How the first generations of luminous baryons established the X-ray and UV backgrounds

Andrei Mesinger∗†
Princeton University
E-mail: mesinger@astro.princeton.edu

The first generations of astrophysical objects made a substantial impact on our Universe with their radiation. X-rays from the first sources, with large mean free paths, likely quickly heated the intergalactic medium (IGM). The second generation of 21cm instruments can provide a unique view into this early epoch. The early stages of reionization likely followed, driven by so-called "mini-halos", i.e. molecularly-cooled halos. These small halos were susceptible to complex feedback mechanisms, especially from the soft-UV background which preceded reionization, resulting in complex and possibly extended early stages of reionization. When atomically-cooled galaxies emerged as the dominant ionizers, reionization could proceed more rapidly, with these being less sensitive to radiative feedback than previously thought. Reionization could have slowed in the final stages when the ionized bubbles grew larger than the separation of Lyman limit absorption systems (LLSs). The final stages likely involved the photo-evaporation of LLSs, which by then regulated the rise of the UV background. I discuss the theoretical underpinnings of this narrative, as well as how future 21cm observations may help shed light on the outstanding uncertainties.
1. Intro

The dawn of the first astrophysical objects and the eventual reionization of the Universe is of fundamental importance, offering insight into many astrophysical processes. Also, since observational surveys only pick out the rarest, most luminous objects, cosmic radiation fields and their effects on the intergalactic medium (IGM) may be our only way of studying the majority of our galactic ancestors. Recent theoretical and observational advances are starting to shed light on a complex and extended reionization epoch. Here I sketch-out three likely stages of this process. I briefly discuss the potential of the redshifted 21cm signal to probe not only reionization, but also the pre-reionization epochs. However, interpreting the upcoming observations requires efficient modeling tools, such as the recently developed 21cmFAST.

2. Once upon a time... the story of cosmic dawn and reionization

2.1 Early stages

Hierarchal structure formation implies that the first astrophysical objects are likely hosted by so-called minihalos (with virial temperatures \( T_{\text{vir}} \lesssim 10^4 \) K). Cooling via atomic hydrogen is inefficient at such low temperatures, so baryons condense into these halos via the molecular hydrogen (H\(_2\)) cooling channel. Without metals to aid in cooling and fragmentation, the first generations of stars (so-called PopIII stars) are likely massive (with masses of \( \sim 100M_\odot \); [1, 5, 81]), yet short-lived (with lifetimes of a few Myr; [61, 62]).

PopIII stars are likely to have very different properties from “normal” PopII stars, with harder spectra, a factor of ten times more ionizing photons per baryon [61], and various exotic, mass-dependent fates (e.g., [29]). Accurately determining their properties, formation efficiency and initial mass functions (IMFs) requires detailed numerical simulations and costly explorations of parameter space (e.g., [73]).

These prehistoric giants have complex interactions with their surroundings, since star formation in such small-mass halos is susceptible to feedback mechanisms. Modeling these requires introducing parameterized prescriptions for stellar properties and extending simulations to larger, \( \sim 1 \) Mpc scales. Feedback processes can come in three forms: (1) \textit{mechanical}: supernovae (SNe) can blow out gas from the surrounding shallow potential well, delaying local star formation until gas is re-accreted (e.g., [74, 75]); (2) \textit{chemical}: evolution in the IMF depends on metal enrichment [68, 65, 63], which is very inhomogeneous making it necessary to model the enrichment on large scales (e.g., [70]); and (3) \textit{radiative}: radiation in various bands can either promote (positive feedback) or suppress (negative feedback) future star formation.

Because it acts on very large scales and involves a large dynamic range in energy, it is likely that radiative feedback is the most pivotal and complex of the feedback mechanisms. Positive radiative feedback can result when the enhanced free-electron fraction from ionizing photons (e.g., [56]) or hydrodynamical shocks [64] catalyzes the formation of H\(_2\), thereby enhancing the H\(_2\) cooling channel. Negative feedback can result from heating by ionizing radiation which can photo-evaporate gas in low-mass halos (e.g., [11]). Also, an active background of Lyman-Werner (LW) radiation (with photon energies in the 11.18–13.6eV range) can dissociate H\(_2\), thus decreasing the gas’s cooling capabilities (e.g., [28]).
The first astrophysical objects can emit radiation in various bands. X-rays, generated from accretion onto remnant black holes or massive X-ray binaries, have long mean free paths, but their feedback effects seem to be mild [38, 34]. However, they can be responsible for early IGM heating and some degree of reionization (e.g., [9, 59, 50]). Feedback resulting from a LW background is by definition negative, and can delay star formation in low mass halos, depending on the strength of the background [37, 38, 82, 58]. Also depending on its strength, a transient (due to the short lifetimes of PopIII stars) ionizing UV background (UVB) can have various feedback regimes [57, 46, 47, 75, 74], especially in conjunction with the negative feedback effects of a LW background [46, 47, 75]. Despite this complexity, it is likely that the final suppression of star formation inside minihalos was regulated by the build-up of persistent UV or LW backgrounds before the bulk of reionization [25, 26]. However, the complexity of radiative feedback and/or an early X-ray background could precipitate an extended early reionization epoch, which could allow a significant amount of reionization to occur at late times ($z \sim 6$), and still be consistent with constraints from the Wilkinson Microwave Anisotropy Probe (WMAP) [35].

2.2 Middle stages

After the initial, “fireworks” stage of PopIII-driven reionization, longer-lived stars begin to establish persistent UVBs. Reionization proceeds in an “inside-out” fashion on large-scales, with biased sources driving HII regions to tens of Mpc in size (e.g., [21, 19, 84]). Larger, atomically-cooled halos ($T_{\text{vir}} \gtrsim 10^4$ K) start dominating the ionization budget. These are likely too faint to image directly with current or upcoming instruments (e.g., [60, 80]), although we can get close with gravitational lensing surveys (e.g., [67, 52, 79]). Likewise, because of the inside-out nature of reionization, most astrophysical objects during the epoch of reionization find themselves inside large HII regions, largely unaffected by the neutral IGM even when it still occupies a substantial volume filling fraction. Thus most of the current observational probes of reionization (derived from quasar proximity regions, damping wings in quasar and gamma ray burst (GRB) spectra, Ly$\alpha$ emitter (LAE) number density and clustering properties; see below) are most effective back when the Universe is predominantly neutral.

Aside from these observational challenges, there are substantial theoretical hurdles to overcome when studying the middle stages of reionization. Ionization structures during this epoch can span volumes over four magnitudes larger than the first-stars simulations discussed in the previous section (see Fig. 1). To statistically capture the middle stages of reionization, simulation boxes need to be hundreds of Mpc in length, yet still be able to account for ionizing radiation from small-mass halos. The required dynamic range is daunting, and state-of-the-art simulations are forced to ignore minihalos and photon sinks, while also using approximate prescriptions to assign ionizing luminosities to source halos (see the recent review in [72]). More approximate techniques have recently been developed in order to overcome the challenge of such a large range of scales, including sub-grid [33, 43] and semi-numerical [83, 49, 23, 2, 7, 69, 50] models.

Additionally, our poor understanding of the early Universe means that large explorations of parameter space are required for any robust conclusion. Analytic models have been very useful in this respect, predicting that HII morphology is not very sensitive to redshift, when normalized to the same mean neutral fraction, $\bar{x}_{\text{HI}}$, and also that sources are more important that ionizing sinks in this regime [21, 19]. These predictions were confirmed by the first numerical parameter studies
How the first generations of luminous baryons established the X-ray and UV backgrounds

Andrei Mesinger

2.3 Final stages

As reionization progresses, HII regions grow to be larger than the mean free path of ionizing photons through them, requiring even larger simulation boxes (see Fig. 1). In this final regime, sinks of ionizing photons, such as Lyman limit systems (LLSs) could regulate the progress of reionization [18], and its morphological structure [19, 7]. These absorption systems are likely found in the filaments of the recently-ionized IGM, where the UVB is the weakest [8]. The final “overlap” stages of reionization could be delayed until either absorbers get photo-evaporated (e.g. [31]), or the remaining neutral islands get ionized from the inside. The time-scales for these processes are of order $\Delta z \sim 1$. This, combined with the large cosmic variance during this epoch (e.g., [45]), suggests
that the overlap stage is much more gradual than was initially predicted by small-box simulations (e.g., [24]). It is also unlikely that these final stages leave a strong observational imprint, such as a sharp rise in the UVB [18]. Despite popular belief, current observations cannot rule-out a very late overlap, even at $z \sim 5$ [36, 45, 41].

2.4 Observational constraints

Several model-dependent constraints on $\bar{x}_{\text{HI}}$ at $z \sim 6$ have been derived from various astrophysical probes such as: (1) the size of the proximity zone around quasars ([78, 6], but see [55, 4, 40]); (2) a claimed detection of damping wing absorption from neutral IGM in quasar spectra ([53, 54], but see [51]); (3) the non-detection of intergalactic damping wing absorption in a GRB spectrum ([71], but see [44]); and (4) the number density and clustering of Ly$\alpha$ emitters ([39, 27, 22, 32, 42], but see [10, 52, 30]). Additionally, a direct lower limit on $\bar{x}_{\text{HI}}$ was recently derived from the dark covering fraction in the Ly$\alpha$ and Ly$\beta$ forests [41].

However, constraining the ionization state of the high-redshift Universe is very difficult with current observations, since they are not direct probes of HII morphology. For example, the Lyman forests begin to saturate at $z \gtrsim 5$, becoming difficult to interpret (e.g., [66, 12, 3, 36, 45]). Similarly, the dominant population of ionizing galaxies is $\sim$ two orders of magnitude fainter than current detection limits; and in any case, galaxies offer a very biased tracer of the cosmic radiation fields. Luckily, there is a plethora of upcoming observations which should help advance our understanding: 21cm tomography, high-redshift IR spectra, wide-field LAE surveys, and improved E-mode CMB polarization power spectra. Of these, arguably the 21cm line from neutral hydrogen provides the greatest potential for directly studying reionization in the near term.

3. The 21cm frontier and 21cmFAST

The cosmological 21cm signal uses the CMB as a back-light. The offset of the 21cm brightness temperature from the CMB temperature, $T_b$, along a line of sight (LOS) at observed frequency $v$, can be written as (c.f. [20]):

$$
\delta T_b(v) = 27x_{\text{HI}} \left( 1 - \frac{T_f}{T_S} \right) \left( 1 + \delta_{\text{nl}} \right) \left( \frac{H}{dv_r/dr + H} \right) \left( 1 + z \right) \left( \frac{0.15}{10} \frac{\Omega_m h^2}{0.023} \right)^{1/2} \text{mK},
$$

(3.1)

where $x_{\text{HI}}$ is the hydrogen neutral fraction, $T_S$ is the gas spin temperature, $\tau_{v_0}$ is the optical depth at the 21cm frequency $v_0$, $\delta_{\text{nl}}(x, z) \equiv \rho/\bar{\rho} - 1$ is the evolved (Eulerian) density contrast, $H(z)$ is the Hubble parameter, $dv_r/dr$ is the comoving gradient of the line of sight component of the comoving velocity, and all quantities are evaluated at redshift $z = v_0/v - 1$. The above equation also assumes that $dv_r/dr \ll H$, which is generally true for the pertinent redshifts and scales.

The redshifted 21cm line carries with it much information, containing both astrophysical ($x_{\text{HI}}$ and $T_S$) and cosmological (last four terms in eq. 3.1) terms (see [16, 17]). However, in order to make sense of the upcoming observations, one must disentangle the various components. Considering the scales and uncertainties involved, this is a daunting task.

Recently “semi-numerical” tools have proven invaluable in overcoming these obstacles. These simulations use more approximate physics than numerical simulations, but are much faster. In particular, the recent launch of 21cmFAST [50] (http://www.astro.princeton.edu/~mesinger/Sim)
How the first generations of luminous baryons established the X-ray and UV backgrounds
Andrei Mesinger

provided the astronomical community a publicly-available tool specialized for fast, large-scale simulations of the 3D cosmological 21cm signal. For example, a 256³ realization of reionization with 21cmFAST is obtained in just a few minutes on a single CPU, with every step agreeing with cosmological hydrodynamic simulations into the quasi-non-linear regime [49, 84, 50]. Such speeds facilitate rapid explorations of the vast parameter space, necessary for interpreting upcoming observations.

Most studies thus far have focused on the reionization term, $x_{HI}$, making various simplifying assumptions in eq. 3.1. The 21cm signal during reionization can indeed tell us much about the relevant physical processes and intergalactic radiation fields. However, the 21cm signal also offers us a precious glimpse into even earlier epochs, via the spin temperature, $T_s$, term.

The spin temperature can be written as (e.g. [20]):

$$T_s^{-1} = \frac{T^{-1}_\gamma + x_\alpha T^{-1}_\alpha + x_c T^{-1}_K}{1 + x_\alpha + x_c}$$

(3.2)

where $T_K$ is the kinetic temperature of the gas, and $T_\alpha$ is the color temperature, which is closely coupled to the kinetic gas temperature, $T_\alpha \approx T_K$ [14]. From this equation, one can see that the spin temperature interpolates between the CMB and the gas temperature. It does this through two coupling coefficients: (1) the collisional coupling coefficient, $x_c$, which requires high densities and is effective in the IGM at $z \gtrsim 40$; and (2) the Wouthuysen-Field ([77, 13]; WF) coupling coefficient which uses the Ly$\alpha$ radiation background and is effective soon after the first sources ignite. If either coefficients is high, the spin temperature approaches the kinetic temperature of the gas. Otherwise, the spin temperature approaches the CMB temperature and there is no signal.

The CMB temperature decreases as $\propto (1 + z)^2$. After decoupling from the CMB, the gas kinetic temperature decreases as $\propto (1 + z)^2$, before being heated, most likely with X-rays from the first sources (see [15] and the discussion therein). Likely before the bulk of reionization, the gas is heated to temperatures far beyond the CMB, $T_K \gg T_\gamma$, and the spin temperature term in eq. 3.1 no longer contributes to the signal. When and where $T_s < T_\gamma$, the 21cm signal is seen in absorption, and when/where $T_s > T_\gamma$, the signal is seen in emission.

Therefore, the 21cm signal and its inhomogeneity could be a powerful probe, not only of reionization, but also of the Ly$\alpha$ background, IGM heating, and cosmology during the collisionally coupled regime before the first sources ignite. Another probe of cosmology would be available at lower redshifts (making it easier to detect) if heating completes before reionization. A movie of 21cm evolution from $z = 250$, computed with 21cmFAST, is available at http://www.astro.princeton.edu/~mesinger/21cm_Movie.html, together with an explanation of the various interesting epochs.

4. Conclusions

- Milestones such as reionization are likely the only practical way of studying the primordial zoo of astrophysical objects in the near future.
- Reionization is likely extended, going through various stages.
- The cosmological 21cm signal is very rich in information, containing both cosmological and astrophysical components.
- The range of scales and unknown parameter space is enormous.
How the first generations of luminous baryons established the X-ray and UV backgrounds
Andrei Mesinger

• We need efficient modeling tools, such as 21cmFAST, to make sense of the upcoming observations.
• We are living in exciting times!

Acknowledgments
I would like to thank all of my collaborators who contributed to the science discussed here. Support for this work was provided by NASA through Hubble Fellowship grant HST-HF-51245.01-A, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555.

References
How the first generations of luminous baryons established the X-ray and UV backgrounds

Andrei Mesinger


How the first generations of luminous baryons established the X-ray and UV backgrounds

Andrei Mesinger


How the first generations of luminous baryons established the X-ray and UV backgrounds


How the first generations of luminous baryons established the X-ray and UV backgrounds

Andrei Mesinger


