Atmospheric Escape

Lecturer: Ray Pierrehumbert
Notes written by: Chris Cawthorn & Yutian Wu

24 June 2008

1 Introduction

The Earth's atmosphere is a comparatively thin layer of a gaseous mixture which is distributed almost uniformly over the surface of the Earth. In the vertical direction, more than 99% of the mass of the atmosphere is found below an altitude of 30km. In contrast, the horizontal scale of the atmosphere is of order 20,000km. The atmosphere is composed of several layers which differ in composition, temperature, stability and energetics. Starting from the surface, the main layers are the troposphere, stratosphere, mesosphere, and thermosphere, separated by conceptual partitions called pauses (e.g., tropopause). The concentrations of nitrogen, oxygen and some inert gases are practically uniform in the atmosphere up to the mesopause. This region constitutes the homosphere. However, above about 100km, the density of gas begins to fall off exponentially with increasing altitude with a rate depending on the molecular mass. Larger mass constituents, such as oxygen and nitrogen, fall off more quickly than lighter ones such as helium and hydrogen. This layer, in which the composition of the atmosphere varies with altitude, is called the heterosphere.

In the following, we introduce atmospheric escape. Due to thermal mechanisms, a lighter molecule is more likely to escape from the atmosphere because of its higher average speed at a given temperature. For example, hydrogen escapes more easily than carbon dioxide. This has numerous applications in astrophysical and planetary science.

1.1 Loss of water from Venus

The atmosphere of Venus contains only $0.1-1\%$ H_2O, a fact revealed by the recent Mariner 5 and Venera 4 missions. The total abundance of H_2O is $20-200\text{gm}\,\text{cm}^{-2}$ in the Venusian atmosphere compared with $320,000\text{gm}\,\text{cm}^{-2}$ in the Earth's atmosphere. The origin of the present atmosphere of Venus is assumed to be the same as that of the Earth, given the similarity in size and mass of the two planets. However, a large amount of H_2O has been lost during Venus' history. In the atmosphere of Venus, water vapor was able to become a major constituent at a high altitude where the atmospheric cold trap (places where the major constituents of the atmosphere condense) was located, and could be steadily photodissociated. The hydrogen atoms freed as a result of this process then flowed outward from the planet due in part to hydrodynamic escape. This process could account for the large loss of water from Venus. See [2, 4] for a much more thorough description.

1.2 Hydrogen content of the early Earth atmosphere

Research has shown that hydrogen was one of the major constituents in the ancient atmosphere. In addition, from a biological view, the existence and efficient production of prebiotic organic compounds on early Earth was necessary for the origin of life. H_2, along with O_2 and CO_2, can absorb extreme ultraviolet (EUV) radiation, but only H_2 can carry energy back to space by hydrodynamic escape. When hydrogen became the major gas in the heterosphere and the major absorber of EUV, the escape rate of hydrogen could have been controlled by the solar EUV flux available to drive the flow. One could consider that a balance then formed between volcanic hydrogen outgassing and the hydrodynamic escape of hydrogen from the atmosphere, helping to maintain the high hydrogen mixing ratio on early Earth. A discussion is given in [6], some of the details of which are given in §3 of these notes.

1.3 Loss of hydrogen from Titan

Methane gas is abundant in Titan's atmosphere, but it can be broken apart by ultraviolet light via the following process

$$\mathrm{CH_4} + h\nu \longrightarrow \mathrm{CH_3} + \mathrm{H}, \tag{1}$$

subsequently reforming to create ethane (C_2H_6). This reaction is common in the atmospheres of giant planets, where the hydrogen remains in the atmosphere due to the high gravitational energy that must be overcome in order for it to escape. However, on Titan, where the gravitational attraction is much lower, hydrogen can escape, causing the observed carbon to hydrogen ratio, which is higher than for pure methane.

1.4 Stellar wind

The phenomenon of atmospheric escape is not confined merely to planetary masses. Indeed, the stellar wind is an escape process that occurs at the outer limits of the Sun's atmosphere. Light elements, particularly hydrogen, gain sufficient energy to escape the Sun's gravity, and are radiated outwards. So great is the magnitude of this release that it can be measured from Earth. The process, while essentially hydrodynamical, is complicated by the influence of the solar magnetic field, and the fact that the high solar atmosphere is a plasma. We shall not discuss the solar wind further in this document, but direct the reader to [5] for a detailed discussion of the subject.

2 Transcritical flows

It is believed that atmospheric flows must be supersonic in order to escape the planet's gravitational field. This viewpoint is justified, somewhat dubiously, by the assertion that the steady-state outflow of subsonic fluid has an infinite density everywhere, whereas the

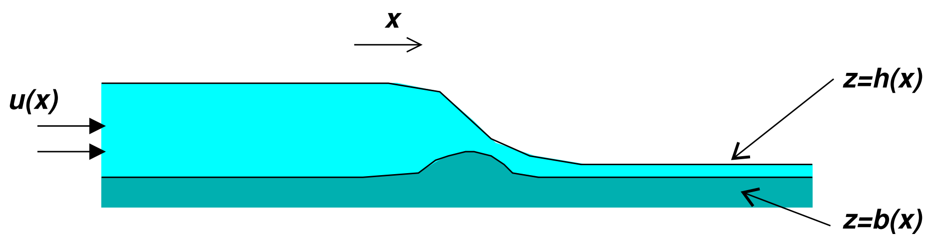

Figure 1: A sketch of shallow-water flow over a gentle bump.

outflow of supersonic fluid can support a more reasonable steady density distribution. Given the unfathomable size of the Universe, however, it seems unlikely that a steady state can be reached, making this argument rather spurious. Much more reasonable is a state in which material gradually radiates away from the planet. However, this transient state cannot be used to draw conclusions about the velocity of escape. Nevertheless, we shall proceed with the belief that flows involved in atmospheric escape must transition between subsonic atmospheric flows and supersonic exospheric flows - that is, they must be *transcritical*. Transcritical flows, characterised by a singularity in the differential equations governing the fluid, are abundant in both modern and classical fluid mechanical literature. In this section, we shall introduce the notion of a transcritical flow via some familiar examples.

2.1 Shallow-water flow over a bump

The flow of a shallow layer of water over a gentle bump is one of the simplest examples of a physical system exhibiting transcritical behaviour, and is often covered in introductory courses in fluid mechanics. The situation is illustrated by Figure 1, with fluid contained between a free surface $z = h(x)$ and a rigid base $z = b(x)$. We denote the fluid velocity (assumed uniform in depth) by $u(x)$. The fluid is incompressible, with constant density ρ.

The equations of mass and momentum conservation applied to the flow, assuming a steady-state solution, are

$$\partial_x\left((h-b)u\right) = \partial_x \Phi = 0 \tag{2}$$

$$u\partial_x u = -g\partial_x h. \tag{3}$$

By differentiating (2) and substituting for $\partial_x h$ in (3), we arrive at the equation

$$u\left(1-\mathrm{Fr}^{-2}\right)\partial_x u = -g\partial_x b, \tag{4}$$

where the Froude number is given by the ratio of the local velocity to the local wave speed, thus

$$\mathrm{Fr} = \frac{u}{\sqrt{g(h-b)}}. \tag{5}$$

If we assume that the basal height $b(x)$ is given, then (4) is an ordinary differential equation for $u(x)$, with forcing $-g\partial_x b$. We note, however, that this equation is singular for $u = 0$ and $\mathrm{Fr} = 1$. The former case is simply the degenerate case of no flow, but the latter is much more important. If at any point the flow conditions are such that $\mathrm{Fr} = 1$, (4) may only have

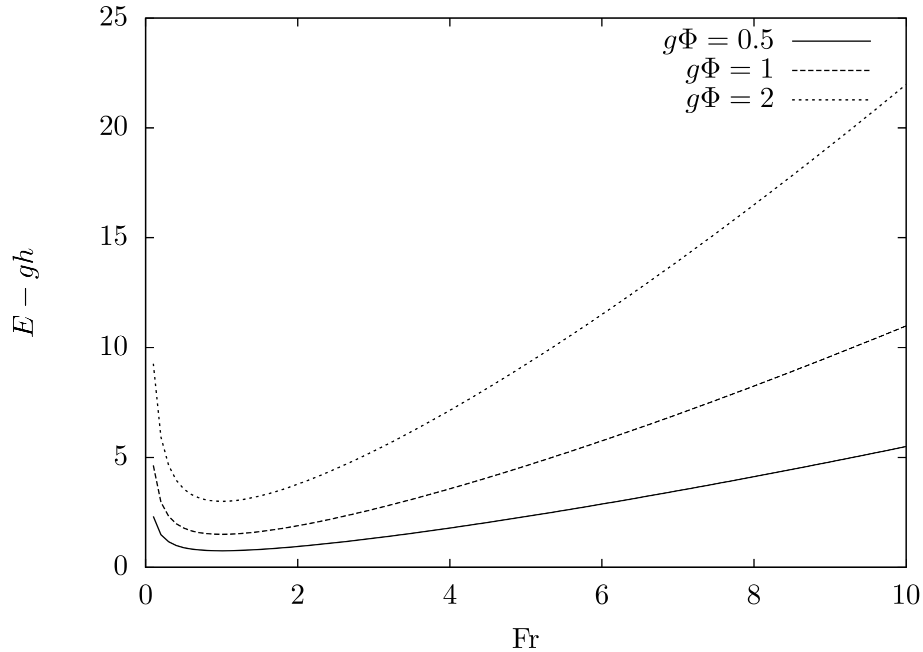

Figure 2: A sketch of the relationship between Froude number and energy, as given by (10)

a smooth solution if the forcing also vanishes at that point - i.e. $\partial_x b = 0$ where $\mathrm{Fr}(x) = 1$. Crucially, this means that the flow may only transition smoothly between a region where $\mathrm{Fr} < 1$ (a *subcritical* region) and a region where $\mathrm{Fr} > 1$ (a *supercritical* region) at a point where $\partial_x b = 0$.

One can think of this transcriticality in the context of energy. If we simply rearrange (3) into the conservative form

$$\begin{aligned} 0 &= \frac{\partial}{\partial x}\left(\frac{1}{2}u^2 + gh\right) && (6)\\ &= \frac{\partial}{\partial x}\left(\frac{1}{2}u^2 + g(h-b) + gb\right) && (7)\\ &= \frac{\partial}{\partial x}\left[(h-b)\left(\frac{1}{2}\mathrm{Fr}^2 + 1 + \frac{b}{h-b}\right)\right] && (8)\\ &= \frac{\partial}{\partial x}\left[\left(\frac{g\Phi}{\mathrm{Fr}}\right)^{2/3}\left(\frac{1}{2}\mathrm{Fr} + 1 + \frac{b}{h-b}\right)\right]. && (9) \end{aligned}$$

If we define the x-independent quantity here to be a specific energy E, then (9) relates E and Fr via

$$E - gh = \left(\frac{g\Phi}{\mathrm{Fr}}\right)^{2/3}\left(\frac{1}{2}\mathrm{Fr}^2 + 1\right). \tag{10}$$

This relationship is shown in Figure 2. Evaluating E by using the upstream boundary conditions, we find that either zero, one, or two solutions for Fr exist at any location x,

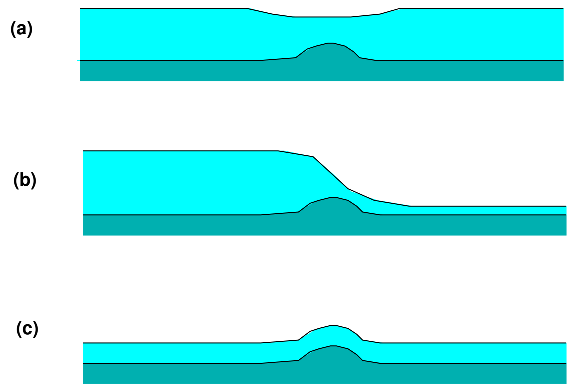

Figure 3: Sketches of the shallow-water flow over a bump for (a) subcritical flow, (b) transcritical flow, and (c) supercritical flow.

given the local bed height $b(x)$. A smooth solution for $\mathrm{Fr}(x)$ must remain on the curve shown in Figure 2, and take the local value of $E - gb$. It should be noted that the minimum energy occurs when $\mathrm{Fr} = 1$, i.e. at the transcritical point. This agrees with our earlier conclusion that, in order to pass smoothly from subcritical to supercritical flow, one must pass through a region where h is extremal. If the geometry and upstream conditions are such that the transcritical regime can be reached, there are three possibilities for the flow, given an initial upstream energy E. These are illustrated in Figure 3.

If the bed elevation $b(x)$ should become sufficiently large so as to forbid any solution of (10), then we can expect no steady solution to exist. In practice, a phenomenon known as 'choking' occurs. Fluid builds up behind the bump, creating a disturbance whose upstream extent increases in time until it has sufficiently modified the upstream boundary conditions, allowing a steady flow to form.

2.2 Compressible flow in a duct

Another important example of transcritical behaviour originates as a problem in supersonic propulsion. Consider the duct illustrated in Figure 4. In a simple, one-dimensional model, the compressible fluid contained in the duct has local density $\rho(x)$, and velocity $u(x)$. The cross-sectional area of the duct is denoted by $A(x)$. Once again, we can write down equations describing the conservation of mass and momentum

$$\partial_x (\rho A u) = \partial_x \Phi = 0, \tag{11}$$

$$u \partial_x u = -\frac{1}{\rho} \partial_x p, \tag{12}$$

but in this case we must supplement these equations with an equation of state. Here, it suffices to use the general form for an adiabatic gas

$$p = p(\rho), \qquad \text{with sound speed} \qquad c_s^2 = \frac{\mathrm{d}p}{\mathrm{d}\rho}. \tag{13}$$

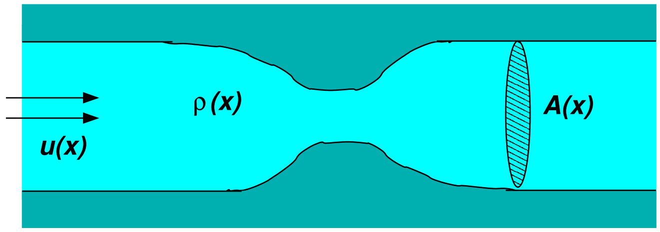

Figure 4: Diagram and notation for compressible duct flow

We can now use (13) to substitute ρ for p in (12), before using conservation of mass to eliminate $\partial_x \rho$, as we did in §2.1. This quickly leads to the ordinary differential equation

$$\frac{1}{u}\left(\mathcal{M}^2 - 1\right)\partial_x u = \frac{\partial_x A}{A}, \tag{14}$$

where $\mathcal{M} = u/c_s$ is the *Mach number*. Once again, for a given geometry defined by $A(x)$, the steady-state velocity obeys a singular ordinary differential equation. Importantly, the singular point occurs when $\mathcal{M} = 1$, separating subsonic (subcritical) flow from supersonic (supercritical) flow. In order to have a transsonic flow, passing smoothly through $\mathcal{M} = 1$ (known as the *Mach point*) we require that the forcing in (14) must vanish. That is, the area $A(x)$ must be extremal at the Mach point. This is known in aerodynamics as the *minimum area rule for transsonic flow*.

Framing this problem in terms of energy requires use of the thermodynamical relation can proceed by finding a conservative form in much the same way as in §2.1

$$0 = \frac{\partial}{\partial x}\left[\frac{1}{2}u^2 + \int^{\rho}\frac{c_s^2(\rho')}{\rho'}\mathrm{d}\rho'\right] \tag{15}$$

$$= \frac{\partial}{\partial x}\left[c_s^2\left(\frac{1}{2}\mathcal{M}^2 + \frac{1}{c_s^2}\int^{\rho}\frac{c_s^2(\rho')}{\rho'}\mathrm{d}\rho'\right)\right]. \tag{16}$$

One would now like to define a specific energy $E(\mathcal{M})$ to be equal to this conserved quantity, but the dependence of c_s on ρ means that this is as far as a general adiabatic theory can take us. If we consider specifically an ideal gas, where $p = p_0(\rho/\rho_0)^{1/\beta}$, then we may write (16) in the form

$$0 = \frac{\partial}{\partial x}\left[\left(\frac{p_0\rho^{-1+1/\beta}}{\beta\rho_0^{1/\beta}}\right)\left(\frac{1}{2}\mathcal{M}^2 + \frac{\beta}{1-\beta}\right)\right]. \tag{17}$$

We can then use conservation of mass (11) to substitute for ρ and arrive at the equation

$$0 = \frac{\partial E}{\partial x} = \frac{\partial}{\partial x}\left[c_{s0}^2\left(\frac{\mathcal{M}_0 A_0}{\mathcal{M}A}\right)^{2\frac{1-\beta}{1+\beta}}\left(\frac{1}{2}\mathcal{M}^2 + \frac{\beta}{1-\beta}\right)\right], \tag{18}$$

where the reference values $(c_{s0}, \mathcal{M}_0, A_0)$ are evaluated using the upstream boundary conditions. Note now that the specific energy E is written only as a function of the local Mach number $\mathcal{M}$, given the variation of cross-sectional area A with x. This relationship, for a representative value of β, is illustrated in Figure 5

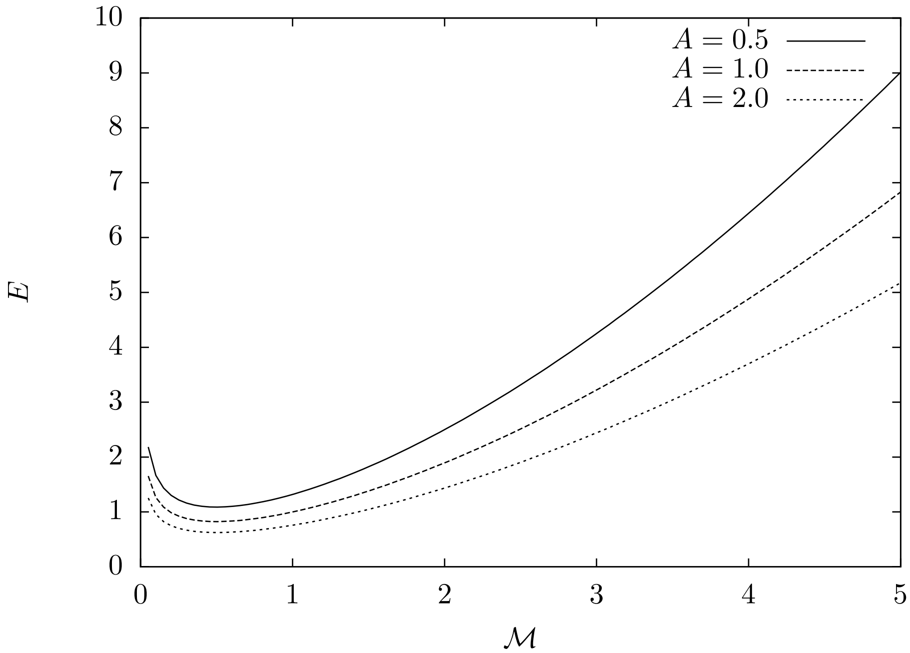

Figure 5: A plot of the relationship between energy and Mach number, as given by (18), with $\beta = 2/3$.

Many similaritites exist between this and the flow over a bump discussed in §2.1. The principal difference is that the varying geometry, specified by $A(x)$, appears as a multiplicative modification to the energy E, rather than as an additive modification in the shallow bump problem. This changes the behaviour surprisingly little. At a given point, there are still zero, one, or two solutions, with one sub- and one super-critical in the latter case. If the cross-sectional area decreases sufficiently, it can be that no steady solution exists, in which case the 'choking' effect causes an upstream disturbance that modifies the upstream boundary conditions.

2.3 Other examples

There are many other examples of transcritical behaviour occuring in many areas of fluid mechanics and physics in general. For example, calculations regarding the stellar wind result in a singular differential equation, forcing flow to pass through a critical point - the *Alfvén point* of magnetohydrodynamics.

An important geophysical example is the flow of a stratified fluid over an mountain. If the atmosphere could be assumed to have a rigid lid, then a weakly nonlinear analysis would proceed analogously to the shallow water flow discussed in §2.1. The same is not true in an unbounded region, but the phenomenology is similar, as shown by Pierrehumbert and Wyman [3]. Figure 6 illustrates their computational solutions, which show that the flow is subcritical upwind of a mountain, but supercritical on the lee-side in a typical geophysically-

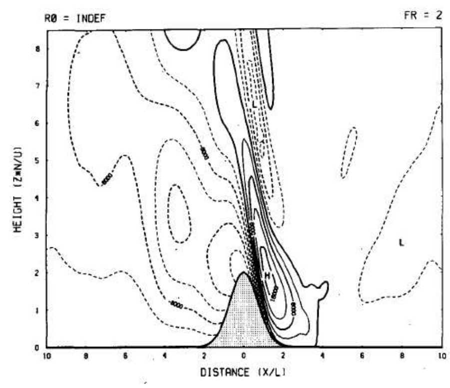

Figure 6: Streamlines of pertubation to uniform left-to-right flow past an idealized mountain, from [3].

reasonable region of parameter space. In their work, Pierrehumbert and Wyman [3] found that the dimensionless Froude number controlling the flow is given by

$$\mathrm{Fr} = \frac{U}{NH_m}, \tag{19}$$

where U is the local velocity scale, N the buoyancy frequency, and H_m a vertical lengthscale comparable to the height of the mountain. For small values of this parameter, a 'blocking' phenomenon occurs. Similar to the 'choking' behaviour described in the earlier examples, blocking generates an upstream-propagating columnar mode, which has the effect of modifying the upstream boundary conditions by creating a layer of stagnant fluid near the ground, effectively reducing the height of the mountain and allowing a steady flow to be reached.

3 Atmospheric escape

We have previously alluded to the fact that atmospheric escape flows are complicated by a number of factors. Solar radiation, gravitational potential and non-adiabatic effects all come into play, among many other effects. A brief discussion of the control of hydrogen in the early Earth atmosphere (see [6]) makes mention of several complications. In this section, we shall consider a simple, spherically-symmetric model for atmospheric escape, before mentioning some possible refinements and future directions for interesting research.

3.1 A spherically-symmetric model

In order to formulate a spherically-symmetric model for atmospheric escape, we follow a very similar approach to the compressible flow problem described in §2.2. In this case, conservation of mass may be expressed in the form

$$\Phi = 4\pi r^2 \rho(r) u(r) = \text{const.}. \tag{20}$$

Note that, in this case, there is no extremal 'cross-sectional area', equivalent to $A(x)$ in the case of the duct. We must therefore rely on a different mechanism to lead to transcritical behaviour. When considering the conservation of momentum, we must therefore include the gravitational potential $\chi = -GM/r$, where M is the mass of the planet, and G is the gravitational constant. Assuming an adiabatic gas, as in §2.2, we can arrive at the following form of the momentum equation

$$\frac{1}{u}\left(\mathcal{M}^2 - 1\right)\frac{\partial u}{\partial r} = \frac{1}{(c_s r)^2}(2rc_s^2 - GM). \tag{21}$$

For this formulation, is the clear that a singular point exists when $\mathcal{M} = 1$. At this point, one may only transition smoothly from a subsonic flow to a supersonic flow provided that

$$c_s(R_s) = \sqrt{\frac{GM}{2R_s}}, \tag{22}$$

where R_s is the *sonic radius*, also referred to as the *sonic point*. In this case, it is the gravitational potential term that plays the role of the bump in §2.1 or the constriction of the duct in §2.2.

Arranging the momentum equation into conservative form yields the equation

$$0 = \frac{\partial E}{\partial r} = \frac{\partial}{\partial r}\left[\frac{1}{2}u^2 + \int^{\rho}\frac{c_s^2(\rho')}{\rho'}\mathrm{d}\rho' - \frac{GM}{r}\right], \tag{23}$$

which allows us to define a specific energy E. For an ideal gas, as described in §2.2, it is possible to write

$$E = c_{s0}^2\left(\frac{\mathcal{M}_0 A_0}{\mathcal{M}A}\right)^{2\frac{1-\beta}{1+\beta}}\left(\frac{1}{2}\mathcal{M}^2 + \frac{\beta}{1-\beta}\right) - g_0 R_0\left(\frac{R_0}{r}\right), \tag{24}$$

where the terms represent specific kinetic, internal and potential energy, respectively. In the case of atmospheric escape, the temperatures are not usually large enough for internal heat to be a significant component of energy, so the principal balance is between kinetic and potential energies. This leads, as every good schoolchild knows, to the notion of an *escape velocity* $u_e = 2g_0R_0$, at which a particle has enough kinetic energy to escape the potential well of the planet. This balance is sufficient, in order-of-magnitude, to drive an outflow like the hydrogen-rich solar wind. However, for typical conditions in the oxygen-rich high terrestrial atmosphere, escape requires an additional input of energy.

3.2 Extensions and generalisations

In real planetary situations, such as that of the Earth, there is little reason to believe that any escape flow will be spherically-symmetric. Indeed, given the difference in solar radiation absorbed on the day- and night-sides of a planet, an entirely symmetric flow would be very surprising. It is possible that the heating the sunward side of the atmosphere by EUV radiation drives a flow to the dark side of the planet, where is then able to escape. The idea is that the radiatively-driven circulation forces the exospheric material into a jet directed away from the Sun. Calculations based on this notion have yet to be carried out.

Furthermore, our assumption that the gas is adiabatic is not valid for the case of atmospheric escape from Earth's atmosphere. This is clear from the calculations, such as that of §3.1, which show that there is insufficient energy for atmospheric escape of oxygen. It is therefore necessary to include some non-adiabatic effects, such as local heating by EUV radiation, in order to arrive at correct scalings for terrestrial atmospheric escape. It should be noted, however, that simple theories of adiabatic hydrodynamic escape perform admirably when applied to the solar wind.

An interesting model through which to consider non-adiabatic effects in hydrodynamical escape might be to return to transsonic flow in a duct, but to supply heat, uniformly or otherwise, to the gas in the duct. This heating should weaken the conditions leading to the minimum area rule, perhaps displacing the sonic point, or allowing for multiple sonic points. When generalised to atmospheric collapse, this could result in enough energy being supplied to allow material to escape.

It should be noted that hydrodynamical escape is not the only mechanism allowing material to leave Earth's atmosphere. The article by Catling [1] argues that hydrodynamic escape may not even be the dominant mechanism for hydrogen removal, and that so-called 'non-thermal' escape mechanisms are more significant. One hopes that a better understanding of the details of the various mechanisms may lead to a resolution of the argument, and a clearer picture of how planetary atmospheres can evolve.

References

[1] D. C. Catling, *Comment on "a hydrogen-rich early earth atmosphere"*, Science, 311 (2005), p. 38a.

[2] J. F. Kasting and J. B. Pollack, *Loss of water from venus. i. hydrodynamic escape of hydrogen*, Icarus, 53 (1983), pp. 479–508.

[3] R. T. Pierrehumbert and B. Wyman, *Upstream effects of mesoscale mountains*, J. Atmos. Sci., 42 (1985), pp. 977–1003.

[4] S. I. Rasool, *Loss of water from venus*, J. Atmos. Sci., 25 (1968), pp. 663–664.

[5] M. J. Thompson, *An introduction to astrophysical fluid dynamics*, Imperial College Press, 2006.

[6] F. Tian, O. Toon, A. A. Pavlov, and H. de Sterck, *A hydrogen-rich early earth atmosphere*, Science, 308 (2005), p. 1014.