The ISM and Galaxies

Contents

1. Continuum radiation

Today we are talking about continuum (or continuous spectrum) radiation.

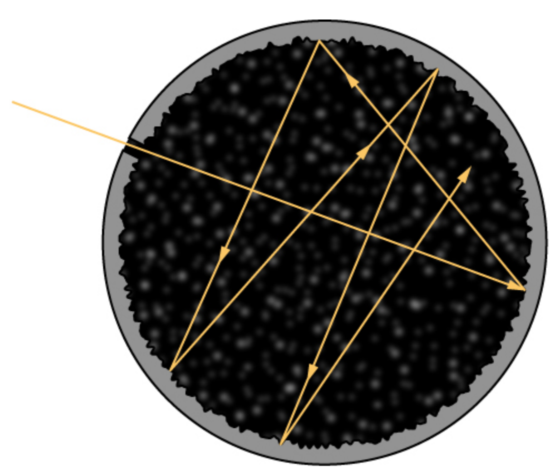
Can you name any types/sources of continuum emission?

- Blackbody (thermal emission, stars and planets and the CMB)
- Bremsstrahlung (free-free emission, nebulae)
- Synchrotron (relativistic magnetobremsstrahlung, SNRs, AGN jets)
- Cyclotron (nonrelativistic gyro radiation, auroras and Jupiter's magnetosphere)

What is the difference between these various types of emission? Consider:

- What are the emitting particles? (Charged particles; generally electrons)
- Are the emitting particles in thermal equilibrium (LTE) with a Maxwellian velocity distribution, or are they nonthermal? (Blackbody radiation is a thermal process, but magnetobrehmsstrahlung involves electrons with a power law velocity distribution)
- Is the velocity distribution of the emitting particles nonrelativistic or relativistic? (Synchrotron originates from highly relativistic electrons, while cyclotron emission originates from sub-relativistic or slightly relativistic electrons)
- What is the opacity of the emission-- is it reabsorbed? (Blackbody radiation generally requires a high opacity).

Today we are going to focus on Blackbody radiation. A classical setup to mimic Blackbody radiation is to make a 3D cavity, into which you input radiation through a small opening.



The radiation bounces around in the cavity (being both absorbed and reflected), but basically cannot escape before it is fully absorbed. If you keep inputting radiation the cavity surface will heat up until eventually an equilibrium is reached where the cavity surface will absorb as much energy as it radiates. This radiation is known as Blackbody radiation. We want to know: what is its spectrum? How much energy is radiated at each wavelength?

If you try to determine this using classical physics, you will be very wrong!!
Classical physics says that the energy of light in a 3D cavity is equally distributed among all of the "modes" of a system (Equipartition Theory; we can think of these modes in this case like standing waves on a string where the nodes of the string are on the edges of the cavity).

For **thermal radiation**, each mode has an energy of

$$E = k_B T \quad (1)$$

and the number of modes per unit frequency is proportional to ν^2.
The constant here is the Boltzmann constant ($k_B = 1.38 \times 10^{-23} J/K$). Note also that this exact equation is correct for *radiation only*; for the thermal motion of particles we will need an additional constant of proportionality)

Using the Equipartion theory led to an expression (correct at low frequencies) called the **Rayleigh-Jeans law** for thermal continuum emission:

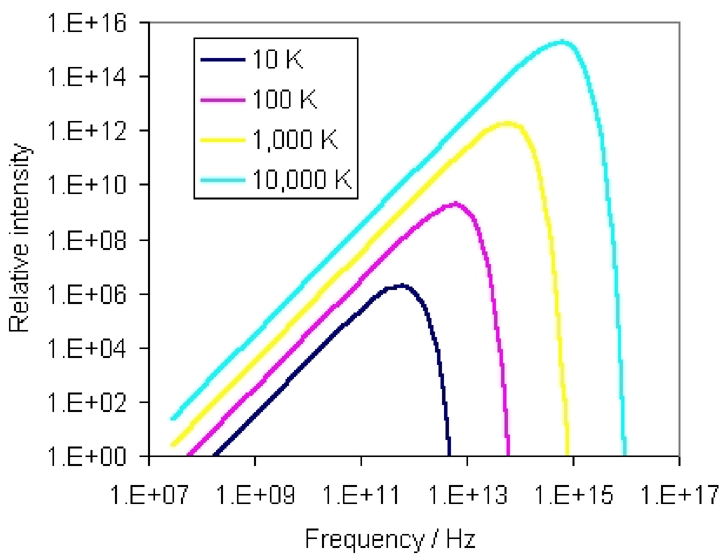
$$B(\nu, T) = \frac{2\nu^2 kT}{c^2} \quad (2)$$

[note that the SI units of B(ν,T) are J s^{-1} m^{-2} Sr^{-1} Hz^{-1}] You can see the problem here: the more modes you have (the higher the frequency of a wave) the more energy you would have-- all the way to infinity at high frequencies. This discrepancy between what was predicted and what was actually observed at high frequencies was known as the ultraviolet catastrophe. Luckily, Planck came along to usher in the age of Quantum Mechanics by saying that ACTUALLY, energy is quantized in packets no smaller than

$$E = h\nu = \frac{hc}{\lambda} \quad (3)$$

This is the equation we use for the **energy of a single photon** and leads to the correct equation for Blackbody emission: **Planck's Law**, which describes emission from an opaque body in thermal equilibrium with its surroundings (no heat transfer; constant temperature)

$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \qquad (4)$$



The straight line part of the spectrum here is the Rayleigh-Jeans tail (**Equation 2**), where the intensity is proportional to $\boldsymbol{\nu}^2$ (a power law or a straight line in log space).

Note that an ideal blackbody then has an intensity determined solely by the source temperature (and the frequency at which you are observing it). Most astronomical objects have some emissivity which either uniformly lowers the source intensity from that of an ideal blackbody, or which modifies the intensity as a function of some variable, like frequency. This is true for sources like dust grains (they can't emit like ideal blackbodies because they are inefficient at radiating at wavelengths larger than their size).

Take this opportunity to remind you of something useful: units are going to be your friend (and enemy, TBH) in this class. When you have a complicated equation, always check your units on both sides. If you don't know what equation to use, ask what units the answer should have, and look at the units of the provided information. And remember: any quantity inside an exponential should be UNITLESS.

2. Spectral Lines and Energy Levels in Atoms and Molecules

	QNs	Energies/ wavelengths	Requirements	Examples
Electronic	n	X-ray-Radio ($h\nu \simeq 1$ eV)	Electron orbital physics (Fermi Statistics)	Hα, Iron K-α, Radio recombination
Vibration	$v_1, v_2, \ldots v_n$ Stretching, Bending	Infrared ($h\nu \simeq 0.1$ eV) Bend ~0.01 eV Stretch ~0.5 eV	Molecules with at least two atoms (stretching along a bond) or at least three atoms (bending) v⇧= higher-order harmonics	CO bandhead, PAH features from small dust grains
Rotation	J, K, K' Angular momentum, Projection along rotation axes	FIR-Radio ($h\nu \simeq 10^{-3}$ eV) m_{mol}⇧ν ⇩ HC_3N 1-0 3.2cm (~10^{-5} eV) H_2 2-0 28μm (~0.04 eV)	Need distinguishable axes to observe rotation (molecules with at least 2 atoms) J⇧ = faster rotation Up to 3 quantum numbers based on symmetry	CO H_2O

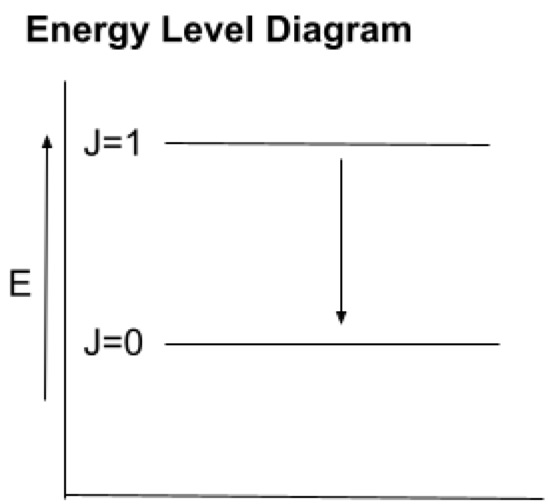
This is a cool time to note: each of these is a different energy mode or **internal energy state**. For a gas of molecules in a perfect thermal equilibrium, the energy in all of these modes or states would be the same (equipartition theory!) and the system would be characterized by a single temperature. We will talk more about temperature soon!

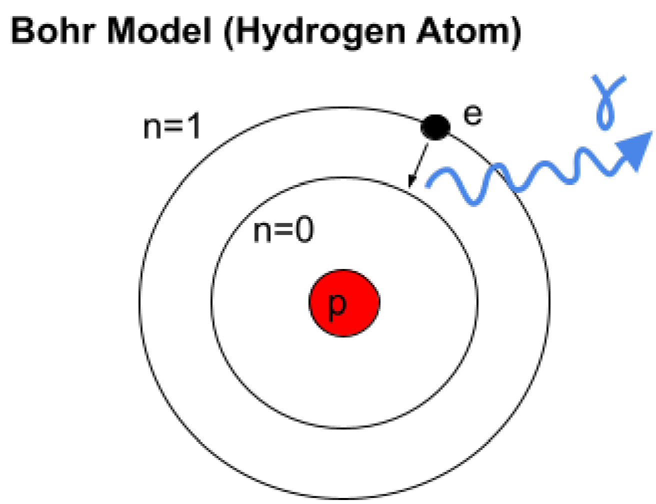
Additional perturbations on the energy structure of atoms/molecules:

- Fine structure: Interactions of electron spin and orbital angular momentum can split the electronic states (H alpha is actually a doublet)
- Hyperfine structure. Magnetic and charge interactions between electrons and nuclei lead to more transitions with smaller energy differences. Famous example: Neutral hydrogen spinflip transition @21 cm between 2 states that H atom can be in: with the spins of the proton and electron aligned and anti-aligned.

You might have noticed– where are the high-energy (e.g. optical, UV, gamma-ray) transitions, particularly for the molecules? Well, for molecules, photon energies (hν) at such short wavelengths start to approach or exceed bond energies. Typical bond energies are on the order of one to a few eV. For atoms, even more energetic photons can exceed ionization threshold (13.6 eV for a hydrogen atom) unless you have electrons in the inner orbitals of heavy atoms with a large nuclear charge (Fe for example has bound-bound transitions out to the X-ray).

How do we observe these different internal energy states? Basically, an atom or molecule has to *change* its internal energy state. If you change the amount of energy a molecule or atom is storing, that energy has to go (or come from) somewhere else. The easiest way for us to observe this is by the emission or absorption of a photon. A transition between energy states is often diagrammed as:





Remembering our definitions of photon energy, the frequency of the photon (γ) involved will be $\Delta E \;=\; h\nu$

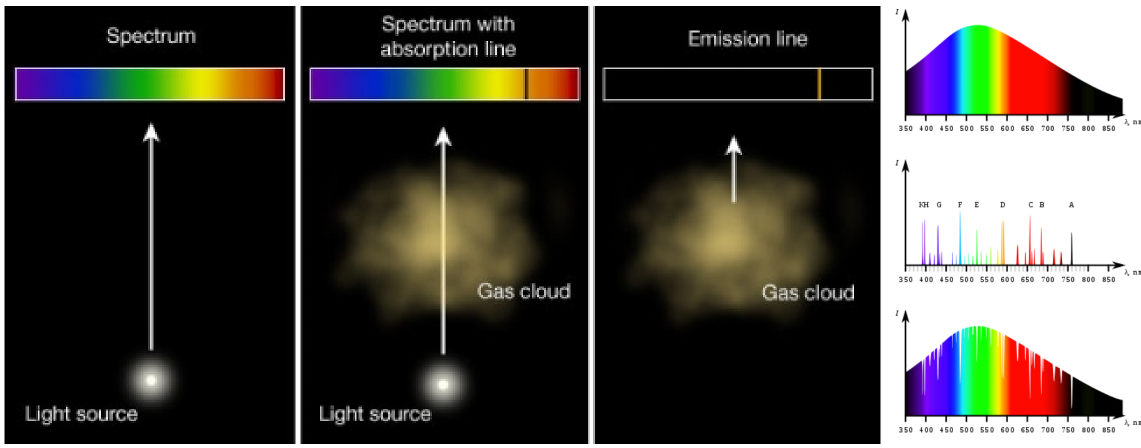
Classically– what do we expect the **width** (frequency spread) of a **spectral line** (all the photons coming from atoms or molecules undergoing a specific energy transition) to be? Basically, a delta function: an infinitely narrow spike. Quantum mechanics however says no to that– each of these energy levels is actually a quantum state with some lifetime (the decay of these states is what leads to the emission of photons we observe!). The Heisenberg uncertainty principle then says for states of finite lifetime, the energy of the state can only be specified as precisely as

$$\Delta E\,\Delta t \;\geq \frac{\hbar}{2} \qquad (5)$$

Finally, as observers of light, we can make a holistic model including both line and continuum emission, and when we expect to see each type (**Kirchoff's Laws**):

- Hot, opaque sources produce Continuum emission (see previous lecture)
- Cool(er) gas in front of a hot source produces Absorption lines
- Just a cloud of gas produces Emission lines

Show these as both spectra and Blackbody plots:



When you put cool gas in front of a continuum-emitting source, you will see dips taken out of a blackbody distribution (e.g. see absorption in stars from the cooler upper photosphere)

Spectral line emission is almost always a (nearly) thermal or more typically sub-thermal process. This means that the spectral lines, in emission, have brightnesses that don't go above the blackbody curve for thermal continuum emission at the same temperature.

3. Collisions and the ISM

What can happen when two gas particles hit each other?

- Transfer energy/momentum from one to the other.
- Form or destroy a molecule.
- Become excited-- start spinning, vibrating, or boost an electron.
- Maybe even become ionized.
- (Not yet really talking about dust, but in that case collisions can knock off surface molecules/ices or even start destroying the grain itself).

You can see that collisions then can be responsible for a lot of things that happen in a gas. How likely is it that these collisions occur, especially for gas clouds in space?

For our first example: let's start by redefining a quantity called a number density.

Number of particles (can be specified to be those of a particular species) in a given volume (with corresponding SI units of m^{-3}):

$$n = N / V \tag{6}$$

Where big N is the number of particles and V is the volume.

Now we can put the interstellar densities we will work with into context with our own experience by calculating the number density of air in this classroom.

What is the density of air? The density of dry air at standard temperature and pressure is 1.28 kg / m^3. But what is air made out of? 78% N and 22% O. In what state is the nitrogen and oxygen? It is molecular, so N_2 and O_2. So then, how many air molecules would we find per cubic meter? This is the number density. We can define a weighted mass for a 'typical' particle of:

$$\overline{m} = \frac{\%A\,(m_A) + \%B(m_B) + \%C(m_C) + ... + \%N(m_N)}{N} \tag{7}$$

Calculating the weighted mass for this situation:

$$\overline{m} = (0.78M_{N2} + 0.22M_{O2}) = 28.8 \text{ x } (1.66\text{x}10^{-27} \text{ kg})$$

We can also define the number density as:

$$n = \rho / \overline{m} \tag{8}$$

Then we can calculate that

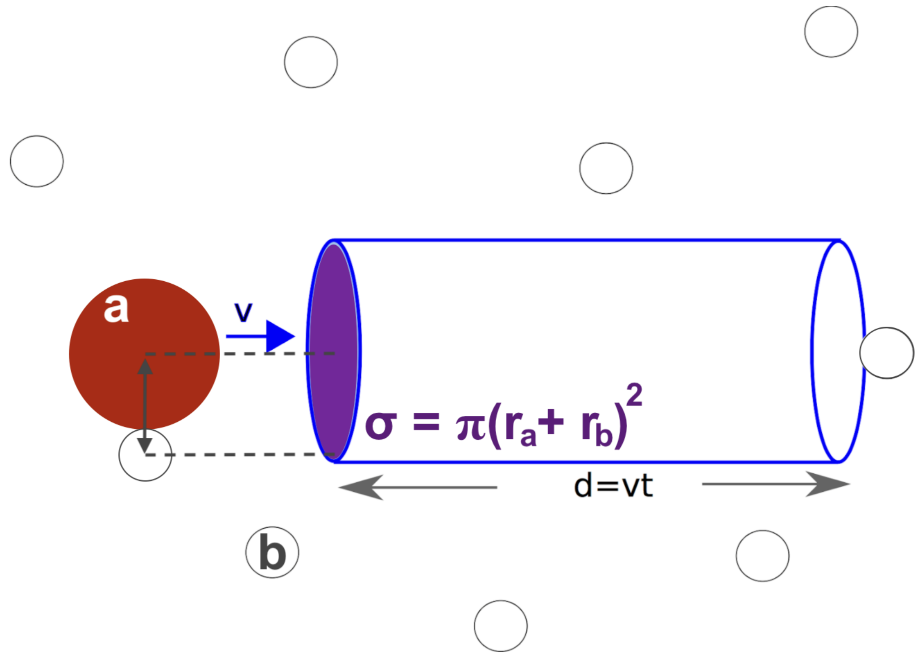
$$n = (1.28 \text{ kg / m}^3) / (28.8\text{x}1.66\text{x}10^{-27} \text{ kg}) = 2.7\text{x}10^{25} \text{ molecules m}^{-3}$$

We often want to calculate the typical time between collisions. For this, we are going to introduce a standard form that will apply to many of the timescales that we will investigate in this class:

$$t = X / \frac{dX}{dt} \tag{9}$$

This just says that to define a timescale (with SI units of s) we can often describe it as some quantity X divided by the rate of change of that quantity with time: dX/dt.

To identify the relevant variables here and calculate the typical time between collisions we can draw a handy diagram:



Here we are going to describe our timescale as d/v, where d is the distance that particle *a* travels before encountering another particle of type *b*, and v is the typical speed of particle *a*, relative to these collision partners (in this example, the particle and its collision partners are both hydrogen atoms, but sometimes they will be different).

It turns out that here d is a quantity that we will see again: the mean free path, with SI units of m. We can define a mean free path for particles of species *a* interacting with particles of species *b* as:

$$\lambda_{MFP} = \frac{1}{n_b \sigma} \qquad (10)$$

Here n_b is the number density of the collision partners (species *b*).
We have also defined σ – the collisional cross section with SI units of m^{-2}. This is a function of the minimum separation between *a* and *b* in order for a collision to occur:

$$\sigma = \pi(r_a + r_b)^2 \qquad (11)$$

Now, we can define the **collision timescale**:

$$t_{collision} = \frac{\lambda_{MFP}}{v} = \frac{1}{n_b v\,\sigma} \qquad (12)$$

For a typical nebular density of $n_H = 10^8$ m^3 and using an H-atom diameter of 1 Å:

$$t_{collision} = \left[(10^8\, m^{-3})(30,000\, m/s)(3.14)(10^{-10}\, m)^2\right]^{-1} = 1 \times 10^7\, s$$

This is ~0.3 years so atoms in space are not regularly hitting each other!

We can use this to define a per-particle collision rate (with SI units s^{-1}) for species *a*:

$$Rate\ per\ particle = \frac{1}{t_{collision}} = n_b v\sigma \qquad (13)$$

We can also define the **volumetric collision rate** (the rate of collisions occurring in a given volume, a quantity which has SI units m^{-3} s^{-1}):

$$Z_{collision} = n_A n_B\,\sigma\, v \qquad (14)$$

The volumetric collision rate turns out to be a very handy way to describe many different processes in the ISM, and we will be coming back to this a lot!

4. Gas Cooling

Why do we care about cooling the gas in the ISM?

- We have to cool gas if we want it to contract and form stars. Otherwise the thermal motion of the gas will support it against gravitational collapse.
- We also need to cool gas if we want to form molecules

What are some ways that we can cool a gas? Ultimately, we need to get energy out of it. How can we do this?

- Though effective in terrestrial conditions (or in stars), conduction and convection are not efficient at the densities we are considering here
- Radiation: spontaneous emission, recombination, or free-free emission
- Collision: (we will talk later about collisions with dust, say, which it turns out can generally cool more efficiently than the gas via continuum radiation)
- Collision-Excitation-Radiation: Gas particles collide with other gas particles, and so the kinetic energy of the gas is transferred to internal energy states, which then usually decay quickly by emitting a photon

To be efficient, collisional cooling in atomic/molecular gas requires the following:

(1) Collisions must be frequent, not requiring a low-abundance partner (say, F)
(2) The excitation energy must be comparable to or less than the thermal kinetic energy of the gas particles. This rules out H_2 for cooling 10 K gas, as though it is the most abundant molecule, its E_{ex} ~500 K makes it not an efficient coolant.
(3) There must be a high probability of excitation from collision (generally this can be assumed to be true)
(4) There must be a high probability of emitting a photon after being excited (H_2 actually fails here: it is symmetric and so lacks a dipole moment, emitting photons more rarely a quadrupole interaction. Only the sheer abundance of H_2 makes it an important coolant for gas at temperatures ~100 K
(5) The gas must be optically thin so that the photons are not immediately re-absorbed (this causes trouble for CO in cool molecular gas at high densities).

Let's do an example, and for those of you following along in the book, this is going to roughly follow Problem 3.1 in Dyson & Williams (note that you can find solutions in the back of this book!)
Recalling **Equation 9**, we will describe the cooling time as:

$$t_c = -T/\left(\frac{dT}{dt}\right) \tag{15}$$

This is approximately the time, given the current rate of cooling, to go from the current temperature to 0 K. Note that this is only an order of magnitude estimate: the cloud will not actually cool to 0 K, and the cooling rate usually has some temperature dependence.

Let us take a gas cloud which has number densities of $n(\mathrm{H}) = 10^8\ \mathrm{m}^{-3}$, $n(\mathrm{e}) = 10^5\ \mathrm{m}^{-3}$, and $n(\mathrm{C}^+) = 4 \times 10^4\ \mathrm{m}^{-3}$. The temperature of this gas cloud is 200 K and we will assume that the only way that it can cool is through the recombination of electrons and C^+ ions:

$$C^+ + e^- \rightarrow C + h\nu \qquad \textbf{(IV)}$$

This results in a somewhat complex function for the C^+ **cooling rate:**

$$\Lambda(C^+) = 8 \times 10^{-33} n_e n_{C+} T^{-0.5} e^{\left(\frac{-92\,K}{T}\right)} \quad (16)$$

Note that this is a **volumetric rate** and as such it has units of J m^{-3} s^{-1}. We could make this into a cooling rate per C^+ molecule (with units J s^{-1}) by dividing by n_{C+}.

This is a good time to notice that volumetric rates based on particle interactions (collisions) are going to have a very characteristic form (see **Equation 14**):

$$Z = k\, n_A n_B \quad (17)$$

Here k is a generic **rate coefficient** (with SI units of m^3 s^{-1}) and n_A and n_B are the number densities of the two interacting species (**reactants**) on the left-hand side of the reaction equation (in this case, e and C^+). Using this, we can define a cooling rate coefficient for this reaction of:

$$k_{C+} = 8 \times 10^{-33} T^{-0.5} exp\,(\frac{-92\,K}{T}) \quad (18)$$

This complicated mess is a function of a number of things: the collisional cross section σ, the relative speed of the interacting particles (which typically introduces a temperature dependency), and the likelihood that a collision results in an excitation.

We can see that we have numerical values for all of the variables in this relation, however it is not yet clear what we do with the cooling rate once we calculate it.

For this, we must match the units in our expression for the cooling time, which we can do by recalling **Equation 3**. Now, it turns out that for particles (as opposed to photons, in **Equation 1**) the equipartition of energy is a little different: you get $\frac{1}{2}kT$ for each degree of freedom. Assuming 3 degrees of translational freedom, we can define the energy density (energy per unit volume) of a gas at a temperature T as:

$$\epsilon \;= \frac{E}{V} = \frac{3}{2} nkT \qquad (19)$$

Here, ϵ has units of J m^{-3} . We can then divide the energy per unit volume by the cooling rate per unit volume to get an answer with units of time: the cooling time.

$$t_c = \frac{3/2\, n_H kT}{8\times10^{33} n_e n_{C+} T^{-0.5} exp(-92\, K\, /\, T)}$$

$$t_c = \frac{3/2\, (10^8)(1.38\times10^{-23})(200)}{8\times10^{33}(10^5)(4\times10^4)(200)^{-0.5} exp(-92\; /\, 200)}$$

$$t_c = \; 1.16\; \times\; 10^{13}\, s \;=\; 3.6\; \times\; 10^5\, yrs$$

5. Heating Rate

What are some different kinds of temperatures a system can have? (Hint: think of temperature as a distribution of energies)

- Kinetic temperature (speed of particles)
- Radiation temperature, T_{eff} (spectrum of light, comparison to a blackbody)
- Ionization temperature (distribution of ionization states in an ensemble)
- Excitation temperature (distribution of excitation states in an ensemble)

Are these always the same? No! This goes back to the idea of the **Equipartion Theory**. *if* the system is in perfect thermal equilibrium, then the energy is distributed equally between all of these modes, and they all have the same temperature. But this is almost never the case in the ISM!

Why do we care about how gas in the ISM gets heated?

- Heating gas can make it less likely that it will form stars (feedback).

- Heat it enough and the gas particles can actually be moving so fast that they will escape the galaxy.
- Heating the gas can greatly change the chemistry (e.g., more energy to remove molecules and ices from dust grain surfaces)

Macroscopically, what sources in a galaxy can be responsible for heating the gas?

- Stars (photons and direct injection of momentum from winds)
- Supernova remnants (cosmic rays, high-energy photons, and shocks)
- Supermassive black holes (release of gravitational potential energy into X-rays)

Microscopically, how does heating happen? Here, I want to call attention to a fundamental difference between heating and cooling, as these processes can look somewhat similar. Situations in which a gas particle gives off a photon almost always results in cooling, because gas is typically optically thin (at least at the long wavelengths at which molecules tend to give off cooling radiation). Any photons emitted are unlikely to interact, and have a large probability of escape. To be effective, a heating process then generally needs to result in an increase of the bulk kinetic energy (not just energy of internal states, which tends to be quickly radiated away).

This is generally accomplished by the incidence of very high-energy photons knocking SOMETHING off of a particle. This can be:

1. Photoionization, creating free electrons having excess energy above the energy used to free them from the atom (the ionization potential)
2. Photodissociation, breaking a chemical bond and increasing the energy of the remaining particles
3. Photoelectric heating in which photons incident on dust grains also release free electrons.

For the gas, we can then generally describe a few important reactions for the balance of heating and cooling in the ISM:

Collision: energy goes into internal energy of particle

$$\mathbf{A + B \rightarrow A + B^{*}} \qquad \mathbf{(I)}$$

De-excitation: collision energy is quickly radiated away

$$B^* \rightarrow h\nu + B \qquad \text{(II)}$$

Ionization: results in a free electron

$$A + h\nu \rightarrow A^+ + e^- \qquad \text{(III)}$$

Recombination: will cool the gas

$$A^+ + e^- \rightarrow A + h\nu \qquad \text{(IV)}$$

Note that for this heating to be efficient, the rate of (III) should exceed the rate of (IV)

Let's do an example, and for those of you following along in the book, this is going to roughly follow Problem 3.3 in Dyson & Williams.

Let's define a **volumetric heating rate** (energy per unit volume and time) based on the ionization of atom X. We will call this heating rate G, with units of J m^{-3} s^{-1}. To determine the net energy gain from ionization, we have to remember that some of the photon energy went into the ionization potential, so define the energy gain from ionization as:

$$E_{ion} = h\nu - I \qquad (20)$$

Where I is the ionization potential (13.6 eV for hydrogen).
Now, assuming the gas is in **ionization equilibrium**, then we can say that the number of ionizations are equal to the number of recombinations, and we can express the heating rate as:

$$G = \beta\, n(e)\, n(X^+)\, E_{ion} \qquad (21)$$

Where β is the recombination rate (SI units of m^3 s^{-1}), or the number of recombinations per unit time and volume, and the dependence on n_e and n_{X+} reflects that a recombination requires the interaction of both of these species: these are the number densities of the species on the left-hand side of **Reaction IV**.

Recall the gas cloud from the last lecture which has number densities of $n(e) = 10^5\ m^{-3}$, and $n(C^+) = 4 \times 10^4\ m^{-3}$. We want to calculate the heating rate in this same cloud due to the ionization of C+. Assume the recombination rate is $\beta = 10^{-17}\ m^{-3}\ s^{-1}$ and that E = 2.1 eV

Then we would calculate

$$G = (10^{-17})(10^{5})(4 \times 10^{4})(2.1)(1.6 \times 10^{-19}) = 1.3 \times 10^{-26}\ J\ m^{-3}\ s^{-1}$$

Where are some environments where the cooling or heating processes we have discussed be inefficient?

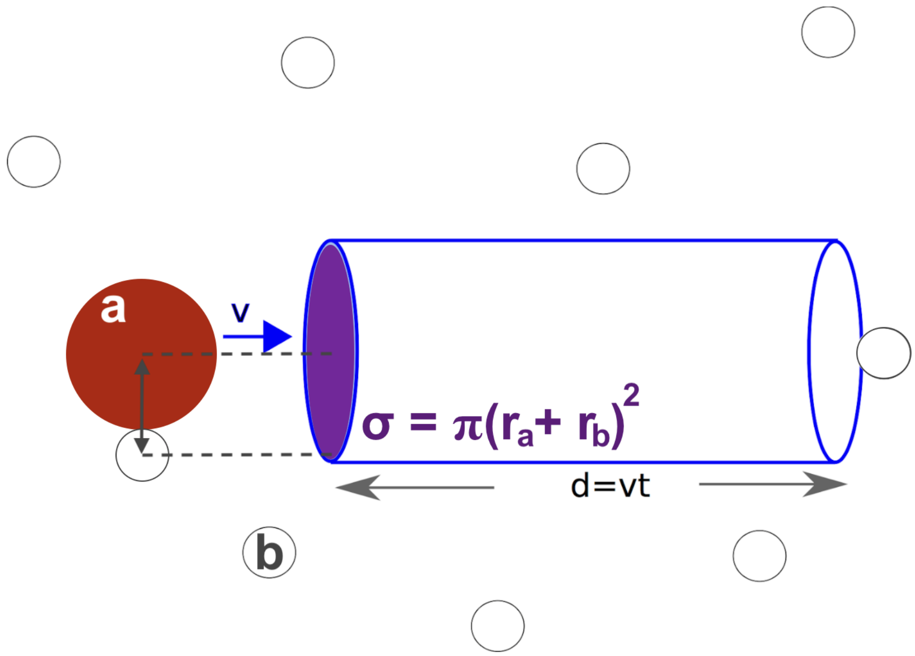
- Heating due to photons: inefficient deep in clouds where ionizing radiation cannot penetrate.
- Cooling due to photons runs into big problems if the cloud becomes optically thick to line radiation.

6. Astrochemistry Part 1

Let's make some molecules!

We know there are molecules in space because we have seen them: H_2 is the most abundant (99.99999% of molecules, with CO a distant second at 0.00001%) and we have detected molecules in interstellar space with up to 13 atoms (and since dust grains are basically big atoms, we suspect they can get bigger).
The main question for astrochemistry is (apart from how you make them)-- how FAST can you make them? What is the rate for every different construction (and deconstruction) of molecules that can explain the abundances of molecules that we observe.

Fundamentally, since you have to get particles together to build bigger ones, this rate is going to depend heavily on the collision rate. We previously defined a volumetric collision rate in **Equation 14** ($Z_{collision} = n_A n_B \sigma v$), which we visualized as follows:



We can compare this to our general rate expression (**Equation 17**) and see that we can then define a typical **collisional rate coefficient** for an interaction like **A + B → A + B***:

$$k_{collision} = \sigma v \qquad (22)$$

Where $k_{collision}$ has the required SI units $m^3\ s^{-1}$.

What is the volumetric collision rate for two H atoms, assuming a typical density of 10^8 m^{-3} and a typical diameter of 1 Å and a speed of 0.4 km/s?

$Z_{collision} = (10^8\, m^{-3})^2 (\pi)\, (10^{-10}\, m)^2 (400\, m/s) = 0.12$ per second per cubic meter.

That's not so bad! Especially because we are talking about HUGE clouds in space. The volume of just this classroom was on the order of 100 cubic meters, so in just this volume you would get multiple collisions per second for these sorts of densities.

Then, the most basic way you might imagine making a molecule: Slam a couple hydrogen atoms together and call it a day.

$$H + H \rightarrow H_2 + \gamma$$

This is an example of a more general type of reaction called radiative association:

$$\mathbf{A + B \rightarrow AB + h\nu} \qquad \textbf{(V)}$$

(note that you can see γ and $h\nu$ used interchangeably to represent an emitted photon, or more specifically the energy it carries away). It turns out this kind of reaction is actually hard because colliding isn't enough: the reaction rate is not just the collision rate. For the atoms to stick together, they have to quickly give off a sufficiently-energetic photon. How quickly? They are generally in contact for 10^{-13} s, or less than a picosecond!! As a molecule, H_2 is especially bad at doing anything in this amount of time because it lacks a permanent dipole moment, and so has to rely on quadrupole moment to give off a photon, which takes basically forever. This additional contribution to the rate basically tanks the likelihood of making H_2 this way, and you will calculate by how much in Problem Set #2. (Note that this likelihood of giving off of a photon would then be a part of the rate coefficient k for this reaction)

How to get around this? You can introduce another particle to take away the excess energy. 3-body reactions are hard in space (again, see Problem Set #2), but you can instead have a case where

$$\mathbf{AB + C \rightarrow AC + B} \qquad \textbf{(VI)}$$

This is called an exchange reaction and it makes a more promising route for forming H_2:

$$H^- + H \rightarrow H_2 + e$$

The electron carries away the excess energy so this proceeds relatively quickly, but the tricky thing is that the rate of this process now also depends on your rate of forming H^-:

$$H + e^- \rightarrow H^- + \gamma$$

This reaction (another radiative association) is slow (except at high T), so it usually limits the rate of forming H_2 this way.

Once you have a molecule, you can also make it even easier to add an atom if you increase the cross section (A or σ) by adding an attractive (Coulomb!) force to the equation. These are Ion-Molecule reactions (the ion induces a dipole moment in the molecule):

$$\mathbf{AB + C^+ \rightarrow AB^+ + C} \qquad \mathbf{(VII)}$$

Note that these reactions can have many different results: they could be a charge-exchange reaction (as seen here in **VII**) or a radiative association (**V**) or even a recombination **(IV**). The description Ion-molecule then applies only to the reactants. These reactions generally proceed much more quickly than reactions between neutral species, and dominate a lot of the chemistry in the interstellar medium.

7. Dust Properties

Let’s talk about dust! How do we know there is dust in space?

- It blocks the light (UV, optical, even IR-- where it makes distinct spectral features)
- It gives off some light (cold dust emits like a blackbody at mm wavelengths, warm dust emits like a blackbody in the IR, and one also sees PAH spectral features)
- It polarizes light (dust grains are paramagnetic)
- Elements that can make dust (e.g., like we see in rocky planets like the earth) are depleted in the interstellar gas

Where does dust come from? Dust formation needs higher densities than typical interstellar clouds (as I will show in an example during the next lecture). It also takes high temperatures, typically T~1000-2000 K.

- AGB stars (winds, all the way to pre-planetary nebulae)
- Massive stellar winds (wolf rayet stars and luminous blue variables)
- Novae and supernovae

Where does it go?

- Any ice mantles that a grain accumulates can be evaporated at high temperatures (either behind shocks, or due to proximity to a strong radiation field)
- The dust itself is most likely destroyed through a process called sputtering: getting hit by a high-energy atom that knocks another atom out of the material lattice. Note: While supernovae are responsible for driving (many of) the shocks that generally destroy dust in clouds, these days supernovae are also regarded as dust factories. It’s complicated!

What is it made of?

- Silicon, Iron, Magnesium, and other metals (elements that can make heat-stable solids).
- There is also Carbon and Oxygen, but at T~ few 1000s K, we may expect most C&O to be locked up in CO molecules.

How much is there?

- The dust-to-gas ratio is typically assumed to be 1:100, by mass.

What is it good for?

- Heating (photoelectric effect)
- Cooling (can deplete molecules from gas phase & radiates like a blackbody)
- Making molecules

Example: For those of you following along in the textbook, this is problem 4.4 from Dyson and Williams. This calls back to **In-Class #5** where we found an equilibrium temperature for cold neutral gas that is cooled and heated solely by C^+.

Now we will consider the dust, and assume that we have an interstellar grain for which the heating is entirely powered by the absorption of UV light, and cooling occurs through blackbody emission (**Equation 4**), with an efficiency of 1%. From this, we want to determine the equilibrium dust grain temperature. We therefore need to set up an expression that balances the heating rate and the cooling rate.

We are told that the flux of UV light is 10^{10} photons per m^2 per s per nm (over a bandwidth of 100 nm).

To make this easier (e.g., not having to explicitly consider grain size) we will balance a heating *flux* and a cooling *flux*, e.g. an energy per time per unit area. (or SI units of J s^{-1} m^{-2}). This is very similar to the units of volumetric heating and cooling rates we have previously defined (**Equations 16 and 21**). We are also going to gloss over some integrals.

Starting with the heating flux: technically, this is an integral of the UV intensity over some wavelength range:

$$F_{Heat} = \int I(\lambda)\, Q_{abs}(a, \lambda)\, d\lambda \qquad (23)$$

Let's assume that a typical UV photon energy is 13.6 eV (the energy to ionize H) and the grain absorbs all the UV light, so that Q=1. Then we can convert from a photon to an energy flux. Further assume the flux is not a function of wavelength, and we can say:

F_{HEAT} = (1) (N_{UV})(E_{UV})($\Delta\lambda$)
= 10^{10} photons s^{-1} m^2 nm^{-1} x (13.6 eV per photon) (1.6 x 10^{-19} J per eV) (100 nm)
= 2.2 x 10^{-6} J s^{-1} m^{-2}

Now for the cooling rate we need to do a slightly nastier integral:

$$F_{Cool} = \int Q_{em}\, B(\lambda, T_g)\, d\lambda \qquad (24)$$

Here, we are EXTREMELY lucky because we happen to remember that the integral of the Planck function over all wavelengths is just the Stefan-Boltzmann law:

$$F = \sigma T^4 \qquad (25)$$

$F_{COOL} = Q\, \sigma\, T_g^{\ 4}$

$F_{COOL} = (0.01)(5.67 \times 10^{-8}$ W m^{-2} K^{-4}) $T_g^{\ 4}$

Setting $F_{COOL} = F_{HEAT}$ we can solve for the grain temperature:

$2.2 \times 10^{-6} = 5.67 \times 10^{-10} (T_g)^4$

$T_g = 8$ K.

This is a bit on the cool side (probably our typical UV photon energy was too low), but not a bad estimate!! This calculation is an example of a 'Rate Balancing' or Equilibrium problem. Here we are trying to find something out about the typical value of a quantity (say a temperature or abundance) that is subject to change according to various rates. In general, the goal is to set up a differential equation for the rate of change of this quantity with time (dX/dt) as a function of the rates of various processes that increase or decrease X, and set that derivative equal to zero:

$$\frac{dX}{dt} = R_1(X) + R_2(X) + ... + R_N(X) = 0 \qquad (26)$$

Note that rates which result in a decrease of X are generally written as negative quantities. As an example, we have several times described the heat exchange of a system in terms of cooling and heating rates. If we define the heat (per unit volume) of a system as Q(T) then we have solved relations like:

$$\frac{dQ(T)}{dT} = G(T) - \Lambda(T) = 0 \tag{27}$$

Thus to find an equilibrium temperature T_{EQ} we set **Equation 16** equal to **Equation 21**.
We did another rate balance calculation in **In-Class #6**, where we found the equilibrium chemical abundance of a molecule by collecting terms for the rates of its formation and destruction due to various processes.

8. Astrochemistry Part 2 (Dust)

As a reminder, chemical reactions in space:

- Are more difficult than reactions on earth (due to low temperatures and densities) although this also means that some species that don't last on earth b/c they are so highly reactive (e.g., radicals) can be present in some quantity in the ISM.
- Do not have an activation energy/ activation barrier. This is because most rate coefficients (k) have some exponential term exp(-E/kT): remember what happens to such a term if E >> kT!
- Favor reactions with large (Coulomb-aided) cross sections (e.g., Ion-molecule and dissociative recombination) because these tend to be faster

Quick example (since it is useful to look at problems from lots of different angles!) in addition to collisional rates and time scales, we can go back to our definition of a mean free path (**Equation 10**): instead of the typical *time* until a particle undergoes a collision, this is the typical *distance* a particle goes until encountering another particle. From the Herbst article (and apologies, here come the centimeters):

A typical collisional rate coefficient is $k = 10^{-10}$ cm^3 s^{-1}. (recall **Equation 22**).
For a typical cold (10 K) and dense (n= 10^4 cm^{-3}) cloud, we find:

- Collisional rate per particle of $k\,n = 10^{-6}$ s^{-1} (**Equation 13**)
- Timescale $t_{collision} = 1/kn = 10^6$ s or 2 weeks (**Equation 12**)
- Mean free path $\lambda_{MFP} = \tau v = 10^5$ km (**Equation 10**)

Note that in order to determine the mean free path we made an assumption that the velocity of the gas was equal to the thermal velocity:

$$v_{th} = \sqrt{\frac{8k_B T}{\bar{m}\pi}} \tag{28}$$

In spite of how hard this is, you can still make a lot of molecules in ISM gas. HOWEVER, saturated organic species require high temperatures and/or 3-body reactions, and so are not easily formed in the gas phase (same with H_2).

Dust grain surfaces play a key role in moderating ISM chemistry: both as a meeting place and a third body (and sometimes a shield against UV light, if reactions take place within and under ice layers).

The basic procedure to make molecules on dust grains is as follows

1. An atom (or molecule) hits a dust grain. This is helped by relatively large dust cross sections (though the number density of dust is fairly low).
2. It then has some probability of sticking (particles sit in small surface potential wells) and remaining on the grain (either due to electrostatic forces, or by forming a chemical bond with the surface).
3. Large nuclei tend to stay put in their potential well, and H atoms hop around, interacting with the particles in other wells.
4. Over time, (saturated) molecules then form on the grain surface: H_2, H_2O, NH_3, CH_4, CH_3OH, H_2CO, and other prebiotic organics like dimethyl ether (CH_3OCH_3) etc.
5. The dust grain lattice acts as a third body to stabilize formation of molecule, allows it to release excess energy of formation.

How do we then get things off of the dust grains?

1. If the dust grain is cold enough, the molecules will stick around, and will tend to form ices on the grain surface (Some common ice constituents are H2O, CH3OH, CO2, CH4, NH3)
2. If it is VERY cold, then molecules already in the gas phase will also adhere to the dust grains (this causes depletion of gas-phase species like CO)
3. Molecules can be driven (back) off of dust grains generally through two main processes: shocks and heating (and combinations of the two). Shock sputtering can even erode and destroy the dust grains themselves, releasing components like Si (molecules like SiO are commonly seen to be tracers of shocks).

9. Ionization and Recombination

To recap what we have covered so far: we have seen many different examples of chemical reactions. Examples of these types include:

1. Ion-molecule reaction	$H_2^+ + H_2$	$\rightarrow$	$H_3^+ + H$
2. Dissociative recombination	$H_3^+ + e^-$	$\rightarrow$	$H_2 + H$
3. Radiative association	$H + H$	$\rightarrow$	$H_2 + h\nu$
4. Photodissociation	$H_2 + h\nu$	$\rightarrow$	$H + H$
5. Collisional excitation	$H + H$	$\rightarrow$	$H^* + H$
6. De-excitation	H^*	$\rightarrow$	$H + h\nu$
7. Ionization	$H + h\nu$	$\rightarrow$	$p + e^-$
8. Recombination	$p + e^-$	$\rightarrow$	$H + h\nu$

Note that photodissociation is just the reverse of radiative association (**Reaction V**), and we could tweak our expression for de-excitation (**Reaction II**) to involve a collision so that it would then be the inverse of collisional excitation (**Reaction I**). Finally, we notice that recombination (**Reaction III**) is also the inverse of ionization (**Reaction IV**). We also define a new reaction here, dissociative recombination:

$$\mathbf{AB^+ + e^- \rightarrow A + B} \qquad \textbf{(VIII)}$$

In this case, it is a molecular ion that is recombining with an electron, and the energy of this process instead of being emitted as a photon goes into breaking a chemical bond. For now, we will focus on the processes of hydrogen ionization and recombination.

Recall that for a hydrogen atom, the ionization energy is 13.6 eV. This means that only energetic (shorter than UV) photons result in ionization.

Here then we are dealing with one main reaction:

$$H + h\nu \leftrightharpoons p + e^-$$

GUESS WHAT-- that means more rate equations!

Let's start with the recombination rate, and write an expression for the rate (per unit volume) at which protons and electrons combine to make hydrogen.

As we might expect, this will be of the form $Z = k\, n_a n_b$ (**Equation 17**)

Specifically,

$$Z_R = \beta_2(T_e)\, n_e n_p = \beta_2(T_e)\, n_e^{\,2} \qquad (29)$$

(In the textbook, this rate variable is written as $\dot{\mathcal{N}}_R$) Note that this simplification is possible because for a pure hydrogen gas your number of free electrons will always equal the number of free protons: $n_e = n_p$.

Here, $\beta_2(T_e)$is the ***recombination coefficient*** of hydrogen.

This is a function of the electron temperature, which is representative of the temperature of the H atoms and protons (H+) as well. A numerical approximation for this coefficient is:

$$\beta_2(T_e) = 2 \times 10^{-16}\, T_e^{-3/4}\, m^3 s^{-1} \qquad (30)$$

Note:

> We use $\beta_2(T_e)$because we are counting all of the recombinations down to n=2.
>
> This is because we run into a problem with recombinations directly down to n=1 because the photon that this reaction gives off is enough to immediately ionize (another) H atom, so the effect is net neutral-- or rather the net effect is to maintain ionization!!

Now, let's look at the ionization rate: H + $h\nu \leftrightharpoons$ p + e^-

What are the reactants on the left side of the equation?

Since one of the 'reactants' is a photon, this means we are dealing with a particle (volume) density times some flux (area) density. As a result, in order for $Z_I$to have the same units as Z_R, it has a coefficient α is then not a volumetric coefficient like β_2, but instead has SI units of m^2 and is what we will call the ***photoionization cross section*** of the atom:

$$Z_I = \alpha n_H J \tag{31}$$

(Note that in the textbook, this rate variable is written as $\dot{\mathcal{N}}_I$) Here we define J as the number flux of ionizing photons (SI units of m^{-2} s^{-1}) which is intercepted at some distance R from the star.

We can define J as

$$J = S_* / (4\pi R^2) \tag{32}$$

where $S_*$is the ***rate of ionizing photons*** from the star (with SI units of s^{-1}).

We are going to look at what happens when we balance the ionization and recombination rate (equilibrium!). Where does this apply? Well, you can imagine that around a star giving off ionizing radiation, close to the star you will have a LOT of energetic photons, and any H atoms will be constantly bombarded, and keep breaking back up any time a proton and an electron get a chance to recombine. In this region, you would be ionization-dominated. Meanwhile, far from the star, you are receiving many less photons: only occasionally does an H atom encounter a sufficiently-energetic photon, and it is much more likely that a free proton will encounter a free electron before it sees another photon. We are looking for the place in-between where these rates are exactly equal. It turns out that this is going to define the size of a so-called HII region: the ionized nebulae we see surrounding massive stars.

To define this balance we will do a little bit of algebraic gymnastics with our rates:

(1) Determine the **total** number of recombinations in a volume around the star
(2) Set this equal to the **total** number of ionizing photons

This condition is expressed with the following relation:

$$S_* = \beta_2(T_e) n_e^2 \frac{4}{3}\pi R^3 \tag{33}$$

This is often rewritten by defining an ionization fraction x:

$$n_e = x\,n \tag{34}$$

Here n is the number density of *all* hydrogen nuclei, including both H and H^+ (protons or p). I am going to tell you without justification (sorry!!!) that we can generally assume $x = 1$. In this case, $n_e = n_p = n$

Solving for R we come up with:

$$R_S = \left(\frac{3S_*}{4\pi n^2 \beta_2}\right)^{1/3} \tag{35}$$

Where this radius is known as the ***Stromgren radius***: the size of an HII region.

Let's plug in some numbers to see exactly how big a nebula we are talking about, in absolute terms. Using $S_* = 10^{49}$ and $x = 1$ and our old favorite $n = 10^8$ m^{-3} and assuming T_e =10000 K:

$$R_S = \left(\frac{3\times10^{49}}{4\pi(10^{16})(2\times10^{-16})(10{,}000)^{-0.75}}\right)^{1/3} = 10^{17} \text{ m or 3 pc!}$$

10. HII Region Cooling (Forbidden lines)

Let's talk about how we heat and cool HII regions! Heating is primarily through photoionization. Thus, we can do a DREADFUL thing and define a heating rate that is itself the function of another rate (the **photoionization rate**).

That is, the rate of energy input per unit volume (the **volumetric heating rate**) is

$$G_I = Z_I Q_I \tag{36}$$

Where Q_I is the average energy of a single photoionization. I am going to state without proof (SORRY) that

$$Q_I = k_B T_* \qquad (37)$$

where $T_* = T_{eff}$, the ***blackbody temperature*** of the star. (See the solution to **Problem 5.3** in the textbook for insight, as well as some help with the last problem in the Problem Set #2!)

Assuming ionization equilibrium:

$$G_I = Z_i k_B T_* = Z_r k_B T_* \qquad (38)$$

The cooling rate is similarly the rate of a rate:

$$L_R = Z_R Q_R \qquad (39)$$

Here Q_R is now the average energy REMOVED from the gas.

I will again state without proof that this is just

$$Q_R = k_B T_e \qquad (40)$$

This is a little simpler to intuit, as this describes a typical electron energy (modulo some likelihood that a given interaction with an electron will result in a recombination)

So, we can then say

$$L_R = Z_R k_B T_e \qquad (41)$$

And if we set $G_I = L_R$ to define an equilibrium electron temperature for the nebula, we would find $T_e = T_*$

Theory then predicts that the temperature of a pure-hydrogen HII region should match that of a typical O star: 30,000 - 60,000 K. However, we observe them to be much cooler! What's up with that?

Answer: our heating rate isn't wrong (we understand the energy spectra of stars pretty well) but our cooling rate *is* wrong: HII regions are not just hydrogen, and line cooling from other ions (He, O, N, etc.) is super-important.

This means if we want to find an equilibrium temperature, we need a different expression for the cooling rate. We are going to choose one based on O^+, which is a good example of line cooling from ions (this should remind us a lot of the C^+ cooling rate in **Equation 16** !!)

$$L_{O+} = 1.833 \times 10^{-30} y_{O+} n_O n_e T_e^{-1/2} e^{\frac{-3.89\times10^4 K}{T_e}} \qquad (42)$$

As before, this is a volumetric cooling rate with SI units J m^{-3} s^{-1}

A quick breakdown of this: y_{O+} is the ***ionization fraction of oxygen***, so that

$$n_{O+} = y_{O+} n_O \qquad (43)$$

This is similar to the ionization fraction x we defined in **Equation 34**, and similarly we can generally just assume $y_{O+} = 1$.

Note:

> We can also replace n_O with $n\left[\frac{n_O}{n}\right]$ where $\left[\frac{n_O}{n}\right]$ is the abundance of oxygen with respect to all hydrogen nuclei (both ionized and neutral). If you assume $n_e = n$ (this is from **Equation 34**; assuming $x = 1$) and adopt an oxygen abundance of $\left[\frac{n_O}{n}\right] = 1.6 \times 10^{-4}$ you get the form of this equation given in the textbook.

The rest of the numbers here are just writing out the temperature dependence of the rate coefficient k in order to match it to the form of **Equation 17**. We could say that

$$k_{O+} = 1.833 \times 10^{-30} T_e^{-1/2} e^{\frac{-3.89\times10^4 K}{T_e}} \qquad (44)$$

Now, to find an equilibrium temperature we need to set $L_{O+} = G_I$, where we now write a full expression for the heating rate (as originally defined in **Equation 36**) using the recombination rate given in **Equation 29** as:

$$G_I = 2 \times 10^{-16} T_e^{-3/4} n_e^2 k_B T_* \qquad (45)$$

We can again assume $n = n_e$ but it still turns out this equivalency needs to be solved numerically for T_e. **Table 5.2** in the book gives some typical values of T_e for various values of T_*.

11. Equations of State (Adiabatic/Isothermal)

Let's talk about pressure!
What are some types (or sources) of pressure in the ISM?

- Radiation pressure
- Thermal gas pressure
- Turbulent gas pressure
- (Degeneracy pressure is another astronomically-relevant pressure, but is not important in the ISM!)

What is an equation of state? Functionally, this is a thermodynamic relationship between quantities in a gas, including temperature, density, and pressure. An example of this that you may have encountered before is the ideal gas law:

$$P = n\,k\,T \qquad (46)$$

This is also sometimes written in astronomy as

$$PV = N\,kT \qquad (47)$$

Where N is the total number of particles in a volume V - our definition of a number density from **Equation 6**.

It is probably useful to be a little more explicit here about how we define *n*, and modify slightly our original expression for it given in **Equation 8**:

$$n = \rho \,/\, \mu m_H \qquad (48)$$

Here m_H is the mass of a hydrogen atom, and we have just introduced a new way to express the average mass of a particle $\overline{m}$ (**Equation 7**) using μ: a quantity called the ***mean molecular weight***.

$$\mu = \overline{m}/m_H \qquad (49)$$

μ is a unitless quantity describing the typical (or number-weighted) mass of a particle. This is important because, as we have found, we often are dealing with inhomogeneous mixtures.

Does the ideal gas law apply for the ISM? The assumptions are:

(1) The gas consists of a large number of molecules that are in random motion and obey Newton's laws of motion. So far, so good!
(2) The particles are far enough apart (so that the volume of the individual particles is much smaller than the volume of the gas). A quick check of this assumption: For a density of 10^8 hydrogen atoms per cubic meter, assuming an effective radius for a hydrogen atom to be 10^{-10} m, the total volume of the atoms would be $10^8\,(\frac{4}{3})\,\pi\,(10^{-10})^3 = 4 \times 10^{-22}$ m^3
(3) The particles are not TOO far apart (so that they undergo many collisions before crossing a region, and the velocity distribution can be described by kinetic temperature). Checking this: we saw λ_{MFP} was large (10^5 km) but this isn't actually so large compared to typical scales (1 pc ~10^{13} km).
(4) All collisions are perfectly elastic. Is this true? Nope! Since molecules have internal energy states, some of the collision energy can be absorbed. Also, chemical reactions occur.
(5) Apart from brief elastic collisions, there are no intramolecular forces. This is also not true! As we have discussed, many of the particles are either charged or even if neutral, have a dipole moment.

So, can we still use the ideal gas law? It turns out: yes! It is good enough: the extreme low density combined moderate temperatures keeps particles far enough apart and interactions sufficiently brief that the intermolecular forces aren't too important.

Two special cases of interest for our ideal gas:

Isothermal, describing changes in a system occurring at a uniform temperature:

$$PV = Constant \qquad (50)$$

Adiabatic, describing changes in a system occurring at a constant entropy with no exchange of heat with its environment or surrounding:

$$PV^{\gamma} = \ Constant \tag{51}$$

While not important for the ISM, there are two other astrophysically-important equations of state you may have seen before: for degenerate gas, both nonrelativistic (in white dwarf stars), and relativistic (in neutron stars).

12. Adiabatic and Isothermal Sound Speed

In the previous lecture, we introduced the equation of state for an ideal gas (**Equations 46 and 47**) and defined two special cases of processes that an ideal gas can undergo: isothermal changes (**Equation 50**) and adiabatic changes (**Equation 51**).

We can rewrite the latter as the following proportionality:

$$P \propto \rho^{\gamma} \tag{52}$$

The quantity γ is known as the ***adiabatic index***, and it has two values that will be of interest to us in this class:

(1) $\gamma = 1$. Combining this with the ideal gas equation, we can see that this requires T to be a constant, leading to the relation PV = constant. This is then the same as **Equation 50** and describes an isothermal process in the gas.

(2) $\gamma = 5/3$. This describes an adiabatic process for a non-relativistic, ideal gas.

Because we love rates in this class, one thing we might want to know is how fast changes happen to the variables in our equation of state.

We actually already know how to describe how the temperature changes: this is the heating (or cooling) rate. Further, if the heating and cooling are able to reach equilibrium ($\Gamma = \Lambda$ or $G = L$) then this equivalency defines an equilibrium temperature of the gas. This is a situation in which the gas is then isothermal.

We would like to be able to compare the temperature change with the speed (and or timescale) on which the pressure changes. Essentially, how fast does a pressure wave move through the gas? This ought to sound familiar: what we are looking for here is the speed of sound!

We can do a light derivation of the speed of sound as follows: Assume that we have a gas at rest, with a constant pressure P_0 and a constant density ρ_0. We then make a small change (a perturbation) in the density and pressure so that:

$$P = P_0 + dP \quad \text{(a)}$$

$$\rho = \rho_0 + d\rho \quad \text{(b)}$$

We are going to do a linear analysis, which means that we assume $dP$and $d\rho$ are so small that we can ignore any terms of dP^2, $d\rho^2$, or higher order as being essentially zero. We then define these perturbations using the relationship of **Equation 52** by taking the derivative of the pressure with respect to density, which gives us the following relationship between the differential quantities $dP$and $d\rho$:

$$dP = \gamma\rho^{\gamma-1}d\rho \quad \text{(c)}$$

Using **Equation b** to substitute for ρ:

$$dP = \gamma(\rho_0 + d\rho)^{\gamma-1}d\rho \quad \text{(d)}$$

Formally, we need to expand the term $(\rho_0 + d\rho)^{\gamma-1}$using the binomial theorem:

$$(\rho_0 + d\rho)^{\gamma-1} = \rho_0^{\gamma-1} + O(d\rho) + O(d\rho^2) + \ldots \quad \text{(e)}$$

Where $O(d\rho^n)$ indicates a term proportional to $d\rho$ raised to the power n.

Luckily, since we are multiplying by $d\rho$ and we get to ignore terms of $O(d\rho^2)$ we only keep the first term of this expansion, yielding:

$$dP = \gamma(\rho_0)^{\gamma-1}d\rho \quad \text{(f)}$$

We then do a little arithmetic trick and say that $x^{n-1} = x^n/x$:

$$dP = \gamma(\rho_0^{\ \gamma} / \rho_0)\, d\rho \qquad \text{(g)}$$

Using **Equation 52** to substitute for $\rho_0^{\ \gamma}$:

$$dP = \gamma(P_0 \,/\, \rho_0)\, d\rho \qquad \text{(h)}$$

Here, the quantity $\gamma(P_0 \,/\, \rho_0)$ happens to have units of m^2/s^2

We can then generally define the speed of sound as:

$$c_s = \sqrt{\frac{\gamma P}{\rho}} \qquad (53)$$

(Note, this is really half a derivation-- for a full derivation you would plug these perturbations into a set of equations for fluid flow, which would yield a differential equation of a form consistent with wave propagation, from which you would see this term defined as the wave speed.)

This is the speed at which changes in pressure propagate through a gas. Recall that for an isothermal gas, $\gamma = 1$ so we can define a special case, the ***isothermal sound speed***, as

$$c_s = \sqrt{\frac{P}{\rho}} \qquad (54)$$

Then the adiabatic sound speed in an ideal non-relativistic gas is just 1.3 times the isothermal speed. We can also rewrite this using the ideal gas equation (**Equation 46**) and our definition of number density (**Equation 8**):

$$c_s = \sqrt{\frac{\gamma k_B T}{\overline{\overline{m}}}} \qquad (55)$$

For an isothermal gas the sound speed is then the same as the root-mean-square of the 1D thermal velocity (very similar to **Equation 28**):

$$v_{thermal(1D)} = \sqrt{\frac{k_B T}{\overline{m}}} \quad (56)$$

Using the sound speed, we can define two useful quantities. One is the ***Mach number***:

$$M = v/c_s \quad (57)$$

This is a unitless quantity that describes how fast the gas in a flow is moving as a ratio with the speed of sound in that gas.

The other is the ***sound crossing time***:

$$t_{SC} = L/c_s \quad (58)$$

(note that this follows the general form for timescales given in **Equation 9**!)

This is the time it takes a sound wave to traverse a region. This is the typical or characteristic timescale on which changes in pressure can occur in a system (or on which a system can respond to some outside stimulus by adjusting the pressure).

To get an idea of whether a system is adiabatic or isothermal, you can compare the cooling time to the sound crossing time.

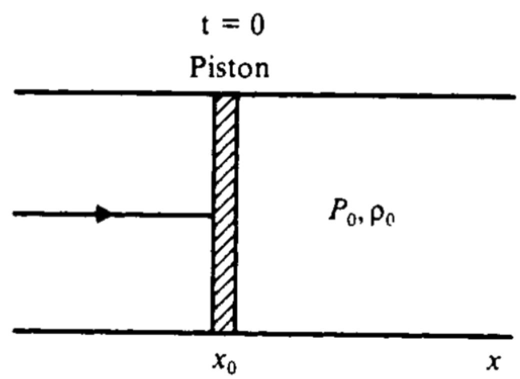
For an adiabatic process, $t_{SC} < t_{cooling}$. This means that either changes to the system are happening fast, or that it just takes a long time to cool. In fact, inefficient cooling is one of the most common reasons you see an adiabatic system. This occurs when radiation cannot escape the source easily, for example when your gas becomes optically thick. The insides of stars are a great example of this!

For an isothermal process $t_{SC} > t_{cooling}$. Now the time scale for pressure adjustment is much longer than the time it takes to cool. This is representative of a lot of what we have seen so far in the ISM, including HII regions and molecular clouds.

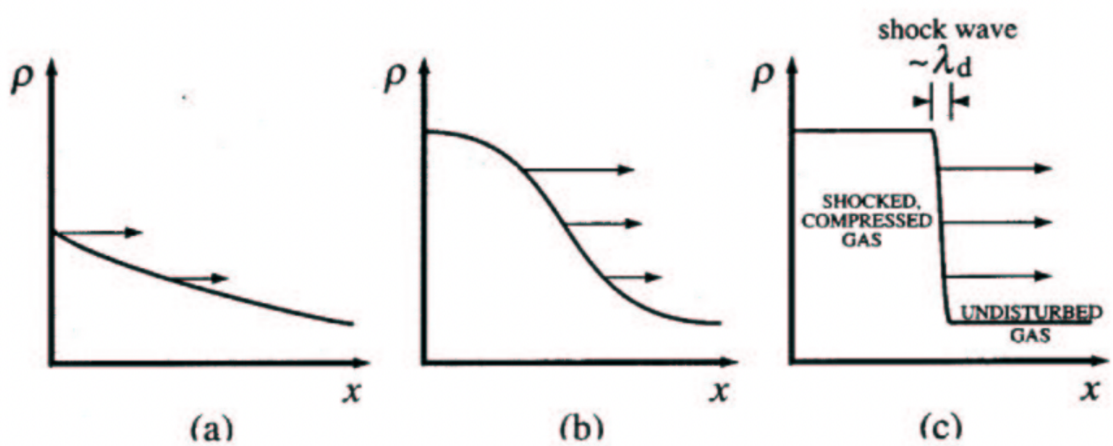
13. Adiabatic Shocks

What is a shock? Like a piston in a car motor, this is a rapid (and so adiabatic) change in the pressure: specifically, a compression wave due to some initially supersonic motion.

Imagine that our gas is in a tube, and we push the piston to the right at a supersonic speed, into stationary gas with an initial pressure P_0 and mass density ρ_0.



Let's imagine that we do this incrementally. Our first push results in an increase of the initial pressure P_0. This pressure wave propagates to the right at the (adiabatic) speed of sound c_s, and results in the compression (increased density) of the gas it moves through. We then push the piston again, and it again leads to a wave of higher pressure, but the gas it is pushing into has already been compressed, and is at higher pressure and density.



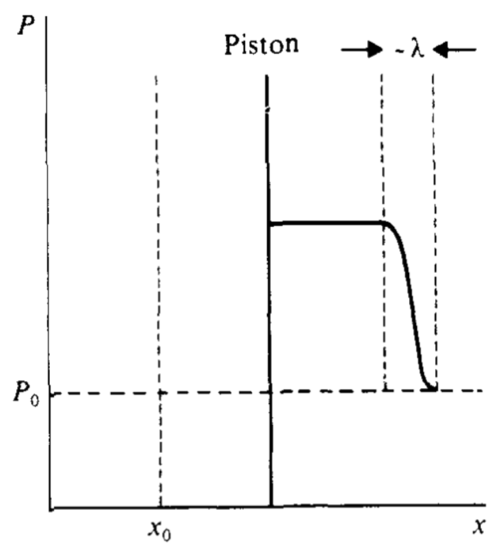
How does this change the speed of sound, since it is related to the ratio of the pressure and density? Well, we can use the adiabatic equation of state (**Equation 50**) and our derivation of the speed of sound (**Equations c and h**) to show that the speed of sound increases as:

$$c_s^{\ 2} \propto \frac{dP}{d\rho} \propto \gamma\rho^{\gamma-1} \qquad (59)$$

For an adiabatic index of $\gamma = 5/3$, we can say

$$c_s^{\ 2} \propto \rho^{2/3} \ \ or \ \ c_s \propto \rho^{1/3} \qquad (60)$$

This means that the speed of sound in this gas is higher than it was, and so this pressure wave moves faster than the one in front of it and starts catching up with it, increasing the magnitude of the pressure change. This keeps happening, and results in a pile-up effect as each of these subsequent pressure waves tries to overtake the first. All of these incremental pressure changes then essentially combine into one drastic pressure change: a shock wave.



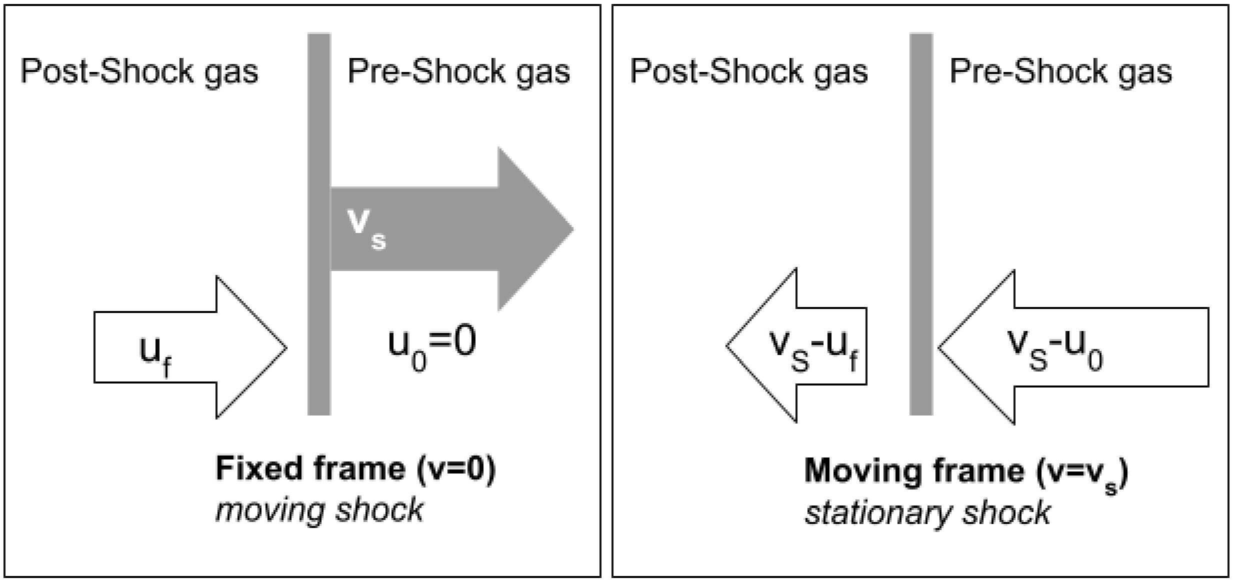
The discontinuity between the initial (pre-shock) and final (post-shock) pressure has a typical size-scale over which the pressure jump occurs that is characteristic of the mean free path in the gas (**Equation 10**). In this area, typical relations used to describe the gas flow break down, and the microscopic processes we have talked about in this class (collisions which increase the viscosity, and energy loss through ionization and even heat transfer through conduction) act to keep the shock from becoming infinitely steep.

An important consideration as we define the variables for our equations describing shocks will be the frame of reference we choose. There are two main ways to look at the shock:

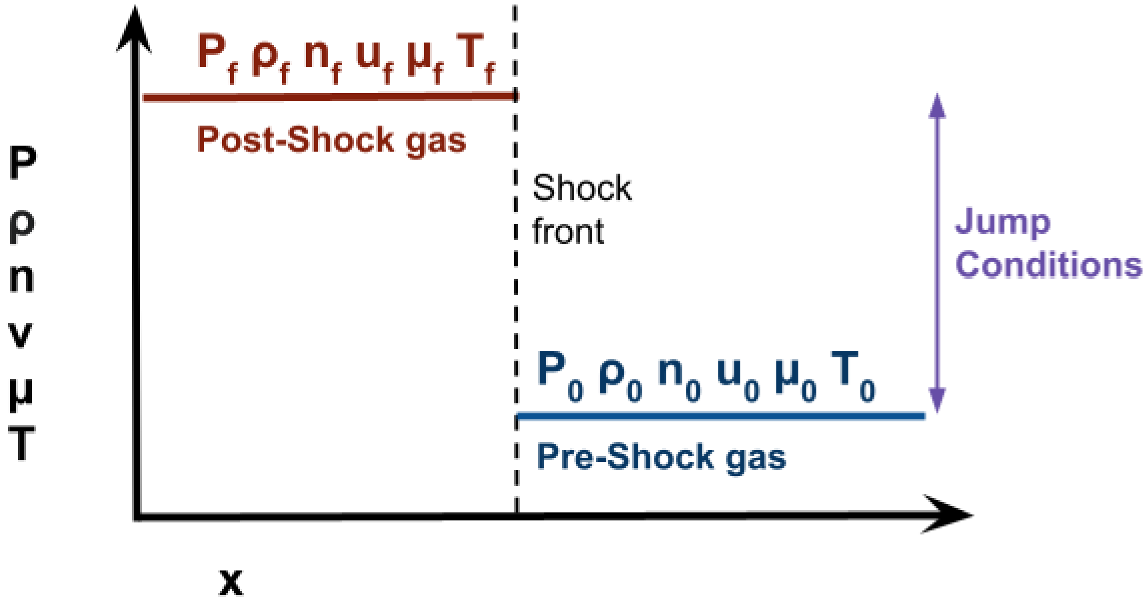
(1) A fixed frame of reference, in which the shock is moving at a speed v_s (for example, propagating outward from an explosion or from a massive star driving a

fast wind)

(2) A frame of reference moving at a speed v_S in which the shock appears to be stationary.



We then want to describe the effects of a shock in terms of the conditions on either side of the shock front: the initial (upstream, or pre-shock) and final (downstream, or post-shock) conditions. The sharp changes in the gas properties – including the pressure (P), mass density (ρ), number density (n), bulk or flow velocity (u), mean molecular weight (μ; **Equation 49**) and temperature (T)– are known as the **jump conditions**.



Although many gas variables have discontinuous values across the shock, there are a number of quantities that must stay constant or conserved: mass (the **mass flux**, or mass per area:

$$\rho_0 v_s = \rho_f (v_s - u_f) \tag{61}$$

momentum (the change in momentum between gas entering the shock front and exiting the shock front must be equal to the net force per unit area – pressure!-- acting on the material as it crosses the shock front):

$$\rho_0 v_s^{\,2} - \rho_f (v_s - u_f)^2 = P_f - P_0 \tag{62}$$

and energy (the difference between the total energy (kinetic and internal; see **Equation 19**) of gas entering and exiting the shock front per unit area per unit time must be equal to the rate at which **work** is done by the pressure difference across the front):

$$\left[\frac{1}{2}\rho_f (v_s - u_f)^3 + \rho\epsilon (v_s - u_f)\right] - \left[\frac{1}{2}\rho_0 v_s^{\,3} + \rho_0 \epsilon v_s\right] =$$

$$P_0 v_s - P_f (v_s - u_f) \tag{63}$$

These can be used to define the jump conditions.

For the sake of time I will spare you a LOT of derivations (see **Section 6.3** for all of the messy details) and define the following jump conditions in the limit of a strong adiabatic shock in which $M >> 1$ (see **Equation 57**) For convenience, we will define these in terms of v_S or the shock velocity in the fixed frame of reference.

The bulk speed of the gas in the *fixed frame* after the shock has passed through is

$$u_f = \frac{3}{4}v_S \tag{64}$$

The relation between the initial and final mass density is

$$\rho_f = 4\rho_i \tag{65}$$

Note:

> To derive the relationship between the initial and final number density from the mass density relationship (**Equation 8**), you will need to know the average particle mass $\bar{m}$. The relationship between the mean molecular weight μ and the average particle mass is given by **Equation 49**

The pressure behind the shock is

$$P_f = \frac{3}{4}\rho_i v_S^2 \tag{66}$$

We can use **Equations 65 and 66** and the ideal gas law (**Equation 46**) to further describe the temperature behind the shock:

$$T_f = \frac{3}{16}\frac{\mu\bar{m}}{k_B}v_S^2 \tag{67}$$

Note that shocks with a speed of at least 100 km/s will generally fully ionize the gas!

14. The input of energy into the ISM

Shocks will arise in a number of cases where energy from stars is input into the ISM. This includes HII region expansion, stellar winds, and supernovae.

We are going to talk about these MUCH MORE GENERALLY than the book does (see all of **Chapter 7** for the gory, gory details).

Qualitatively, HII regions are expected to go through 2-3 stages of evolution.

In the first stage (an ***ionization front*** expanding to an equilibrium) the newly-formed O star outputs a constant flux of ionizing photons (S_*; see the solutions to #3 in **Problem Set 2**), and begins to ionize the surrounding gas. Assuming that the surrounding gas has a uniform density then during this process we can expect that the *mass density* (ρ) of the gas remains uniform, and is the same inside and outside of the nebula. Eventually the star reaches an equilibrium of ionization and recombination. This occurs on a timescale that is approximately the time at which the recombination process can 'respond' to the ionization of the gas. Unsurprisingly, this is known as the recombination time and we can derive it using our general form of a timescale from **Equation 9**:

$$t_R = n_e/Z_R = \frac{1}{\beta_2(T_e)n_e} \qquad (68)$$

While the mass density initially remains constant, the NUMBER density is in fact changing significantly: the number of particles behind the ionization front has doubled. Not only that, but these particles have absorbed a lot of UV energy and they are moving WAY faster: as we noted, the temperature of the ionized gas is characteristically about 10^4 K, compared to 10^2 K in the surrounding neutral gas. We can then use the ideal gas law (**Equation 44**) to compare the pressure of the ionized gas to the pressure of the surrounding neutral gas:

$$\frac{P_{ion}}{P_n} = \frac{n_{ion}T_{ion}}{n_n T_n} = \frac{2n_n(10^4)}{n_n(10^2)} = 200$$

This then triggers the second phase: a pressure-dominated expansion, which drives a shock into the surrounding medium. The speed of the shock is equal to the speed of sound in the ionized medium (c_{ion}), and this is strongly supersonic compared to the sound speed in the neutral medium (c_n). Using **Equations 53 and 56** we can show (assuming that initially $\rho_i = \rho_n$) :

$$M = c_{ion}/c_n = \sqrt{\frac{P_{ion}}{P_n}} = 14$$

The third phase, which most massive stars don't live long enough to see (unless they were born in a really dense cloud) is pressure equilibrium, where the interior pressure equilibrates to match the exterior pressure. As we might expect, the timescale for pressure equilibrium is proportional to sound crossing time (**Equation 57**), and is actually:

$$t_{EQ} \sim 300 \frac{R_S}{c_i} \tag{69}$$

For a density of $n = 10^8$ we found a typical R_S of 10^{17}m. For $T = 10^4$ K we can also use **Equation 54** to calculate that

$$c_{ion} = \sqrt{\frac{(1.38\times10^{-23})(10^4)}{0.5(1.67\times10^{-27})}} = 10^4 \text{ m/s}$$

Then $t_{EQ} \sim 10^8$years, much longer than the lifetime of an O star.

If HII regions aren't reaching pressure equilibrium, what is the O star doing?
A: dumping a lot of energy into the ISM!

The winds from massive stars are fast (2000 km/s) and drive strong shocks. Assuming a shock velocity of 2000 km/s, we can calculate a post-shock temperature using Equation 63 :

$$T_f = \frac{3}{16}\frac{(0.5)(1.67\times10^{-27})}{(1.38\times10^{-23})}(2 \times 10^6)^2 = 7.5 \times 10^8 \text{ K}$$

Fast shocks like those from stellar winds and supernovae are then the origin of the coronal gas (the hot ionized medium) in the ISM, and are strong sources of (thermal) X-ray emission:

$$E = kT$$
$$= (1.38 \times 10^{-23})(7.5 \times 10^{8})(6.2 \times 10^{18}) = 6.2 \text{ keV}$$

15. The Virial Theorem

Molecular clouds are an extremely important part of the ISM because it is here that new stars will form. The physics of star formation is in large part a struggle between two opposing forces: the inward-acting force of gravity from the assembled mass of all of the molecules in the cloud, and the outward-acting (thermal) pressure from the motion of these same molecules. (While other forces also come into play, including centripetal forces from rotation, magnetic pressure, and surface pressure, we will focus on just these two for now!)

To describe the action of these forces on a spherical object (and the distribution of kinetic and potential energies resulting from their balance), we will first define the problem by thinking of a sphere as a series of shells. Each shell will have an infinitesimal thickness dr and a mass dM(r):

$$dM(r) = 4\pi r^2 \rho(r)\, dr \qquad (70)$$

Note that this definition also lets us describe the object's mass as a function of radius. Now you can get DANGEROUSLY fancy with your density structure, but right now we will just consider a constant-density case. Then:

$$M(r) = \int_0^r 4\pi r^2 \rho\, dr = \frac{4}{3}\pi r^3 \rho \qquad (71)$$

So, now we want to use this to describe the forces that are felt on a shell of gas in the object at a radius of r.

The gravitational force will be:

$$F_g(r) = -\frac{GM(r)dM(r)}{r^2} \quad (72)$$

Remembering that $P = F/A$ we will define the pressure force over the whole shell as:

$$F_P(r) = 4\pi r^2\, dP(r) \quad (73)$$

To define an equilibrium, we will then set both of these equal to each other:

$$4\pi r^2\, dP(r) = -\frac{GM(r)dM(r)}{r^2} \quad (74)$$

Now, we are going to do some slight shenanigans, because while this balance of forces defines the equilibrium, what we really want to know is how the ENERGY of the system is distributed into kinetic energy and potential energy when the forces are balanced. This is the end goal.

Reaching back again into the depths of our physics knowledge, we can define the energy as work: $E = W = F\,r$. Multiplying both sides by r:

$$4\pi r^3\, dP(r) = -\frac{GM(r)dM(r)}{r}$$

We then notice that the left hand side reminds us a lot of an expression for volume, so we will substitute variables and define the volume at a radius r as V(r):

$$3\,V(r)\,dP(r) = -\frac{GM(r)dM(r)}{r} \quad (75)$$

Now we will integrate both sides from the center to the edge of the object to determine the total amount of each type of energy in the system. We will assume the pressure at the surface is zero and the pressure at the center is the central pressure (P_c). We will also assume that the enclosed mass at the center is 0, and at the surface is M.

$$\int_{P_c}^{0} 3V(r)\, dP(r) \;=\; -\int_{0}^{M} \frac{GM(r)dM(r)}{r} \qquad (76)$$

For the left side of **Equation 76** we will integrate by parts:

$$\int_{P_c}^{0} 3V(r)\, dP(r) \;=\;$$

$$3\,V(r)\;\, P(r) \;-\; 3\int_{0}^{V} P(r)dV(r) \qquad (77)$$

The first term becomes zero (because the volume at r=0 is 0 and we are ignoring external pressure to say that the pressure at r=R is 0). For the second term, we will do a little more trickery and recall that the thermal (kinetic) energy of a single particle, free to move in 3 dimensions, is

$$E \;=\; \frac{3}{2}kT \qquad (78)$$

Note again that this is different from the thermal energy of a *photon*, which is $E = kT$ (**Equation 1**). The energy per unit volume is then given by **Equation 19** and is $\epsilon = \frac{3}{2}nkT$.

We recall from the ideal gas law (**Equation 46**) that nkT is just the pressure so we can rewrite the second term as

$$-\;3\int_{0}^{V} \frac{2}{3}\varepsilon\, dV(r) \;=\; 2E \qquad (79)$$

And this is just twice the total (thermal) **kinetic energy** E. Note that we can also define the total thermal kinetic energy as

$$E = \frac{3}{2}Mv^2 \tag{80}$$

For the right-hand side of **Equation 76**, this is just equal to the **gravitational potential energy** U.

$$U = -\int_0^M \frac{GM(r)dM(r)}{r} \tag{81}$$

For a uniform sphere, you can integrate this to find:

$$\begin{aligned} U &= \int_0^R \frac{-G\, 4\pi r^3 \rho (4\pi r^2\, \rho)\, dr}{3r} \\ &= \int_0^R \frac{-G\, 16\pi^2 \rho^2\, r^4\, dr}{3} \\ &= \frac{-G\, 16\pi^2 \rho^2}{3} \int_0^R r^4 dr \\ &= \frac{-G\, 16\pi^2 M^2}{3(4/3)^2 \pi^2 R^5}\, (\frac{1}{5}R^5) \end{aligned}$$

$$U = -\frac{3}{5}\frac{GM^2}{R} \tag{82}$$

Putting these results back together into Equation 77 we get:

$$U = -2E \tag{83}$$

This result is called the **Virial Theorem** and we can do cool things with it!

16. Scales of interest for Star Formation

On Wednesday we derived the Virial Theorem (**Equation 83**) which describes the distribution of energies in a system that is in hydrostatic equilibrium (**Equation 74**).

Now we want to ask what happens when a cloud is NOT in equilibrium: on what timescales will it change? We want then to define two characteristic time scales: a timescale for changes due to gravity, and a timescale for changes due to pressure. We actually already have a timescale for changes due to pressure: the sound-crossing time! (**Equation 58**). We are going to derive a timescale on which changes occur due to gravity. This is basically the time it would take (if you turned off pressure in the cloud) for the cloud to completely collapse under its own gravity.

The textbook gives a derivation of the free-fall time and it is, unsurprisingly, complicated as all heck. I am going to present an order-of-magnitude derivation, and then give you the exact form as well.

Consider the situation in which we have a cloud in which the magnitude of the gravitational potential energy is slightly more than twice the kinetic energy. We are going to focus on a tiny particle of gas at the very edge of this cloud. Like all good timescales, we are going to consider the representative (but actually unrealistic) scenario in which we take the initial value of a quantity (here the radius of the cloud) and ask how long it would take for it to become zero (no, we are not making a black hole here).

Now, unlike some timescales we have derived, we know that just taking the radius of the cloud and dividing by some velocity (using our typical timescale form given in **Equation 9**) is not going to work here. This is because the collapse is being driven by gravity, and gravity doesn't produce a constant velocity: it causes a constant acceleration. So, we have to get a little bit fancier, and cast our memories back to some equations of projectile motion. We will start with

$$d = v_0 t + \frac{1}{2}at^2 \tag{84}$$

In this case we know that the distance we want to travel is the total cloud radius R, we will assume that the initial velocity is 0, the acceleration in question is the gravitational acceleration, and the time is the free-fall time we are trying to define. Then:

$$R = \frac{1}{2} g t_{ff}^{2} \tag{85}$$

We need an equation for the gravitational acceleration, and this is just

$$g = \frac{GM}{R^{2}} \tag{86}$$

Putting these together:

$$R \approx \frac{1}{2} \frac{GM}{R^{2}} t_{ff}^{2}$$

$$t_{ff} \approx \sqrt{\frac{2R^{3}}{GM}}$$

Finally, substituting $\rho = M / \left(\frac{4}{3}\pi R^{3}\right)$:

$$t_{ff} \approx \sqrt{\frac{8\pi}{3G\rho}} \propto \frac{1}{\sqrt{G\rho}} \tag{87}$$

The reason that this is not precise and there is a slightly different constant is that as the cloud collapses, the radius (and density) are changing (though the mass enclosed remains constant), which introduces a slight perturbation on this more simple result. The precise expression for the free-fall time is

$$t_{ff} = \sqrt{\frac{3\pi}{32\, G\rho}} \tag{88}$$

Note that if a system is in hydrostatic equilibrium, then these two time scales (the free-fall time and the sound-crossing time) should be approximately equal: gravity has time to respond to any pressure changes and vice versa.

Finally, we want to ask-- what are the conditions that would trigger a cloud to collapse? Here we are looking for a typical mass at which a cloud (or part of a cloud) becomes unstable.

We start with the virial theorem (**Equation 83**), and the gravitational potential energy for a uniform-density cloud (**Equation 82**):

$$3NkT = \frac{3}{5}\frac{GM^2}{R} \tag{89}$$

Then, we are going to make two substitutions, first for the total number of particles:

$$N = \frac{M}{\overline{m}} \tag{90}$$

And then for the radius, as a function of mass and density:

$$R = \left(\frac{3M}{4\pi\rho}\right)^{1/3} \tag{91}$$

Then we can write:

$$\frac{3MkT}{\overline{m}} = \frac{3}{5}\frac{(4\pi\rho)^{1/3}GM^6}{3M^{1/3}} \tag{92}$$

Simplifying this expression to cancel out terms in common:

$$\frac{kT}{\overline{m}} = \frac{1}{5}GM^{2/3}\left(\frac{4\pi\rho}{3}\right)^{1/3} \tag{93}$$

We can solve this for M, yielding the **Jeans Mass**: the mass at which a given fragment of gas at this density would start to collapse under its own gravity.

$$M_J = \left(\frac{5kT}{G\bar{m}}\right)^{3/2}\left(\frac{3}{4\pi\rho}\right)^{1/3} \qquad (94)$$

We can scale this to some typical molecular cloud conditions to express the Jeans mass in terms of solar masses:

$$M_J = 2.3\, M_\odot \left(\frac{T}{10\, K}\right)^{3/2}\left(\frac{n}{10^{11}\, m^{-3}}\right)^{-1/2} \qquad (95)$$

17. The Physics of Star Formation

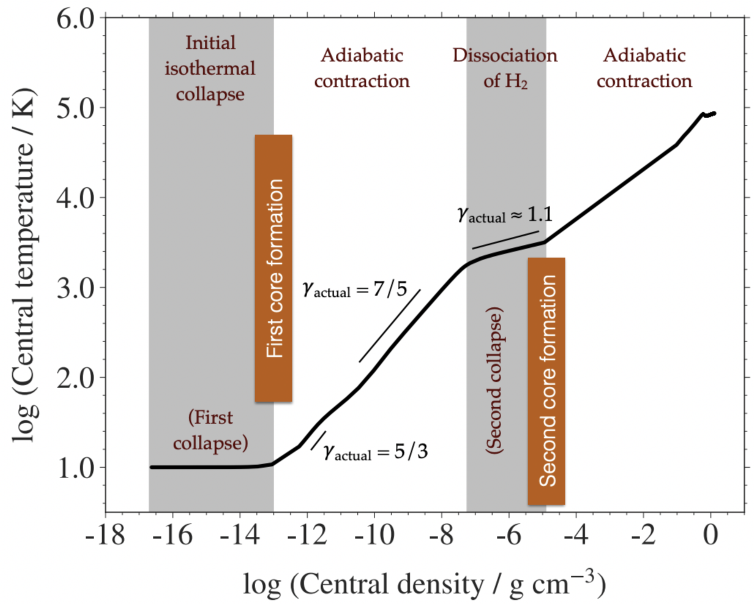
Star Formation! We talked about some of the physics involved already (with the Virial theorem and the free-fall time scale), so now let's get into the real gory details!

How do we change a molecular cloud into a star? Specifically, how will the size, density, and temperature change?

→ It is going to get a lot smaller!

→ In doing so, it will also get much more dense. How much more? Well, a bunch. Remember that a number density of $10^8\, m^{-3}$ means that the typical MASS density of a molecular cloud is $10^{-19}\, kg\, m^{-3}$-- less than a quintillionth of a kg of mass in each cubic meter. Now, the air you are breathing right now has a density of about a kilogram per cubic meter. The average density of the sun is 1400 kilograms per cubic meter.

→ Ultimately, it will get much hotter as the potential energy is turned into kinetic energy (it's like dropping a penny from a skyscraper: the further it falls, the more of the potential energy is converted into kinetic energy). And in this case, our penny is going to fall almost a trillion miles, to get from the typical size of a prestellar cloud fragment (a few thousand AU) to the size of something like the sun (700 million meters)



Stage 1: First collapse

$T_{central} = 10\ K$	$n_{central} = 10^{12}\ m^{-3}$	$R = 4 \times 10^{16}\ m\ (3000\ AU)$

During the first collapse, we are dealing with a fragment of a cloud that has built up a mass greater than the **Jeans mass**. The inward gravitational force is overwhelming the outward pressure produced by the motion of the gas particles-- this fragment is no longer in **virial equilibrium** and it collapses into a structure we will call a core. As gravitational energy is turned into kinetic energy, the core tries to heat up: but it is actually really bad at it (or rather, the dust grains are really good at radiating away their heat). Because the core can't heat up, the pressure force stays much smaller than the gravitational force, and the collapse proceeds virtually unhindered. Looking at this another way: with a short cooling time compared to a long free fall time, the collapse is isothermal: and isothermal means that the gas temperature basically stays constant. For a 1 solar mass core this stage lasts ~10^4 years (approximately the free-fall time) Everything changes, however, once it hits a number density of about $10^{16}\ m^{-3}$ which leads us to

Stage 2: The first hydrostatic core

$T_{central} = 10\ K$	$n_{central} = 10^{16}\ m^{-3}$	$R = 10^{12}\ m\ (10\ AU)$

At this point, the dust particles are packed so closely together that the photons they emit start to be blocked (re-absorbed) by other dust grains. So, the free fall time hasn't really changed, but now the cooling time goes to hell: it is going to take those photons that the dust grains are busy radiating FOR. EVER. to get out of the cloud. The heat is now trapped, and the temperature of the cloud starts to go up: the isothermal collapse phase ends, and the pressure finally has a chance to build up and start pushing back against the inward gravitational pull. This leads to a near-balance between the two: a quasistatic equlibrium called the first hydrostatic core. In reality, this is a phase of slow adiabatic contraction. Gravity wins (because the surface gas still slowly cools as heat radiates away from the core) but it is winning much more slowly. This impasse continues for about 1000 years until the core reaches a temperature of 2000 K, and then suddenly things change.

Stage 3: The second collapse

$T_{central} = 2000\ K$	$n_{central} = 10^{22}\ m^{-3}$	$R = 10^{12}\ m\ (4\ AU)$

Why is this collapse happening? There is suddenly a new energy sink INSIDE of the star: instead of just radiating energy out of its surface, suddenly there is a place for the energy to go inside of the star itself. This is because a temperature of 2000 K is enough for a typical collision to destroy (dissociate) the molecular hydrogen atoms. All of the collapse energy then goes into this destruction process instead of heating up the star: the star then once again becomes nearly isothermal and its radius rapidly decreases. About how long do we expect this collapse to take? We can compare the free-fall time for the protostellar core at the start of the second collapse to the initial free-fall time of the core before its first collapse. The free-fall time is proportional to $1/\sqrt{\rho}$. The density has changed by ten orders of magnitude, so t_{ff} is $\sqrt{10^{10}} = 10^{5}$ times shorter.

Eventually the star runs out of molecular hydrogen, and the collapse halts to reach

Stage 4: The second hydrostatic core

$T_{central} = 2000\,K$	$n = 10^{24}\,m^{-3}$	$R = 3 \times 10^{9}\,m\,(4\,R_{\odot})$

Finally, we have a true protostar! At these temperatures the dust has also long ago started to be destroyed (at about 1000 K) so the protostar is somewhat less opaque and finds it a bit easier to get rid of its energy, but it still takes time, during which the protostar continues to contract. The contraction slows drastically once it reaches a temperature of about 3000 K, when the atomic hydrogen begins to be ionized. Free electrons are EXCELLENT at scattering photons, so the opacity again rises quickly, and the protostar stabilizes a little more. At this point we should mention that all of these processes aren't happening to all regions of the protostar at the same time. The protostar begins to have an onion skin structure, where the innermost regions are ionized, surrounded by a dust-free zone of atomic and then molecular hydrogen, which continues to collapse toward the central core. As you might imagine, when this still collapsing material reaches a region that is contracting more slowly, you set up a shock. This accretion shock becomes the dominant source of luminosity of the star, and as this is a stochastic and variable process, the protostar luminosity becomes quite variable during this phase.

What does this whole process look like to astronomers? Well, it takes millions of years, so as usual we can't just watch this start to finish, instead we have to try to observe lots of different forming stars, and piece together the different stages of their evolution.

One tool that astronomers use for this is a Hertzsprung-Russell diagram (commonly abbreviated as just HR diagram). This is a great tool for plotting the properties of stars, and we will talk about it more next time, as it is going to be our guide to following and understanding the evolution of stars.

18. The Physics of Stars

Galaxies are endgame in this class, and we are getting closer! We know they are made of stars (and lots of other things). But, for all but the most nearby galaxies, we do not observe them as individual stars: we see them as stellar populations: unresolved light from all of the stars of all types. To interpret this light, we need to know more about the visible properties of stars.

For an individual star with some surface temperature T_{eff} the radiation field is well approximated by a Planck function (blackbody). We've seen this before!

$$B_{\nu} = \frac{2h\nu^{3}}{c^{2}} \frac{1}{e^{h\nu/kT} - 1}$$

With SI units of W m^{-2} Sr $^{-1}$ Hz^{-1}

This is kind of a units salad– there is a lot going on here! This is the energy per time ($dP;\ W$) emitted in a frequency range ($d\nu;\ Hz^{-1}$) in all directions into space ($d\Omega;\ Sr^{-1}$) by a tiny patch on the surface of the star ($dA;\ m^{-2}$).

We want to start getting rid of some of these units to get to the intrinsic quantity we are most interested in: the luminosity (energy per time, or power).

The first step, to deal with the solid angle dependency, is a bit subtle. Imagine each little patch on the surface of the star radiating its energy in all directions (4π Sr). However, half of this we never see: it goes back inside the star. This means that we only care about the hemisphere above the surface (2π Sr). However, we are not done: we actually only care about the component of the radiation that is perpendicular to the surface, which means we reduce this by an additional factor of 2. Mathematically, this is represented by the integral:

$$\int_{0}^{2\pi} \int_{0}^{\pi/2} B_{\nu} cos\phi d\phi d\theta = \pi B_{\nu}$$

If we now integrate the result over all frequencies, we get the equation for the bolometric flux:

$$F = \int_{0}^{\infty} \pi B_{\nu} d\nu = \sigma_{SB} T_{eff}^{4} \qquad (97)$$

This has SI units of W m^{-2} and σ_{SB} is the Stefan-Boltzmann constant

To get the total energy emitted per time, we need to multiply by the surface area of the star:

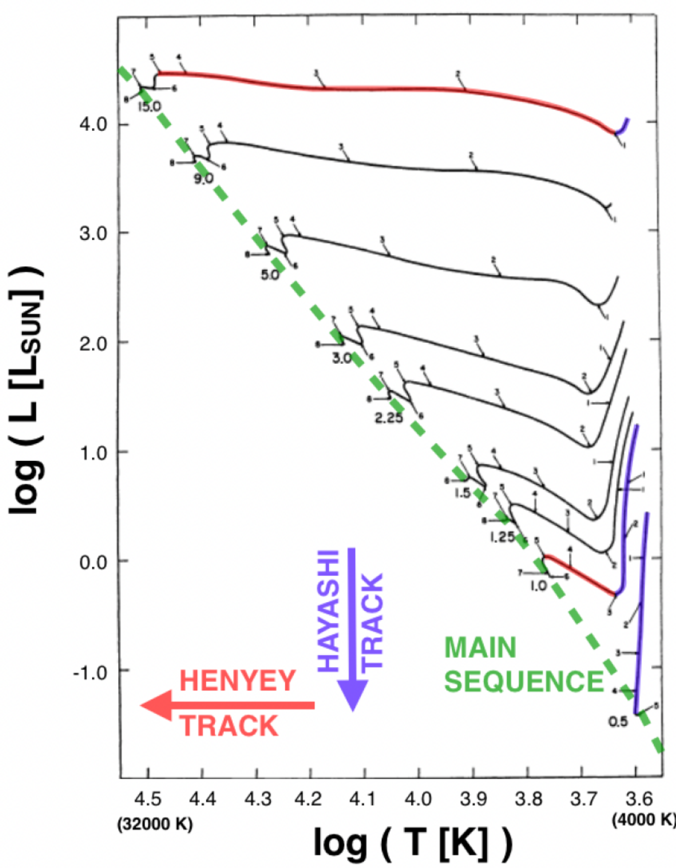
$$L = 4\pi R^2 \sigma T_{eff}^4 \tag{98}$$

This includes some of the key *observable* properties of stars:

- surface temperature (T_{eff}),
- luminosity (L)
- radius (R), though this is almost never directly measured, except for the most enormous stars.

We will focus on the first two, the most easily observed for a wide range of stars.

Enter the H-R diagram:



Let's start with the locations of the protostars we discussed on Friday. Stars of a fixed mass move through this diagram on evolutionary tracks. The protostellar tracks are known as the Hayashi and Henyey tracks. Hayashi: the protostar moves down. Why does this happen? This is one of the isothermal collapse phases but the star is getting dimmer because it is contracting! ($L \propto R^2$). The star then enters one of its quasi-hydrostatic equilibrium stages. It moves to the left (getting hotter) and luminosity increases slightly: If R decreases by factor of 2 (L smaller by $2^2 = 4$) but temperature goes up by factor of 1.5 (L larger by $1.5^4 = 6.$), the luminosity ultimately increases.

Once stars reach true hydrogen-burning energy equilibrium (they make as much energy in their core as they radiate at their surface), stars of different masses are found on the main sequence along a line of nearly constant radius.

How do we know what a constant radius line looks like on an H-R diagram? Going back to our expression for luminosity, if we hold R constant then $L \propto T^4$. In a log-log plot, this is then a slope of 4 ($log\, L \propto 4\, log\, T$)

We have talked about L, R, and T. A fundamental property we are missing is M, the mass. It is hard to directly measure the mass of stars! The best tool for this is generally binary systems, which can then be used to derive an empirical relationship between observed luminosity and the mass (so mass can be determined from a measurement of luminosity). We can sketch a justification for this relationship using some equations for stellar structure:

Hydrostatic equilibrium equation:

$$\frac{dP}{dr} = \frac{GM(r)}{r^2}\rho \qquad (99)$$

Radiation transport equation:

$$\frac{dT}{dr} = \frac{-L(r)}{4\pi r^2}\frac{3}{16}\frac{\kappa\rho}{\sigma T^3} \qquad (100)$$

We will radically simplify things by replacing derivatives with ratios, replacing density by its average, approximating M(r)/L(r) with their surface values, and getting rid of constants.

$$\frac{P}{R} \propto \frac{M}{R^2}\frac{M}{R^3} \propto \frac{M^2}{R^5}$$

$$\frac{T}{R} \propto \frac{L}{R^2}\frac{M}{R^3T^3}$$

Solve this second equation for L:

$$L \propto \frac{R^4T^4}{M}$$

Need to bring in one more equation to deal with pressure: the ideal gas law ($P \propto \rho T$)

$$\frac{TM}{R^4} \propto \frac{M^2}{R^5} \text{ or } T \propto \frac{M}{R}$$

$$L \propto \frac{R^4M^4}{M\,R^4} \propto M^3$$

Actual main sequence values range from 2 - 4:

$$\begin{aligned} L &\propto M^{2.3} && M < 0.43\,M_\odot \\ L &\propto M^{4} && 0.43\,M_\odot < M < 2\,M_\odot \\ L &\propto M^{3.5} && 2\,M_\odot < M < 50\,M_\odot \end{aligned} \tag{101}$$

19. The Initial Mass Function

The initial mass function is the number of stars that are born with each mass, for any group of stars (from a small association of a few stars to an enormous globular cluster with millions of stars). The initial mass function is defined as:

$$\Phi(M) = \frac{dN}{dM} \tag{102}$$

and it is generally assumed to take a power law form:

$$\Phi(M) \propto M^{-\alpha} \tag{103}$$

The most commonly-adopted form of the initial mass function is the Salpeter initial mass function, with an exponent of $\alpha = -2.35$. Technically, this IMF is only valid for massive stars, as observationally it does a poor job describing the number of stars with masses less than about a solar mass (where the slope becomes flatter, and you make more of these stars than the Salpeter IMF would predict).

Let's put this function through its paces a little bit. The first thing we need to do is to normalize the function, and the easiest way to do this is to normalize the total number of stars to 1. In this case, we will then be able to look at the fraction of the total number of stars that are produced at each mass. In order to normalize, we need to choose some upper and lower bounds of integration. We will choose

$$M_{min} = 0.1\, M_{\odot} \text{ and } M_{max} = 120\, M_{\odot}$$

Then we can write (solving for A as the normalization constant):

$$1 = \int_{0.1}^{120} A\, M^{-2.35} dM$$

$$1 = \left[\frac{-A}{1.35}(120)^{-1.35}\right] - \left[\frac{-A}{1.35}(0.1)^{-1.35}\right]$$

$$A = 0.06$$

With this normalization constant defined, we can now do cool things like find what fraction (by number, or F_N) of stars are formed with a mass greater than some mass m:

$$F_N = \int_m^{120} 0.06\, M^{-2.35}\, dM \tag{104}$$

We can even get fancier and ask, what fraction of the MASS do these stars have? To do this, we have to re-do our normalization because now we are integrating a slightly different function: $M\Phi(M)$. This even has its own ridiculous variable, and is defined as:

$$\xi(M) \;=\; M\,\Phi(M) \tag{105}$$

Let's call this normalization constant B, and once again we will do the trick of normalizing to a mass of one (so that the integrals we do will return the fraction of the total (cluster) mass that can be found in stars with some range of masses:

$$1 \;=\; \int_{0.1}^{120} B\, M\, M^{-2.35} dM \;=\; \int_{0.1}^{120} B\, M^{-1.35}\, dM$$

$$1 \;=\; \left[\frac{-B}{0.35}(120)^{-0.35}\right] \;-\left[\frac{-B}{0.35}(0.1)^{-0.35}\right]$$

$$B \;=\; 0.17$$

Now we can set up the same kind of problem, for example asking what fraction of the mass of a cluster (F_M) is found in stars with a mass less than some mass m:

$$F_M = \int_m^{120} 0.17\, M^{-1.35}\, dM \tag{106}$$

20. Star Clusters and Gravitational Interactions

When we begin to look at larger systems of stars, and how they evolve, an important consideration is how frequently the stars in a system interact. We can broadly divide systems into two groups: **collisional**, in which stars regularly undergo 2-body interactions (close approaches that cause a change in their kinetic energy and orbit) and **collisionless**, in which stars essentially never interact, and their motions are primarily determined by the overall gravitational field of the system. Note that even in "collisional" systems the actual stars themselves rarely actually collide, though this can occur in the cores of especially dense globular clusters, where the resulting stellar mergers are known as blue stragglers.

Today, we are going to focus on the dynamics of **collisional systems**, starting with the effects of so-called **strong gravitational encounters**. We are going to define these as encounters that could lead to significant change in a star's velocity: essentially, resulting in $\Delta v \sim v$. Since the only place that the star is going to pick up extra kinetic energy is from the gravitational potential energy of the interaction, we can rewrite this condition as:

$$\frac{Gm^2}{r} \sim \frac{1}{2}mv^2 \tag{107}$$

We can then use this in order to figure out how close together two stars (assuming they have an identical mass m) have to get in order to interact strongly. We will define this separation as the **strong encounter radius**. Solving for this variable yields:

$$r_s = \frac{2Gm}{v^2} \tag{108}$$

Note that the slower the stars are moving relative to each other, the larger the distance between them can be in order to still have a strong gravitational impact. We can do a quick example for stars passing near the solar system. Assume a typical star mass of $0.5\ M_\odot$ (remember most stars are red dwarfs!) and that a typical random speed for stars in a globular cluster is 10 km s^{-1}

Then:

$$r_s = \frac{2(6.67\times10^{-11}\ m^3 / kg / s^2)(0.5)(2\times10^{30}\ kg)}{(10{,}000\ m/s)^2}$$

$$r_s = 4.5 \times 10^{11}\ m$$ or about 3 AU

To determine how regular (or are not!) these encounters are, you would need to take the typical strong encounter radius and plug it into an equation we have worked with before: the **collisional timescale**. Substituting $\sigma = \pi r_s^2$ gives us:

$$t_{strong} = \frac{1}{n_{stars}\pi r_s^2 v}$$

This would give you a strong interaction timescale.

In the opposite regime as strong encounters are the cumulative effects of many smaller **weak gravitational encounters**. These are encounters with separations of $b > r_s$ which essentially act like little tugs on the star, slightly modifying its velocity perpendicular to its direction of motion. We will define the separation b as the **impact parameter**. Each weak encounter leads to a change in the perpendicular velocity of:

$$\Delta v_{perp} = \frac{2\,G\,m}{b\,v} \tag{109}$$

Since these velocity `kicks' have random orientations, to determine the net displacement in the vertical direction, we have to take the sum of the square of all of the perpendicular velocity perturbations:

$$< \left(\Delta v_{perp}\right)^2 > = \int_{b_{min}}^{b_{max}} \left(\frac{2\,G\,m}{bv}\right)^2 dN \tag{110}$$

The quantity dN is the number of interactions that will occur in some time t for stars with an impact parameter between b and $b + db$. To determine dN as a function of the variable b, we want to again visualize a single star moving at a velocity v through a field of stars with some number density n. During a time t it travels a distance vt, and will interact with stars within the volume of a cylindrical shell with radius b and thickness db. Putting this all together, we can define

$$dN = n\, vt\, 2\pi b\, db \tag{111}$$

Plugging this into the integral, we can solve it to find

$$< \left(\Delta v_{perp}\right)^2 > = \frac{8\pi G m^2 n t}{v} ln\left(\frac{b_{max}}{b_{min}}\right) \tag{112}$$

We will then define the timescale on which a star `loses all knowledge of its original orbit', because the change in its perpendicular velocity is as large as its original orbital velocity. This is known as the **relaxation time** (and this process is known as **2-body relaxation**, as although the interactions are weak, they still represent “collisional” encounters between stars) . Setting $v^2 =< \left(\Delta v_{perp}\right)^2 >$ we can solve for the relaxation time t_{relax}:

$$t_{relax} = \frac{v^3}{8\pi G^2 m^2 n\, ln[b_{max}/b_{min}]} \tag{113}$$

Typically, you can assume that b_{min} is just the radius for a strong gravitational interaction, and b_{max} is the size of your system (the radius of a globular cluster, or the thickness of the Milky Way disk). This means the factor $ln\,[b_{max}/b_{min}]$ has typical values of 18-23.

It turns out if you compare the relaxation timescale to the timescale on which strong interactions happen, the 2-body relaxation time will be shorter: This means all of these cumulative, weak interactions will more typically significantly change a star's velocity faster than the time it would take to undergo a strong interaction.

21. Galaxies and Collisional Dynamics[1]

As we calculated, the weak interaction timescale is so long for typical galaxies that we enter the regime of **collisionless systems**. Effectively the collision rate is zero and the mean free path of a star is infinite. Here, an individual star's motions are governed by the ensemble of all stars, rather than by individual encounters. To better quantify this, we introduce the concept of the **gravitational potential**.

The gravitational potential at some point in space a distance r from a point mass M is defined as the amount of work per unit mass that it takes to move a new mass m from infinity to r:

$$\Phi(r) \;=\; \frac{W}{m} \;=\; \frac{1}{m}\int_{\infty}^{r} F(r)dr \;=\frac{-G\,M}{r} \qquad (114)$$

Or more generally in 3D coordinates:

$$\frac{F(x,y,z)}{m} \;=\; \nabla\Phi(x,y,z) \qquad (115)$$

For a continuous distribution of mass in 3D space we can redefine Φ as

$$\Phi(x,y,z) \;=\; \int\int\int \frac{G\,\rho(x',y',z')}{|(x,y,z)-(x',y',z')|}dx\,dy\,dz$$

The gravitational potential is determined by the distribution of mass in a system – in a galaxy, this includes both the visible and dark matter. Applying ∇^2 to both sides of the above integral equation and invoking the divergence theorem gives us Poisson's Equation, which can be a more convenient way to describe the relationship between the potential Φ at a point and the local mass density ρ:

$$\nabla^2\Phi \quad (\equiv\; \nabla\cdot\nabla\Phi) \;=\; 4\pi G\,\rho \qquad (116)$$

[1] This lecture borrows heavily from the notes of others, especially Frank C. van den Bosch and Richard Mushotzky

A couple important and illustrative examples of the gravitational potential:

(1) For a spherical shell, the gravitational force at any point inside a shell of uniform density is zero. This means that the potential anywhere inside the shell is constant.
(2) Outside any spherically symmetric object, the force is the same as if all of the mass were concentrated at the center of that object, and the potential behaves the same way, reducing to the form given in Equation.

Note that both of these break down if the density distribution varies in any way from spherical symmetry!

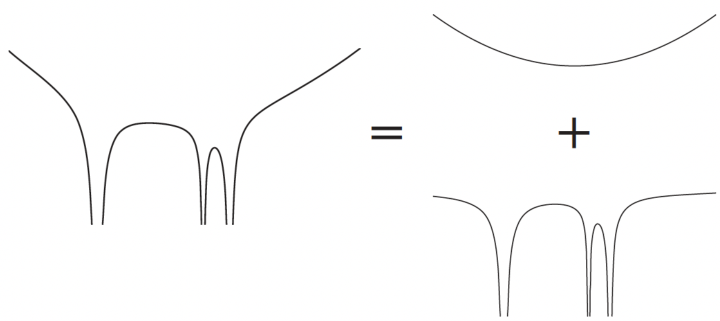


Fig. 3.3. The potential $\Phi(\mathbf{x})$ of a stellar system, represented here by vertical height, can be split into a smoothly varying averaged component and a steep potential well near each star.

The gravitational potential in a system of stars like a galaxy can be represented as essentially having two components. The first of these is the broad, smooth, underlying potential due to the entire galaxy. This is the sum of the gravitational potential of all the mass in the galaxy– the stars, and also the dark matter and the interstellar medium. The second component is the localized deeper potentials due to individual stars. Stars in a galaxy (or other collisionless system) only interact with the smooth part of the potential (the deeper potential wells are relevant to strong/weak individual encounters).

Now, we are going to go one step further than thinking about each individual star interacting with this potential, and think about all of the stars collectively as a fluid. For large systems in particular, this will be an alternative to thinking of stars as objects on individual orbits[2]. In some ways, this will be similar to how we describe atoms in a gas[3]: not by following the path of each atom, but by asking about the density of atoms in a particular region *and* about their average motion– introducing a six-dimensional construct known as the **phase space**.

Why do this? Take a galaxy as an example. To describe the dynamics of a galaxy, we could use the position of each star ($x_i\, y_i\, z_i$) and the velocity of each star ($v_{x,i}\, v_{y,i}\, v_{z,i}$) where i goes from 0 to N and $N \sim 10^6 - 10^{12}$. This is clearly computationally impractical. If we tried to store these data on a computer as 4-byte numbers for every star in a galaxy having $N = 10^{12}$ stars, we would need $6 \times 4 \times 10^{12}$ bytes or 20 terabytes. This is such a large data size that the storage requirements are prohibitive. Furthermore, to actually simulate the motion of stars in the galaxy over time, we would have to compute the gravitational forces between these particles for a large number of time steps. Simply storing the complete set of data for, say, $10^3 - 10^6$ time steps would be impossible. Observationally, meanwhile, it is impossible to determine the positions and motions of every star in any galaxy, even our own. In practice, therefore, people represent the stars in a galaxy using the **distribution function** $f(\vec{x}, \vec{v}, t)$ over the position vector $\vec{x}$ and the velocity vector $\vec{v}$, at a time t. This is the probability density in the 6-dimensional phase space of position and velocity at a given time. It is also known as the **phase space density**.

We can describe how stars move through some volume of (phase) space, which will depend on how the external forces (from the smoothed gravitational potential) accelerate the stars to different velocities. We assume here that stars are conserved, which means ignoring star formation and the deaths of stars, but it will still yield a useful relation.

---

[2] Note: the term orbit is used to describe the trajectories of stars within galaxies, even though they are very different to Keplerian orbits such as those of planets in the Solar System. The orbits of stars in a galaxy are usually not closed paths (think: a rosette) and in general they are three dimensional (they do not lie in a plane). They are often complex. In general they are highly chaotic, even if the galaxy is in equilibrium.
[3] Not only is a collisionless fluid not an ideal gas (encounters are governed solely by long range forces) but there is no equation of state relating the “constitutive” properties of the fluid. Ultimately, the problem is that collisionless fluids do not have constitutive equations such as the equation of state for a collisionless fluid.

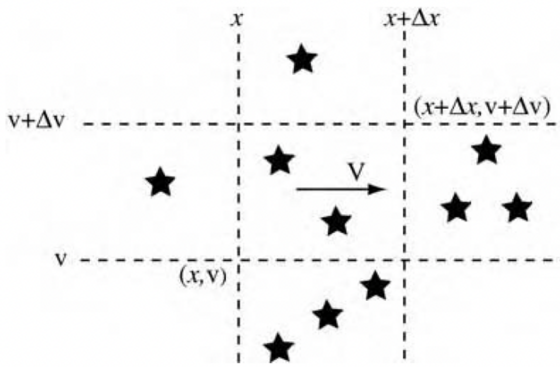


Fig. 3.13. Flow in and out of a box in the phase space (x, v) is described by the collisionless Boltzmann equation.

The defining equation is the **Collisionless Boltzmann Equation**, essentially a continuity equation describing the local flow of stars with velocity v through some volume element**:**

$$\frac{df}{dt} = \frac{\partial f(\vec{x},\vec{v},t)}{\partial t} + \vec{v} \cdot \nabla f - \nabla\Phi \cdot \frac{\partial f}{\partial \vec{v}} = 0$$

What does this mean? The collisionless Boltzmann equation is an extremely complicated way of telling us that $f(\vec{x}, \vec{v}, t)$– the density in phase space– does not change with time for a test particle– it is a conserved quantity. Therefore if we follow a star in orbit, the density f in 6-dimensional phase space around the star is constant. This is a deceptively simple result. If a star moves inwards in a galaxy as it follows its orbit, the density of stars in space increases (because the density of stars in the galaxy is greater closer to the center). Requiring that f be conserve means that the spread of stellar velocities around the star will increase to keep f constant. Therefore the velocity dispersion around the star increases as the star moves inwards. The velocity dispersion is therefore larger in regions of the galaxy where the density of stars is greater. Conversely, if a star moves out from the center, the density of stars around it will decrease and the velocity dispersion will decrease to keep f constant.

All this said, the phase space distribution is usually very difficult to measure observationally, because of the challenges of measuring the distribution of distant stars in 3D space and particularly over velocity. Velocity components along the line of sight can be measured spectroscopically from a Doppler shift, but transverse (along the plane of the sky) velocity components cannot be measured directly for distant sources (

(although the Gaia satellite has made amazing improvements for stars in our own galaxy!). As a function of seven variables (six of the phase space, plus time), the phase space distribution can be awkward to compute theoretically. It is therefore more convenient to use quantities related to $f(x, v, t)$:

- The number density n of stars in space can be measured observationally by counting more luminous stars for nearby galaxies, or from the observed intensity of light for more distant galaxies.

- Spectroscopy provides mean velocities $< v_{los} >$ along the line of sight through a galaxy

- Spectroscopy also yields widths of absorption lines which provide velocity dispersions σ_{los} along the line of sight.

It is typically much more convenient to calculate quantities involving n, $< v_{los} >$, and σ_{los} from an expression for f. These quantities can then be compared with observations more directly.

22. Dwarf Galaxies and the Local Group

As we have previously discussed, unlike clusters, galaxies are **collisionless** systems– any given star in a galaxy is extremely unlikely to have a strong or even a weak gravitational encounter (let alone actually collide head on!)

Another key difference is that for most star clusters (with a few key exceptions), all of the stars were born in a single episode of star formation from the same gas, and share a common age and metallicity. This is known as a **single stellar population.** Galaxies in contrast have a more continuous star formation history

Typically (but not always), galaxies contain many more stars than are found in star clusters

Object	Mass	Description	Distinguishing features
Pleiades	$800\ M_\odot$	Classic open cluster	
R136	$9 \times 10^4\ M_\odot$	Biggest young cluster in the local group	Young stars, high metallicity, located in galaxy disk
M22	$3 \times 10^5\ M_\odot$	First discovered globular cluster	Old stars, low metallicity, located in galaxy halo
Mayall II	$10^7\ M_\odot$	Largest globular in the local group	No Dark matter
Carina Dwarf	$2 \times 10^6\ M_\odot$	A small dwarf galaxy	Dark matter-dominated
LMC	$10^{10}\ M_\odot$	A large dwarf galaxy	Irregular shape, satellite
M33	$5 \times 10^{10}\ M_\odot$	A small spiral galaxy	Well-defined spiral shape
Milky Way	$10^{12}\ M_\odot$	Our home!	Supermassive black hole

And, finally, unlike star clusters galaxies contain an additional source of mass: all galaxies have a **dark matter** component. Unlike star clusters, most galaxies also contain some gas, though this is not required (spirals and irregulars have a lot of gas for forming new stars, ellipticals and spheroidals can have little to none).

With galaxies, we are also beginning to deal with complex **hierarchical structures**: stars are organized into clusters, stars and clusters are organized into galaxies (even the Fornax dwarf galaxy has 5 globular clusters of its own!), and galaxies are organized into groups and clusters, with large galaxies having dozens of smaller satellite galaxies.

Our galaxy, the Milky Way, is part of the **Local group**: a gravitationally-interacting collection of galaxies. Besides us, there are 2 other spiral galaxies: M31 (Andromeda, which is probably a little bigger), and M33 (the triangulum galaxy) which is much smaller). There are also a bunch of dwarf galaxies: some are satellites of the big 3, and some are “field galaxies”, just hanging out by themselves. These dwarf galaxies come in many, many flavors: we see **irregular** galaxies, just blobs of star formation and gas. We see **dwarf elliptical** and **dwarf spheroidal** galaxies, which are kind of like oversize globular clusters (mostly stars, not much dust and gas), except with dark matter. We even see some like the Large Magellanic Cloud (**Magellanic type**), which are *really trying* to be like the big galaxies– a sort of dwarf spiral with a weak bar and one sad spiral arm. Like low-mass stars, these dwarf galaxies greatly outnumber their full-sized counterparts. Unlike low-mass stars, dwarf galaxies are often destined to be eaten by their larger companions, merging with each other or with the host galaxy they are orbiting, becoming part of a greater whole.

We can see many remnants of past meals as coherent streams of stars in the halo of the milky way. The Sagittarius stream, the Virgo stream: these are former dwarf galaxies that have been assimilated into the Milky Way. There is even a Magellanic stream, because even these famous satellites are not out of reach, and they are being tidally distorted by the pull of the Milky Way (and it's mutual: their pull is warping the Milky Way's disk!). Their eventual fate is not clear: they may merge with each other in an act of self-cannibalization, and they may or may not then merge with the Milky Way, perhaps participating in the larger merger event between the Milky Way and Andromeda. Tune in 4 billion years from now to find out!

How do we tell what is what, especially in areas where our classification overlaps in mass (e.g., between an open and a globular cluster, or a satellite dwarf galaxy and a globular cluster?). The difference often comes down to two things: the age of the stars, and the presence of non-luminous matter. Both of these things can be inferred by comparing the ratio of the mass (determined from the stellar or gas dynamics) and the luminosity.

As an example, we can calculate a minimum mass-to-light ratio for stars from the Eddington limit:

$$L_{Edd} = \frac{4\pi G c m_p}{\sigma_T} M$$

$$\left[\frac{M}{L}\right]_{Edd} = \frac{\sigma_T}{4\pi G c m_p}$$

$$= \frac{6.65\times10^{-29}\, m^2}{4\pi\,(6.67\times10^{-11}\, m^3/kg/s^2)(2.99\times10^{8}\, m/s)(1.67\times10^{24}\, kg)}\; 0.16\, kg\,/\,W$$

$$= 0.16 \frac{kg}{W} \frac{1 M_{\odot}}{1.99\times10^{30} kg} \frac{3.8\times10^{26} W}{1 L_{\odot}} = 0.00003$$

We can also calculate a Mass-to-light ratio for a young stellar population, assuming $L \propto M^3$ and a Salpeter initial mass function $dN = M^{-2.35} dM$:

$$\frac{M}{L} = \frac{\int_{0.1}^{120} M (M^{-2.35})\, dM}{\int_{0.1}^{120} L (M^{-2.35})\, dM} = \frac{\int_{0.1}^{120} M^{-1.35} dM}{\int_{0.1}^{120} M^{0.65}\, dM} = \frac{5.9}{1600} = 4 \times 10^{-3}$$

Mass to light ratios of reference objects:

		M ($M_{\odot}$)	L ($L_{\odot}$)	M/L
Proxima centauri	Red dwarf	0.1	0.002	50
Sun	Main-sequence G star	1	1	1
Theta 1 Orionis C	O star	30	200,000	0.00015
Aldebaran	Red giant	1	400	0.0025
Procyon B	White dwarf	0.6	0.0005	1200
R136 (Age ~ 1 Myr)	Young Cluster	9×10^4	3×10^6	0.03
	Globular Cluster			2
	Dwarf Irregular	$10^8 - 10^{10}$	$10^7 - 10^9$	10
Carina	Dwarf Spheroidal	2×10^6		40
	Spiral Galaxy			

How would you calculate a mass-to-light ratio for an older stellar population? This becomes complicated, thanks to red giant stars and stellar remnants, but to get an idea of how it would evolve, you could reduce the maximum mass that you integrate over. For example, doing the same calculation but integrating to a maximum mass of 1 solar mass would give you an idea of how a 10 billion year old stellar population might behave.

23. Dark Matter and Galaxy Rotation Curves

Some of the earliest evidence for the existence of Dark Matter came from observations of the motions of galaxies, particularly the groundbreaking work done by Vera Rubin to study the rotation curves of spiral galaxies. To understand why the rotation speed of a galaxy is such an important parameter, we are going to jump backward a bit into solar system dynamics, and Kepler's Laws.

Based on the observed orbits of the planets in our solar system around the sun, Kepler's third law states that:

$$P^2 \propto a^3 \qquad (117)$$

where P is the orbital period and a is the semimajor axis distance of the planet from the sun. This proportionality is exact if you base it on units defined by the earth's own orbit (units of years and AU) for which each variable has a value of 1. So, for example, you can check that Venus's year is equal to $(0.724\ AU)^{3/2} = 0.61$ years

Adding Newton's law of gravity to the mix, we can write Kepler's third law as:

$$P^2 = \frac{4\pi^2 a^3}{GM} \qquad (118)$$

It turns out that when one is looking at galaxies (rather than planets) it can be useful to be able to solve for the enclosed mass as a function of orbital velocity, rather than the orbital period (since after all, it takes the sun about 200 million years to go around the Galaxy!). To do this, we recall that the distance traveled over one orbit is the circumference of a circle: $2\pi a$. Then, we can write:

$$P^2 = \frac{(2\pi a)^2 a}{GM_{enc}}$$

And

$$M_{enc} = \frac{v^2 a}{G} \qquad (119)$$

For completeness, we can derive the same relation by setting the gravitational force on an object (say a planet, or a star in a galaxy, with mass m) due to the total mass enclosed by its orbit (M_{enc}) to be equal to the centripetal force:

$$\frac{GM_{enc}m}{r^2} = \frac{mv^2}{r}$$

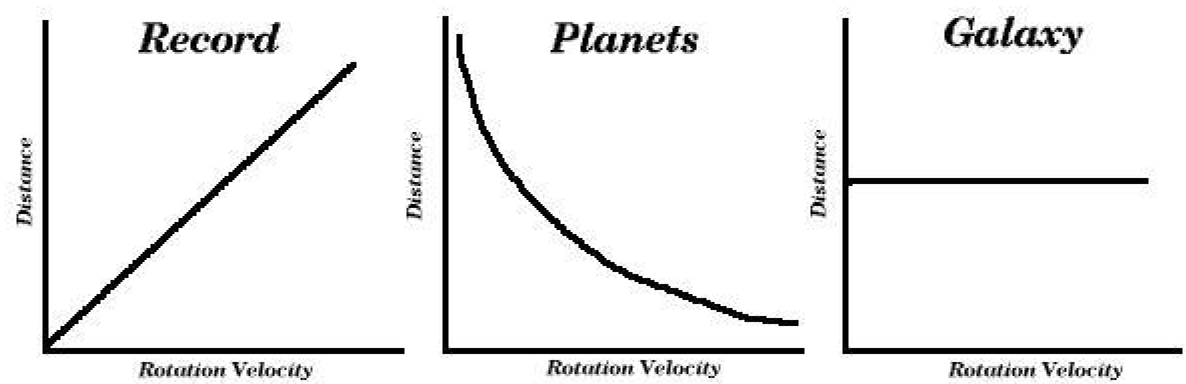
Let's try this out for the Earth. The Earth, according to that great reference ``Monty Python's Galaxy Song", is orbiting the Sun at 19 miles a second (so it's reckoned). For a radius of 1 AU ($\sim 10^{11}$ m) this would give us:

$$M_{enc} = \frac{(19\ miles/s \times 1609\ m/mile)^2 (1.496\times 10^{11}\ m)}{6.67\times 10^{-11}\ m^3\ kg^{-1} s^{-1}}$$

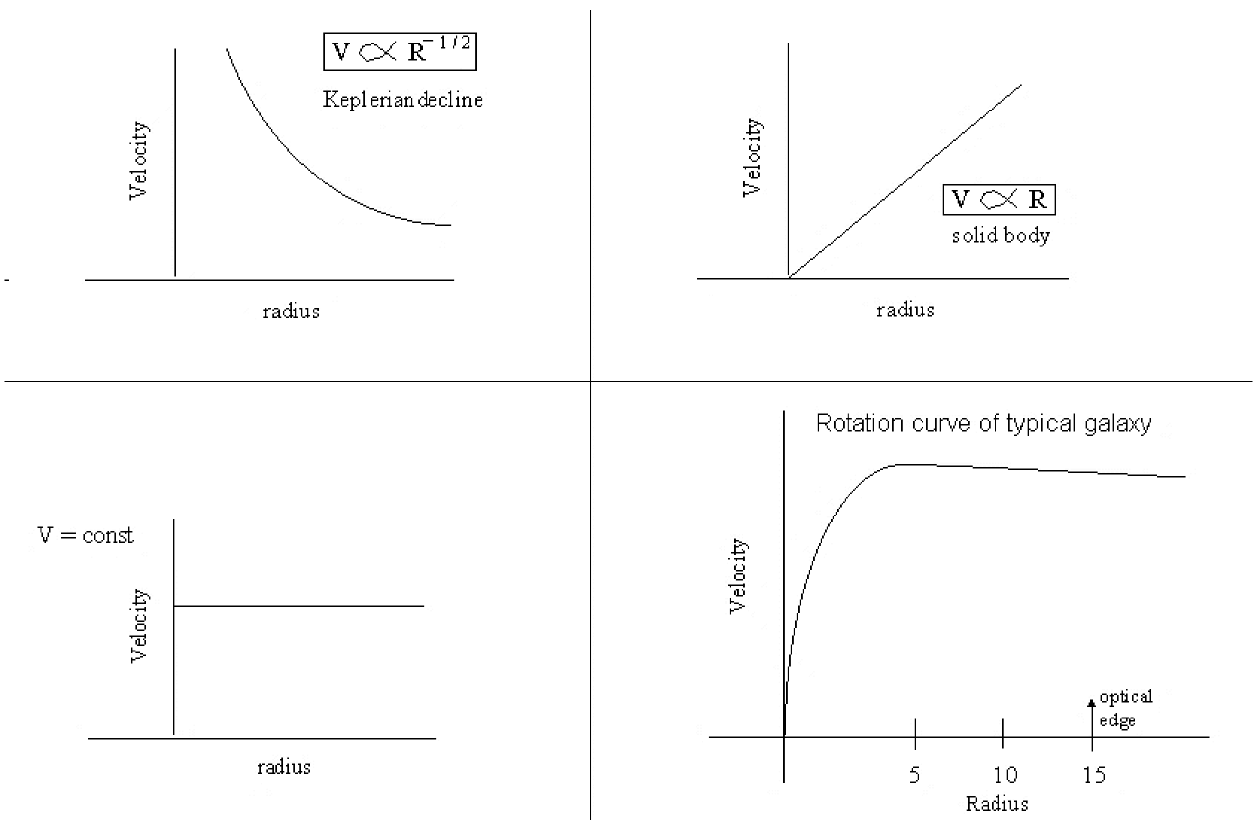
$$M_{enc} = 2.1 \times 10^{30}\ kg$$

NOT BAD! Given the imprecision of our velocity (it came from a comedy song) this is a pretty great estimate of the sun's mass.

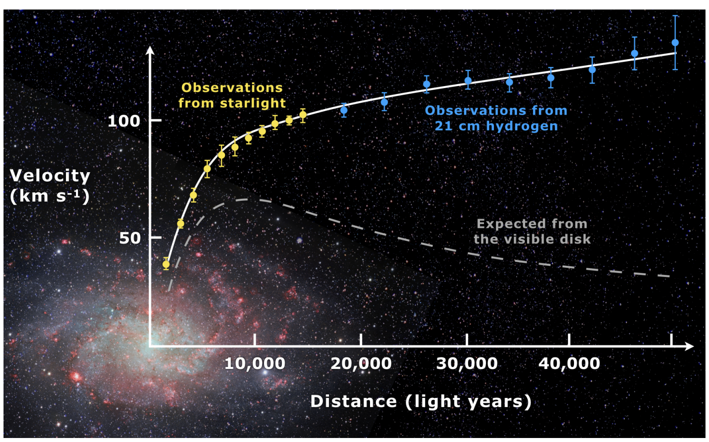
Moving on now, we can look more generally at spinning systems and ask ourselves, what is our intuition for how the rotational speed will change with distance? As humans, we are probably most used to the idea of solid-body rotation. This is like a merry-go-round, or a DVD player (or if you are old, a record player): the object has a constant angular speed, but the farther from the center you look, the faster the tangential speed. This type of relationship between velocity and radius is, unsurprisingly, a good description for things like planets, or even the sun. What about things orbiting under the influence of gravity?

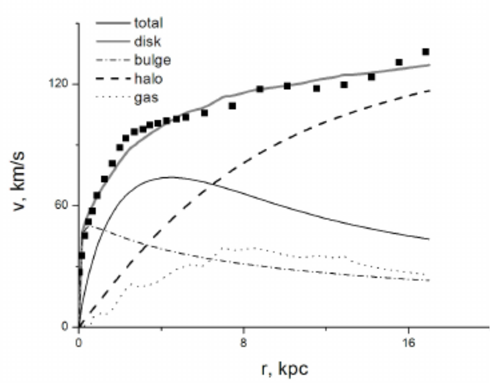


When we look at the solar system, and compare the orbital speeds of the planets, what we find is that the velocity decreases like $1/r^2$. This is easily explained: just look at our expression for the enclosed mass. If the enclosed mass stays constant (and it basically does: the mass of Jupiter is less than 1/1000 the mass of the Sun!), then this is exactly what you would predict for the velocity. This is called a Keplerian rotation curve. What about galaxies?



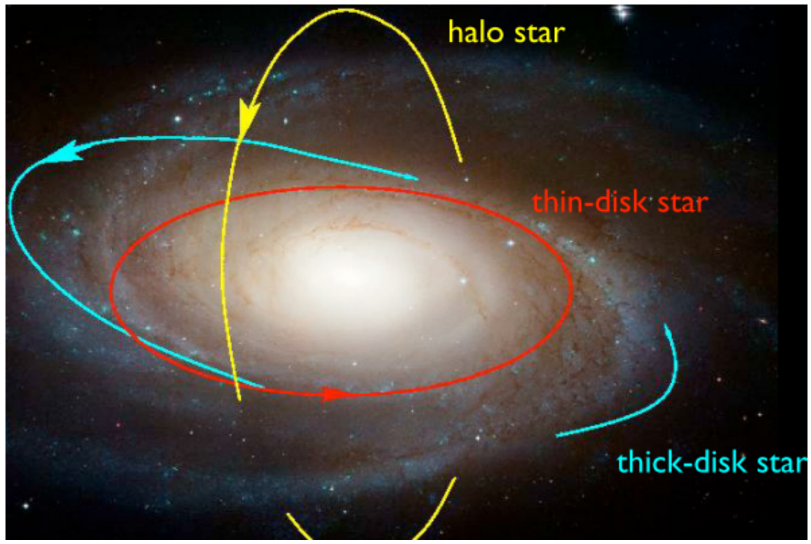
Enter Vera Rubin, who started systematically looking at the rotation curves of entire galaxies: the stars and gas orbiting around their centers. What she found was unexpected and puzzling: the rotation curves did not decline. And Kepler's laws say that as soon as the enclosed mass becomes constant-- as soon as you are not adding any more mass -- the rotation curve should rapidly drop. Yet, even when you get to the edge of the galaxy, and you apparently run out of stars, the curve stays flat (or even increasing!). How can this be explained? The outer regions of galaxies must somehow contain more mass. Eventually astronomers came to accept that while the outer parts of galaxies lacked {\em luminous} matter, there was still a lot of matter present, just in a form we could not see. In fact, this `Dark Matter' actually dominates all of the mass in the universe.
However, the dominance of dark matter in galaxies doesn't tend to kick in until their outer regions (hence why it took so long to notice this). The inner regions of galaxies tend to be dominated by normal matter: the stars and gas are concentrated in the center. There is also dark matter in the center, however its distribution is that of a much larger spherical halo. Thus, it keeps going (and begins to dominate the mass) far out beyond where the stars stop. You can see this in M33: at a distance of about 8 kpc, the mass profile of the dark matter overtakes the mass profile of the stars.





24. Bars, Bulges, and Black Holes

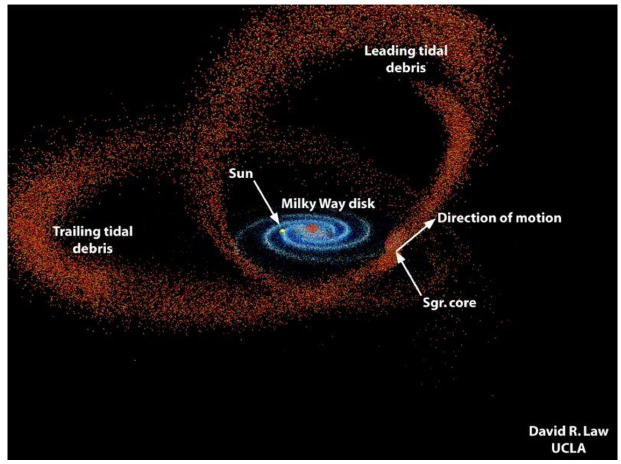
Major components of the Milky Way:



- **Thick/thin Disk**. This is the first evidence some kind of merger likely happened in our galaxy's long-distant past. Our galaxy has two disks: a younger, more metal-rich "thin" disk and an older more metal-poor "thick" disk. One theory for the existence of both is that a minor merger in the distant past stirred up our galaxy's original disk. Because the stars are a collisionless system, the energy injected into the stars that let them travel farther above and below the plane on their orbits, increased excursions that effectively thickened the disk. In contrast, the gas affected by the merger is collisional, and so shocks and falls to the midplane, where it could form a second generation of stars: today's thin disk. Note that the present-day molecular ISM has an even smaller scale height than the thin disk, and as a result we see smaller scale heights for massive (& thus young) stars compared to low-mass (and so typically, old) stars. This is because the longer a star is around, the more time it has for its kinematics to be perturbed: the oldest stars will have the most random (highest above & below the plane) orbits, which will eventually approach the fully random orbits seen in halo stars.

- **Bulge/Bar.** Overall, classical bulges are a denser, more spherical concentration of stars than in the disk, with more randomized orbits: their random motions (sigma) are comparable to their rotation velocities (v_{rot}). Bulge stars tend to be older than disk stars. Our bulge is a bit weird. While Andromeda's bulge is a more classical bulge, the Milky Way bulge is likely something called a "pseudobulge" in that it has a boxy shape with a strong sense of rotation that is dominated by the bar component. As opposed to classical bulges, such pseudobulges may form as a result of instabilities in the disk. However, bulge formation overall is still a mystery, and is linked to the formation of central supermassive black holes.

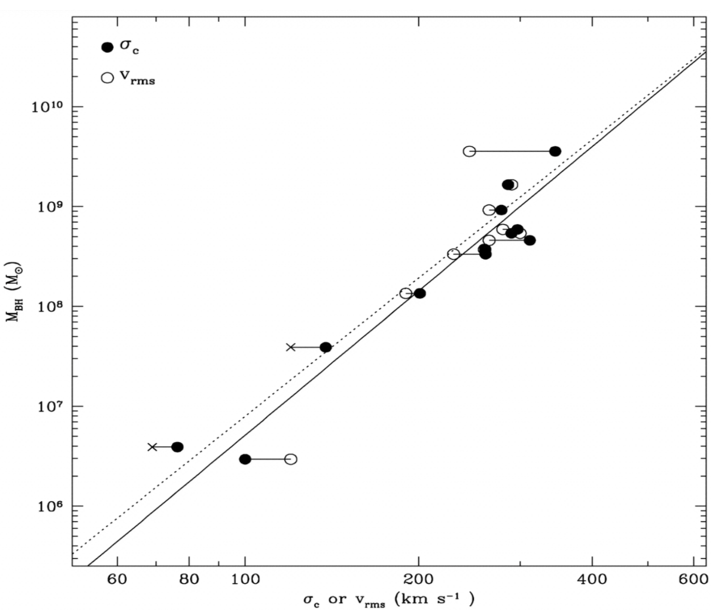
- **Halo.** This is the dark-matter dominated outskirts of the galaxy, with the most metal-poor, oldest (and so most energetic/ randomized/ fastest moving) stars. These are often referred to as Population 2 stars (disk stars are Population 1 stars, and the theorized but not yet observed "first stars" are often referred to as Population 3 stars). The halo includes the most metal-poor systems in our galaxy: globular clusters, which like basically all clusters are believed to have formed in a single burst of star formation, a so-called "simple stellar population". Note that there is a population of slightly more metal-rich globular clusters closer to the plane of the galaxy, which may actually be part of the thick disk, as the distribution of halo stars is roughly spherical. It also contains streams of stars from past mergers, like the Sagittarius stream: a tiny satellite galaxy snapped up by the Milky Way.



- **Supermassive black hole.** Like many galaxies, the Milky Way also has a nuclear star cluster surrounding the SMBH, kind of like a central globular cluster, except with more continuous star formation. It can be seen as a bright central concentration of infrared light, with a radius out to 10 pc from the central SMBH.

Today, the idea that every sufficiently-massive galaxy has a central supermassive black hole is a fundamental assumption that guides studies of galaxy evolution. However, we did not always know that this was the case. The early 2000s marked two big discoveries that cemented this understanding.

Toward the center of the Milky Way, astronomers had started using new capabilities of the largest telescopes (the 10 m Keck telescope in Hawaii, and the 10 m VLT in Chile) to monitor the stars orbiting the Sgr A* radio source (named for its location in the constellation Sagittarius). Earlier observations had shown elevated radial velocities for stars within a few hundred light-days of this source, which meant that there had to be something big and dark (apart from the weak radio and X-ray emission, Sgr A* had no other known multi-wavelength counterpart) they were orbiting. However, at these distances, this could still be a cluster of neutron stars or some other exotic type of object: it did not have to be a black hole. However, with the high resolution now enabled by the size of these telescopes, combined with the ability to observe at infrared wavelengths, where the light is less impacted by foreground dust, astronomers could begin to see that the positions of the stars closest to Sgr A* were changing, year after year. In 2002, independent teams led by Andrea Ghez and Reinhard Genzel measured the orbit of the closest star, S02, which passes within 17 light-hours(!) of the central source. At this radius, there is no known physics that could support a million solar mass object against gravitational collapse: Sgr A* must be a supermassive black hole. In 2020, these teams received the Nobel prize for their work.



Around the same time, astronomers were looking at other galaxies for which there was also evidence of a massive central object (though with less rigorous constraints on its size). In these distant galaxies, they found a fascinating relationship: the mass of the central object was related to the mass of the central stellar bulge (as measured from the velocity dispersion of stars in the bulge). Today, this correlation is known as the "M-sigma" relation, as it relates the mass of the central supermassive black hole (M_{BH}) and the velocity dispersion of the bulge ($< v_{RMS} >$ or σ).

The first measurement of the $M - \sigma$ correlation between the central supermassive black hole mass (M_{BH}) and the velocity dispersion of stars in the galaxy's bulge (σ).

The best current fit to the data yields a relation that can be expressed as:

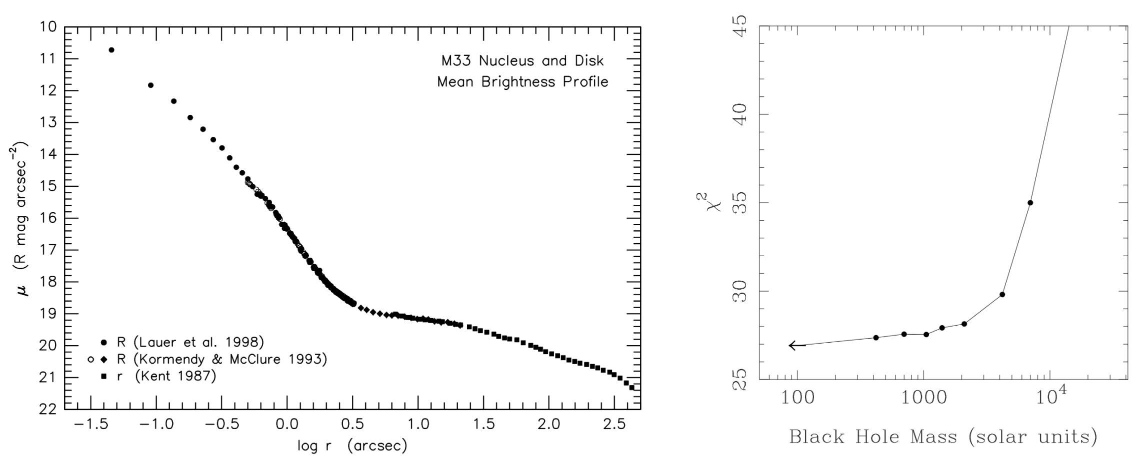
$$\frac{M}{10^8 M_\odot} = \left(\frac{\sigma}{200\ km/s}\right)^{5.1} \tag{120}$$

First and foremost, the $M - \sigma$ relationship proved to be an essential tool, as it meant that if the velocity dispersion of the bulge stars can be measured (a relatively easy task for distant galaxies) then it is possible to ascertain not only whether a galaxy has a central supermassive black hole (now the most likely physical interpretation of the central mass, thanks to measurements in our own Galaxy), but what its mass must be. The relationship also raises interesting possibilities as to whether this is purely a correlation, or if this also implies causation: some action of the black hole or its growth contributes to the mass of the bulge. Either option is rich with possibility: either the means of growing a black hole and growing a bulge are the same (for example, both could be dominated by the inflow of mass from galaxy mergers) or feedback from the black hole shapes the galaxy, contributing to the growth of the bulge.

In the Local Group, we can see that the galaxies fairly well follow the $M - \sigma$ relation. M31 (the Andromeda Galaxy) is measured to have a more substantial bulge ($\sigma_{M31} = 160\ km/s$; $\sigma_{MW} = 75\ km/s$) and black hole mass ($M_{BH(M31)} = 10^8 M_\odot$; $M_{BH(MW)} = 4 \times 10^6 M_\odot$) than seen in the Milky Way. We can check this with the proportionality suggested by the $M - \sigma$ relation, finding that doubling the velocity dispersion should increase the mass by a factor of about 34, very consistent with these measurements.

The late-type spiral galaxy M33, with no bulge. Right: The early-type spiral galaxy M101, with a prominent central bulge

However, unlike the other spiral galaxies in the Local Group, M33 (the Triangulum Galaxy) does not actually have a bulge at all! In contrast to M31 (Hubble type SA-b) and the Milky Way (Hubble type SB-b/c), it has a type of SA-c/d, classifying it as a ``late-type" spiral with no prominent bulge. Early-type galaxies like SA-a (no bar) or SB-a (bar) have a substantial bulge, making them more similar to elliptical galaxies which can be broadly termed to be all bulge (no disk).

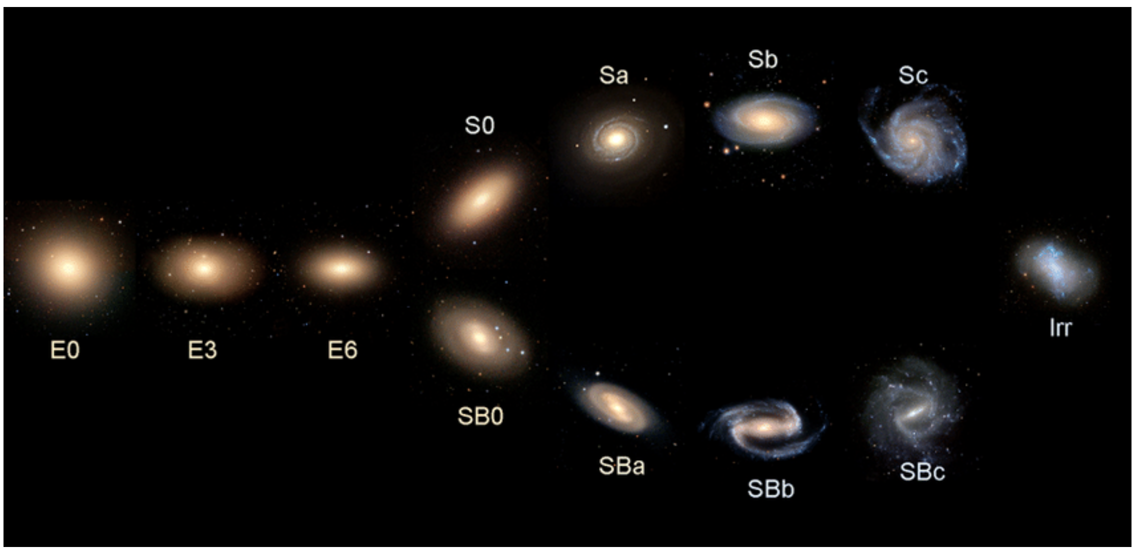


Left: Surface brightness profile of M33. The galaxy consists only of a disk and a central star cluster with kinematics similar to a globular cluster. There is no bulge. Right: Chi-squared fitting of models for the mass of a central supermassive black hole in M33 from Gebhardt et al. (2001). A small χ^2 indicates a good fit.

Consistent with the lack of a bulge, observations of the central stars in M33 have limited the presence of any black hole to be less than a few thousand solar masses: not supermassive at all, if it even exists! (note that the best-fit models from Gebhardt et al. (2001), in this figure suggest a supermassive black hole is not needed at all to replicate the observed stellar kinematics).

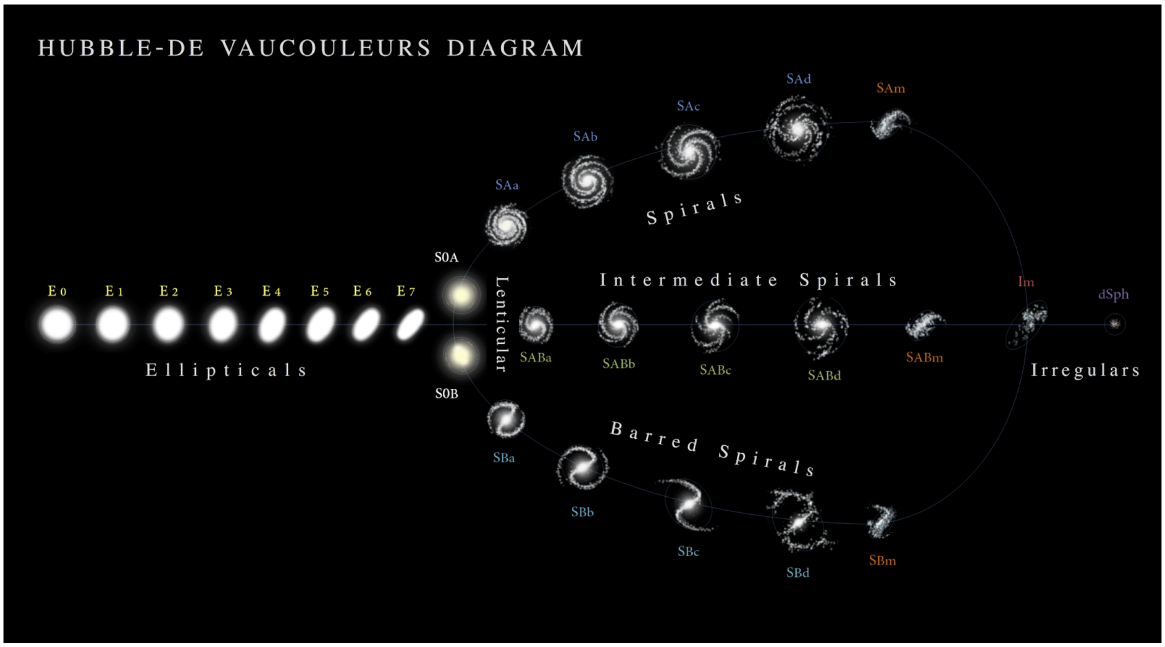
25. Galaxies, Groups, and Mergers

Hubble sequence:



To see how our Milky way compares to other galaxies, we look to an organization of galaxy types made by Edwin Hubble: The Hubble Sequence. In Hubble's ordering, galaxies on the left side of the sequence are known as "early type"-- this can just mean ellipticals, or can also refer to S0 galaxies and SAa/SBa galaxies. Galaxies on the right are known as late type galaxies, and are characterized by having mostly young stars, and little if any bulge. As a reminder, M33 is classified as SAcd, reflecting its non-existent bulge. Andromeda (M31) which has a prominent bulge, is an SAb galaxy. The Milky Way is given a totally ambiguous classification of SABbc, indicating its relatively weak but present bar, and its moderate (pseudo)bulge.

Which of these galaxies do you find in a group environment? Mostly spirals and generally "late-type" galaxies. Note that when we start talking about galaxy evolution, this naming system is a CLASSIC example of observers fucking it up for everyone because they just classified things arbitrarily, by looks, without understanding the underlying physics. In fact, we think the evolutionary sequence of galaxy evolution through mergers proceeds in the opposite direction: from so-called “late-type” gas-rich (low-mass) bulge-less spirals to the more and more gas-poor, bulge-dominated “early-type” galaxies, ending with the giant elliptical galaxies, the eating machines that hang out in the center of dense cluster environments (more on these next week).

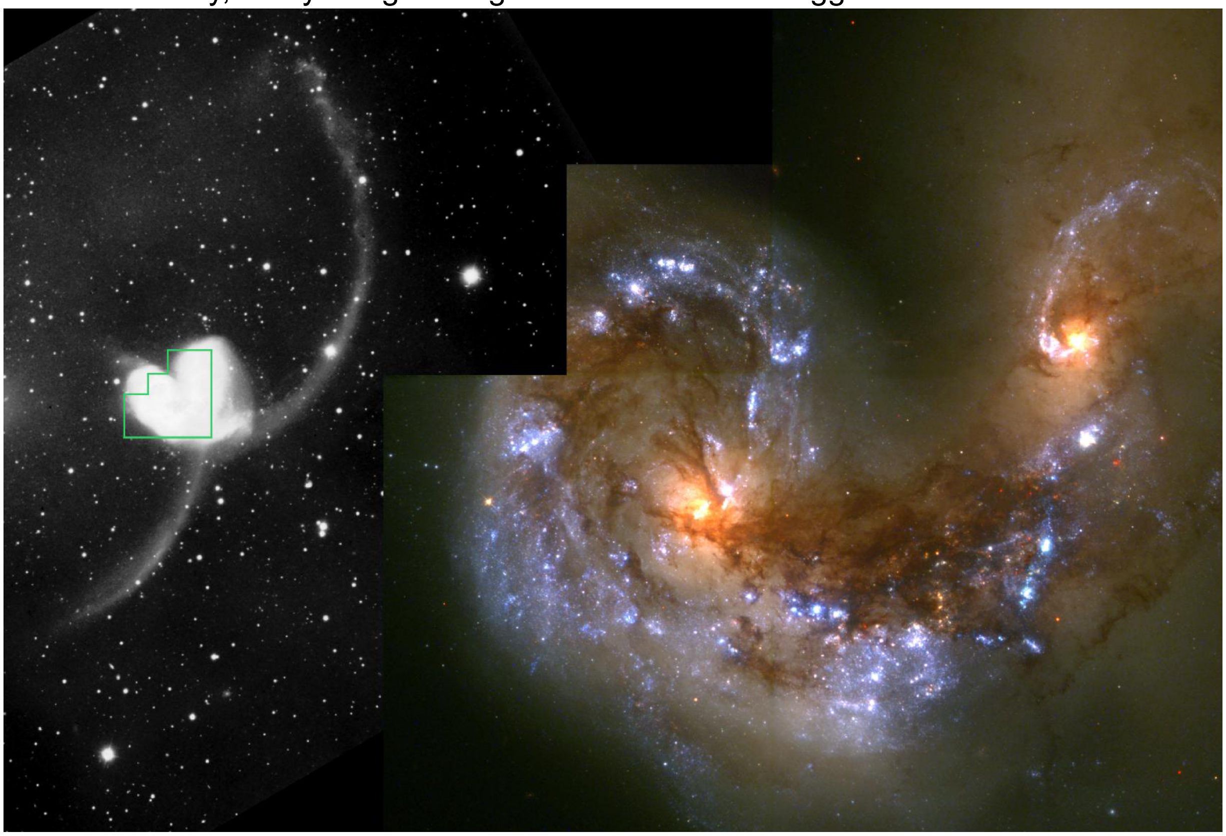


How do galaxy-galaxy interactions work? Even when you only consider the stars (and dark matter) in galaxies, these are more complicated than the star-star interactions we discussed for clusters, these are more like an infinitely-scaled up version of interactions between stars in binary systems. This is because the energy exchanged during an encounter can go not just into the system energy (the overall speed and trajectory of the galaxy as a whole) but into the “internal” energy states of the stars orbits. Let’s just say: when we are talking about systems of galaxies, this is generally FAR from the assumptions of an ideal gas. The galaxies themselves take up a substantial volume. Interactions are not fast: they can last millions to billions of years. And they are hardly elastic.

Some mergers, like the classic example of the Antennae galaxies, are “major” mergers of comparable-sized galaxies.

Some, like M82, are more likely the result of a minor merger, where a smaller galaxy fell into the nucleus, triggering a violent central starburst. This, as well as the thick disk of the Milky Way, and the fact that the Magellanic clouds are even now warping the Milky Way's disk, are good reminders that a "minor" merger can still fuck things up.

Observationally, many mergers of galaxies are seen to trigger star formation:



The Antennae galaxy is again a classic example of this, with the zoomed-in image showing bright blue spots, each of which are baby super-star clusters. To trigger star formation, a merger typically has to be a "wet merger", in which the galaxies contain gas along with the stars and dark matter. Many irregular/starburst galaxies are believed to be the result of a recent wet merger (or sometimes just the tidal disruption leading up to one). In contrast, dry mergers involve no gas, and only stars/dark matter: these can be mergers like the incorporation of a small gas-poor dwarf spheroidal into the Milky Way halo, or more classically the merger of two elliptical galaxies.

Let's get back to why we see predominantly disk galaxies in group environments. Number one of course, groups are relatively low-mass structures, so this could be a result of how they are formed, like observations that high-mass stars are only found to form in relatively large star clusters. Another explanation however involves their evolution, and the relative frequency (or infrequency) of mergers. Note when talking about such interactions that (1) galaxies are messy and complicated, so interactions can be drawn out over hundreds of millions to a billion years and require simulations to really understand -- it gets really complicated when you get mutually orbiting systems-- and (2) galaxy groups are relatively small (N<50) systems for which the assumptions that go into calculating a relaxation time do not necessarily apply (for example, that the only gravitational potential galaxies experience is the smoothed potential due to the summed contribution of all objects in the system).

All that said, we can still try to define a “collision rate” for galaxies in a group environment:

$$t = 10^9\, yrs \left(\frac{1000\ km/s}{\sigma}\right)\left(\frac{1000\ Mpc^{-3}}{n}\right)\left(\frac{100\ kpc^2}{R^2}\right) \qquad (121)$$

To do this, we need to know three variables: typical speed, typical radius, and the galaxy density. We will say: Local group has a radius of 3 Mpc, and 50 constituent galaxies. Typical velocity dispersion is 60 km/s. Typical size is a bit trickier, but let's just say everything is LMC-sized: 2 kpc radius. Given these properties, we can see that the collision rate is going to be much more than the age of the universe, which helps explain why disk galaxies survive so well in a group environment!

26. Galaxy Scaling relations

As we start talking about elliptical galaxies (the giants of galaxies!) it is useful to break down the key observable and intrinsic parameters of galaxies as a whole.

(1) **Size.** Unlike (most) stars, the hugeness of galaxies means we can actually measure a physical size, even though they are hundreds to thousands of times farther away. As observers, we measure an angular size on the sky (θ) which we can relate to its actual physical size (R) using the relation $R = d\theta$. Here, R and d (the distance) should be in the same units, and θ should be measured in radians.
(2) **Brightness**. As observers, we measure an apparent brightness f of all stars in the galaxy. This can be related to the total luminosity L using the inverse square law: $L = 4\pi d^2 f$
(3) **Mass.** Understandably, this one is a little harder to observe directly, but luckily we have equations like the virial theorem (for objects dominated by random motions), or the enclosed mass relation (for rotation-dominated objects) which lets us relate the observable quantity of the velocity dispersion σor the velocity v to the total mass, using either $M = Rv^2/G$ or $M = R\sigma^2/G$.

Now we can use these relations to justify some of the trends we observe when we look at galaxies.

We start by defining the surface brightness of a galaxy (which is a quantity independent of distance). Putting this together with (1) and (2) yields:

$$I = \frac{f}{\theta^2} = \frac{L\,d^2}{4\pi d^2 R^2} = \frac{L}{4\pi R^2} \qquad (122)$$

We can then use (3) to express R in terms of M and v:

$$I = \frac{L\,v^4}{4\pi M^2 G^2}$$

Here is where we do a little bit of ridiculous mathematical gymnastics and note that we can multiply the top and the bottom of this fraction by $1/L^2$, and introduce a quantity called the mass-to-light ratio M/L:

$$I = \frac{v^4}{L\, 4\pi G^2 (M/L)^2}$$

Rearranging to solve for L:

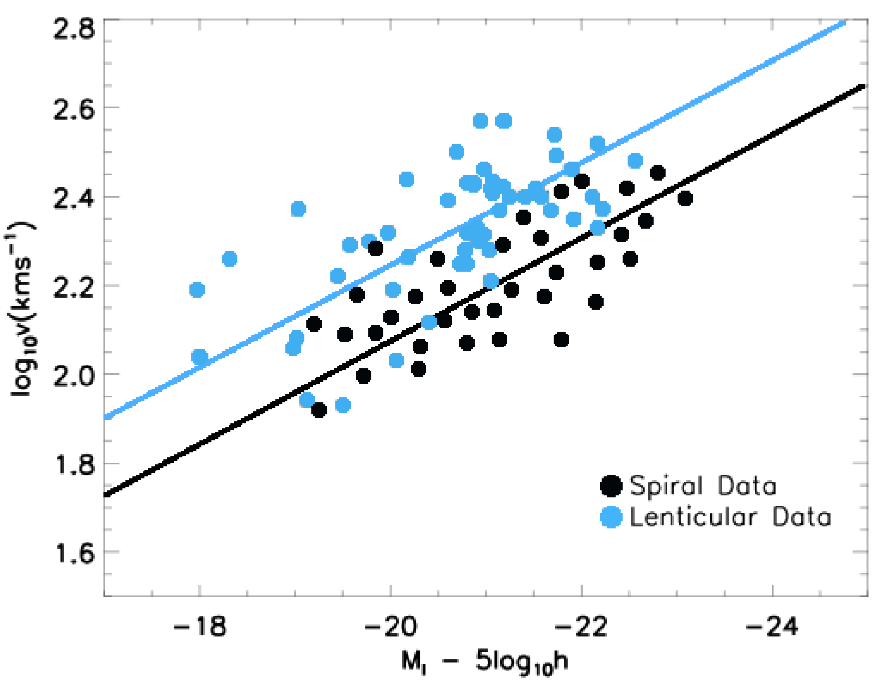
$$L = \frac{v^4}{4\pi G^2 I\, (M/L)^2}$$

Here you can make two big and key assumptions:

(1) For galaxies of a similar type, the central surface brightness I can be assumed to be constant
(2) For galaxies of a similar type, the mass-to-light ratio is the same. Note that this does *not* take into account dark matter, and so essentially is a measurement of the typical age of the stellar population in the galaxy)

Then, you can say:

$$L \propto v^4 \tag{123}$$



This relation (valid for spiral galaxies), is known as the Tully-Fisher relation (named after Brent Tully and Richard Fisher), and is one of the fundamental observed galaxy scaling relations. This means that galaxies mostly (but not entirely!) are observed to follow this relation, which is broadly justified by the expected underlying physics, but which can have deviations due to variations in quantities (like e.g., I or M/L). As an example, systematic differences are seen between spirals and lenticular (S0) galaxies (which are an earlier type than typical spirals). Still, when properly calibrated by observational data, this is a super-useful relation for roughly assessing the distance to galaxies, as it says that the intrinsic luminosity is proportional to the observed rotation speed.

A similar relation exists for elliptical galaxies, just substituting σ (the velocity dispersion) for v:

$$L \propto \sigma^{4} \tag{124}$$

This relation is known as the Faber-Jackson relation, named for Sandra Faber and Robert Jackson. It is part of a set of relationships between the correlated properties of elliptical galaxies (mass, luminosity, velocity dispersion, and central surface brightness) known as the Fundamental Plane. This is often expressed as a power-law relationship between R, σ and I:

$$R \propto \sigma^{1.4} I^{-0.9} \tag{125}$$

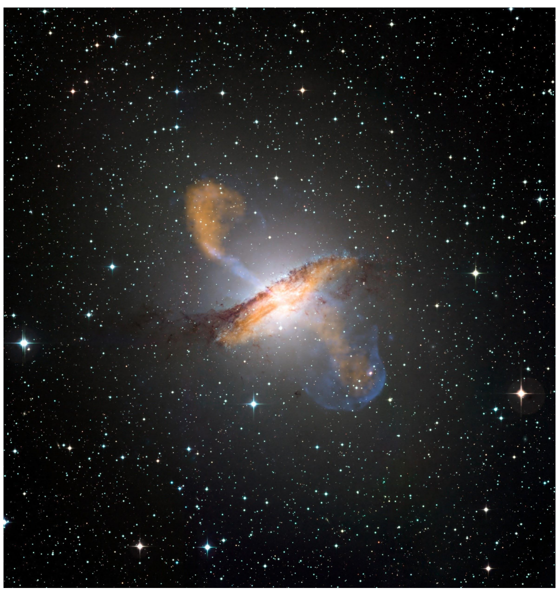
This can be re-expressed in terms of the luminosity if you apply another of the associated scaling relationships for elliptical galaxies, the one for surface brightness, which in this form is sometimes known as De Vaucouleurs' law:

$$L \propto I\, R^{2} \tag{126}$$

27. Galaxy Clusters and Hot Gas

Elliptical galaxies are most commonly found in a cluster (or compact/dense group) environment, like the rich Coma cluster seen above. This reflects the increased density of sources in a cluster environment, which makes mergers (and other dynamical interactions) more frequent. Consistent with this, particularly massive ellipticals are seen at the centers of clusters. Special examples of massive ellipticals are cD galaxies and BCGs (brightest cluster galaxies). cD galaxies are more than just extremely massive ellipticals: they are seen to be structurally different, with a more extended faint halo, likely indicative of their origin through many major mergers. cD galaxies and BCG galaxies are a not-entirely overlapping set (like rectangles and squares): Typically a cD galaxy will be the BCG, but there may be several cD galaxies in a cluster (for example, as seen in the Coma cluster above, which may indicate a past merger, and resulting complex structure that means there is no clear center to the cluster).

Although elliptical galaxies are generally thought of as passively evolving, and most lack recent star formation, that doesn't mean there is no ISM in these galaxies. Many contain dust, likely from recent mergers (check out Centaurus A below left as an extreme example of an S0/ elliptical galaxy that has undergone a huge merger, and has enormous radio lobes from an AGN jet: such jets are seen almost exclusively in ellipticals). Elliptical galaxies also contain hot gas, which comes from the winds of all of the stars. Some of this gas may escape the galaxy, contributing to the hot X-ray emitting cluster gas that fills the space between the galaxies. Interestingly, some elliptical galaxies can also have cold and even molecular gas! Often this is seen in ellipticals outside of a rich cluster environment, but it is also sometimes seen in a dominant central galaxy of a cluster, like in Perseus (below right, where you can also see huge ionized gas filaments surrounding the massive central elliptical galaxy). There is still generally very little star formation associated with this gas. Its origin may be from gas-rich mergers, or 'cooling flows' where the dense hot x-ray gas in the cluster cools rapidly and moves inward.



Overall, a typical galaxy cluster has:

- A total mass of $10^{14} - 10^{15}\,M_{\odot}$
- 1% of the mass is from the galaxies (100s to 1000s of large galaxies)
- 9% of the mass is also baryonic but in the form of the hot x-ray cluster gas
- 90% of the mass is Dark matter

The fraction of mass that is not dark matter (the **baryon fraction**) that is measured in galaxy clusters is less than expected from cosmology, which predicts that it should be about 17%. This is called the **missing baryons problem**. While this has not been definitively solved, a likely solution is that the missing baryons lie outside of galaxy halos and galaxy clusters, in hot x-ray gas associated with filaments in the “cosmic web” of large-scale structure.

This is also a good point to be clear that there are two (or more!) types of Mass-to-Light ratios (M/L) that are typically discussed in galactic and extragalactic astronomy.

(1) $(M/L)_*$ This refers only to the mass of the stars compared to the luminosity of the stars. A large M/L implies an old stellar population (recall our previous calculations of M/L for stars more and less massive than the sun).
(2) $(M/L)_{tot}$ This is all of the sources of mass in an object (gas, dust, stars, and dark matter) compared to just the luminosity of the stars. This M/L is typically used to discuss galaxies as a whole, in the context of how dark-matter-dominated they are.

28. Supermassive Black Holes and AGN

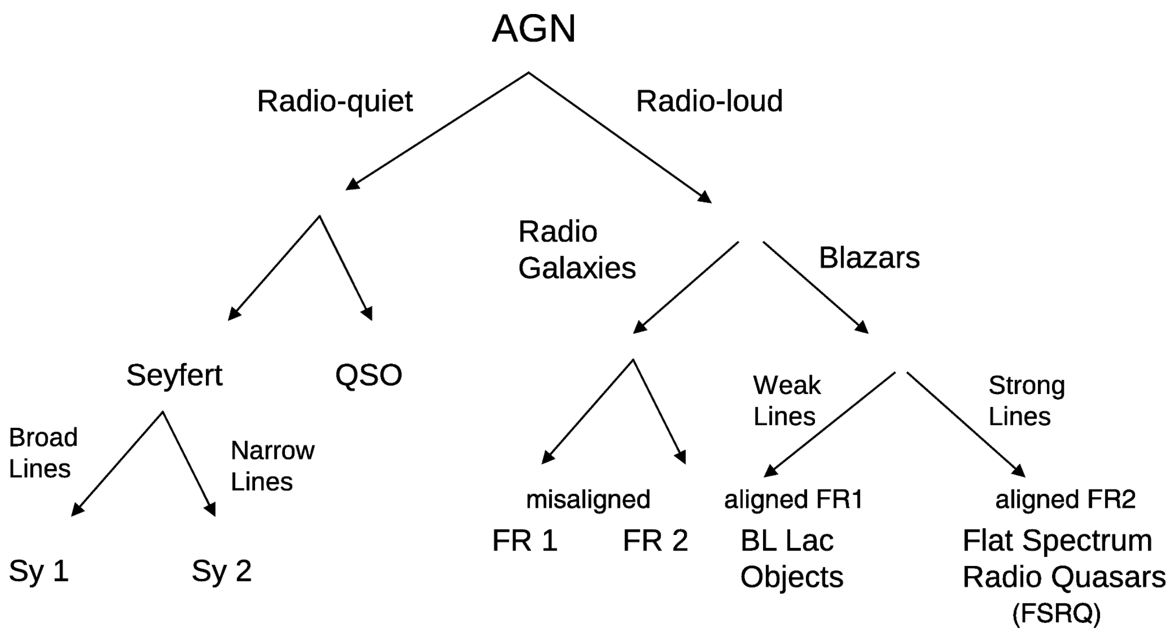
How big is the event horizon? A closely related quantity is the Schwarzschild radius (R_s) To derive this, equate the speed of light with the escape speed at some distance from the black hole:

$$R_s = \frac{2GM_{BH}}{c^2} \tag{127}$$

Note that the masses of supermassive black holes (and thus their effective radii) can vary by orders of magnitude:

- Milky way black hole mass is $4 \times 10^6 M_\odot$
- M87 black hole mass is $7 \times 10^9 M_\odot$

Many properties of AGN scale with the black hole mass and the resulting Schwarzschild or gravitational radius.
Observationally, AGN have many different characteristics and are subdivided into a range of categories based on the properties of their emission:



Some important observational properties of AGN include:

- Wavelengths at which emission is observed (from radio to gamma rays)
- Luminosity (overall or at a specific wavelength, like X-rays)
- Ionization state of the observed emission lines
- Widths of the observed emission lines
- Variability time scales

We can get an (order of magnitude!) estimate of the **accretion luminosity** from the virial theorem. Assume that a small mass of gas m originally starts at a far enough distance from a black hole of mass M that its initial gravitational potential energy is zero. The final gravitational potential energy will be the gravitational potential at the inner radius the gas can reach: the Schwarzschild radius. The change in potential energy is:

$$\Delta U = -\frac{GMm}{R_s}$$

Conservation of energy says that energy is not lost in this process: for the potential energy to have become more negative, an equivalent, positive amount of energy must have been ‘released’. The virial theorem says that of the gravitational potential energy which is released, half of it goes into the kinetic energy of the gas, and half of it is radiated away.

To get the luminosity (energy per time) instead of a single point mass m, we will imagine many of these coming in over time, so that we can describe them as the accretion rate $\dot{M}$. Then, we can write the accretion luminosity as

$$L_{acc} = \frac{1}{2} GM\dot{M}\frac{c^2}{2GM} = \frac{1}{4} \dot{M}c^2$$

$$L_{acc} = \epsilon \dot{M}c^2 \quad (128)$$

The exact fraction out front varies due to the geometry of the accretion flow and the black hole, and is typically described with an accretion efficiency factor ϵ (or η). Like with stars, there is a theoretical maximum to this luminosity: the Eddington luminosity:

$$L_{edd} = \frac{4\pi c\, GM}{\kappa}$$

You can set the accretion luminosity equal to the eddington luminosity to find the maximum accretion rate for a black hole:

$$\dot{M}_{edd} = \frac{4\pi G}{\epsilon \kappa c} M \quad (129)$$

Seyfert AGN are divided into **Type 1** and **Type 2** based on whether broad emission lines (**the broad line region**) can be observed. Going back to our equation for enclosed mass, we can make an *order-of-magnitude approximation*, and say that the width of the line (sigma), which is basically the spread in observed velocities, can be related to a (circular) rotation velocity at some radius by $\sigma = 2\, v_{rot}$ (e.g., if we viewed a disk of gas edge-on we would see half the gas coming toward us at $-\, v_{rot}$ and half the gas receding at $+\, v_{rot}$). Then, we can say that for a given observed line width and black hole mass, the distance of the gas from the black hole is:

$$r = \frac{4GM}{\sigma^2}$$

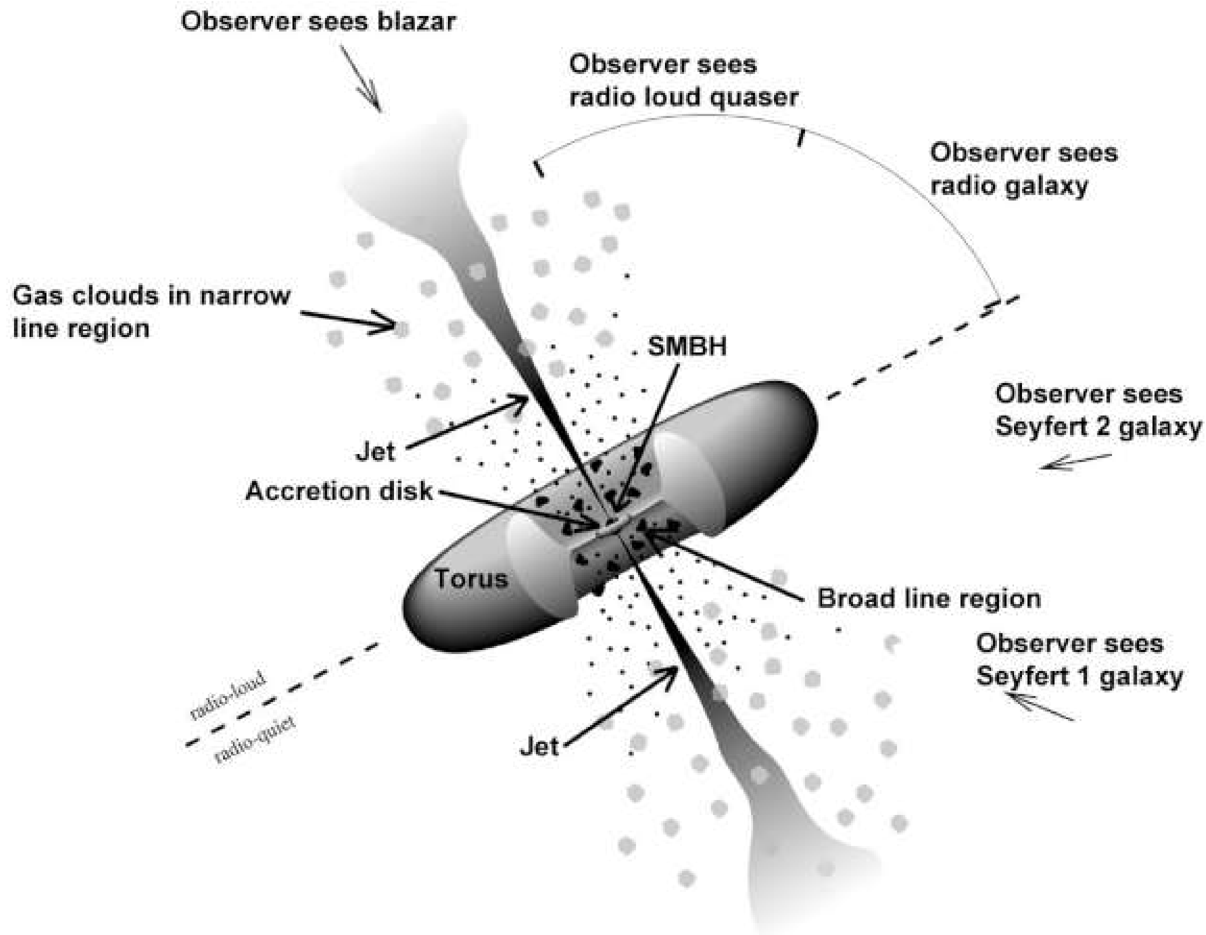
A typical “broad” line width (as observed from the Circinus AGN, which has a supermassive black hole mass of $10^6 M_\odot$) is 3000 km/s:

$$r = \frac{4\,(6.67\times10^{-11}\,m^3 kg^{-2}\,s^{-2})(10^6)(1.99\times10^{30}\,kg)}{(3000\times10^3\,m/s)^2} = 6\times10^{13}\,m$$

This is equivalent to about 30,000 times the Schwarzschild radius, or 2 light days.

AGN are observed to be highly time-variable. The shortest time scale on which variability can be observed will be the time it takes light to travel from one side of the object to the other – the **light crossing time**. Thus, for AGN variability, this will be set by the Schwarzschild radius (and the speed of light). Variability time scales range from milliseconds, for stellar-mass black holes, to days for large supermassive black holes.

An important insight into all of these observed characteristics is the unified model, which postulates that many (though not necessarily all) of these categories of AGN are actually the same type of object, and are visually different only because of the angle at which they are viewed.



Observer sees blazar
Observer sees
radio loud quaser
Observer sees
radio galaxy
Gas clouds in narrow
line region
SMBH
Observer sees
Seyfert 2 galaxy
Jet
Accretion disk
Torus
Broad line region
Observer sees
Seyfert 1 galaxy
radio-loud
radio-quiet
Jet