Class 5 Notes: The 21 cm line

Now that we have gotten the basics of radiative transfer and radiation-matter interaction, we are ready to examine the most important astronomical applications of these ideas. This will occupy us for the next three classes. This class is devoted to the hydrogen 21 cm line, the next to absorption lines and absorption spectroscopy, and the third to the continuum emission processes that occur in a plasma.

I. Physics of the 21 cm line

We will begin with what is likely the simplest emission and absorption system, the 21 cm line of neutral hydrogen. To remind you, this line arises from the hyperfine splitting of the ground electronic state of hydrogen. Not only is this the simplest case, it is one of the most important for studying the ISM.

A. Emissivity and attenuation coefficient

The 21 cm line occurs because in the ground electronic state of neutral hydrogen, the state with the proton and electron spins parallel differs in energy from the state with them antiparallel by $E_{u\ell} = 5.87 \times 10^{-6}$ eV. The anti-parallel lower energy state has total spin $S = 0$, so its degeneracy is $g_\ell = 2S + 1 = 1$. The upper state has total spin $S = 1$, so $g_u = 3$. The Einstein spontaneous emission coefficient is $A_{u\ell} = 2.8843 \times 10^{-15}$ s^{-1}, or 1/(11.0 Myr).

The energy difference between the two state is corresponds to a temperature $T_{u\ell} = E_{u\ell}/k = 0.0682$ K. The excitation temperature, also known as the spin temperature for this particular transition, will depend on the region, but under any normal circumstances it cannot be lower than the CMB temperature, since all interstellar hydrogen has had an essentially infinite amount of time to be heated by CMB photons. Thus $T_{\rm spin} > T_{\rm CMB} = 2.72$ K. Since $T_{\rm spin} \gg T_{u\ell}$, the Boltzmann factor $e^{-E_{u\ell}/kT_{\rm spin}} \approx 1$, and the levels are populated simply in proportion to their degeneracy:

$$\frac{n_u}{n_\ell} = \frac{g_u}{g_\ell} e^{-E_{u\ell}/k_{\rm B}T_{\rm spin}} \approx 3. \tag{1}$$

Thus at any given time we expect 3/4 of atoms to be in the upper hyperfine state, and 1/4 to be in the lower state. This has the important implication that the emissivity is independent of temperature:

$$j_\nu = \frac{h\nu_{u\ell}}{4\pi} n_u A_{u\ell} \phi_\nu \approx \frac{3}{16\pi} h\nu_{u\ell} n_{\rm H} A_{u\ell} \phi_\nu, \tag{2}$$

where $n_{\rm H}$ is the total number density of atomic hydrogen.

The attenuation coefficient is similarly easy to compute. Recall that we showed

$$\kappa_\nu = \frac{h\nu}{4\pi} n_\ell B_{\ell u} \phi_\nu \left(1 - e^{-E_{u\ell}/k_{\rm B}T_{\rm spin}}\right) = \frac{3}{8\pi} n_\ell A_{u\ell} \lambda_{u\ell}^2 \phi_\nu \left(1 - e^{-E_{u\ell}/k_{\rm B}T_{\rm spin}}\right), \tag{3}$$

where $\lambda_{u\ell} = c/\nu_{u\ell}$, and in the last step we used the relationship between the Einstein coefficients to rewrite everything in terms of $A_{u\ell}$. Note that, since $E_{u\ell}/kT_{\rm spin} \ll 1$, the term in parentheses is nearly zero: stimulated emissions almost exactly balance absorptions. To find the difference, we can Taylor expand the exponential. Doing so and also setting $n_\ell = n_{\rm H}/4$, we find

$$\kappa_\nu \approx \frac{3}{32\pi} A_{u\ell} \frac{hc\lambda_{u\ell}}{k_{\rm B} T_{\rm spin}} n_{\rm H} \phi_\nu. \tag{4}$$

To get a sense of the level of opacity, it is helpful to compute the optical depth of a slab of hydrogen of length L, which is simply

$$\tau_\nu = \kappa_\nu L = \frac{3}{32\pi} A_{u\ell} \frac{hc\lambda_{u\ell}}{k_{\rm B} T_{\rm spin}} N_{\rm H} \phi_\nu, \tag{5}$$

where $N_{\rm H} = n_{\rm H} L$ is the column density of H atoms. If we take the line profile function to be a Gaussian (which is usually a good approximation, since the column density is almost never high enough to see the damping wings), then we have

$$\phi_\nu = \frac{1}{\sqrt{2\pi}} \frac{\lambda_{u\ell}}{\sigma_V} e^{-v^2/2\sigma_v^2}, \tag{6}$$

where σ_v is the velocity dispersion. Plugging this in and putting in some numerical values typical of the ISM in the Galaxy, we have

$$\tau_\nu = 2.19 \left(\frac{N_{\rm H}}{10^{21}\ {\rm cm}^{-2}}\right) \left(\frac{100\ {\rm K}}{T_{\rm spin}}\right) \left(\frac{{\rm km\ s}^{-1}}{\sigma_v}\right) e^{-v^2/2\sigma_v^2}. \tag{7}$$

Thus we see that the 21 cm line can be marginally optically thick along typical sightlines through the ISM of the Milky Way and similar galaxies in places where the gas is cold, with a spin temperature ~ 100 K. Fortunately for us we will see that much of the H I mass of the Milky Way and other galaxies is actually in a warm component with $T_{\rm spin}$ of several thousand K and $\sigma_v \sim 10$ km s^{-1}, and this gas is optically thin. Self-absorption by H I is a concern only for the cold, low velocity dispersion component.

B. Optically thin emission

If we observe an optically thin object in the 21 cm line, we can use this theoretical treatment to compute its mass from the 21 cm flux that we observe. This calculation is the basis for almost all estimates of the atomic gas masses of galaxies. For an optically thin line of sight, we can ignore absorption, and the transfer equation is trivial to integrate:

$$\int dI_\nu = \int j_\nu \, ds \quad \Longrightarrow I_\nu = I_\nu(0) + \frac{3}{16\pi} A_{u\ell} h\nu_{u\ell} \phi_\nu N_{\rm H}, \tag{8}$$

where $N_{\rm H} = \int n_{\rm H}\, ds$. If we are looking at a line of sight where there is no background object, then the background intensity $I_\nu(0)$ is simply the intensity of the CMB, which we can compute from the Planck function. We know that the integral of ϕ_ν is unity, so if we integrate over frequency then we have

$$N_{\mathrm{H}} = \frac{16\pi}{3A_{u\ell}h\nu_{u\ell}} \int [I_\nu - I_\nu(0)]\, d\nu. \tag{9}$$

Thus, given a measured I_ν, we can solve for N_{H}. In practice the background intensity is generally small, so we can drop the $I_\nu(0)$ term.

Since this is a radio observation, it is common to work with antenna temperature instead of intensity, and with velocity instead of frequency. We define the antenna temperature in the usual way, $T_A = (c^2/2k\nu^2)I_\nu$, and we relate velocity to frequency via the Doppler shift: $\nu = \nu_{u\ell}(1 - v/c)$. With these definitions, we can define the velocity-integrated antenna temperature by

$$\int [T_A - T_A(0)]\, dv = \int \frac{c^2}{2k\nu^2}[I_\nu - I_\nu(0)]\frac{c}{\nu}d\nu = \frac{3}{32\pi}\frac{hc\lambda_{u\ell}^2}{k}A_{u\ell}N_{\mathrm{H}} \tag{10}$$

Plugging in the constants, we have

$$\int [T_A - T_A(0)]\, dv = 54.89 \text{ K km s}^{-1} \left(\frac{N_{\mathrm{H}}}{10^{20}\text{ cm}^{-2}}\right). \tag{11}$$

Thus for optically thin H I, we have a direct method of measuring its column density – simply point a radio antenna tuned to 21 cm off the source to measure the background antenna temperature $T_A(0)$, then point it at the source and measure T_A. Do this for a range in frequencies / velocities and add, and the result immediately gives you the H I column density.

If the emitting object is unresolved, by measuring the total flux inside a radio beam we can measure the mass, at least for an object of known distance. To figure out the total flux observed in a radio beam, we need to integrate the intensity over the unresolved solid angle occupied by the object. The total frequency-integrated flux observed is

$$F = \int I_\nu\, d\nu\, d\Omega = \int I_\nu\, d\nu\, \frac{dA}{D^2} \tag{12}$$

where we understand that the intensity I_ν is a function of position on the sky. In the second step we have taken $d\Omega = dA/D^2$, where D is the distance to the object and dA is the area element on the emitting object that subtends the solid angle $d\Omega$. (If the object is at cosmological distance D should be replaced by the luminosity distance D_L, which accounts for relativistic effects.) Dropping the background term for simplicity, we can substitute to obtain

$$F = \frac{3A_{u\ell}h\nu_{u\ell}}{16\pi D^2}\int N_{\mathrm{H}}\, dA = \frac{3A_{u\ell}h\nu_{u\ell}}{16\pi D^2}\frac{M_{\mathrm{H}}}{m_{\mathrm{H}}}, \tag{13}$$

where m_{H} is the mass of a hydrogen atom and $M_{\mathrm{H}} = m_{\mathrm{H}} \int N_{\mathrm{H}}\, dA$ is the total hydrogen mass of the emitting object. Plugging in and solving, we have

$$M_H = \frac{16\pi m_{\mathrm{H}}}{3A_{u\ell}h\nu_{u\ell}} D^2 F = 4.95 \times 10^7\, M_\odot \left(\frac{D}{\text{Mpc}}\right)^2 \left(\frac{F}{\text{Jy MHz}}\right), \tag{14}$$

where 1 Jy = 10^{-23} erg s^{-1} cm^{-2} Hz^{-1} is the standard unit of flux used in radio astronomy.

II. Absorption in the 21 cm line

We can learn considerably more from the 21 cm line if the background intensity is not tiny, and instead we have a relatively bright background source. There are two prominent cases where this situation is realised: when we see neutral hydrogen in the foreground against a background radio quasar, and when we cold H I in absorption as a foreground in front of brighter warm H I in the background.

A. 21 cm absorption

This generally occurs when a bright quasar sits behind a cloud of H I. Since quasars usually occupy a small angular extent, we can often observe the same object on two parallel but slightly offset lines of sight, one of which hits the background quasar and one of which does not. Since the separation of the two lines of sight is small, we assume that the foreground H I cloud is the same along the two lines of sight. We refer to the line of sight that hits the quasar as on-source, and the one that does not as off-source.

This is the problem of radiative transfer through a uniform-temperature slab that we solved a while ago. Recall that the solution is

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + B_\nu(T_{\rm spin})(1 - e^{-\tau_\nu}). \tag{15}$$

Since $k_{\rm B}T_{\rm spin} \gg h\nu$ for the 21 cm line, we can simplify this by Taylor-expanding the Planck function:

$$B_\nu(T_{\rm spin}) = \frac{2h\nu^3}{c^2}\frac{1}{\exp(h\nu/kT_{\rm spin}) - 1} \approx \frac{2k\nu^2}{c^2}T_{\rm spin} \tag{16}$$

Plugging this into the intensity equation and converting to antenna temperature, we obtain

$$T_A(\tau_\nu) = T_A(0)e^{-\tau_\nu} + T_{\rm spin}(1 - e^{-\tau_\nu}). \tag{17}$$

Although we have suppressed the subscript, recall that T_A does still depend on frequency / velocity.

Applying this equation to the two lines of sight, we have

$$T_A^{\rm on} = T_{\rm QSO}e^{-\tau_\nu} + T_{\rm spin}(1 - e^{-\tau_\nu}) \tag{18}$$

$$T_A^{\rm off} = T_{\rm sky}e^{-\tau_\nu} + T_{\rm spin}(1 - e^{-\tau_\nu}), \tag{19}$$

where $T_{\rm QSO}$ is the antenna temperature of the QSO, and $T_{\rm sky}$ is that of the blank sky. We can solve these two equations for $T_{\rm spin}$ and τ_ν:

$$\tau_\nu = \ln\left(\frac{T_{\rm QSO} - T_{\rm sky}}{T_A^{\rm on} - T_A^{\rm off}}\right) \tag{20}$$

$$T_{\rm spin} = \frac{T_A^{\rm off}T_{\rm QSO} - T_A^{\rm on}T_{\rm sky}}{(T_{\rm QSO} - T_{\rm sky}) - (T_A^{\rm on} - T_A^{\rm off})} \tag{21}$$

We can usually measure $T_{\rm QSO}$ by measuring the intensity of emission from the QSO on either side of the 21 cm line, i.e. far from the absorption feature, and fitting a line between the two sides. We can measure $T_{\rm sky}$ by measuring the blank sky away from the hydrogen cloud. Thus every quantity on the right hand side is directly measurable, and we can simultaneously solve for the optical depth and the spin temperature.

B. H I self-absorption (HISA)

As we have shown, the optically thin limit applies when the temperature is high and/or the velocity dispersion is large. For the CNM, however, velocity dispersion can be < 1 km s^{-1} and temperatures can be below 100 K, and in this case the H I becomes optically thick. In this case the cloud is likely to be very dim in emission. To see why, compare two clouds of equal column density $N_{\rm H}$, one consisting of optically thick cold gas at temperature T_c, and the other consisting of optically thin warm gas at temperature T_w. The transfer equation for both clouds is the same:

$$T_A(\tau_\nu) = T_A(0)e^{-\tau_\nu} + T_{\rm spin}(1 - e^{-\tau_\nu}). \tag{22}$$

For $\tau_\nu \gg 1$, clearly $T_A(\tau_\nu) \to T_{\rm spin} = T_c$. Thus the antenna temperature will simply be T_c at line center, and will fall off at frequencies far enough from line center so that the cloud is optically thin.

For the warm optically thin cloud we have $\tau_\nu \ll 1$, and we can series expand the exponentials. Assuming $T_A(0) = 0$, i.e. there is no background source behind the warm cloud, we obtain

$$T_A(\tau_\nu) = T_w \tau_\nu. \tag{23}$$

Plugging into our earlier formula for τ_ν, we have

$$T_A = \frac{3}{32\pi} A_{u\ell} \frac{hc\lambda_{u\ell}}{k} N_H \phi_\nu = 220\ {\rm K} \left(\frac{N_{\rm H}}{10^{21}\ {\rm cm}^{-2}}\right) \left(\frac{{\rm km\ s}^{-1}}{\sigma_v}\right) e^{-v^2/2\sigma_v^2}. \tag{24}$$

Notice that the result is independent of T_w, consistent with our earlier result that optically thin emission depends only on the total column of H I, not on its temperature.

The problem with observing cold H I in emission now becomes clear. Typical values for T_c are at most 100 K, and are often lower, producing a brightness temperature of no more than that. In contrast, even a relatively modest column of warm, optically thin H I will produce a brightness temperature twice as high. Thus it is often hard to see cold H I in emission.

However, it is possible to see it in absorption against warm H I behind the cloud. Consider a background of warm H I with a small cold cloud in front of it, and, as in the QSO case, consider two lines of sight – one passing through the cold cloud and one not. Assume that the background warm gas is relatively smooth, so the column density (and thus the antenna temperature) of the warm background gas is the same along the two lines of sight. In this case along the line of sight that misses the cold cloud, we see an antenna temperature as we computed above for pure warm H I:

$$T_A^{\rm off} = 220\ {\rm K}\left(\frac{N_{\rm H}}{10^{21}\ {\rm cm}^{-2}}\right)\left(\frac{\rm km\ s^{-1}}{\sigma_v}\right)e^{-v^2/2\sigma_v^2}. \tag{25}$$

Along the line of sight that hits the cold cloud, we have $T_A(0) = T_A^{\rm off}$, so the resulting antenna temperature is

$$T_A^{\rm on} = T_A^{\rm off}e^{-\tau_\nu} + T_c(1 - e^{-\tau_\nu}). \tag{26}$$

Since $T_c < T_A^{\rm off}$, the result is an absorption feature. This is known as H I self-absorption (HISA).

Unfortunately it is not easy to use HISA measurements to constrain cold H I masses. The most favorable situation is when T_c is known. This can be from observations of some other species that is co-located with the cold H I, for example. In this case one can observe $T_A^{\rm on}$ and $T_A^{\rm off}$ and solve for τ_ν and thence for $N_{\rm H}$ for the cold cloud.

III. 21 cm cosmology

The final application of the 21 cm line that we will consider today is in cosmological contexts, where substantial efforts are underway, in Australia and elsewhere, to measure redshifted 21 cm radiation from neutral hydrogen in the early universe. In some ways this analogous to the cases we have just considered, in that it is a problem of H I absorption against a backlight. In this case, however, the backlight is the CMB, and we exploit the fact that the 21 cm line is spread out in frequency by the expansion of the Universe to probe different epochs in cosmic history.

A. The 21 cm optical depth

The starting basis for 21 cm cosmology is the fundamental equation we have already used repeatedly:

$$T_A = T_{\rm bg}e^{-\tau_\nu} + T_{\rm spin}(1 - e^{-\tau_\nu}), \tag{27}$$

where the background temperature $T_{\rm bg}$ is the temperature of the the CMB. The first step in working out how 21 cm absorption behaves is figuring out the optical depth τ_ν, which, as always is given by

$$\tau_\nu = \frac{3}{32\pi}A_{u\ell}\frac{hc\lambda_{u\ell}}{k_{\rm B}T_{\rm spin}}N_{\rm H}\phi_\nu. \tag{28}$$

The question is, in the cosmological context, what are the column density $N_{\rm H}$ and the line shape function ϕ_ν?

To figure this out, we need to think a bit about Doppler shifting and cosmology. Hydrogen atoms always absorb radiation at a wavelength of 21 cm in their rest frame, but when this radiation reaches us it will have been redshifted to a wavelength that is longer by a factor of $(1+z)$, where z is the redshift. Thus we can relate the observed frequency ν to the redshift at which that absorption was generated via

$$\nu = \frac{\nu_0}{1+z}, \tag{29}$$

where $\nu_0 = 1.42$ GHz is the rest frequency of the line. This means that, if we observe at frequency ν, then we must care about the density of hydrogen $n_\mathrm{H}(z)$ at $z = (\nu_0/\nu) - 1$.

To deal with the column density and the line shape function, we can make what is known as the Sobolev, or large velocity gradient, approximation. The basic idea is as follows: we had previously computed the line shape function by assuming that the gas has a thermal velocity distribution, but in the cosmological context (and in some other places), this is not the case, because there are bulk velocities that greatly exceed the sound speed. We will therefore make the opposite approximation, and neglect the thermal velocity in comparison to the bulk flow.

In the case of cosmology, the main bulk flow that we have to worry about is just the Hubble flow. We can therefore reason as follows. Suppose we focus on the redshift z that corresponds to our chosen observing frequency, and then consider gas slightly further away, by a distance ds. This gas will have a velocity that is larger by an amount $dv = H(z)\,ds$, where $H(z)$ is the Hubble constant at redshift z. In turn, this means that its absorption will be Doppler shifted to a frequency that is lower by an amount $d\nu = (\nu/c)H(z)\,ds$. This is what we need to compute the line shape function, because recall that the line shape function is supposed to tell us how much of the column is absorbing or emitting per unit frequency interval $d\nu$. Thus we can approximate the product of the column density and the line shape function as

$$N_\mathrm{H}\phi_\nu \approx \frac{n_\mathrm{H}(z)\,ds}{(\nu/c)H(z)\,ds} = \frac{n_\mathrm{H}(z)c}{\nu H(z)} \tag{30}$$

In practice τ_ν is always small, so we can Taylor expand and write out an expression for the change in temperature that we will observe relative to the CMB, ΔT. This is

$$\Delta T \approx \frac{T_\mathrm{spin} - T_\mathrm{bg}}{1+z}\tau_\nu, \tag{31}$$

where the $1+z$ factor comes from the fact that the temperature shift will be redshifted by this factor, i.e., if the temperature shift is 1 K at redshift $z = 9$, this will be reduced to 0.1 K by the time we observe it at $z = 0$. Putting this all together, and substituting in some cosmological variables for the density and $H(z)$, we can write the temperature shift as

$$\Delta T \approx 25\text{ mK}\; x_\mathrm{HI} \left(\frac{\Omega_b h}{0.03}\right)\left(\frac{\Omega_m}{0.3}\right)^{-1/2}\left(\frac{1+z}{10}\right)^{1/2}\frac{T_\mathrm{spin} - T_\mathrm{bg}}{T_\mathrm{spin}}, \tag{32}$$

where $x_{\rm HI}$ is the fraction of hydrogen that is neutral, Ω_b is the baryon density of the universe, Ω_m is the matter density of the universe, and h is the Hubble constant at $z = 0$ scaled to 100 km s^{-1} Mpc^{-1}. Thus a measurement of the temperature shift relative to the CMB tells us about the abundance of neutral hydrogen and the evolution of the spin temperature.

B. What sets the spin temperature

The 21 cm cosmological signal has yet to be observed, but we can write down some qualitative expectations as to what it will look like. At redshifts $\lesssim 7-10$, we expect the signal to vanish, because the IGM is mostly ionised, so $x_{\rm HI} \ll 1$. At higher redshift, on the other hand, $x_{\rm HI} \approx 1$, and the signal will be determined by the evolution of the spin temperature, since everything else in the expression is constant. Note that, at high redshift, we cannot just assume $T_{\rm spin} \gg T_{\rm bg}$, since the CMB temperature may be quite high: $T_{\rm bg} \propto 1 + z$.

The spin temperature in turn evolves in response to three main effects: interactions of H I with the CMB, collisions of H I with other H I and with free electrons, and interactions with Lyman α photons. Thus if we let n_u be the number density of HI atoms in the upper state, we can write the condition of equilibrium schematically as

$$\left(\frac{dn_u}{dt}\right)_{\rm CMB} + \left(\frac{dn_u}{dt}\right)_{\rm coll} + \left(\frac{dn_u}{dt}\right)_{\rm Ly\alpha} = 0 \tag{33}$$

Let us now try to write down each of these rates, starting with the CMB. We have shown previously that the rate of change in the upper state density n_u due to interaction with a background radiation field can be written as

$$\left(\frac{dn_u}{dt}\right)_{\rm CMB} = -n_u A_{u\ell} - n_u \langle n_\gamma \rangle A_{u\ell} + \frac{g_u}{g_\ell} n_\ell \langle n_\gamma \rangle A_{u\ell} = A_{u\ell}\left[\frac{g_u}{g_\ell} n_\ell - \left(1 + \langle n_\gamma \rangle\right) n_u\right] \tag{34}$$

It is convenient to rewrite this expression in terms of the spin temperature and the CMB temperature. From our definition of the spin temperature, which is implicitly given by $n_u/n_\ell = (g_u/g_\ell)e^{-E_{u\ell}/k_B T_{\rm spin}}$, we can write

$$n_\ell = \frac{g_\ell}{Z(T_{\rm spin})} n_{\rm H} \qquad n_u = \frac{g_u e^{-E_{u\ell}/k_{\rm B} T_{\rm spin}}}{Z(T_{\rm spin})} n_{\rm H} \qquad Z(T_{\rm spin}) = g_\ell + g_u e^{-E_{u\ell}/k_{\rm B} T_{\rm spin}}, \tag{35}$$

where $n_{\rm H}$ is the total density of H. Similarly, since the CMB is a blackbody, we can write out the direction-averaged photon occupation number $\langle n_\gamma \rangle$ very simply:

$$\langle n_\gamma \rangle = \frac{c^2}{2h\nu_{u\ell}^3} B_\nu(T_{\rm CMB}) = \frac{1}{\exp\left(E_{u\ell}/kT_{\rm CMB}\right) - 1} \tag{36}$$

If we substitute these two expressions in, and simplify a bit by Taylor expanding the exponentials in the limit $E_{u\ell} \ll k_{\rm B} T_{\rm spin}$ and $E_{\rm u\ell} \ll k_{\rm B} T_{\rm CMB}$ (both of which are certainly the case), then we get

$$\left(\frac{dn_u}{dt}\right)_{\rm CMB} = A_{u\ell} n_{\rm H} \left(\frac{g_u}{g_\ell + g_u}\right) \frac{T_{\rm CMB} - T_{\rm spin}}{T_{\rm spin}} \tag{37}$$

Note that this gives the result we expect: if interaction with the CMB were the only process occurring, so this term had to be zero, then the solution is that $T_{\rm spin} = T_{\rm CMB}$.

Next consider collisions; the rate at which the density of H atoms in the upper state changes due to collisions with some species i (where here the significant species are H, e^-, and p) is

$$\left(\frac{dn_u}{dt}\right)_{\rm coll} = -k_{u\ell} n_i n_u + k_{\ell u} n_i n_\ell, \tag{38}$$

where the first term represents collisional de-excitations out of state u to state ℓ, and the second term represents the reverse process. Our first step here is to rewrite the excitation coefficient $k_{\ell u}$ in terms of the de-excitation rate coefficient $k_{u\ell}$, using the relationship we proved from detailed balance:

$$k_{\ell u} = \frac{g_u}{g_\ell} k_{u\ell} e^{-E_{u\ell}/k_{\rm B} T_K}. \tag{39}$$

Note that the temperature that appears here is the kinetic temperature T_K, which describes the distribution of particle velocities – this is *not* necessarily the same as the spin temperature or the CMB temperature! If we substitute this in, rewrite n_u and n_ℓ in terms of $T_{\rm spin}$ and $n_{\rm H}$ as before, and again expand in the limit of small $E_{u\ell}$, with a bit of algebra we get

$$\left(\frac{dn_u}{dt}\right)_{\rm coll} = k_{u\ell} n_i n_{\rm H} \frac{g_u}{g_\ell + g_u} \left(\frac{E_{u\ell}}{k_{\rm B} T_K}\right) \frac{T_K - T_{\rm spin}}{T_{\rm spin}} \tag{40}$$

Note how similar this is to the functional form for coupling to the CMB. Indeed, it is informative to write down the solution if this and the CMB are the only significant effects. In this case we have

$$A_{u\ell} n_{\rm H} \frac{g_u}{g_\ell + g_u} \frac{T_{\rm CMB} - T_{\rm spin}}{T_{\rm spin}} + \sum_i k_{u\ell,i} n_i n_{\rm H} \frac{g_u}{g_\ell + g_u} \left(\frac{E_{u\ell}}{k_{\rm B} T_K}\right) \frac{T_K - T_{\rm spin}}{T_{\rm spin}} = 0, \tag{41}$$

where the sum runs over all species i with which collisions occur. Clearly the degeneracy factors and $n_{\rm H}$ drop out, and the solution is

$$T_{\rm spin} = \frac{1 + x_c}{T_{\rm CMB}^{-1} + x_c T_K^{-1}} \tag{42}$$

where $x_c = \sum_i k_{u\ell,i} n_i E_{u\ell} / A_{u\ell} k_{\rm B} T_{\rm CMB}$. We can see that, for $x_c \ll 1$, the solution approaches $T_{\rm CMB}$, while for $x_c \gg 1$ it approaches T_K; thus the parameter x_c

tells us whether collisions or the CMB are more important in setting the spin temperature.

The final effect is a bit more subtle. The underlying physical process is as follows: Lyman α photons can excite H atoms from the $n = 1$ state to the $n = 2$ state. When the atoms then de-excite by emitting a Lyman α photon, they are not guaranteed to go back into the same spin state from which they started. Thus if there are a large number of Lyman α photons around, then can induce changes in the spin temperature. This phenomenon is known as the Wouthuysen-Field Effect.

It turns out that the spin temperature towards which this drives the H atoms is determined by the colour temperature of the Lyman α photons, where the colour temperature here is defined as the ratio of number of photons at the very slightly different energies corresponding to the different transitions between the two spin-substates. When this effect is included, the solution we have written down is modified to

$$T_{\text{spin}} = \frac{1 + x_c + x_\alpha}{T_{\text{CMB}}^{-1} + x_c T_K^{-1} + x_\alpha T_\alpha^{-1}}, \tag{43}$$

where T_α is the Lyman α colour temperature and x_α is another dimensionless parameter, this one describing the relative importance of Lyman α photons compared to CMB photons.

C. Expected signal

Now that we know the physical mechanisms that set the spin temperature, we can sketch out, at least roughly, how we expect the 21 cm signal from cosmological distances to behave. Starting at very high redshift, $z \gtrsim 200$, the density is high and thus $x_c \gg 1$; there are no Lyman α photons, so $x_\alpha \ll 1$. The spin temperature therefore equals the gas kinetic temperature. However, at these high redshifts it also turns out that Compton scattering of CMB photons by free electrons in the gas forces the gas and CMB temperatures to match, so the system is in equilibrium, and $T_{\text{spin}} = T_K = T_{\text{CMB}}$. Thus there is no detectable signal.

At $z \lesssim 200$, the gas begins to drop below the CMB temperature, because cosmological expansion causes its temperature to fall as $T_K \propto (1 + z)^2$, whereas radiation only cools as $T_{\text{CMB}} \propto 1 + z$. Since $T_K < T_{\text{CMB}}$, and x_c is not negligibly small, we have $T_{\text{spin}} < T_{\text{CMB}}$, and we expect the redshifted 21 cm line to be seen in absorption. The amount of absorption drops as the expansion continues, because density and thus x_c are dropping, so T_{spin} rises toward T_{CMB} again; by $z \sim 40$ the absorption is almost gone.

That remains the state of affairs until the first stars and quasars turn on, at which point Lyman α photons begin to appear. These produce $x_\alpha \gg 1$, so the spin temperature becomes locked to the Lyman α colour temperature. However, the Lyman α photons are also coupled to the gas by collisions, so $T_\alpha \approx T_K$, and as a result we have $T_{\text{spin}} \approx T_K < T_{\text{CMB}}$. Thus we see absorption again.

At that point the evolution becomes more uncertain, and measuring it is one of the goals of the current experiments. Generically, we expect that radiation from stars and galaxies will start heating up the gas and we will see the 21 cm line in emission. However, in places the gas is also become ionised, at which point the signal vanishes. The evolution is a competition between these effects. Eventually ionisation wins, and the 21 cm signal vanishes.