Class 15 Notes: H I: Thermodynamics and Phase Structure

We now turn to the next phase of the ISM: atomic hydrogen. This is the most abundant phase of the ISM in the disk of the Milky Way, and in most other galaxies like it. Our goal today will be to understand the heating and cooling processes in these gas that control its thermodynamics, and to understand the reason we speak of H I as having distinct phases. We will then examine the interfaces between those phases.

I. Heating processes

As with H II, we will start by examining the heating processes that take place in this gas. There are only two significant ones in most cases: cosmic rays and the grain photoelectric effect. Other heat sources, such as x-rays, are of importance only in small regions around strong sources.

A. Cosmic rays

Galactic disks are filled with energetic cosmic rays that have long mean free paths. When these strike atoms, they cause ionizations, a process we studied several weeks back in our discussion of ionization processes. The free electrons created by this process typically have energies around 35 eV. Some of this energy will go into secondary ionizations, or into collisional excitation of H, H_2, or He followed by radiative de-excitation, but some of it will go into thermal energy as the electron scatters off other particles.

The amount that goes into heat depends on the chemical state of the gas with which the electron is interacting. If the gas is wholly or partially ionized, the cosmic ray-produced electron will Coulomb scatter off the large number of free electrons, rapidly thermalizing. As a result, almost all of its energy will go into heat. In predominantly neutral gas, the interactions will be charged-neutral instead. If the scatterings do not excite the target atom or molecule, then there will be negligible energy transfer. The electron loses energy mainly by causing excitations in the atoms or molecules it hits.

In atomic gas, the primary collision parters for the electron will be H atoms. When such an atom is excited, it is far more likely to radiatively de-excite than to collisionally de-excite, and if this happens the photon will usually escape and thus no heat will be added to the gas. Heat is added only on the rare occasions when the H atom collisionally de-excites. Calculations of the many relevant cross sections and branching ratios by Dalgarno & McCray suggest that the energy yield per cosmic ray-produced electron may reasonably be approximated by

$$E_h = 6.5\text{ eV} + 26.4\text{ eV}\left(\frac{x_e}{x_e + 0.07}\right)^{1/2}, \quad (1)$$

where $x_e = n_e/n_{\rm H}$. In the atomic ISM x_e is generally quite small, so the yield is close to 6.5 eV per CR ionization.

Recent calculations by Glassgold et al. (2012) suggest that the yield is a factor of $\sim$ 2 larger in regions with significant molecular content, because the many rotational and vibrational levels of H_2 provide additional conduits into which electrons can dump their energy, and because the lower Einstein A's of these transitions make it much more likely that molecules will collisionally rather than radiatively de-excite. Chemical reactions facilitated by the presence of free electrons and H_2^+ ions (created when the primary electron is knocked free) provide an additional heating source.

In the atomic gas, the cosmic ray heating rate per unit volume due to interactions with neutral atoms is

$$\Gamma_{\rm CR,n} = (n_{\rm H^0} + n_{\rm He})\, \zeta_{\rm CR} E_h, \tag{2}$$

where $\zeta_{\rm CR}$ is the primary cosmic ray ionization rate (i.e. not including secondary ionizations) that we computed earlier. For the observed galactic cosmic ray population, this is about 10^{-16} s^{-1}. Plugging this in, we obtain

$$\Gamma_{\rm CR,n} \approx 1.0 \times 10^{-27} n_{\rm H} \text{ erg s}^{-1} \zeta_{\rm CR,-16} \left[1 + 4.06 \left(\frac{x_e}{x_e + 0.07}\right)^{1/2}\right], \tag{3}$$

where $\zeta_{\rm CR,-16}$ is the cosmic ray ionization rate in units of 10^{-16} s^{-1}.

Cosmic rays can also interact with free electrons and heat them, but this effect is small compared to heating due to interactions with H and He atoms unless the ionization fraction is high.

B. Dust photoelectric heating

Another source of free electrons that can heat gas is dust grains. These have the great advantage over atoms and molecules that they are continuum absorbers, and the great advantage over cosmic rays that photons are much more abundant than cosmic rays in the ISM. The work function (the minimum energy required to eject an electron) for graphite is about 4.5 eV, so all photons in this energy range can produce photoelectrons, although the yield is very low until photon energies reach about 8 eV.

As a rough estimate, let $n(8 - 13.6 \text{ eV})$ be the number density of photons from 8 eV, where the photoelectric yield becomes significant, up to 13.6 eV, where the spectrum cuts off in the neutral ISM due to absorption by neutral hydrogen. For the observed interstellar radiation field of the Milky Way, this is about 3×10^{-3} cm^{-3}. The grain cross section to photons in this energy range is about $\langle \sigma_{\rm abs} \rangle \sim 10^{-21}$ cm^2 per H atom at Milky Way dust abundances, and the photoelectric yield per absorption is typically $\langle Y \rangle \sim 0.1$. Note that it is this small largely because the grains are on average quite negatively charged; for neutral grains it would be closer to unity In this case the rate of dust photoelectric heating per unit volume is

$$\Gamma_{\rm pe} = 1.4 \times 10^{-26} n_{\rm H} \text{ erg s}^{-1} \left(\frac{n(8-13.6 \text{ eV})}{3\times 10^{-3} \text{ cm}^{-3}}\right) \left(\frac{\langle \sigma_{\rm abs}\rangle}{10^{-21} \text{ cm}^2}\right) \frac{\langle Y\rangle}{0.1} \frac{\langle E_{\rm pe}\rangle - \langle E_c\rangle}{1 \text{ eV}}, \tag{4}$$

where $\langle E_{\rm pe}\rangle$ and $\langle E_c\rangle$ are the mean kinetic energy per photoelectron and the mean kinetic energy of electrons captured from the gas onto the grain, respectively. These depend on somewhat uncertain energy-dependent electron capture rates and photoelectric yields, and the estimate of a 1 eV difference is based on a combination of laboratory experiments and simple models. More sophisticated approaches first model the charge distribution of the grains, then use the charge-dependent photoelectric yields to evaluate this function, but our simple estimate is accurate to the order of magnitude level. The heating rate tends to be dominated by small grains, both because these provide most of the UV absorption, and because the large charges on larger grains tend to reduce their photoelectric yield and photoelectron energies.

The interesting thing to notice here is that this rate is an order of magnitude larger than the cosmic ray heating rate. Thus the grain photoelectric effect is the dominant heat source in the atomic ISM, at least under Milky-Way like conditions.

Because the grain photoelectric heating rate scales linearly with both the number density of hydrogen atoms (assuming a fixed dust to gas ratio) and the number density of UV photons, it is common to write it as

$$\Gamma_{\rm pe} = n_{\rm H} G_0 g(G_0/n_e, T_e), \tag{5}$$

where G_0 is the UV radiation field strength normalized to the Solar neighborhood value (or to be more precise the Solar neighborhood value as it was estimated by Habing (1968); current estimates are a factor of 1.6 higher, but we retain the estimate $G_0 = 5.29 \times 10^{-14}$ erg cm^{-3} from $6-13.6$ eV as our unit of measurement for historical reasons). The function g incorporates all the messy grain behavior. It depends on G_0/n_e, where n_e is the free electron number density, and on the electron temperature T_e, because these two quantities determine the charge equilibrium of the grains. The function g also depends on the grain size distribution.

II. Cooling processes: atomic lines

A. Dominant cooling lines

The main cooling process in the atomic ISM is, as in ionized gas, collisionally excited lines. Which lines dominate is determined by a number of factors

- More abundant species are obviously more important than less abundant ones.
- The ambient starlight spectrum contains many photons below 13.6 eV, and almost none above it. Thus the relevant ionization state is neutral for elements

with first ionization potential above 13.6 eV (e.g. O) or once ionized for elements with first ionization potential below 13.6 eV (e.g. C and Si).

- In order to cool effectively, the energy of the upper state of a particular line must be well matched to kT. Higher energy levels are not collisionally excited often enough to contribute significantly, and lower energy levels are excited often but do not remove much energy per collision. As a result, the dominant lines tend to be forbidden infrared fine structure lines, which generally have upper state energy levels at $\sim 100 - 1000$ K.

The most important coolant at low temperatures ($\ll 10^4$ K) is generally C II. Carbon is the most abundant element after H and He, and its first ionization potential is 11.26 eV, so in the atomic ISM most of the carbon is C II. This species has five electrons, so its ground electronic state is $1s^2 2s^2 2p^1$. Since there is only one electron in the outer shell, the $L-S$ state corresponding to this ground electronic state is trivial. The one electron has orbital angular momentum $\ell = 1$ and spin angular momentum $s = 1/2$, so the total orbital angular momentum is $L = 1$ and the total spin angular momentum is $S = 1/2$. The total angular momentum is either 1/2 or 3/2, depending on whether the spins are aligned or anti-aligned, so the $L - S$ state corresponding to the ground electronic state consists of a doublet $^2\mathrm{P}^o_{1/2}$ and $^2\mathrm{P}^o_{3/2}$. As one expects for a fine structure splitting, the energy difference is small: the $J = 3/2$ level lies 92 K above the $J = 1/2$ state. This energy difference is well matched to temperatures ~ 100 K.

The second most important low-temperature coolant is O I. Like C it is very abundant, and its first ionization potential is 13.618 eV, as opposed to 13.598 eV for H, so there are few photons capable of ionizing oxygen. Moreover, the nearly identical ionization potentials means that O and H can charge exchange easily, and this generally forces the oxygen ionization fraction to be closely tied to the H ionization fraction. O I has 8 electrons, so the ground electronic state is $1s^2 2s^2 2p^4$. The ground electronic state in this cases consists of three terms: ^{3}P, ^{1}D, and ^{1}S. The ^{1}S and ^{1}D terms are singlets, and both lie $\gtrsim 10^4$ K above ground, so they are not important at neutral gas temperatures. The ^{3}P state is a triplet that breaks into $^3\mathrm{P}_0$, $^3\mathrm{P}_1$, and $^3\mathrm{P}_2$, with the latter being the lowest energy state. The $^3\mathrm{P}_1$ is 228 K above ground, and the $^3\mathrm{P}_0$ is 326 K above ground. Since the direct $^3\mathrm{P}_0 \rightarrow {}^3\mathrm{P}_2$ transition has $\Delta J = 2$ it is strongly suppressed, and collisional excitations to $^3\mathrm{P}_0$ almost always decay by going into $^3\mathrm{P}_1$ and then $^3\mathrm{P}_2$ The first step produces a 14 μm photon, but most of the power comes out in the $^3\mathrm{P}_1 \rightarrow {}^3\mathrm{P}_2$ decay, which produces a photon at 63 μm. This process starts to contribute significantly compared to carbon once the temperature is above a several hundred K.

At temperatures that approach 10^4 K the cooling rate shoots way up because non-fine structure transitions and electronic transitions become possible. In particular, the Lyα line of hydrogen is 1.2×10^4 K above ground, and can provide a huge cooling effect for gas that reaches temperatures approach 10^4 K. This is generally the dominant cooling line for the warmest phases of H I.

In addition to these major lines, there are a host of minor lines that contribute at the $\sim 10\%$ level. Numerical calculations can take them into account.

B. Critical densities

The two dominant low-temperature cooling lines we have discussed have fairly high critical densities. For C II, the critical density for collisions with neutral H atoms is 3×10^3 cm^{-3}, and for collisions with free electrons it is 10 cm^{-3}. For O I, the critical densities are 2×10^5 cm^{-3}, 1×10^5 cm^{-3}, and 1×10^5 cm^{-3} for neutral hydrogen atoms, free electrons, and free protons, respectively. All of these densities are significantly higher than typical densities in atomic regions, so to good approximation we can simply assume that every collisional excitation is followed by an immediate radiative de-excitation that removes energy.

Since the collision rate varies with density as n^2, it is common to write the cooling rate as

$$\Lambda = n_{\mathrm{H}}^2 \lambda(T), \tag{6}$$

where $\lambda(T)$ is a function that depends on temperature only, and incorporates the temperature dependence of the collisional excitation rate coefficient for a particular line. Note that, in some sources, one instead sees the cooling rate written Λn^2, with Λ being the quantity that we have written $\lambda(T)$. Unfortunately there is no uniformity on this in the literature, so check units carefully to make sure you know which cooling function you're dealing with!

The temperature dependence $\lambda(T)$ behaves in the manner one would expect. For the neutral-neutral collisions $\lambda(T) \propto T^{1/2} e^{-E_{\mathrm{lev}}/kT}$ from the usual integration of the collision rate coefficient over the Boltzmann distribution. For ion-neutral collisions the variation is $\lambda(T) \propto e^{-E_{\mathrm{lev}}/kT}$, while for ion-electron collisions it is $\lambda(T) \propto T^{-1/2} e^{-E_{\mathrm{lev}}/kT}$.

III. Thermal equilibrium and the two phase model

A. The three phases

Given these heating and cooling processes, we can solve for the equilibrium temperature for a given density. In the process we must solve for the equilibrium ionization state too, since excitations due to collisions with free electrons can be significant even at low electron fractions, due to the higher thermal velocities and collision strengths of the electrons compared to the H atoms. Thus we simultaneously balance the heating rate against the cooling rate, and the ionization rate against the recombination rate, for a given density and abundance of elements. Such calculations have been carried out by a number of authors, most recently by Mark Wolfire and collaborators.

[Slide 1 – temperature vs. density from Wolfire+ 1995]

We can understand the general trend here by recalling that the main heating source, cosmic rays, provides an energy input per unit volume that varies as n_{H}, whereas the main cooling effects from collisionally-exicted lines remove energy at a rate that varies as n_H^2. We roughly expect

$$n_\mathrm{H} G_0 g(G_0/n_e, T_e) \approx n_\mathrm{H}^2 \lambda(T) \qquad \Longrightarrow \qquad \lambda(T) \approx \frac{G_0 g(G_0/n_e, T_e)}{n_\mathrm{H}}. \tag{7}$$

The UV field G_0 and the grain properties represented by the function g do not vary much with density or temperature, so the temperature-density relation is essentially dictated by the shape of $\lambda(T)$.

In the regime from $\sim 10^2 - \text{few} \times 10^3$ K where C II and O I cooling dominate, $\lambda(T)$ is a slowly increasing function of T. This means that the temperature must drop as the density increases, and that it must do so faster than $1/n_\mathrm{H}$.

This trend breaks at both low and high T. On the high T side, the equilibrium temperature cannot get much higher than $\sim 10^4$ K, because at that point the immense effects of Lyα cooling kick in and prevent the gas from getting any warmer as long as neutral H is present. Similarly, the equilibrium temperature in atomic gas stops declining with density once the temperature is $\sim 30 - 50$ K because the C II cooling rate has a temperature dependence $e^{-92\,\mathrm{K}/T}$. Once T drops significantly below 92 K, the line ceases to be able to cool effectively, and $\lambda(T)$ becomes an extremely steep function of T. As a result, only a small change in temperature is required to offset a large change in density.

It is instructive to plot this in a somewhat different way: pressure ($\propto nT$) versus density rather than temperature versus density. The motivation for this is that gas at a given position in a galaxy generally has a pressure that is fixed by its environment, e.g. the weight of the gas on top of it. Thus the density and temperature of the gas at that point have to adjust to produce the required pressure. If they cannot, the gas density will be changed by hydrodynamic motions until an equilibrium is found. The resulting plot of pressure versus density is a characteristic S-shaped curve, the general form of which was first obtained by Field, Goldmsith, and Habing in 1969. As a result, this is sometimes referred to as an FGH curve. You will derive a simplified version of this on your homework.

[Slide 2 – FGH curve from Wolfire+ 1995]

At low density, $\log n \lesssim -0.3$, the temperature is not far from its $\sim 10^4$ K ceiling, and the pressure increases close to linearly with density, since $P \propto nT$. As the density reaches $\log n \sim -0.5$, the temperature drops below 10^4 K, mostly due to the [O I] 63 μm line. Thus the pressure stops rising linearly with temperature. As the density continues to rise the cooling gets stronger and stronger, and eventually the equilibrium temperature declines faster than $1/n$. Thus the pressure decreases as the density increases. This continues until the temperature gets to ~ 100 K at a density $\log n \sim 0.5$. At this point the exponential dependence of the C II cooling curve starts to play a role, and $\lambda(T)$ varies sharply with T. Thereafter the temperature changes only fairly slowly, so the pressure-density relation becomes close to $P \propto n$ again.

If we move vertically above the line then T increases at fixed n. This does not change the heating rate, but it does raise the cooling rate, so in the region above the line cooling occurs faster than heating and the gas tends to cool. Conversely, in the region below the line, the reverse holds, and the gas tends to heat faster than it cools.

We can identify three phases based on the maxima and minima in this curve. The low density, higher temperature phase is known as the warm neutral medium, or WNM. The high density, lower temperature phase is the cold neutral medium, CNM. The intermediate density and temperature phase, we will see in a moment, is unstable.

Which phases are actually present depends on the pressure in the ISM, which is a function mostly of the large scale properties of the galaxy, i.e. how much pressure is required to maintain hydrostatic equilibrium for a given galactic potential and total amount of gas. Of course there are significant local variations around this number. For the Milky Way near the Solar circle the estimated mean ISM mid-plane pressure is $P/k \sim 3000 - 4000$ K cm^{-3}, and in this regime all three phases can exist. That appears to be the case for most spiral galaxies today.

B. Stability considerations

For a given pressure we have seen that there are three possible densities that can be in thermal equilibrium. However, only two of those equilibria are stable. To see this, it is helpful to write down the first law of thermodynamics:

$$de = T\,ds - P\,d(\rho^{-1}), \tag{8}$$

where e is the specific internal energy of a fluid, T is the temperature, s is the specific entropy, P is the pressure, and ρ is the density (so that $1/\rho$ the volume per unit mass).

The $P\,d(\rho^{-1})$ term represents work done by or on the gas, while the $T\,ds$ term represents heating or cooling processes, such as radiation. Thus for our H I that is subject to radiative heating and cooling, in a time dt during which radiation adds or removes an amount of heat dq, we can write

$$ds = \frac{dq}{T} = -\frac{\mathcal{L}}{T}\,dt, \tag{9}$$

where $\mathcal{L} = \Lambda - \Gamma$ is called the loss function. It is the net rate of radiative energy loss, with units of erg s^{-1} g^{-1} in cgs.

For a parcel of gas in thermal equilibrium, $\mathcal{L} = 0$, since heating and cooling balance, so the change in specific entropy ds is zero. Now consider perturbing a parcel of gas away from equilibrium by changing its entropy by an amount δs at time $t = 0$, in such a way as to leave some other thermodynamic variable A (for example pressure or density) fixed. How will the gas respond? Based on what we have just written down, we can write an evolution equation for the specific entropy perturbation:

$$\frac{d}{dt}(\delta s) = \delta\left(\frac{ds}{dt}\right) = -\delta\left(\frac{\mathcal{L}}{T}\right). \tag{10}$$

In words, the rate at which the entropy perturbation grows or shrinks is simply minus the rate at which the perturbation induces the gas to lose energy, divided by the gas temperature.

Suppose that δs and $\delta(\mathcal{L}/T)$ have the same sign, meaning that when we increase the specific entropy of the gas, the loss rate increases as well. In this case the evolution equation tells us that δs goes down. Thus radiation acts like a restoring force, and the entropy oscillates about its equilibrium value. On the other hand suppose that δs and $\delta(\mathcal{L}/T)$ have opposite signs. In this case an increase in entropy causes a drop in the loss rate, meaning that the gas heats up. In this case the gas is pushed further from equilibrium, and is unstable.

Thus the condition for instability is simply that

$$\left[\frac{d}{ds}\left(\frac{\mathcal{L}}{T}\right)\right]_A < 0, \tag{11}$$

where the subscript A indicates that the derivative is to be taken while holding the quantity A constant. This just says that the gas is unstable if $\mathcal{L}/T$ decreases with increasing specific entropy. Since T is always positive, we can equivalently write the stability condition as

$$\left(\frac{d\mathcal{L}}{ds}\right)_A < 0. \tag{12}$$

To apply this to the H I in the ISM, let us consider perturbing the entropy of the medium at constant pressure, since the constancy of the pressure, at least over long time scales and on average, will be enforced by hydrostatic balance in the galactic disk. In our FGH diagram, this corresponds to starting from a point on the equilibrium curve at a given pressure, then sliding left or right off the curve.

At constant pressure, a change in entropy ds is related to a change in temperature by

$$T\,ds = C_P\,dT, \tag{13}$$

where C_P is the specific heat at constant pressure. However, in order to keep the pressure constant, a change in temperature dT must also involve a change in density. Since $P = \rho kT/\mu$, where μ is the mean particle mass, we must have

$$T\,d\rho = -\rho\,dT \qquad \Longrightarrow \qquad d\rho = -\frac{\rho}{T}\,dT. \tag{14}$$

Thus a change perturbation to the entropy ds produced a corresponding perturbation to the temperature $dT = (T/C_P)\,ds$, and a perturbation to the density

$d\rho = -(\rho/T)\, dT = -(\rho/C_P)\, ds$. Plugging this in, we have

$$\left(\frac{d\mathcal{L}}{ds}\right)_P = \frac{T}{C_P}\left(\frac{\partial \mathcal{L}}{\partial T}\right)_\rho - \frac{\rho}{C_P}\left(\frac{\partial \mathcal{L}}{\partial \rho}\right)_T, \tag{15}$$

and the condition for instability becomes

$$\left(\frac{\partial \mathcal{L}}{\partial T}\right)_\rho - \frac{\rho}{T}\left(\frac{\partial \mathcal{L}}{\partial \rho}\right)_T < 0. \tag{16}$$

We can write this in an even more compact form if we recall that, since $P = (k/\mu)\rho T$, so

$$\frac{dP}{P} = \frac{d\rho}{\rho} + \frac{dT}{T}. \tag{17}$$

If we make this substitution, then it is clear the instability condition is simply

$$\left(\frac{\partial \mathcal{L}}{\partial T}\right)_P < 0. \tag{18}$$

Now let us apply this to the FGH diagram. Suppose we start at the leftmost equilibrium point, labelled F. We slide to the right at fixed density, so n increases. Since the pressure is fixed, this means T decreases. We move into a region where heating is stronger than cooling, so $\mathcal{L}$ decreases as well, and we conclude that, at this point $(\partial\mathcal{L}/\partial T)_P > 0$. This point is stable. If we slide left, the conclusion is the same: density decreases, temperature increases, and the loss function increases as well.

Clearly point H behaves exactly the same as point F. Point G is a different story. From there if we slide right n increases, T decreases, and $\mathcal{L}$ increases, because we move into the region where cooling is stronger than heating. Thus this point is unstable. Moreover, this analysis immediately provides us with the general principle: anywhere the curve of equilibrium pressure versus density has a negative slope is unstable. Thus we can identify that the central region of the curve is unstable.

C. Consequences of thermal instability

The existence of two stable phases and an intermediate temperature unstable one has important consequences for the structure of the atomic ISM. Most profoundly, it provides a mechanism for driving continual turbulent motions of gas, in effect converting the thermal energy of stellar FUV photons into kinetic energy of the gas. Thus the phase structure can be thought of as a principle driver of ISM turbulence. The typical velocities produced by this turbulence are of order the sound speed in the warm phase, which is highly supersonic with respect to the cold phase. A number of numerical simulations have explored this effect.

[Slide 3 – Koyama & Inutsuka (2002) simulation]

[Slides 4 and 5 – Piontek & Ostriker (2005) simulation]

As a result of this turbulent stirring, a significant fraction of the gas at any time is not in one of the stable phases, but is in the intermediate unstable phase. Indeed, obtaining the correct mass fraction in each of the two stable phases and in the unstable phase has proven to be a significant challenge for global models of the ISM, one to which we shall return once we have discussed supernova blast waves.

D. Structure of shocks

A final topic for today is the structure of the shocks produced by thermal instability and other drivers of supersonic motion in the neutral ISM. These turn out to be somewhat complicated by the presence of ions like C II and their associated electrons. The significance of the free electrons and ions is that they respond to the magnetic field that is present in the gas. We may think of the gas as consisting of two separate fluids: a magnetically-inert one consisting of neutral atoms and molecules, and a magnetized plasma consisting of the ions, electrons, and magnetic field. When the ionization fraction is low, the neutrals will encounter one another much more often than they encounter ions, and as a result they develop a Maxwellian velocity distribution. The ions interact with one another via long-range Coulomb forces, so they also adopt their own Maxwellian velocity distribution. The neutral and ions are only coupled weakly, via the rare ion-neutral scattering events.

Now consider the implications of the presence of this magnetized fluid. Since the ionization fraction is low, its density ρ_i is much lower than the density of the neutral fluid, ρ_n, and this gives it a very large signal speed in the ions. The Alfvén speed is

$$v_A = \frac{B}{\sqrt{4\pi\rho_i}} = 112\text{ km s}^{-1}\frac{B}{5\,\mu\text{G}}\left(\frac{100\text{ cm}^{-3}}{n_\text{H}}\frac{10^{-4}}{x}\right)^{1/2}, \tag{19}$$

where x is the ion fraction and we have adopted an ion mass of $m_i = 12m_\text{H}$, appropriate if most of the ion mass comes from carbon atoms. In contrast, the neutral gas has a signal speed equal to the sound speed,

$$c_n = \sqrt{\frac{kT}{m}} = 0.8\text{ km s}^{-1}T_2^{1/2}, \tag{20}$$

where we have used a mean particle mass $m = 1.4m_\text{H}$, appropriate for H and He in the standard cosmic abundance.

Thus we see that motions at a few km s^{-1}, the typical warm gas sound speed, are supersonic with respect to cold neutrals, but are highly sub-Alfvenic with respect to the ions. Motions at a few km s^{-1} in the ions are a factor of ~ 100 slower than the signal speed, comparable to speeds of ~ 10 m s^{-1} in air – comparable to the speed of a bicyclist. There are certainly no shocks associated with such motions.

The question then arises: what happens in this ion-neutral fluid when gas parcels collide at speeds that are supersonic with respect to the neutrals but sub-Alfvenic with respect to the ions? Consider an idealized problem in which we sit at a stationary interface between an upstream region and a downstream region. Upstream the ions and neutrals move together at speed v_s into the interface, at which they decelerate. The magnetic field strength is B_0. Far downstream of they move away from the interface at some smaller speed $v_p < v_s$, and the magnetic field strength is $B_p > B_0$. The difference in speeds $v_s - v_p$ is greater than c_n but smaller than v_A.

If the neutral and ion gasses were uncoupled, what happens at the interface would be clear. The ions would smoothly decelerate across the interface, since the motions are below their signal speed. The neutrals, since they would not find out about the change in velocity until they hit the interface, would experience a shock, producing a sharp jump in velocity and density at the interface. In reality, however, the ions and neutrals do collide. As a result, the neutrals upstream get an advanced warning about the upcoming interface via collisions with the ions, which find out via magnetically-mediated waves.

The results in this case can be calculated rigorously, but we will content ourselves with an order-of-magnitude calculation that exposes the basic physics. First, we can estimate the distance in front of the shock over which the ions and neutrals communicate by writing down the equation of momentum conservation for the ions. We do so treating them as a separate fluid, with the ion-neutral collisional coupling handled as a source term. Thus the equation is

$$\rho_i \left[\frac{\partial \mathbf{v}_i}{\partial t} + (\mathbf{v}_i \cdot \nabla)\mathbf{v}_i\right] = -\nabla p_i - \nabla \frac{B^2}{8\pi} + \frac{(\mathbf{B}\cdot\nabla)\mathbf{B}}{4\pi} + n_n n_i \langle \sigma v \rangle_{in} \mu \left(\mathbf{v}_n - \mathbf{v}_i\right). \quad (21)$$

Here p_i is the ion pressure, $\mathbf{B}$ is the magnetic field, $\mathbf{v}$ is the velocity for ions or neutrals as indicated, and the final term represents the rate at which collisions between ions and neutrals exchange momentum between them. The term in angle brackets is the ion-neutral collision rate coefficient, and μ is the ion-neutral reduced mass. Taking the simplest case where the velocities are solely in the x direction, and magnetic field is parallel to the interface (so it is in the yz plane), this reduces to

$$\rho_i v_i \frac{dv_i}{dx} = -\frac{dp_i}{dx} - \frac{1}{8\pi}\frac{d}{dx}B^2 + n_n n_i \langle \sigma v \rangle_{in} \mu (v_n - v_i), \quad (22)$$

where the v's now refer to the x components of the velocity.

Since the ions have very little inertia or pressure, the two dominant terms are the ion-neutral collision term and the magnetic pressure term, and they must approximately balance:

$$\frac{1}{8\pi}\frac{d}{dx}B^2 \approx n_n n_i \langle \sigma v \rangle_{in} \mu (v_n - v_i). \quad (23)$$

If the region over which the ions and neutrals interact has characteristic size L, the characteristic magnetic field strength in this region is $\sim B_0$, and the characteristic difference in speeds between ions and neutrals in this region is $\sim v_s$, then at the order of magnitude level we have

$$\frac{B_0^2}{8\pi L} \sim n_n n_i \langle \sigma v \rangle_{in} \frac{m_n m_i}{m_n + m_i} v_s. \tag{24}$$

$$L \sim \frac{1}{2}\left(\frac{B_0^2}{4\pi n_n m_n}\right) \frac{m_n + m_i}{m_i} \frac{1}{n_i \langle \sigma v \rangle_{in} v_s} \tag{25}$$

$$= \frac{v_{A,n}^2}{2v_s} \frac{m_n + m_i}{m_i} \frac{1}{n_i \langle \sigma v \rangle_{in}} \tag{26}$$

$$\approx 10^{15} \text{ cm} \left(\frac{v_{A,n}}{\text{km s}^{-1}}\right)^2 \left(\frac{10 \text{ km s}^{-1}}{v_s}\right) \left(\frac{0.01 \text{ cm}^{-3}}{n_i}\right), \tag{27}$$

where $v_{A,n}$ is the Alfvén speed computed with respect to the neutral gas density (rather than the ion density), and in the numerical evaluation we have taken $m_n = 1.4 m_{\mathrm{H}}$, $m_i = 12 m_{\mathrm{H}}$, and $\langle \sigma v \rangle \approx 10^{-9}$ cm^3 s^{-1}. Thus we see that the interaction region gets larger as the magnetic field strength increases (larger $v_{A,n}$), the shock velocity decreases, or the ion density decreases.

The total momentum transmitted by the ions to the neutrals in this interaction region is, to order of magnitude, just the ion-neutral momentum exchange rate multiplied by the time it takes the neutrals to traverse the interaction region. Thus the change in neutral velocity in this region is roughly

$$\Delta v_n \sim (n_i \langle \sigma v \rangle_{in}) \left(\frac{\mu}{m_n} v_s\right) \frac{L}{v_s} \sim \frac{v_{A,n}}{v_s} v_{A,n}. \tag{28}$$

We may think of the first term as representing the ion-neutral collision rate, the second as the mean velocity change per collision (where we assume that the relative ion-neutral speed is of order v_s), and the third as the amount of time for which a neutral is in the interaction region.

Thus we see that if $v_{A,n} \gtrsim v_s$, the change in velocity in the interaction region is comparable to the total velocity change v_s required to traverse the interface. In this case all the momentum transfer required to decelerate the neutrals occurs in the interaction region, and there is no shock at all. Instead, the momentum is transfered by ion-neutral collisions. We refer to this case as a C-shock, for continuous, since in this case all the fluid variables change continuously. If $v_{A,n} \lesssim v_s$, the momentum transfer in the pre-shock region is insufficient to decelerate the neutrals fully, and there is still a shock in the neutrals. We refer to this as a J-shock, for jump, since the neutral velocity still jumps discontinuously.

[Slide 6 – C and J shocks from Draine & McKee (1993)]

Radiative cooling can further complicate this picture, allowing a third type of shock, called C*. We will not discuss shocks of this type here.