ASTR 4020 / 8020: Diffuse Matter and Star Formation

Class 1 Notes: Collisional processes

The subject of this class is the low density gas between the stars and galaxies. Our goal will be to develop a physical understanding of how this material works. The physics that governs it is all familiar and has been understood for at least the last fifty years – gas dynamics, radiation, some quantum mechanics – but the low densities found in interstellar space provide a completely novel context far removed from any other physical regime we're used to thinking about. For this reason, many of the familiar behaviours we expect are absent or altered in the interstellar context. This makes the ISM a wonderfully complex and challenging problem.

The class is roughly divided into three parts. The first third covers the physics of the ISM, developing theories for the important processes that occur in low density gas. The second third applies those models to understand the behaviour of the major constituents of the ISM and IGM. What we will cover does not even come close to being exhaustive. The ISM is still a young and rapidly-developing field of study. The last third of the course will focus on how the self-gravitating portion of the ISM, and the process by which it transforms into new stars.

Finally, a note on nomenclature: because saying "ISM and IGM" repeatedly would be tiresome, I am simply going to say "ISM" most of the time, with the understanding that this encompasses the IGM as well. I will try to be clear when I am saying something that applies only to the ISM or the IGM.

I. Components of the ISM

As we will learn later on in the course, the ISM is made up of a range of components: solid particles, several distinct phases of gas, and radiation fields. Although we will be starting with ISM physics and only later getting into applications, it is helpful to have in mind a rough phenomenology to guide us in picking characteristic numbers and scales, and in suggesting what problems are interesting.

A. Dust

A significant fraction of the refractory elements, i.e., those whose solid forms vaporise at temperature ~ 1000 K rather than ~ 100 K, exist in interstellar space in the form of small dust grains. Much of the carbon, silicon, iron, and similar elements that exist in the ISM are in grain form. The exact fraction in grain rather than gas form varies with the ambient density and temperature, ranging from a majority of material being in grain form in dense and cold environments to almost none of it in hot environments. However even in extremely hot gas at temperatures of $\sim 10^6 - 10^7$ K some dust grains survive, mainly because the grains are not necessarily heated to anything close to the gas temperature – a point we will return to later in the course.

The size distribution of grain is somewhat uncertain, but it can be constrained by observations of the grains' absorption, scattering, and emission properties, since these are closely correlated with size. Roughly speaking the cross-section of a grain to light reaches a maximum for light whose wavelength is comparable to the grain size. For light of shorter wavelengths grain cross-sections are simply their physical sizes, while for light of longer wavelengths the cross-section declines roughly as the square of the ratio of grain size to wavelength. We will also return to this discussion later on. Based on these observations, the typical size of interstellar dust grains must be less than ~ 1 μm in size.

B. Cosmic rays

The gas in interstellar space mostly has a Maxwellian velocity distribution. However, there are also ions and electrons that have much larger, typically relativistic velocities. These are referred to as cosmic rays. The most energetic ones detected have energies of $\sim 10^{21}$ eV, but most of the energy is rests in protons with kinetic energies of ~ 1 GeV. The total energy density in these particles in the Milky Way is ~ 1 eV cm^{-3}.

The mechanism by which cosmic rays are accelerated to such speeds is still not completely understood, but it is thought to involve high-speed shocks produced by supernovae, massive star winds, or similar phenomena. Regardless of the acceleration mechanism, though, once launched cosmic rays propagate long distances through the ISM. We will not discuss CRs much in this class, because they are covered extensively in other courses, except when they become important for understanding the way the ISM and IGM behave.

C. Photons

The ISM is also pervaded by photons of various frequencies. In addition to the ubiquitous CMB, there is mostly optical and UV light from stars, infrared radiation from dust grains, radio emission from hot gas, synchrotron radiation from relativistic gas, and numerous line photons at frequencies ranging from radio to gamma rays produced by a huge number of molecules, atoms, ions, and nuclei.

In terms of energy, in the Milky Way the dominant component of interstellar energy is in starlight at near-IR to near-UV wavelengths. The energy density of this light is of order an eV cm^{-3}, similar to the cosmic rays. Infrared radiation, produced when starlight photons interact with dust, is a close second.

D. Magnetic fields

As we shall see later on, much of the ISM is occupied by ionised gas, and even in regions where the gas is mostly neutral there is usually some weak residual ionisation produced by a number of processes we will discuss later. This means that the ISM can generally be thought of as an ideal plasma. Plasmas have the property that they can sustain magnetic fields, and any moving plasma invariably generates fields. The ISM is not exception.

Evidence for the magnetic field arises from numerous sources. First, we observe that light from interstellar dust is polarised – both the light it emits, and the light it absorbs from background stars. The most natural explanation for this is that dust grains are not perfectly spherical, and in the presence of a magnetic field they will line up in preferred orientations relative to the field like little bar magnets, which will produce polarisation.

Second, there are a number of atomic and molecular emission lines that are magnetically sensitive, meaning that the line shape, polarisation, or some other property of the line changes in the presence of a magnetic field. Such lines do generally indicate that a field is present.

The strength of the interstellar magnetic field is $\sim 5 - 10$ μG. This makes the magnetic energy density also about an eV cm^{-3}, comparable to photons and cosmic rays.

E. Gas

The bulk of the ISM by mass consists of hydrogen and helium gas. For reasons we will come to understand through this course, the gas naturally segregates itself into a series of distinct phases.

1. Hot gas (HIM)

   The hottest component, and the dominant one outside the discs of galaxies and into the IGM, is hot gas, or HIM. This gas is also found within the galactic disk in places where the gas has been shocked by supernova blast waves, and it may occupy several tens of percent of the volume of the galactic disk.

   Typical temperatures in this gas are 10^6 K or more. As these temperatures hydrogen is collisionally ionised, and numerous highly-ionised species of heavy elements are also present, for example O VI. The high temperatures ensure that the gas is able to expand easily until it reaches low densities, typically $\sim 10^{-3}$ cm^{-3} within the disc of the Galaxy and even lower outside it.

   At these temperatures the gas emits mostly in X-rays and UV, and by radio synchrotron emission from free electrons. X-ray emission also provides the main channel by which this gas is able to cool, although the cooling times can be extremely long due to the low densities. For much of the IGM, the cooling time is longer than the Hubble time.

2. Warm ionised gas (WIM) / H II

   The next hottest phase, called the warm ionised medium or H II regions, is gas at temperatures of $\sim 10^4$ K. At this temperature $k_B T \sim 1$ eV, so the gas is not moving gas enough for the typical collision to induce ionisation (since the hydrogen ionisation potential is 13.6 eV). Instead, WIM gas is found near hot stars that provide high energy photons to photoionise the hydrogen. Regions in this state occupy $\sim 10\%$ of the volume of the Galactic disc, and their typical densities are $1 - 100$ cm^{-3}, although there is a huge range of variation.

The photons that ionise the gas also provide the major source of energy to it. Each time an ionisation occurs by a photon that has an energy a bit above 13.6 eV, the resulting free electron acquires some excess kinetic energy, which it then thermalises by bouncing off the surrounding ions.

Countering this heating, gas at these temperatures cools by numerous processes. These include recombination radiation, which is produced when ionised hydrogen recombines into an excited state and then radiatively decays to the ground state, free-free emission from free electrons, and also a great deal of line emission produced by collisions between free electrons and partially ionised metal atoms. The line radiation produces spectacular optical emission, and for this reason H II regions are some of the most visually spectacular objects around.

In addition to the visually spectacular H II regions, which represent the densest parts of the WIM, there are also numerous diffuse, low-density regions of ionised gas. These are not necessarily associated with young hot stars, and instead represent places where the gas was ionised once, and the density is low enough that only a little radiation is needed to keep it ionised, or where the recombination has not yet had time to occur.

3. Warm neutral gas (WNM)

   In the absence of a local heat source like a hot star or a supernova blast, interstellar gas tends to become neutral, which brings us to our next phase: WNM, or warm neutral medium. The gas in this phase is generally at temperatures of $5,000 - 10,000$ K, and has a density not very different from that of the H II regions, $\sim 0.1 - 1$ cm^{-3}. Gas in this state occupies a large fraction of the volume of the Galactic disc, $\sim 40\%$.

   In this gas there are no photons above 13.6 eV to provide energy, but photons at somewhat lower energies provide a similar heating mechanism. Although such lower energy photons cannot ionise hydrogen, then can knock electrons off dust grains via the photoelectric effect, and this proves to be the dominant heating source.

   Countering this heating, the WNM also contains numerous weakly ionised or neutral metal atoms that can be collisionally excited much like those in H II regions. Since there are few free electrons to collide with, and the metal atoms are more weakly ionised and thus tend to have lower energy scales, most of this line emission is in the infrared rather than the visible. The 158 μm line observed by ALMA is an example of this sort of emission.

   Although these IR emission lines can be used to study the WNM, as can optical and UV absorption lines, by far the most common tool is the 21 cm hyperfine transition of the hydrogen itself. This is present everywhere, and has numerous favourable features that we will discuss.

4. Cold neutral gas (CNM)

   Neutral gas can be warm, but it can also be cold. The cold neutral medium, or CNM, is similar in ionisation state and energy balance to the WNM, but it is found at much lower temperatures, ~ 100 K, and much higher densities, ~ 10 cm^{-3}. Because of its high density, it has a much lower volume filling fraction, $\sim 1\%$ of the Galactic disc. Nonetheless, it contains an amount of mass not much smaller than the mass of the WNM.

5. Diffuse molecular gas

   In the densest, coldest parts of the CNM, hydrogen can become molecular. Like the transition from H I to H II this is a process driven by photons. Gas goes from neutral to ionised when there are hot stars around to provide photons above 13.6 eV. It goes from atomic to molecular regions where there is enough absorption to *exclude* photons with energies of above ~ 10 eV. Only the densest and coldest regions of CNM do this, so H_2 is found only in gas at densities ~ 100 cm^{-3} and at temperatures ~ 50 K. The volume occupied by these regions is tiny, $\sim 0.1\%$ of the disc.

   Despite the transition from atomic to molecular, in the diffuse H_2 clouds the energetics are quite similar to those in CNM or WNM. Molecular hydrogen does not provide much of a source of heating or cooling, for reasons we will discuss. For the same reason, observing these regions is hard – the 21 cm H I line is unavailable, and H_2 is hard to see. The main way we know about this gas is via UV absorption lines of H_2, which are available only when there is a conveniently-located background star.

6. Dense molecular gas / GMCs

   The final, densest phase is the dense molecular gas. In this part of the ISM the temperature falls to ~ 10 K or even a little lower, and the density is at least 100 cm^{-3}, and often more. Most of the gas in this state is organised in structures known as giant molecular clouds. These clouds occupy only $\sim 10^{-4}$ of the volume of the Galactic disc, but constitute $\sim 20-30\%$ of its total gas mass. In some other galaxies that fraction is even higher. These clouds are also where star formation occurs.

   The change that occurs between diffuse and dense molecular gas that causes this change in properties is the appearance of molecules. Whereas in the diffuse molecular gas and all the less dense phases most of the carbon, oxygen, and other species are either atomic or in dust grains, in the dense molecular gas significant fractions of these atoms end up in diatomic and polyatomic molecules, the most prominent of which is CO.

   The appearance of these molecules is significant because, unlike H_2, these molecules are strong emitters, capable of cooling the gas to temperatures of ~ 10 K. These molecular transitions also provide the most common way of studying the dense molecular gas, particularly the rotational levels of CO. Other methods include mm emission by cold dust grains, and infrared absorption of background starlight.

II. Collisional processes

With that brief tour of the ISM completed, we will begin to build the theoretical tools we will need to understand the behaviour of the ISM. The first goal will be an understanding of collisions between particles. In the materials we're used to, collisions occur so frequently that we don't usually worry about them, and we simply assume that they occur frequently enough for distributions to reach thermodynamic equilibrium. In the ISM we cannot safely make this assumption, and so we must worry about calculating the rates of collisions.

A. The collision rate coefficient

Consider an interaction between two particles A and B, which we write in the general form

$$A + B \rightarrow \text{products} \tag{1}$$

Depending on the type of interaction, the products that appear on the right hand side can be many different things. Simple elastic scattering is a trivial case, for which products is simply $A + B$ again. For inelastic scattering, where either A or B is left in an excited state after the encounter, it might be $A^* + B$ or $A + B^*$, with the asterisk indicating an excited state. If a chemical reaction occurs, it might be an entirely different species C. For now, the exact identity of the right hand side does not matter.

We wish to compute the rate at which the given reaction / collision occurs in a gas that contains species A and B with number densities n_A and n_B, respectively. The dependence of the densities is fairly obvious. Suppose we imagine a beam of particles of type A being fired at a static grid of particles of type B. Clearly the rate of collisions will be linearly proportional to both the density of targets and the density of particles in the beam. Thus we expect to have a rate per unit volume that varies as $n_A n_B$. We rate the rate as

$$\text{rate per unit volume} = n_A n_B k_{AB}, \tag{2}$$

where k_{AB} is the rate coefficient for the reaction. It has units of cm^3 s^{-1}.

By analogy one can also define three-body collision rate coefficients, for reactions involving three species such that the rate per unit volume is $n_A n_B n_C k_{ABC}$. In practice the low density of the ISM implies that three-body processes are only very rarely important.

B. Calculation of rate coefficients

To figure out k_{AB}, we can roughly divide it into two parts: the "internal" part that has to do with the physical properties of the colliding particles and the quantum mechanical probabilities of a given interaction producing a given outcome, and the "external" part that has to do with the kinematics of particles running into one another.

To see how the external part works, we can first return to the easier-to-picture case of a beam of particles directed at a static grid of targets. Clearly if the beam moves faster, more particles per unit time will go through the array of targets, so the collision rate will be linearly proportional to the relative velocity of the targets and the beam. In the more realistic case of two interacting species mixed together, we need to integrate over all possible reaction velocities. Thus we can write

$$k_{AB} = \int_0^\infty v f_v \sigma_{AB}(v)\, dv = \langle \sigma v \rangle_{AB}, \tag{3}$$

where v is the relative velocity, f_v is the fraction of particle pairs that have that relative velocity, and $\sigma_{AB}(v)$ is the velocity-dependent interaction cross-section. This encapsulates all the internal information about particle sizes and quantum transition probabilities. The angle brackets indicate an average over collision velocities.

We can compute f_v from the Boltzmann distribution, under the assumption (which we'll check later) that the particles follow this distribution. To remind you: the Boltzmann distribution says that in a system with temperature T, the probability of finding a particle in a state with energy E is proportional to e^{-E/k_BT}. Thus the probability of having a given vector velocity $\mathbf{v} = (v_x, v_y, v_z)$ is proportional to $e^{-mv^2/(2k_BT)}$, where m is the particle mass and $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$, and we have

$$\frac{d^3 f}{dv_x\, dv_y\, dv_z} = \left(\frac{m}{2\pi k_B T}\right)^{3/2} e^{-mv^2/(2k_BT)}. \tag{4}$$

The normalisation constant in front has been chosen to ensure that the integral of $d^3 f/(dv_x\, dv_y\, dv_z)$ over all possible velocities is unity.

This is the velocity distribution for individual particles. In other words, it applies separately to A and B, and we have

$$\frac{d^3 f_A}{dv_{x,A}\, dv_{y,A}\, dv_{z,A}} = \left(\frac{m_A}{2\pi k_B T}\right)^{3/2} e^{-m_A v_A^2/(2k_BT)} \tag{5}$$

and similarly for B. We want to know what the probability that a randomly chosen pair of particles has relative velocity v. To compute this, first note that the probability of a given velocity combination $\mathbf{v}_A, \mathbf{v}_B$ is the just the product of the individual probabilities, the same as for any independent pair of events. Thus

$$\frac{d^6 f(\mathbf{v}_A, \mathbf{v}_B)}{d\mathbf{v}_A\, d\mathbf{v}_B} \propto \left(\frac{\sqrt{m_A m_B}}{2\pi k_B T}\right)^3 e^{-(m_A v_A^2 + m_B v_B^2)/(2k_BT)}. \tag{6}$$

This is a six-dimensional probability distribution function that gives us the probability of picking a given sextuplet of values $(\mathbf{v}_A, \mathbf{v}_B)$. What we want to know is the probability of picking a sextuplet that has the particular property that $|\mathbf{v}_A - \mathbf{v}_B| = v$, since that is the definition of f_v. Thus we want to compute

$$\int e^{-(m_A v_A^2 + m_B v_B^2)/(2k_B T)} \delta(|\mathbf{v}_A - \mathbf{v}_B| - v)\, d^3\mathbf{v}_A\, d^3\mathbf{v}_B. \tag{7}$$

We've dropped the leading constants because we're only interested in the dependence on velocities – other coefficients we can recompute at the end just by requiring that our integrated probability be unity.

This integral can be evaluated by making a change of variables. Let $\mathbf{v} = \mathbf{v}_A - \mathbf{v}_B$, and $\mathbf{v}_{\text{CM}} = (m_A\mathbf{v}_A + m_B\mathbf{v}_B)/(m_A + m_B)$, or equivalently

$$\mathbf{v}_A = \frac{\mu}{m_A}\mathbf{v} + \mathbf{v}_{\text{CM}} \qquad \mathbf{v}_B = -\frac{\mu}{m_B}\mathbf{v} + \mathbf{v}_{\text{CM}} \qquad \mu = \frac{m_A m_B}{m_A + m_B} \tag{8}$$

First hold $\mathbf{v}$ and $\mathbf{v}_A$ fixed and substitute for $\mathbf{v}_B$ in terms of $\mathbf{v}_{\text{CM}}$; in this case $d^3\mathbf{v}_B = -(\mu/m_B)^3 d^3\mathbf{v}_{\text{CM}}$, and the integral becomes

$$\int \exp\left[-\frac{m_A v_A^2 + m_B|(\mu/m_B)\mathbf{v} - \mathbf{v}_{\text{CM}}|^2}{2k_B T}\right] \delta\left(\left|\mathbf{v}_A + \frac{\mu}{m_B}\mathbf{v} - \mathbf{v}_{\text{CM}}\right| - v\right) d^3\mathbf{v}_A\, d^3\mathbf{v}_{\text{CM}}, \tag{9}$$

where we have again dropped leading constants that don't depend on velocity. Now hold $\mathbf{v}_{\text{CM}}$ fixed and substitute for $\mathbf{v}_A$ using $\mathbf{v}$. Thus $d^3\mathbf{v}_A = (\mu/m_A)^3 d^3\mathbf{v}$, and the integral becomes

$$\int \exp\left[-\frac{m_A|(\mu/m_A)\mathbf{v} + \mathbf{v}_{\text{CM}}|^2 + m_B|(\mu/m_B)\mathbf{v} - \mathbf{v}_{\text{CM}}|^2}{2k_B T}\right] \delta\left(|\mathbf{v}| - v\right)\, d^3\mathbf{v}\, d^3\mathbf{v}_{\text{CM}}. \tag{10}$$

Now it's just a matter of algebra to evaluate the integral. The term in the exponential can be expanded to

$$m_A\left|\frac{\mu}{m_A}\mathbf{v} + \mathbf{v}_{\text{CM}}\right|^2 + m_B\left|\frac{\mu}{m_B}\mathbf{v} - \mathbf{v}_{\text{CM}}\right|^2 = \frac{\mu^2 v^2}{m_A} + \frac{\mu^2 v^2}{m_B} + (m_A + m_B)v_{\text{CM}}^2 \tag{11}$$

$$= \mu v^2 + (m_A + m_B)v_{\text{CM}}^2, \tag{12}$$

and if we substitute this in we get

$$\int \exp\left[-\frac{\mu v^2}{2k_B T}\right] \exp\left[-\frac{(m_A + m_B)v_{\text{CM}}^2}{2k_B T}\right] \delta\left(|\mathbf{v}| - v\right)\, d^3\mathbf{v}\, d^3\mathbf{v}_{\text{CM}} \tag{13}$$

The part that depends on v and not v_{CM} is now trivial to evaluate thanks to the δ function, which just gives us a v^2 dependence:

$$v^2 \exp\left(-\frac{\mu v^2}{2k_B T}\right) \int \exp\left[-\frac{(m_A + m_B)v_{\text{CM}}^2}{2k_B T}\right] d^3\mathbf{v}_{\text{CM}} \tag{14}$$

The remaining integral is just a number that does not depend on velocity, and so we've arrived at the fundamental dependence we were after: $f_v \propto v^2 e^{-\mu v^2/2k_BT}$. Inserting the appropriate normalization factor to ensure that the integral over all velocities gives unity, we have

$$f_v = 4\pi \left(\frac{\mu}{2\pi k_B T}\right)^{3/2} v^2 e^{-\mu v^2/2k_BT}. \tag{15}$$

The two-body collision rate coefficient therefore is

$$k_{AB} = 4\pi \left(\frac{\mu}{2\pi k_B T}\right)^{3/2} \int_0^\infty v^3 e^{-\mu v^2/2k_BT} \sigma_{AB}(v)\, dv. \tag{16}$$

Alternately, it is sometimes helpful to write things in terms of the energy of the collision in the center of mass frame instead of the relative velocity. The distribution in energy is just given by the fundamental rule for the transformation of probabilities: $f_v\, dv = f_E\, dE$, i.e. since E is a monotonic function of v, the probability of measuring a velocity between v and $v + dv$ must be the same as the probability of measuring an energy between E and $E + dE$. Since $E = \mu v^2/2$ and $v = \sqrt{2E/\mu}$, we have $dE = \mu v\, dv$, and plugging in we get

$$k_{AB} = \sqrt{\frac{8k_BT}{\pi\mu}} \int_0^\infty x e^{-x} \sigma_{AB}(x k_B T)\, dx, \tag{17}$$

where $x = E/k_BT$. This is often the most practical form for computation.

The function we have just written down already carries an important point. Suppose we have a cross section that is not highly velocity- or energy-dependent. In this case σ_{AB} is a constant and comes out of the integral. The remaining part, the integral of xe^{-x} from 0 to ∞, trivially evaluates to 1, and we have

$$k_{AB} = \sqrt{\frac{8k_BT}{\pi\mu}} \sigma_{AB}. \tag{18}$$

This gives us a simple formula to evaluate any reaction coefficient with a constant cross section, and shows us that such reactions proceed at a rate that varies as $T^{1/2}$. In practice it turns out that there are reasonably large number of collisional processes where the cross section is indeed not very energy-dependent, so in practice many rate coefficients do vary as close to $T^{1/2}$.

C. Cross sections and rate coefficients for varying reactant types

Now that we have a general framework, we are in a position to work out reaction rate coefficients for a variety of interaction types.

1. Neutral-neutral scattering

The simplest case to consider is scattering of one neutral species off another. For now we will not worry if the interaction is elastic or not, and we will simply compute the overall collision rate; the elastic and inelastic interaction rates will each be a fraction of this total rate. This collision rate is particularly important because interactions of this sort are responsible for establishing a Boltzmann distribution of velocities among a population of neutral particles, which is what we assumed existed for the purposes of computing rate coefficients.

At large distances the only force between two neutral particles is a van der Waals attraction, produced when fluctuations in the electric dipole moment of one particle induce a corresponding electric dipole in the other. Since the dipole electric field of the first particle varies as $1/r^3$, so does the strength of the dipole in the second particle. The potential then varies as the product of the two dipoles, giving rise to an overall potential that varies as $1/r^6$. Moreover, since the dipole is only due to fluctuations, the coefficient is quite small. When the two particles get within ~ 1 Å of one another, i.e. when the separation is comparable to the total sizes of the interacting molecules or atoms, the electron clouds of the two neutrals begin to repel one another, and the force becomes very strongly repulsive.

This combination of very weak attraction at large radii, coupled with a sudden transition to strong repulsion at small separations, can be modeled reasonably well as a "hard sphere" interaction. We simply think of the two neutrals as balls with a physical size of $r_A = r_B = 1$ Å; if they get closer than 2 Å they collide, and otherwise they do not. Thus

$$\sigma_{AB} = \pi(r_A + r_B)^2 = 1.2 \times 10^{-15}\ \mathrm{cm}^2. \tag{19}$$

Plugging this into our trivial expression for the rate coefficient when the cross section is constant, we get

$$k_{AB} = 1.81 \times 10^{-10} \left(\frac{T}{100\,\mathrm{K}}\right)^{1/2} \left(\frac{m_\mathrm{H}}{\mu}\right)^{1/2} \left(\frac{r_A + r_B}{2\ \text{Å}}\right)^2\ \mathrm{cm}^3\ \mathrm{s}^{-1}. \tag{20}$$

We can think of this as a generic, order-of-magnitude estimate for the collision rate coefficient for any process where the particles are neutral and there are no chemical reactions involved, just scattering.

2. Charged-neutral scattering

Now let us consider the interaction of a neutral particle with a charged one. This could be a neutral atom or molecule interacting either with an ion or with a free electron. The main difference here is that the force when the two particles are far apart is no longer completely negligible. Let us suppose that the charged particle has a charge Ze, where e is the electron charge. The electric field of the charged particle will polarize the neutral particle, inducing a dipole moment $\mathbf{P} = \alpha \mathbf{E}_\mathrm{ch}$, where α is the polarizability of the neutral particle and $\mathbf{E}_\mathrm{ch}$ is the electric field created by the charged particle. Typical polarizabilities are of order a_0^3, where $a_0 = \hbar^2/m_e e^2 = 5.29 \times 10^{-9}$

cm is the Bohr radius. These can be computed quantum mechanically or measured in the lab fairly easily.

The attractive force that the polarized atom experiences is given by the standard formula for the force on a dipole in an electric field: $F = P(dE_{\rm ch}/dr) = -2\alpha Z^2 e^2/r^5$. The corresponding interaction potential is

$$U(r) = -\frac{1}{2}\frac{\alpha Z^2 e^2}{r^4}. \tag{21}$$

Scattering in an r^{-4} potential has the property that there is a critical impact parameter b_0 (defined as the distance of closest approach that the two particles would have if there were no force between them) below which the separation between the two particles goes through zero exactly. (Proving this is left as an exercise to the reader – it's fairly easy to show simply by writing down the Largangian for the system.) Deflections are relatively weak for larger impact parameters.

The value of b_0 depends on the relative energy of the two particles at infinity in the center of mass frame, E. It is given by

$$b_0 = \left(\frac{2\alpha Z^2 e^2}{E}\right)^{1/4} = 6.62 \times 10^{-8} Z^{1/2} \left(\frac{\alpha}{\alpha_{\rm H}}\right)^{1/4} \left(\frac{0.01 \text{ eV}}{E}\right)^{1/4} \text{ cm}, \tag{22}$$

where $\alpha_{\rm H} = 4.5 a_0^3$ is the polarizability of neutral hydrogen. Thus we see that in general b_0 is significantly larger than the $\sim 10^{-8}$ cm geometric cross section of the particles. This means that πb_0^2 provides a natural estimate for the collision cross section of an ion and a neutral, since any interaction in which the initial impact parameter is below b_0 will necessarily bring the ion and neutral extremely close, while more distant interactions will not produce significant interaction. Plugging this in, we have

$$\sigma_{AB} = \pi b_0^2 = \pi Z e \sqrt{\frac{2\alpha}{E}} \tag{23}$$

Plugging this into our expression for the rate coefficient, we have

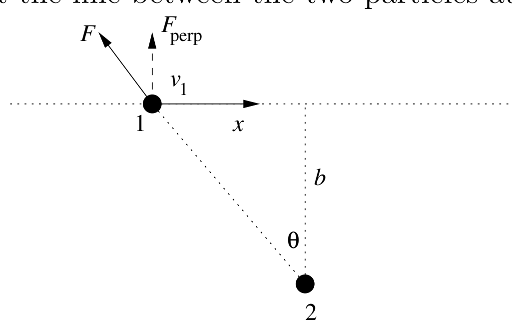
$$\begin{aligned}
k_{AB} &= \sqrt{\frac{8 k_B T}{\pi \mu}} \int_0^\infty x e^{-x} \left(\pi Z e \sqrt{\frac{2\alpha}{x k_B T}}\right) dx && (24)\\
&= 4 Z e \sqrt{\frac{\pi \alpha}{\mu}} \int_0^\infty x^{1/2} e^{-x}\, dx && (25)\\
&= 2\pi Z e \sqrt{\frac{\alpha}{\mu}} && (26)\\
&= 8.98 \times 10^{-10} Z \left(\frac{\alpha}{a_0^3}\right)^{1/2} \left(\frac{m_{\rm H}}{\mu}\right)^{1/2} \text{ cm}^3 \text{ s}^{-1}. && (27)
\end{aligned}$$

Note that the result is independent of temperature. This means that, even at low temperatures where neutral-neutral collisions are very rare (due to the $T^{1/2}$ dependence), ion-neutral collisions continue to remain common. This makes ion-neutral reactions a critical driver of chemistry in low-temperature gas.

3. Charged-charged collisions

   Collisions between two charged particles are a bit more complicated, because there we have a long-range force, and even at large distances there will be some non-negligible transfer of momentum. A useful approximation in this case is the impact approximation. The basic idea of the impact approximation is simple: we neglect changes in particle velocities during the encounter, and simply add up the change in transverse momentum that a projectile particle experiences as a result of the forces exerted by the electric field of the target. (The net change in momentum along the direction of the encounter is zero.)

   The setup is simple: consider two particles of charges $Z_1 e$ and $Z_2 e$, and work in the reference frame where particle 2 is at rest. Particle 1 approaches it, moving with velocity v_1 and with impact parameter b. Following the impact approximation, it moves in a straight line at constant velocity v_1 independent of how close it gets to particle 2. Let x be the distance between particle 1 and the point of closest approach, and θ be the angle between the line of closest approach at the line between the two particles at any given time.



   At a time when the angle between the two particles is θ, the distance between them is $b/\cos\theta$, so the total force is $F = Z_1 Z_2 e^2/(b/\cos\theta)^2$. The component of this in the perpendicular direction is $\cos\theta$ of the total, so the total perpendicular force is

$$F_\perp = \frac{Z_1 Z_2 e^2}{b^2}\cos^3\theta. \tag{28}$$

   To figure out the total momentum imparted over the entire encounter, we must find out how much time the particle spends at each angle θ, since $dp/dt = F$. This is easy to compute: clearly $x = b\tan\theta$, and

$$v_1 = \frac{dx}{dt} = b\frac{d}{dt}\tan\theta = \frac{b}{\cos^2\theta}\frac{d\theta}{dt} \qquad \Rightarrow \qquad d\theta = \frac{v_1}{b}\cos^2\theta\, dt. \tag{29}$$

Now it's just a matter of integrating over time to get the total momentum change:

$$\Delta p_\perp = \int_{-\infty}^{\infty} F_\perp \, dt = \frac{Z_1 Z_2 e^2}{b^2} \int_{-\infty}^{\infty} \cos^3\theta \, dt = \frac{Z_1 Z_2 e^2}{b v_1} \int_{-\pi/2}^{\pi/2} \cos\theta d\theta = 2\frac{Z_1 Z_2 e^2}{b v_1}. \tag{30}$$

So what does this tell us about the cross section? It tells us that we need to think carefully about what exactly we mean. Notice that the transverse momentum change varies as $1/b$, so more distant encounters have less effect. That's good. However, the total area available for encounters goes as $2\pi b \, db$, i.e. there's also more area available at large b than small b. The product of area times effect, $2\pi b \, db \cdot \Delta p_\perp$, does not depend on b, and if we integrate over all impact parameters from zero to infinity, things diverge! This means that the encounter cross section as we've been thinking about it in the neutral-neutral and neutral-charged cases is not really a meaningful concept as applied to the charged-charged case. We need to be a bit more specific about what sort of encounter we're interested in.

4. Electron-ion collisions and collision strengths

One particular case of a charged-charged collision where we can be more specific, which is one of the most important in the ISM, is a collision between an ion or atom and an electron that induces a change in the quantum state of the ion, either an excitation or de-excitation. (We will see in a moment that if you know the rate of one, you automatically know the other as well.) Such collisionally-induced changes in state, followed by radiative decay of excited states, are responsible for most of the visually-spectacular emission we need from ionised nebulae. We specialise to the case of an electron because electrons generally move much faster than ions, so most collisions are electron-ion rather than ion-ion.

Consider an encounter between an ion of charge Z and an electron. The unperturbed ion has a potential $U(r)$ in which the electrons move. We'll do the case of collisional de-exciation, so let the ion be in some excited eigenstate u of the potential $U(r)$, with energy E_u. We want to know the rate at which collisions cause it to transition to a lower energy state ℓ with energy E_ℓ.

We can calculate this rate up to a factor of order unity using a semi-classical approach. Here we will not use the impact approximation, and we will include deflection of the electron by the ion potential. Suppose the approaching electron moves classically in the potential provided by the ion. What is its closest approach? If the electron approaches with initial velocity v and impact parameter b, we show in the practice problems that its distance of closest approach r_{min} obeys

$$b = r_{\text{min}} \left(1 + \frac{2Ze^2}{m_e v^2 r_{\text{min}}}\right)^{1/2}. \tag{31}$$

How close does the electron have to get to have a significant chance of inducing a state change? At the order of magnitude level, the answer is that the perturbation in the potential δU must be comparable to or larger than the difference in energy between the two levels $E_{u\ell} = E_u - E_\ell$. Thus we want the distance of closest approach to obey

$$\frac{e^2}{r_{\rm min}} \sim E_{u\ell}. \tag{32}$$

It is convenient to normalise the energy difference to typical energy differences for electronic states. The typical energy scale for two electronic states is of order the potential of an electron at a distance of order a Bohr radius, i.e. $E_{u\ell} \sim e^2/a_0$ for a typical pair of electronic states. We therefore adopt a minimum distance

$$r_{\rm min} = W a_0, \tag{33}$$

where W is a constant whose value will depend on the exact transition. For electronic transitions that produce optical lines, we expect it to be of order unity.

Plugging this in for $r_{\rm min}$, we obtain a maximum impact parameter required to have a reasonable chance of inducing a change in state:

$$b_{\rm max} \approx W a_0 \left(1 + \frac{2Ze^2}{m_e v^2 W a_0}\right)^{1/2}. \tag{34}$$

The corresponding cross section is

$$\sigma_{u\ell} = \pi b_{\rm max}^2 = W^2 \pi^2 a_0^2 \left(1 + \frac{2Ze^2}{m_e v^2 W_0 a_0}\right). \tag{35}$$

Now we just have to integrate over the Maxwellian velocity distribution to get the rate coefficient:

$$\begin{aligned} k_{u\ell} &= 4\pi \left(\frac{\mu}{2\pi k_B T}\right)^{3/2} \int_0^\infty v^3 e^{-\mu v^2/2k_B T} W^2 \pi^2 a_0^2 \left(1 + \frac{2Ze^2}{m_e v^2 W_0 a_0}\right) dv && (36) \\ &= \pi W^2 a_0^2 \left(\frac{8 k_B T}{\pi m_e}\right)^{1/2} \left(1 + \frac{Ze^2}{W a_0 k_B T}\right), && (37) \end{aligned}$$

were we have again set $\mu = m_e$.

The term $Ze^2/(a_0 k_B T)$ that appears in the second parentheses has the numerical value $15.78(Z/T_4)$, where T_4 is the temperature in units of 10^4 K. Thus unless $Z/T_4 \ll 1$, which is generally not the case in optical nebulae, we can drop the 1. Doing so and recalling that $a_0 = \hbar^2/(m_e e^2)$, we obtain

$$k_{u\ell} \approx \frac{h^2}{(2\pi m_e)^{3/2}} \frac{1}{(k_B T)^{1/2}} 2WZ. \tag{38}$$

Thus we have written the collision rate in terms of the unknown parameter W, which is a measure of how easy it is to perturb the atom. Based on this argument, we formally define the dimensionless **collision strength** $\Omega_{u\ell}$ of a particular interaction by

$$k_{u\ell} \equiv \frac{h^2}{(2\pi m_e)^{3/2}} \frac{1}{(k_B T)^{1/2}} \frac{\Omega_{u\ell}}{g_u}. \tag{39}$$

The reason for including the factor of g_u, the degeneracy of the upper state, will become apparent momentarily. The advantage of this definition of $\Omega_{u\ell}$ is that it "factors out" the dependence of the reaction rate on the kinetics of the plasma, which is essentially the same for any reaction, and isolates the quantum-mechanical part that is reaction-specific.

Collision strengths must be calculated quantum mechanically or measured in the laboratory. Exact values are generally known only to the $\sim 10\%$ level, except for a few very well-studied cases. Unfortunately many of the astrophysically relevant collisions are quite difficult to study on Earth, because the lines they correspond to are very weak and hard to see under terrestrial conditions. Also note that in general $\Omega_{u\ell}$ can be a function of temperature, but in practice it is at most a very weak one, and the temperature dependence can be dropped.

D. Inverse collision rates

For the case of a collision that causes a change in quantum state, we generically have a forward process and a reverse process. That is, suppose that we have some species X that can be in a lower state $X(\ell)$ or an upper state $X(u)$, and that can transition between these states due to collision with another species Y – in the case we just considered Y is an electron, but the result we are about to demonstrate is more general, and applies regardless of what type of particle Y is. Thus we have the generic forward-backward reaction pair

$$X(\ell) + Y \leftrightarrow X(u) + Y. \tag{40}$$

The rates at which the left-hand reaction occurs is $k_{\ell u} n_{X(\ell)} n_Y$, and the rate at which the right-hand one occurs is $k_{u\ell} n_{X(u)} n_Y$, where $k_{\ell u}$ and $k_{u\ell}$ are the rate coefficients for the excitation and de-excitation reactions. Now consider a system that is statistical steady state, so these two reaction rates are equal:

$$k_{\ell u} n_{X(\ell)} n_Y = k_{u\ell} n_{X(u)} n_Y \qquad \Longrightarrow \qquad k_{\ell u} = k_{u\ell} \frac{n_{X(u)}}{n_{X(\ell)}}. \tag{41}$$

Now we know that if the system is in thermal equilibrium at temperature T, the ratio of particles in state u to particles in state ℓ is not free: it must be described by a Boltzmann factor. Thus in thermal equilibrium, we must have

$$k_{\ell u} = k_{u\ell} \frac{g_u}{g_\ell} e^{-E_{u\ell}/k_B T}, \tag{42}$$

where $E_{u\ell}$ is the difference in energy between the upper and lower states, and g_u and g_ℓ are the degeneracies of states u and ℓ.

Thus if we know the excitation or de-excitation rate coefficient and the gas temperature T, we can immediately calculate the other one. For the specific case of collisions between electrons and ions, substituting in our result above, we have

$$k_{\ell u} = \frac{h^2}{(2\pi m_e)^{3/2}} \frac{1}{(k_B T)^{1/2}} \frac{\Omega_{u\ell}}{g_\ell} e^{-E_{u\ell}/k_B T}. \tag{43}$$

Note that the g_u in the denominator in equation 41 has cancelled and turned into a g_ℓ, which explains why we inserted it in the first place: doing so maintains symmetry between the expressions for collisional excitation and de-excitation.

A further point worth making is that the relationship between $k_{\ell u}$ and $k_{u\ell}$ applies only if the particles have a Maxwellian velocity distribution, since this is required for thermodynamic equilibrium – we can only define a temperature in the first place if this assumption holds. However, we can make an even more general statement as applied to the microphysical cross sections for the forward and backward reactions. Recall that, from equation 17,

$$k_{u\ell} = \sqrt{\frac{8k_B T}{\pi\mu}} \int_0^\infty x e^{-x} \sigma_{u\ell}(x)\, dx, \tag{44}$$

and similarly

$$k_{\ell u} = \sqrt{\frac{8k_B T}{\pi\mu}} \int_{E_{u\ell}/k_B T}^\infty x e^{-x} \sigma_{u\ell}(x)\, dx, \tag{45}$$

where $x = E/k_B T$, and in the second integral we have set the lower limit to $E_{u\ell}/k_B T$ because clearly the cross section for excitation must be zero if the collision energy is smaller than the energy difference between the levels.

If we use our relationship between $k_{u\ell}$ and $k_{u\ell}$, we immediately get

$$\frac{\int_0^\infty E e^{-E/k_B T} \sigma_{u\ell}(E)\, dE}{\int_{E_{u\ell}}^\infty E e^{-E/k_B T} \sigma_{\ell u}(E)\, dE} = \frac{g_\ell}{g_u} e^{E_{u\ell}/k_B T}. \tag{46}$$

Re-arranging, we have

$$\int_{E_{u\ell}}^\infty E e^{-E/k_B T} \sigma_{\ell u}(E)\, dE = \int_0^\infty \frac{g_u}{g_\ell} E e^{-(E+E_{u\ell})/k_B T}\, \sigma_{u\ell}(E) dE. \tag{47}$$

For the integral on the left, let us make a change of variable $E' = E - E_{u\ell}$. This gives

$$\int_0^\infty (E' + E_{u\ell}) e^{-(E'+E_{u\ell})/k_B T} \sigma_{\ell u}(E' + E_{u\ell})\, dE = \int_0^\infty \frac{g_u}{g_\ell} E e^{-(E+E_{u\ell})/k_B T}\, \sigma_{u\ell}(E) dE. \tag{48}$$

Clearly these two integrals can be equal for arbitrary T, as they must be, only if

$$(E + E_{u\ell}) \sigma_{\ell u}(E + E_{u\ell}) = \frac{g_u}{g_\ell} E \sigma_{u\ell}(E). \tag{49}$$

This rule applies to the energy-dependent cross section itself, which is a function solely of the microphysical properties of the atoms in question. Thus we have managed to constrain even atomic physics based on our statistical equilibrium arguments.