**Statistical Equilibria: Saha Equation**

**Initial questions:** What sets the temperature of the warm interstellar medium? What determines the epoch from which we see the CMB?

We are now going to consider the statistics of equilibrium. Specifically, suppose we let a system stand still for an extremely long time, so that all processes can come into balance. We then measure the fraction of particles that are in each of several states (e.g., if we have a pure hydrogen gas, we might want to know the ionization fraction). How can we compute this fraction? We'll take a look at this in several different ways. We will then step back and ask ourselves when these equilibrium formulae do *not* apply. This is important! To paraphrase Clint Eastwood, "An equation's got to know its limitations". In virtually all astronomical problems we deal with approximations, so we always need a sense of when those approximations fail. This serves a double purpose: it keeps us from making errors, but it also keeps us from applying too complex an approach to a simple problem.

Let's start out with a simple case. We consider a column of isothermal gas (temperature $T$) in a constant gravitational acceleration $g$. Each molecule has mass $m$. If we can treat this like an ideal gas, **Ask class:** how do we figure out how the pressure varies with height $h$? An ideal gas has pressure $P = nkT$, where $n$ is the number density. In equilibrium the pressure gradient balances the weight of the gas, so $dP/dh = -\rho g$. Using $\rho = nm = Pm/kT$ and the constancy of $T$, we have $dP/dh = -(mg/kT)P$, so $P = P_0 \exp(-mgh/kT)$, where $P_0$ is defined as the pressure at $h = 0$. This is the familiar Boltzmann formula. As usual, we can check various limits: when $h$ increases the pressure decreases (correct); when $T$ decreases the pressure decreases at a fixed $h$ (correct); and so on. Note that the decrease is faster for heavier molecules. This is one reason why at high altitudes we have an oxygen deficit: our natural breathing rate is controlled by the level of $CO_2$ in the bloodstream, and this decreases rapidly at high altitude so we don't breathe as often as we should until we acclimate.

This is one special case of a much more general formula of statistical distribution in thermal equilibrium. I highly recommend you look at the Feynman Lectures on Physics if you want physical insight on just about anything; Feynman had a remarkably original perspective on physics, which is different enough that it often provides the key to understanding. In it, he shows that in *any* potential, the distribution of states is proportional to $\exp(-\Delta E/kT)$, where $\Delta E$ is the difference in energy from some arbitrary level (it's the ratio that matters, so we don't need to specify the reference level). There may be other factors (e.g., the multiplicity of the state), but this basic one is always obtained in a classical equilibrium. Another example is the Maxwell-Boltzmann distribution. At temperature $T$, the fraction in a volume $d^3p$ around momentum $\mathbf{p}$ is $\propto \exp(-E/kT)d^3p$, which (for nonrelativistic motion and isotropic distribution in momentum space) is $\propto \exp(mv^2/2kT)p^2dp$. Yet another example

is the population of states in an atom. Compared to some reference state (e.g., the ground state), the fraction of atoms in a state of energy $E$ is $\propto \exp(-E/kT)$.

Now let's return to the subject of ionization. This is important because the ionization fraction has consequences for the equation of state (pressure related to density and temperature) and opacity (e.g., which sources matter most), among other things. Let us assume that we have a cavity in which atoms and radiation are in thermal equilibrium with the same temperature $T$. For simplicity, consider a hydrogen atom with a single energy level (ground state), and consider the ionization reaction

$$H^+ + e^- \longleftrightarrow H^0 + \chi_H \,, \tag{1}$$
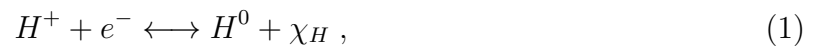
where $\chi_H$=13.6 eV is the ground state binding energy. In equilibrium, what can be said about the abundances? In this case, let the reference energy ($E = 0$) be for the free electron and proton. Statistical weights: 2 for the hydrogen atom, two combined for the proton and electron (spins in same direction, spins in opposite directions). From the previous lecture notes, integration of the distribution function over the momenta yields the number densities

$$n_e = \frac{2[2\pi m_e kT]^{3/2}}{h^3} e^{\mu^-/kT} \,, \tag{2}$$

$$n^+ = \frac{[2\pi m_p kT]^{3/2}}{h^3} e^{\mu^+/kT} \,, \tag{3}$$

and

$$n^0 = \frac{2[2\pi(m_p + m_e)kT]^{3/2}}{h^3} e^{\mu^0/kT} e^{\chi_H/kT} \,. \tag{4}$$

**Ask class**: if we were being *really* accurate, would the actual mass in the third equation be $m_p + m_e$? (No, because we'd have to subtract the binding energy. But this is only 13.6/511,000 of even the electron mass, which would be a 1/40,000 correction, and can be dumped. In the same spirit, the electron mass is only a 1/1,800 correction and can be ignored here.)

Forming the product $n^+ n_e/n^0$, we find

$$\frac{n^+ n_e}{n^0} = \left(\frac{2\pi m_e kT}{h^2}\right)^{3/2} e^{(\mu^- + \mu^+ - \mu_0)/kT} e^{-\chi_H/kT} \,. \tag{5}$$

**Ask class**: what is $(\mu^- + \mu^+ - \mu_0)$ in equilibrium, which we've been assuming? It's zero. The numerical factor is about $2.4 \times 10^{15} T^{3/2}$ cm$^{-3}$. **Ask class**: is this equation sufficient to determine the abundances of electrons, protons, and neutral hydrogen? No, because we have only one equation for three unknowns. **Ask class**: what are reasonable auxilliary conditions? Charge balance implies $n^+ = n^-$, and conservation of nucleons means that $n^0 + n^+ = n$, the total number density of protons (which equals the total number density of electrons). With

these extra conditions, the degree of ionization $y \equiv n^+/n = n_e/n$ is, numerically,

$$\frac{y^2}{1-y} = \frac{4 \times 10^{-9}}{\rho} T^{3/2} e^{-1.6 \times 10^5/T} \,, \tag{6}$$

where as usual quantities such as $\rho$ and $T$ are referenced to the appropriate cgs unit.

An important thing to remember about this equation and many like it is that the exponential usually dominates. What this means specifically is that for "interesting" levels of ionization (say, around 50%), $T$ is so much lower than $1.6 \times 10^5$ K that a tiny change in $T$ makes a huge difference in the ionization fraction. Operationally, this means that you can get a remarkably good first guess at the temperature for a specified value of $y$ by *ignoring* the $T^{3/2}$ factor, which doesn't change as rapidly. If you like, you can then substitute your value of $T$ in the $T^{3/2}$ factor and solve again by iteration. This is an easy way to converge rapidly to a solution without having to write a computer program.

In fact, let's do a practice example. Suppose we take a number density of 1 cm$^{-3}$, which is about average for our interstellar medium. Then $\rho = 1.7 \times 10^{-24}$ g cm$^{-3}$, so the Saha equation becomes

$$\frac{y^2}{1-y} \approx 2.4 \times 10^{15} T^{3/2} e^{-1.6 \times 10^5/T} \,. \tag{7}$$

We would like to solve for the temperature where $y = 0.5$, which is about the halfway ionization point. Let us also say that we only need the answer to within a factor of 3. Thus we get

$$T^{3/2} e^{-1.6 \times 10^5/T} \approx 2 \times 10^{-16} \,, \tag{8}$$

where $T$ is in Kelvin. Getting an exact solution would require a computer, but the fact that the right hand side is so small gives us an easy way to get an approximate answer. Let us first ignore the $T^{3/2}$ factor. Then we can use the approximate equation

$$e^{-1.6 \times 10^5/T} \approx 2 \times 10^{-16} \,. \tag{9}$$

Since $e^{10} \approx 2 \times 10^4$, if the exponent is about -40 then we get something in the vicinity of $10^{-16}$. This would suggest $T = 4000$ K. Using that value, $T^{3/2}$ is about $4^{3/2}(10^3)^{3/2} \approx 8 \times 3 \times 10^4 \approx 2 \times 10^5$, so dividing through we would need $e^{-1.6 \times 10^5/T} \approx 10^{-21}$, or about -50 for our exponent. This would suggest $T \approx 3200$ K, which is already within our factor of 3. 3300 K is an excellent approximation. Behold the power of exponentials! It should be said, by the way, that in reality the warm neutral medium has a temperature of about 8000 K. This is an indication that one or more of our assumptions is not correct. That assumption is that the radiation field is in thermal equilibrium with the matter. In fact, the ionizing radiation field is provided mainly by UV from hot stars, rather than a constant thermal bath. Thus even though we did not get quite the right answer, (1) we did get within a factor of a few, and (2) the fact that comparison with observations shows we're not quite there tells us that we need to know something extra.

**Ask class**: what does the Saha equation imply about the degree of ionization for a fixed temperature but varying density? (degree of ionization goes up when density goes down). Why is that? This is thermal ionization, so you can imagine some radiation field pervading the region. The ionization rate per volume, therefore, just goes like the number density. However, the recombination rate involves two particles, and hence must go as the product of their densities. Therefore, the recombination rate increases more rapidly when the density goes up, and the resulting ionization fraction goes down.

Let's test how well this equation works in practice. **Ask class to compute ionization fraction in the center of the Sun**. Suppose **Ask class** $T = 1.5 \times 10^7$ K and $\rho = 100$ g cm$^{-3}$, for round numbers. Then $y^2/(1-y) = 2.3$. This has a solution of $y = 0.75$, so the center of the Sun is 25% neutral (!). **See if this poses a problem for anyone in class!**. This is nonsense, of course! The problem is that at such densities, the separation between atoms is much less than the size of an individual free atom, so the electrons get squeezed up to continuum states and *pressure* ionization dominates. In reality, the center of the Sun has fully ionized hydrogen. As Shu points out, this also means that when you sometimes hear people say that the full ionization is because of the high temperature, that's not completely accurate.

Okay, so let's return to the Saha equation. We derived it by assuming that hydrogen atoms are single-level systems (either ground state or ionized). 'Tain't so in reality. There is, in principle, an infinity of excited bound states. Each of these contributes a term that is like the multiplicity of the state times $\exp(-\Delta E/kT)$ times the original term, where $\Delta E$ is the energy with respect to the energy of the ground state. But wait! This means we have an infinite sum of finite terms. **Ask class:** how do we get around this? In reality, the size of the orbital increases like $n^2$, so in most circumstances only the first term needs to be kept, as we did.

The Saha equation is important enough that I'd like to come at it from another perspective, given in Shu (Chapter 7, pgs. 64-65). Let's define the thermal de Broglie wavelength of the electron

$$\lambda_{Te} \equiv h/(2\pi m_e kT)^{1/2} \ . \tag{10}$$

Then the Saha equation between ionization states $i$ and $i+1$ can be written

$$\frac{n_{i+1}}{n_i} = \frac{2Z_{i+1}}{(n_e\lambda_{Te}^3)Z_i} \exp(-I_i/kT) \ . \tag{11}$$

Here $I_i$ is the ionization energy from state $i$ to state $i+1$ and $Z_i$ and $Z_{i+1}$ are partition functions (these are sums of the multiplicity of each atomic level times $\exp(-\Delta E/kT)$). Even better, we note that the chemical potential for nondegenerate free electrons is $-\mu_e = kT\xi$, where $\xi \equiv \ln(2/n_e\lambda_{Te}^3)$. Note that $n_e\lambda_{Te}^3$ is the number of electrons in a cubic thermal de Broglie wavelength, which is really tiny if our assumption of nondegeneracy works. Typically

$\xi \sim 10 - 30$ (remember, "all logarithms are 10"!). Then we rewrite the Saha equation as

$$\frac{n_{i+1}}{n_i} = \frac{Z_{i+1}}{Z_i} e^{-(\mu_e + I_i)/kT} \ . \tag{12}$$

What does this tell us? The ratio of partition functions is usually of order unity, and as before the exponential will dominate. This means that the two states are populated equally roughly when $-\mu_e = I_i$. This implies that the thermal energy at this transition temperature satisfies $kT \approx I_i/\xi$, which is a good order of magnitude less than one's first guess $kT \approx I_i$. Why the difference? You can think of it like this: at a temperature $kT = I_i/\xi$ there is a substantial energy penalty to be paid by ionizing an electron, but the electron then has lots more phase space to run around in. There are some additional consequences. First, the transition between ionization states is usually pretty sharp, on the order of $\Delta T \approx kT/\xi$. Therefore, if you fix the density and look at the state that is most populated, it changes abruptly with the temperature (and there is usually one state that dominates). Second, this means that it is in some sense "easier" to ionize an atom than to raise it to an excited bound state (since that really does need $kT \approx I_i$, if the excitation energy is roughly $I_i$ as it is in hydrogen).

The Saha equation is useful in lots of circumstances, but now we need to examine the assumptions that go into it. **Ask class:** what are two circumstances under which the Saha equation will give incorrect ionization fractions? One we just saw: when the density is high enough to infringe on the atoms, thermal ionization isn't as relevant as pressure ionization. **Ask class:** what would this mean, physically, about the size of atoms and their separations? Qualitatively, if the separation between atoms is comparable to the size of their electron clouds, one would expect significant extra ionization. This is the same thing we used to argue that we could cut off the partition function after a finite number of terms.

**Ask class:** what is another way in which the Saha equation could be invalid? If there is any other source of ionization, it could modify the ionization fraction. **Ask class:** Examples? Gas in a molecular cloud at 100 K that is being illuminated by UV from a hot star has a much different ionization structure than if it were not illuminated. Cosmic rays help keep protoplanetary disks far more ionized than they would be in thermal equilibrium at their temperatures. In fact, this may be crucial to the birth of stars. There is strong evidence that a nonnegligible level of ionization is crucial to allowing angular momentum transport in protoplanetary disks, so without the influence of cosmic rays the matter would just pile up until something catastrophic occurred.

**Ask class:** what is yet another way in which the ionization fraction could be different from the Saha prediction? The ionization fraction could in principle be anything at all if there hasn't been enough time to establish equilibrium. **Ask class:** is this more likely in low-density or high-density environments? Low, because interaction rates are less. An example is the very early universe (prior to any star formation), where there is a residual ionization

fraction because the density drops due to universal expansion. **Ask class:** how, then, would one compute the ionization fraction? Your only recourse is to have some initial conditions, then make calculations based on the fundamental rate equations rather than being able to assume equilibrium. As always, equilibrium is a great simplifier, but you have to be able to justify that it is valid.