Formation of the first galaxies. They are actually simpler questions theoretically, because the lack of significant feedback, this is a well posed question which we can solve from the initial conditions.

1 First Halos in the Universe

Figure 1 shows $\sigma(M)$ and $\delta_{\text{crit}}(z)$, with the input LCDM power spectrum. The solid line is $\sigma(M)$ for the cold dark matter model with the parameters specified above. The horizontal dotted lines show the value of $\delta_{\text{crit}}(z)$ at $z = 0, 2, 5, 10, 20$ and $30$, as indicated in the figure. From the intersection of these horizontal lines with the solid line we infer, e.g., that at $z = 5$ a $1 - \sigma$ fluctuation on a mass scale of $2 \times 10^7 M_\odot$ will collapse. On the other hand, at $z = 5$ collapsing halos require a $2 - \sigma$ fluctuation on a mass scale of $3 \times 10^{10} M_\odot$, since $\sigma(M)$ on this mass scale equals about half of $\delta_{\text{crit}}(z = 5)$. Since at each redshift a fixed fraction (31.7\%) of the total dark matter mass lies in halos above the $1 - \sigma$ mass, Figure 1 shows that most of the mass is in small halos at high redshift, but it continuously shifts toward higher characteristic halo masses at lower redshift. Note also that $\sigma(M)$ flattens at low masses because of the changing shape of the power spectrum. Since $\sigma \to \infty$ as $M \to 0$, in the cold dark matter model all the dark matter is tied up in halos at all redshifts, if sufficiently low-mass halos are considered.

Figure 2 shows the cooling rates as a function of temperature for a primordial gas composed of atomic hydrogen and helium, as well as molecular hydrogen, in the absence of any external radiation. We assume a hydrogen number density $n_H = 0.045$ cm$^{-3}$, corresponding to the mean density of virialized halos at $z = 10$. The plotted quantity $\Lambda/n_H^2$ is roughly independent of density (unless $n_H > 10$ cm$^{-3}$), where $\Lambda$ is the volume cooling rate (in erg/sec/cm$^3$). The solid line shows the cooling curve for an atomic gas, with the characteristic peaks due to collisional excitation of H1 and He2. The dashed line shows the additional contribution of molecular cooling, assuming a molecular abundance equal to 1\% of $n_H$. From the curve, it requires $T \sim 10^3$ for molecular H to be important in cooling, and $10^4$ for atomic H to be important. Therefore, dark matter halo can’t cool at virial temperature less than 1000K. Also note that if the universe were ionized, no H2 therefore things have to wait.

In figures 3, 4, 5 we show explicitly the properties of collapsing halos which represent $1 - \sigma$, $2 - \sigma$, and $3 - \sigma$ fluctuations (corresponding in all cases to the curves in order from bottom
to top), as a function of redshift. No cutoff is applied to the power spectrum. Figure 2 shows the halo mass, Figure 3 the virial temperature (as well as circular velocity. In figures 2 and 3, the dotted curves indicate the minimum virial temperature required for efficient cooling with primordial atomic species only (upper curve) or with the addition of molecular hydrogen (lower curve). Figure 4 shows the binding energy of dark matter halos. The binding energy of the baryons is a factor $\sim \Omega_b/\Omega_m \sim 15\%$ smaller, if they follow the dark matter. Except for this constant factor, the figure shows the minimum amount of energy that needs to be deposited into the gas in order to unbind it from the potential well of the dark matter. For example, the hydrodynamic energy released by a single supernovae, $\sim 10^{51}$ erg, is sufficient to unbind the gas in all $1-\sigma$ halos at $z > 5$ and in all $2-\sigma$ halos at $z > 12$.

At $z = 5$, the halo masses which correspond to $1-\sigma$, $2-\sigma$, and $3-\sigma$ fluctuations are $1.8 \times 10^7 M_\odot$, $3.0 \times 10^{10} M_\odot$, and $7.0 \times 10^{11} M_\odot$, respectively. The corresponding virial temperatures are $2.0 \times 10^3$ K, $2.8 \times 10^5$ K, and $2.3 \times 10^6$ K. The equivalent circular velocities are 7.5 km s$^{-1}$, 88 km s$^{-1}$, and 250 km s$^{-1}$. At $z = 10$, the $1-\sigma$, $2-\sigma$, and $3-\sigma$ fluctuations correspond to halo masses of $1.3 \times 10^3 M_\odot$, $5.7 \times 10^7 M_\odot$, and $4.8 \times 10^9 M_\odot$, respectively. The corresponding virial temperatures are 6.2 K, $7.9 \times 10^3$ K, and $1.5 \times 10^5$ K. The equivalent circular velocities are 0.41 km s$^{-1}$, 15 km s$^{-1}$, and 65 km s$^{-1}$. Atomic cooling is efficient at $T_{\text{vir}} > 10^4$ K, or a circular velocity $V_c > 17$ km s$^{-1}$. This corresponds to a $1.2-\sigma$ fluctuation and a halo mass of $2.1 \times 10^8 M_\odot$ at $z = 5$, and a $2.1-\sigma$ fluctuation and a halo mass of $8.3 \times 10^7 M_\odot$ at $z = 10$. Molecular hydrogen provides efficient cooling down to $T_{\text{vir}} \sim 300$ K, or a circular velocity $V_c \sim 2.9$ km s$^{-1}$. This corresponds to a $0.81-\sigma$ fluctuation and a halo mass of $1.1 \times 10^6 M_\odot$ at $z = 5$, and a $1.4-\sigma$ fluctuation and a halo mass of $4.3 \times 10^5 M_\odot$ at $z = 10$.

From these calculations, we see why the first objects, probably cooled by $H_2$, formed at $z \sim 20$.

## 2 Fragmentation of the First Gaseous Objects to Stars

As mentioned in the preface, the fragmentation of the first gaseous objects is a well-posed physics problem with well specified initial conditions, for a given power-spectrum of primordial density fluctuations.

### 2.1 Star Formation

Recently, two groups have attempted detailed 3D simulations of the formation process of the first stars in a halo of $\sim 10^6 M_\odot$ by following the dynamics of both the dark matter and the gas components, including $H_2$ chemistry and cooling. Bromm, Coppi, & Larson (1999)
have used a Smooth Particle Hydrodynamics (SPH) code. Abel et al. (2002) isolated a high-density filament out of a larger simulated cosmological volume and followed the evolution of its density maximum with exceedingly high resolution using an Adaptive Mesh Refinement (AMR) algorithm.

The collapsing region forms a disk which fragments into many clumps. The clumps have a typical mass $\sim 10^2$–$10^3 M_\odot$. This mass scale corresponds to the Jeans mass for a temperature of $\sim 500$K and the density $\sim 10^4$ cm$^{-3}$ where the gas lingers because its cooling time is longer than its collapse time at that point. Each clump accretes mass slowly until it exceeds the Jeans mass and collapses at a roughly constant temperature (isothermally) due to H$_2$ cooling that brings the gas to a fixed temperature floor. Bromm (2000) has simulated the collapse of one of the above-mentioned clumps with $\sim 1000 M_\odot$ and demonstrated that it does not tend to fragment into sub-components. Rather, the clump core of $\sim 100 M_\odot$ free-falls towards the center leaving an extended envelope behind with a roughly isothermal density profile. These calculations indicate that each clump may end as a single massive star; however, it is conceivable that angular momentum may eventually halt the collapsing cloud and lead to the formation of a binary stellar system instead.

In their simulation, Abel et al. (2000) adopted the actual cosmological density perturbations as initial conditions. The simulation focused on the density peak of a filament within the IGM, and evolved it to very high densities (Fig. 5). Following the initial collapse of the filament, a clump core formed with $\sim 200 M_\odot$, amounting to only $\sim 1\%$ of the virialized mass. Subsequently due to slow cooling, the clump collapsed subsonically in a state close to hydrostatic equilibrium. Unlike the idealized top-hat simulation of Bromm et al. (2001) [?], the collapse of the different clumps within the filament is not synchronized. Once the first star forms at the center of the first collapsing clump, it is likely to affect the formation of other stars in its vicinity.

As soon as nuclear burning sets in the core of the proto-star, the radiation emitted by the star starts to affect the infall of the surrounding gas towards it. The radiative feedback involves photo-dissociation of H$_2$, Ly$\alpha$ radiation pressure, and photo-evaporation of the accretion disk. Tan & McKee studied these effects by extrapolating analytically the infall of gas from the final snapshot of the above resolution-limited simulations to the scale of a proto-star; they concluded that nuclear burning (and hence the feedback) starts when the proton-star accretes $\sim 30 M_\odot$ and accretion is likely to be terminated when the star reaches $\sim 200 M_\odot$.

If the clumps in the above simulations end up forming individual very massive stars, then these stars will likely radiate copious amounts of ionizing radiation and expel strong winds. Hence,
the stars will have a large effect on their interstellar environment, and feedback is likely to control the overall star formation efficiency. This efficiency is likely to be small in galactic potential wells which have a virial temperature lower than the temperature of photoionized gas, $\sim 10^4$K. In such potential wells, the gas may go through only a single generation of star formation, leading to a “suicidal” population of massive stars.

The final state in the evolution of these stars is uncertain; but if their mass loss is not too extensive, then they are likely to end up as black holes. The remnants may provide the seeds of quasar black holes. Some of the massive stars may end their lives by producing gamma-ray bursts. If so then the broad-band afterglows of these bursts could provide a powerful tool for probing the epoch of reionization.

Show Heger picture.

Where are the first stars or their remnants located today? The very first stars formed in rare high-$\sigma$ peaks and hence are likely to populate the cores of present-day galaxies. However, the bulk of the stars which formed in low-mass systems at later times are expected to behave similarly to the collisionless dark matter particles and populate galaxy halos.

### 2.2 The Mass Function of Stars

Currently, we do not have direct observational constraints on how the first stars, the so-called Population III stars, formed at the end of the cosmic dark ages. It is, therefore, instructive to briefly summarize what we have learned about star formation in the present-day Universe, where theoretical reasoning is guided by a wealth of observational data.

Population I stars form out of cold, dense molecular gas that is structured in a complex, highly inhomogeneous way. The molecular clouds are supported against gravity by turbulent velocity fields and pervaded on large scales by magnetic fields. Stars tend to form in clusters, ranging from a few hundred up to $\sim 10^6$ stars. It appears likely that the clustered nature of star formation leads to complicated dynamics and tidal interactions that transport angular momentum, thus allowing the collapsing gas to overcome the classical centrifugal barrier. The most important feature of the observed IMF is that $\sim 1M_\odot$ is the characteristic mass scale of Pop I star formation, in the sense that most of the mass goes into stars with masses close to this value. In Figure 6, we show the result from a recent hydrodynamical simulation of the collapse and fragmentation of a molecular cloud core. This simulation illustrates the highly dynamic and chaotic nature of the star formation process.

*How did the first stars form?* A complete answer to this question would entail a theoretical
prediction for the Population III IMF, which is rather challenging. Let us start by addressing the simpler problem of estimating the characteristic mass scale of the first stars. As mentioned before, this mass scale is observed to be \( \sim 1M_\odot \) in the present-day Universe.

Bromm & Loeb (2004) carried out idealized simulations of the protostellar accretion problem and estimated the final mass of a Population III star. Using the smoothed particle hydrodynamics (SPH) method, they included the chemistry and cooling physics relevant for the evolution of metal-free gas. They focused on an isolated overdense region that corresponds to a 3\( \sigma \)-peak: a halo containing a total mass of \( 10^6M_\odot \), and collapsing at a redshift \( z_{\text{vir}} \approx 20 \). In these runs, one high-density clump has formed at the center of the minihalo, possessing a gas mass of a few hundred solar masses. Soon after its formation, the clump becomes gravitationally unstable and undergoes runaway collapse. Once the gas clump has exceeded a threshold density of \( 10^7 \text{ cm}^{-3} \), it is replaced by a sink particle which is a collisionless point-like particle that is inserted into the simulation. This choice for the density threshold ensures that the local Jeans mass is resolved throughout the simulation. The clump (i.e., sink particle) has an initial mass of \( M_{\text{cl}} \approx 200M_\odot \), and grows subsequently by ongoing accretion of surrounding gas. High-density clumps with such masses result from the chemistry and cooling rate of molecular hydrogen, \( \text{H}_2 \), which imprint characteristic values of temperature, \( T \sim 200 \text{ K} \), and density, \( n \sim 10^4 \text{ cm}^{-3} \), into the metal-free gas. Evaluating the Jeans mass for these characteristic values results in \( M_J \sim \text{a few } 10^2M_\odot \), which is close to the initial clump masses found in the simulations.

**How massive were the first stars?** Star formation typically proceeds from the ‘inside-out’, through the accretion of gas onto a central hydrostatic core. Whereas the initial mass of the hydrostatic core is very similar for primordial and present-day star formation, the accretion process – ultimately responsible for setting the final stellar mass – is expected to be rather different. On dimensional grounds, the accretion rate is simply related to the sound speed cubed over Newton’s constant (or equivalently given by the ratio of the Jeans mass and the free-fall time): \( \dot{M}_{\text{acc}} \sim c_s^3/G \propto T^{3/2} \). A simple comparison of the temperatures in present-day star forming regions (\( T \sim 10 \text{ K} \)) with those in primordial ones (\( T \sim 200–300 \text{ K} \)) already indicates a difference in the accretion rate of more than two orders of magnitude.

**Can a Population III star ever reach this asymptotic mass limit?** The answer to this question is not yet known with any certainty, and it depends on whether the accretion from a dust-free envelope is eventually terminated by feedback from the star. The standard mechanism by which accretion may be terminated in metal-rich gas, namely radiation pressure on dust grains, is evidently not effective for gas with a primordial composition. Recently, it has been speculated that accretion could instead be turned off through the formation of an \( \text{H II} \) region,
or through the photo-evaporation of the accretion disk. The termination of the accretion process defines the current unsolved frontier in studies of Population III star formation. Current simulations indicate that the first stars were predominantly very massive ($> 30M_\odot$), and consequently rather different from present-day stellar populations. The crucial question then arises: *How and when did the transition take place from the early formation of massive stars to that of low-mass stars at later times?* We address this problem next.

The very first stars, marking the cosmic Renaissance of structure formation, formed under conditions that were much simpler than the highly complex environment in present-day molecular clouds. Subsequently, however, the situation rapidly became more complicated again due to the feedback from the first stars on the IGM. Supernova explosions dispersed the nucleosynthetic products from the first generation of stars into the surrounding gas, including also dust grains produced in the explosion itself. Atomic and molecular cooling became much more efficient after the addition of these metals. Moreover, the presence of ionizing cosmic rays, as well as of UV and X-ray background photons, modified the thermal and chemical behavior of the gas in important ways.

Early metal enrichment was likely the dominant effect that brought about the transition from Population III to Population II star formation. Recent numerical simulations of collapsing primordial objects with overall masses of $\sim 10^6M_\odot$, have shown that the gas has to be enriched with heavy elements to a minimum level of $Z_{\text{crit}} \simeq 10^{-3.5}Z_\odot$, in order to have any effect on the dynamics and fragmentation properties of the system. Normal, low-mass (Population II) stars are hypothesized to only form out of gas with metallicity $Z \geq Z_{\text{crit}}$. Thus, the characteristic mass scale for star formation is expected to be a function of metallicity, with a discontinuity at $Z_{\text{crit}}$ where the mass scale changes by $\sim$ two orders of magnitude. The redshift where this transition occurs has important implications for the early growth of cosmic structure, and the resulting observational signature.

### 3 Basic Models of Reionization

The baryonic pre-galactic medium (PGM) evolves in three distinct phases. At high redshifts ($z > 1100$) the PGM is hot, fully ionized, and optically thick to Thomson scattering, and hence coupled to the photon field. As the universe expands, the PGM cools, and eventually recombines, leaving a surface of last scattering (the cosmic microwave background, CMB), plus a neutral PGM. This neutral phase lasts from $z = 1100$ to $z \sim 14$. At some point between $z \sim 14$ and 6, hydrogen in the PGM is ‘reionized’, due to UV radiation from the
first luminous objects, leaving the fully reionized intergalactic medium (IGM) seen during the ‘realm of the galaxies’ \((6 > z > 0)\). The ionized, dense PGM at very high redshift has been well studied through extensive observations of the CMB. Likewise, the reionized, rarified IGM at low redshift has been well characterized through QSO absorption line studies. The middle phase – the existence of a neutral IGM during the so-called ‘dark ages’ (Rees 1998), and the process of reionization of this medium, is the last directly observable phase of cosmic evolution that remains to be verified and explored. The epoch of reionization (EoR) is crucial in cosmic structure formation studies, since it sets a fundamental benchmark indicating the formation of the first luminous objects, either star forming galaxies or active galactic nuclei (AGN).

Assuming reionization is driven by UV photons from the first luminous sources (stars or AGN), analytic and numerical, calculations suggest that reionization follows a number of phases. The first phase will be relatively slow, with each UV source isolated to its own Stromgren sphere. Eventually, these spheres grow and join, leading to much faster reionization, due to the combination of accelerating galaxy formation, plus the much larger mean free path of ionizing photons in the now porous IGM. This stage is known as ‘overlap’, or ‘percolation’. The last phase entails the etching-away of the final dense filaments in the IGM by the intergalactic UV radiation field, leading to a fully ionized IGM (neutral fraction, \(x_{HI} \sim 10^{-5}\), at \(z = 0\)).

Figure: overlap picture from Loeb and Haiman.

Some of the key questions regarding reionization include:

- **When** did reionization happen? Whether it is early \((z \sim 15)\) or late \((z \sim 6 – 8)\), whether it has an extended history or resembles a phase transition.

- **How** did reionization proceed? Whether it is homogeneous or with large scatter? How did HII regions grow during reionization and how did overlap happen?

- **What** did it? Whether the main sources of reionization photons come from dwarf galaxies, AGNs, or even more exotic sources such as decay particles.

4 Gunn-Peterson Effect

First detection in 60’s with Maarten Schmidt’s observations of fast-receding point sources called “quasi-stellar objects”. If recession is cosmological (debated for some time), these were at great distances and highly luminous. Two notable features:
1. Flux is seen bluewards of Lyα emission peak.
2. Occasionally, absorption lines are seen along the LOS.

Led to two influential back-to-back papers in 1965 ApJ, by Gunn & Peterson, and Bahcall & Salpeter. Gunn and Peterson argued that it’s remarkable that any flux is seen at all, because if cosmos was neutral then HI absorption should strongly absorb that flux; hence diffuse gas must be highly ionized.

Bahcall & Salpeter argued that absorption lines were actually redshifted Lyα along the line of sight (rather than some unknown element associated with the quasar). The picture developed that the IGM was highly ionized but dotted by pockets of neutral gas (possibly associated with ISM of other galaxies).

Gunn & Peterson’s result still gives valuable insight today, while Bahcall & Salpeter’s turned out not to be the right picture; this wasn’t fully realized until the 90’s.

Gunn & Peterson’s argument was that a smooth neutral IGM would have enormous optical depth below 1216 Å (HI Lyα, 2-1 transition). That we observe flux bluer than this means that the IGM is almost entirely ionized. For a number density of neutral hydrogen atoms \( n(z) \), with cross-section \( \sigma(\nu(1+z)) \), we get an optical depth of

\[
\tau = \int_0^z n(z) \sigma(\nu(1+z))(dl/dz)dz
\]

Now \( n(z) = n_{\text{HI}}(1+z)^3 \) for a constant comoving number density of HI atoms. \( dl/dz = c/H(z) = cH_0^{-1}(1+z)^{-3/2} \) for a EdS cosmology. \( \sigma = \pi e^2 f/m_e c \) (where \( f \) is oscillator strength=0.416 for Lyα) and is highly peaked at the (redshifted) resonance of HI, namely 1216 Å.

\[
\tau(z) \approx n_{\text{HI}} \sigma(\nu_0(1+z))cH_0^{-1}(1+z)^{3/2}
\]

Plugging in some numbers and noting that \( n_H \approx \rho_c\Omega_b/m_p \) gives

\[
\tau = 6.4 \times 10^5 h^{-1} \left( \frac{\Omega_b h^2}{0.02} \right) \left( \frac{1+z}{3} \right)^{3/2} \left( \frac{n_{\text{HI}}}{n_H} \right)
\]

Schmidt observed 3C9 to be at \( z = 2.01 \). The mean optical depth was \( \sim 0.5 \). For \( H_0 = 10^{10} \text{ yr}^{-1} \), \( \Omega_m = 1 \), Gunn & Peterson calculated that \( \Omega_{\text{HI}} \approx 2 \times 10^{-7} \). For WMAP parameters, this gives a neutral fraction of \( 3.8 \times 10^{-6} \). This turns out to be remarkably close, given the simplicity of the argument.

There must be a sea of UV photons (above 13.6 eV) in the universe to keep the IGM ionized. So: universe becomes neutral at \( z = 1000 \), else the optical depth to \( z = 1000 \) would be far larger than we observe (they’ve done this calculation in homework). However, we observe
the IGM to be ionized, certainly out to $z = 6$ and probably beyond. Hence, the IGM must be reionized at some redshift between 6 and about 50.

Figure, SDSS quasar spectra

Figure, deepest troughs

Figure, $\tau$ evolution

Figure, neutral fraction evolution.

Fan et al. measured the evolution of Gunn-Peterson optical depths along the line of sight of the nineteen $z > 5.7$ quasars from the SDSS. We found that at $z_{\text{abs}} < 5.5$, the optical depth can be best fit as $\tau \propto (1 + z)^{4.3}$, while at $z_{\text{abs}} > 5.5$, the evolution of optical depth accelerates: $\tau \propto (1 + z)^{>10}$. There is also a rapid increase in the variation of optical depth along different lines of sight: $\sigma(\tau)/\tau$ increases from $\sim 15\%$ at $z \sim 5$, to $> 30\%$ at $z > 6$, in which $\tau$ is averaged over a scale of $\sim 60$ comoving Mpc. Assuming photoionization equilibrium and a model of IGM density distribution, one can convert the measured effective optical depth to IGM properties, such as the level of UV ionizing background and average neutral fraction. We find that at $z > 6$ the volume-averaged neutral fraction of the IGM has increased to $> 10^{-3.5}$, with both ionizing background and neutral fraction experiencing about one order of magnitude change over a narrow redshift range, and the mean-free-path of UV photons is shown to be $< 1$ physical Mpc at $z > 6$. However, with the emergence of complete Gunn-Peterson troughs at $z > 6$, it becomes increasingly difficult to place stringent limits on the optical depth and neutral fraction of the IGM.

Briefly, flavors of these include:

- Quasar stromgren sphere
- Ly$\alpha$ galaxies

5 Other Observational Constraints

5.1 CMB

Thomson scattering produces CMB polarization when free-electron scatterers are illuminated by an anisotropic photon distribution. Large scale CMB polarization therefore probes the ionization history by measuring the total optical depths to be CMB produced by free electrons generated from reionization process. The strongest polarization constraint comes from the new WMAP results, with the best-fit values of $\tau = 0.084 \pm 0.016$, and $z_{\text{reion}} = 10.8 \pm 1.4$,
assuming an instantaneous reionization at \( z_{\text{reion}} \). However, the CMB polarization only measures an integrated reionization signal. Figure illustrates the degeneracy of reionization redshift and detailed reionization history based on CMB data alone, using a simple reionization model assuming that the universe was partially reionization at \( z_{\text{reion}} \) to an ionization fraction of \( \delta_e^0 \) and then became fully ionized at \( z = 7 \).

### 5.2 21 cm Probe

The 21cm line of neutral hydrogen presents a unique probe of the evolution of the neutral intergalactic medium, and cosmic reionization. Furlanetto & Briggs (2004) point out some of the advantages of using the HI line in this regard: (i) unlike Ly\( \alpha \) (ie. the Gunn-Peterson effect), the 21cm line does not saturate, and the IGM remains ‘translucent’ at large neutral fractions. And (ii) unlike CMB polarization studies, the HI line provides full three dimensional (3D) information on the evolution of cosmic structure, and the technique involves imaging the neutral IGM directly, and hence can easily distinguish between different reionization models. HI 21cm observations can be used to study the evolution of cosmic structure from the linear regime at high redshift (ie. density-only evolution), through the non-linear, ‘messy astrophysics’ regime associated with luminous source formation. As such, HI measurements are sensitive to structures ranging from very large scales down to the source scale set by the cosmological Jeans mass, thereby making 21cm the “richest of all cosmological data sets”

In the Raleigh-Jeans limit, the observed brightness temperature (relative to the CMB) due to the HI 21cm line at a frequency \( \nu = \nu_{21}/(1 + z) \), is given by:

\[
T_B \approx \frac{T_S - T_{\text{CMB}}}{1+z} \tau \approx 7(1 + \delta)\delta_e(1 - \frac{T_{\text{CMB}}}{T_S})(1 + z)^{1/2} \text{mK},
\]

(1)

The conversion factor from brightness temperature to specific intensity, \( I_\nu \), is given by:

\[
I_\nu = \frac{2k_B}{(\lambda_{21}(1+z))^2}T_B \approx 22(1 + z)^{-2} T_B \text{ Jy deg}^{-2}.
\]

This shows that for \( T_S \sim T_{\text{CMB}} \) one expects no 21cm signal. When \( T_S >> T_{\text{CMB}} \), the brightness temperature becomes independent of spin temperature. When \( T_S << T_{\text{CMB}} \), we expect a strong negative (ie. absorption) signal against the CMB.

The interplay between the CMB temperature, the kinetic temperature, and the spin temperature, coupled with radiative transfer, lead to a number of interesting physical regimes for the HI 21cm signal: (I) At \( z > 200 \) equilibrium between \( T_{\text{CMB}}, T_K \), and \( T_S \) is maintained by Thomson scattering off residual free electrons and gas collisions. In this case \( T_S = T_{\text{CMB}} \) and there is no 21cm signal. (II) At \( z \sim 30 \) to 200, the gas cools adiabatically, with temperature
falling as \((1 + z)^2\), i.e. faster than the \((1+z)\) for the CMB. However, the mean density is still high enough to couple \(T_S\) and \(T_K\), and the HI 21cm signal would be seen in absorption against the CMB. (III) At \(z \sim 30\), collisions can no longer couple \(T_K\) to \(T_S\), and \(T_S\) again approaches \(T_{CMB}\). However, the Lya photons from the first luminous objects (Pop III stars or mini-quasars), may induce local coupling of \(T_K\) and \(T_S\), thereby leading to some 21cm absorption regions. At the same time, Xrays from these same objects could lead to local IGM warming above \(T_{CMB}\). Hence one might expect a patch-work of regions with no signal, absorption, and perhaps emission, in the 21cm line. (IV) At low-\(z\), all the physical processes come to play. The IGM is being warmed by hard Xrays from the first galaxies and black holes, as well as by weak shocks associated with structure formation, such that \(T_K\) is likely larger than \(T_{CMB}\) globally. Likewise, these objects are reionizing the universe, leading to a fundamental topological change in the IGM, from the linear evolution of large scale structure , to a bubble dominated era of HII regions.

Difficulties: small signal, strong foreground, radio frequency interference, ionosphere variation.

Needs: large area, quite site, model astrophysical foreground.

Many programs have been initiated to study the HI 21cm signal from cosmic reionization, and beyond. The largest near-term efforts are the Mileura Wide Field Array (MWA) and the Low Frequency Array (LOFAR). These telescopes are being optimized to study the power spectrum of the HI 21cm fluctuations. In the long term the Square Kilometer Array (SKA)\(^1\) should have the sensitivity to perform true three dimensional imaging of the neutral IGM in the 21cm line during reionization.

6 Source of Reionzation

Regardless of the detailed reionization history, the IGM has been almost fully ionized since at least \(z \sim 6\). This places a minimum requirement on the emissivity of UV ionizing photons per unit comoving volume required to keep up with recombination and maintain reionization:

\[
\dot{N}_{\text{ion}}(z) = 10^{51.2} \text{s}^{-1} \text{Mpc}^{-3} \left( \frac{C}{30} \right) \times \left( \frac{1 + z}{6} \right)^3 \left( \frac{\Omega_b h^2}{0.02} \right)^2,
\]

where \(C \equiv \frac{\langle n_e^2 \rangle}{\langle n_H \rangle^2} \) is the clumping factor of the IGM. It is difficult to determine \(C\) from observations, and has large uncertainty when estimated from simulations (\(C = 10 - 100\)). At \(z < 2.5\), the UV ionizing background is dominated by quasars and AGN. At \(z > 3\), the density of luminous quasars decreases faster than that of star-forming galaxies, and the

\(^1\)www.skatelescope.org/
ionizing background has an increased contribution from stars.

Quasars and AGN are effective emitters of UV photons. Luminous quasar density declines exponentially towards high-redshift: it is $\sim 40$ times lower at $z \sim 6$ than at its peak at $z \sim 2.5$. However, quasars have a steep luminosity function at the bright end – most of the UV photons come from the faint quasars that are currently below the detection limit at high-redshift. Based on the derived luminosity function, we find that the quasar/AGN population cannot provide enough photons to ionize the intergalactic medium (IGM) at $z \sim 6$ unless the IGM is very homogeneous and the luminosity at which the QLF power law breaks is very low.

Due to the rapid decline in the AGN populations at very high $z$, most theoretical models assume stellar sources reionized the universe. Figure shows the present constraints on the evolution of UV luminosity and star formation rate of Lyman Break galaxies. Comparing to AGNs, galaxy population show only mild evolution at $z > 5$. However, despite rapid progress, there is still considerable uncertainty in estimating the total UV photon emissivity of star-forming galaxies at high-redshift, especially the IMF and the UV escape fraction from dwarf galaxies. Given these uncertainties, the current data are consistent with star forming galaxies, in particular, relatively low luminosity galaxies, as being the dominant sources of reionizing photons, although more exotic sources, such as high-redshift mini-quasars, can not yet be ruled out as minor contributors of reionization budget. Issue with escape fraction.

Summary plot. Complex and extended reionization history.

7 Helium reionization

In simple models of ionizing background evolution, stars dominate the early growth of the background, then quasars become progressively more important at $z < 4$. Since quasar light is much harder, it can ionize HeII into HeIII (which needs 4 Ry photons, i.e. 54.4 eV). This process is known as Helium reionization, and has several consequences:

1. The global ionizing background becomes harder. This is important for metal species such as OVI whose ionization potential is $\sim 9$ Ry.

2. Latent heat is released, causing energy injection into the IGM. If this occurs on timescales faster than Hubble expansion can equilibrate, then the IGM temperature may spike. Some claims to see this at $z \sim 3.2$, but controversial.

3. Offers an opportunity to study reionization at an observationally accessible redshift ($z \sim 3$),
perhaps giving insights to how H reionization proceeds at $z > 6$. Downside is that photons are 304Årest, so need far UV spectroscopy to see at $z \sim 3$. COS will help.

8 Baryon census:

Baryons seen in emission make up a small fraction of BBN or CMB inferred total ($\Omega_b \approx 0.045$). Most baryons today are not observed to be in stars or cold, dense gas in galaxies: Integrated LF gives 8-10% stars; 21cm/CO emission gives 1% in cold, dense gas; X-ray emission from clusters gives 10% in hot ICM. Where is the rest? Intergalactic!

Universe begins with most baryons in “intergalactic medium”, i.e. gas that has not (yet) collapsed into galaxies. Physical properties:
1. Gas pressure forces are low, hence gas mostly traces dark matter.
2. Gas densities are low, so radiative cooling is unimportant.
3. Optically thin, so radiation mostly passes thru.
4. Highly ionized owing partly to background radiation from various sources.

IGM comes in three basic forms:
1. Cool photoionized intergalactic gas tracing uncollapsed LSS ($T \sim 10^4$K), detectable as HI absorption to background sources.
2. “Warm-hot” intergalactic medium (WHIM); gas shock heated by collapse on LSS to $T \sim 10^{4.5} - 10^{6.5}$K), which is too hot to contain HI and too cold to emit in X-rays. Sometimes called the “missing baryons”.
3. Intracluster medium (ICM); gas shock heated by virialization within large halos to $T \sim 10^7 - 10^8$K), emitting in X-rays.

Simulations indicate currently roughly equally divided between “bound” baryons (including ICM), Ly$\alpha$ forest, and “warm-hot” gas. But in the past, most in Ly$\alpha$ forest (Figure).

9 Ly$\alpha$ Forest

Ly$\alpha$ cloud models

An alternative hypothesis was that Ly$\alpha$ absorbers were ejected from the quasar. Energetically feasible since quasars were highly luminous. Indeed, today ejected absorption is seen up to 10’s of thousands of km/s away. But Sargent et al. (1980) showed that Ly$\alpha$ forest was not ejected but intervening, by mainly showing that Ly$\alpha$ absorbers were distributed uniformly
along LOS (while an ejection model would predict many more at low velocities), and there was no correlation between line strength and (presumed) ejection velocity.

Sargent et al. also showed that Ly$\alpha$ lines were uncorrelated, unlike what would be expected if they arose in galaxies along LOS. Hence they proposed cosmologically distributed cold clouds, pressure-confined by a warm IGM ($n \sim 10^{-5}, T \sim 10^5 K$). By measuring $N_{HI}$ and assuming sphericity, estimated mass density in such cold clouds to be quite small, $\Omega_c \sim 10^{-3}$. This turned out to be wrong, but initiated a bevy of Ly$\alpha$ absorber models.

The main competing idea in the 80’s was that Ly$\alpha$ clouds were “minihalos”, i.e. gravitationally-confined pockets of gas. As data improved, both confinement models needed fine-tuning to what eventually became rather elaborate levels, with non-spherical geometries and no obvious source of such pockets.

Another difficulty with confinement models is that nearby LOS showed correlated Ly$\alpha$ absorption features over 0.5 – 1 Mpc away. Hence Ly$\alpha$ absorbers are large!

Advent of high-resolution quasar spectroscopy put to death such confinement models, while advancing numerical simulations offered a new and elegant structure-formation based interpretation. We will discuss observations today, modern Ly$\alpha$ models next time.

Observations of the Ly$\alpha$ forest

Echelle spectroscopy (dispersing light into orders on a CCD) on large telescopes provided first fully resolved spectra of Ly$\alpha$ forest in distant quasars. In problem set you will compute that the thermal linewidth of Ly$\alpha$ absorbers is $\sim 10$ km/s, hence need this spectral resolution in order to resolve every absorber.

High-z quasar (Figure).

Quasar’s Ly$\alpha$ emission peak at $\lambda = (1 + z) \times 1216 Å$, Bluewards see Ly$\alpha$ forest absorption due to intervening HI.

Bluewards of $\lambda = (1 + z) \times 1025 Å$ we get Ly-β forest superimposed.

Note decline in flux bluewards: Intrinsic quasar power law + falling detector efficiency.

See truncation at $\lambda \approx 4000 Å$ due to LLS at $z \approx 3.4$.

Note absorption redwards of Ly$\alpha$ : Metals!

Taxonomy of Lyman alpha absorbers and the connection to $N_{HI}$:
1. $N_{HI} < 10^{14}$ cm$^{-2}$: no saturation
2. $10^{14} < N_{HI} < 10^{17}$: saturation depending on line width in velocity space
3. $N_{HI} > 10^{17}$: Lyman limit system (LLS)
4. $N_{\text{HI}} > 2 \times 10^{20}$: Damped Lyman alpha (DLA) system

Explain Voigt profile and curve of growth.

DLAs are so named because at those $N_{\text{HI}}$ the Lorentzian damping wing of the Voigt profile becomes prominent. In fact, the division between DLA & LLS is purely historical, as modern spectrographs can detect damping wings down to $N_{\text{HI}} \sim 10^{19}$ (now called “sub-DLAs”). Looking thru ISM of MW would give $N_{\text{H}} \gg 10^{21}$ cm$^{-2}$, i.e. a DLA.

LLS are so named because at $N_{\text{HI}} > 10^{17}$ gas is optically thick at the Lyman limit, i.e. all incident ionizing photons are absorbed; this is a physical definition. Below this is known as Ly$\alpha$ forest, although “partial” LLS at $N_{\text{HI}} > 10^{16}$ are still generally thought to be associated with galaxies’ ISMs (e.g. seen to contain low-ionization metals).

Metal lines and BALs.

Characterizing Ly$\alpha$ forest absorbers

Each absorber can be fit with three parameters: redshift, column density $N_{\text{HI}}$, and line width (called $b$-parameter). High-resolution spectra are fit with a large number of superposition of Voigt profiles. Note for $N_{\text{HI}} \ll 10^{20}$ the Voigt profile is just a Gaussian, as the Lorentzian damping wing is unimportant.

The most basic counting statistic is column density distribution: Number of lines per unit redshift interval per unit log column density. Observed to be a power law $f(N_{\text{HI}}) \propto N_{\text{HI}}^{-\beta}$, with $\beta \sim 1.5$ at $z \sim 3$.

Aside: Note that Ly$\alpha$ forest is much better characterized at $z > 2$ where Ly$\alpha$ falls into optical, than at $z < 2$ where space-based UV observations are required!

The statistic associated with $b$-parameter is just a histogram called the linewidth distribution. At $z \sim 3$, this peaks at $\sim 25$ km/s, with a roughly truncated Gaussian to smaller $b$ and a long tail to large $b$.

Generally this is understood in terms of a thermal component and a bulk flow component:

$$b^2 = \frac{2kT}{m_p} + b^2_{\text{bulk}}$$

(3)

As we will see next time, $b_{\text{bulk}}$ arises because Ly$\alpha$ clouds are large and still expanding with Hubble flow, causing an intrinsic redshift spread within the cloud.

Minimum $b$ (i.e. the low-$b$ truncation) corresponds to pure thermal broadening. Measuring this is complicated by the fact that lines are heavily blended and not necessarily pure Gaussian (as we will see next time), hence small $b$ lines can be wings of a larger line rather than a true
low-\(T\) absorber. Nevertheless, it is possible to estimate \(b_{\text{thermal}} \approx 1.5 - 2 \times 10^4\) K (Schaye et al 2000).

The statistic associated with redshift is line-of-sight clustering. High-quality data confirmed Sargent et al’s result that Ly\(\alpha\) lines have low clustering, much less than e.g. galaxies. In detail, clustering increases with \(N_{\text{HI}}\), with a slight signature first appearing at \(N_{\text{HI}} \sim 10^{14}\).

**Redshift evolution**

Ly\(\alpha\) forest undergoes a transition at \(z \sim 1 - 1.5\), from rapidly-evolving to slowly-evolving. Initially postulated as two absorber populations. Absorbers expanding with Hubble follow, or are self-gravitating clouds. Now let’s consider a static population of small neutral clouds, with fixed comoving number density \(\phi_0\) and comoving radius \(r_0(1 + z)\) (i.e. constant physical radius \(r_0\)). Then in a path length \(\Delta z\) one expects to find the following number of absorbers:

\[
\frac{dN}{dz} \Delta z = \pi r_0^2 \phi_0 (1 + z)^2 \frac{dS}{dz} \Delta z,
\]

where \(S(z) = \int c dz / H(z)\) is the comoving path length. So

\[
\frac{dN}{dz} \propto (1 + z)^2 \frac{dS}{dz} \propto \frac{(1 + z)^2}{H(z)}
\]

Putting in a rough value of \(H(z) \propto (1 + z)^{3/2}\) (for \(\Omega = 1\)), we get \(\gamma = 0.5\). This agrees with low-\(z\) evolution.

Hence the idea was that high-\(z\) absorbers were dominated by large, uncollapsed clouds while low-\(z\) were essentially galaxy/halo gas (whose size and comoving number density evolve slowly since \(z = 1\)). It turns out this interpretation, like confinement models, was shown to be incorrect in the “new” model for the Ly\(\alpha\) forest. While the low-\(z\) scaling was correct in the cloud model, the amplitude for realistic sizes of gaseous halos was too small; basically, it is very rare for a random LOS to pass thru a galactic halo.

Instead the change in evolution reflects mostly an evolution in the photoionization rate \(\Gamma_{\text{HI}}\), which drops rapidly since \(z \sim 2\) owing to the rapid decline in the quasar number density.

**FGPA**

We have Gunn-Peterson optical depth:

\[
\tau = 6.4 \times 10^5 h^{-1} \left(\frac{\Omega_b h^2}{0.02}\right) \left(\frac{1 + z}{3}\right)^{3/2} \left(\frac{n_{\text{HI}}}{n_H}\right)
\]
Changing the density alters the ionization balance and the amount of neutral Hydrogen. Hence, density fluctuations produce variations in the optical depth and produce the Lyman alpha forest. Ignoring peculiar velocities and thermal broadening, then the optical depth depends only on the underlying baryon density. Because pressure forces as small, the baryon density traces the dark matter density.

Thus the Lyα forest represents a 1-D probe of the spectrum of mass fluctuations. Lyα lines should not be thought of as individual features, but as a continuous flux distribution. This is the Fluctuating Gunn Peterson Approximation.

The Lyα forest arises from absorption of Lyα photons by neutral hydrogen gas in the intergalactic medium (eq. 2). Assuming local photoionization-recombination equilibrium, we have:

\[ n_{\text{HI}} \Gamma = n_{\text{HII}} n_e \alpha(T), \tag{6} \]

where \( n_{\text{HI}} \), \( n_{\text{HII}} \) and \( n_e \) are the local densities of neutral, ionized hydrogen and electrons in the IGM, respectively, \( \Gamma \) is the photoionization rate, and \( \alpha(T) \) is the recombination coefficient at temperature \( T \) (Abel et al. 1997),

\[ \alpha(T) = 4.2 \times 10^{-13} (T/10^4 \text{K})^{-0.7} \text{cm}^3\text{s}^{-1}. \tag{7} \]

The photoionization rate \( \Gamma \) is related to the ionizing background flux \( J_\nu \) by,

\[ \Gamma = 4\pi \int \frac{J_\nu}{h\nu} \sigma_\nu d\nu, \tag{8} \]

where the integral includes the ionizing photons from the HI Lyman Limit to the HeII Lyman Limit assuming that the He in the IGM is singly ionized, and \( \sigma_\nu \) is the HI cross section of ionizing photons, \( \sigma_\nu \sim \nu^{-3} \).

If the IGM is mostly ionized by a uniform ionizing background, the evolution of the optical depth can be expressed as (Weinberg et al. 1997),

\[ \tau_{\text{GP}} \propto \frac{(1 + z)^6 \Omega_b h^2)^2 \alpha(T)}{\Gamma H(z)} \propto \frac{(1 + z)^{4.5} \Omega_b h^2)^2 \alpha(T)}{h \Gamma \Omega_m^{0.5}} \tag{9} \]

The Lyα absorption increases rapidly with increasing redshift even if the ionizing background remains constant with redshift.

\[ \tau \propto n_{\text{HI}} \propto n_{\text{HI}}^2 T^{-0.7}/\Gamma_{\text{HI}} \tag{10} \]

And the combination of adiabatic cooling and photoionization heating in an expanding universe gives a relation:

\[ T \propto \rho^{0.6}. \]
Use FGPA to measure:

- **Matter power spectrum** on intermediate scales at high redshifts. Since \( \rho_{\text{gas}} \propto \rho_{\text{dark}} \propto \tau_{\text{HI}}^{1.6} \), we can compute fluctuations in \( \rho \) from fluctuations in \( \tau_{\text{HI}} \). Obtaining \( \tau_{\text{HI}} \) from line-of-sight fluxes is complicated by saturation; fortunately, at \( z \sim 2-4 \) saturated lines comprise a tiny portion of the spectrum.

Scales: Can’t probe too close to thermal broadening, since assumption of gas-follows-DM breaks down in velocity space. So hard limit < 10 km/s (30 kpc comoving at \( z = 3 \)). For precision work, limit is much larger, so this limits to >several hundred kpc.

To obtain large samples, need faint targets hence low-res spectra. For ex SDSS spectra good enough for this (cf. BOSS), with resolution of \( \sim 200 \) km/s, i.e. \( \sim 1 \) Mpc.

On large scales, no limit in principle, but one must know that \( \Gamma_{\text{HI}} \) is spatially uniform (or know its variations). Since QSOs dominate, it is probably non-uniform on clustering scale of QSOs, which is tens of Mpc.

So FGPA effective on scales of \( \sim 1-10 \) Mpc. Also probes an interesting \( z \) range, between LSS (\( z < 1 \)) and CMB (\( z = 1088 \)), and generally smaller scales. Hence Ly\( \alpha \) forest, along with WMAP and LSS (e.g. BAO, clusters, or WL), provides the three pillars for constraining cosmological parameters and (eventually) dark energy.

- **Baryon density** (\( \Omega_b h^2 \approx 0.02 \)). Since \( \bar{\tau}_{\text{HI}} \propto \bar{\rho}_{\text{gas}}^{1.6} \Gamma_{\text{HI}}^{-1} \), by measuring mean optical depth and estimating photoionization rate it is possible to get mean baryon density, i.e. \( \Omega_b \).

Estimate by Rauch et al. (1997) found \( \Omega_b > 0.017 h^{-2} \), consistent with WMAP and BBN.

BBNS ([D/H]) can constrain \( \Omega_b \), but in early 90’s gave conflicting values. Deuterium lies 82 km/s bluewards of Ly\( \alpha \), but is down by 5 orders of mag in \( \tau \). So must find strong but narrow line. In late 90’s, Burles & Tytler had series of papers giving \( \Omega_b \sim 0.02 h^{-2} \), confirming Ly\( \alpha \) forest results. Then WMAP clinched it.

These days, argument can be turned around to measure \( \Gamma_{\text{HI}} \) given \( \Omega_b \); in fact, this gives the most precise measure of \( \Gamma_{\text{HI}} \) at both high and low redshift!

- **Metallicity of IGM.** Heavy elements are observed in the diffuse IGM as metal absorbers, e.g. CIV. From this can measure ratios like CIV/HI. What is desired is metallicity, e.g. [C/H]. Need ionization corrections. Using \( \tau_H I \), one can estimate \( \rho_{\text{gas}} \) and \( T \) using FGPA, which then give the ionization corrections for C and H, allowing one to correct CIV/HI into C/H.

**IGM metals**
Observed IGM metal lines: CIV, SiIV, NV, OVI, CIII, CII, MgII, others. Usually identified by doublet pair (CIV, SiIV, OVI); others identified by correlation with another line. CIV is most commonly observed at high-z, followed by SiIV. OVI is expected to be common as well, but it lies in the Lyα forest so at high-z is difficult to pick out; better characterized at low-z. Low-ionization lines like CII and MgII are typically associated with strong (near LLS) absorbers, hence are probably dense halo or ISM gas.

Use FGPA to get ionization corrections ($\tau_{HI} \rightarrow \rho \rightarrow T$), allowing conversion from e.g. CIV/HI to C/H in the data. Metallicity is about $[C/H]_\odot \approx -2.5$ in CIV, which traces moderately overdense gas.

**Warm-Hot IGM**

Can we use FGPA at low-z? Not so fast. First, fewer lines. Second, data is lower quality. But also phase space character of IGM changes.

At low-z, structure formation shock heats IGM gas. Perhaps heating from winds as well. This causes WHIM to contain approx half of baryons today (model prediction).

How to detect/confirm? No H/He absorption, too low $\rho$ for emission. So need to rely on metals. Strongest absorbers are high-ionization oxygen lines: OVI(1032,1037 Å), OVII(43 Å), OVIII(23 Å). Latter two are soft X-ray. OVII traces bulk of WHIM temperature range.

OVI observed, wide range of conditions (some w/HI, some w/CIV, some with neither). Simulations suggest $dN/dz$ of OVI absorbers best explained by a WHIM metallicity of $\sim 0.1Z_\odot$.

Measure $N_{OVI}$, assume physical extent to get $\rho_{OVI}$, assume max ionization correction 20% to get $\rho_O$, assume $0.1Z_\odot$ to get $\rho_{WHIM}$. Gives $\Omega_{WHIM} \approx 0.2 - 0.5\Omega_b$. Plausible, but lots of assumptions.