5 Gas Dynamics

The second part of this course is devoted to understanding the gas in galaxies. While the Milky way has nearly ten times as much mass in stars as in gas, it is clear that all of the stars must have formed out of gas, so we can not ignore this component. Moreover, even now stars are forming out of gas, and so much of the interesting physics in galaxy evolution and formation is in the evolution of the gas (and dust) component.

As we discussed in the introduction, when the mean free path between collisions for a set of atoms is much less than any other length scale of interest, then we can treat the set of atoms as a gas, and use the gas dynamics equations. The equations themselves are more complicated than the simple Newton equations but in many respects, they are easier to understand. For example, the velocity dispersion of atoms in a gas is well-represented by the Maxwell-Boltzmann distribution.

In the following sections, we will first examine the components of the Interstellar medium and then derive the gas equations (I will sometimes refer to these as the fluid equations). Then we will examine simple solutions that can be written down for propagating sounds waves and investigate the gravitational instability.

5.1 Observations of the ISM

First, we will briefly summarize the primary constituents of the interstellar medium, and how it is observed. The gas itself is largely hydrogen (74% by mass) and helium (24%) and other, heavier elements (2%). Often the heavier elements combine to form dust grains, which are micrometer to millimeter sized concentrations. Dust of course is quite opaque at optical wavelengths. In dense regions, the gas can be largely molecular (e.g., H_2, CO).

How the ISM is observed

The ISM is observed through a large range of techniques, the more common of which are summarized here.

Absorption of starlight was one of the first ways to observe the ISM, and remains very important. Dust, of course, obscures background objects at UV and optical wavelengths, but there are also absorption lines of individual atoms, such as CaII and NaI (for neutral gas), and CIV and OVI (for ionized gas). The advantage of this sort of approach is that it can be quite sensitive to small amount of intervening material. The downsides are that it requires complicated corrections to know how much of the gas is in the form of that particular ionized state (e.g. CaII vs. CaI), and that we measure only the velocity of the absorbing gas, not its distance directly. Both of these make it difficult to work out the distance and amount of interstellar gas.

Gas which is a bit denser and warmer often emits line radiation directly. These optical and UV **emission lines** include Hα, which is the $n = 3$ to $n = 2$ transition of hydrogen. It turns out that this emission usually comes from gas which is being ionized by massive stars (HII regions, which we will discuss in more detail later) and so is an excellent diagnostic for the amount of ongoing star formation in a region. Other emission lines are important as diagnostics of the density and temperature of the emitting gas. The line ratios of OIII and OII, for example, can be used to work out the gas temperature, because of the different energies required to excite the electrons.

Radio emission lines are also important, and one of the most important is the 21 cm hyperfine transition of neutral hydrogen. We have mentioned this line previously in the context of measuring galactic rotation curves, but it is also a good probe of the total amount of neutral hydrogen gas. While molecular hydrogen is difficult to observe directly, the rotational transitions of CO are a good tracer for dense molecular gas. These transitions fall in the radio spectrum.

While the 21 cm line traces the HI distribution, the dust can be distributed somewhat differently (in particular, it also traces dense regions where the HI has been converted to H_2). The dust can be observed both due to its obscuring effects, but more directly from it's **infrared emission**. The dust is usually quite cold ($T \sim 10-500$ K) and emits mostly black-body radiation, although there are some dust line features which can be used to probe the chemical composition of the dust.

The **X-ray emission** mostly traces hot gas with temperatures of 10^6 K and above. At these temperatures, the emission is both through lines of high-energy atomic transitions, but also bremsstrahlung radiation. In spiral galaxies, hot gas is mostly generated by supernovae explosions, but elliptical galaxies are dominated by their hot gas and generally have very little cold gas and dust.

Synchrotron radiation is produced by electrons spiraling along magnetic fields lines. To be observed, the electron energies have to be quite large, and cosmic-rays are generally the most important source of synchrotron emission in the galaxy. Therefore, synchrotron traces both the high-energy electron population and the magnetic fields.

How the gas and dust is distributed in disk galaxies

The above observations have led us to the following basic picture for how matter is distributed in our galaxy. First, the gas is distributed in a way much like the thin disk, with a similar scale-length extending out to probably even larger distances than the stellar distribution (this is certainly true for most other spiral galaxies). The width of the gas disk is even thiner than the stars with a scale height of around 100 pc (compared to 300 pc for the stars).

The gas in the disk is not smoothly distributed. In particular, it is clear that small clouds form with varying temperatures and densities. Observations indicate that these clouds generally have fall into one of three different typical states, or **phases**. These phases are designated as the **Cold Neutral Medium** (CNM), the **Warm Neutral Medium** (WNM, or sometimes if the gas is ionized, the WIM), and the **Hot Ionized Medium**, (HIM), which is more commonly known as the coronal gas. These three phases are summarized in the following table.

phase	density (cm^{-3})	Temperature (K)
CNM	10	10^2
WNM	10^{-1}	10^4
HIM	10^{-3}	10^6

Although this is an oversimplified description, we shall see that there is some use to this description. In fact, it is immediately clear, that assuming an ideal gas where the pressure is $P = nkT$, the three phases are in pressure equilibrium.

The HI is distributed throughout the disk, although it is particularly dense in the spiral arms. In the spiral arms, the density is high enough that it can form molecules, and we see the formation of **Giant Molecular Clouds** (GMCs) largely in the spiral arms. These clouds have masses in the range $10^3 M_\odot$ to $10^6 M_\odot$. Note that the center of the Milky Way and other galaxies has gas which is sufficiently dense that most of the hydrogen is in molecular form, giving the impression of a hole in the HI distribution.

Finally, it is worth pointing out that not all gas is in galaxies, in fact, probably the majority of all gas is outside of galaxies. Evidence for this comes from absorption seen in the spectra of distant quasars; this absorption is almost surely due to clouds of intergalactic hydrogen (and helium, and some small fraction of heavy elements).

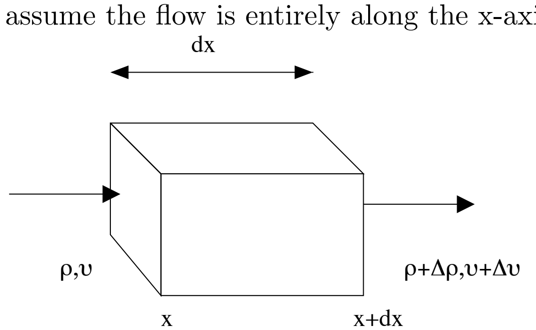
5.2 The Equations of Gas Dynamics

In this section, we will derive the equations which describe how gas moves under the influence of internal and external forces. We describe the gas with three related quantities. These are the gas density $\rho(\mathbf{x})$, velocity $\mathbf{v}(\mathbf{x})$, and pressure $P(\mathbf{x})$ (and possibly energy $e(\mathbf{x})$. Through most of this course we will deal with the gas dynamic equations in only one dimension, although there three-dimensional form is not much more complicated. Note that in order to define the density, velocity and pressure of a parcel of gas we must restrict ourselves to regions which contain enough atoms that these quantities can reasonably be defined.

Our guiding principle, or rather guiding principles, will be the conservation of mass, momentum and energy. These simple ideas will prove to be very powerful.

Mass conservation

We begin with mass conservation, because it is the simplest and most basic. We apply it to a box and imagine gas flowing in one side of the box with density ρ and velocity $u(x)$. The box has a length Δx and gas flows out the other side with density $\rho + \Delta\rho$ and velocity $u + \Delta u$. The area of the box sides is A. We will assume the flow is entirely along the x-axis.



Mass conservation tells us that the change in mass ΔM in some time Δt must be the mass flowing in one side minus the mass flowing out the other. In other words,

$$\Delta M = \rho A u \Delta t - (\rho + \Delta\rho) A (u + \Delta u) \Delta t$$

Or, re-arranging this and defining $\Delta M = A\rho\Delta x$, we find

$$\frac{\Delta\rho}{\Delta t} = -\left[\frac{(\rho + \Delta\rho)(u + \Delta u) - \rho u}{\Delta x}\right]$$

If we take Δ to be an infinitesimally small change, then this becomes a partial differential equation (partial because $\rho(x,t)$ and $u(x,t)$ depend on more than one variable):

$$\frac{\partial\rho}{\partial t} = -\frac{\partial(\rho u)}{\partial x}$$

Or, as this is sometimes written:

$$\frac{\partial\rho}{\partial t} + \rho\frac{\partial u}{\partial x} + u\frac{\partial\rho}{\partial x} = 0 \tag{56}$$

This is our first equation and tells us how the density at some point in space changes due to the flow of gas into and out of that spot. It is not a complete description because we do not have any equation which tells us how the velocity changes. This is next.

Momentum conservation

We repeat nearly the same line of argument as before, but this time replace mass with momentum S, so that the change in momentum during some time Δt is given by

$$\Delta S = S_{\text{in}} - S_{\text{out}} + F\Delta t - (F + \Delta F)\Delta t.$$

In this case we not only have the flow of momentum, but the change in momentum due to the pressure force acting on the sides of the box (recall that $dS/dt = F$). Since the pressure $P = F/A$ (force per unit area), and $S = Mu = \rho A\Delta x u$, this becomes

$$\frac{\Delta(\rho u)}{\Delta t} = -\left[\frac{(\rho + \Delta\rho)(u + \Delta u)^2 - \rho u^2}{\Delta x}\right] - \frac{(P + \Delta P) - P}{\Delta x}$$

Once again, we can take Δx and Δt to be arbitrarily small so that

$$\frac{\partial(\rho u)}{\partial t} = -\frac{\partial(\rho u^2)}{\partial x} - \frac{\partial P}{\partial x}$$

or we can expand this, and use the mass equation (eq. 56) to simply, to get an equation for the time derivative of u alone:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -\frac{\partial P}{\partial x}$$

This then is the second of our fluid dynamics equations. We can add gravity on to this by examining the change in momentum due to the gravitational force. When we work it through, we find

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -\frac{\partial P}{\partial x} - \frac{\partial\Phi}{\partial x} \tag{57}$$

Energy conservation

We could repeat this process for the conservation of energy and write down an equation for the evolution of the specific energy e (energy per unit mass). We would find,

$$\frac{\partial e}{\partial t} + u\frac{\partial e}{\partial x} = \frac{P}{\rho}\frac{\partial u}{\partial x}$$

Notice that we still do not have a complete set of equations because we have four unknowns ρ, u, e, P and only three equations. A fourth relation must be established for the pressure, and generally this depends on the type of gas or fluid. A common assumption is that the gas is ideal, in which gas

$$P = (\gamma - 1)\rho e$$

where γ is the ratio of specific heats and is equal to 5/3 for a a monatomic gas. A formula of this type which relates P to e and ρ is known as an **equation of state**. It is often common to define a temperature T which, for an ideal gas, is related to the pressure via

$$P = nkT \tag{58}$$

In fact, we will not use the energy equation in this course, and will instead depend on an simpler description of the pressure. In general we will assume that the pressure is directly related to the density via:

$$P = K\rho^\gamma$$

where K is a constant and γ will be either 5/3 or 1. This type of relation between pressure and density is known as the **polytropic** equation of state. The case of $\gamma = 5/3$ is sometimes known as the *adiabatic* equation of state. When $\gamma = 1$, the pressure is directly proportional to the density, and from eq.(58, this is the same as saying the temperature is constant. Therefore, $\gamma = 1$ is also refereed to as an **isothermal** equation of state.

Fluid Equations: summary

Here we repeat the (simplified) fluid equations that we will use:

$$\frac{\partial \rho}{\partial t} + \rho\frac{\partial u}{\partial x} + u\frac{\partial \rho}{\partial x} = 0 \tag{59}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = -\frac{\partial P}{\partial x} - \frac{\partial \Phi}{\partial x} \tag{60}$$

$$P = K\rho^\gamma \tag{61}$$

Imperfect gases: Heat conduction and viscosity

So far, we have assumed the gas is *perfect* in the sense that there are no particle effects that might introduce forces between neighboring gas elements beyond what we have written. In fact, this is not generally true. The particle nature of gases mean that particles typically move a mean free path $l_{\rm mfp}$ between collisions. When the gas properties change over this length-scale, an atom of gas will collide will other atoms with a different temperature or density. This will conveying information.

When the particle collides with atoms of a different temperature, this is known as **heat conduction**. Heat (energy) is communicated from one part of the fluid to another. Heat conduction is very familiar in solids, but it also occurs in gasses.

When the particle collides with atoms moving at a different bulk velocity u, then momentum is conveyed and this effect is known as viscosity. Once again, viscosity is a familiar property of a fluid.

In general, in astrophysics, these sorts of collisional effects are not important, although there are a number of important examples. One of these comes about when there are discontinuities in the flow – in other words, when the gas properties such as density and temperature, change very rapidly over a very short space. As we will see in following lectures, an important example is a shock. In turns out that although we will represent shocks as arbitrarily thin changes in the fluid properties, the density and temperature in a shock change over a mean free path length $l_{\rm mfp}$. This is because if the shock were to be narrower than this length, viscosity would act to smear it out.

5.3 Sound Waves

We now turn to a number of solutions of these equations. The first is the effect of small perturbations in the fluid equations *without* gravity. Let's take a uniform gas with density ρ_0, pressure P_0 and zero velocity. This is a trivial solution to the fluid equations (check it!).

Now we perturb this solution slightly and write the solution as the sum of the uniform part plus the perturbation:

$$\begin{aligned} \rho(x,t) &= \rho_0 + \rho_1(x,t) \\ u(x,t) &= u_1(x,t) \\ P(x,t) &= P_0 + P_1(x,t) \end{aligned}$$

where in each case we assume the perturbation is small so that $\rho_1 \ll \rho_0$ and $P_1 \ll P_0$. This appears to make the solution more complex but in fact it simplifies things be removing "non-linear" terms. Let's see how this works by putting these solutions into eq. (61). The density equation (derived from mass conservation) becomes:

$$\frac{\partial \rho_1}{\partial t} + u_1 \frac{\partial \rho_1}{\partial x} + (\rho_0 + \rho_1)\frac{\partial u_1}{\partial x}$$

Notice that the middle term is the product of two of these "perturbed" values (ρ_1 and u_1). If each one is very small, then the product of two of them will be negligibly small. Therefore we can neglect this term. The same is true of the ρ_1 part of the third-term so that this equation becomes

$$\frac{\partial \rho_1}{\partial t} + (\rho_0)\frac{\partial u_1}{\partial x} \tag{62}$$

Now, let's turn to the second fluid equation, that we derived from momentum conservation. This is

$$\frac{\partial u_1}{\partial t} = -\frac{1}{\rho_0}\frac{\partial P_1}{\partial x} \tag{63}$$

where once again we have neglected two non-linear terms. We are looking to write down a single closed equation for ρ_1 so let's re-write P_1 in terms of ρ_1 by writing:

$$\frac{\partial P_1}{\partial x} = \frac{\partial P_1}{\partial \rho_1}\frac{\partial \rho_1}{\partial x} = c_s^2 \frac{\partial \rho_1}{\partial x}$$

where we have defined c_s for convenience. We can remove u_1 from these two equations by differentiating eq. (62) with respect to t and differentiating eq. (63) with respect to x to get

$$\begin{aligned} \frac{\partial^2 \rho_1}{\partial t^2} &= -\rho_0 \frac{\partial^2 u_1}{\partial x \partial t} \\ \frac{\partial^2 u_1}{\partial t \partial x} &= -\frac{c_s^2}{\rho_0}\frac{\partial^2 \rho_1}{\partial x^2} \end{aligned}$$

Flipping the order of the x and t differentials for u_1 allows us to eliminate u_1 entirely, and we find

$$\frac{\partial^2 \rho_1}{\partial t^2} = c_s^2 \frac{\partial^2 \rho_1}{\partial x^2} \tag{64}$$

which is the **wave equation**. To show these, we can take a sinusoidal solution of the form

$$\rho_1(x,t) = A\cos(kx - \omega t)$$

Substituting this into the wave equation, we find that this is a solution for arbitrary A as long as

$$\omega = c_s k$$

. It is fairly easy to convince yourself that this is the solution for a traveling wave with velocity c_s (to do so, note that the peak of cosine wave occurs at $kx - \omega t = 0$, or when $x = c_s t$). The wave repeats itself when $kx = 2\pi$, so the wavelength of the wave is $\lambda = 2\pi/k$. Similarly, the period $T = 2\pi/\omega$.

This equation has a number of nice properties. It is linear so it supports the principle of superposition – that is, if you have two solutions ρ_A and ρ_B (with different wavelengths), then the sum $\rho_A + \rho_B$ is also a solution.

What about the speed of sound? Remember that $c_s^2 = \partial P/\partial\rho$, which we can evaluate for a given equation of state. For example, let's take $P = K\rho^\gamma$, in which case

$$c_s^2 = \frac{\gamma P}{\rho} = \frac{\gamma kT}{\mu}$$

where for the last equality we have assumed the ideal gas law $P = nkT$ and $\rho = \mu n$.

5.4 The Gravitational Instability

In the above analysis, we *linearized* the fluid equation by neglecting the non-linear terms. This produced an equation for how small perturbations propagate. We can repeat this including gravity. We have the same set of equations as above, except for two changes. First, we include the $\partial\Phi/\partial x$ term in the momentum equation, and second we include the one-dimensional Poisson equation:

$$\frac{\partial^2\Phi}{\partial x^2} = 4\pi G\rho$$

The actual analysis progresses very much as before. When the momentum equation is differentiated with respect to x, the gravitational acceleration term becomes equal to $\partial^2\Phi/\partial x^2$ and we get:

$$\begin{aligned}\frac{\partial^2\rho_1}{\partial t^2} &= -\rho_0\frac{\partial^2 u_1}{\partial x\partial t}\\ \frac{\partial^2 u_1}{\partial t\partial x} &= -\frac{c_s^2}{\rho_0}\frac{\partial^2\rho_1}{\partial x^2} - \frac{\partial^2\Phi_1}{\partial x^2}\end{aligned}$$

This last term can be replaced with the Poisson equation. But note that there is subtle point here, because th Poisson equation should really have ρ and Φ replaced with $\rho = \rho_0 + \rho_1$, and $\Phi = \Phi_0 + \Phi_1$. However, there is no solution of the Poisson equation with a non-zero ρ_0 (this implies an infinite universe with matter everywhere and so Φ_0 should be infinite), so we must assume that we can simply write

$$\frac{\partial^2\Phi_1}{\partial x^2} = 4\pi G\rho_1$$

If this is true, we can combine the above three equations to get a single equation for ρ_1 again:

$$\frac{\partial^2 \rho_1}{\partial t^2} = c_s^2 \frac{\partial^2 \rho_1}{\partial x^2} + 4\pi G \rho_0 \rho_1 \tag{65}$$

This is similar to the wave equation we derived earlier, but has an additional term on the right-hand side.

We can once again postulate a solution of the form $A\cos(kx - \omega t)$ in which case we get the dispersion relation

$$\omega^2 = c_s^2 k^2 - 4\pi G \rho_0$$

If the wavelength λ is small so that $k = 2\pi/\lambda$ is large, then ω will be large and the density term will be small and we will get back the same sound waves as before. Another way to put this is if the density ρ_0 is small enough, then the last term will be small. This means that in the small wavelength limit (or, equivalently, the low density limit), gravity is unimportant and we just get oscillating waves.

On the other hand, if the λ is increased (or ρ_0 is increased), then k gets smaller and smaller until

$$c_s^2 k^2 - 4\pi G \rho_0 < 0$$

so that $\omega^2 < 0$ (which implies that ω is an imaginary number). At this point a better solution is one of the form $\rho_1 = A\exp(\alpha t)\exp(kx)$ where $\alpha = -\omega$ so that α is positive. This is an exponentially growing solution and so the density ρ_1 keeps on getting larger and larger with time. This, of course, is known as the gravitational instability, or the Jeans instability.

There is a critical wavelength when $\omega = 0$ and the solution is balanced between oscillations and growing. This occurs when $k_J = 4\pi G \rho_0 / c_s^2$ or in terms of the wavelenth:

$$\lambda_J = \left(\frac{\pi c_s^2}{G \rho_0}\right)^{1/2}. \tag{66}$$

This is known as the Jeans wavelength. Waves with wavelengths smaller than this value oscillate as sounds waves, while larger waves grow in amplitude. We can define the Jeans mass as

$$M_J = \frac{4\pi G}{3} \lambda_J^3 \rho_0 = \frac{\pi \rho_0}{6} \left(\frac{\pi c_s^2}{G \rho_0}\right)^{3/2} \tag{67}$$

Applications of the Jeans Equation

We can now ask the question, what is the predicted Jeans length for the three phases of the ISM discussed earlier. First we note that because the sound speed is proportional to $c_s \propto T^{1/2}$, the Jeans length scales as

$$\lambda_J \propto \left(\frac{T}{\rho}\right)^{1/2}$$

and $M_J \propto T^{3/2} \rho^{-1/2}$. We begin with the hot phase, and for the typical densities and temperature listed earlier, we find $\lambda_J \sim 500$ kpc for the hot phase, much larger than the galaxy. Therefore, to a very good approximation, the hot phase is not gravitationally unstable. The warm phase has a Jeans length of about 5 kpc, which is about the same size as the galaxy. Therefore, only large wavelength modes in the warm phase are unstable.

The cold phase on the other hand, has typical jeans lengths of 50 pc (or even smaller – a few pc – for the coldest, densest part of this gas). In fact, this, as we shall see, is the typical length scale for giant molecular clouds, and we can conclude that it is the cold phase which is the source of star formation, since it is only this phase which can possibly have wavelengths larger than this critical value

This process of gravitational fragmentation and star formation forms an important part of the life-cycle of gas in the ISM. A simplified picture might be what hot gas cools, first to warm and then to cold gas. This cold gas becomes gravitationally unstable and forms stars. The massive stars then generate winds, ionizing photons and eventually explode as supernovae, all of which return energy back into the ISM, heating some of the cold gas back into the warm and hot phases, to begin the cycle again.

5.5 Shock Waves

In the previous section, we discussed solutions of the fluid equations involving small perturbations away from a uniform state, but there are of course, many other types of solutions. One important set involves discontinuities in the flow – that is places where the density, temperature, pressure and velocity change very suddenly. There are two classes of discontinuities – those in which the pressure and velocity change suddenly and those in which the they does not. The first type are called **shocks**, while the second are known as **contact discontinuities**. This second type is very easy to understand – they are places where, for some reason, the density and temperature change but their product — the pressure — does not. Since there is no change in the pressure, the pressure gradient term is zero and there is no net force on the case. In fact, it is this kind of solution which occurs at the edges of clouds between the different phases.

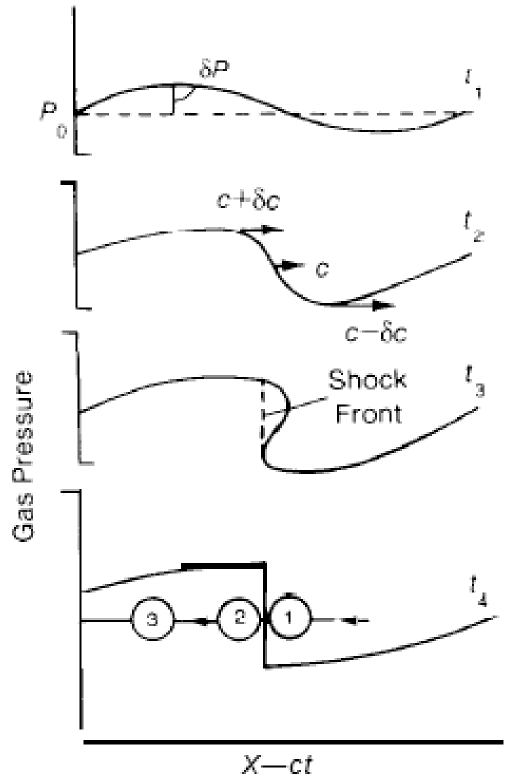
Waves steepening into shocks

Shocks on the other hand are quite different and more complex. There are a number of reasons shocks can occur, for example supernovae explosions or collisions between clouds. However, it is worthwhile first considering another wave to form a shock, from the steepening of a sound wave. Let's go back to the cosine sound wave we considered before. In the wave equation we derived, we assumed that the sound speed could be effectively given by the unperturbed pressure and density (P_0 and ρ_0), but in fact a more careful derivation shows that the wave travels at slightly different speeds along its length. In particular, the sound speed is

$$c_s^2 = \frac{\partial P}{\partial \rho} = \gamma K \rho^{\gamma-1}$$

where we have used our simplified equation of state to get the second equality. This shows that (provided $\gamma \neq 1$) that the sound speed is faster at the peak where the density is higher and slower in the trough. This is shown schematically in the figure at right (taken from a Los Alamos National Lab publication), where we can also see the steepening of the wave as it continues to propagate. Eventually, if it were allowed to, the peak would overtake the trough of the wave and look like a breaking water wave. In fact, because a sound wave is longitudinal rather than transverse, this is not permitted and instead a discontinuity forms in the pressure (and in all the other quantities as well).

In fact, in the ISM, this steepening process is not a significant source of shocks, but it does illustrate one important point which is that there is a process which is acting to keep the discontinuity sharp. Of course, as we highlighted earlier, the gas is really composed of discrete atoms and this particle nature prevents the shock from becoming too steep — viscosity acts to smear it out. The competition between these two processes keeps the shock stable with a thickness of a few mean free paths.



The Shock Jump Conditions

To derive relations between the density, pressure and velocity on either side of a shock, we go back to the mass, momentum and energy conservation equations that we used to derive the fluid equations. We denote the pressure, density and velocity on the pre-shock side with P_0, ρ_0 and u_0, and use P_1, ρ_1 and u_1 for the post-shock quantities. We assume that we are in a frame in which the shock itself is stationary.

We begin with mass conservation. Clearly all the matter flowing into a shock has to flow out the other side so

$$\rho_0 u_0 = \rho_1 u_1. \tag{68}$$

We can see what must be going on by imaging drawing a box with side area A that cuts across the shock. During a time Δt the amount of mass that flows into the box on the right is $\Delta t u_0 A \rho_0$, while the amount of mass flowing out is $\Delta t u_1 A \rho_1$. Canceling Δt and A gives us the earlier result.

Momentum is more complicated because the pressure is different on either side, but once again we can appeal to our box. The momentum flowing in is just the mass times the velocity, but now each side of the box feels a force $P_0 A$ and $P_1 A$ in the other direction, so the momentum changes by the difference between these forces multiplied by Δt. Once again we can cancel $A\Delta t$ and we get

$$\rho_0 u_0^2 + P_0 = \rho_1 u_1^2 + P_1 \tag{69}$$

Finally, the energy is very similar in that we have energy conservation, but now we have both kinetic energy which is the mass times $u^2/2$ and also the thermal energy which is (for an ideal gas with $\gamma = 5/3$) equal to the mass times $3P/2\rho$. We add to this the energy added due to the pressure forces (work equals force times distance so this is $F u \Delta t = P u A \Delta t$. The result is (after canceling all the $A\Delta t$ terms):

$$\frac{1}{2}u_0^2(\rho_0 u_0) + \frac{3}{2}P_0 u_0 + P_0 u_0 = \frac{1}{2}u_1^2(\rho_1 u_1) + \frac{3}{2}P_1 u_1 + P_1 u_1$$

or dividing through by $\rho_0 u_0$ and simplifying we find

$$\frac{1}{2}u_0^2 + \frac{5P_0}{2\rho_0} = \frac{1}{2}u_1^2 + \frac{5P_1}{2\rho_1} \tag{70}$$

Finally, we can use the simplified equation of state $P = K\rho^{\gamma}$ to write the sound speed as $c_s^2 = 5P/3\rho$. It is also typical to define the **Mach number** as the ratio of the velocity of the pre-shock gas to the sound speed in the gas

$$M_0 = \frac{u_0}{c_0}$$

These equations, in particular, eq. (68) to (70) are three equations for six unknowns (the pressure, density and velocity on either side of the shock) . This means that we can only solve for some quantities in terms of others. As long as we are given 3 values, we can derive the other 3. The usual approach is to determine the ratios between the quantities on either side, i.e. ρ_1/ρ_0, u_1/u_0 and P_1/P_0.

The process of determining these ratios involves only algebra and is "relatively easy". It involves a solving a quadratic equation which gives two distinct solution; however, only one of these solutions involves an increase in the pressure across the shock (in the other the pressure drops and this is known as a rarefaction shock – it also involves a decrease in the entropy and so violates the third law of thermodynamics, implying that such a shock does not occur in nature). The solutions can be written in terms of the Mach number and are:

$$\frac{\rho_1}{\rho_0} = \frac{u_0}{u_1} = \frac{4M_0^2}{M_0^2+3} \tag{71}$$

$$\frac{P_1}{P_0} = \frac{5M_0^2-1}{4} \tag{72}$$

These relations show that the ratios are all a function of just one parameter – the Mach number of the shock. This number must be larger than one ($M_0 = 1$ implies that the ratios are all unity so there is no change in the properties, and so there is no shock; $M_0 < 1$ is physically not allowed – see the discussion above).

Strong shocks

If M_0 is just slightly above 1, then the shock is *weak* and the density and pressure barely change, but when $M_0 \gg 1$, the density and velocities ratios go the limiting case

$$\frac{\rho_1}{\rho_0} = \frac{u_0}{u_1} = 4$$

while the pressure jump goes to $5M_0^2/4$ (the temperature jump becomes $T_1/T_0 = 5M_0^2/16$ in this limit). Because of this simple result it is often useful to simplify the discussion to such **strong shocks**. Many astrophysical shocks are strong.

However, throughout the above derivation we have worked in the frame in which the shock is stationary and the pre-shocked gas is flowing into the shock. While it is easy to derive the jump conditions in this frame, it is not the most natural frame for most situations. For example, in a supernova explosion the pre-shock gas is the ISM outside of the blast wave and so should be stationary. We can transform our results into this frame by performing a simple Galilean transformation (we are assuming the shock is non-relativistic), in which case we transform the velocities such that the pre-shock gas is motionless. This can be done by subtracting u_0 from all velocities, and the shock velocity becomes $V_s = -u_0$. We can define u_0' and u_1' as the pre- and post-shock velocities in the new frame. By definition $u_0' = 0$ and so $u_1' = u_1 - u_0 = u_0/4 - u_0 = -3u_0/4$ or $u_1\prime = 3V_s/4$. We will call this the "rest" frame (of the pre-shocked gas).

As gas flows through the shock, kinetic energy is converted to thermal energy, and it is interesting to investigate the just how much of the energy is converted. We can find this by computing the post-shock kinetic and thermal energies (per particle of mass μ in the rest frame).

$$\begin{aligned} e_K &= \frac{1}{2}\mu(u_1')^2 = \frac{9}{32}\mu V_s^2 \\ e_T &= \frac{3}{2}kT_1 = \frac{3}{2}\frac{5M_0^2}{16}kT_0 = \frac{9}{32}\mu V_s^2 \end{aligned} \tag{73}$$

The last step uses the definition for $M_0 = u_0/c_0$ and the fact that $c_0^2 = 5kT_0/3\mu$ for $\gamma = 5/3$. Therefore we see that (in this frame and for a strong shock) one-half of the post-shock energy is in the form of thermal energy and one-half in the form of kinetic energy.

Isothermal shocks

In a strong shock, the density increase is a maximum of four, but the temperature may increase a large amount. In some cases, the resulting temperature and density result in a very high rate of radiative cooling (we will address radiative cooling in more detail later in these notes). If the cooling is strong enough, the gas will quickly radiative its thermal energy and the temperature will return to the pre-shock temperature. This is known as an **isothermal** shock.

In this case, we can lump the shock and the cooling region into one single region and treat the whole thing like one big modified shock. The modification is largely in the energy condition, which is now simplified to be $T_2 = T_0$ (we denote the post-cooling region with a subscript '2'). We will also immediately assume the shock is strong so that we can neglect the pre-shock pressure in the momentum equation (this is not necessary but simplifies the algebra). In this case we have

$$\begin{aligned} \rho_0 u_0 &= \rho_2 u_2 \\ \rho_0 u_0^2 &= P_2 + \rho_2 u_2^2 \\ \frac{P_0}{\rho_0} &= \frac{P_2}{\rho_2} = c_0^2 \end{aligned}$$

The last equality comes from the isothermal relation $T_2 = T_0$. We can replace P_2 in terms of ρ_2 in the momentum equation, and the eliminate the ρ_2/ρ_0 ratios that result using the mass conservation relation, to obtain the following quadratic equation:

$$u_0 u_2 = c_0^2 + u_2^2$$

which has the solution

$$\begin{aligned} u_2 &= \frac{u_0}{2}\left(1 \pm \sqrt{1 - \frac{4c_0^2}{u_0^2}}\right) \\ &\approx \frac{u_0}{2}\left(1 - (1 - \frac{2c_0^2}{u_0^2})\right) = \frac{c_0^2}{u_0} \end{aligned}$$

In terms of the Mach number $M_0 = u_0/c_0$, we see that the density ratio is

$$\frac{\rho_2}{\rho_0} = \frac{u_0}{u_2} = M_0^2. \tag{74}$$

This should be compared to the strong shock, for which the maximum density increase is a factor of 4. In a strong isothermal shock the density increase can be very large. In fact, such shocks describe the late stages of supernovae remnants, and we can often see the shells of gas that such shocks sweep up.

5.6 Supernova Remnants and the ISM

When massive stars have exhausted all of their fuel, they may explode as supernovae. These Type II supernovae are thought to occur in stars with initial masses larger than about 8 solar masses. There are also other classes of supernovae, of which the next most important for the ISM is the Type Ia supernova, which results from a runaway thermonuclear burst when enough mass from a companion star is accreted onto a white dwarf. Type II supernovae do show hydrogen spectral lines during the outburst, while Type Ia do. The other difference is that Type II are associated with star formation and so often explode in dense regions, sometimes within the molecular cloud within which they were born. Type Ia SN typically occur hundreds of million years after the stars' births and so are no associated directly with star forming regions.

Despite the difference, the energy produced by both supernovae is similar – about 10^{51} erg of energy liberated over a very short time. Some of this energy goes into radiation but much of it goes into expelling heavy elements at high speed into the ISM.

We will model this injection of energy as three stages of evolution. The first phase is the **free-expansion** of the ejecta away from the star and lasts generally a few hundred years. During this time, the velocity of the ejecta is nearly constant and this phase ends when the ejecta has swept up a mass approximately equal to the amount of material ejected from the star. At this point, the remnant enters the **energy-conserving** (or Sedov) phase of the evolution.

As we noted earlier, for a strong shock, about 1/2 of the energy of the shock is converted into thermal energy and 1/2 into kinetic energy. Therefore, at any time, the total energy contained inside the shocked bubble is

$$E_T = \frac{4}{3}\pi R^3 \rho_0 (e_K + e_T) = \frac{3}{4}\pi n_0 \mu R^3 V_S^2$$

where in the second part we have used equation (73) for the thermal and energy content of the post-shock gas. Since the bubble is expanding with the shock velocity, we can write the shock velocity as $V_S = dR/dt$, and equate the total energy in the bubble to the supernovae energy E_* (assuming energy conservation, as the name of this phase implies), to get a differential equation for R:

$$R^3 \left(\frac{dR}{dt}\right)^2 = \frac{4E_*}{3\pi\rho_0}$$

The solution to this equation is a power-law of the form $R(t) = At^\alpha$. It is straightforward to show that it is given by:

$$R = \left(\frac{25}{3\pi}\right)^{1/5} \left(\frac{E_*}{\rho_0}\right)^{1/5} t^{2/5} \tag{75}$$

Since the radius is increasing as $t^{2/5}$, the shock velocity goes as $V_s \propto t^{-3/5}$, and is steadily slowing. Since the shock velocity is decreasing, we know that the Mach number of the shock is dropping and, from the jump conditions, the post-shock temperature is also decreasing. After a time of about 10^4 years, the shock has grown to a size of about 10 pc (for a supernovae exploding into a medium with $n = 1$ cm^{-3}). During this phase, the gas is hot enough that it emits primarily in the X-ray domain, and so supernovae are mostly easily seen as X-ray bubbles (although it should be pointed out that due to instabilities in the shock front, supernovae remnants are rarely round).

As the remnant expands, its temperature drops and the energy emitted in X-rays becomes more and more important, until we can no longer ignore this loss of energy, and our principle of energy conservation no longer applies. Fortunately, the cooling due to radiation soon becomes so important that the gas behind the shock loses all of its energy and forms a thin, dense, cold shell right behind the shock. At this point, the shock is no longer being driven by the pressure of the hold gas behind the shell but it is still expanding because of momentum conservation. This marks the beginning of the third phase, usually known as the **momentum conserving**, or snow-plow phase.

We are now guided by the principle of momentum conservation so that

$$\frac{4}{3}\pi\rho_0 R^3 V_s$$

is the momentum of the swept-up material and is a constant. This can be integrated to find the position of the shock as a function of time, and matched on to the end of the energy-conserving phase. If the beginning of this phase occurs at a time t_0 when the shock is at a radius R_0 and is moving at a velocity V_0, then the solution is

$$R = R_0 \left[1 + r\frac{V_0}{R_0}(t - t_0)\right]^{1/4}$$

and we can see that at late times $R \propto t^{1/4}$. The expansion is slower than the energy-conserving phase and generally occurs when the post-shock temperature is about a million degrees (and the shock is moving at about 250 km/s). For our standard conditions, this occurs about 40,000 years after the explosion at a radius of about 25 pc. At this point, the shell has swept up almost 1400 solar masses of material in the ISM.

The supernovae remnant will continue to expand until its motion becomes part of the general motion of the ISM (or it mergers with another SNR). Typically about 10% of the energy of the supernovae is available to heat the ISM (after account for the energy lost due to radiative cooling). This energy is available to heat the ISM. Although this does not seem like that much it represents the largest source of energy for the gas in the ISM, and is probably the driving force behind the turbulent motions seen in the ISM gas.

5.7 Radiative heating and cooling

Quite generally, there are three ways to transfer energy: (i) convection, (ii) conduction, and (iii) radiation. Convection involves the exchange of energy via the exchange of mass: hot and cold parcels of gas mix directly. In conduction, heat flows via particle collisions from the hot gas to the cold gas, but the gas itself does not move. Finally, radiation involves the transmission of photons.

We have remarked on a number of occasions that gas, when heated, can emit photons and so lose energy. Equivalently, of course, photons can be absorbed by the atoms in the gas and be heated. These processes are known as radiative cooling and heating and play a hugely important role in the ISM. We will treat them separately, first examining cooling and then heating.

5.7.1 Cooling

In general, cooling works by a simple three step process. Two atoms collide, which leads to the excitation in one of the atoms and eventually the decay of the excited state, leading to the production of a photon. In order for this process to be an efficient coolant, four things must occur: (i) collisions must be sufficiently frequent; (ii) the collision energy must be larger than the threshold energy for excitation; (iii) the collision must actually lead to an excitation, and finally (iv) the decay must happen quickly enough that another collision doesn't lead to de-excitation without a photon.

The collisional nature of this process means that the rate of cooling depends on the rate of collisions between two species. Therefore, quite generally, the cooling rate scales as the product of the densities of both species (this fails if the density is so high that collisional de-excitation is important), so that the cooling rate is often written in the form $n_i n_j \Lambda(T)$. In the following, we will consider a number of different physical cooling mechanisms.

Cooling by ions and atoms

In this case, the collision leads to an electronic excitation, generally from the ground state to the first excited state. There are a large number of such transitions, and here we just highlight a few of the more important ones in the ISM:

ion/atom	$\Delta E/k$
HI	$\sim 10^4$ K
SiII	413 K
OI	228 K
CII	92 K

This table also include the excitation energy divided by the Boltzman constant k to make it have units of degrees K. This is the energy required in order to excite the electron and generally the gas must have this temperature or larger in order for cooling to be effective. We see that there is a sequence of atoms which permit cooling to lower and lower temperatures (in fact, there are many more transitions than in this table). For example, hydrogen line cooling (the Lyman-alpha transition between the $n = 2$ and $n = 1$ states) is very effective down to temperature of about 10^4 K, but cannot cool the gas below that. As the table shows, other atomic states can pick up where hydrogen leaves off, although we expect the cooling rate to be lower, in large part because the other atoms and ions are much rarer than hydrogen. The last entry in the table, corresponding to once ionized carbon, is one of the most important coolants for the low-temperature ISM but even this does not cool the gas below about 100 K. To go to the lowest temperatures, we must use molecules.

The discussion above focuses on electronic transitions from one bound state to another (so-called "bound-bound" transitions), but there are also bound-free transitions (e.g. recombination) and free-free transitions. There are a wide range of possible mechanisms, but we focus here on two which are particularly common. The first is the close passage between an electron and a proton in which the electron is not captured but does suffer an acceleration. From electromagnetic theory we know that an accelerated charge will emit radiation and this particular form is known as **Bremsstrahlung**. It is particularly important for high temperature gas and is the dominant cooling mechanism for $T > 10^7$ K. If a magnetic field is present, an electron will spiral along field lines, again leading to photon emission in a process known as **synchrotron** emission. Synchrotron emission is a good indicator of the presence of both magnetic fields and high-energy electrons.

Cooling by molecules and dust

Molecules generally have weaker bonds between the atoms than for electrons within an atom and so the excited transitions are generally lower in energy. This means that molecules can cool the gas to lower temperatures. These excited states are generally rotational or vibrational in nature (where the bond acts like a spring between the atoms). Two important molecules in the ISM are H_2 and CO. These molecules form in dense, cold gas clumps and are commonly referred to as molecular clouds. While H_2 is the most common, it is hard to observe (because it has no dipole and so emits only via quadrupole transitions) and so most observations are of CO. The rotational energies of CO are quantized and have the form $E_J = BJ(J+1)$ where $J = 0, 1, 2\ldots$ is the quantized angular momentum. The transitions between these states are low enough energy that they fall in the radio part of the spectrum and are typically labeled by the transition (e.g. the $J = 2$ to 1 transition is quite common).

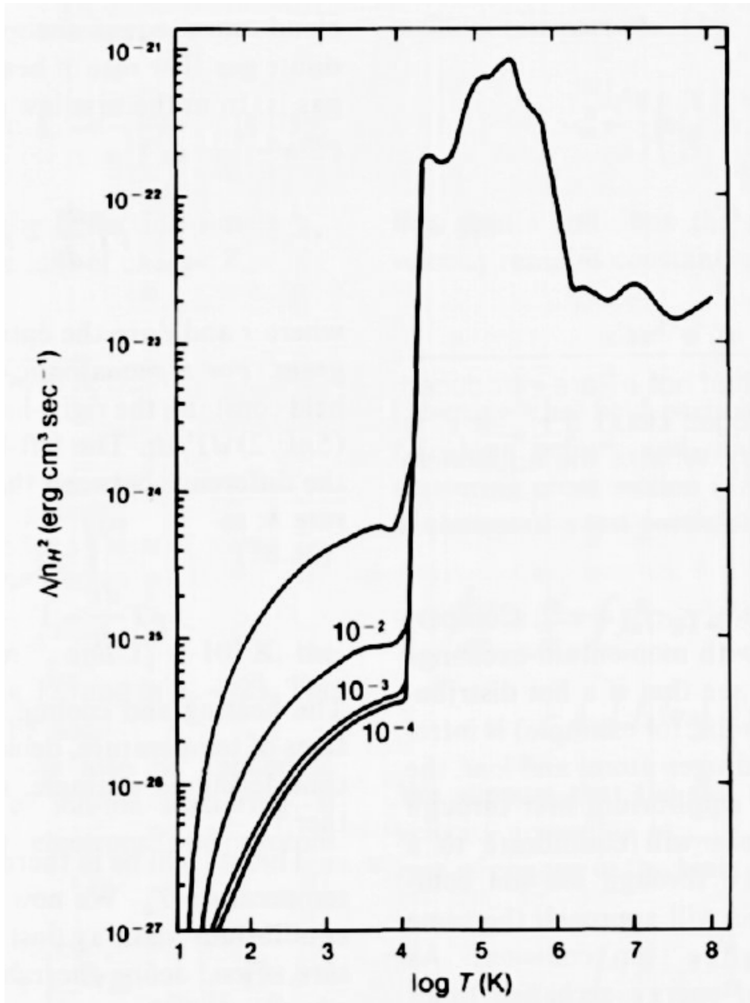
Dust is formed in the dense environments around stars (often in stellar outflows) and consists of micrometer to millimeter-sized agglomerations of atoms. We have discussed the importance of dust grains in blocking visible and UV light, which they do very effectively (considering that they are usually only a few percent of the gas content by mass). However dust grains also emit radiation in the infrared and although they are not thought to be an important coolant for the Milky Way's ISM, they do serve as a reasonable tracer for the distribution of gas in the ISM. In starburst galaxies, they can block large amounts of UV and optical radiation and in an extreme population of such objects nearly all of the radiation is emitted by dust (known as Ultra Luminous Infrared Galaxies, or ULIRG's for short).

The cooling curve: putting it all together

The cooling processes discussed above are, in general, very complicated and required detailed calculations based on the density, temperature and ionization state of the gas. However, it is possible to construct a simplified picture based on observation we made earlier, that generally the cooling rate depends on the multiplication of two densities, n_i and n_j. If we assume that $n_i = f_i(T)n$, then we can write the cooling rate (energy loss per unit time per unit volume) as $n^2\Lambda_{ij}(T)$ where $\Lambda_{ij}(T)$ depends only on the temperature. Since all the processes are like this, we can sum the terms together to get the overall cooling rate:

$$C = n^2\Lambda(T)$$

This **cooling curve**, as it is often called, encodes all of the information about the atomic physics and the ionization state of the atoms. It assumes that the current abundances can be determined uniquely based on the temperature, which is clearly incorrect in some cases, but is sufficiently accurate that it provides a useful guide. In the figure below (from Dalgarno & McCray 1972) we show this cooling curve (note that the notation in the figure differs slightly from ours – in our notation, what is plotted is Λ; also there are four different curves in which assumptions about the properties of the gas are varied – in particular the electron fraction at low temperatures).



The shape of the cooling curve in different temperature ranges is due the different cooling mechanisms which are important at that temperature. For example, the sharp peak just above 10^4 K comes from HI cooling, while the cooling below this depends on CII and other ions with lower excitation energies. At the very highest temperatures, the emission is from Bremsstrahlung. This figure is from Spitzer (1968).

5.7.2 Heating

Heating in the ISM is partially accomplished by shocks and other mechanisms, but radiative heating also plays a role, particularly at low temperature. We briefly summarize the primary heating mechanisms below.

Heating by **photo-ionization** is one of the most important heating mechanisms, particularly near massive stars. A photon with energy larger than that required to unbind an atom's electron (e.g., for hydrogen this is $I_H = 13.6$ eV) will deposit the extra energy it has (i.e. $E - I_H$) in the form of kinetic energy. This kinetic energy will be shared with the rest of the atoms in the gas, resulting in an increase in the temperature. Photonionization of hydrogen is very important, but heating can also come from the ionization of other atoms (indeed, molecules can by photo-disassociated as well). Generally, this sort of heating is most important inside HII regions where high energy photons are common.

The diffuse gas in the ISM is also heated through collisions with high energy photons (X-rays) and cosmic-rays. In both cases, the gas is photoionized and some of the extra energy of the photon is shared as heat. Finally, dust absorbs optical and UV photons, which can free an electron from the surface of the grain (the photoelectric effect). This electron will generally have some extra energy which will heat the gas. These two heating mechanisms (cosmic-ray heating and dust-grain heating) are thought to be the most important heating source in the ISM.

As before, the heating rate is proportional to the density of the two things that need to collide, but in this case one is an atom in the gas, while another is a photon or cosmic-ray. Therefore, the heating rate is proportional to a single power of the density: $H = \Gamma n$, where Γ depends on the flux of photons or cosmic-rays (and possibly other factors).

5.7.3 Thermal equilibrium

If radiative heating and cooling are the most important heating and cooling mechanisms, then we can determine the equilibrium temperature of the gas by equating the heating and cooling rates: $H = C$. Using our earlier expressions for these rates, we find:

$$\Gamma = \Lambda(T)n$$