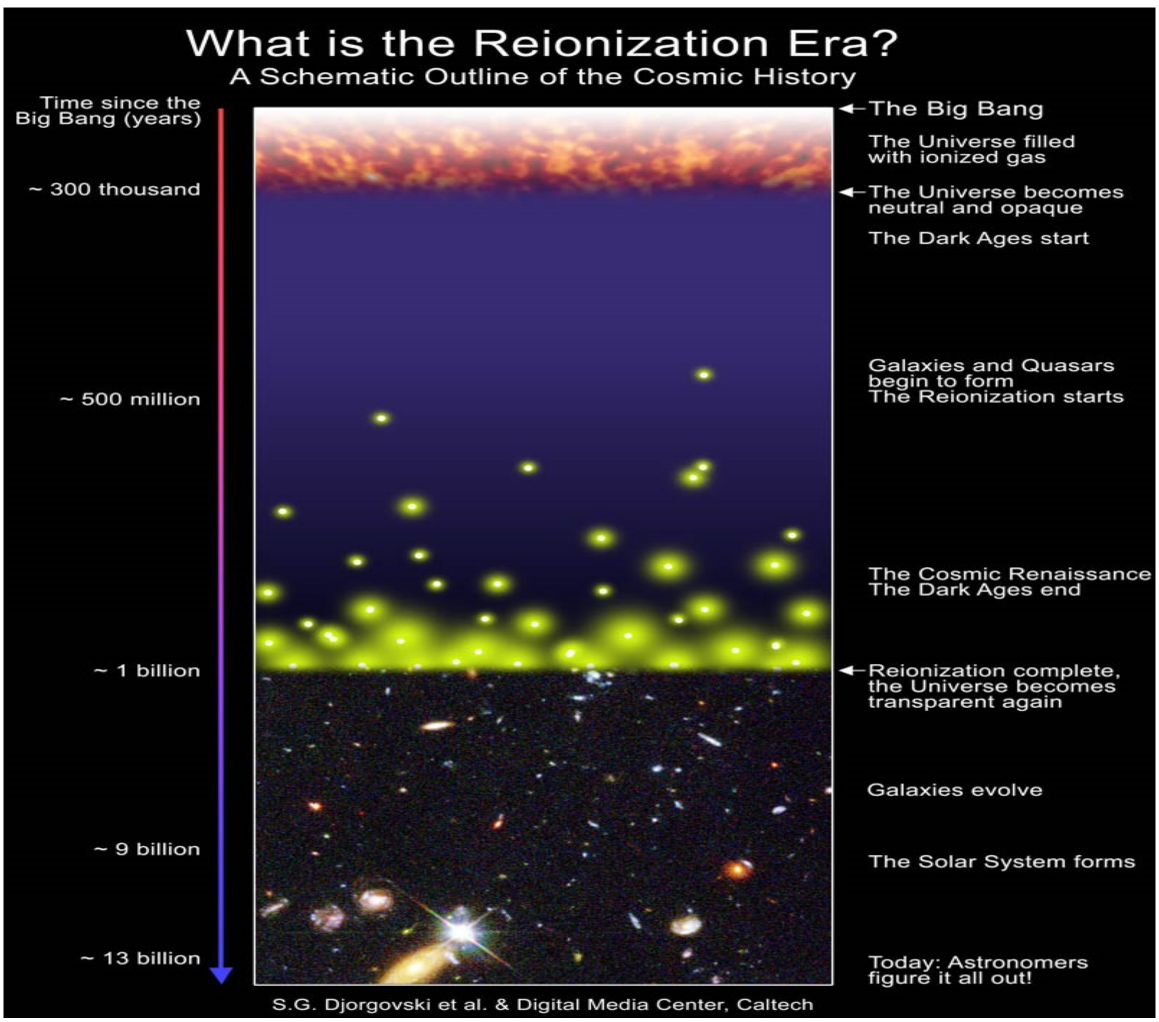
Edmond Cheung
Low Density Astrophysics
December 11, 2008
Jason X. Prochaska

Reionization: 21 cm

I. Introduction

For those who are interested in what the universe looked like in the early ages, about a z~1000 to z~6, there are not many options available. However, one that has been extensively studied is the 21 cm transition of neutral hydrogen. To understand why this is important, we must take a quick trip back to the beginning of the universe as we understand it so far.



As seen in the previous diagram, the universe starts with a big bang, which creates an extremely hot, dense soup of ionized plasma. Since everything is ionized at this point, there will be no 21 cm absorption/emission from neutral hydrogen, hence we could not probe back this far. However, in the next era, everything recombines, which means that neutral hydrogen is present since hydrogen is the most abundant element in the universe. Here one would think we can see the 21 cm signal and be able to trace the whole era, but we will see that this is not the case, in fact, we will only be able to see the very end of it; the part leading into the epoch of reionization, where everything is ionized again due to very hot stars and possibly quasars.

II. Physics of 21 cm

What is the 21 cm transition? In a hydrogen atom, there is both a proton and an electron; both have spins, which can either be the same or opposite to one another. This difference is what creates the hyperfine transition in the ground state of hydrogen, which is separated by an energy of 5.9 x 10^{-6} eV, corresponding to a wavelength of 21 cm.

Before we can apply the 21 cm transition to the bigger scheme of things, we need to delve into the details of what makes this tick. The first thing we need is the concept of a brightness temperature. The brightness temperature is the temperature in which the specific intensity is equal to the blackbody radiator (with spectrum B_ν) at the brightness temperature. In other words: $I_\nu = B_\nu(T_b)$ where $T_b(\nu) \approx \frac{I_\nu c^2}{2k_B \nu^2}$. Armed with this we can change the equation of radiative transfer along a line of sight through a cloud of uniform excitation temperature into: $T_b^{'}(\nu) = T_{ex}(1-e^{-\tau_\nu}) + T_R^{'}(\nu)e^{-\tau_\nu}$. Here we identify

$\tau_\nu \equiv \int \kappa_\nu ds$ as the optical depth and $T_R^{'}$ is the brightness of the background radiation field. T_{ex} is the spin temperature of the 21 cm transition, which is defined as:

$\frac{n_1}{n_2} = \frac{g_1}{g_2} e^{-E_{10}/k_B T_S} = 3e^{-T_*/T_S}$ which is just the Boltzmann distribution between two hyperfine levels of hydrogen. The next step is to find an expression for the optical depth, which after through some manipulation we get:

$$\tau_{\nu o} = \frac{3}{32} \frac{hc^3 A_{10}}{k_B T_S \nu_o^2} \frac{x_{HI} n_H}{(1+z)(dv_{||}/dr_{||})} \approx .0092(1+\delta)(1+z)^{3/2} \frac{x_{HI}}{T_S} \left[\frac{H(z)/(1+z)}{dv_{||}/dr_{||}} \right].$$

The $(1+\delta)$ factor here represents the fractional overdensity of baryons, $dv_{||}/dr_{||}$ is the gradient of the proper velocity along the line of sight, A_{10} is the spontaneous emission coefficient, and *H(z)* is the Hubble flow. What's important to note here is that the neutral hydrogen fraction density represented by x_{HI} . The point of introducing all these equations is to calculate the differential brightness temperature between the high-redshift hydrogen clouds and the Cosmic Microwave Background (CMB). The reason we want the differential brightness temperature and not just the brightness temperature is to isolate the 21 cm signal. If we take the brightness temperature for a high-redshift hydrogen cloud at the cloud (there is a cosmological redshift we must take account of, hence the prime) we have: $T_b^{'}(\nu) = T_S(1-e^{-\tau_\nu}) + T_\gamma^{'}(\nu)e^{-\tau_\nu}$. The first term represents the 21 cm signal while the second term is the CMB. Since we just want the 21 cm, we subtract this brightness temperature by the brightness temperature of a sightline with a clear view of the CMB: $T_b^{'}(\nu) = T_\gamma^{'}(\nu)$ to get

$$\delta T_b(\nu) = \frac{T_s(1-e^{-\tau_\nu}) + T_\gamma(\nu)e^{-\tau_\nu} - T_\gamma(\nu)}{1+z} = \frac{T_s - T_\gamma(z)}{1+z}(1-e^{-\tau_\nu}) \approx \frac{T_s - T_\gamma(z)}{1+z}\tau_{\nu o}$$

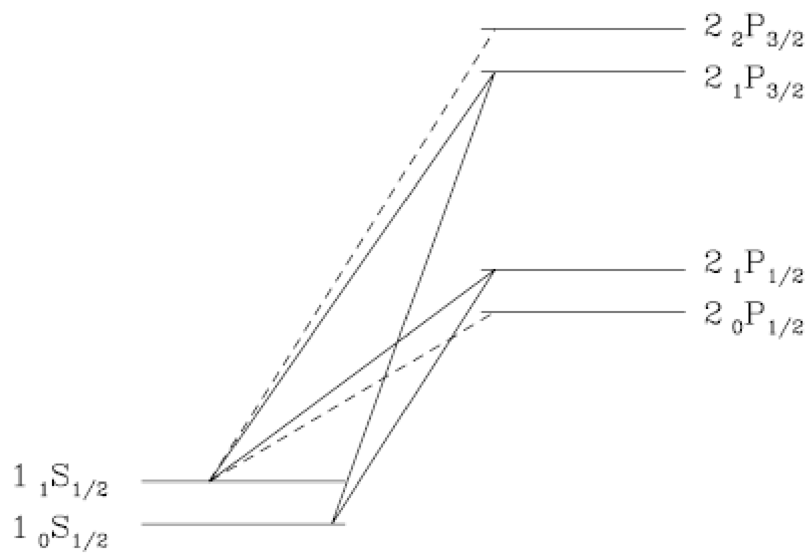
$$\approx 9 x_{HI}(1+\delta)(1+z)^{1/2}\left[1 - \frac{T_\gamma(z)}{T_s}\right]\left[\frac{H(z)/(1+z)}{dv_{\|}/dr_{\|}}\right]\text{mK}.$$

From these equations, it is clear that we need to understand the spin temperature, in particular, how it is set, in order to know when we will get a signal. Furthermore, we see that the neutral fraction of hydrogen is something we can explicitly measure if $T_s >> T_\gamma$, meaning when we have emission with respect to the CMB, $\delta T_b(\nu)$ is independent of T_s, allowing us to calculate the neutral hydrogen fraction. It is this aspect of the 21 cm transition that makes it so attractive; if one were to measure the differential brightness temperature at multiple redshifts, one can create a 3-D map (tomography) of the neutral and ionized hydrogen fractions, giving us an unparalleled view of the epoch of reionization.

There are three processes that determine the T_s: (1) absorption of CMB photons; (2) collisions with other hydrogen atoms, free electrons, and protons; and (3) scattering of UV photons. This gives us an equation: $T_s^{-1} = \frac{T_\gamma^{-1} + x_c T_K^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha}$, where x_c and x_α are coupling coefficients for collisions and UV scattering, respectively, T_K is the gas kinetic temperature, and T_c is the effective color temperature of the UV radiation, and in most of our situations we can just approximate $T_c \rightarrow T_K$. It is important to note that there are three relevant temperatures: T_K, T_s, T_γ, Every process above effectively acts to couple T_s to T_K , without these processes, T_s will reach equilibrium with the CMB and take on the same temperature, leaving the universe invisible to us.

The coupling coefficient, x_c , is different for each type of particle that hits the hydrogen atom. The details for this calculation are not enlightening, essentially, it all hinges on the computation of the rate coefficient for spin de-excitation for a particular species. The first and most known interaction is the hydrogen-hydrogen collisions. The dominant interaction is the electron (and hence spin) exchange in atomic collisions. There are several types of permissible spin-exchange collisions that I will not go into. The next important collision is neutral hydrogen and free electrons, which also induces spin-exchange. These collisions are usually unimportant in the early universe compared to H-H collisions due to the small fraction of ionization during this epoch (after recombination and before reionization). Neutral hydrogen can also collide with bare protons, deuterium atoms, helium atoms or ions, and other trace elements. However, they are all not as significant as H-H collisions.

Another important coupling mechanism is the Wouthuysen-Field Effect. This is illustrated in the following diagram where the hyperfine sublevels of both the 1S and 2P states are displayed.

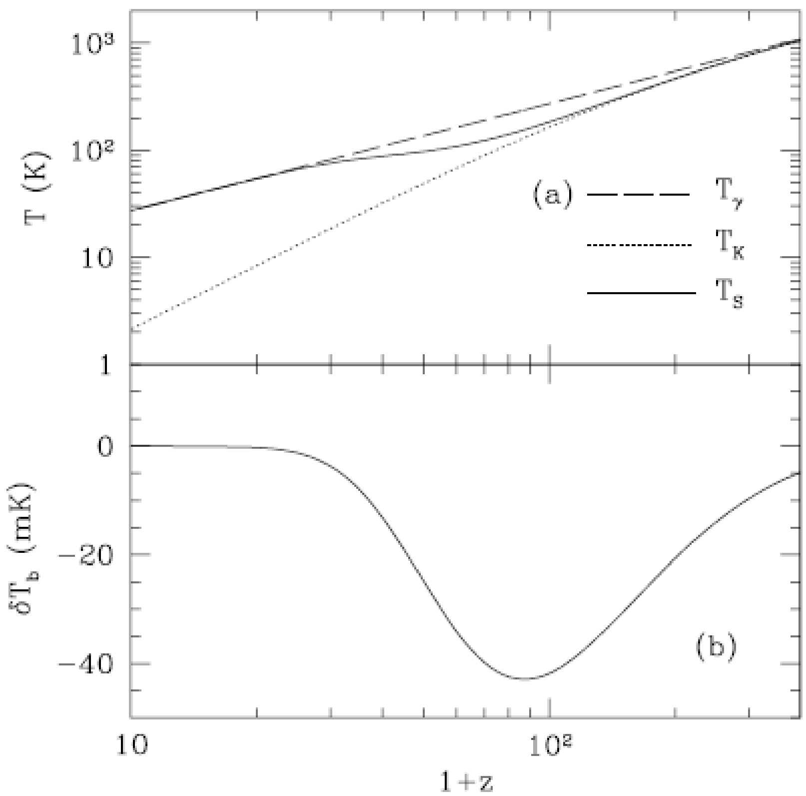


The notation that is used here is: ${}_F L_J$, where F is the total angular momentum of the atom (instead of electron) and L and J is the orbital and total angular momentum of the electron. Here is how it works, suppose the hydrogen atom was in its hyperfine singlet state and absorbs a $Ly\alpha$ photon. The electric dipole selection rule allows $\Delta F = 0,1$ except that $F = 0 \rightarrow 0$ is prohibited. Thus the atom will jump up into one of the central 2P states. But instead of having to decay back to the singlet, the atom is allowed to decay into the hyperfine triplet level, and hence changing hyperfine levels. It is basically this effect that determines the coupling coefficient from UV scattering, x_α.

However, the $Ly\alpha$ photon is not the only one that contributes to this effect, all the Lyman line photons also make contributions, including those that redshift into a Lyman series resonance. Suppose a *Lyn* photon is absorbed by an atom. Now it can either scatter through a direct decay to the ground state, or vanish if the atom decays into an intermediate level. Typically *Lyn* photons scatter ~ 5 times before being consumed by a decay cascade, hence everything leads down the road of the second option. And this is significant because some *Lyn* photons can cascade to form $Ly\alpha$ photons. For example, the $Ly\beta$ photon is only allowed to decay to the ground state or to the 2S level. The Hα photon produced in the decay from 3P $\rightarrow$ 2S transition escapes into infinity. And then it would decay to the ground state through a forbidden two photon process. Thus $Ly\beta$ will never form a $Ly\alpha$ photon, but $Ly\gamma$ will. $Ly\gamma$ can also decay directly to the ground state, or to the 2S level, or to the 2P level. If it finds itself in the 2P level, essentially, it'll become a $Ly\alpha$, and hence the Wouthuysen-Field mechanism can take place. The coupling from *Lyn* photons is about one-third as efficient as from $Ly\alpha$ photons.

III. Global Evolution of the IGM

In the beginning of the universe, all three temperatures that we are concerned with (T_K , T_s , T_γ) were all coupled to one another. As it got less dense, the Compton scattering between the CMB photons and residual free electrons in the IGM became less efficient to the point in which T_K and T_γ became decoupled. At this point, at z~150, $T_K \propto (1+z)^2$, as is expected for an adiabatically expanding non-relativistic gas, while $T_\gamma \propto (1+z)$, which means we'll see T_K become significantly cooler than the CMB, T_γ, as can be seen here:



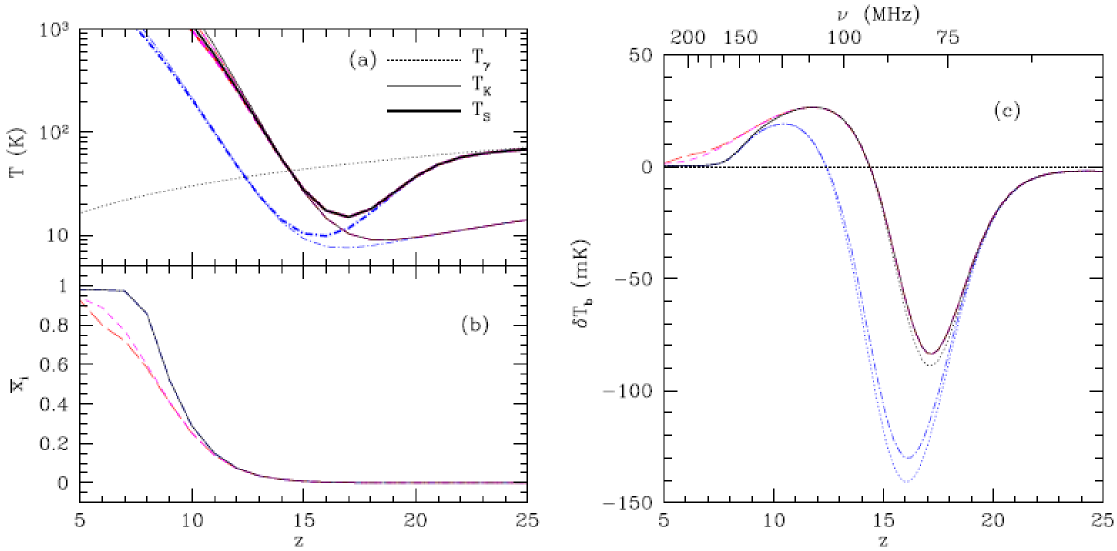
It can be seen that T_K starts to decouple from T_γ at around z~300, which corresponds to Compton heating becoming inefficient, and at z~150, it is negligible. As for our protagonist, T_s, we can see that as T_K decouples from the CMB, so does T_s. This is because the universe at these high redshifts was still dense enough for collisional coupling to be efficient. Up until z~70, $x_c = 1$, but after that, $T_s \longrightarrow T_\gamma$, which corresponds to the universe being invisible because the differential brightness

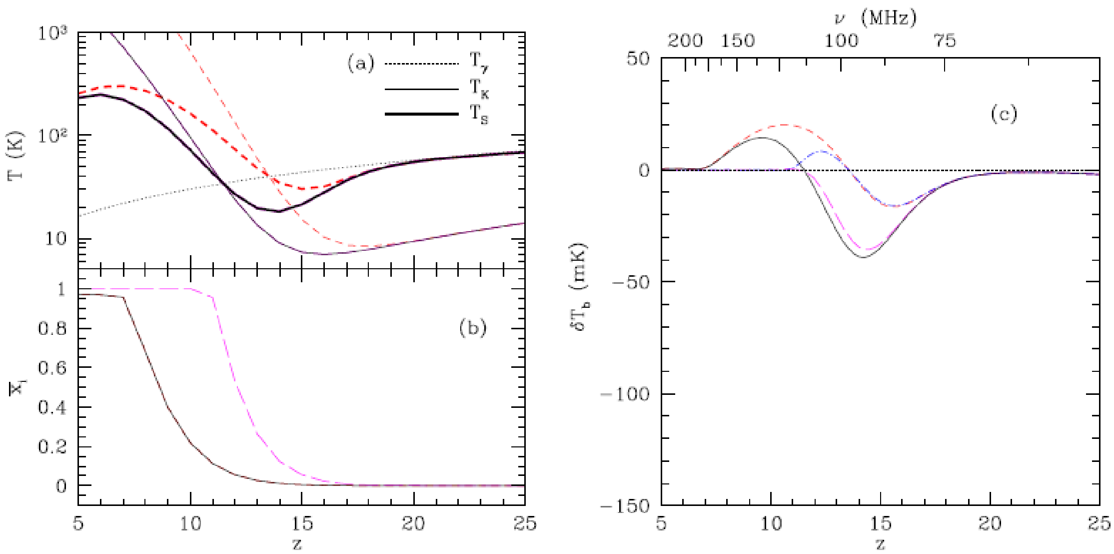
temperature, $\delta T_b(\nu) \approx \dfrac{T_s - T_\gamma(z)}{1+z}\tau_{\nu o}$, becomes zero as $T_s \longrightarrow T_\gamma$. It's interesting to note that during this epoch called the Dark Ages ($1000 \le z \le 30$), 21 cm observations are the only proposed means to probe it.

In the graph above, we can see that the universe is predicted to stay invisible after the absorption feature at z~80. This would be the case if there were no luminous objects, fortunately for us, the first luminous objects are believed to have formed around z~30, leading to the re-heating of the IGM. X-rays from these first objects are thought to be the most important source of heating because of its relatively long mean free paths, in particular, photons with $E > 1.5\bar{x}_{HI}^{1/3}[(1+z)/10]^{1/2}$ keV have mean free paths greater than the Hubble length. X-rays are formed through two major sources, inverse-Compton scattering off of relativistic electrons accelerated by supernovae and high-mass X-ray binaries. The latter consists of a massive main sequence star accreting onto a compact neighbor, these systems would be born as soon as the first massive stars die. There are three main ways to heat the IGM with these X-rays: (1) collisional ionizations, producing more secondary electrons, (2) collisional excitations of HeI (which produce photons capable of ionizing HI) and HI (which produces a $Ly\alpha$ background), and (3) Coulomb collisions with free electrons. $Ly\alpha$ can also heat the IGM through atomic recoil, but is negligible compared to X-ray heating. It is important to keep in mind that these heating mechanisms act in a such a way that decouples T_K and T_γ, as a matter of fact, through these heating mechanisms, T_K will surpass T_γ before reionization starts.

IV. Representative Models

Armed with the thermal history of the IGM and the physics of 21 cm, we can now put this together to form a rough picture of how the universe looked like from z~25, the latter part of the Dark Ages, to z~6, after reionization is complete.





Both are representative models of how T_K , T_s , T_γ evolve with respect to redshift. We also have the ionization fraction on the bottom left and the differential temperature on the right. The difference between these two models is that the first one is based on a fiducial set of Population II star parameters, while the bottom one is with Population III star parameters. Important features to note are: the redshift, z_c , in which the coupling coefficient for UV scattering equals one ($x_\alpha = 1$), meaning the

Wouthuysen-Field effect couples T_s to T_K; redshift at which the IGM is hotter than the CMB ($T_K > T_\gamma$), z_h; and the redshift of reionization, z_r. The dot-dashed, short- and long-dashed curves are curves that include feedback effects, which we will not explore. We are only concerned with the solid curves of T_K, T_s and the dot curve of T_γ. Taking a closer look at the first graph, with the Population II parameters, we see at the beginning (z=25) the spin and CMB temperature are coupled together while the kinetic temperature is decoupled due to adiabatic cooling. This is almost a continuation of the graph we saw before during the epoch of the Dark Ages. As redshift decreases, we see that T_s is decoupling from T_γ and going towards T_K. The reason for this is because $x_\alpha > 0$, which means that the Wouthuysen-Field effect is becoming efficient. At this point, the universe is extremely diffuse compared to the Dark Ages due in part to expansion; thus x_c is essentially zero unless one were to be in a cosmological filament. Therefore, the main mechanism to couple T_s to T_K is the Wouthuysen-Field effect. The UV photons, in particular the $Ly\alpha$ photons, are generated from stars, either Population II or III. The redshift where $x_\alpha = 1$, z_c, takes place around $z \approx 18$; since there is a discrepancy between T_s and T_γ, we expect to see something in the differential brightness temperature, and that is what the graph on the right tells us. z_h, the redshift where the IGM is hotter than the CMB takes place at $z \approx 14$, and we can see on the brightness temperature plot that at this point, the signal changes from absorption to emission. The fact that it changes into emission is important because that allows us calculate the neutral hydrogen fraction independent of T_s.

$\delta T_b(\nu) \approx 9 x_{HI}(1+\delta)(1+z)^{1/2}\left[1-\frac{T_\gamma(z)}{T_s}\right]\left[\frac{H(z)/(1+z)}{dv_{\|}/dr_{\|}}\right]$mK. We can see that if

$T_s \gg T_\gamma$, then the $\left[\frac{T_\gamma(z)}{T_s}\right]$ term drops out and we have an expression with x_{HI} and some other quantities we can measure. And finally, the redshift of reionization, z_r, starts around $z \approx 13$ and ends around $z \approx 7$. As expected, δT_b disappears as reionization goes underway because of the lack of HI present. But what's more important is that there is an emission signal before reionization, and hence we are presented with the hope of actually observing this signal and how reionization evolved in terms of neutral hydrogen fraction across the whole sky; this allows us to see how the neutral and ionized gas was distributed at a given ionization fraction, $\bar{x}_i(z)$. And if we do this for a range of redshifts, we can form a 3-D map, giving us unimaginable information during this rather unknown period.

Similar things are observed in the Population III star parameters. z_c ~13 and z_h ~11, which is closer than the Population II star case. Looking at the corresponding δT_b graph, we see that the absorption is also weaker, this is due to Population III stars producing fewer $Ly\alpha$ photons compared to Population II stars. z_r ~11, which is quite close to z_h, meaning that the δT_b doesn't turn into emission and saturate before reionization starts. This means that it would be difficult to separate out T_s and x_i at the beginning of reionization. However, the significant point is that both models predict that we should be able to see a signal from δT_b and learn something about reionization.

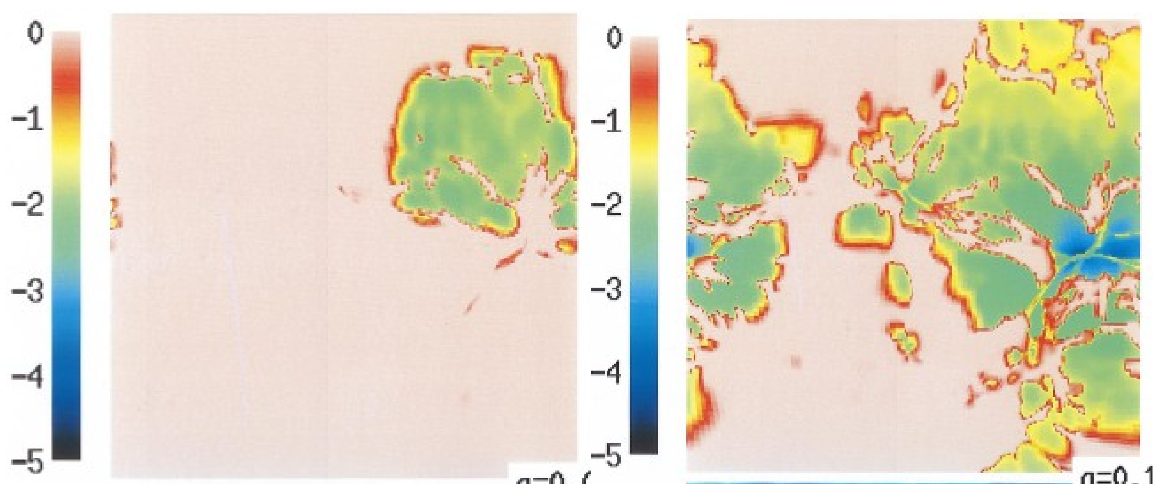
V. Reionization

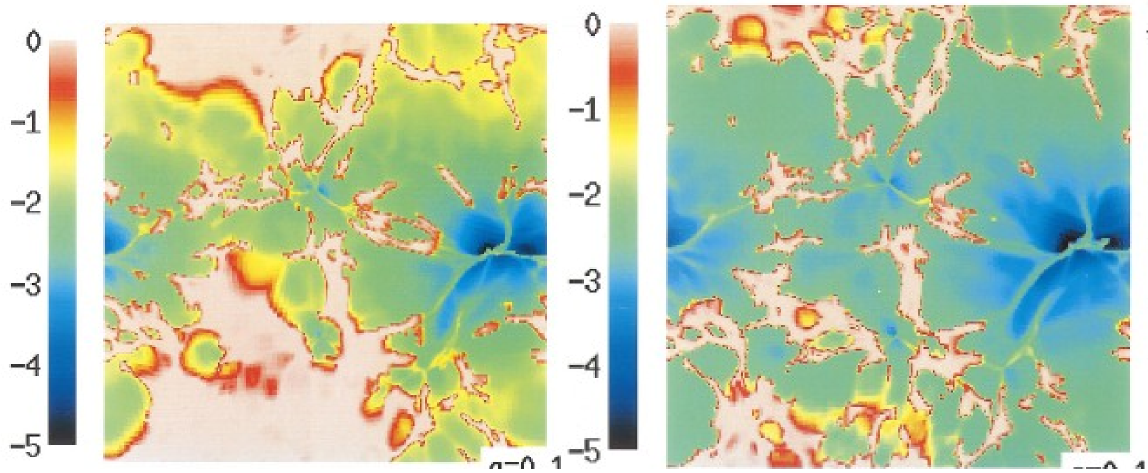
The ionization history has been a subject of extensive study for the past three decade and is described with this equation: $\frac{d\bar{x}_i}{dt} = \zeta(z)\frac{df_{coll}}{dt} - \alpha C(z,\bar{x}_i)\bar{x}_i(z)n_e(z)$,

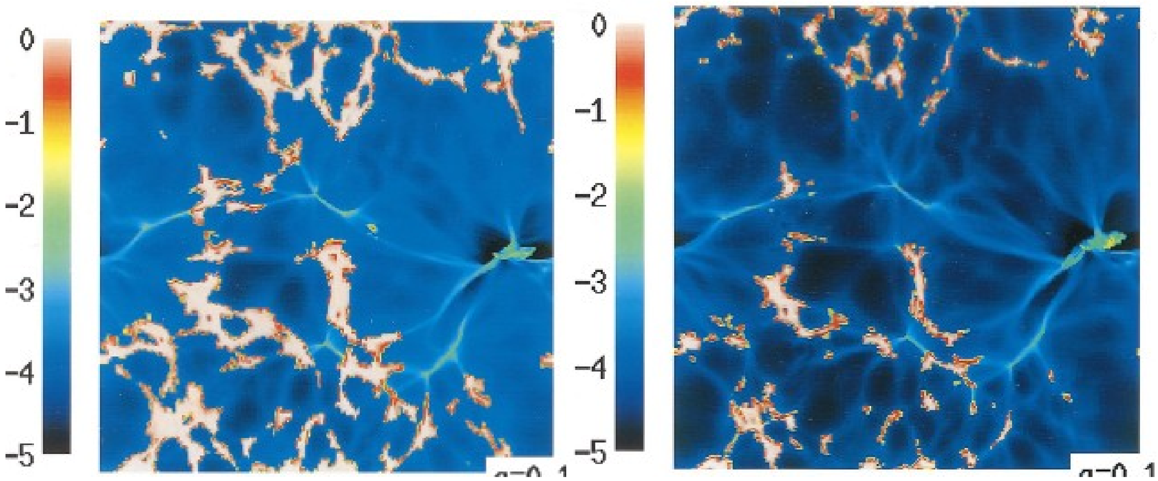
$\frac{df_{coll}}{dt}$ is the rate at which gas collapses onto virialized halos, α recombination coefficient, C is the clumping factor. The ionizing efficiency, $\zeta = A_{HE} f_* f_{esc} N_{ion}$ where f_* is the star formation efficiency, f_{esc} is the fraction of ionizing photons that escape their host galaxy into the IGM, N_{ion}, is the mean number of ionizing photons produced per stellar baryon, A_{HE} is the correction factor to convert the number of ionizing photons per baryon stars to the fraction of ionized hydrogen. We see in this equation that the rate at which ionization occurs is due to the balance between the source and sink terms. This equation is the basis for simulations and semi-analytical models on reionization.

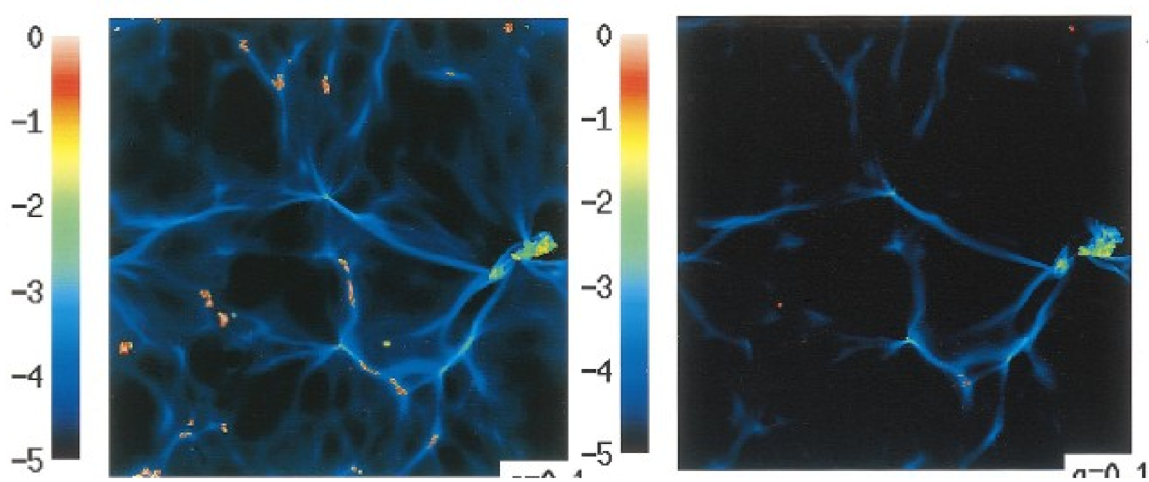
The most problematic factor is the clumping factor. This factor acts as a sink to ionization because as gas in the IGM clump up, they become very dense. If an ionization source turns on inside this clump, it will ionize the inner regions, but the outer regions will counter this by providing neutral gas for it to recombine and hence create a shield between this ionizing source and the outer parts of the clump and the rest of the IGM; the greater the clumpiness, the less ionization from this source, and hence a sink. We don't know this factor and we are left with trying to compute it with numerical simulations. This requires great resolution to resolve the density fluctuations in the smallest scales, and also we need to correctly trace how ionized and neutral gas was distributed.

Semi-analytical models are based on a simple clumping factor to account for the inhomogeneity of the IGM. While these simulations seem to be able to produce the general features of reionization, the simplicity of the models do not allow them to inspect in detail the key dynamics between dark matter, gas, stars, and the radiation field. Thus to look at these incredibly complicated dynamical interactions between the different matter contents of the universe, people have turned to simulations. In particular, these simulations take into account: dark matter, gas dynamics, star formation, atomic physics, molecular hydrogen, and three-dimensional radiative transfer. One numerical simulation gives the following results. The scale on the left is logarithm of neutral hydrogen, so the bluer/blacker it is, the more ionized it is; this takes place at z=11.5, 9, 7.7, 7, 6.7, 6.1, 5.7, respectively.




0
−1
−2
−3
−4
−5
0
−1
−2
−3
−4
−5





We can see as reionization starts, HII regions are formed throughout the box with small bubbles. These bubbles cross into each other and fully ionizes the universe. When reionization starts and ends is not exactly determined, but is generally agreed to be around z~12 to z~6. Throughout these slices of the reionization epoch, we can see that HI is present until the very end. Thus we should be able to see how the neutral and ionized hydrogen fraction evolves with time by taking “all-sky” spectrum of the brightness temperature at different redshifts.

VI. Conclusion

The 21 cm transition is an exciting probe of the early universe, especially during reionization and the Dark Ages. In fact it’s the only suggested probe for the Dark Ages as of right now. Although we have the CMB to tell us something about reionization, the 21 cm transition is a spectral line, which means redshift information can be used to trace the entire three-dimensional history, hence tomography. However, there are challenges before 21 cm can reach its potential. Trying to extract a clear signal at such low frequencies is a challenge in itself, but we must also deal with the high foreground noise, which as of right now, is insurmountable. There is, of course a silver lining to this cloud, the first generation 21 cm experiments are being constructed right now, for example, in Australia where they are building the Mileura Wide-Field Array (MWA); there is also one in China the Primeval Structure Telescope (PaST).

References

[1] S. R. Furlanetto, et al., Phys. Rep. 433 (2006) 181-301

[2] A. Loeb, astro-ph/0603360

[3] N. Gnedin, Astrophys. J. 535 (2000) 530-554

[4] Z. Haiman, et al., Astrophys J. 551 (2001) 599-607

[5] P. Madau, et al., Astrophys J. 475 (1997) 429-444

[6] J. Pritchard, et al., Phys. Rev. 78 (2008) 10