The Infancy of Cosmic Reionization

Rennan Barkana\textsuperscript{1,2,3,*}

\textsuperscript{1} Institute for Cosmic Ray Research, University of Tokyo, Kashiwa 277-8582, Japan
\textsuperscript{2} Division of Physics, Mathematics and Astronomy, California Institute of Technology, Mail Code 130-33, Pasadena, CA 91125, USA
\textsuperscript{3} Guggenheim Fellow; on sabbatical leave from the School of Physics and Astronomy, Tel Aviv University, Israel

3 September 2008

ABSTRACT

We consider the early stages of cosmic hydrogen or helium reionization, when ionizing sources were still rare. We show that Poisson fluctuations in the galaxy distribution substantially affected the early bubble size distribution, although galaxy clustering was also an essential factor even at the earliest times. We find that even at high redshifts, a significant fraction of the ionized volume resided in bubbles containing multiple sources, regardless of the ionizing efficiency of sources or of the reionization redshift. In particular, for helium reionization by quasars, one-source bubbles last dominated (i.e., contained 90% of the ionized volume) at some redshift above \(z = 7\), and hydrogen reionization by stars achieved this milestone at \(z > 23\). For the early generations of atomic-cooling halos or molecular-hydrogen-cooling halos, one-source ionized regions dominated the ionized volume only at \(z > 31\) and \(z > 48\), respectively. To arrive at these results we develop a statistical model for the effect of density correlations and discrete sources on reionization and solve it with a Monte Carlo method.

Key words: galaxies:high-redshift – cosmology:theory – galaxies:formation

1 INTRODUCTION

The earliest generations of stars are thought to have transformed the universe from darkness to light and to have reionized and heated the intergalactic medium. Knowing how the reionization process happened is a primary goal of cosmologists, because this would tell us when the early stars formed and in what kinds of galaxies. The strong fluctuations in the number density of galaxies, driven by large-scale density fluctuations in the dark matter, imply that the dense regions reionize first, producing on large scales an inside-out reionization topology (Barkana & Loeb 2004). This basic picture has been studied and confirmed with detailed analytical models (Furlanetto et al. 2004), semi-numerical methods (Mesinger & Furlanetto 2007), and by a variety of large numerical simulations (Mellema et al. 2006; Zahn et al. 2007; Trac & Cen 2007) that solve gravity plus radiative transfer. The distribution of neutral hydrogen during reionization can in principle be measured from maps of 21-cm emission by neutral hydrogen (Madau et al. 1997), although upcoming experiments such as the Murchison Widefield Array (MWA\textsuperscript{1}) and the Low Frequency Array (LOFAR\textsuperscript{2}) are expected to be able to detect ionization fluctuations only statistically (for reviews see, e.g., Furlanetto et al. 2006; Barkana & Loeb 2007).

The infancy of cosmic reionization, when only a small fraction of the volume of the universe was ionized, is of interest for a number of reasons. First, when ionizing sources were rare at early times, they are expected to have formed separate \(\text{H II}\) bubbles which if observed can be used to study directly the properties of individual sources and their surroundings (Cen 2006), without the complications of later times, when overlapping bubbles imply that galaxy clustering dominates the ionization distribution and the 21-cm power spectrum. Second, when ionization fluctuations disappear over much of the universe, it becomes possible to use the 21-cm technique for other applications including those of fundamental cosmology, without the complications of ionization fluctuations which are intrinsically non-linear (since the ionization fraction varies from 0 to 1). Major such applications include measurements of the density power spectrum (Hogan & Rees 1974; Scott & Rees 1990), of fluctuations in the Ly\(\alpha\) radiation emitted by the first galaxies (Barkana & Loeb 2005; Pritchard & Furlanetto 2006; Chuzhoy, Alvarez & Shapiro 2006), and of fluctuations in the rate of heating from early X-rays (Pritchard & Furlanetto 2007). If ionization fluctuations are negligible then the angular anisotropy of the 21-cm power spectrum makes it possible to measure separately various fluctuation sources, including in particular the cosmologically-interesting baryonic density power spectrum.

\textsuperscript{*} E-mail: barkana@wise.tau.ac.il
\textsuperscript{1} http://www.haystack.mit.edu/ast/arrays/mwa/
\textsuperscript{2} http://www.lofar.org/
(Barkana & Loeb 2005a). On small scales, the existence of H II bubbles (even when rare) affects the fluctuations in Lyα and X-ray radiation, producing a small-scale cutoff in the 21-cm power spectrum that can be used to detect and study the population of galaxies that formed just 200 million years after the Big Bang (Naoz & Barkana 2008).

While analytical models and numerical simulations exist that can be used to study the later epochs of reionization, the early times are very difficult to investigate. Simulations, which in general must overcome the huge disparity between the large characteristic scales of galaxy clustering at high redshift and the small scales of individual galaxies (Barkana & Loeb 2004), are stretched even further at early times, when ionizing sources become very rare and even larger cosmological volumes are required in order to assemble a reasonable statistical sample. As discussed in detail below, current analytical models based on the model of Furlanetto et al. (2004) account for galaxy clustering but are based on continuous variables and cannot account for the fact that galaxies are discrete sources. This discreteness becomes a crucial factor in the early stages of reionization, when the number of ionizing sources per bubble is small. In this limit, Poisson fluctuations also become substantial, weakening the correlation between the galaxy distribution and the underlying large-scale density fluctuations in the dark matter. Discreteness can also play a significant role during the central stages of reionization, particularly in the case of He reionization by quasars, which are rare sources believed to form only in massive halos that correspond to many-σ density fluctuations at high redshift. These various aspects of discrete sources are not accounted for in current analytical models. Furlanetto & Oh (2008) considered helium reionization and showed that the continuous models break down when discreteness is important. They suggested to instead use a pure stochastic Poisson model, without halo correlations, when He is less than ∼ 50% ionized globally.

In this paper we develop a model that accounts for discrete sources as well as density correlations. We solve the model with a Monte Carlo method and use it to show that galaxy correlations play a major role even in the infancy of cosmic reionization. Isolated one-source bubbles do dominate at sufficiently high redshifts, but the pure stochastic Poisson model is essentially never a good description of the bubble size distribution. In the next section we first review the basic setup of the random walk problem in the context of reionization. We then show how the standard approach can be generalized to solve for the bubble size distribution including Poisson fluctuations.

2 MODEL

Analytical approaches to galaxy formation and reionization are based on the mathematical problem of random walks with barriers. The statistics of a random walk with a barrier can be used to calculate various one-point distributions, including the distribution of ionized bubble sizes during reionization (Furlanetto et al. 2004). This distribution indicates how likely it is for each scale to determine whether a given point is ionized. As such, it indicates the relative importance of various scales in reionization, yielding important intuition about the internal dynamics of reionization. If bubbles of a given radius $R$ are common, this produces a strong correlation in the neutral fraction (and thus 21-cm emission) on a scale $\sim R$, since the ionization states of two points separated by up to $R$ are then often coupled. Calculations of the 21-cm correlation function using two-point extensions of the model yield reasonable agreement with numerical simulations (Furlanetto et al. 2004; Zahn et al. 2005; Barkana 2007) and indicate that the main feature of the power spectrum during reionization, i.e., enhanced large-scale power, indeed appears on scales corresponding to the most likely bubble sizes.

In this section we first review the basic setup of the random walk problem in the context of reionization. We then show how the standard approach can be generalized to solve for the bubble size distribution including Poisson fluctuations.

2.1 Reionization: basic setup

The basic approach for using random walks with barriers in cosmology follows Bond et al. (1991), who used it to rederive and extend the halo formation model of Press & Schechter (1974). In this approach we work with the linear overdensity field $\delta(x, z) \equiv \rho(x, z)/\bar{\rho}(z) - 1$, where $\rho$ is the mean value of the mass density $\rho$. In the linear regime, the overdensity grows in proportion to the linear growth factor $D(z)$ (defined relative to $z = 0$), making it possible to extrapolate the initial density field at high redshift to the present by multiplication by the relative growth factor. Thus, in this paper the density $\delta$ and related quantities refer to their values linearly-extrapolated to the present. In each application there is in addition a barrier that signifies the critical value (as a function of scale) which the linearly-extrapolated $\delta$ must reach in order to achieve some physical milestone, which here corresponds to having a sufficient number of galaxies within some region in order to fully reionize it.

Considering an arbitrary point $A$ in space (at a given $z$), we calculate as follows its probability of being inside H II bubbles of various sizes (Furlanetto et al. 2004). We consider the smoothed density around this point, first averaging over a large scale or, equivalently, including only small comoving wavenumbers $k$. We then average over smaller scales (i.e., include larger $k$) until we find the largest scale on which the averaged overdensity is higher than the barrier; in the application to reionization, we then assume that the point $A$ belongs to an H II bubble of this size. Mathematically, if the initial density field is a Gaussian random field and the smoothing is done using sharp $k$-space filters, then the value
of the smoothed $\delta$ undergoes a random walk as the cutoff value of $k$ is increased. Instead of using $k$, we adopt the (linearly-extrapolated) variance $S$ of density fluctuations as the independent variable. While the solutions are derived in reference to sharp $k$-space smoothing, we follow the traditional extended Press-Schechter approach and substitute real-space quantities in the final formulas. In particular, $S$ is calculated as the variance of the mass $M$ enclosed in a spatial sphere of comoving radius $r$.

The appropriate barrier for reionization was derived by Furlanetto et al. (2004), who noted that the ionized fraction in a region is given by $x^i = \zeta F_{\text{coll}}$, where $F_{\text{coll}}$ is the collapse fraction (i.e., the gas fraction in galactic halos) and $\zeta$ is the overall efficiency factor, which is the number of ionizing photons that escape from galactic halos per hydrogen atom (or ion) contained in these halos, divided by the number of times each hydrogen atom in the intergalactic medium must be reionized (where this number is assumed to be spatially uniform). In the extended Press-Schechter model (Bond et al. 1991), in a region containing a mass corresponding to variance $S_R$,

$$F_{\text{coll}} = \text{erfc}\left(\frac{\delta_c(z) - \delta}{\sqrt{2(S_{\text{min}} - S_R)}}\right),$$

(1)

where $S_{\text{min}}$ is the variance corresponding to the minimum mass $M_{\text{min}}$ of a halo that hosts a galaxy, $\delta$ is the mean density fluctuation in the given region, and $\delta_c(z)$ is the critical density for halo collapse at $z$. In reality, the cosmic mean halo distribution in simulations is better described by the halo mass function of Sheth & Tormen (2002) (with the updated parameters suggested by Sheth & Tormen 2002). However, an exact analytical generalization is not known for the biased $F_{\text{coll}}$ in regions of various sizes (corresponding to $S_R$) and mean density fluctuations $\delta$.

Barkana & Loeb (2004) suggested a hybrid prescription that adjusts the abundance in various regions based on the extended Press-Schechter formula (Bond et al. 1991), and showed that it fits a broad range of simulation results. In general, we denote by $f(\delta_c(z), S)dS$ the mass fraction contained at $z$ within halos with mass in the range corresponding to variance $S$ to $S + dS$, where $\delta_c(z)$ is the critical density for halo collapse at $z$. Then the biased mass function in a region of size $R$ (corresponding to density variance $S_R$) and mean density fluctuation $\delta$ is (Barkana & Loeb 2004)

$$f_{\text{bias}}(\delta_c(z), \delta, S_R, S) = \frac{f_{\text{ST}}(\delta_c(z), S)}{f_{\text{PS}}(\delta_c(z), S)} f_{\text{PS}}(\bar{\delta}_c(z) - \delta, S - S_R),$$

(2)

where $f_{\text{PS}}$ and $f_{\text{ST}}$ are, respectively, the Press-Schechter and Sheth-Tormen halo mass functions. The value of $F_{\text{coll}}(\delta_c(z), \delta, S_R, S)$ is the integral of $f_{\text{bias}}$ over $S_R$ from 0 up to the value $S_{\text{min}}$ that corresponds to the minimum halo mass $M_{\text{min}}$ or circular velocity $V_c = \sqrt{GM_{\text{min}}/R_{\text{vir}}}$ (where $R_{\text{vir}}$ is the virial radius of a halo of mass $M_{\text{min}}$ at $z$). We then numerically find the value of $\delta$ that gives $F_{\text{coll}} = 1$ at each $S_R$, yielding the exact barrier. Also, in order to compare with a simpler, analytically-solvable model, we derive a linear approximation to the barrier, $\delta(S_R) \approx \nu + \mu S_R$, by numerically finding the value of the barrier function at $S_R = 0$ and its derivative with respect to $S_R$. In general, photon conservation implies that the mean global ionized fraction should equal $\bar{x}^i = \zeta f_{\text{ST}}$ in terms of the cosmic mean collapse fraction.

Barkana (2007) and Barkana & Loeb (2008) used an approximation in which effectively each factor on the right-hand side of equation (2) was integrated separately over $S$, yielding a simple analytical formula for the effective linear barrier. This approximation was also assumed by Furlanetto et al. (2006a) when they stated that this hybrid prescription does not change the bubble size distribution from the pure Press-Schechter case (for a fixed redshift, minimum halo mass, and cosmic mean ionized fraction). Here we solve numerically for the barrier using the exact formulas. We show that the previously-used approximation is not too accurate, especially at the early stages of reionization that are our focus in this paper.

### 2.2 The statistics of a random walk with a barrier and discrete sources

The standard approach presented above treats the random walks as functions of a continuous variable $S_R$, and assumes a one-to-one correspondence between the value of $\delta$ and the ionized fraction $x^i$ at each scale. The statistical distribution of first barrier crossing, which physically corresponds to the bubble size distribution, can be derived analytically for the approximate linear barrier (Furlanetto et al. 2004), and for the exact barrier can be solved with Monte Carlo simulations of random walks or by solving an integral equation (Zhang & Hui 2006).

In reality, there are two additional physical constraints that are neglected in the standard approach: the ionizing sources are discrete, and the ionized fraction (for a given value of $\delta$ in a region) fluctuates due to Poisson fluctuations in the number of galaxies. The discreteness of ionizing sources means that the possible volume of bubbles has a minimum value $V_{\text{bubble}}$ corresponding to the bubble due to a single galaxy hosted by a halo of mass $M_{\text{min}}$. Also, the expected ionized fraction $x^i$ given by the continuous model is subject to Poisson fluctuations, as the actual ionized fraction depends on the number of galaxies. Unlike the standard random walk approach, in which the statistics of the walk depend only on the barrier expressed as a function $\delta(S_R)$, Poisson fluctuations introduce an explicit dependence on the mapping between $S_R$ and scale $R$.

In order to include these discrete aspects in the bubble distribution, we begin with the standard analytical approach, which considers the statistics of spherical volumes of various sizes $R$, all about a point $A$. Given a value $\delta$ on a scale $R$ (with corresponding variance $S_R$), we now treat the continuous ionized fraction $x^i$ of the previous subsection only as an average expectation value. To find its real distribution, we first calculate the mean expected value $\langle j \rangle$ of the number of ionizing sources within the sphere of radius $R$. This (non-integer) value can be calculated from the integral of $f_{\text{bias}}$ $dS$ weighted by $1/M$ (which yields halos weighted by number rather than mass); it depends on $z$, $S_{\text{min}}$, $\delta$, and $S_R$.

The actual value of $j$ is given by a Poisson distribution with mean equal to $\langle j \rangle$. To find the actual $x^i$, we find the mass of each of the $j$ halos according to the halo mass distribution given by $f_{\text{bias}}$; note that this procedure does not involve a single, fixed mass distribution since $f_{\text{bias}}$ is a function of $\delta$ and $S_R$. 
The complicating factor in this procedure is that we cannot treat each scale \( R \) independently, since the ionizing sources are correlated among the various volumes. This is the case first because the densities \( \delta \) are correlated, and second because the Poisson fluctuations are correlated, since each sphere contains all the galaxies that lie within all smaller enclosed spheres. The correlation of the densities is dealt with in the standard way reviewed above, where small-scale power is added gradually as smaller spheres are considered. This makes \( \delta(S_2) \) dependent on \( \delta(S_1) \) if \( S_2 > S_1 \), forcing us to start on large scales \( S_R = 0 \) and go to smaller ones. However, the Poisson fluctuations are correlated in the other direction, since a region \( S_2 \) contains a region \( S_1 \) if \( S_2 < S_1 \).

The solution is a two-step Monte Carlo method: first, we generate the random walk \( \delta(S_R) \), going from \( S_R = 0 \) to its maximum value (corresponding to the minimum bubble volume \( V_{\text{bub}} \)) in equal steps. At each \( S_R \) step, we find the mean expected number of galactic halos \( \langle j \rangle(S_R) \) and the mean expected total mass of these halos, \( \langle M_{\text{tot}} \rangle(S_R) \). Note that the mean expected ionized fraction is \( \langle \alpha' \rangle(S_R) = \zeta \langle M_{\text{tot}} \rangle(S_R)/M(S_R) \), where \( M(S_R) \) is the total mass contained within the spherical volume of radius \( R \). In the second step, we generate the actual ionized fractions starting from the smallest scale, \( V_{\text{bub}} \), and working outwards. At \( V_{\text{bub}} \), we generate an instance of a Poisson distribution with mean \( \langle j \rangle(S_1) \), yielding an actual integer number \( j \) of halos, for each of which we find its mass from the appropriate distribution of halo number versus mass, derived from \( f_{\text{bias}} \).

Then, for each larger scale (i.e., smaller \( S_R \) value), the additional number of galaxies from the last step is on average expected to be the difference \( \langle \Delta j \rangle = \langle j \rangle(S_1) - \langle j \rangle(S_2) \), where \( S_1 < S_2 \) are two consecutive steps in \( S_R \). We find the actual difference \( \Delta j \) from a Poisson distribution with a mean of \( \langle \Delta j \rangle \). However, while an actual number of galaxies cannot be negative, sometimes the random walk in \( \delta \) gives a value \( \langle \Delta j \rangle < 0 \). In this case we assume that \( \Delta j = 0 \), since the number of galaxies already enclosed in a smaller volume (corresponding to \( S_2 \)) must also be found in the larger, enclosing volume (\( S_1 \)). We do not discard the negative value of \( \langle \Delta j \rangle \) but add it in the next step to the next value of \( \langle \Delta j \rangle \), continuing until we reach a positive expected mean value on which we can operate a Poisson distribution. In each step, we also keep track of the expected total mass difference, \( \langle \Delta M \rangle = \langle M_{\text{tot}} \rangle(S_1) - \langle M_{\text{tot}} \rangle(S_2) \). We slightly modify \( f_{\text{bias}} \) for each distribution to use the individual halo masses so as to give the correct expected \( \langle \Delta M \rangle \). This procedure ensures that on each scale we obtain the correct average number of galaxies and correct average galaxy mass, both to high accuracy. We note also that in each step in \( S_R \), even if \( \delta \) at the end of the step is below the barrier, there is a chance that the random walk hit the barrier during the step. We estimate this probability using a linear barrier approximation applied separately to each step, and if the walk hit the barrier then we raise \( \delta \) at the end of the step to the barrier. This procedure greatly accelerates the convergence of the results as a function of the total number of steps adopted in \( S_R \).

### 2.3 Summary of models

We summarize here the various models for the bubble size distribution that we consider and compare below.

(i) Model A: The correct distribution as given by our full model. The bubble size distribution is calculated with our Monte Carlo method with discrete sources and Poisson fluctuations, as detailed in section 2.2. We also keep track of how many sources are contained in each generated bubble, which allows us to find the cumulative volume fraction contained in bubbles with at least \( N \) sources, as a function of \( N \).

(ii) Model B: The exact, continuous barrier (without Poisson fluctuations or discreteness). We calculate the nonlinear barrier \( \delta(S_R) \) numerically, as detailed in section 2.2. We then derive the bubble size distribution with a Monte Carlo method that generates random walks and tracks where they first cross the barrier.

(iii) Model C: A continuous linear barrier approximation. We calculate a linear barrier approximation \( \delta(S_R) \approx \nu + \mu S_R \) numerically, as detailed in section 2.2. We then derive the bubble size distribution analytically as in Furlanetto et al. (2004).

(iv) Model D: The previously-used continuous linear barrier approximation. Here we apply the additional approximation mentioned at the end of section 2.2, where we noted that it gives the same bubble size distribution as in the linear barrier approximation of the pure Press-Schechter (rather than Sheth-Tormen) model. In this case we calculate analytically a linear barrier approximation \( \delta(S_R) \approx \nu + \mu S_R \) and then derive the resulting bubble size distribution analytically as in Furlanetto et al. (2004).

(v) Model E: The pure stochastic Poisson model suggested by Furlanetto & Oh (2008). This model, which neglects halo correlations and assumes randomly placed, equal-intensity sources, yields an analytical result for the cumulative volume fraction contained in bubbles with at least \( N \) sources.

Note that the minimum bubble scale is \( V_{\text{bub}} \) for models A and E, \( V_{\text{bub}}/\zeta \) (corresponding to the scale of the minimum halo mass \( M_{\text{min}} \)) for models B–D. Also note that we have tested our barrier-crossing Monte-Carlo code by comparing it to the analytical solution of a continuous linear barrier (Models C and D). We have confirmed precise convergence, to within a relative error below 1% in the total ionization probability, i.e., the total probability of crossing the barrier.

### 3 RESULTS

We illustrate our results for a wide range of possible parameters for either hydrogen reionization or helium (full) reionization. In the latter case, \( \zeta \) is simply interpreted as the overall efficiency factor of producing helium-ionizing photons in halos. For hydrogen, minimum halo masses \( M_{\text{min}} \) that are often considered are the minimum mass for atomic cooling (corresponding to a circular velocity \( V_c \approx 16.5 \text{ km/s} \), where

---

3 We scale the input \( M \) value of the cumulative distribution of halo mass \( M \) (so that the total probability of having \( M > 0 \) remains unity), with the scaling factor (typically close to unity) chosen to yield the correct mean halo mass.
However, in this paper we do not consider the late stages of reionization except as a convenient fiducial mark for normalizing $\zeta$ through $z_{\text{red}}$. Note that our simple model does not include the likely added photoionization feedback (which should affect most of the universe by the time reionization is well advanced), or even larger values if internal supernova feedback strongly decreases the star formation efficiency of low-mass halos at high redshift. For helium reionization, assuming it occurs much later, photoionization feedback affects the source halos from an early stage when the density of the assembling matter is still low, resulting in a larger $V_c \sim 80$ km/s (i.e., of order the Jeans mass). Furthermore, if the observed super-linear local relation between halo and black hole mass holds at high redshift, then quasars are relatively much brighter in more massive halos, increasing the typical halos of helium-ionizing sources to $V_c \sim 200$ – 300 km/s. For a given $V_c$, the efficiency $\zeta$ can be chosen to give complete reionization $x^i = 1$ at various redshifts $z_{\text{red}}$. For a fixed $z_{\text{red}}$, a larger $M_{\text{min}}$ implies that rarer halos caused reionization, resulting in larger Poisson fluctuations.

3.1 Basic results and comparison with previous models

We begin by considering examples corresponding to an early stage (mean ionized fraction $x^i = 1\%$) of hydrogen or helium reionization. For hydrogen, we assume atomic cooling ($V_c = 16.5$ km/s), with efficiency set to complete reionization (i.e., $x^i = 1$) at $z_{\text{red}} = 7$ (implying $\zeta = 19$). For helium we assume $z_{\text{red}} = 3$ and $V_c = 285$ km/s (implying $\zeta = 95$), which gives the bubbles a minimum size at $z \sim 3$ of $R = 10$ Mpc, about the expected size for the quasars that are observed to dominate the ionizing photon production at that redshift (Furlanetto & Oh 2008). Figure 1 shows that the bubble size distribution obtained from our full model is substantially different from the predictions of previous models that are based either on a continuous barrier or on a purely stochastic Poisson distribution.

In general these models, based as they are on spherical statistics, do not precisely conserve photons, and thus do not yield precisely the desired $x^i \equiv \zeta f_{\text{ST}}$ which we have set to 1%. Indeed, the raw total ionization probability yielded by the models is 2.1% (H) and 2.0% (He) for model D, 0.90% (H) and 1.4% (He) for model C, and 1.3% (H) and 1.4% (He) for model A. Thus, in the figure we compare the relative distributions, expressed in terms of the fraction $F_1$ of the ionized volume contained in various bubbles. Note that model E is defined according to the desired $x^i$, and model B (the exact continuous barrier) is mathematically consistent in the sense that it yields the correct total $x^i$ if the probability is integrated down to $V = V_{\text{bub}}/\zeta$ (We have numerically verified this mathematical consistency to a relative error of $\sim 1\%$).

Discreteness strongly fails for the continuous barrier models (both linear and non-linear), in the sense that much of the ionized volume in these models is predicted to occur inside bubbles below the minimum volume $V_{\text{bub}}$, especially in the case of helium reionization. Indeed, $F_1(\geq V)$ is only 55% (H) and 8.2% (He) for model D, 70% (H) and 13% (He) for model C, and 82% (H) and 19% (He) for model B. Thus, the continuous barrier models fail since they assign a substantial probability to the unphysical case of fractional bubbles that are produced by less than one source. Expressed differently, the continuous barrier models underpredict $F_1(\geq V_{\text{bub}})$ since they do not include the Poisson fluctuations that allow large regions to sometimes reach $x^i = 1$ even when their mean expected ionized fraction $\langle x^i \rangle (S_R)$ is below unity.

Figure 1 also illustrates the continuous model with a linearly approximated barrier, a model used very commonly because it yields analytical predictions (Furlanetto et al. 2004). The error of the linear barrier approximation grows at small scales, and becomes a 10% error in the barrier height at $V \sim 0.07V_{\text{bub}}$ (H) or $V \sim 0.02V_{\text{bub}}$ (He). However, the linear barrier approximation becomes relatively accurate on scales larger than the scale $V_{\text{bub}}$ corresponding to a one-source bubble. On that scale, the height of the linear barrier in the examples shown here is only slightly below the height of the real barrier (by 2.6% for H and just 0.05% for He),
though when $x_i \ll 1$ the barrier corresponds to a rare $\sim 3-\sigma$ fluctuation on this scale (and rarer still at larger scales), and thus small differences in barrier height translate to larger differences in $F_i$. The figure also shows that the pure Press-Schechter model (model D) is a rather poor approximation to model C. The Sheth-Tormen hybrid model yields more large bubbles than the Press-Schechter model, which agrees with the expectation based on the Sheth-Tormen mean halo mass function, which yields more rare, massive halos than does the Press-Schechter mass function.

While the continuous barrier model extends unphysically to $V < V_{\text{bub}}$, it does indicate correctly the fact that $F_i(\gtrsim V)$ declines much more rapidly with $V$ for the He case we consider than for H reionization. In fact, we find that if we simply cut off the $V < V_{\text{bub}}$ portion and renormalize the continuous models relative to $V = V_{\text{bub}}$ (which is not a standard way of interpreting these models), then the exact and linear barrier models yield nearly identical results, and they both yield a reasonable rough estimate to the true bubble size distribution in the full model. This is illustrated in Figure 2, which shows the same quantities as in Figure 1 except that all the continuous models have been renormalized and are plotted only for $V \gtrsim V_{\text{bub}}$. For instance, the ratio $V_{1/2} \equiv F_i(V \gtrsim V_{\text{bub}})/F_i(V \gtrsim 2V_{\text{bub}})$ is 1.18 (H) and 2.33 (He) in the full model (model A), 1.14 (H) and 2.54 (He) for the continuous exact barrier (model B), and 1.15 (H) and 2.58 (He) for the continuous linear barrier (model C). This approach to the continuous models provides a reasonable estimate of the full bubble size distribution in the case of He reionization; e.g., $V_{1/100} \equiv F_i(V \gtrsim V_{\text{bub}})/F_i(V \gtrsim 100V_{\text{bub}})$ for H is 6.73 in model A, 6.46 in model B, and 6.21 in model C, so that here the linear model C, calculated analytically (i.e., without using Monte Carlo random walks or Poisson fluctuations), yields an estimate of $V_{1/100}$ that is within 8% of the true answer according to model A. However, for He reionization this approach is much less successful in predicting ratios involving large volumes; e.g., $V_{1/5} \equiv F_i(V \gtrsim V_{\text{bub}})/F_i(V \gtrsim 5V_{\text{bub}})$ for He is 9.9 in model A, 16.0 in model B, and 17.6 in model C, and these differences increase with $V$ (Figure 2).

With our full model (model A), we can also separately predict the distribution by number $F_i(\gtrsim N)$. This drops more rapidly with $N$ than the distribution by volume $F_i(\gtrsim V)$ does with $V$, since large-volume bubbles can be produced either by having many sources of mass $\sim M_{\text{min}}$ or with a smaller number of individually massive halos taken from the high-mass end of the halo mass function. Still, $F_i(\gtrsim N)$ declines with $N$ much more slowly than a pure Poisson model would predict. Indeed, a purely stochastic model as suggested by Furlanetto & Oh (2008) for the early stages of He ionization (or even as late as $x_i \sim 50\%$), where Poisson fluctuations are assumed that are uncorrelated with the underlying density distribution, completely fails to describe the results. The analytical predictions of this model (Furlanetto & Oh 2008) yield, for $x_i = 1\%$ (for either H or He), $F_i(N \gtrsim 2) = 2.0 \times 10^{-2}$ and $F_i(N \gtrsim 3) = 4.4 \times 10^{-4}$ (with the latter already outside the range of Figures 1 and 2). In particular, the ratio from the previous paragraph (but applied to the number of sources), $N_{1/2} \equiv F_i(N \gtrsim 1)/F_i(N \gtrsim 2)$, is 1.32 (H) or 3.8 (He) in the full model, compared to $N_{1/2} = 51$ for model E. Clearly, density correlations play a substantial role in determining the abundance of multi-source bubbles, even early on in reionization and even when the process is driven by large, rare ionizing sources (such as quasars).

To help understand the relation of the full model to the pure Poisson and to the continuous barrier models, we show in Figure 3 the relation between ionization in bubbles and the underlying linear density $\delta$. Density fluctuations are strongly correlated with ionization, so that the density of ionized regions is strongly biased high, and the distribution is very different from the standard Gaussian that would be expected in a pure Poisson model. However, Poisson fluctuations allow regions to fully ionize themselves even if their density is significantly lower than the barrier, which in a continuous model would set the minimum needed $\delta$ for ionization by internal sources. In particular, the median $\delta$ for regions ionized by exactly $N$ sources (where ‘exactly’ means not contained in any larger H II region) represents a fluctuation of 2.4-$\sigma$, 2.57-$\sigma$, and 2.61-$\sigma$ (for H) or 1.9-$\sigma$, 2.4-$\sigma$, and 2.7-$\sigma$ (for He), for $N = 1, 2, 3, \ldots$. The corresponding (median) barriers, on the other hand, are 2.907-$\sigma$, 2.909-$\sigma$, and 2.922-$\sigma$ (for H), or 3.2-$\sigma$, 3.5-$\sigma$, and 3.7-$\sigma$ (for He). Thus, the barriers do give a good rough indication in each case of whether the $\delta$ distributions for various $N$ are spaced out or squeezed together. This in turn determines whether one-source bubbles are dominant and $N > 1$ is rare, or if multi-source bubbles are at least as common as $N = 1$.

The continuous model indicates that the main parameters controlling the relative dominance of single-source bubbles are the effective efficiency $\zeta$ and the effective slope of the power spectrum on the scale $R_{\text{bub}}$ of a one-source bubble. The efficiency sets the ratio between the scale $R_{\text{bub}}$ of a one-source bubble and the scale $R_{\text{min}}$ from which a galactic...
the barrier, the density has the variance $S_n$ achieving self-ionization, when we consider different scales. On the variance ($S_n$ formula in equation (1), requires a value of fraction, which according to the extended Press-Schechter ratio, which indicates how much harder (in terms of number values of $S_n$ and halo scales differ by a large factor (which requires large case when $R_{\text{min}}$ leading to the dominance of one-source bubbles. This is the fractional decline in ($S_n R_{\text{min}}$ $S_n$), and also that the variance depend significantly over the relevant range of scales, then this ratio, which indicates how much harder (in terms of number of $\sigma$ of the fluctuation) it is to ionize larger scales, is approximately $\zeta + (n/3) - 1$. On small scales, $n$ approaches the asymptotic value of $-3$, making all scales behave roughly equally even when $\zeta$ is relatively large. Note, though, that increasing $\zeta$ increases $R_{\text{min}}$ and thus brings larger scales into play, making the effective $n$ less negative and thus boosting the effect of the increased $\zeta$ on making few-source bubbles dominant. This puts a quantitative face on the intuition that rare sources tend to create bubbles with small numbers of sources. To illustrate, in our H example $R_{\text{min}} = 64$ kpc and $R_{\text{bub}} = 169$ kpc, giving $n \sim -2.5$, while in the He example $R_{\text{min}} = 1.7$ Mpc and $R_{\text{bub}} = 7.8$ Mpc, giving $n \sim -2$. Thus, He reionization by quasars has both a high efficiency and corresponds to a relatively large scale, both of which contribute to making small bubbles more dominant, in particular the smallest bubbles created by single sources.

### 3.2 Approximate calculation

The current lack of observations at high redshifts leaves basic parameters of the galaxy population unconstrained at early times. While our model can be used to calculate the bubble distribution in any particular case, the need to run Monte Carlo trials makes it difficult to explore a large parameter space. Thus, an approximate but quick calculation is useful for this purpose. In developing such an approximation, we focus on determining when one-source bubbles dominate the ionizing volume. This can be investigated with the ratio $N/M$, which is close to unity when multi-source bubbles dominate, and $N \geq 1$ when one-source bubbles do. Specifically, this ratio is related to the fraction $F_1(N = 1)$ of the ionized volume that is contained in one-source bubbles through $N_{1/2} = 1/ [1 - F_1(N = 1)]$.

To construct an approximate calculation of this ratio we first adopt the approximation of having equal intensity sources, all corresponding to halos having a mass equal to the mean expected mass $\langle M \rangle$. While this approximation does not work well for obtaining information on the bubble size distribution, we find that it works reasonably for our desired ratio involving the distribution of number of sources per bubble. We first consider in general the self-ionization probability on the scale of a bubble containing $j$ sources (with a variance $S_n$ that we approximate as that corresponding to a volume $jV_{\text{bub}}$), i.e., the probability that a region of this size contains at least $j$ sources (regardless of whether or not it is contained in some larger bubble). A first attempt to calculate this quantity $P_{\text{self}}(j)$ is to calculate the Poisson probability of having at least $j$ sources, averaged over the normal distribution of $\delta$ on the scale $S$: $P_{\text{self}}(j) = \int d\delta \frac{1}{\sqrt{2\pi S}} e^{-\delta^2/(2S)} P_{\text{Pois}}(j\tilde{\delta}(\delta, S); \geq j)$, \( \tag{3} \)

where $P_{\text{Pois}}(\alpha; \geq j)$ denotes the probability of having at least $j$ sources in a Poisson distribution with mean $\alpha$, and $\tilde{\delta}$ (which also depends on $z$ and $S_{\text{min}}$) is an approximate estimate of $\langle \tilde{\delta} \rangle$ where we use the same approximation as in model D in order to obtain a simple formula. For large bubbles, equation \( \ref{3} \) for $P_{\text{self}}(j)$ understimates the self-ionization probability, since for a given mean $\delta$ in the region, internal density fluctuations increase the variance of the number of sources beyond a pure Poisson distribution. For $j = 2$ we can instead calculate a more accurate self-

---

5 The Sheth-Tormen hybrid model alters things slightly, but we use the simpler formula here as a rough guide for a qualitative understanding.
ionization probability by calculating a double integral over the joint normal distribution of $d_1$ and $d_2$, the mean densities inside a one-source volume $V_{\text{bub}}$ and inside the surrounding two-source volume, respectively. Given $d_1$ and $d_2$, the mean expected number of sources in the two regions is $n_1 = 2\bar{x}(z, S_{\text{min}}, d_1, S_1)$ and $n_2 = 2\bar{x}(z, S_{\text{min}}, d_2, S_2)$, respectively, where $S_1$ and $S_2$ are the corresponding variances.

The probability of self-ionization of the two-source volume is then the probability of having at least 2 total sources from the sum of a Poisson distribution of mean $n_1$ plus a Poisson distribution of mean $n_2 - n_1$ (except that the latter quantity is restricted to be non-negative, a key point which allows the larger fluctuations in $n_1$ to contribute).

Calculating $P_{\text{self}}(j)$ exactly for $j > 2$ would require at least a triple integration, but since $j = 1$ and $j = 2$ are most important for estimating $N_{1/2}$, we simply estimate the self-ionization probability for all $j > 2$ with equation (3). Now, $P_{\text{self}}(j)$ for any $j$ is itself only a lower limit for the ionization probability $P(N \geq j)$, since the region may be part of a larger H II bubble even if it cannot fully ionize on its own. Actually, when one-source bubbles dominate and $P(N \geq j)$ drops rapidly with $j$, regions are much more likely to self-ionize than to get ionization help from larger scales, and then $P_{\text{self}}(j)$ becomes an accurate estimate of $P(N \geq j)$. However, in order to achieve reasonable accuracy also when multi-source bubbles are important, we add a correction to each $P_{\text{self}}(j)$ based on the values of $P_{\text{self}}(k)$ for $k > j$. Indeed, instead of just calculating $P_{\text{self}}(k)$, which is the probability of having at least $k$ sources in a region of size corresponding to $k$ sources, we can separately estimate $P(k)$, the probability of having exactly $l$ sources in that region, using a formula just like equation (3) but using the Poisson probability of finding $l$ sources. Then, for any number $l > k$ sources, we calculate the additional ionization probability that was not previously included in $P(N \geq j)$ (for each internal volume $j < k$) using the approximation that the $l$ sources are uniformly distributed within the volume $k$. In this way, we estimate the probabilities $P(N \geq 1)$ and $P(N \geq 2)$ including the contributions of larger volumes with $j > 2$. When one-source bubbles dominate, higher-$j$ volumes have a small effect, but when multi-source bubbles dominate the effect adds up, and we cut off $j$ so that $P(N \geq 1)$ does not rise above the global ionization fraction $\bar{x}^i$. Actually, we find that while the correction from higher-$j$ volumes can change each of $P(N \geq 1)$ and $P(N \geq 2)$ by up to a factor of a few (giving results much closer to the full model A), the relative effect on their ratio is $\sim 15\%$ at most.

Our estimate for $N_{1/2}$ is simply $P(N \geq 1)/P(N \geq 2)$. The approximate calculation becomes exact in the limit $N_{1/2} \rightarrow \infty$, where our estimated probabilities $P_{\text{self}}(j)$ become very small for all $j \geq 2$, while in the opposite limit, when $N_{1/2} \rightarrow 1$ all quantities become nearly independent of $j$ and thus our estimate for the ratio approaches unity, also correctly. In practice, from direct comparison with the Monte Carlo method at $\bar{x}^i$ ranging from $10^{-6}$ to 1, and at ratios $N_{1/2}$ ranging from 1 to 200, we find that our approximation for this ratio is accurate to $\sim 15\%$ (though below we extrapolate it beyond the tested range).

Having developed a quick, relatively accurate calculation method, we can use it to explore which areas of parameter space will be dominated by one-source bubbles and which will form many multi-source bubbles. Figures 4 and 5 show the ratio $N_{1/2}$ in the approximate calculation, for $\bar{x}^i$ ranging from 1 down to $10^{-9}$, over the whole relevant range of source masses, i.e., assuming a minimum $V_c = 4.5, 16.5, 35, 80$, or 285 km/s (solid curves, from bottom to top). We compare to the case of a pure stochastic Poisson distribution (model E; dotted curves). Also shown are the locations corresponding to half of the volume being in one-source bubbles (horizontal long-dashed line), and to 90% in one-source bubbles (horizontal short-dashed line); redshifts are indicated at these locations for each case (if it lies within the range of the plot). Note also that the various curves are not continued below $z = 3$ (for $V_c = 285$ km/s) or $z = 6$ (for the other cases).

Figure 4. Sweep of the parameter space using our approximate calculation, showing the relative dominance of one-source compared to many-source bubbles as indicated by the ratio $N_{1/2} = P(N \geq 1)/P(N \geq 2)$. For $\zeta = 19$ or 95, as indicated, we consider galactic halos with minimum $V_c = 4.5, 16.5, 35, 80$, or 285 km/s (solid curves, from bottom to top). We compare to the case of a pure stochastic Poisson distribution (model E; dotted curves).
The Infancy of Cosmic Reionization

90% (N_{source} bubbles equals 50% (corresponding to ζio in the stochastic model) is an highly unlikely combination) approaches the results ex-
simulations suggest that in each value is indeed the efficiency expected for the very first, pri-
but with a star formation efficiency of only 1%. The latter
ities: perhaps only 10% of photons escape, or 100% escape
but we assume Pop II stars, or we assume Pop III stars but with a star formation efficiency of only 1%. The latter
value is indeed the efficiency expected for the very first, pri-
mordial Pop III stars in molecular cooling halos. Numerical
simulations suggest that in each \( \sim 10^8 \text{M}_\odot \) halo (contain-
ing, therefore, \( \sim 10^4 \text{M}_\odot \) in baryons), it is likely that only a
single 100\text{M}_\odot star forms \( \text{Yoshida et al.} \text{2006} \) before its
feedback disrupts the rest of the halo gas and prevents the
formation of additional stars, at least for some time. Note
also that these values of \( ζ \) neglect recombinations, which can
only lower the effective \( ζ \) further.

Figures 4 and 5 imply the general conclusion that ion-
izing sources produce isolated, single-source bubbles only
quite early in reionization, when \( \bar{x}_i \ll 1 \). This is a result
of the fact that while Poisson fluctuations are large when
we consider just one or two sources, they are strongly mod-
ulated by halo bias due to the underlying density fluctu-
ations. Thus, sources are usually found in high-density re-
ions, which makes it relatively likely to find other sources
nearby. As sources become rarer at high redshift, the increas-
ing correlation strength between halos partially compensates
for the overall low number density of sources, though event-
tually the sheer rarity of sources does come to dominate.
As discussed above, increasing \( V_c \) or \( ζ \) at a given \( \bar{x}_i \) makes
sources rarer and brings larger scales into play, making it
easier to form one-source bubbles relative to multi-source
bubbles. However, only the most extreme case we consider of
rare, extremely bright sources (\( V_c = 285 \text{ km/s} \) and \( ζ = 5800 \), an
highly unlikely combination) approaches the results ex-
pected for a pure stochastic Poisson distribution; the ra-
tio in the stochastic model is \( N_{1/2} = 1/[1 - \exp(-2\bar{x}_i)] \)
\( \text{Furlanetto & Oif} \text{2008} \).

The Figures also indicate the redshifts when the fraction
\( F_1(N = 1) \) of the ionized volume that is contained in one-
source bubbles equals 50% (corresponding to \( N_{1/2} = 2 \)) or
90% (\( N_{1/2} = 10 \)). In particular, for \( V_c = 16.5 \text{ km/s} \) normal-
ized to produce II reionization at \( z_{rei} = 7 \) (i.e., \( ζ = 19 \)), one-
source bubbles dominate (i.e., \( F_1(N = 1) > 90\% \)) only above
\( z = 57 \) (outside the plot range), while multi-source bubbles
become equally important (i.e., \( F_1(N = 1) = 50\% \)) at red-
shift 30. Primordial Pop III stars with \( ζ = 580 \) and \( V_c = 4.5 \)
km/s also tend to form multi-source bubbles at rather high
redshifts, with one-source bubbles remaining dominant only
down to \( z = 50 \), and with multi-source bubbles becoming
equally important at \( z = 37 \). On the other hand, for He
reionization at \( z_{rei} = 3 \) with \( V_c = 285 \text{ km/s} \) (i.e., \( ζ = 95 \)),
these milestones are reached at \( z = 7.7 \) and \( z = 5.2 \), respec-
tively. Additional cases where these milestones occur outside
the plot range of the Figures include \( ζ = 95 \) and \( V_c = 4.5 \)
km/s, which reaches \( N_{1/2} = 10 \) at \( z = 62 \); \( ζ = 19 \) and
\( V_c = 35 \text{ km/s} \), which reaches \( N_{1/2} = 10 \) at \( z = 37 \); and the
faintest example we consider for individual sources, \( ζ = 19 \)
and \( V_c = 4.5 \text{ km/s} \), which reaches \( N_{1/2} = 2 \) at \( z = 53 \) and
does not reach \( N_{1/2} = 10 \) even at the most likely redshift
(\( z = 65 \)) of the very first star (\( \text{Naoz et al.} \text{2006} \)).

If we consider a range of values of \( ζ \) for halos of a
given \( V_c \), the global ionized fraction \( \bar{x}_i \) corresponding to
a particular milestone (as defined by a particular value of
\( F_1(N = 1) \)) increases with \( ζ \), since increasing \( ζ \) at a fixed
\( \bar{x}_i \) makes sources rarer, while increasing \( x_i \) (with a fixed \( ζ \))
compensates for this by increasing the source number den-
sity. For each milestone, however, the redshift, which observa-
tionally is the most directly relevant quantity, behaves in
a more complicated way, since it is directly related to the
number density of sources, and thus depends on the ratio
\( \bar{x}_i/ζ \). We find that sources with a given \( V_c \) can only achieve
a dominance of one-source bubbles at high redshift, almost
regardless of the efficiency \( ζ \) (and thus, regardless of the
reionization redshift).

Figure 5 shows the minimum \( z \) required to achieve vari-
ous values of \( F_1(N = 1) \) (as a function of \( V_c \)), assuming
only that the value of \( ζ \) lies within some wide range. The
figure shows that while high values of \( ζ \) do have a larger
effect on low-\( V_c \) halos, the minimum redshift is overall rel-
atively insensitive to the particular range assumed. In par-
ticular, assuming \( 10 < ζ < 1000 \), for He reionization by
quasars (assuming \( V_c \leq 300 \text{ km/s} \)), the volume fraction in
one-source bubbles \( F_1(N = 1) \) can be greater than 50% only
at \( z > 4.9 \), 90% at \( z > 7.3 \), and 99% at \( z > 9.1 \). For II rei-
onization by stars (assuming \( V_c \leq 35 \text{ km/s} \)), these milestones
require \( z > 18 \), \( z > 23 \), and \( z > 28 \), respectively. The gener-
ation of atomic-cooling halos (\( V_c = 16.5 \text{ km/s} \)) can achieve
\( F_1(N = 1) > 50\% \) only at \( z > 24 \), 90% at \( z > 31 \), and
99% at \( z > 38 \). Finally (again assuming \( 10 < ζ < 1000 \)),
the earliest generation of molecular-hydrogen-cooling halos
(\( V_c = 4.5 \text{ km/s} \)) can achieve these milestones only at \( z > 36 \),
\( z > 48 \), and \( z > 61 \), respectively.

3.3 Later stages

As reionization advances, eventually the typical bubble size
encompasses a large number of ionizing sources, reducing
the importance of discreteness and of Poisson fluctuations.
Figures 4 and 6 show the cumulative bubble size distribution
as in Figure 2 but for later stages of reionization. At these
times, the continuous barrier models still have a significant
probability at \( V < V_{bub} \), especially for He reionization by
quasars; however, if only the \( V > V_{bub} \) portion is considered

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Same as Figure 4 but for \( ζ = 580 \) or 5800, as indicated.}
\end{figure}
as in these figures (see also the discussion in section 3.1), then the linear barrier predictions become essentially identical to those of the exact barrier, and the predicted bubble size distributions of these continuous models are reasonably accurate. Specifically, when $\bar{x}_i = 10\%$, for the example of H reionization, $V_{1/2}$ and $V_{1/100}$ equal 1.08 and 2.28, respectively, in model A (the full model), 1.06 and 2.18 in model B (continuous barrier), and 1.06 and 2.16 in model C (linear barrier). For He reionization, $V_{1/2}$ and $V_{1/5}$ are 1.42 and 2.56 in model A, 1.44 and 2.98 in model B, and 1.44 and 2.99 in model C. When the universe is 10% ionized, bubbles with a small number of sources still play a major role, e.g., one and two-source bubbles together account for $19\%$ (H) or $60\%$ (He) of the total ionized volume, and the small-$N$ regime is still quite important.

At the midpoint of global reionization ($\bar{x}_i = 50\%$), the continuous barrier models approach the full model even more (note that the figures at different $\bar{x}_i$ have different $y$-axis ranges). For the example of H reionization, $V_{1/2}$ and $V_{1/100}$ equal 1.03 and 1.22, respectively, in model A (the full model), 1.01 and 1.19 in model B (continuous barrier), and 1.01 and 1.19 in model C (linear barrier). For He reionization, $V_{1/2}$ and $V_{1/5}$ are 1.08 and 1.23 in model A, 1.08 and 1.24 in model B, and 1.08 and 1.24 in model C. For H reionization only 5.2% of the ionized volume lies in one and two-source bubbles, but for He this fraction is still 17%. As we found in section 3.1, at $\bar{x}_i = 1\%$, at $\bar{x}_i = 10\%$ and $50\%$ we again see that the pure Press-Schechter model (Model D) is a rather poor approximation to model C, and that the pure Poisson model (Model E) predicts a distribution by number $F_i(\geq N)$ that falls off much faster with $N$ than do the true distributions (for H or He reionization) according to model A.

4 CONCLUSIONS

We have developed a model of reionization that adds discrete ionizing sources and Poisson fluctuations to the continuous model of Furlanetto et al. (2004). We have shown how to obtain the distribution of ionized bubbles, versus both bubble size and number of ionizing sources, with a two-step Monte Carlo method that accounts for both density and Poisson correlations among regions of various sizes surrounding a given random point in the universe. The bubble size distribution we obtained differs substantially from previous models, but if the continuous barrier model is cut off below $V_{\text{bub}}$ (the minimum bubble volume corresponding to a single halo of mass $M_{\text{min}}$) then it yields a reasonable rough estimate to the true bubble size distribution. More specifically, this estimate is generally accurate for H reionization even as early as a mean ionized fraction $\bar{x}_i = 1\%$, while for He reionization it works best for small volumes and at later times, and at $\bar{x}_i = 1\%$ is accurate only up to $V \sim 3V_{\text{bub}}$. Note that with the cutoff at $V_{\text{bub}}$, the linear barrier approximation (which
The Infancy of Cosmic Reionization

redshift. In particular, for He reionization by quasars, one-source bubbles can dominate (i.e., contain 90% of the ionized volume) only at $z > 7.3$, and fill half the ionized volume at $z > 4.9$, while H reionization by stars can achieve these milestones only at $z > 23$ and $z > 18$, respectively (assuming $10 < \zeta < 1000$). The generation of atomic-cooling halos can place 90% of the ionized volume in isolated bubbles only at $z > 31$ and 50% at $z > 24$, while the earliest generation of molecular-hydrogen-cooling halos can achieve the same only at $z > 48$ and $z > 36$, respectively.

We note that reality likely includes even more fluctuations than included in our Poisson model, since we have still assumed that the number of ionizing photons emitted from a galactic halo is proportional to its mass. In reality, variations in the ionizing efficiency (through spatial or temporal fluctuations in the star formation efficiency and in the escape fraction of ionizing photons), and in the merger histories of halos of a given mass (even within a given environment, as measured by the average density of a surrounding region) will increase the role of (now generalized) Poisson fluctuations compared to that of galaxy bias due to the underlying large-scale density fluctuations. Simple forms of such variability can be included in a model of the type that we presented, since the ionizing photon outputs from sources are added as individual units (which could be generated from additional distributions for a given halo mass). In general, the model we developed can be used to investigate helium reionization and observational prospects for 21-cm observations during the infancy of hydrogen reionization.

ACKNOWLEDGMENTS

The author thanks Andrei Mesinger for very useful comments, and is grateful for support from the ICRR in Tokyo, Japan, the Moore Distinguished Scholar program at Caltech, and the John Simon Guggenheim Memorial Foundation, as well as Israel Science Foundation grant 629/05.

REFERENCES

Barkana R., 2007, MNRAS, 376, 1784
Barkana R., Loeb A., 2004, ApJ, 609, 474
Barkana R., Loeb A., 2005a, ApJ, 624, L65
Barkana R., Loeb A., 2005b, ApJ, 626, 1
Barkana R., Loeb A., 2007, Rep. Prog. Phys., 70, 627
Barkana R., Loeb A., 2008, MNRAS, 384, 1069
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Cen R., 2006, ApJ, 648, 47
Chuuzzhoy L., Alvarez M. A., Shapiro P. R., 2006, ApJ 648 L1
Furlanetto S. R., McQuinn M., Hernquist L., 2006a, MNRAS, 365, 115
Furlanetto S., Oh S. P., 2008, ApJ, 681, 1
Furlanetto S. R., Oh S. P., Briggs, F., 2006b, Phys. Rep., 433, 181
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Hogan C. J., Rees M. J., 1979, MNRAS, 188, 791
Komatsu E., et al., 2008, ApJS, submitted (arXiv:0803.0547)
Madau, P., Meiksin, A., Rees, M. J., 1997, ApJ, 475, 429
Mellema, G., Iliev, I. T., Pen, U.-L., Shapiro, P. R. 2006, MNRAS, 372, 679
Mesinger, A., & Furlanetto, S. 2007, ApJ, 669, 663
Naoz, S., Barkana, R., 2008, MNRAS, 385, 63
Naoz S., Noter S., Barkana, R., 2006, MNRAS, 373, L98
Press W. H., Schechter P., 1974, ApJ, 187, 425
Pritchard J. R., Furlanetto S. R., 2006, MNRAS 367, 1057
Pritchard J. R., Furlanetto S. R., 2007 MNRAS 376, 1680
Scott D., Rees M. J., 1990, MNRAS, 247, 510
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Trac H., Cen R., 2007, ApJ, 671, 1
Yoshida N., Omukai K., Hernquist L., Abel T., 2006, ApJ, 652, 6
Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12
Zhang J., Hui L., 2006, ApJ, 641, 641