column is negligible, and A-curves are also position-dependent unless the column has constant density. Hence, by focusing on constant-density slabs, a stability analysis requiring multiple S-curves is equivalent to one using a single A-curve. Returning to the left panel of Figure 1, notice that, at some depth corresponding to $E_r/E_{r,0} \approx 0.5$, gas occupies the last stable point on the cold branch of the S-curve (see the open circle left of the black dot). At greater depths, it occupies the unstable branch. This conclusion, i.e., that attenuation causes the slab to become unstable at some distance within, can alternatively be arrived at by analyzing the A-curve: a transition to instability occurs when this curve passes above its corresponding Balbus contour (see Appendix D), shown here as the dashed line.

To demonstrate what is likely a physical scenario that can be studied numerically, consider an initially thermally stable slab that begins to evolve as $T_{C,eff}$ increases with time. At t=0, the face of the slab will occupy a single point on the S-curve. We show such a point as the black dot on the $\log(T_{C,h}) = 7$ S-curve in the right panel of Figure 1. In this particular example, we increase $T_{C,eff}$ with time by raising $T_{C,h}$ with f_h held fixed at 0.1. The density of the slab will remain constant, at least initially, while the temperature (and hence pressure) will

⁵ There is no analytic expression for A-curves, due to the nonlinear E_r -dependence of $T_{\text{C,eff}}$ in Equation (8).

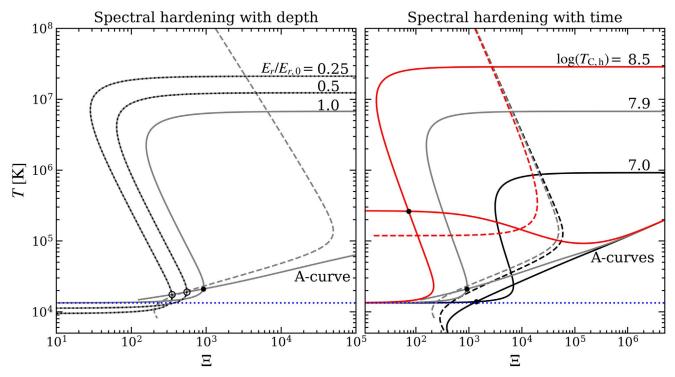


Figure 1. Examination of the two ways Compton heating can lead to TI in RHD. Either A-curves, along with their corresponding Balbus contours (dashed lines), or S-curves can be used to assess stability. Balbus contours for S-curves are not shown, because negative slopes indicate instability. Horizontal dotted lines denote the initial radiation temperature, $T_{r,0}$, of the thermal soft X-ray flux. Right panel: Route (i) to TI: $T_{C,eff}$ is increased with time by raising $T_{C,h}$ for a fixed f_h (set to 0.1). The dot on the black S-curve marks the irradiated face of an initially stable slab. It will move to the black dot on the gray (red) S-curve if $\log T_{C,h}$ is increased to 7.9 (8.5). In the $\log T_{C,h} = 7.9$ (8.5) case, the face of the slab remains stable (becomes unstable) because the S-curve has a positive (negative) slope at that point; equivalently, it beneath (above) the Balbus contour for that A-curve. Left panel: Route (ii) to TI: $T_{C,eff}$ increases with depth as a consequence of spectral hardening due to attenuation. The black dot marks the face of the slab corresponding to the $\log T_{C,h} = 7.9$ S-curve and A-curve from the right panel. The two open circles mark the (T, E) values at the depths where the smaller values of $E_r/E_{r,0}$ are reached. Thermal stability can be assessed using a single A-curve: a transition to instability occurs between the two open circles once the Balbus contour crosses below the A-curve. Alternatively, multiple S-curves can be computed to see this transition. They are labeled by the value of $E_r/E_{r,0}$ and explicitly show the spectral hardening with depth. The $E_r/E_{r,0} = 0.25$ point is seen to occupy the unstable branch.

increase until the slab reaches the new equilibrium temperature corresponding to $T_{\text{C,eff}}(t)$. We note that Ξ will decrease because $\Xi \propto p^{-1}$.

We consider two cases: the final value of the free parameter $\log(T_{\text{C,h}})$ is (i) 7.9 and (ii) 8.5. The slab settles on the gray S-curve in case (i) and the red S-curve in case (ii), at the locations marked with a black dot. In case (ii), the entire slab is now unstable: its face occupies the unstable branch of the S-curve, and the A-curve shows that the attenuated layers at lower Ξ have temperatures slightly greater than at Ξ_0 , thus satisfying the negative slope criterion for instability (see Equation (24)). The analysis of case (i) was already presented when discussing the left panel.

Notice that, for a slab to rise to the temperature of case (ii), it must first pass through the parameter space of case (i). Whether or not the unstable portion of the slab can undergo TI before the full slab becomes unstable (upon occupying the red S-curve) depends on the timescales involved during this period of heating. Explicit slab structure calculations are needed to determine whether there can be multiple dynamical outcomes.

4. Discussion and Conclusions

Numerous attempts made over the decades to explain the presence of broad emission lines in AGN spectra are posited on TI being the process that forms the BLR clouds within which many of the lines originate (e.g., Beltrametti 1981; Perry & Dyson 1985; Shlosman et al. 1985; Krolik 1988; Mathews &

Doane 1990; Goncalves et al. 1993; Wang et al. 2012; Elvis 2017). Perhaps the most promising cloud-based BLR models have been developed in the framework of magnetohydrodynamics (MHD; Emmering et al. 1992; de Kool & Begelman 1995; Begelman & Silk 2017); of these, only the model by Begelman & Silk (2017) is focused on TI, which is envisioned to occur in the atmosphere of a magnetically elevated disk. Testing the viability of this model will require solving the equations of radiation MHD (RMHD). Advances in understanding of the dynamics of TI in MHD have been led by the solar physics community in recent years (see reviews by Antolin & Froment 2022; Soler & Ballester 2022), especially in the context of solar prominence formation and the coronal rain phenomenon (e.g., Claes & Keppens 2019; Claes et al. 2020; Hermans & Keppens 2021). Because these studies have all assumed an optically thin formalism where attenuation is neglected, we stress that the RHD framework and analysis methods presented here are equally applicable to the equations of RMHD.

We also emphasize that TI is a process inherent to *somewhat* optically thin media: the growth of condensations will be suppressed in an optically thick, radiation diffusion regime, just as condensation growth is suppressed in optically thin media for wavelengths less than the Field length. In the latter case, the cooling of an unstable entropy mode cannot take place when the heat flux due to thermal conduction is strong enough to maintain a flow of heat from hot to cold across the entire perturbation; this circumstance is what the Field length

identifies (Begelman & McKee 1990). In a radiation diffusion regime, TI will also be suppressed because radiation will establish LTE and hence wipe out any temperature inhomogeneities.

The theory developed in this paper extends the optically thin qualification to encompass situations where attenuation is too important to completely ignore. We retain the key approximation made by Balbus (1986): the relevant wavenumbers are large enough that gas can be treated as homogeneous over length scales that are much smaller than those on which the background flow varies. In such a regime, the familiar criteria for TI still apply but must be evaluated layer by layer through an attenuated slab. In lieu of this position-dependent analysis using S-curves, it is possible to use a single A-curve to assess the stability of a constant-density column of gas. As with S-curves, the slope of A-curves is mathematically related to Field's instability criterion. However, there is no longer a oneto-one correspondence between a negative slope on the (T, Ξ) plane and instability for A-curves. Instability is less restrictive: an A-curve with a negative slope is unstable to TI, but the threshold is actually at some small positive slope, as shown in Figure 1 and discussed in Appendix C. Remarkably, this analysis reveals that a negative slope on an A-curve has the physical implication that a constant-density slab with an increasing temperature profile is thermally unstable.

We furthermore demonstrated, again in the case of a constant-density slab, how S-curves and A-curves are complementary in understanding how the slab will evolve under temporal changes in the hardness of the radiation field. This type of analysis may prove especially useful in understanding the dynamics of more complicated gas distributions. If, for example, we envision a Compton-thick, multiphase slab, all the gas with a given density will occupy a single A-curve, while all the gas at a given mean intensity will occupy a single S-curve. In subsequent papers of this series, we will apply this analysis to slab structure calculations to understand the nonlinear outcome of TI in RHD.

We are indebted to Yan-Fei Jiang for pointing out that the discrepancy between $4 k T_{C,s}$ and $\langle h \nu \rangle_B$ in Equation (B6) is due to the neglect of stimulated scattering. Support for this work was provided by the National Aeronautics and Space Administration under ATP grant NNX14AK44G and TCAN grant 80NSSC21K0496.

Appendix A Applicability of Established Linear Theory

Equation (16) has a functional dependence $\mathcal{L} = \mathcal{L}(\rho, T, E_r)$, whereas the stability criteria for local TI were derived assuming $\mathcal{L} = \mathcal{L}(\rho, T)$ under the optically thin approximation. While we aim to reach an RHD regime in the nonlinear phase of TI where the gas within BLR clouds is optically thick and radiating strongly, here we are only concerned with the background flow out of which the clouds form. This gas is seen to be optically thin locally, meaning that the wavelengths of unstable entropy modes are much smaller than the mean free path of the X-ray photons. Specifically, adopting $A_{\rm ff} = 5.19 \times 10^{24} \, {\rm cm}^5 \, {\rm g}^{-2} \, {\rm K}^{3.5}$ for the bremsstrahlung constant in Equation (15), the opacity is dominated by electron scattering at the densities and

temperatures considered:

$$\frac{\kappa_e}{\kappa_{aP}} = 1.2 \times 10^5 \, n_{11}^{-1} \, T_5^{3.5}.$$
 (A1)

Here, $n_{11}=n/10^{11}\,\mathrm{cm}^{-3}$ and $T_5=T/10^5\,\mathrm{K}$ are characteristic values for the BLR. In examining the stability of an irradiated column of gas, $\mathcal{L}(\rho,T,E_r)$ will be reduced back to $\mathcal{L}=\mathcal{L}(\rho,T)$ if we can treat E_r as constant over portions of the column Δx satisfying $\Delta x \ll \lambda_{\mathrm{mfp}}$, with $\lambda_{\mathrm{mfp}}=(\kappa_e\rho)^{-1}=5.0\times10^{13}\,n_{11}^{-1}\,\mathrm{cm}$ the mean free path of the soft X-ray photons. Applying local TI theory also requires the gas to be homogeneous over scales where the growth rate of TI peaks, namely for $\Delta x \gtrsim 10\,\lambda_{\mathrm{cool}}$, where $\lambda_{\mathrm{cool}}=c_s t_{\mathrm{cool}}$ is the cooling length. Combining the above requirements, the local approximation holds when $\lambda_{\mathrm{cool}}\ll\lambda_{\mathrm{mfp}}$. With $t_{\mathrm{cool}}=c_v T/(c\,a\,A_{\mathrm{ff}}\rho T^{1/2})$ for free–free cooling, we obtain $\lambda_{\mathrm{cool}}=2.6\times10^9\,T_5\,n_{11}^{-1}\,\mathrm{cm}$; thus, $\lambda_{\mathrm{cool}}\ll\lambda_{\mathrm{mfp}}$ is satisfied independent of density. We further note that the radiation force will not alter the linear theory, as the flow will be uniformly accelerated when $\kappa_e\gg\kappa_{aP}$, thereby permitting an analysis in the comoving frame.

The above length scale hierarchy validates the use of the socalled eikonal approximation (or WKBJ theory) that was used by Balbus (1986) and permits applying Equation (1) to a timedependent background flow such as the one considered here (see Section 3.3 and the right panel of Figure 1).

Appendix B Parameterization of the Mean Photon Energy

In our gray treatment of the X-ray irradiation, the time-dependent mean photon energy defines an effective Compton temperature $T_{\rm C,eff}$ through

Approximate Te,eff unrough
$$\langle h\nu \rangle \equiv 4 \ k \ T_{\text{C,eff}} = \frac{\int_0^\infty h\nu J_\nu \ d\nu}{\int_0^\infty J_\nu \ d\nu}$$

$$= \frac{\int_0^{\nu_\text{h}} h\nu J_\nu \ d\nu + \int_{\nu_\text{h}}^\infty h\nu J_\nu \ d\nu}{J_\epsilon + J_\text{h}}. \tag{B1}$$

Here, $\nu_{\rm h}$ divides the incident spectrum into two components: (i) $J_{\rm s} \equiv \int_0^{\nu_{\rm h}} J_{\nu} \ d\nu$, the thermal component spanning the far-UV and soft X-ray energy range for which we solve the radiation transfer equation; and (ii) $J_{\rm h} \equiv \int_{\nu_{\rm h}}^{\infty} J_{\nu} \ d\nu$, the high-energy tail (i.e. keV range) that is parametrized by the two free parameters ($f_{\rm h}$ and $T_{\rm C,h}$) defined below. Both components have an associated Compton temperature, as the terms in the numerator are related to mean photon energies, viz.

$$4 k T_{\text{C,s}} = \langle h\nu \rangle_{\text{s}} = J_{\text{s}}^{-1} \int_{0}^{\nu_{\text{h}}} h\nu J_{\nu} d\nu; \tag{B2}$$

$$4 k T_{C,h} = \langle h\nu \rangle_h = J_h^{-1} \int_{\nu_h}^{\infty} h\nu J_{\nu} d\nu.$$
 (B3)

Below, we explain how to modify this two-band formalism to make $T_{\rm C,s}$ exactly equal to $T_r = (E_r/a)^{1/4}$, the radiation temperature that is evolved using the equations of RHD. Meanwhile, $T_{\rm C,h}$ is taken as a free parameter and controls the thermal stability of the gas.

To express $T_{C,eff}$ in terms of $T_{C,s}$, $T_{C,h}$, and E_r , we first note that, at t = 0, the second line in Equation (B1) can be written as

$$T_{\text{C,eff}}(t=0) = \frac{T_{\text{C,s}} + f_{\text{h}} T_{\text{C,h}}}{1 + f_{\text{h}}},$$
 (B4)

where $f_{\rm h}=J_{\rm h}/J_0$ parameterizes component (ii) in terms of the initial value of component (i), $J_0\equiv J_{\rm s}(t=0)$. Further recognizing that $J_{\rm h}=(E_r/E_{r,0})J_0$, with $E_{r,0}=(4\pi/c)J_0$, the time-dependent decomposition then becomes

$$T_{\text{C,eff}} = \frac{T_{\text{C,s}}E_r + f_{\text{h}}T_{\text{C,h}}E_{r,0}}{E_r + f_{\text{h}}E_{r,0}}.$$
 (B5)

This reduces to $T_{\rm C,s}$ for $f_{\rm h}=0$, and thus corresponds to the simplified treatment of Compton scattering that has been used in the past (e.g., Hirose et al. 2009; Jiang et al. 2013) when $T_{\rm C,s}=T_r$. However, the equality $T_{\rm C,s}=T_r$ is not obtained using the definition of $\langle h\nu \rangle$ in Equation (B1). The mean photon energy for a Planck spectrum, $J_{\nu}=B_{\nu}$, is

$$\langle h\nu \rangle_B \equiv \frac{\int_0^\infty h\nu B_\nu \, d\nu}{\int_0^\infty B_\nu \, d\nu} = \frac{360 \, \zeta(5)}{\pi^4} k \, T_r \approx 3.823 \, k \, T_r, \quad (B6)$$

where $\zeta(s)$ is the Riemann zeta function. With this definition, net heat exchange will take place between gas and radiation at temperature T_r . When Compton scattering is treated using the Kompaneets equation (the frequency diffusion limit), the mean photon energy is instead given by (see Blaes & Socrates 2001)

$$\langle h\nu \rangle = \frac{\int_0^\infty h\nu J_\nu [1 + (c^2/2h)J_\nu / \nu^3] \, d\nu}{\int_0^\infty J_\nu \, d\nu},$$
 (B7)

with the factor $(c^2/2h)J_{\nu}/\nu^3$ accounting for stimulated scattering. With this definition in place of Equation (B1), we still arrive at the parameterization given in Equation (B5) but now with $T_{\text{C,s}} = T_r$, which brings us to Equation (8).

Appendix C The Slopes of Equilibrium Curves in RHD

Beginning with the identity quoted in Section 3.2, $(\partial T/\partial p)_{\mathcal{L}}(\partial p/\partial \mathcal{L})_T(\partial \mathcal{L}/\partial T)_p = -1$, we use the chain rule on the first derivative to introduce the slope along equilibrium curves, $(\partial \log T/\partial \log \Xi)_{\mathcal{L}}$:

$$\left(\frac{\partial T}{\partial p}\right)_{\mathcal{L}} = \frac{T}{p} \left(\frac{\partial \log T}{\partial \log \Xi}\right)_{\mathcal{L}} \left(\frac{\partial \log \Xi}{\partial \log p}\right)_{\mathcal{L}}.$$
 (C1)

By equations (20) and (21), we have $(\partial \log \Xi/\partial \log p)_{\mathcal{L}} = \Xi^{-1}(\partial E_r/\partial p)_{\mathcal{L}} - 1$. Meanwhile, for the derivative $(\partial p/\partial \mathcal{L})_T$, the chain rule gives

$$\left(\frac{\partial \mathcal{L}}{\partial p}\right)_{T} = \left(\frac{\partial \mathcal{L}}{\partial \rho} \frac{\partial \rho}{\partial p}\right)_{T} + \left(\frac{\partial \mathcal{L}}{\partial E_{r}} \frac{\partial E_{r}}{\partial p}\right)_{T}.$$
(C2)

By the ideal gas law, $(\partial \rho/\partial p)_T = \rho/p$. With these substitutions, the original identity becomes

$$\left[1 - \frac{1}{\Xi} \left(\frac{\partial E_r}{\partial p}\right)_{\mathcal{L}}\right] \left(\frac{\partial \log T}{\partial \log \Xi}\right)_{\mathcal{L}} \\
= \frac{\rho(\partial \mathcal{L}/\partial \rho)_T + p(\partial \mathcal{L}/\partial E_r)_T(\partial E_r/\partial p)_T}{T(\partial \mathcal{L}/\partial T)_p}.$$
(C3)

This equation gives the relation between Field's instability criterion, $(\partial \mathcal{L}/\partial T)_p < 0$, and the slopes of S-curves and A-curves. In the case of S-curves, E_r is independent of p and this equation reduces to

$$\left(\frac{\partial \log T}{\partial \log \Xi}\right)_{\mathcal{L}} = \frac{\rho(\partial \mathcal{L}/\partial \rho)_T}{T(\partial \mathcal{L}/\partial T)_p}.$$
 (C4)

The derivative $(\partial \mathcal{L}/\partial \rho)_T$ is readily seen to be positive via Equation (16), and this is generically true for astrophysical cooling functions (Balbus & Soker 1989; see also Appendix B of Waters & Proga 2019), giving a one-to-one relation between Field's criterion and the sign of the slope of S-curves. For A-curves, the bracketed term on the left-hand side of Equation (C3) can be shown to be positive, but the derivative $(\partial \mathcal{L}/\partial E_r)_T$ can be negative and change the sign of the numerator on the right-hand side. The one-to-one correspondence is therefore lost. However, in the optically thin limit, $\partial E_r/\partial p$ approaches zero, showing that Equation (C4) holds for A-curves in this limit, a fact relied upon in Section 3.2.

Appendix D Balbus Contours

Diagnosing instability using Equation (1) entails computing the location in a slab where the quantity $[\partial(\mathcal{L}/T)/\partial T]_p$ changes sign, which can be done by plotting the Balbus contour, defined as locations in phase space where $[\partial(\mathcal{L}/T)/\partial T]_p = 0$. Dividing Equation (16) by T and replacing ρ with p using the ideal gas law (so that p can be held constant while taking the derivative with respect to T) gives

$$\left[\frac{\partial(\mathcal{L}/T)}{\partial T}\right]_{p} = -\frac{3}{2}c A_{\rm ff} \rho T^{-5.5} \left(aT^{4} - \frac{11}{3}E_{r}\right) + A_{C} T^{-2} (E_{r} + f_{\rm h} E_{r,0})T_{\rm C,eff}.$$
(D1)

Both S-curves and A-curves have corresponding Balbus contours. The analytic expression that accompanies Equation (23) is found by setting $[\partial(\mathcal{L}/T)/\partial T]_p = 0$ in Equation (D1), solving for ρ , and as before, setting this expression equal to $\rho = \bar{m}(k \ T \ \Xi)^{-1}(E_r + f_h E_{r,0})$ from Equation (20), to give

$$\Xi_B(T) = \frac{3}{2} \frac{\bar{m} c A_{\rm ff}}{A_C} \frac{a T^4 - (11/3)E_r}{k T_{\rm Ceff}} T^{-4.5}.$$
 (D2)

By instead using Equation (20) to eliminate E_r from Equation (D1), the equation that results from setting $[\partial(\mathcal{L}/T)/\partial T]_p = 0$ yields the Balbus contours corresponding to A-curves that are plotted in Figure 1. The utility of using Balbus contours to identify where gas can be thermally unstable is that they also reveal the parameter space for TI

when gas departs from the equilibrium curve (the contour $\mathcal{L} = 0$).

ORCID iDs

Daniel Proga https://orcid.org/0000-0002-6336-5125 Tim Waters https://orcid.org/0000-0002-5205-9472 Sergei Dyda https://orcid.org/0000-0002-1954-8864 Zhaohuan Zhu https://orcid.org/0000-0003-3616-6822

References

```
Antolin, P., & Froment, C. 2022, FrASS, 9, 820116
Balbus, S. A. 1986, ApJL, 303, L79
Balbus, S. A., & Soker, N. 1989, ApJ, 341, 611
Begelman, M. C., & McKee, C. F. 1990, ApJ, 358, 375
Begelman, M. C., & Silk, J. 2017, MNRAS, 464, 2311
Beltrametti, M. 1981, ApJ, 250, 18
Blaes, O., & Socrates, A. 2001, ApJ, 553, 987
Claes, N., & Keppens, R. 2019, A&A, 624, A96
Claes, N., Keppens, R., & Xia, C. 2020, A&A, 636, A112
Collin-Souffrin, S., & Lasota, J.-P. 1988, PASP, 100, 1041
Davis, S. W., Stone, J. M., & Jiang, Y.-F. 2012, ApJS, 199, 9
de Kool, M., & Begelman, M. C. 1995, ApJ, 455, 448
Dyda, S., Proga, D., & Reynolds, C. S. 2020, MNRAS, 493, 437
Elvis, M. 2017, ApJ, 847, 56
Emmering, R. T., Blandford, R. D., & Shlosman, I. 1992, ApJ, 385, 460
Field, G. B. 1965, ApJ, 142, 531
```

```
Gonçalves, A. C., Collin, S., Dumont, A. M., & Chevallier, L. 2007, A&A,
Gonçalves, D. R., Jatenco-Pereira, V., & Opher, R. 1993, ApJ, 414, 57
Hermans, J., & Keppens, R. 2021, A&A, 655, A36
Hirose, S., Krolik, J. H., & Blaes, O. 2009, ApJ, 691, 16
Jiang, Y.-F. 2021, ApJS, 253, 49
Jiang, Y.-F., Stone, J. M., & Davis, S. W. 2012, ApJS, 199, 14
Jiang, Y.-F., Stone, J. M., & Davis, S. W. 2013, ApJ, 778, 65
Jiang, Y.-F., Stone, J. M., & Davis, S. W. 2014, ApJS, 213, 7
Krolik, J. H. 1988, ApJ, 325, 148
Krolik, J. H. 1999, Active Galactic Nuclei: From the Central Black Hole to the
  Galactic environment (Princeton, NJ: Princeton Univ. Press)
Krolik, J. H., McKee, C. F., & Tarter, C. B. 1981, ApJ, 249, 422
London, R., McCray, R., & Auer, L. H. 1981, ApJ, 243, 970
Mathews, W. G., & Doane, J. S. 1990, ApJ, 352, 423
Netzer, H. 2015, ARA&A, 53, 365
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae
  and Active Galactic Nuclei (Mill Valley, CA: Univ. Science Books)
Perry, J. J., & Dyson, J. E. 1985, MNRAS, 213, 665
Proga, D., Jiang, Y.-F., Davis, S. W., Stone, J. M., & Smith, D. 2014, ApJ,
  780, 51
Proga, D., & Waters, T. 2015, ApJ, 804, 137
Ryan, B. R., & Dolence, J. C. 2020, ApJ, 891, 118
Ryan, B. R., Dolence, J. C., & Gammie, C. F. 2015, ApJ, 807, 31
Shields, J. C., Ferland, G. J., & Peterson, B. M. 1995, ApJ, 441, 507
Shlosman, I., Vitello, P. A., & Shaviv, G. 1985, ApJ, 294, 96
Snedden, S. A., & Gaskell, C. M. 2007, ApJ, 669, 126
Soler, R., & Ballester, J. L. 2022, FrASS, 9, 789083
Wang, J.-M., Du, P., Baldwin, J. A., et al. 2012, ApJ, 746, 137
Waters, T., & Proga, D. 2016, MNRAS, 460, L79
Waters, T., & Proga, D. 2019, ApJ, 875, 158
```