ABSTRACT

Title of Dissertation: SIMULATING BURSTY AND CONTINUOUS REIONIZATION USING GPU COMPUTING

Blake Hartley
Doctor of Philosophy, 2023

Dissertation Directed by: Massimo Ricotti
Department of Astronomy

Reionization is the process by which the neutral intergallactic medium of the early universe was ionized by the first galaxies, and took place somewhere between roughly redshift 30 and redshift 6, or from 100 Myr into the universe to 1 Gyr. The details of this transition are still not well understood, but observational constraints suggest that reionization happened faster than naive estimates would suggest. In this thesis, we investigate the theory that galaxies which form their stars in short bursts could complete reionization faster than galaxies which emit their photons continuously over their lifespans.

We began investigating this theory with a semi-analytic model of the early universe. We used analytic methods to model the expansion of H II (ionized hydrogen) regions around isolated galaxies, as well as the behavior of the remnant H II regions after star formation ceases. We then compiled assortments of galaxies matching dark matter simulation profiles and associated each with an H II region that could either grow continuously or grow quickly before entering a

dormant period of recombination. These tests indicated that the remnants of bursty star formation had lower overall recombination rates than those of continuously expanding H II regions, and that these remnants could allow for ionizing radiation from more distant sources to influence ionization earlier.

We decided that the next step towards demonstrating the differences between continuous and bursty star formation would require the use of a more accurate model of the early universe. We chose a photon conserving ray tracing algorithm which follows the path of millions of rays from each galaxy and calculates the ionization rate at every point in a uniform 3D grid. The massive amount of computation required for such an algorithm led us to choose MPI as the framework for building our simulation. MPI allowed us to break the grid into 8 sub-volumes, each of which could be assigned to a node on a supercomputer. We then used CUDA to track the millions of rays, with each of the thousands of CUDA cores handling a single ray. Creating my own simulation library would afford us complete control over the distribution and time dependence of ionizing radiation emission, which is critical to isolating the effect of bursty star formation on reionization.

Once we had completed, we conducted a suite of simulations across a selection of model parameters using this library. Every set of model parameters we selected corresponds to two models, one continuous and one bursty. This selection allowed us to isolate the effect of bursty star formation on the results of the simulations. We found that the effects we hoped to see were present in our simulations, and obtained simple estimates of the size of these effects.

SIMULATING BURSTY AND CONTINUOUS REIONIZATION USING GPU COMPUTING

by

Blake Teixeira Hartley

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Massimo Ricotti, Chair/Advisor
Professor Alberto Daniel Bolatto
Professor Richard Mushotzky
Professor Zackaria Chacko
Professor ¡

Preface

My goal for this thesis is to build a story of my research over the course of my time at the University of Maryland, College Park. Despite this, some of the the individual chapters are either published (Chapter 2 and Chapter 3) or intended for publication (Chapter 4) in peer-reviewed journals, and as such should stand alone. My hope is that the Introduction and Conclusions (Chapters 1 and 5) add enough of a framework that the thesis as a whole tells a coherent story.

Dedication

To my parents, for always supporting me in my never ending quest to answer every question my curiosity leads me to.

Acknowledgments

This thesis is culmination of more than a decade of research, brainstorming, crumpled scratch paper, HPC terminal logins, inexplicable errors, emails, and paperwork. It would not have been possible without countless people who deserve more than my thanks, but here I will attempt to thank some of them.

First and foremost, I want to thank Professor Massimo Ricotti. This thesis was born from a simple idea he had years ago, and without his constant guidance, willingness to answer questions, and lateral thinking, I would never have gotten off the ground with this work. Many times throughout my research I found myself at an impasse, and he was categorically ready to provide a perspective that would break up the clot. He supported me above and beyond the call of duty through many setbacks, and I have been honored to have worked with him all this time.

None of my simulations would have been possible without input data generated by Emil Polisensky, who I am deeply grateful to. I also want to thank Professor Cole Miller, who may not realize just how important his insights into subtle physics questions were to the completion of this thesis. I'm grateful to Professor Derek Richardson also aided me several times while dealing with HPC issues. I want to thank the UMD HPC support staff, and in particular Kevin Hildebrand and Thomas Payerle, who helped me get the supercomputer time needed to run my simulations and helped me any time I had issues with my code.

I want to thank the UMD Astronomy Department Staff – including but not limited to Andy

Harris, Eric McKenzie, Mary Ann Phillips, Barbara Hansborough, Olivia Dent, Lauren Miles, Natalie Rowe, Susan Lehr, Dorinda Kimbrell, Mark Wolfire, Kevin Rauch, and Peter Teuben – I was surely not one of the easiest students to deal with and tested their patience many times, and I'm grateful to them for persevering.

I'm also grateful to my friends and classmates. I'm grateful to my classmates Amy Steele, Zeeve Rogoszinski, Myra Stone, Alex Fitts, Kyle Sheppard, Qian Wang, Chongchong He, Drew Hogg, Tiara Hung, Corbin Taylor, and Andrew Gallagher, with whom I share many fun memories and without whom I would surely be less sane. Beyond this group, I want to thank the entire UMD Astronomy Department. They created an environment which was fun, engaging, and conducive to science and learning in a way I wouldn't have believed possible if I didn't experience it myself. I would also like to thank ADHD for incredible amount I was able do, experience, and accomplish besides finish my thesis. Beyond the department, I'm also grateful to Lennon Grimm, Cody Reynolds, Marshall Asfoor, Rick Taylor, Cheyenne Coffy, Melissa Barber, Keiichiro Amemiya, Luke Roddis, Kevin Robinson, Ricky James, Patrick Ferguson, Josh D, Joshua Gang, Patrick C, Judith Penaloza, Carlos, and Bryson. Among the names mentioned here, I want to give a special thanks to Amy Steele and Joshua Gang, both of whom helped me pick over and polish the non-peer-reviewed bits of the thesis.

Though it may be unusual, I want to thank Akasaka Aka and Eiichiro Oda, two manga artists who wrote countless characters and stories that entertained me throughout the later years of my work. Of particular note are Shinomiya Kaguya and Yamato, characters whose struggle to find a place in their worlds I found extremely motivational.

Finally, I want to thank my mother Roberta Hartley, father Jeff Hartley, and brother Maxwell Hartley for their unconditional (maybe conditional in Maxwell's case) love and support, and for

always at least smiling and nodding while I waved my hands talking about the issues of hierarchical ray splitting in a Cartesian 3-torus computational grid. They have given me more than I ever could have asked for along my journey through life. I dedicate this thesis to them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
EoR	Epoch of Reionization
GC	Globular Cluster
H	Hydrogen
FLRW	Friedmann–Lemaître–Robertson–Walker
He I	Neutral helium
He II	Singly ionized helium
He III	Double ionized helium
H I	Neutral hydrogen
H II	Ionized hydrogen
HPC	High Performance Computing
IGM	Inter Galactic Medium
kpc	Kiloparsec
MNRAS	Monthly Notices of the Royal Astronomical Society
h^{-1} $M_\odot$	Stellar mass
Mpc	Megaparsec
MPI	Message Passing Interface
pc	parsec
UV	Ultra Violet

Chapter 1: Introduction

The development of astronomical tools has allowed us to peer further and further back into the history of the cosmos. The history of the Universe before the formation of the first stars and galaxies is well understood through the study of cosmological models and observations of the anisotropies of the Cosmic Microwave Background (*e.g.*, Planck Collaboration et al., 2015). However, between these two periods lies a span of cosmic history that is too far for instruments to observe and too complicated for modern models to accurately simulate. Reionization is a process which took place during this period, and it is this portion of cosmic history that we are interested in investigating.

We begin this chapter with an introduction to the relevant aspects of cosmic history that sets the stage for our work, including recombination, dark matter, the nature of the first stars and galaxies, the thermal/chemical evolution of the Universe, and reionization. We then introduce the computational methods and languages which underlie modern research in this area.

With this thesis, we hope to emphasize that a simple adjustment to the way in which the stars form during reionization can make a large impact on how the Universe developed during this period. Newer and bigger simulations of reionization are constantly being performed, and we believe that this adjustment may be a piece in the puzzle to explaining how the Universe got to be the way it is today.

1.1 Early Cosmic History

After the Big Bang, the Universe was composed of very dense and hot plasma, radiation, and dark matter, all of which were distributed uniformly. The Universe as a whole was expanding and cooling rapidly, with the physics of this expansion governed by the Friedmann–Lemaître–Robertson–Walker (FLRW) equations. These equations define a metric of space-time which is homogeneous and isotropic, but changes in scale with time. We further refine this model into the Λ-CDM (Λ Cold Dark Matter) model, which assumes that space is flat and permeated by a combination of matter, radiation, and dark energy (Λ). We refer to the present day balance of these species with Ω_{m} (matter), Ω_{R} (radiation), Ω_{Λ} (dark energy). The C in Λ-CDM refers to the fact that the dark matter is "cold", meaning that it doesn't have much kenetic energy, which would tend to prevent it from forming clumps. In this model, the rate of expansion is referred to as H, or Hubble's parameter. This parameter is fundamentally in units of $[s^{-1}]$, but is often given in units of km/s/Mpc, so that points which are at rest x Mpc away from each other will move away from each other at xH km/s.

Cosmologists measure time in this model using the time since the Big Bang, which is called the Hubble time. There are other relevant measurements, such as the redshift (z), which is a measure of how much light rays emitted at a given time have stretched by universal expansion by the time they reach us in the present. There is also the scale factor (a), which is a measure of how much smaller the Universe was at a given time compared to today. These are related by:

$$a = \frac{1}{1+z}, \qquad z = \frac{1}{a} - 1.$$

The rate at which the Universe expands can also be calculated from the FLRW equations, and in general depends on the values of Ω_m, Ω_R, and Ω_Λ, or the form of matter/radiation/energy that dominates the Universe's composition. The values of these constants are measured by the Planck experiment (Planck Collaboration et al., 2015):

$$\Omega_{\rm m} = 0.27$$

$$\Omega_{\rm R}\ 0$$

$$\Omega_\Lambda = 0.73$$

$$\Omega_{\rm b} = 0.044$$

$$H_0 = 67.3\ {\rm km/s/Mpc}.$$

Intuitively, these parameters tell us that today, as measured by the Planck experiment, the Universe is composed of 73% of dark energy and 27% of matter, and that matter is mostly dark matter, with only 4.4% of the overall composition being visible matter or baryonic matter. These parameters may then be used to determine the expansion rate of the Universe using one of the Friedmann equations:

$$\frac{H^2}{H_0^2} = \Omega_{\rm R} a^{-4} + \Omega_{\rm m} a^{-3} + \Omega_\Lambda, \tag{1.1}$$

where the Hubble parameter $H = d\ln a/dt$ describes the expansion rate of the Universe and H_0 is the value of H measured at the present time. While the redshift and scale factor are simply related, the expansion rate of the Universe needs to be calculated over its entire history to calculate the relationship between time and redshift. If the Universe's composition is dominated by a single

species, or when one term in Equation 1.1 dominates over the others, the right-hand side of the Friedmann equation becomes a simple power law of the scale parameter, and the relationship between time and redshift can be approximated easily. For example, in a matter dominated Universe $a \propto t^{\frac{2}{3}}$, while in a radiation dominated Universe $a \propto t^{\frac{1}{2}}$. The value $1/H$ is known as the Hubble time and has units of seconds, and can be thought of as the time it would take for the Universe to expand to its current size at its current expansion rate. We also take this moment to introduce the idea of a "comoving" distance unit, which is the given unit multiplied by the (dimensionless) scale factor a. Thus, "cpc" is a comoving parsec, which at redshift $z = 9$, $a = 1/(1 + z) = 0.1$, would be $0.1 \sim$ pc in proper or physical units. These units are useful because objects which grow with the expansion of the Universe have a constant size in comoving units.

1.1.1 Recombination

As the Universe evolved, the balance of matter, dark matter, and radiation shifted. Inspecting Equation 1.1 reveals that the radiation term will dominate for the smallest a values, which lasted for the first 50,000 years. In terms of redshift, this would be $z \sim 3400$ or $a \sim 0.0003$, meaning the Universe was 3400 times smaller than it is today. After 50,000 years, matter began dominating the composition, but the radiation which permeated the Universe was still energetic enough to maintain the ionization of the matter. After roughly 370,000 years, or at roughly $z \sim 1100$, the radiation had cooled enough that the hydrogen and helium that dominate the Universe's composition could recombine en masse. All the thermal radiation that permeated the Universe was then not energetic enough to interact with matter. This radiation continued to

permeate the Universe until the present day, with its wavelength continually being stretched by universal expansion (i.e. redshifting). We now see these photons as the cosmic microwave background (CMB), which has a nearly perfect thermal spectrum (black body) with a temperature of roughly 2.7 Kelvin. The CMB appears to come from a spherical surface, which is the matter that the radiation last interacted with. This surface is known as the last scattering surface, and is effectively the horizon of the observable Universe, as any photons from outside this sphere (earlier in time) would have interacted with the the ionized matter that composed the Universe before recombination.

1.1.2 Dark Matter

The matter that we can see only composes about 15% of the matter within the Universe. The rest of it is known as dark matter, which only interacts with visible matter through gravitation. As the Universe cooled and expanded, over-dense regions of matter began to collapse under their own gravity, and dark matter dominated the development of these regions. The fact that dark matter only interacts through gravity makes it significantly easier to model, a fact which has allowed astronomers to study these processes very accurately. The over-dense regions clumped and fragmented, developing into what are known as dark matter halos. This process proceeded hierarchically, with smaller mass halos forming first and and massive halos forming later. This means that objects which form in small mass halos (such as the first stars (Pop III) or dwarf galaxies) likely formed before objects which form in massive halos (such as galaxies larger than the Milky Way or galaxy clusters). How small galaxies could be when they formed, or how numerous Pop III stars were, are is not well known, an uncertainty which is a key factor of our

motivation for studying this topic.

1.2 Thermodynamics of the expanding Universe

As the Universe expands under the FLRW model, the volume of a given portion increases as $(1+z)^{-3}$ (redshift decreases with time), while little thermal energy flows. This means that the Universe as a whole is a closed system, which means that this expansion is adiabatic as long as there are no sources and sinks of heat. The thermal properties of the components of the Universe thus obey the equation of adiabatic expansion:

$$TV^{\gamma-1} = \text{constant},$$

where γ is the adiabatic index of the material. For monatomic gas, which the Universe was mostly composed of at this time, $\gamma = 5/3$, so that $TV^{2/3} =$ constant, or $T \propto V^{-2/3} \propto (1+z)^2$. For radiation, $\gamma = 4/3$, so that $TV^{1/3} =$ constant, or $T \propto V^{-1/3} \propto (1+z)^1$.

While the Universe was expanding, the self gravitation of gas clouds and dark matter halos causes them to collapse. As this happens, the time scale of energy flow becomes lower than the Hubble time, and the adiabatic assumption no longer applies. The cloud thus seeks an equilibrium which can be determined using the Virial Theorem, which relates the average kinetic and potential energy of the cloud by $< K >= -\frac{1}{2} < V >$. The kinetic energy of the gas is in its thermal energy, so this balance tells us that larger clouds will reach equilibrium at higher temperatures. The temperature at which a given cloud will be able to collapse is referred to as its Virial temperature.

1.3 First Sources of Light

The collapse of dark matter halos accelerated the collapse of matter that was bound within these halos. As these clouds grew in mass, they would begin to reach their Jean's Mass, at which point they could collapse under their own self gravity and form light emitting objects. The properties of these objects, including mass, luminosity, lifespan and spectrum, show a large variety. Some of the observed and proposed objects fall into the following categories:

- Population III Stars: The first generation of stars, composed of unenriched gas from the Big Bang. None of these stars remain today, which makes the understanding of their properties difficult. However, it is generally believed that these stars were extremely massive, burning very hot and living short lives.

- Hypernovae: A hypernova is a proposed phenomenon that results at the end of the life of an extremely massive star. This process could emit very high energy photons which would have a long mean free path through the IGM. This could result in a "pre-ionization" of the IGM, which could accelerate the process of reionization overall.

- Massive Galaxies: Large dark matter halos, with $M_{DM} > 10^8\ h^{-1}\ \mathrm{M}_\odot$, the virial temperature of the cloud is above 10^4 K. H I emission is able to cool gas above this temperature, meaning the cloud is able to cool and collapse without first forming molecular hydrogen. These galaxies are known as large-halo protogalaxies.

- Dwarf Galaxies: For dark matter halos with mass $M_{DM} < 10^8\ h^{-1}\ \mathrm{M}_\odot$, the gas in the cloud is cooler than the virial temperature of 10^4 K, meaning that the gas cannot cool further without the presence of molecular hydrogen, or H_2. During the virialization process,

free electrons can bind to neutral hydrogen to form H^-, which can then bind to another hydrogen atom to form H_2 (and a free electron). This H_2 then allows the gas to cool enough to form small galaxies.

- Black Holes: An issue of tension in modern cosmology is the size of black holes today, which are larger than standard accretion models would allow them to become. Quasars, which are generally thought to be powered by a central black hole, also existed very early in the Universe. Both of these facts indicate that black holes may have been present during this time period. Active black holes are capable of emitting very high energy radiation, similar to hypernovae.

- Compact Star/Globular Clusters: A Globular Cluster (GC) is a tightly bound group of thousands to millions of stars. The globular clusters present within and in the vicinity of the Milky Way indicate that GCs formed with a bimodal distribution in time, with one of the modes happening roughly 13 Gyr ago, which makes them candidates for being one of the first sources of ionizing radiation (Ricotti, 2002a).

Another important feature associated with these objects is their escape fraction $f_{\rm esc}$, or the fraction of ionizing radiation which escapes the neutral gas that surrounds the regions where they form and enters the IGM. Most of these objects are expected to have relatively low escape fractions of $f_{\rm esc} \sim 0.05 - 0.2$, though it is possible that some (such as GCs) could have had $f_{\rm esc} \sim 1$. Estimates of $f_{\rm esc}$ are difficult to make in practice and form one of the biggest sources of uncertainty in studying how these objects effect the IGM.

1.3.1 Reionization

Now that we have laid a foundation in cosmology, we are ready to discuss the main topic of this thesis, *reionization*. The first sources of light are thought to have begun appearing at roughly a redshift of $z \sim 30$, or 100 Myr into the life of the Universe. At this point, the IGM was still fully neutral. Over the next several hundred million years, some combination of the light emitting objects we discussed fully ionized the IGM, as direct observations have shown that the IGM is fully ionized up to a redshift of at least $z \sim 5$, or a Hubble time of just over 1 Gyr. This period, also known as the Epoch of Reionization (EoR) is a time during which the complexity of the Universe's structure developed massively, yet it is still generally beyond the observational capabilities of modern observatories, which makes it the most poorly understood part of the Universe's history. Understanding this period is critical, however, as it bridges the gap between the high redshift Universe and the Universe we see today, and tells us about the processes that drove galaxy formation.

Over the past several decades, a coherent picture of the EoR has been gradually developed by astronomers. The bedrock of this picture is the understanding of cosmological parameters gained from modeling anisotropies in the Cosmic Microwave Background (Planck Collaboration et al., 2020), which can provide the basis for a cosmological simulation. The properties of the IGM between the Earth and distant quasars can be studied by way of a variety of experimental methods, including:

1. Lyman-α Forest: The scattering of Lyman-α emission line (a resonant emission line produced by hydrogen recombination) from distant quasars, is known as the Lyman-α forest and gives information about the distribution of Lyman-α absorbers (Becker et al., 2015).

The number density of Lyman-α emitters can also be studied (Mason et al., 2018). The damping wing of the Gunn-Peterson Trough of this emission can also yield information about the presence of neutral hydrogen at high redshifts (Davies et al., 2018; Hsiao et al., 2023).

2. 21 cm tomography: The 21 cm line produced by the spin-flip transition in neutral hydrogen can be used to probe the distribution of neutral hydrogen before the end of reionization by comparing its emission/absorption against the CMB (Furlanetto, 2016).

3. Optical Depth to Thompson scattering: The free electrons, which began appearing at the start of reionization, scatter photons from the CMB. The cumulative optical depth of this effect can be measured and compared to its predicted value from simulations (Planck Collaboration et al., 2015). This effect is more pronounced at high redshift when the IGM has a higher density, so its value is relatively sensitive to the early conditions of reionization.

4. High-z galaxies: Infrared observatories such as HST and JWST are able to observe the rest-frame (without redshift) ultraviolet emissions from high redshift galaxies using photometry. These observations will allow for the development of statistics on the luminosity function of these galaxies (Donnan et al., 2022).

These methods combined have created a generally accepted picture of the EoR spanning ~ 5.5 to $z \sim 10$, with individual ionized regions coalescing in the range of $5 \lesssim z \lesssim 6$. It is also generally accepted that the process was mostly powered by objects lower in the mass range, with stellar mass $M < 10^8 M_\odot$, which had relatively high escape fractions (Gnedin, 2000a).

While this picture is generally accepted, confirming it with direct observations is only possible with an understanding of the IGM between galaxies and the formation of the first structures,

a task which has proven daunting. To date, less than 10,000 galaxy candidates are known to exist in the $6 \lesssim z \lesssim 8$ range and much fewer at higher redshift (Bouwens et al., 2015). Despite this, the catalog of galaxies above $z \sim 6$ is growing due to efforts such as REBEL (Bouwens et al., 2022) and ALPINE (Fè vre et al., 2020). Looking forward, many instruments that are currently or soon to be operational will help build this picture, including the Atacama Large Millimiter Array (ALMA), the James Webb Space Telescope (JWST), the Cerro Chajnantor Atacama Telescopeprime (CCAT-p), and the Spectro-Photometer for the history of the Universe, Epoch of Reionization and Ices Explorer (SPHEREx), the Low-Frequency Array (LOFAR), the Murchison Widefield Array (MWA), the Hydrogen Epoch of Reionization Array (HERA), and the Square Kilometer Array (SKA).

1.3.2 Computational Cosmology

To help guide the understanding of the influx of data from such observatories, astronomers will lean on knowledge that has been gained from computational simulations of the EoR. However, it is extremely difficult to simulate the overall process of reionization, because such simulations require: a) covering large enough volume for statistical representation; b) resolving galactic or even sub-galactic scales, to understand the properties of the objects that form; and c) treating radiative transfer (RT) accurately. Simulations to date tend to either fail in one of these categories, For example, simulations use volumes which are too small to represent the EoR as a whole (Trebitsch et al., 2021) or resolutions which are not fine enough (Ocvirk et al., 2020). Of particular interest to us is the third factor, which is the expensive cost of accurate radiative transfer simulations.

The physics of radiative transfer is is extremely difficult to simulate due to a number of factors, including but not limited to: a) the complexity of Maxwell's Equations; b) the amount of information that needs to be tracked to follow radiation of every frequency from every source at every time, and its interaction with every species of matter; and c) the small time scales necessitated by the high value of the speed of light. Generally, all modern computational approaches to this problem make the simplifying assumption of following a handful of discrete frequencies (known as "bins"), but beyond this there are generally two paradigms for radiative transfer simulations: moment methods and ray-tracing. To understand these, we first note the continuity equation for specific intensity:

$$\frac{1}{c}\frac{\partial I_\nu}{\partial t} + \mathbf{n} \cdot \nabla I_\nu = j_\nu - \kappa_\nu \rho I_\nu, \tag{1.2}$$

where I_ν is the intensity of frequency bin ν, j_ν is the emission term, and κ_ν is the absorption coefficient (Mihalas & Mihalas, 1984). Intuitively, this equations says that the change in intensity (first term) decreases as the radiation spreads (second term), increases when there is new emission (third term), and decreases when there is absorption (fourth term). Now, we can discuss the most important prevalent methods of numerically approximating this equation in cosmic simulations:

Moment Methods: These methods track the radiation field at each location in the simulation grid without tracking the detailed information about how the radiation moves away from each source. This is accomplished by focusing only on the zeroth and first moments of Equation 1.3.2 (Kannan et al., 2019). This simplification effectively reduces the radiation field to a fluid, which is a massive simplification. These methods have the benefit of having a computational workload that

that doesn't scale with the number of sources. However, these methods are generally diffusive, meaning that the radiation tends to "flow" away from sources in ways that can be unphysical. They also tend to make use of a reduced speed of light simplification (from c/1000 or c/100 up to $c/5$ for modern methods, see Kannan et al. 2021), which allows for larger time steps at the cost of some accuracy.

Ray Tracing: These methods model the radiation emitted by a source as a collection of light rays that begin at the source. Equation 1.3.2 for each ray is then tracked individually, with its effect on every grid cell it passes through individually evaluated and compiled (Abel et al., 1999). The second term of Equation 1.3.2 is ignored not included in these calculations, as the spreading of radiation is inherently handled by the geometric spread of propagating rays. These methods are generally the most accurate, as in the limit of an infinite number of rays, they model the process perfectly. However, these methods require $O(N^2)$ rays to cover an expanding sphere of radiation, with N growing as the sphere of influence grows.

1.3.3 High Performance Computing

Most modern EoR simulations use moment methods, as the scale and resolution required for these simulations make ray-tracing methods prohibitively expensive, and the use of ray-tracing is limited to smaller scale simulations. With the size of processors shrinking and their speed increasing rapidly, the amount of computational resources that can be devoted to a given problem is growing exponentially, allowing astronomers to push the boundaries of resolution and scale for cosmic simulations. Aggregated computational resources today are known as supercomputers, and these supercomputers are composed of a combination of Central Processing

Units (CPUs) and Graphics Processing Units (GPUs). The computational frameworks associated with these architectures are MPI (CPUs) and CUDA (GPUs):

MPI: MPI, or Message Passing Interface, is a framework for creating a library which can be run on parallel computing architectures. MPI based methods function by allowing the user to define "messages" which can be flexibly sent and and received by any node in the parallel computational network. It is portable, scalable, and high performance, and is the dominant framework for high performance computing today (Message Passing Interface Forum, 2021).

CUDA: Cuda, or Compute Unified Device Architecture, is a parallel computing platform and API (application programming interface) which allows software to leverage the massive and ever growing amount of computational power in GPUs, or Graphics Processing Units (NVIDIA et al., 2020). Modern GPUs have 2000+ cores, which means that a well designed CUDA program is able to divide its calculations into 2000+ fragments, which can be computed in parallel.

While working on our project, we realized that it would be possible to harness MPI and CUDA simultaneously. This means we would be able to divide the computational work between supercomputer nodes using MPI, then perform the math on thousands of rays simultaneously with CUDA. The principle of superposition means that the math being performed by each supercomputer node and GPU core is often independent, which minimizes the amount information sharing the MPI cores and GPU threads would have to share. This information sharing is often the bottleneck when it comes to the speed of computational method.

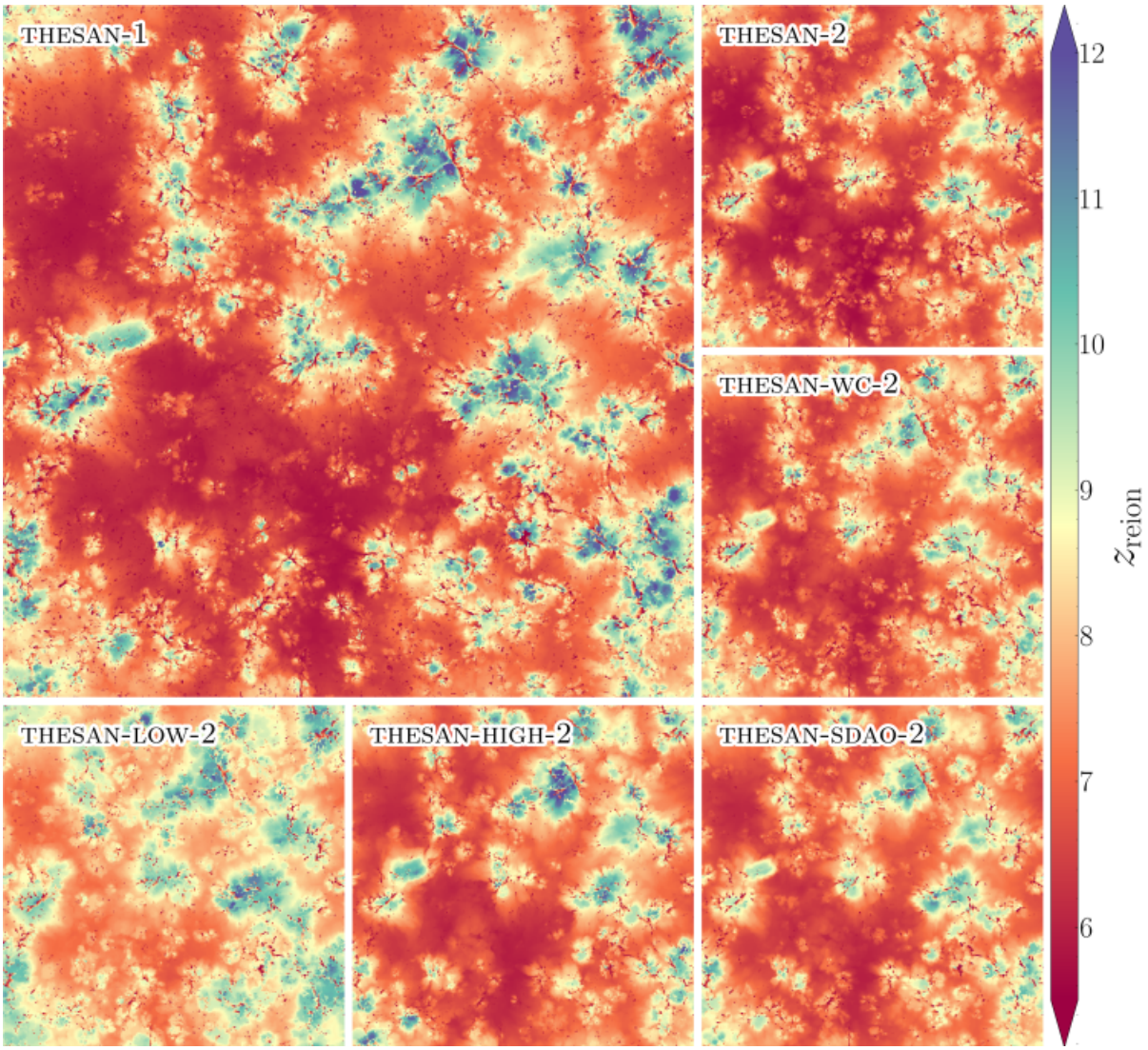

Figure 1.1: Redshift of reionization for slices through the seven THESAN simulations The large scale of the simulation and high resolution produce amazing and beautiful results.

1.3.4 Recent Reionization Simulations

With an idea of the computational methods in mind, we now present a snapshot of modern computational reionization research. One of the more recent large scale EoR simulations is the THESAN Project (Kannan et al., 2021). This project performed a suite of 7 simulations over the course of reionization on a 95.5 cMpc (comoving megaparsec) box. These simulations were fully hydrodynamic and tracked down to a distance scale of 10 pc, with dark matter particles as small as $3.12 \times 10^6\ \mathrm{M}_\odot$ and gas particles as small as $5.82 \times 10^5\ \mathrm{M}_\odot$. These simulations used the moment method for radiative transfer, with a reduced speed of light factor of $c/5$, which is one of the largest of recent simulations. In Figure 1.1 we copy one of the more stunning visualizations from Kannan et al. (2021), which shows the course of reionization over a slice through each of the volumes.

1.3.5 Our Contribution to Reionization Research

This project began with a simple question: "does star formation happening in rapid bursts during the EoR lead to a more efficient or rapid reionization?" Our initial research into this question consisted of a semi-analytical model of the behavior of the IGM during and after star formation. After these investigations into the topic, we decided to devote the bulk of our research efforts into testing this idea in a robust and self consistent way. The amount of computational resources that modern EoR simulations have utilized is staggering, so we do not hope to be able to compete with the scale or fidelity that these simulations have achieved. We thus hope to achieve the following objectives with this thesis:

1. Introduce an algorithm that makes use of parallel GPU computing to tackle the ray tracing

RT problem. The inherently parallel nature of GPU computing, compounded by using multiple GPUs, allows users of our algorithm to iterate rapidly on different ideas and parameters. We believe this will allow users to test ideas which would otherwise be prohibitively expensive in terms of time and computational resources when using other algorithms.

2. Perform a suite of simulations using our algorithm to test our idea that short bursts of star formation can ionize the IGM more effectively and rapidly than continuously growing sources of star formation. Modern reionization simulations typically use a single model of star formation across their suites of multiple simulations, so we hope that by isolating and varying the mode of star formation, we can demonstrate an effect which is otherwise not well studied.

1.4 Thesis Outline

This thesis tells the story of our attempts to investigate how the mode of star formation affects the course of reionization. In Chapter 2, we discuss our initial investigation, which consisted of a semi-analytic model that I was able to use to write my Master's Thesis. This project gave us the motivation to continue on this line of research and pursue a more accurate and robust model of the effects we were investigating.

In Chapter 3, we present ARC, a fully functional ray-tracing radiative transfer library that we wrote with the intention of testing our hypotheses. We demonstrate that our library is able to match state of the art radiative transfer simulations in terms of accuracy, while also harnessing the power of parallel GPU computing. We built this library with features designed specifically to allow us to implement the types of bursty star formation we were interested in, which made it the

ideal test bed for our work.

In Chapter 4, we present the improvements made to this algorithm in the following years, as subsequent testing showed us that we needed more ways to evaluate and utilize data generated by the library. We also present the results of a suite of simulations run using ARC, with each set of simulations consisting of a pair of continuous and bursty models for direct comparison. These simulations showed that the differences we hoped to see were present in these simulations, and gave us various estimates of the magnitude of this effect on different parameters and observable quantities.

In Chapter 5, we briefly discuss the future of our research and ARC. We only scratched the surface of ideas for what bursty star formation look like during the EoR, so we present a selection of hypotheses that could be used for future simulations. We also believe that ARC is capable of answering many questions more quickly than would otherwise be possible because of our use of GPU computing. The individual modules within the library can also be adapted for use in other simulation platforms to allow them to harness GPU computing for their radiative transfer calculations.

Finally, In Appendix 5.2, we include relevant calculations to Chapter 2.

1.5 Summary of Facilities and Software

Computational resources used in this dissertation:

1. University of Maryland: Deepthought2 HPC cluster
2. University of Maryland: Zaratan HPC cluster
3. University of Maryland: Astronomy Department computer resources

4. Indiana University: Jetstream 2, through ACCESS

Software used in this dissertation:

1.

2. `NumPy`, `Matplotlib`, `yt`

3. `ARC`

Chapter 2: Modeling Reionization in a Bursty Universe

In this chapter, we present semi-analytic models of the epoch of reionization focusing on the differences between continuous and bursty star formation (SF). Our model utilizes physically motivated analytic fits to 1D radiative transfer simulations of H II regions around dark matter halos in a representative cosmic volume. Constraining our simulations with observed and extrapolated UV luminosity functions of high redshift galaxies, we find that for a fixed halo mass, stellar populations forming in bursty models produce larger H II regions which leave behind long-lived relic H II regions which are able to maintain partial ionization in the intergalactic medium (IGM) in a manner similar to an early X-ray background. The overall effect is a significant increase in the optical depth of the IGM, τ_e, and a milder increase of the redshift of reionization. To produce $\tau_e = 0.066$ observed by Planck and complete reionization by redshift $z_{\rm re} \sim 6$, models with bursty SF require an escape fraction $f_{\rm esc} \sim 2\% - 10\%$ that is 2-10 times lower than $f_{\rm esc} \sim 17\%$ found assuming continuous SF and is consistent with upper limits on $f_{\rm esc}$ from observations at $z = 0$ and $z \sim 1.3 - 6$. The ionizing photon budget needed to reproduce the observed τ_e and $z_{\rm re}$ depends on the period and duty cycle of the bursts of SF and the temperature of the neutral IGM. These results suggest that any remaining tension between observed and predicted ionizing photon budget for reionization can be alleviated if reionization is driven by short bursts of SF, perhaps relating to the formation of Population III stars and compact star clusters such as proto-globular

clusters.

2.1 Introduction

Reionization is one of the least understood epochs of the history of the universe, mainly because the dominant sources of ionizing radiation during this epoch are currently unknown. After recombination, the density perturbations throughout the universe grew until the first stars and galaxies formed. A fraction of the UV photons produced by these objects escaped from their host halos and began ionizing the inter-galactic medium (IGM). Some combination of stars, galaxies, and quasi-stellar objects (QSOs) produced enough ionizing photons to fully ionize the IGM by a redshift of $z \sim 6$, as required by a variety of observations Fan et al. (2002, 2006); Becker et al. (2007); Bolton et al. (2011); Mortlock et al. (2011); Planck Collaboration et al. (2015). The details of this phase transition are poorly understood, as few direct observations of the universe at high redshifts are possible.

Observations of the Lyman-α forest absorption lines in the spectrum of distant quasars have shown unambiguously that the IGM was ionized to a very high level by a redshift not much later than $z \sim 6$ (*e.g.*, Fan et al., 2006). Observations of Lyman-α emissions from sources at redshifts $z > 6$ indicate rapid changes in their abundance at these redshifts Ouchi et al. (2010); Kashikawa et al. (2011); Caruana et al. (2014). These observations, as well as the optical depth to Thomson scattering on free electrons of the IGM, $\tau_e \sim 0.066$, derived from CMB observations Planck Collaboration et al. (2015) and observations of UV light from redshift $z \sim 10$ galaxies McLure et al. (2010); Pentericci et al. (2011); Bouwens et al. (2013); Oesch et al. (2014), suggest that reionization was well underway significantly earlier. The duration of reionization and the nature

of the sources which supplied the necessary photons remains the subject of observational and theoretical research Venkatesan et al. (2001); Hansen & Haiman (2004); Madau et al. (2004); Ricotti & Ostriker (2004b,a); Ricotti et al. (2005, 2008); Volonteri & Gnedin (2009); Trenti et al. (2010); Wise et al. (2014); Boylan-Kolchin et al. (2015).

A number of different semi-analytic approaches exist for modeling reionization, including the excursion set formalism that takes into account clustering of sources (inside-out reionization, see Furlanetto et al., 2004; Alvarez & Abel, 2007) and inhomogeneous reionization models (outside-in reionization, see Miralda-Escudé et al., 2000; Wyithe & Loeb, 2003; Choudhury & Ferrara, 2005), that take into account the inhomogeneous density distribution of the IGM. These methods have historically required relatively high escape fractions ($> 10\% - 20\%$) or an escape fractions which increase with increasing redshift. Recently, (Mitra et al., 2015) have shown that their semi-analytic inhomogeneous reionization model is consistent with Planck's optical depth $\tau_e = 0.066$ and other observational constraints, assuming a non evolving $f_{\rm esc} \sim 10\%$. This is interesting and encouraging, however high resolution cosmic zoom-in simulations of galaxy formation have found $f_{\rm esc} < 5\%$ (Ma et al., 2015), and many authors found that $f_{\rm esc}$ is likely to decrease with increasing redshift Ricotti & Shull (2000); Kimm & Cen (2014). Observationally, measurement of the escape fraction are hard, but in many cases upper limits $f_{\rm esc} \leq 4\% - 8\%$ are found in local starburst galaxies (*e.g.*, Hurwitz et al., 1997; Boutsia et al., 2011; Nestor et al., 2013).

Numerical methods have also been used with some success, including radiative transfer simulations over static density fields (Zahn et al., 2011; Sokasian et al., 2001; Ciardi et al., 2003; Iliev et al., 2006b; McQuinn et al., 2007; Croft & Altay, 2008; Trac et al., 2008; Aubert & Teyssier, 2010; Ahn et al., 2012). More recently, full radiation hydrodynamics simulations which follow matter and radiation simultaneously and self consistently have been developed (Gnedin, 2000a;

Petkova & Springel, 2011; Paardekooper et al., 2013; So et al., 2014; Gnedin, 2014; Gnedin & Kaurov, 2014; Norman et al., 2015; Pawlik et al., 2015). However, all of these simulations treat the ionizing radiation photons as monochromatic: the 3D transfer of radiation is done on a low number (often one) of frequency bands.

In this paper we employ a simple model to explore the possibility that the sources of reionization had an intermittent UV emissivity. Sources that are definitively characterized by a bursty mode of star formation include Population III stars and the first small-mass dwarf galaxies Ricotti et al. (2002a,c); Schaerer (2003); O'Shea et al. (2015). Our model also includes radiation transfer of ionizing radiation well sampled in the frequency domain (about 400 logarithmically spaced frequency bins), allowing us to properly reproduce the width of ionization fronts and specific intensity of ionizing background radiation, that are affected by high energy photons. At low redshift, the halo matching technique has proven to be a good ansatz to match observed galaxies to dark matter halos from simulations Vale & Ostriker (2004, 2006); Guo et al. (2010); Moster et al. (2010). This method works by successively placing the brightest stellar populations within the most massive halos, neglecting the possibility that some halos may become significantly brighter for a brief period of time. However, at high redshift the high merger rate and the small masses of the first galaxies suggest that star formation in galaxies should be rather bursty, as confirmed by simulations. In particular, it has been suggested Ricotti (2002b); Katz & Ricotti (2013, 2014) that the formation of compact stellar systems before reionization, which may lead relics such as globular clusters, ultra-compact dwarfs and dwarf-globular transition objects, may dominate reionization. In these scenarios we expect an effective duty cycle for UV luminosity of the first galaxies, leading to a large fraction of halos of any given mass to be nearly dark in between short lived bursts.

The simplest semi-analytic method typically used in literature to investigate the ionization evolution of the IGM is, essentially, to keep a budget of hydrogen ionizing photons needed to maintain a fraction of the volume in IGM, $Q_{\rm HII}(t)$, fully ionized at a time t Madau et al. (1999); Kuhlen & Faucher-Giguère (2012). This filling fraction evolves according to a simple differential equation:

$$\frac{dQ_{\rm HII}}{dt} = \frac{\dot{n}_{\rm ion}}{\bar{n}_{\rm H}} - \frac{Q_{\rm HII}}{\bar{t}_{\rm rec}}. \tag{2.1}$$

Here, $\dot{n}_{\rm ion}$ is the rate of ionizing photon production per comoving volume, $\bar{n}_{\rm H}$ is the mean comoving cosmic number density of atomic hydrogen, and

$$\begin{aligned} \bar{t}_{\rm rec} &= \frac{1}{C_{\rm H\,II}\alpha_{\rm B}(T_0)\bar{n}_H(1+Y/4X)(1+z)^3} \\ &\approx 0.93\,{\rm Gyr}\left(\frac{C_{\rm H\,II}}{3}\right)^{-1}\left(\frac{T_0}{2\times10^4\,{\rm K}}\right)^{0.7}\left(\frac{1+z}{7}\right)^{-3} \end{aligned} \tag{2.2}$$

is a time scale of hydrogen recombination in fully ionized bubbles of H II . Here, $\alpha_{\rm B}$ is the case-B hydrogen recombination coefficient, T_0 is the IGM temperature at mean density, $C_{\rm H\,II}$ is the effective clumping factor in ionized gas, $X = 0.75$ is the hydrogen mass fraction and $Y = 0.25$ the helium mass fraction. We assume that helium is singly ionized at the same time as hydrogen, but only fully ionized at $z < 4$ (see § 2.2.4). However, the rate of hydrogen recombination is proportional to x_e^2, where x_e is the electron fraction in the IGM (so that $n_e = x_e\bar{n}_{\rm H}$). Thus, after a burst of star formation recombination proceeds quickly at first, but slows down as lower levels of partial ionization are reached. The method described by Equation (2.1) lumps the complicated spatial dependence of electron fraction and recombination rate together into $\bar{t}_{\rm rec}$, and as such does not take into account the volume filling fraction of partially ionized gas and the reduced

recombination rates in regions of partial ionization produced by intermittent star formation. Here, we focus on investigating the observational implications of assuming either a continuous or a bursty mode of star formation (SF) in the first galaxies, and how each affects the evolution of the IGM during the epoch of reionization (EOR). A sudden burst of star formation produces a much higher luminosity of ionizing photons than the same mass of stars forming continuously over a period significantly longer than the lifetime of the brightest stars ($\sim 5 - 10$ Myr). This increased luminosity produces a larger H II region than in the continuous case. Once the brightest stars have died, however, the larger region of ionization begins to recombine. The rate of recombination is proportional to the electron fraction, so that the electron fraction within these regions remains non-trivially boosted for long periods of time. We wish to investigate whether these relatively long lived relic H II regions of partially ionized gas have a signature similar to hypothetical X-ray preheating of the IGM Venkatesan et al. (2001); Ricotti & Ostriker (2004b); Ricotti et al. (2005) and if they produce an observable effect on the optical depth to Thompson scattering of the IGM τ_e. The model simulates a population of dark matter halos hosting star forming populations in a representative cosmic volume between the redshifts of $z = 30$ and $z \approx 5.8$, a period of 900 Myr. Luminosities are assigned through a halo matching process such that the galaxies UV luminosity functions (and the mean ionizing emissivity) in both cases are identical and match observations. In the Appendix we present physical models of cosmological H II regions and the evolution of the relic H II regions in the presence of bursty star formation. These physical models, calibrated to reproduce 1D radiation transfer simulations, provide useful equations for the evolution of the volume filling fraction of partially ionized gas. We use the results of these simulations to calculate the average electron fraction at a given time and the optical depth to Thompson scattering of the IGM and analyze how different modes of star formation affect these

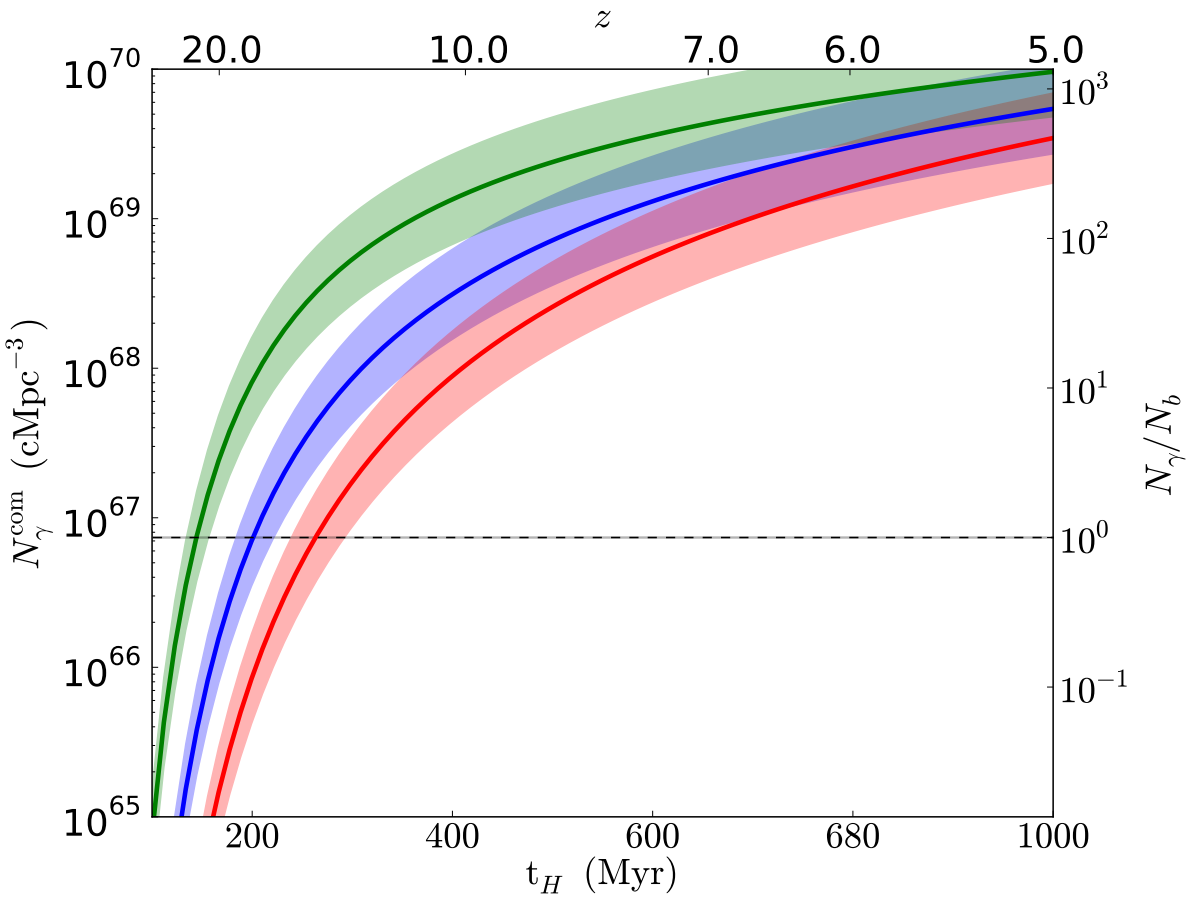

Figure 2.1: Cumulative photon count per comoving cubic Mpc for the FIT (blue), MIN (red), and MAX (green) models presented in Kuhlen & Faucher-Giguère (2012). The shaded regions represent the full range of emission hardness ($\zeta = 0.5$ to $\zeta = 2.0$; see text). The dotted line and accompanying shaded region shows the comoving number density of baryons as derived from the Planck results Planck Collaboration et al. (2015). We interpret the ratio of photon count to baryon count (plotted on the right axis) as the upper limit of $1/f_{\rm esc}$.

observable quantities. Near the redshift of reionization the overlap of individual H II regions produces ionized bubbles containing many UV sources. Hence, in this final phase the time-average UV emissivity within bubbles approaches the continuous limit even if the individual galactic sources are bursty. The analytic approximation utilized by our model also breaks down when H II of independent sources overlap. However, this is does not affect our main results as we find that the ionization history in bursty models differs from the continuous models mostly at

high-redshift, when the H II regions around individual sources hardly overlap. However, in order to be more quantitative, in Section 2.3.3 we use the results of our simulations to quantify the redshift at which these effects might become dominant, providing plots of the redshift at which a given bursty star formation model becomes better described by the continuous star formation model.

The paper is organized as follows. In section 2.2 we describe the methods used to simulate a cosmic volume during the EOR and how observable quantities are derived from the results. Most of the technical aspects of the method and a physically motivated analytic model describing recombining H II regions are presented in the Appendix. In section 2.3, we present results of a fiducial set of simulations and a parameters study to asses how the results depend on the free parameters in the model. In section 2.4, we present a summary and concluding remarks. Throughout this paper we assume a flat ΛCDM cosmology with $\tau_e = 0.066 \pm 0.012$, $H_0 = 67.51$ km s^{-1} Mpc^{-1}, $(\Omega_\Lambda, \Omega_m, \Omega_b, n_s, \sigma_8)$=$(0.691, 0.309, 0.0489, 0.9667, 0.816)$, as presented by the Planck Collaboration Planck Collaboration et al. (2015). A fiducial value for the redshift of reionization $z_{\rm re} = 6.0$ will be assumed as a constrain for parameter studies.

2.2 Overview and Methodology

Our simulation is a numerical representation of the evolution of a cosmic volume between the redshifts of $z \sim 30$ and $z \sim 5.8$, or a time of roughly 1 Gyr. The volume is a cube with side length 100 comoving $\rm Mpc$ (cMpc), which is enough to expect convergence of our results (*e.g.*, Iliev et al., 2014). Dark matter halos, extracted from a Press-Schechter distribution, are placed in the volume randomly (neglecting clustering). Newly virialized halos are added as the Hubble

time increases and star formation begins with the addition of the halo. In the continuous case, star formation proceeds continuously throughout the life of the halo, whereas in the bursty case star formation occurs periodically with a given period ΔT and duty cycle $f_{\rm duty}$. In both modes of SF the time averaged ionizing emissivity of the halos is the same.

2.2.1 Halo matching

We assign a luminosity to any given halo within our cosmic volume according to a halo matching procedure (see, Guo et al., 2010). We randomly generate halos of mass M at redshift z with a mass distribution $\phi_M(M, z)$ using the Press-Schechter formalism with the modification by Sheth-Tormen Press & Schechter (1974); Sheth & Tormen (2002) using Planck 2015 cosmological parameters. We limit the halos to a minimum mass $m_{\rm dm}$ (which we will vary to understand which halos contribute the most to our results.)

We assign a luminosity to each halo so that the population has a luminosity function described by a Schechter function $\phi_L(L, z)$ consistent with HST deep field observations. We use fits to the Schechter parameters evolution as a function of z from Kuhlen & Faucher-Giguère (2012). As in their study, we consider three models: "FIT," "MAX, and "MIN." These models are extrapolations of Schechter function parameters fit to published luminosity functions at different redshifts in the rest frame UV band at 1500 Å presented in Bouwens et al. (2015). The "FIT" model is the best linear regression for the time dependence of the Schechter parameters of the form $\{M^*, \log_{10}\phi^*, \alpha\} = A + B(z-6)$, while the "MAX" and "MIN" models independently adjust these parameters by $\pm 1\sigma$ (see Kuhlen & Faucher-Giguère (2012) for a discussion). In Figure 2.1 we plot the cumulative comoving hydrogen ionizing photon density derived from these models

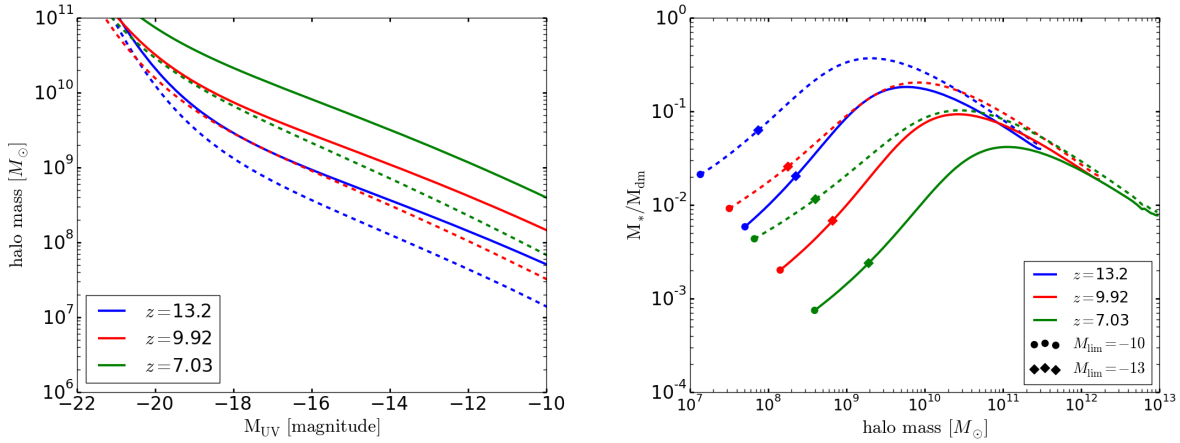

Figure 2.2: (*Left*). Typical halo mass hosting a galaxy of UV magnitude M_{UV} produced by the halo matching method in equation (2.4). The solid lines represent the continuous star formation, and the dashed lines represents bursty star formation with $f_{\rm duty} = 5\%$. (*Right*). Star formation efficiency $f_* \equiv M_*/M_{\rm dm}$ as a function of halo mass derived by our halo matching procedure. Halo matching in the bursty star formation models places brighter stellar populations in less massive halos.

(blue, green, and red lines representing "FIT," "MAX," and "MIN," respectively) plotted alongside the total comoving baryon density of the universe (dotted lines) and $M_{UV,lim} = -13$. We have used the same definition for the conversion between UV magnitudes and ionizing photon luminosity S_0 as in Kuhlen & Faucher-Giguère (2012):

$$S_0 = 2 \times 10^{25}\ {\rm s}^{-1} \left(\frac{L_{\nu,1500}}{\rm ergs\ s^{-1} Hz^{-1}} \right) \zeta, \tag{2.3}$$

where $\log_{10}\left(L_{\nu,1500}/({\rm ergs\ s^{-1} Hz^{-1}})\right) = 0.4(51.63 - M_{\rm UV})$. The shaded regions represent the range of possible spectral hardness ($0.5 < \zeta < 2.0$). We interpret the crossing of the shaded regions with the dotted line as the earliest epoch at which the universe may be fully reionized (neglecting recombinations in the IGM). The addition of an escape fraction $f_{\rm esc} < 1$ reduces the total number of photons reaching the IGM. So, this figure can be used to infer the absolute lower limit for $f_{\rm esc}$ in the tree models we consider for the emissivity to reionize by $z_{re} = 6$. Shifting the

cumulative photon counts down by $f_{\rm esc}$ until the shaded region crosses the dashed line at $z = 6$ gives $f_{\rm esc}$ for which each ionizing photon escaping into the IGM is used to ionize a hydrogen atom only once by redshift $z = 6$.

Halos are matched so that in both continuous and bursty models the overall luminosity function is identical. However, in the case of bursty star formation, we assume a periodic UV luminosity with period ΔT with a simple step function functional form within each period:

$$L(t) = \begin{cases} L_{UV,1500}/f_{\rm duty} & 0 < t < f_{\rm duty}\Delta T \\ 0 & f_{\rm duty}\Delta T < t < \Delta T, \end{cases}$$

where $L_{UV,1500}$ is the time averaged rest frame UV luminosity at 1500 Å over a period ΔT, and the burst has peak luminosity $L_{UV,1500}/f_{duty}$ of duration $T_{\rm on} = f_{\rm duty}\Delta T$. We assign a luminosity $L(z)$ to a halo of mass $M(z)$ so that:

$$\int_{L(z)}^{\infty} \phi_L(L', z)dL' = \int_{M(z)}^{\infty} f_{\rm duty}\phi_M(M', z)dM'. \tag{2.4}$$

Here, we interpret $f_{\rm duty} \leq 1$ as the fraction of halos emitting ionizing radiation at a given time. We thus take the number of available luminous halos for halo-matching to be $f_{\rm duty}\phi_M(M, z)$. In the case of continuous star formation, $f_{\rm duty} = 100\%$.

In Figure 2.2 (left) we show the result of this halo matching procedure at three sample redshifts indicated in the figure's legend. The UV luminosities are given in terms of UV magnitudes (M_{UV}) and the halo masses in solar masses. The solid lines represent the result of the procedure assuming continuous star formation ($f_{\rm duty} = 100\%$), while the dashed lines represent the result

of the procedure for bursty star formation with $f_{\rm duty} = 5\%$. The figure shows that if galaxies are bursty, an observed galaxy at a given UV magnitude observed in HST deep fields lives in dark matter halo that is less massive than it would be inferred if their stars formed continuously. The reverse argument is also true: a dark matter halo of a given mass that hosts a bursty galaxy has significantly higher UV luminosity (during the star burst) when compared to the luminosity of the same mass halo forming stars continuously. This also has indirect consequences on the theoretical expectations for $f_{\rm esc}$ in the continuous vs bursty models. Assuming that for a given mass halo the ISM structure is similar in the two models, the number of recombinations during the burst is proportional to the burst duration: $N_{\rm rec} = t_{\rm burst}/t_{\rm rec}$. The mean escape fraction over a burst cycle is $f_{\rm esc}$ =$1 - N_{rec}/N_{\rm ph}$, where $N_{\rm ph}$ is the given total number of ionizing photons emitted in one cycle (which is same in the two models by construction). Thus, the escape fraction will be higher for shorter burst of star formation and the smallest for a continuous mode of star formation. The right plot of Figure 2.2 shows the mean star forming efficiency M_*/M (stellar mass per unit dark matter mass) as a function of halo mass for the same three sample redshifts. The symbols toward the small mass end of the curves show the typical halo mass of galaxies with UV magnitudes $M_{\rm UV,lim} = -10$ (circles) and $M_{\rm UV,lim} = -13$ (squares). This is going to be relevant for models in which the faint end of the luminosity function is extrapolated to $M_{\rm UV,lim}$ and assumed to be zero at fainter magnitudes. The figure shows that galaxies of a given total mass have higher star formation efficiency when assuming a bursty model instead of a continuous star formation model.

2.2.2 Analytic approximation of Strömgren spheres

The starting point to derive our physical model for recombining H II regions are the radiation transfer simulations presented in Ricotti et al. (2001). This paper presents 1D radiative transfer simulations around a point source of given spectrum (Population II and Population III stars or miniquasars) in an expanding universe following the ionization state of hydrogen, helium and the formation of H_2 via the H^- catalyst. We used the same code to generate the electron fraction around a halo with a given time dependent spectral energy distribution (SED). We derived simple analytic models of these outputs for both continuous and bursty star formation SEDs (see Appendix A.1). The models allow to estimate the ionization fraction x_e at a distance R from a source or the distance from a source at which the electron fraction is x_e:

$$x_e(R) = f\left(z, R, L, z_{\rm on}(, z_{\rm off})\right) \tag{2.5}$$

$$R(x_e) = g\left(z, x_e, L, z_{\rm on}(, z_{\rm off})\right), \tag{2.6}$$

where L is the luminosity of the halo, z is the redshift (independent variable), $z_{\rm on}$ is the redshift at which the star formation began, $z_{\rm off}$ is the redshift at which star formation ends in the case of bursty star formation. The analytic description of our physical model can be divided into two regimes:

1. During the burst of star formation the state of the IGM around an isolated halo is simulated as a cosmological Strömgren sphere. The relatively low density of the IGM causes such spheres to have non-trivial transition regions. We found that the profile of these boundaries are well approximated by a simple analytic formula dependent only on a scale radius

$R_S(z, L, z_0)$ (see Appendix A.1.2). The functional form of $R_S(z, L, z_0)$ is taken to be that of a Strömgren sphere around a constant UV luminosity source in an expanding universe (see, Donahue & Shull, 1987; Shapiro & Giroux, 1987).

2. After the burst of star formation: we assume that the ionization rate becomes zero and the gas recombines in the expanding universe. We solve analytically the equations for the evolution of the electron fraction under these assumptions at a given distance from the source as a function of time (see Appendix A.1.1). We take the state of the IGM at the moment when star formation ends as the initial condition for the electron fraction and gas temperature. Interestingly the results are quite sensitive to the assumed IGM temperature outside the H II region, determined by the X-ray background radiation. With our analytic model we can therefore explore different scenarios beyond what we could do using pre-computed tables from a grid of 1D radiation transfer simulations.

Our simulation models a cosmic volume by randomly placing halos within the volume and assigning to each a set of parameters $(L, z_{\rm on}(, z_{\rm off}))$ such that the luminosity functions at each redshift match observations at $z < 10$ (see, Bouwens et al., 2015) and extrapolations to $z > 10$ (see, Faucher-Giguère et al., 2008, 2009; Kuhlen & Faucher-Giguère, 2012). The volume filling fractions at a given electron fraction $0 < x_e < 1$ (sampled uniformly with 60 bins in linear scale) are calculated using Equation (2.6) and output every 5 Myr.

2.2.3 Ionizing background

The evolution of the IGM ionization fraction obtained using the method presented in the previous section approaches unity asymptotically as a function of cosmic time, but reionization

is not fully complete (see Fig. 2.3). Thus, the value of τ_e can be calculated quite precisely but the redshift of reionization remains undetermined. Methods which track the photon budget such as Equation (2.1), assume ionizing photons are absorbed instantaneously and H II regions have sharp boundaries. Our method allows for regions near ionizing sources to be fully ionized while large regions of gas further from the sources are only partially ionized. When the average electron fraction of the universe is small, our method (for continuous star formation) reproduces rather closely Equation (2.1), as photons are effectively absorbed locally and the mean free path is shorter than the typical distance between sources. As the universe expands and the average electron fraction of the universe increases, the average density of neutral hydrogen decreases and the mean free path of photons in the IGM increases. These photons build up a ionizing background which becomes more dominant as more ionizing sources appear. The harder photons of the ionizing spectrum build up a background earlier than the softer photons due to their longer mean free path. Individual halos never produce large enough regions where the gas is fully ionized to completely reionize the cosmic volume, so the derived average electron fraction underestimates the true electron fraction. We correct for this underestimation by calculating and including in the photon budget the ionizing background and its effect on the cosmic ionization history. We quantify this effect solving the equation of radiation transfer in an homogeneous expanding universe as in Gnedin (2000a); Ricotti & Ostriker (2004b), which we briefly summarize here. We begin with the number density of ionizing background photons n_ν at a redshift z and evolve it to $z - \Delta z$. During each code timestep Δz, we add to the initial background at redshift z (appropriately redshifted and absorbed by the neutral IGM) the photons produced by ionizing sources within our simulation between the redshifts of $z - \Delta z_0$ and $z - \Delta z$ including absorption and redshift effects (source term), where $\Delta z_0 \equiv H(z)R_0/c$ and R_0 is the minimum comoving

distance of any emitting source that contributes to the radiation background. Mathematically, we solve the equation:

$$n_\nu(z-\Delta z) = n_\nu(z)\exp\left[-\int_z^{z-\Delta z} dz' \alpha_{\nu'}(z')\right] + \int_{z-\Delta z_0}^{z-\Delta z} dz' S_{\nu'}(z')\exp\left[-\int_{z'}^{z-\Delta z} dz'' \alpha_{\nu''}(z'')\right], \tag{2.7}$$

where $\nu' = \nu(1+z')/(1+z)$ we have defined a dimensionless absorption coefficient and source function:

$$\alpha_\nu = \frac{(1+z)^2}{H(z)} c\overline{n}_H \sigma_\nu(\mathrm{H\,I})(1-x_e(z)), \tag{2.8}$$

$$S_\nu = \frac{\dot{n}_{\rm ion}\langle h\nu\rangle g_\nu/h\nu}{(1+z)H(z)}, \tag{2.9}$$

where the sources spectra are normalized as $\int_{\nu_0}^\infty g_\nu d\nu = 1$, with $h\nu_0 = 13.6$ eV and $\langle h\nu\rangle^{-1} \equiv \int_{\nu_0}^\infty (g_\nu/h\nu)d\nu$.

Sources at $R < R_0$ are local and their radiation is used to produce individual ionization bubbles and are thus excluded from the background calculation. We take R_0 to be the average distance between the dimmest (and most numerous) objects, so that the background begins at the distance of the nearest luminous objects. We also note that this distance decreases as more collapsed structures form, allowing the background to become more prominent at later times. The redshift of reionization $z_{\rm re}$ is rather sensitive to the choice of R_0 and the spectrum of the sources. Since it is difficult to make a precise estimate of R_0 (clustering of sources and other subtleties will affect the value of R_0) and because the frequency dependence of $f_{\rm esc}$ is unknown from either theory or observations, the redshift at which reionization is completed remains somewhat uncertain in our

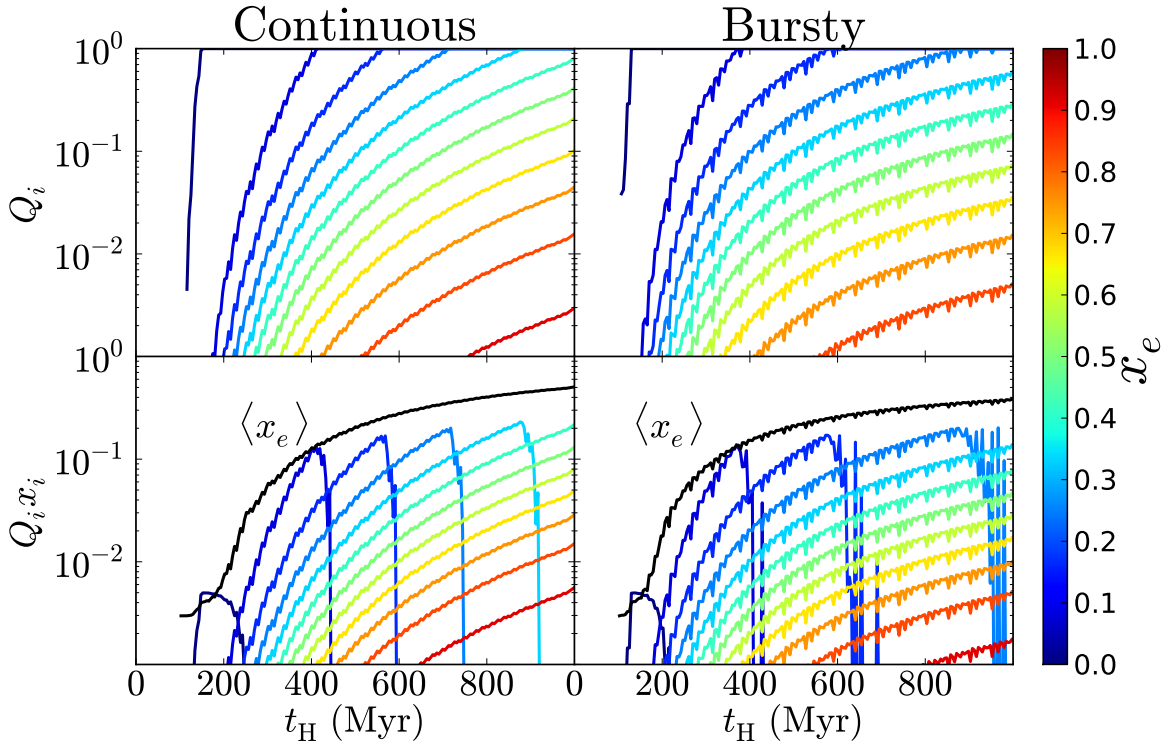

Figure 2.3: (*Top panels*). Volume filling fractions Q_i of partially ionized gas with $x_e < x_i$ as a function of time for the continuous (left) and fiducial bursty (right) cases. (*Bottom panels*). Weighted electron filling fractions ($Q_i' x_i$) for the same models. The solid black lines represent the derived x_e from equation (2.11) for both models. Note that a given electron fraction Q_i' increase towards 1, but decrease as the next higher electron fraction Q_{i+1} begins filling out the space indicated by Q_i.

approximate separation between background and local sources. However, the relative difference in $z_{\rm re}$ between continuous and bursty models with different duty cycles is robust.

For our background calculation, we consider sources with a simple power law spectrum $g_\nu \propto \nu^{-1}$ truncated at $h\nu = 100$ eV, absorbed by a column density $N_{\rm H\,I}$ of neutral gas. In order to consider the possibility that the preferential absorption of soft UV photons by neutral hydrogen inside the galaxy halo may result in a hardening of the spectrum emitted into the IGM, we consider two cases in our results: (i) $N_{\rm H\,I} = 0$, *i.e.*, the spectrum is not affected by hydrogen absorption (ii) $N_{\rm H\,I} = 3.8 \times 10^{18}$ cm^{-2}, which corresponds to an escape fraction of $f_{\rm esc} \sim 5\%$ and a significantly

larger $\langle h\nu \rangle$ than in the case of an purely stellar spectrum.

2.2.4 Calculating observable quantities

Our simulation produces an array $Q_i(t)$ which represents the volume filling fraction of gas with electron fraction $x_e < x_i$ with $i = 0, ..., n-1$. We assume that regions of higher ionization are nested within regions of lower ionization, so that $Q_0 < Q_i < ... < Q_{(n-1)}$. The filling fraction of gas with $x_{j-1} < x_e < x_j$ is

$$Q_i' = \begin{cases} Q_i - Q_{i-1} & \text{if } 1 < i < n-1, \\ Q_0 & \text{if } i = 0. \end{cases} \tag{2.10}$$

We compute the average electron fraction as:

$$\langle x_e(t) \rangle = \sum_{i=0}^{n-1} x_i Q_i'(t) \tag{2.11}$$

We let $\tau_e = \tau_0 + \Delta\tau$, where τ_0 is the contribution to τ_e from $z = 0$ to the time our simulation ends ($z_0 = 5.8$) and reionization has happen, and $\Delta\tau$ is the contribution to τ_e from the simulation. The average optical depth of reionization may now be calculated using the formula:

$$\Delta\tau_e(t) = c\sigma_T \int_{t_0}^{t} dt\, n_e(t) = c\sigma_T \int_{t_0}^{t} \left(\sum_{i=0}^{n-1} x_i n_H(t') Q_i'(t') \right) dt', \tag{2.12}$$

$$\tau_0 = \int_0^{z_0} dz \frac{c(z+z)^2}{H(z)} \sigma_T \left(1 + \eta(z) T/4X\right). \tag{2.13}$$

Here, $H(z)$ is the Hubble parameter and σ_T is the cross section of Thomson scattering. We let helium be singly ionized ($\eta = 1$) for $z > 4$ and doubly ionized ($\eta = 2$) at lower redshifts so that we may directly compare our results with those presented in Kuhlen & Faucher-Giguère (2012).

2.3 Results

For the results presented in the following sections, we chose UV luminosity functions with Schechter parameters from the FIT model. We take a fiducial choice of $M_{\text{UV, lim}} = -13$ and $m_{dm} = 10^7$ $M_\odot$. For the instantaneous star formation case, we make fiducial choices of $\Delta T = 100$ Myr and $t_{\text{on}} = 5$ Myr (thus, $f_{\text{duty}} = 5\%$). The top panels in Figure 2.3 show the primary output of our simulation, $Q_i(t) \equiv Q(x_e, t)$, for both continuous and instantaneous SF. We see that the filling fraction for higher levels of partial ionization are smaller in the instantaneous SF universe relative to those in the continuous SF universe. This is a result of the duty cycle of star formation in the case of instantaneous SF, wherein only the fraction of the halos actively producing stars are able to maintain high levels of ionization. We also see that the volume filling fraction at lower electron fractions rise at earlier times in the instantaneous SF case. This is due to the presence of relic regions of partial ionization around halos which have stopped forming stars and the higher luminosity of the burst at a given halo mass. The neutral-ionized transition regions surrounding star forming halos are not sharp for either star formation modes (see Appendix A.1.1 for a discussion), so there are large volumes of partially ionized H II in both cases. These regions may be missed in full radiative transfer simulations which utilize only one or two frequency bins for ionizing radiation.

The bottom panels of Figure 2.3 show the exclusive filling factors weighted by electron

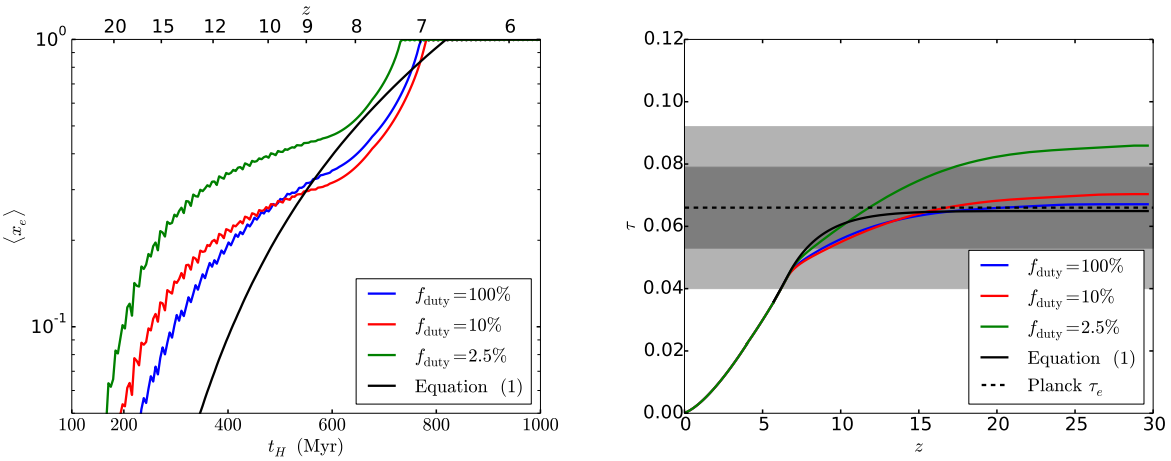

Figure 2.4: Average electron fraction (left panel) and integrated τ_e (right panel) plotted as a function of time (and redshift) for the fiducial continuous model and two bursty models ($f_{duty} = 10\%$ red line, $f_{duty} = 2.5\%$, blue line) with the same $\Delta T = 100$ Myr and $f_{\rm esc} = 17\%$. The dashed line in the right panel shows the measured optical depth to Thompson scattering of the IGM due to reionization Planck Collaboration et al. (2015), with the shaded regions representing the 68% and 95% confidence regions. The plotted simulation results include the ionizing background, so that reionization is complete at the expected time. The black line in both plots represents a sample result produced by Equation (2.1) with $f_{\rm esc} = 20\%$.

fraction, $Q'(x_e, t)x_e$, plotted for both continuous and instantaneous SF. We plot 20 of the 60 bins color coded such that higher electron fractions are more red and lower electron fractions are more yellow ($0.001 < x_e < 0.9$ in 20 uniform intervals as noted above). Each Q' increases as more halos are formed until one of two things happen: (i) a large halo which dominates the region stops forming stars, in which case the higher electron fraction filling factors decrease more rapidly than the lower level filling factors; (ii) once a filling factor $Q(x_j, t)$ for a given electron fraction x_j reaches unity: its corresponding exclusive filling factor $1 - Q(x_{j+1}, t)$ will only decrease as Q for lower levels of ionization grow. The black lines in the bottom panels of Figure 2.3 show the average electron fraction $\langle x_e \rangle$, which is simply a sum of all of the curves below it. The left plot of Figure 2.4 shows the average electron fraction as calculated with equation (2.11) for the continuous star formation and two choices of bursty star formation (both with $\Delta T = 100$ Myr

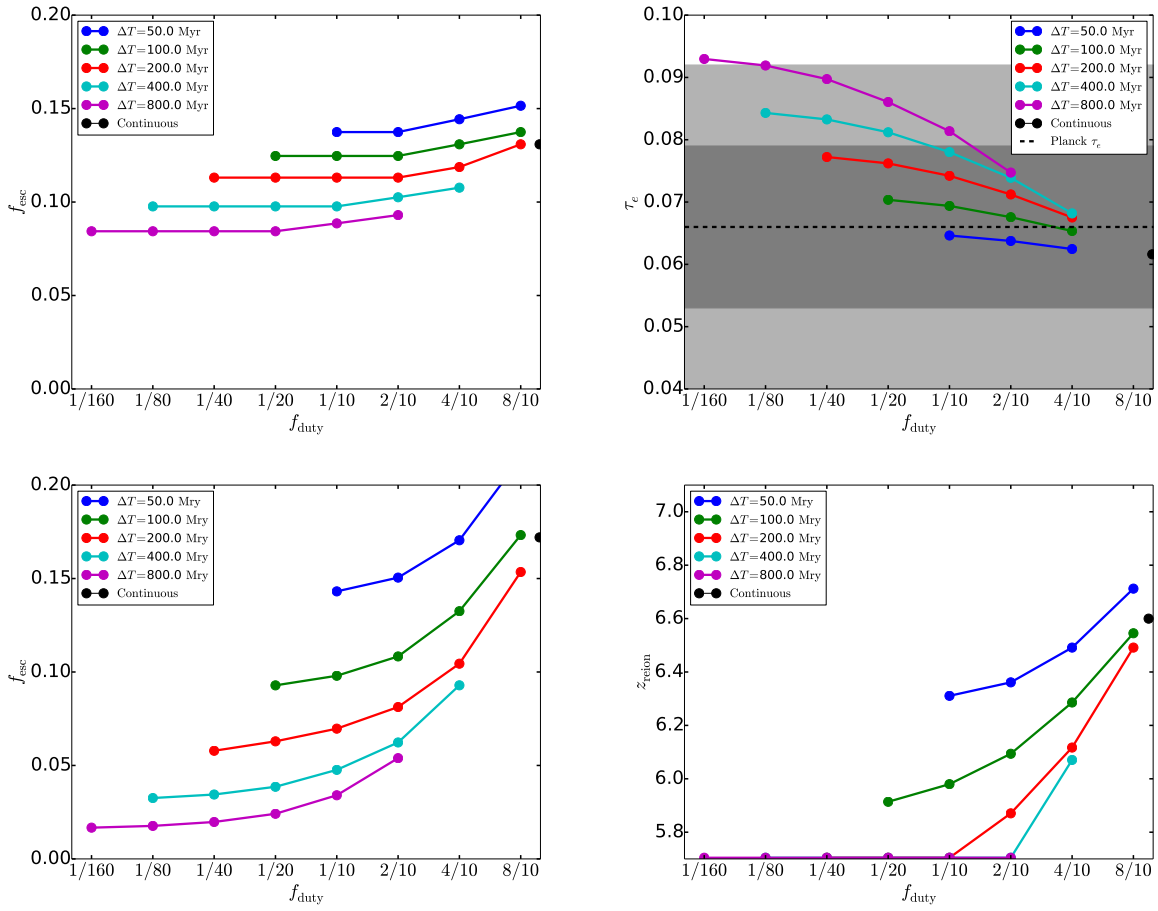

Figure 2.5: Suite of runs for various values of $(\Delta T, T_{\rm on})$ matched to fixed observational constraints. All plotted points assume spectra produced within the halos are unaffected by absorption within the halos ($N_{HI} = 0$ cm^{-2}). (*Top-Left*). Escape fraction $f_{\rm esc}$ as a function of duty cycle, $f_{\rm duty}$, for five sets of runs completing reionization at $z_{\rm re} = 6.25$. Points connected by a line have the same burst periods ΔT, as shown in the legend. We see here that runs with lower duty cycles require lower escape fractions to produce the same optical depth of reionization. (*Top-right*). IGM optical depth τ_e as a function of duty cycle for the same five sets of runs on the top-left. The dashed line and shaded regions refer to Planck's measurement of τ_e with 68% and 95% confidence intervals for the measurement error. (*Bottom-Left*). Escape fraction $f_{\rm esc}$ as a function of duty cycle, for the same five sets of runs but keeping the optical depth to Thompson scattering fixed at $\tau_e = 0.066$. (*Bottom-right*). Reionization redshift $z_{\rm re}$ plotted as a function of duty cycle for the same five sets of runs on the bottom-left. In all plots, the continuous case is represented by the black point at $f_{\rm duty} = 1.0$.

and $f_{duty} = 10\%$ and 2.5% respectively). We see that the electron fraction in the bursty case is higher at higher redshifts than in the continuous case. In the right panel of Figure 2.4 we plot the integrated optical depths to reionization as a function of redshift for the same models as in the left panel. For redshifts less than $z \sim 5.8$ at the end of our simulation, we integrated the formula for τ_e assuming a fully ionized universe as in Equation (2.13). The shaded region shows the 1σ and 2σ confidence levels from Planck Collaboration et al. (2015). Here we see that τ_e is significantly larger in the case of instantaneous SF even though $f_{\rm esc}$ and all other parameters are kept constant.

2.3.1 Variation of parameters

Our simulations allow us to compare the continuous star formation model to the bursty star formation model in a universe with freely chosen parameters including the burst duty cycle, $f_{\rm duty}$, burst period ΔT, minimum halo luminosity $M_{\rm UV,lim}$, and minimum halo mass $m_{\rm dm}$. To help us understand the role of these parameters, we take either τ_e or $z_{\rm re}$ as observational constraints and vary $f_{\rm esc}$ for a given set of starting parameters to match the desired observational constraint. The results of this parameter study are shown in Figure 2.5. In the top panels we show a set of runs where we adjusted $f_{\rm esc}$ to match $z_{\rm re} = 6.0$, allowing us to examine the effect of the burst period and duty cycle on the best fit $f_{\rm esc}$ (top-left panel) and resulting optical depth of reionization τ_e (top-right panel). We ran suite of simulations with ΔT, $T_{\rm on}$ selected from the sets $\{50, 100, 200, 400\}$ Myr and $\{5, 10, 20, 40\}$ Myr. $T_{\rm on} = 5$ Myr was chosen as the shortest burst length as this is the timescale for the life of stars in a truly instantaneous starburst. Here we see that the duty cycle of the burst, that mainly affect the peak luminosity of the burst and the halo mass in which such luminosity is observed, does not have a strong effect on $f_{\rm esc}$ and therefore the

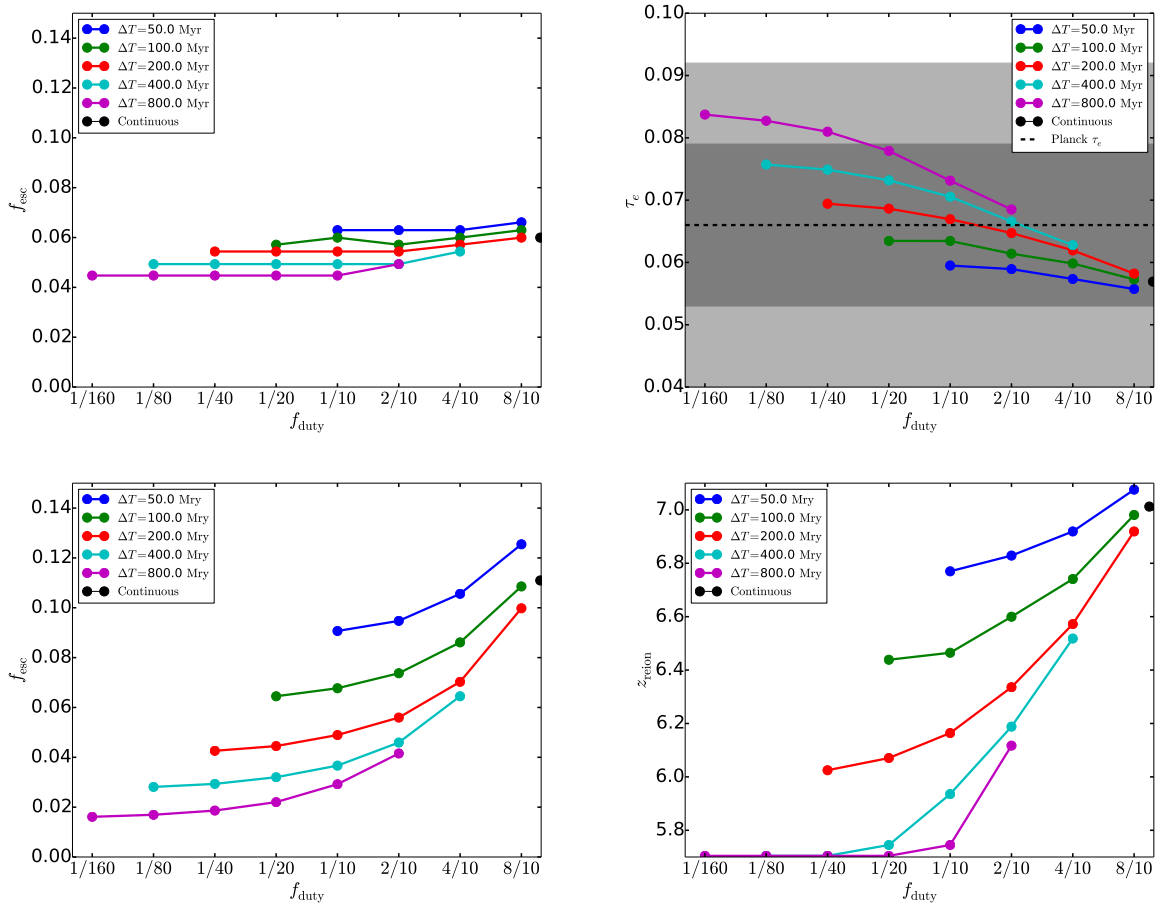

Figure 2.6: Same plots as in Fig. 2.5 but assuming a spectrum of the sources absorbed by a fixed column density of neutral hydrogen $N_{HI} = 3.8 \times 10^{19}$ cm^{-2}. Neutral hydrogen preferentially absorbs softer photons, thus resulting in a harder final spectrum. The overall result is a lower optical depth (for fixed $z_{\rm re}$), earlier completion of reionization (for fixed τ_e), and a lower escape fraction (for fixed τ_e, $z_{\rm re}$.)

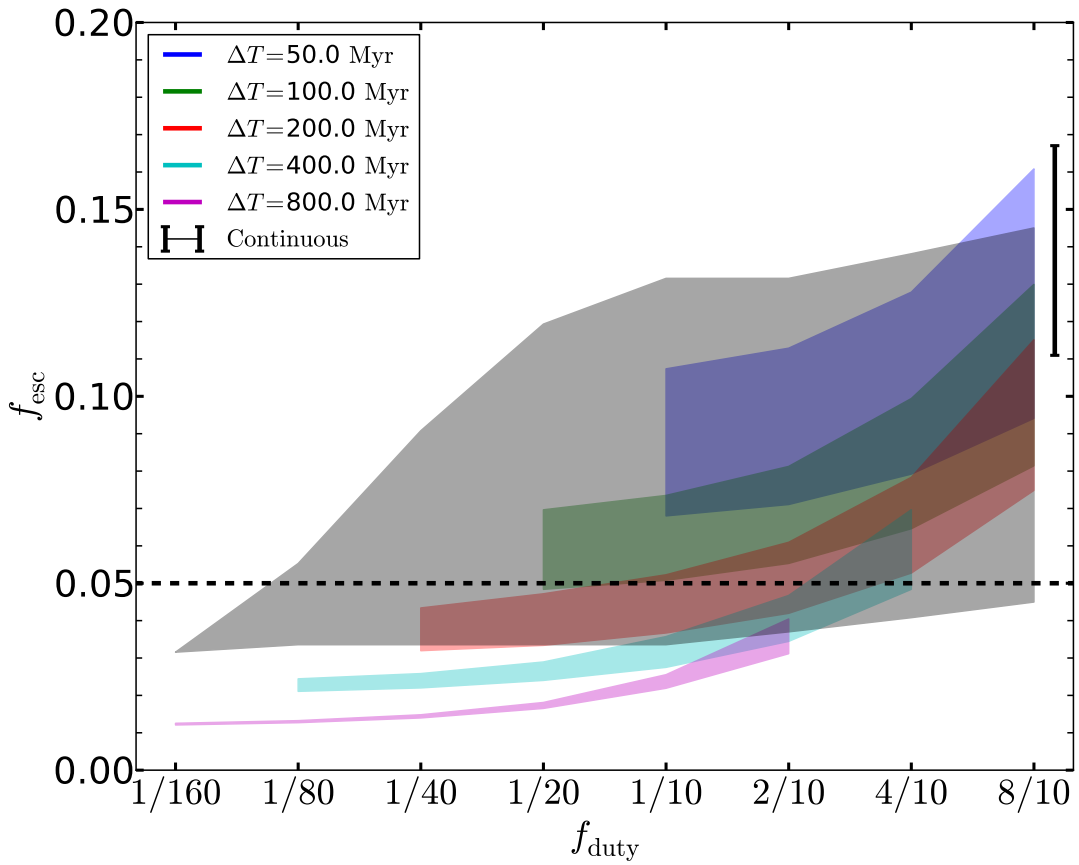

Figure 2.7: Combined observational constraints on $f_{\rm esc}$ in bursty (shaded regions) and continuous (black bar on the top-right) models. The shaded gray region represents the set of $(f_{\rm duty}, f_{\rm esc})$ values which, for a value of the absorption column density $0 \leq N_{HI} \leq 3.8 \times 10^{19}$ cm^{-2}, satisfy the observational constraints $6.0 \leq z_{\rm re} \leq 6.3$ and $0.053 \leq \tau_e \leq 0.079$. The colored regions represent values of $f_{\rm esc}$ which produce $\tau_e = 0.066$ for the same range of N_{HI}. The dashed line represents the upper limit $f_{\rm esc} \leq 5\%$ that would match all observational constraints on reionization, including Lyman-α forest data at $z < 6$ presented in Figure 2.9 and Table 2.1, assuming $f_{\rm esc}$ does nor evolve with redshift throughout the cosmic history.

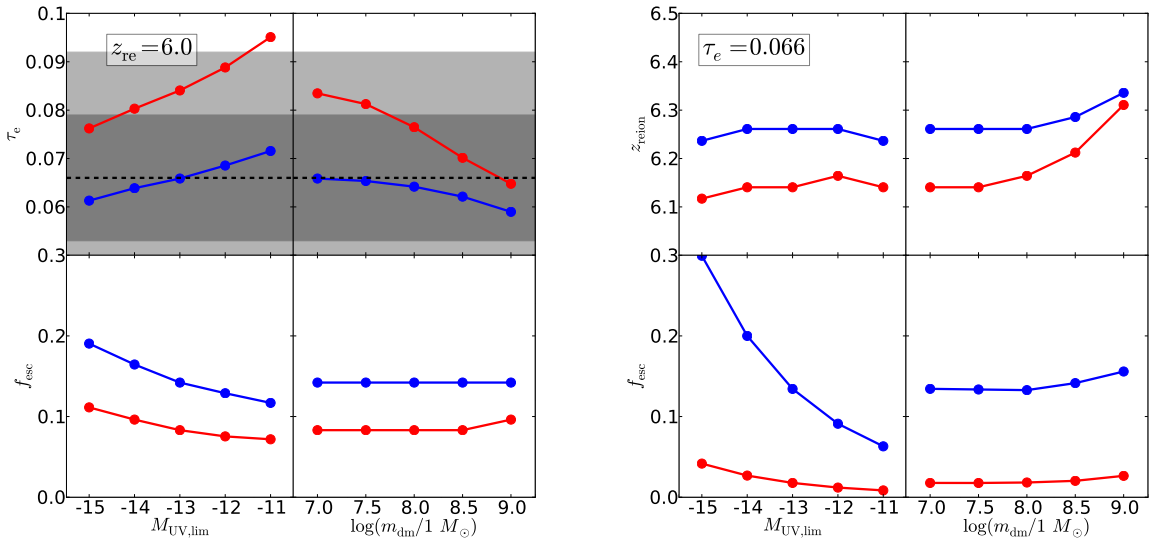

Figure 2.8: Behavior of the complete model constrained to observable quantities under different choices of $M_{\rm lim}$ and $m_{\rm dm}$. The blue and red lines represent the continuous and bursty cases, respectively. For the bursty model we adhere to our original fiducial choice of $(\Delta T, T_{\rm on}) = (50\,{\rm Myr}, 5\,{\rm Myr})$. (*Left*). Model constrained to $z_{\rm re} = 6.0$. We see that increasing $M_{\rm lim}$, thus adding more photons to our simulation, increases τ_e while requiring a smaller $f_{\rm esc}$ to maintain $z_{\rm re} = 6.0$. We also see that increasing $m_{\rm dm}$, thus removing lower mass halos from our simulation, decreases τ_e and increases $f_{\rm esc}$. Note that the bursty mode is more affected by this change, as in our simulation bursty stellar populations inhabit lower mass halos. (*Right*). Model constrained to $\tau_e = 0.066$. We note that this plot exhibits similar trends as in the previous plot, with decreasing $z_{\rm re}$ replacing increasing τ_e. We also note that the effect of altering $M_{\rm lim}$, $m_{\rm dm}$ is greater in this case. This reflects the fact that our $z_{\rm re}$ is dominated by the ionizing background, and is thus less sensitive to variations in these parameters.

redshift of reionization (that is constrained). This makes sense because the redshift of reionization is determined by the ionizing background that is sensitive to the average halo luminosity rather than its bursty nature. However increasing the period between bursts has the effect of reducing $f_{\rm esc}$ needed to reionize at $z_{rmre} = 6$. This is likely because the enhance electron fraction of the IGM in bursty models makes the IGM transparent to ionizing radiation earlier. The increase of the electron fraction in bursty models with increasing ΔT is evident as an increase in τ_e shown in the top-right panel. In the bottom panels we show a second set of runs where we constrained $f_{\rm esc}$ to produce $\tau_e = 0.066$, this time to examine the effect of ΔT and $f_{\rm duty}$ on $f_{\rm esc}$ (bottom-left panel) and redshift of reionization $z_{\rm re}$ (bottom-right panel). Here we see that decreasing the duty cycle decreases the required escape fraction to produce $\tau_e = 0.066$. However, the decrease in $f_{\rm esc}$ is more sensitive to an increase of the period between bursts. Hence, the parameter that affects τ_e most significantly appears to be $T_{\rm off} \equiv \Delta T - T_{\rm on} = \Delta T(1 - f_{\rm duty}^{-1})$, allowing for the greater contribution from the relic regions of partial ionization to the ionization history. To summarize, we see that decreasing the duty cycle of bursty star formation (or increasing the period between bursts) produces the same τ_e with a lower required escape fraction and given that our results for continuous star formation agree with previous predictions found in the literature, we believe that this effect may alleviate the need for high escape fractions or extrapolations of the faint end of the luminosity function to very low values to explain the observed optical depth of reionization.

The results of Figure 2.5 assume a photon spectrum which is a pure stellar spectrum unaffected by absorption by neutral hydrogen in the ISM or gas within halos. In Figure 2.6, we present the same results as Figure 2.5, this time with a fixed absorption of a column density $N_{HI} = 3.8 \times 10^{19}$ cm^{-2} of neutral hydrogen. Lower energy photons have a shorter mean free path in neutral hydrogen, so that they are preferentially absorbed. The results presented in Figure 2.6 have the same total

photon count as before, but with a higher $\langle h\nu \rangle$ of the harder spectrum. The top plots, for which $z_{\rm re} = 6.0$, show a lower escape fraction and lower optical depth of reionization than Figure 2.5. The harder photon spectrum produces a stronger ionizing background, thus lowering the required escape fraction and resulting optical depth of reionization. The bottom plots, for which $\tau_e = 0.066$, show a lower escape fraction and higher redshift of reionization than Figure 2.5. The ionizing background dominates the redshift of reionization, so that the earlier build up of the ionizing background due to higher average photon energy of the sources anticipates the redshift of reionization $z_{\rm re}$.

We may combine the results presented in these plots to construct a simple constraint on the allowed values of $f_{\rm esc}$ as a function of $f_{\rm duty}$. The gray shaded region of Figure 2.7 represents the values of $(f_{\rm duty}, f_{\rm esc})$ which satisfy the constraints $0.054 < \tau_e < 0.078$ and $6.0 < z_{\rm re} < 6.3$ for some value of $N_{HI} < 3.8 \times 10^{19}$ cm^{-2}. The separate error bar represents the same quantity for $f_{\rm duty} = 1$, i.e. continuous star formation. The dashed line represents the upper limit on $f_{\rm esc}$ which allows a constant $f_{\rm esc}$ as a function of redshift to reproduce Lyman-α observations (see Section 3.2). To emphasize the dependence of $f_{\rm esc}$ on $f_{\rm duty}$ and the period of the bursts ΔT, we have also plotted the values in the $(f_{\rm duty}, f_{\rm esc})$ plane which produce $\tau_e = 0.066$ in addtion to reionize between $6.0 < z_{\rm re} < 6.3$ for some value $N_{HI} < 3.8 \times 10^{19}$ cm^{-2}, represented by the shaded colored regions of this figure. We see that for a given ΔT, a shorter duty cycle allows for a lower $f_{\rm esc}$, and that a larger ΔT has also the effect of lowering $f_{\rm esc}$.

We also ran two suites of simulations to examine the effect of altering the parameters $M_{\rm UV,lim}$ and $m_{\rm dm}$. For the the first suite of runs we fixed $\Delta T = 50$ Myr, $T_{\rm on} = 5$ Myr, $m_{\rm dm} = 10^7$ M$_\odot$ and let the luminosity cut at the faint end of the luminosity function vary: $M_{\rm UV,lim} = (-15, -14, -13, -12, -11)$. For the second suite of runs we fixed $\Delta T = 50$ Myr, $T_{\rm on} =$

5 Myr, $M_{\rm UV,lim} = -13$ and let the cut of the halo mass function vary: $\log(m_{\rm dm}/1\,M_\odot) = (7.0, 7.5, 8.0, 8.5, 9.0)$. In Figure 2.8 we plot the results of these simulations, with τ_e and $z_{\rm re}$ constrained as in Figures 2.5-2.6. In both figures, the blue lines represent the continuous model, while the red line represents the bursty model. The left plot of Figure 2.8 shows the results of both suites with the constraint $z_{\rm re} = 6.0$. We see that increasing $M_{\rm UV,lim}$, thus increasing number of halos emitting ionizing photons in the simulation, has the expected of effect of decreasing the escape fraction necessary to produce the fixed reionization redshift. Similarly, we see that increasing $m_{\rm dm}$, thus decreasing the number of low mass halos in the simulation, increases the escape fraction necessary to match the constraint reionization redshift. In the left panel we note that the increase in $f_{\rm esc}$ is marginal, while the decrease in τ_e is significant, so that low mass halos contribute significantly to τ_e, but not as significantly to the ionizing background and $z_{\rm re}$. We also note that the change in both quantities is more significant for the bursty case, a result of the fact that bursty stellar populations in our simulation inhabit lower mass halos. The right plot of Figure 2.8 shows the results of both suites, now subject to the constraint $\tau_e = 0.066$. The results of this plot reflect those of the previous plot; increasing $M_{\rm UV,lim}$ (adding photons) or decreasing $m_{\rm dm}$ (adding halos) decreases the escape fraction necessary to reach the constrained τ_e. We note that the changes in $f_{\rm esc}$ and $z_{\rm re}$ are more significant than in the left panel, a reflection of the fact that τ_e is more sensitive to these parameters. We also note that increasing $M_{\rm UV,lim}$ too much for this bursty model means that τ_e can be nearly or completely achieved before the background completes reionization, so that these models can't complete reionization within the span of our simulation.

Table 1
Observational constraints.

Constraint #	Redshift	Constraint	Technique	References
1a.	5.9	$Q_{\rm H\,II} > 0.89$	Dark Gaps in Quasar Spectra	McGreer et al. (2015)
2a.	6.24-6.42	$Q_{\rm H\,II} < 0.9\,(2\sigma)$	Lyα Damping Wing of Quasars	Ota (2008)
3a.	7.0	$Q_{\rm H\,II} < 0.5$	Clustering of Lyα Emitting Galaxies	Tilvi et al. (2014)
1b.	6.0	$\dot{n}_{\rm ion}^{\rm com} < 2.6\,(< 2.6)$	Lyα forest	BH07, SC10
2b.	5.0	$\dot{n}_{\rm ion}^{\rm com} = 4.3 \pm 2.6\,(\pm 2.6)$	Lyα forest	BH07, SC10
3b.	4.2	$\dot{n}_{\rm ion}^{\rm com} = 3.5 \pm 0.8\,(^{+2.9}_{-2.2})$	Lyα forest	FG08, P09
4b.	4.0	$\dot{n}_{\rm ion}^{\rm com} = 3.2 \pm 0.4\,(^{+2.2}_{-1.9})$	Lyα forest	FG08, P09
5b.	3.8	$\dot{n}_{\rm ion}^{\rm com} = 2.8 \pm 0.3\,(^{+1.8}_{-1.6})$	Lyα forest	FG08, P09
6b.	3.6	$\dot{n}_{\rm ion}^{\rm com} = 2.6 \pm 0.3\,(^{+1.7}_{-1.5})$	Lyα forest	FG08, P09
7b.	3.4	$\dot{n}_{\rm ion}^{\rm com} = 2.8 \pm 0.7\,(^{+2.5}_{-1.8})$	Lyα forest	FG08, SC10

Table 2.1: (a) Observational Constraints on the reionization history of the universe (b) Observational Constraints on the high redshift ionizing emissivity. Total uncertainties for $\dot{n}_{\rm ion}^{\rm com}$, which include systematic effects due to the spectral shape of the observational UV background and the thermal history of the IGM, are shown in parenthesis. $\dot{n}_{\rm ion}^{\rm com}$ is shown in units of $10^{50} s^{-1}\,{\rm cMpc}^{-3}$. Constraints are taken from Bolton & Haehnelt (2007)(BH07), Songaila & Cowie (2010)(SC10), Faucher-Giguère et al. (2008)(FG08), and Prochaska et al. (2009)(P09).

2.3.2 Consistency with Lyman-αorest observations

We have thus far focused on how the calculated τ_e or $z_{\rm re}$ may be constrained for a model with fixed parameters $(\Delta T, T_{\rm on}, M_{\rm lim}, m_{\rm dm})$ by varying $f_{\rm esc}$. For any given set of these parameters, our model produces an average electron fraction $\langle x_e(z) \rangle$ and ionizing emissivity $\dot{n}_{\rm ion}^{\rm com}(z)$ at all times throughout our simulation. The value of these quantities may be estimated from to redshifts as high is $z = 6 - 7$ using observations of the Lyman-α forest. We included a collection of such data in Table 2.1. We may compare these observations to the outputs of our simulation (which ends at $z \sim 5.8$) to see how well our outputs compare to real data, and if the inclusion of partially ionized precursor of H II regions and relic H II regions improves the agreement with observations.

In the left plot of Figure 2.9 we plot the derived $\langle x_e \rangle$ for three simulations against the observational constraints on the average electron filling fraction $Q_{\rm H\,II}$ in Table 2.1. The model

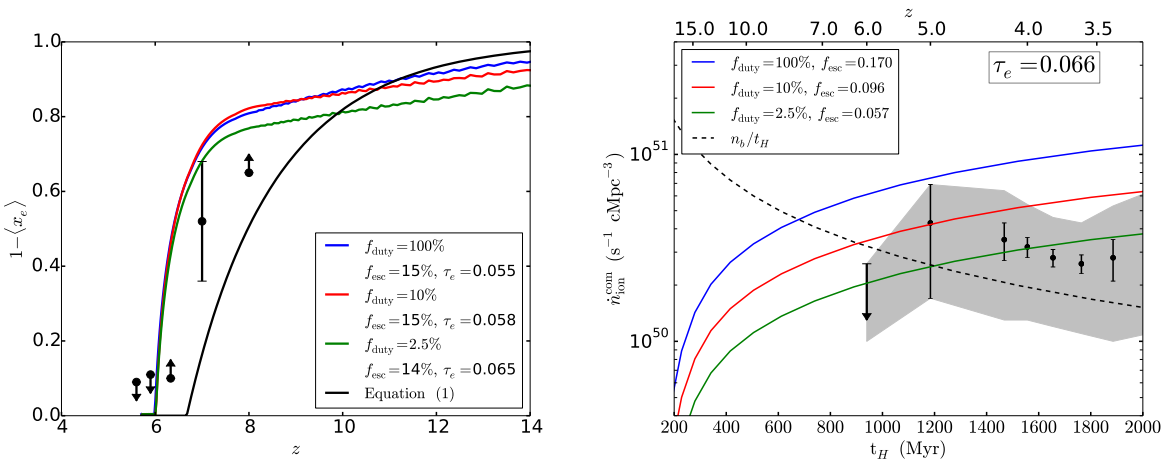

Figure 2.9: (left) A comparison of the simulated average electron filling fraction of the continuous model and two bursty models to observational data presented in Table 2.1. All three runs have $f_{\rm esc}$ adjusted so that $z_{\rm re} \sim 6.0$. The black line represents a sample calculation using equation (2.1). The ionizing background dominates when reionization is completed, and is most sensitive to the escape fraction, so that $f_{\rm esc}$ is similar for all three models. The bursty models, however, have higher electron fractions at high redshifts, so that τ_e is higher for these models. (right) A comparison of the integrated photon production rate for the continuous model and two bursty models with observational data presented in Table 2.1. All three runs have $f_{\rm esc}$ adjusted so that $\tau_e = 0.066$. We see that a lower duty cycle allows for a lower escape fraction to produce the same τ_e, allowing for better agreement with observational data.

with $f_{\rm duty} = 10\%$ has $(\Delta T, T_{\rm on}) = (100, 10)$ Myr while the model with $f_{\rm duty} = 2.5\%$ has $(\Delta T, T_{\rm on}) = (200, 10)$ Myr. All three models have $f_{\rm esc}$ scaled as in the previous section so that $z_{\rm re} = 6.25$ for best agreement with the data. The ionizing background dominates the completion of reionization, as can be seen from the sharp decrease in $1 - \langle x_e \rangle$ around $z \sim 8$. The ionizing background is proportional to $f_{\rm esc}$, so that all three models have similar matched escape fractions (as in the fixed $z_{\rm re}$ plots of Figures 2.5 and 2.6). The primary effect of decreasing the duty cycle is a boost in $\langle x_e \rangle$ at high redshifts, resulting in an increase in τ_e, which we note in the figure. In the right plot of Figure 2.9 we plot the derived ionizing emissivity $\dot{n}_{\rm ion}^{\rm com}(z)$ for the same three models as a function of Hubble time against the observational constraints on this quantity shown in Table 2.1b. All three models have $f_{\rm esc}$ scaled as in the previous section so that $\tau_e = 0.066$ to emphasize the effect of altering the duty cycle on τ_e. The three models thus represent three points shown in Figure 2.5; the escape fractions of the continuous, $f_{\rm duty} = 10\%$, and $f_{\rm duty} = 2.5\%$ models are $f_{\rm esc} = 11.5\%$, $f_{\rm esc} = 6.5\%$, and $f_{\rm esc} = 3.9\%$, respectively. The short-dashed line shows $n_b(z)/t_H(z)$, the average number of baryons in the universe per Hubble time. The point at which the solid lines in each model crosses the short-dashed line gives a rough estimate of the lower bound for the redshift of reionization. The universe should be fully ionized once the average rate of ionizing photon production is a few times greater than n_b/t_H, which is in good agreement with our findings from the previous section. We see that continuous models must have an evolving $f_{\rm esc}$, higher at high-z and decreasing toward lower redshifts in order to agree with the data points at $z < 4$. An attractive feature of the bursty models is that a $f_{\rm esc}$ nearly constant with redshift agrees with observational data, because of the lower values of $f_{\rm esc}$ at $z > 6$ needed to to produce $\tau_e = 0.066$ and reionize by $z = 6$.

2.3.3 Transition to continuous star formation due to overlap of H II egions and massive galaxies

A limitation of our simulations is the inability to accurately account for the overlap of nearby H II regions and source clustering. At early redshifts, such overlap will be negligible, while near the completion of reionization several individual H II regions will overlap into larger ionized regions containing many sources, and the ionizing background will start dominating. We may estimate the redshift at which overlap of H II regions becomes important by determining when the volume filling fraction $Q(x_e')$ of gas with electron fraction $x_e < x_e'$ becomes larger than a given threshold. We examine the outputs from our simulations for different duty cycles and burst periods to find an approximate redshift at which ionized bubble overlap becomes important. In the left plot of Figure 2.10, the solid lines show the redshift at which the volume filling fraction of gas with $x_e \leq 0.5$ reaches $Q = 10\%$ in the $N_{HI} = 0$ simulations constrained to reionize at $z_{\rm re} = 6.25$ (i.e. the same outputs presented in the top plots of Figure 2.5). We take a conservative $Q = 10\%$ as the threshold at which overlap may begin to be important, and thus find an indicative redshift at which star formation begins to gradually transition to a more continuous mode of star formation. We see that only for the most bursty models the overlap becomes relevant before reionization is complete. In the same figure we also address the effect of source clustering in overdense regions, neglected in our simulations. The growth of halos within overdense regions is accelerated, so that the overlap becomes relevant at higher redshifts. Roughly, in a region with an overdensity $1 + \delta$, the redshift of overlap z' is higher by a factor $1 + z' = (1 + \delta)(1 + z)$. The dashed lines in left plot of Figure 2.10 are the same as the solid lines but for an overdense region of the universe with $\delta = 0.2$.

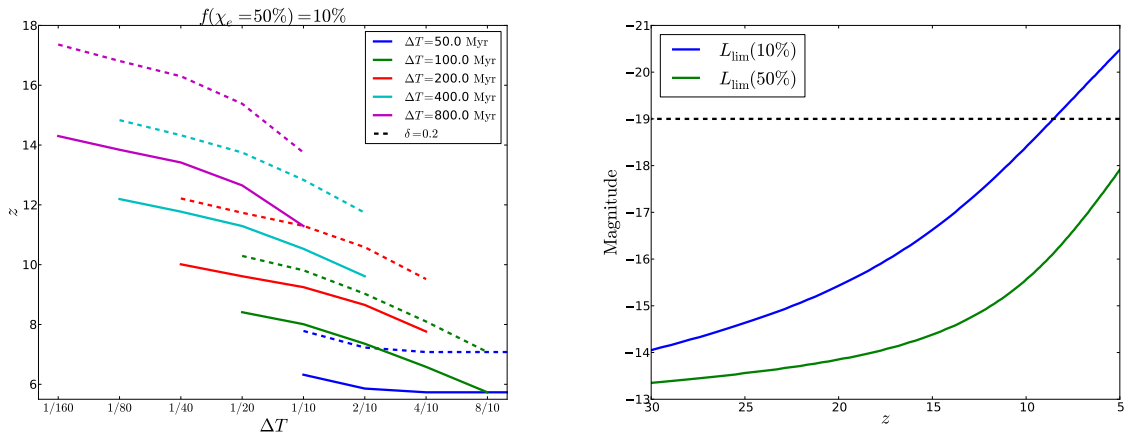

Figure 2.10: (*Left.*) Redshift at which the volume filling fraction of gas with $x_e \leq 50\%$ reaches $Q = 10\%$ for the same models as in Fig. 2.5. We interpret this as the redshift at which overlap of H II regions begins to be relevant and star formation transitions to a more continuous mode. We see that for all but the most bursty scenarios, the time between the beginning of general overlap and the domination of the ionizing background is relatively short. The dashed lines represent the same results in a region with an overdensity $1 + \delta = 1.2$. (*Right.*) Limiting UV magnitude above which 10% (blue line) and 50% (green line) of the total ionizing photon budget is emitted. If we take $M_{UV} = -19$ (dashed line) as the luminosity above which a halo will always produce stars continuously, we see that such large halos do not dominate the total luminosity until very late.

Finally, we note that massive halos ($M > 10^{11} M_\odot$) host bright galaxies with many star forming regions, and satellite sub-halos forming stars independently. Even though star clusters may form in short bursts, the overall effect is that the halo behaves as though it was forming stars continuously. We assume that this effect becomes dominant for halos with stellar masses greater than $10^6\ M_\odot$, which corresponds to a UV magnitude of roughly $M_{UV} \sim -19$. In the left plot of Figure 2.10, we plot the UV magnitude M_{UV} of galaxies at which brighter galaxies contribute 10% (blue line) and 50% (green line) of the total ionizing photon budget. We see that halos which are too massive to be "bursty" don't dominate the total ionizing luminosity until close to the completion of reionization, so that we believe that galaxies dominating the reionization process are bursty nearly until reionization.

2.4 Summary and Discussion

The model and analysis presented in this paper builds on the idea that bursty star formation in the early universe may increase significantly the average electron fraction of the universe when compared to continuous star formation, thus alleviating possible tensions between the observed high redshift UV luminosity functions and measurements of τ_e. Stellar populations which form in bursts are luminous in the UV for about 5 Myr and mostly dark afterwards. Constraining models using high redshift UV luminosity functions and the halo matching method assuming bursty star formation, we found that the Strömgren spheres which formed around halos of a fixed mass are larger than assuming continuous star formation, but begin to recombine once star formation ends. The recombination rate of ions in these halos is proportional to the square of the electron fraction, so that relatively long lived relic regions of ionization will be left behind in the IGM. Our goal was to build models to test whether this effect would have a non-negligible contribution to the average electron fraction of the universe, thus altering predictions of the volume filling fraction of H II regions, $Q_{\rm H\,II}(z)$, the Thompson optical depth τ_e, and the redshift of reionization $z_{\rm re}$, all of which are constrained to varying accuracy by observation.

In Section 2.2 and in the Appendix, we presented the details of the model we used to compare ionization histories in cosmic volumes containing continuous and bursty star formation. The semi-analytic model tracks the evolution of stellar populations within statistically generated dark matter halos uniformly distributed throughout a volume and constrained to reproduce observed and extrapolated UV luminosity functions at high redshift. The electron fraction as a function of time around a given halo hosting bursty or continuous star formation is derived from writing down and solving the basic physics equations calibrated using one-dimensional radiative transfer

simulations presented in Ricotti et al. (2001). The analytic equations we derived in the continuous and bursty cases are used to calculated volume filling fractions of partially ionized IGM gas for a uniformly sampled array of electron fractions. To calculate the cosmic ionization history we also include the contribution of the ionizing background. These quantities allowed us to derived the average electron fraction $\langle x_e \rangle$ throughout the volume as a function of time, from which we were able to calculate the relevant observable quantities.

We presented an in-depth analysis of the behavior of the our model and its predictions in Section 2.3. Our model allows us to freely choose the length of the starburst, $T_{\rm on}$, the period of repeated bursts, ΔT, the escape fraction of ionizing photons, $f_{\rm esc}$, the hardness of the ionizing spectrum, the lower limit of halo luminosity function, $M_{\rm UV,lim}$, and the lower limit of dark matter halo mass, $m_{\rm dm}$. We took $f_{\rm esc}$ as an adjustable parameter to constrain either τ_e or $z_{\rm re}$ keeping fixed each time the other parameters, but exploring a range of possible parameter combinations. The main results of this study are the following:

- By performing halo matching with the assumption of a duty cycle of star formation ($f_{\rm duty} = T_{\rm on}/\Delta T$), we find that the dark matter halos hosting galaxies of a given magnitude in the HST deep fields are less massive than they would be if we assume that stars formed continuously.

- We speculate that the inferred smaller masses of these halos and the short duration of the luminosity burst should result in a larger fraction of the ionizing radiation produced by the stars escaping into the IGM.

- For a fixed halo mass, stellar populations forming in the bursty mode are more luminous and produce larger H II regions than in models where star formation is continuous. The

relatively long-lived relic regions of partial ionization left behind by these bursts are able to maintain partial ionization throughout much of the IGM in a manner similar to X-ray pre-ionization, but heating the IGM less efficiently than X-rays. The overall effect is an increase in $\langle x_e \rangle$ at high redshifts, resulting in an increase in τ_e and a slightly higher redshift of reionization $z_{\rm re}$.

- By constraining the galaxy UV luminosity density in our simulation, we find that to produce the $\tau_e = 0.066$ observed by Planck and complete reionization by redshift $z_{\rm re} \sim 6.0$, models with bursty star formation require an $f_{\rm esc} \sim 2\% - 10\%$ that is $2 - 10$ times lower than in continuous star formation models ($f_{\rm esc} \sim 17\% - 20\%$).

- The ionizing photon budget needed to reproduce the observed τ_e depends strongly on the duty cycle of star formation but also the temperature of the partially ionized IGM that affects the life time of relic H II regions. Thus even a relative low intensity of X-ray background radiation sufficient to increase the temperature of the IGM but not its ionization fraction may have an important indirect effect on the life span of relic H II regions and therefore τ_e.

- The hardness of the ionizing spectrum and radiation background instead affects $z_{\rm re}$. The hardness of the spectrum for stellar sources is determined mainly by the dependence of $f_{\rm esc}\ (\nu)$ on the frequency. We find that increasing the column density $\langle N_{\rm H\,I} \rangle$ of neutral gas within the halo seen on average by the sources, produce a higher mean energy of escaping photons. A more energetic photon spectrum has a greater mean free path in the IGM, allowing the ionizing background to become dominant at earlier times and complete the reionization process earlier.

- Our results suggest that any shortcoming of the ionizing photon budget suggested from extrapolated observations of the UV luminosity functions at high redshift and low values of $f_{\rm esc} \sim 5\%$ typically measured in local starburst galaxies (*e.g.*, Hurwitz et al., 1997; Boutsia et al., 2011; Nestor et al., 2013), would be alleviated if reionization was driven by short bursts of star formation, perhaps relating to the formation of Population III stars and compact star clusters such as proto-globular clusters, as suggested by previous studies Katz & Ricotti (2013, 2014). A non-evolving $f_{\rm esc}\ (z) = 5\%$ in our bursty model is consistent not only with local observations of $f_{\rm esc}$, τ_e from Planck, redshift of reionization $z_{\rm re} \sim 6$ from quasars, but also with the ionizing photon emissivity between redshifts $z \sim 2$ and 6 inferred from observation of the Lyman-alpha forest.

The present study is only the first step to asses the effects of a bursty mode of star formation on the reionization history and on the the properties of halos hosting high-redshift galaxies. Our model is fundamentally very simple and has several limitations but also some advantages and important improvements even with respect to full 3D radiative transfer simulations.

3D radiative transfer in cosmological simulations is basically monochromatic: because of computational limitations one can only afford to consider one frequency band for H I ionizing radiation (sometimes also He I , He II ionizing bands are considered, *e.g.*. Gnedin, 2000a, 2014). This may lead to an underestimate of the width of cosmological ionization fronts and therefore miss large volumes of partial ionization around H II regions, in addition to the relic H II regions left behind in a bursty mode of SF. Our analytic model is instead calibrated to reproduce 1D radiation transfer simulations in which the radiation field is sampled with more than 400 logarithmically spaced frequency bins. We are therefore able to capture the true width of cosmological H II regions that

is significant in the low density IGM (see Figure 1(left) of the Appendix: the radius of the H II region defined where $x_e = 90\%$ is 0.2 Mpc, while the radius where x_e drops to 20% is 1 Mpc; that means that the region of partial ionization around H II regions extends at least 5 times further than the Strömgren radius and the volume is 125 times larger). This may explain why our continuous SF model requires a slightly lower $f_{\rm esc}$ to produce $\tau_e = 0.066$ than that found in other semi-analytic models that do not take into account regions of partial ionization.

The main limitations of our treatment are: i) we neglect clustering of sources and therefore the topology of reionization and size distribution of H II regions Furlanetto et al. (2004). Realistically, the redshift of reionization will have a variance along different lines of sight. ii) Our treatment of the radiation background is an approximation. This is because without full radiative transfer calculation and clustering of the sources it is difficult to precisely account for the fractional contribution of local sources vs background sources (in practice because of this somewhat artificial separation, photons in the background can be slightly underestimated or overestimated). The background calculation is important in determining the redshift of reionization $z_{\rm re}$ but affects the value of τ_e only weakly. This is because the electron fraction produced by individual sources of reionization neglecting the background reaches values of $\langle x_e \rangle \sim 80\% - 90\%$ while the background completes reionization bringing $\langle x_e \rangle$ to unity with a sharp rise similar to a phase transition Gnedin (2000a). Therefore in our model the absolute value of $z_{\rm re}$ may not be accurate. However, the trend of $z_{\rm re}$ as a function of the various free parameters in the model are robust.

Theoretical estimates of $f_{\rm esc}$ are very uncertain (*e.g.*, Ricotti & Shull, 2000; Gnedin, 2008; Gnedin et al., 2008; Wise & Cen, 2009; Yajima et al., 2011, 2014) and observations are only possible in a limited number of cases, particularly local starburst galaxies and Lyman brake galaxies at $z \sim 3$ that in most cases set upper limits $f_{\rm esc} \lesssim 4\% - 8\%$ (*e.g.*, Hurwitz et al., 1997; Steidel et al., 2001;

Fernández-Soto et al., 2003; Inoue et al., 2006; Shapley et al., 2006; Siana et al., 2007; Iwata et al., 2009; Vanzella et al., 2010; Boutsia et al., 2011; Nestor et al., 2013). Our study found that values of $f_{\rm esc} \sim 5\%$ are consistent with the observed values of τ_e and $z_{\rm re}$ and with direct or indirect measurements at $z = 0$ and $z \sim 1.3 - 6$, even assuming truncation of the faint end of the high-z luminosity functions at $M_{UV,lim} = -15$ or truncation of the dark matter mass function at 10^9 $M_\odot$. We thus believe that bursty star formation, that is likely to be prevalent in the early universe, is an important physical process neglected or overlooked in previous works on reionization. Once this effect is included, the need for missing sources of reionization suggested in past works is either eliminated or greatly reduced. This different result follows from taking particular care in modeling partially ionized gas in front of H II regions and in relic H II regions, thus reducing the number of ionizing photons used up by hydrogen recombinations (therefore increasing the effective $\bar{t}_{\rm rec}$ in Equation (2.1)). We therefore caution of simply using Equation (2.1) to constrain the ionizing emissivity and $f_{\rm esc}$ from the observed τ_e. We also point out that the redshift of reionization $z_{\rm re}$ cannot be used to constrain the ionizing emissivity at $z \sim 6 - 8$ because it is determined by the radiation background emitted by sources that might be at higher redshift and not yet observable, and its value is sensitive to the hardness of the sources spectra.

Chapter 3: ARC: Adaptive Ray-tracing with CUDA, a New Ray Tracing Code for Parallel GPUs

We present the methodology of a photon-conserving, spatially-adaptive, ray-tracing radiative transfer algorithm, designed to run on multiple parallel Graphic Processing Units (GPUs). Each GPU has thousands computing cores, making them ideally suited to the task of tracing independent rays. This ray-tracing implementation has speed competitive with approximate momentum methods, even with thousands of ionization sources, without sacrificing accuracy and resolution. Here, we validate our implementation with the selection of tests presented in the "cosmological radiative transfer codes comparison project," to demonstrate the correct behavior of the code. We also present a selection of benchmarks to demonstrate the performance and computational scaling of the code. As expected, our method scales linearly with the number of sources and with the square of the dimension of the 3D computational grid. Our current implementation is scalable to an arbitrary number of nodes possessing GPUs, but is limited to a uniform resolution 3D grid. Cosmological simulations of reionization with tens of thousands of radiation sources and intergalactic volumes sampled with 1024^3 grid points take about 30 days on 64 GPUs to reach complete reionization.

3.1 Introduction

The propagation of ionizing and dissociating radiation from stars and black holes and its effect on the interstellar medium (ISM) and the intergalactic medium (IGM), is one of the most fundamental and computationally difficult problems in theoretical astrophysics. A number of schemes have been implemented for tackling radiative transfer. As the computational power available for astrophysical simulations has increased over the past few decades, the full seven dimensional (three spatial, two angular, one frequency, one time) radiative transfer problem has been solved in earnest. There are multiple popular methods for approaching this problem, including:

1. Moment Methods: The first three moments of the radiation intensity, which are the energy density, flux, and radiation pressure, are tracked by the simulation (Auer & Mihalas, 1970; Norman et al., 1998; Stone et al., 1992). Simulations of this type have been implemented with short characteristics (Stone et al., 1992), long characteristic (Finlator et al., 2009), utilizing the optically thin variable Eddington Tensor method (Gnedin & Abel, 2001; Ricotti et al., 2002b,d; Petkova & Springel, 2009), and with a two moment model using a closure relation (González et al., 2007; Aubert & Teyssier, 2008). These methods have the advantage of being fast, with computation times that don't depend on the number of radiation sources. However, these methods are fundamentally diffusive, so that radiation will in some cases flow around occluding regions in non-physical ways, resulting in incorrect behavior for shadows. In addition, since the radiation is treated as a photon fluid with characteristic speed equal to the speed of light, the condition for stability of the integration

typically requires taking very small time step (similarly to the CFL condition). A reduced speed of light approximation is often adopted in order to be able to take larger timesteps (Deparis et al., 2018).

2. Ray-tracing Methods: Radiation from a point source is approximated as a collection of linear rays propagating away from the point source. Along the rays the heating/ionization of the gas and extinction of the radiation are tracked through a grid (Abel et al., 1999; Razoumov & Scott, 1999; Ciardi et al., 2001; Sokasian et al., 2001; Alvarez et al., 2006; Mellema et al., 2006; Rijkhorst et al., 2006; Whalen & Norman, 2006; Krumholz et al., 2007; Trac & Cen, 2007; Paardekooper et al., 2010) or collection of particles (Susa, 2006; Johnson et al., 2007; Altay et al., 2008; Pawlik & Schaye, 2008, 2011; Hasegawa et al., 2009). These methods are computationally more expensive, as the amount of computation scales linearly with the number of sources and the number of grid points or gas particles. The results, however, are a better approximation of the true solution.

3. Non-simulation methods, such as the excursion set formalism (inside-out reionization, see Furlanetto et al., 2004; Alvarez & Abel, 2007), inhomogeneous reionization models (outside-in reionization, see Miralda-Escudé et al., 2000; Wyithe & Loeb, 2003; Choudhury & Ferrara, 2005), and semi-analytic inhomogeneous reionization methods (Mitra et al., 2015). These methods require much fewer computational resources than simulation methods, allowing for the exploration of a much larger range of simulation parameters. However, they lack the accuracy of computational methods.

Our approach is of the second type, based on the photon-conserving, spatially-adaptive ray-tracing method originally presented in Abel & Wandelt (2002), but designed for use on parallel

Graphics Processing Units (GPUs). GPUs have slower clock speeds than CPUs of the same generation, but they possess thousands of parallel cores which operate independently of each other. The cost of calculations for each individual ray is computationally small and independent of other rays, which makes GPUs ideally suited for the task. Despite the large ongoing effort to model the epoch of reionization (EOR) by various teams, our proposed approach is complementary to previous and ongoing investigations, and has new and unexplored aspects to it as explained below. We are writing our numerical code with the future goal of incorporating it in widely used hydrodynamical cosmological codes. At the moment, the focus of our effort is on ray-tracing radiative transfer, meaning we limit ourselves to simulations with fixed grid (*i.e.*, the adaptive mesh refinement (AMR) grid structure is not implemented at this point) and no coupling to hydrodynamics, thus making this code a purely post-processing method on previously run cosmological simulations. This approximation is not poor when focusing on simulations of cosmological reionization on scales much larger than galaxy-halo scales, which are the focus of our test runs and future science investigations.

This paper is organized as follows. In Section 3.2 we describe the underlying physics and computational implementation used by our code. In Section 3.3 we present the results produced by our code when running the battery of tests laid out in the radiative transfer codes comparison project (Iliev et al., 2006a). We also present a selection of benchmarks demonstrating the speed of the code in Section 3.4. Finally, in Section 3.5 we summarize the methods and results presented in this paper.

3.2 Methodology

Our code utilizes the power of GPU processing to attack the computationally intensive problem of adaptive ray-tracing radiative transfer. The tracing of a single ray of photons is a very simple task, but a ray tracing radiative transfer simulation requires the tracking of a large number of such rays, making the task well suited to GPU computing. Classically, memory sharing between GPUs is only possible if those GPUs are all located on the same node of a computer. This limits the scale of a purely GPU algorithm, and lead us to add MPI parallelization. MPI parallelization allows us to share the results of GPU calculations between GPUs on different nodes, thus allowing for a highly scalable code and also overcoming the problems associated with the limited memory of a single GPU. In addition to the relatively straightforward parallelization of radiation from different sources, we break up the computational volume and therefore the ray tracing and the ionization/heating calculations into equal portions and distribute them among each node of the process using MPI. There, the GPU algorithm performs the calculations, sharing ray data between nodes using MPI as necessary. The results of the calculations are then consolidated using MPI and the next time step can be taken. An earlier and simpler version of the code shared copies of the whole the grid data among nodes and distributed only the work from different ionization sources using MPI. This way, there was no need for ray data sharing among nodes. However, this method was severely limited by the available memory of each GPU. With current GPUs, the maximum resolution of reionization simulations we could run without breaking up the volume was between 256^3 and 512^3 grid points. The current code is indeed faster and not limited by these memory restrictions.

3.2.1 Program Design

Each GPU in our code calculates the radiation field produced by sources within the sub-volume assigned to the node which hosts that GPU. In addition each sun-volume can share the work with multiple GPUs. The direction of the rays is assigned using the HEALPix (Górski et al., 2005) scheme to distribute rays in all directions. The HEALPix scheme assigns to each ray equal solid angle by dividing the sphere around each sources in equal areas. In order to maintain constant spatial resolution as the rays travel further from the source, each subsequent level of the scheme breaks this area into four sub-areas, allowing the algorithm to easily keep track of the splitting of rays. We initialize an array of photon packets for each ray in the initial HEALPix array (typically HEALPix level 1, or 48 photon packets, or rays). Each photon packet traced by our code contains the ID of the ionizing source which emitted it, the unique HEALPix PID which determines the direction of the photon's motion, the distance each photon has traveled, and the optical depth of each frequency bin along the photons path. The position of the ionizing source is stored in the GPU's shared memory, allowing for a reduction of the required per-photon packet memory. Once the calculation of the radiation fields by the GPUs is complete, we combine the radiation fields from other sources distributed on different nodes linearly using MPI. We then divide and distribute all these fields to the processes to evolve the gas fields over a single time step. The algorithm's overall structure is described in steps below.

1. Initialize radiation field arrays to zero at the start of a new time step.

2. Use MPI to divide distribute photon source data and source array to all nodes of the process.

3. Loop through each photon source one at a time using the GPU kernel as follows:

(a) Initialize all photon packets to radius zero and zero optical depth in all frequency bins.

(b) Trace all photon packets at a single HEALPix level until the ray moves to an adjacent sub-volume, terminates or splits. If the photon moves into an adjacent sub-volume, we store it in a sharing buffer to be sent later. If the photon terminates, set the radius of the packet to zero, signaling that it should be ignored in future calculations. Otherwise, leave the radius and optical depth at the point of splitting unchanged.

(c) Create a new array of child rays populated by the unterminated rays from the previous step. The radius of all four child rays is the same as the parent rays, but the newly populated array have directions based on the next level of the HEALPix scheme.

(d) If any rays are unterminated, return to step (b).

(e) Use MPI to share rays between adjacent sub-volumes, and return to step (a).

4. Use MPI to linearly combine radiation fields produced by all of the kernels, thus giving us the overall radiation field produced by all the sources.

5. Apply the ionization/heating calculation to the current radiation field and gas fields to calculate the changes to the grid over a single time step.

6. Use the execution times of each GPU to redistribute the sources between GPUs handling the same sub-volume to balance the computational load.

7. Return to (i) until the simulation reaches the desired time.

Each node and corresponding GPU loops through this code to advance the simulation. We can also send the same sub-volume to multiple GPU's, allowing us to break a set of sources in the same sub-volume between multiple GPUs. When any timing information about the entire

simulation, or outputs of the full grid, are required, MPI is used to consolidate the data from all of the nodes.

In the next subsections we describe in more detail the implementation of the physics and equations solved by each module of the code.

3.2.2 Photon Data

Photon packets in our code are polychromatic. As each ray is traced through the grid, we calculate the total optical depth due to all atomic species (hydrogen and helium) and their ions, for each frequency bin. Each bin is a single frequency which represents the frequency averaged ionization cross section of all present species. We are able to track an arbitrary number of frequency bins, subject to memory limitations.

3.2.3 Grid Data

The basis of our code is a uniform 3-dimensional cubic grid of N^3 cells which store the neutral/ionized fraction, temperature, and photoionization rate of a given set of atomic/molecular species as a function of time. The current version of our simulation tracks the neutral fractions $x_{\mathrm{H\,I}}$ and $x_{\mathrm{He\,I}}$ (future versions will include tracking of more species as required). We assume that helium is either neutral or singly ionized over the course of the simulation. Thus, we calculate the electron number density as:

$$n_e = n_b \left[\frac{1 - Y_p}{A_{\mathrm{H}}} x_{\mathrm{H\,II}} + \frac{Y_p}{A_{\mathrm{He}}} (x_{\mathrm{He\,II}} + 2x_{\mathrm{He\,III}}) \right],$$

where $n_b \equiv n_p + n_n$ is the baryon number density, $Y_p \equiv \rho_{He}/\rho_b$ is the Helium fraction, $A_H = 1$ and $A_{\rm He} = 4$ are the atomic weights of hydrogen and helium, respectively.

We are able to sub-divide this volume as desired. The first science runs of our code will divide the volume evenly into 8 cubic regions, each of which will be handled by a subset of the available GPUs. The current implementation of the code is limited to a fixed grid; however, the structure of the code would allow for a transition into an adaptive mesh refinement scheme (AMR) should it be required in the future.

3.2.4 Photon Transmission

Ionizing sources within our volume are represented by a set of photon packets, or rays, which are initialized at the location of the source. These rays are traced through the grid of our simulation one step at a time, with each step corresponding to the distance through the volume it takes to reach the next X, Y, or Z cell boundary within the volume. Each ray is initialized with an ionizing photon production rate (S_ν [s^{-1}] for each frequency bin ν) which corresponds to the luminosity of the source divided between the rays. For each step, we increment the total optical depth along the ray τ_ν for each frequency bin by $\Delta\tau_\nu$, the optical depth of ray within a single simulation cell. We use this optical depth to calculate the fraction of ray's photons absorbed by gas within the cell, and thus calculate the ionization rate of the cell for each frequency bin ν per absorbing atom:

$$s_\nu = \frac{S_\nu(1 - e^{-\Delta\tau_\nu})}{V\ n_{\rm abs}}, \tag{3.1}$$

where $n_{\rm abs}$ is the physical number density of absorbers, and V is the volume of the simulation cell. More specifically, in our code $n_{\rm abs}$ is the neutral number density n_{SI} of a species S, where S refers to hydrogen and helium. The flux along the ray in each frequency bin is reduced by this amount, so that this method conserves the total number of photons. Each ray in the simulation is traced by a single CUDA core, with the work distributed as evenly as possible by the CUDA kernel. Each GPU contains all of the grid data necessary to trace the rays, and the ionization rate at a given cell in the grid is altered "atomically" by each core, eliminating the chance for interference between simultaneous processes.

The directions of the rays are chosen according to the HEALPix scheme (Górski et al., 2005), which assures that each ray is assigned an equal solid angle relative to the source. This means that we can assign a photon emission rate for each ray by dividing the overall $f_{{\rm esc},\nu}S_\nu$ of the source evenly between all rays (where $f_{{\rm esc},\nu}$ is the mean ionizing radiation escape fraction from each source), or we can adopt a more sophisticated (and realistic) scheme in which the escape fraction is anisotropic. We can, for instance, obtain the wanted escape fraction by only assigning the mean S_ν per-ray, to a fraction $f_{{\rm esc},\nu}$ of the rays, while completely blocking the radiation escaping from the remaining rays.

Each ray is traced independently by a CUDA core until it reaches a set distance from the source, at which point the ray is split into four child rays corresponding to the next level of the HEALPix scheme. The distance at which the ray is split is a free parameter which allows us to control how many rays pass through the average simulation cell at a given distance, giving us control over the spacial accuracy of the ray tracing method. In Figure 3.1 we plot a schematic of the adaptive ray tracing process. The two trees represent the progress of a single pixel at the lowest HEALPix level traced out through three branches. The two trees demonstrate the

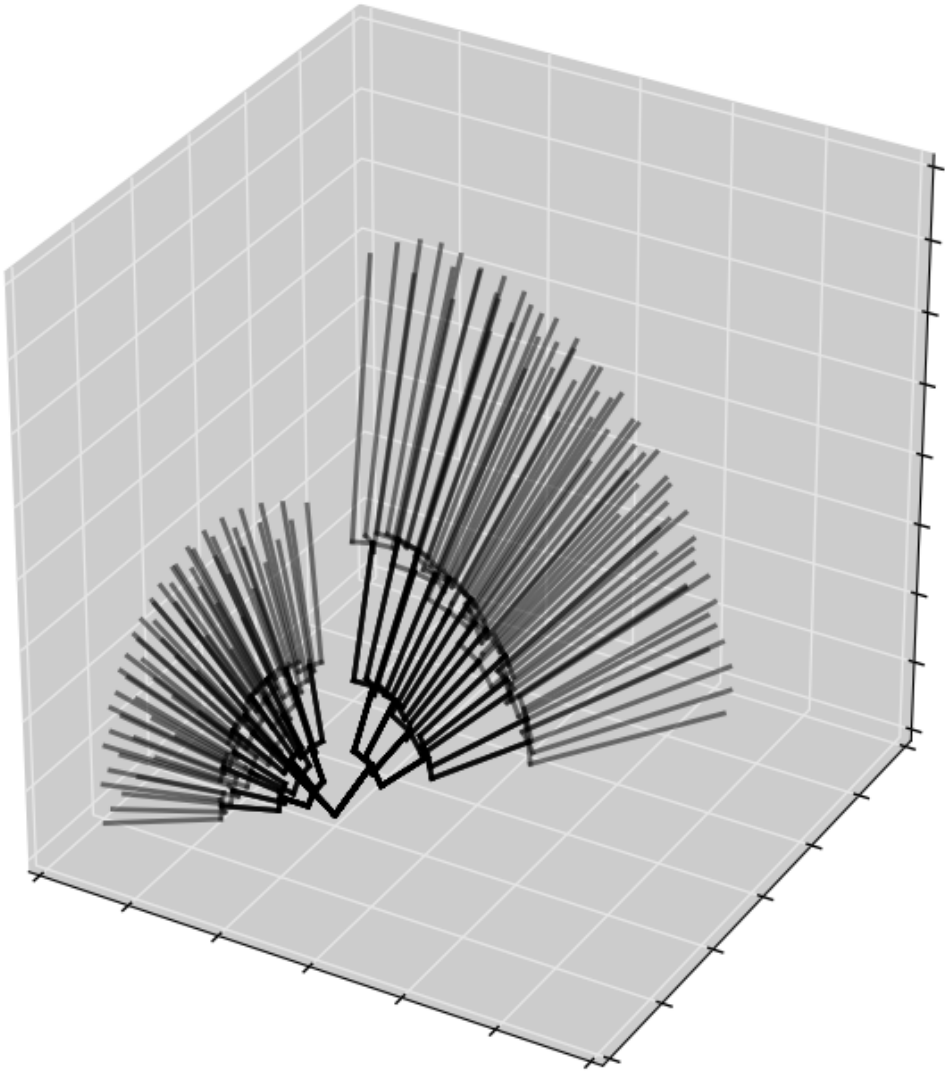

Figure 3.1: Schematic example of the adaptive ray tracing method based on HELPix. As rays propagate outwards, they split to maintain a constant number of rays per unit area crossing the surface of the sphere centered on the source. The left and right trees illustrate the difference between assuming at least 3.0 and 1.0 rays intersect a given simulation cell for the left and right ray trees, respectively.

difference between thresholds for ray splitting, with the left tree guaranteeing three times as many pixels through a given cell than the right tree.

It is important to emphasize that, as mentioned above, HEALPix method gives us the ability

to control to any desired accuracy the directions into which the ionizing source emits photons, allowing simulations in which ionizing photons leak out of a galaxy anisotropically. Numerical simulations of galaxy formation have shown that "chimneys" of low optical depth through which most ionizing photons escape into the intergalactic voids are a more realistic description of the radiation escaping into the IGM than an overall isotropic attenuation of the emission. We can control these directions on a source by source basis, giving us a versatile tool to study different types of sources in different cosmic environments. This desirable feature of the code is only possible because the ray tracing method is non-diffusive and can capture shadows accurately. The faster "radiation moments" methods discussed in the introduction, are probably too diffusive to allow the implementation of anisotropic emission from the sources.

3.2.5 Geometric Correction

The rays in our code are one-dimensional lines which represent a cone of radiation extending from a point source, or from a splitting ray, with a fixed solid angle. The intersection of the ray volume with a given grid cell is a complicated geometric shape whose volume is impratical to calculate exactly. The ray tracing method approximates this volume as a truncated cone created by the segment of the ray whose length is the distance between the points where the ray enters and exits the grid cell. Thus, when the ray remains closer to the edge of the grid cell than the radius of the ray (along the length of the ray), a portion of the ray remains entirely outside of the grid cell. We make a first order correction to this problem following the method presented in Wise & Abel (2011). We let $L_{\rm Pix}$ be the width of the pixel and let $D_{\rm edge}$ be the distance from the midpoint of the segment of the ray within the cell to the nearest edge of the cell. We reduce the

ionization within the cell by a factor f_c defined by:

$$f_c = \begin{cases} \frac{1}{2} + \frac{D_{\rm edge}}{L_{\rm Pix}} & D_{\rm edge} < L_{\rm Pix}/2, \\ 1 & D_{\rm edge} \geq L_{\rm Pix}/2. \end{cases} \tag{3.2}$$

This correction factor is slightly different from the one used in Wise & Abel (2011) in which $(1/2 + D_{\rm edge}/L_{\rm Pix})$ is squared. We remove the square from the first expression, so that $f_c = 1/2$ when $D_{\rm edge} = 0$, which physically corresponds to half of the ray lying within the grid cell when the ray travels along the edge of the cell. However, this correction factor only serves to reduce the ionization rate in a given cell, and thus has the effect of reducing the global ionization rate in the simulation and breaks photon conservation, that is one of the most desirable properties of the method. We find that adopting this correction makes a non-negligible reduction to the volume filling fraction in the simulation, which is more pronounced at the points where the rays split (as may be seen in the radial profiles of Wise & Abel (2011) Figure 5 and Figure 6).

We decided to keep the correction factor as it mitigates small spatial artifacts in azimuthal directions (see Section 3.3.6), but we also compensate in order to maintain photon conservation by adding the ionization removed by the correction factor to the nearest cell using a secondary correction factor:

$$f_c' = (1 - f_c)\frac{x_{\rm abs}}{x_{\rm abs}'}, \tag{3.3}$$

where $x_{\rm abs}$ and $x_{\rm abs}'$ are the absorber fraction in the cell intersected by the ray, and in the nearest adjacent cell to the ray, respectively. Here we have multiplied by the ratio of the density of

absorbers between the cells, as the ionization rate in Equation 3.1 is an overall absorption rate, and needs to be corrected in case the the densities of absorbers vary between the cells. This second correction factor is another slight difference with respect to the method used in Wise & Abel (2011), but we find it very beneficial. The combination of these correction factors corrects for geometric artifacts while maintaining photon conservation and the correct volume filling factor of the H II regions (see Section 3.3.6). This secondary correction relies on the regularity of the Cartesian grid, and in the case of non-regular grids or AMR it would need to be generalized or removed.

3.2.6 Optically Thin Approximation

In regions of low neutral fraction surrounding active sources of ionization, the ray tracing process tracks an optical depth which is almost unchanging over the grid cells. This allows us to implement a procedure for calculating the ionization within a certain volume without tracing rays, and thus start the ray tracing process at a greater distance and save significant computation time. Our optically thin approximation proceeds as follows:

1. To each ionizing source within the simulation we assign a radius $R_{\tau<0.1}$, which we initially set to zero.

2. As rays are traced from the ionizing source, we set $R_{\tau<0.1}$ to be the minimum of all the ray lengths for which $\tau_\nu < 0.1$ in the softest frequency band.

3. When $R_{\tau<0.1} > 0$, we assume all cells within a distance of $R_{\tau<0.1}$ of the ionizing source to be directly exposed to the ionizing source without absorption. We approximate the

ionization rate as:

$$s_\nu \approx \frac{S_\nu x_{\rm abs} \sigma_{\rm abs}(\nu)}{4\pi R^2},$$

which is the approximation of Equation 3.1 for small τ_ν and replacing the dilution of spreading rays with a $4\pi R^2$ term.

4. We initialize rays at a radius of $R_{\tau>0.1}$. This allows us to skip ray tracing within the optically thin region.

This approximation is trivial at early times in simulation, when $R_{\tau>0.1}$ is small for most sources, but the amount of calculation is then also small. However, when the average neutral fraction of the simulation decreases, this approximation becomes more and more efficient, saving significant computational resources. Beginning the rays outside of these these highly ionized regions decreases the ray tracing computation time by a factor of the order of the mean neutral fraction in the simulation box, while the calculation itself is orders of magnitude faster than the ray tracing module.

3.2.7 Ionization Calculation

Each GPU in our code loops through all sources it is assigned using the procedure detailed in the previous section. Once all the GPUs have completed their assigned work, they return the radiation array which represents the ionization rate due to all of the sources assigned to that GPU. The host node then sums these arrays using MPI, giving us the overall ionization rate at each point in the grid due to all sources in the simulation.

Once the ionization rate, Γ_ν, is calculated, we compute the change in the neutral density, n_{SI} of species S of the gas according to the differential equation:

$$\dot{n}_{SI} = -n_{SI} \sum_{\nu} s_\nu - C_S(T) n_e n_{SI} + \alpha_S(T) n_e n_{SII}, \tag{3.4}$$

where C_S is the collisional ionization rate, $\alpha_S(T)$ is the recombination rate, n_e is the electron number density and n_{SII} is the ionized number density of species S.

The ray tracing algorithm is computationally the most expensive part of the simulation, so we have designed the code to limit the number of calls to the ray tracer as much as possible. We solve the non-equilibrium chemistry and energy equations for all species under consideration sub-cycling between the radiative transfer time steps (Anninos et al., 1997). The non-equilibirum chemistry equations are stiff ordinary differential equations (ODEs). Thus, for individual steps in the sub-cycle, we tested several integration methods, including predictor-corrector methods, Runga-Kutta methods, semi-implicit methods, backwards difference, etc. We found that the backwards difference formula (BDF) gave the best combination of speed, accuracy, and computational stability in our tests, in agreement with the results presented in (Anninos et al., 1997). Effective use of the BDF requires writing the non-equilibrium chemistry equations in the form:

$$\dot{n}_{SI} = D - C \cdot n_{SI}, \tag{3.5}$$

where D represents source terms which do not depend on n_{SI} (to the first order) and C represents sink terms which are linear in n_{SI} (to the first order). The photon conserving method we use (Equation 3.1) calculates the photo ionization rate per absorbing atom within a cell at a given

neutral fraction. In order to use Equation 3.5, we must assume that the photo-ionization (s_ν) and heating per absorber within the cell, which depend on $(1 - e^{-\Delta\tau})/n_{SI}$, is either independent of the neutral fraction or is $\propto 1/n_{SI}$, so that $n_{SI}s_\nu \sim \text{const}$. Since $\Delta\tau$ is proportional to $n_{\rm SI}$, we are effectively treating single cells as having either $\Delta\tau \ll 1$ (*i.e.*, $1 - e^{-\Delta\tau} \propto \Delta\tau \propto n_{\rm SI}$) or $\Delta\tau \gg 1$, over a set of sub-cycles. We note that where this approximation is less accurate, the cells ionize faster than in the accurate case, becoming rapidly optically thin and transitioning into the more accurate regime. We also have the option of treating the cells as optically thick in cases of particularly high density. However, in the bulk of the cases we consider, high density cells at the ionization front are pre-ionized by photons with a long mean-free-path because of a spectrum which has been hardened by absorption along the ray. In summary, the optically thin approximation is almost always sufficiently accurate.

We choose sub-cycle time steps for the ionization/energy ODEs to limit the fractional change in any quantity which our equations is tracking, including the energy of the cell. Thus:

$$dt = \min\left(\frac{\epsilon E}{|dE/dt|}, \frac{\epsilon n_{\rm H\,I}}{|dn_{\rm H\,I}/dt|}\right).$$

The choice of ϵ is free; we find that a choice of $\epsilon = 0.1$ following the convention of similarly written codes (Anninos et al. (1997), Wise & Abel (2011)) gives a good balance of speed and accuracy.

The solution of the ODEs for ionization/heating is also parallelized. The computational volume is divided into sub-volumes and distributed between MPI processes, with multiple copies distributed to independent processes when there are a large number of sources. Each process then calls a CUDA kernel to assign these cells to the CUDA cores of each GPU. While the processes

operate independently, MPI is used to consolidate data as necessary between ray tracing steps.

Once the calculation in each process is complete, the sub-volumes are recombined in the host process and redistributed between all processes. In some cases, the time required to distributed and recombine the grid data may be larger than the time required to actually perform the full ionization calculation; in these cases, we are also able to perform the ionization calculation locally in each process, eliminating the need to send any data between the processes during this step of the simulation.

3.2.8 Heating and Cooling

The heating of gas within each cell is calculated at the same time ionization rate is calculated. The energy per unit volume and per unit time added to each cell by a given ray is calculated based on the ionization rate:

$$\dot{E}_\nu = \sum_{\mathrm{S}} h(\nu - \nu_{\mathrm{S}}) s_\nu n_{SI},$$

where ν_{S} is the ionization edge frequency of the species being tracked. We then use the available physical quantities (such as temperature, free electron density, ionized/neutral hydrogen/helium density) to calculate the cooling rate for a variety of physical processes, including collisional ionization, recombination, collisional excitation, bremsstrahlung, and adiabatic expansion due to the Hubble flow (Hui & Gnedin, 1997). In Figure 3.2, we plot the cooling function we use in our simulations. In order to save memory, the simplest version of the code only tracks hydrogen and singly ionized helium fractions, meaning we assume a soft spectrum of radiation (stellar spectrum) in which He III number densities is negligible. In our initial simulations, all of the sources

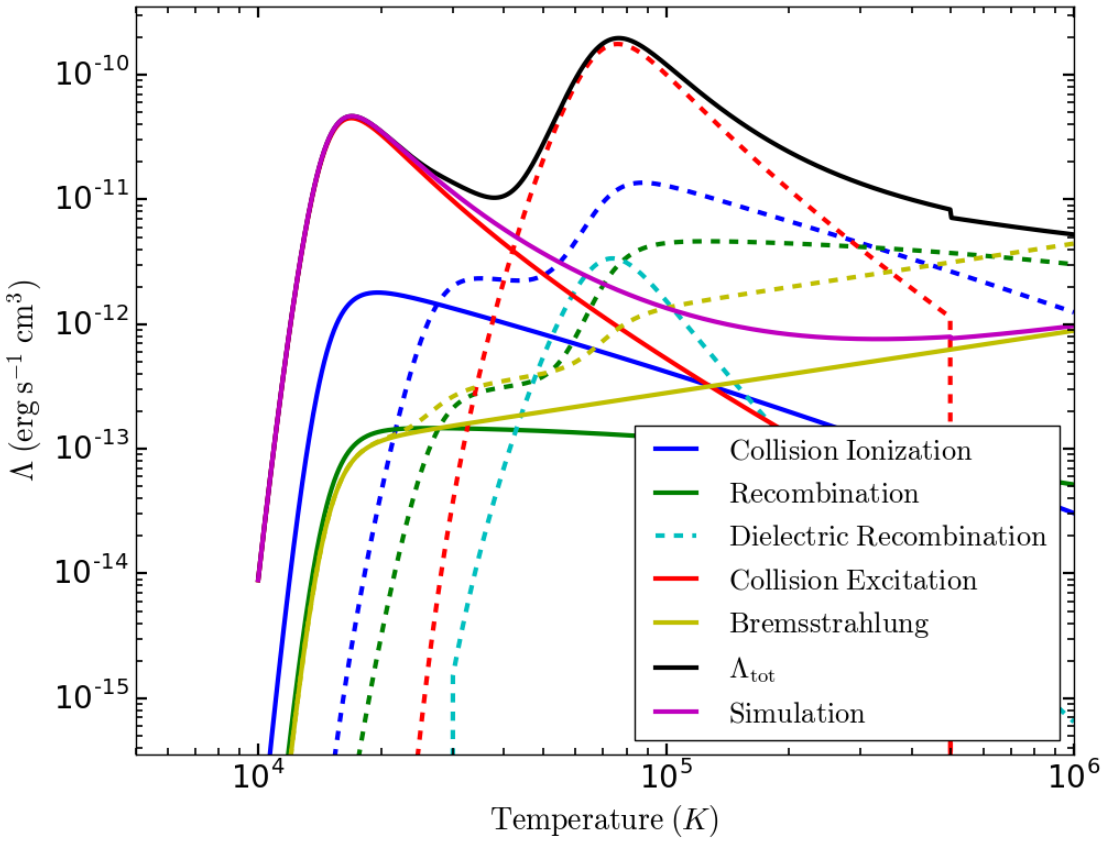

Figure 3.2: Cooling function $\Lambda(T)$ (black solid line) as a function of temperature for a gas of primordial composition (Hui & Gnedin, 1997). The solid lines represent the contribution to the cooling from different processes for hydrogen and the dashed lines for helium as shown in the legend.

are initialized with spectra corresponding to stellar sources, so that temperatures above 2×10^4 K is rare, and we thus expect our cooling rates to be accurate. We initially omit tracking He III and primordial H_2 chemistry to save on computer memory; however, we have the ability to include other relevant chemistry, for instance when we include sources of X-rays and Population III stars that emit harder photons and H_2 dissociating radiation (cite).

3.2.9 Radiative Transfer Time Step

The previous section describes how we dynamically sub-cycle the time integration between radiative transfer calls; however, the choice of time step between calls of the radiative transfer routine is more difficult. The radiation field does not evolve between ray tracing calls, so that the velocity of the I-front is limited by the choice of this time step. This means that if the time step is chosen to be too large, the speed of the I-front is unphysically reduced. However, the I-front still approaches the correct asymptotic solution at large times regardless of choice of time step.

The ray tracing step is by far the most computationally expensive part of the simulation, so we wish to maximize the time step between ray tracing calls while keeping the required accuracy. In cases where the velocity of the ionization front is not important, we can opt to use a constant time step chosen based on physical considerations. In cases where higher accuracy is required, we have implemented two different schemes for determining the optimal time step adaptively:

1. **Minimum neutral fraction change:** The simplest method for adaptively correcting the time step is regulate the maximum rate at which the neutral fraction in any given cell

changes. For a single cell:

$$dt_n = \frac{\epsilon n_{\rm H\,I}}{|dn_{\rm H\,I}/dt|},$$

where ϵ is the maximum fractional change in the neutral fraction. We calculate this quantity for every cell within the computational volume and calculate the minimum time step. We find that this method constrains the time step to be unnecessarily small in regions with small neutral fraction, where a large fractional change in neutral fraction results in a negligible change in the neutral density. We thus add the constraint to only consider cells with a relatively high optical depth:

$$\tau = n_{\rm H\,I}\sigma_i d > 0.5,$$

where d is the width of the cell and σ_i is the cross section of the lowest frequency bin. This condition also limits us to cells near the ionization front.

2. **Minimum intensity change:** Our second scheme for adaptively controlling the time step is to regulate the maximum rate at which the photon flux changes in a single cell. Similar to above, we define the time step for a given cell as:

$$dt_I = \frac{\epsilon I_\nu}{|dI_\nu/dt|},$$

Where $I_\nu = A\exp(-\tau_\nu)$. We calculate this expression at runtime by using the following

simplification:

$$\begin{aligned}\left|\frac{dI_\nu}{dt}\right| &= \left|\frac{d}{dt}A\exp(-\tau_\nu)\right| \\ &= \left|A\exp(-\tau_\nu)\frac{d\tau_\nu}{dt}\right| \\ &= \left|I_\nu\sigma_i d\frac{dn_{\mathrm{H\,I}}}{dt}\right|, \\ dt_I &= \frac{\epsilon}{\sigma_i d|dn_{\mathrm{H\,I}}/dt|}.\end{aligned}$$

We find that this method gives significantly larger time steps than the previous methods, with a minimal loss of accuracy when the ionization rate is very high.

We note here that for a given choice of time step, smaller H II regions are underrepresented compared to larger H II regions. We also note that the choice of time step makes much less of a difference in regions without radiation (as chemistry subcycles accurately solve the equations) and when H II region overlap becomes prevalent (as most H II regions are large and I-front velocities are low).

3.2.10 Ionizing Background

As the universe expands and the average electron fraction of the universe increases, the average density of neutral hydrogen decreases and the mean free path of photons in the IGM increases. These photons build up a ionizing background which becomes more dominant as more ionizing sources appear. The harder photons of the ionizing spectrum build up a background earlier than the softer photons due to their longer mean free path. Individual halos never produce large enough regions where the gas is fully ionized to completely reionize the cosmic volume, so

the derived average electron fraction underestimates the true electron fraction. We correct for this underestimation by calculating and including in the photon budget the ionizing background and its effect on the cosmic ionization history. We quantify this effect solving the equation of radiation transfer in an homogeneous expanding universe as in Gnedin (2000a); Ricotti & Ostriker (2004b), which we briefly summarize here. We begin with the number density of ionizing background photons n_ν at a redshift z and evolve it to $z - \Delta z$. During each code timestep Δz, we add to the initial background at redshift z (appropriately redshifted and absorbed by the neutral IGM) the photons produced by ionizing sources within our simulation between the redshifts of $z - \Delta z_0$ and $z - \Delta z$ including absorption and redshift effects (source term), where $z - \Delta z_0$ represents the redshift at which we begin adding the contribution of the sources to the background radiation. This parameter is used, as explained later, to avoid double counting the emission from low redshift (local) sources, that is already included in the radiation transfer calculation. Mathematically, we solve the equation:

$$n_\nu(z - \Delta z) = n_\nu(z) \exp\left[-\int_z^{z-\Delta z} dz' \alpha_{\nu'}(z')\right] + \int_{z-\Delta z_0}^{z-\Delta z} dz' S_{\nu'}(z') \exp\left[-\int_{z'}^{z-\Delta z} dz'' \alpha_{\nu''}(z'')\right], \tag{3.6}$$

where $\nu' = \nu(1+z')/(1+z)$ we have defined a dimensionless absorption coefficient and source function:

$$\alpha_\nu = \frac{(1+z)^2}{H(z)} c\overline{n}_H \sigma_\nu(\mathrm{H\,I})(1 - x_e(z)), \tag{3.7}$$

$$S_\nu = \frac{\dot{n}_{\rm ion} \langle h\nu \rangle g_\nu / h\nu}{(1+z)H(z)}, \tag{3.8}$$

where the sources spectra are normalized as $\int_{\nu_0}^{\infty} g_\nu d\nu = 1$, with $h\nu_0 = 13.6$ eV and $\langle h\nu \rangle^{-1} \equiv \int_{\nu_0}^{\infty} (g_\nu / h\nu) d\nu$. We use this equation to calculate two quantities which we track throughout the simulation:

1. The overall background $n_\nu^{\rm all}(z - dz)$ is calculated using $\Delta z_0 = 0$, so that we include background and local radiation. This quantity is used as $n_\nu(z)$ in Equation 3.6 for the next time slice.

2. The mean local radiation emission $n_\nu^{\rm loc}(z - dz)$ is calculated assuming $\Delta z_0 = H_0 R_0 / c$, where R_0 represents the size of the simulation box. By combining the local radiation background (from stars inside the box but within Δz) with the overall background $n_\nu^{\rm all}$ from previous slices, we find the background radiation from all stars at higher redshifts in the simulation without double counting the contribution of sources inside the box.

3.2.11 MPI/CUDA Parallelization

The N^3 uniform grid of our full simulation represents a simple first application of our ray tracing algorithm. In the first version of ARC (v1) every GPU requires the full grid data to perform the ray tracing calculations, which limits the possible size of the simulation grid. In the current version (v2), if the grid required for the simulation is larger than can be stored on a single GPU, the code allow the code to break the volume into smaller regions. When rays reach the edge of a sub-volume, they are sent to the adjacent sub-volume into which they move, much in the same way as AMR methods perform the calculation. This means our method is applicable in AMR settings, though our current iteration of the code is limited to Cartesian grids which may be subdivided. Regardless of these considerations, GPUs excel at performing these ray tracing

calculations as long as the number of rays within a single computational volume is large and the CUDA core occupancy is near optimal ($> 10^4$).

3.3 Radiative Transfer Tests

Consistency tests are paramount in demonstrating that a new radiative transfer code produces results that are consistent with well established methods or scenarios that have exact analytical solutions. In this section we present the results produced by our code when running the tests presented in (Iliev et al., 2006a) (RT06). These tests include 0) Tracking chemistry in a single cell 1) expansion of an isothermal H II region in pure hydrogen gas 2) expansion of an H II region in a pure hydrogen gas with evolving temperature 3) I-front trapping and the formation of a shadow 4) multiple sources in a cosmic density field. We show that our code performs well when compared to CPU codes that use similar ray-tracing methods. The various codes used for comparison in RT06 show a good deal of variability for several of the tests. The method implemented in our code most closely resembles C^2-Ray (Mellema et al., 2006) and MORAY (Wise & Abel, 2011). Where possible we compare our results to the publicly available RT06 results for C^2-Ray. Since at the moment our ray tracing algorithm is used in a post-processing setting, it neglects hydrodynamic evolution. We therefore limit ourselves to tests which do not require accurate tracking of the hydrodynamic properties of the gas.

3.3.1 Test 0 - Chemistry in a Single Cell

The test presented in RT06 applies a constant radiation field of plane-parallel radiation to a single cell of the simulation. Our code is limited to tracking point sources of radiation, so we

place a single source of photons at one side of a 6.6 kpc cube with 128^3 cells. The source is given a luminosity 5.2×10^{57} photons s^{-1} so that the flux at the cell is 10^{12} photons s^{-1} cm^{-2}. The density of the cell is $n = 1$ cm^{-3} and it is initially neutral at temperature $T = 100$ K. The radiation is approximated to be a blackbody of temperature 10^5 K; following Wise & Abel (2011), we use four frequency bins with central energies $E_i = (16.74, 24.65, 34.49, 52.06)$ and relative luminosities $L_i/L = (0.277, 0.335, 0.2, 0.188)$. The radiation is applied for 0.5 Myr, after which the radiation is turned off and the cell is tracked for 5 Myr.

In figure 3.3 we plot the neutral fraction and temperature of the cell as a function of time. We find that these results agree with those presented in RT06. The only discrepancy between our results and those presented in RT06 is in the electron fraction for the first time step. This is a result of the optically thin approximation for sub-cycles. The first ray tracing calculation occurs at $x_{\mathrm{H\,I}} = 1$, at which point the cell absorbs more soft photons, so that the chemistry calculator doesn't reach the correct equilibrium point. After this step, however, the optically thin approximation is satisfied and the solution agrees from the second step on.

3.3.2 Test 1 - Expansion of Isothermal H II egion in Pure Hydrogen

The most fundamental test for a radiative transfer code is the simulation of an expanding H II region around a constant luminosity source in a constant density medium of pure hydrogen Assuming the medium is initially neutral and that the H II region has a sharp boundary, we have

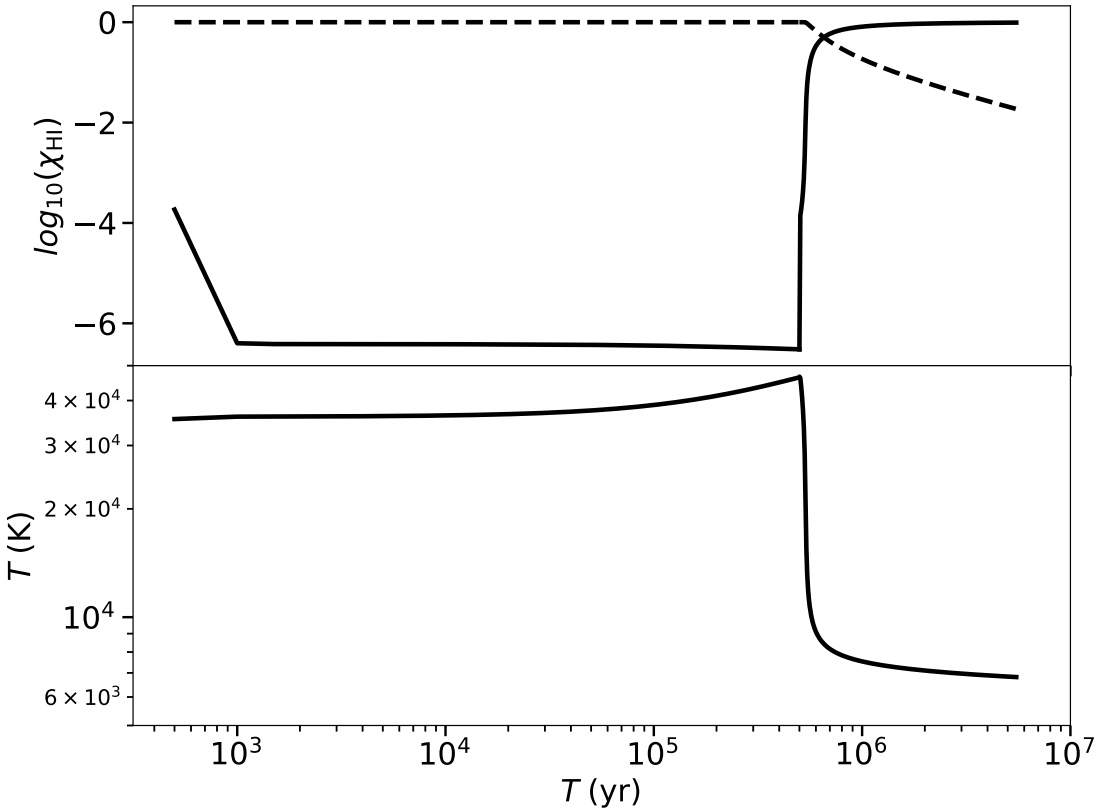

Figure 3.3: Test 0 - Chemistry in a single cell. In this test, a single cell is subject to intense radiation for 5×10^5 yr, at which point the radiation stops. (*Top.*) Plot of the neutral fraction (solid line) and ionized fraction (dashed line) as a function of time. (*Bottom.*) Plot of the temperature as a function of time. These results agree well with the results presented in RT06, except at the first time step, which is expected given our use of a fixed time step.

the well known analytic solution:

$$r_I(t) = r_S \left(1 - \exp\left(\frac{t}{t_{\rm rec}}\right)\right)^{1/3}, \tag{3.9}$$

$$v_I(t) = \left(\frac{r_S}{3t_{\rm rec}}\right) \frac{\exp(t/t_{\rm rec})}{(1 - \exp(t/t_{\rm rec}))^{2/3}}, \tag{3.10}$$

where:

$$r_S = \left(\frac{3\dot{N}}{4\pi\alpha_B(T)n_H^2} \right)^{1/3},$$

$$t_{\rm rec} = [\alpha_B(T)n_H]^{-1}.$$

The test domain is a 6.6 kpc cube with a 128^3 element cubic grid. The gas has a fixed density $n_H = 1.0 \times 10^{-3}$ and temperature $T = 10^4$ K and an initial equilibrium ionization fraction of 1.2×10^{-3}. We place a single ionizing source in the corner of the simulation box, and assume a monochromatic spectrum with energy $h\nu = 13.6$ eV and luminosity of $S_0 = 5 \times 10^{48}$ photons s^{-1}. The chosen physical parameters give a Strömgren radius $R_S = 5.4$ kpc and a recombination time $t_{\rm rec} = 122.4$ Myr. We track the evolution of this simulation for 500 Myr, or roughly four recombination times.

In Figure 3.4 we plot a slice through the computational volume at t=500 Myr (top left). This panel shows that the code produces a spherically growing region of ionization, as expected. In the top-right panel of Figure 3.4 we plot the radially averaged profile of the H I fraction as a function of radius at four times throughout the simulation. The shape of this profile agrees well with the results of RT06. In the bottom-left panel of Figure 3.4 we plot the comparison of our simulation with the model in Equation 3.9. We see that our model agrees with the analytic fit within 5% for the majority of the simulation. The underestimation of the radius at early times is a result of using a fixed time step for these tests; with an adaptive time step, the agreement is much better (see Figure 3.11). Finally, in the bottom-right panel we show the histogram of H I fraction for three times in the simulation (solid lines). The dashed lines represent the same plots produced using the C^2-Ray data from RT06. We see that our code produces the expected results

for this test.

3.3.3 Test 2 - Expansion of H II egion in Pure Hydrogen with Evolving Temperature

The second test is similar to the first test, but with the reintroduction of heating and cooling processes to the simulation. We give the source a blackbody spectrum with $T = 10^5$ K, using the same spectrum and luminosity as Test 0. The gas is given an initial temperature $T = 100$ K.

In the top and middle left panels of Figure 3.5 show a slice through the neutral fraction at $t = 10$ Myr and $t = 100$ Myr, respectively. We see that the boundaries of the H II region in these plots is less sharp than those of Fig. 3.4, as anticipated with the harder radiation present in the $T = 10^5$ K black body spectrum. We plot $t = 100$ Myr instead of $t = 500$ Myr for this test because the edge of the H II region and heating reaches the boundary by then, meaning the plots show less relevant information. In the top and middle right panels of the same figure, we plot a slice through the temperature at the same times to show how the radiation is able to heat at a larger radius than the ionization front. In Figure 3.6 we show the respective radial profiles of the neutral fraction (top-left) and the temperature (top-right). We plot our results (solid lines) against the RT06 results for C^2-Ray (dashed lines) for comparison. In contrast to Figure 3.4, we see that the neutral fraction increases more gradually, in good agreement with the results of C^2-Ray. We note that our model heats the gas outside of the Strömgren sphere more than the C^2-Ray results from RT06. This is a result of the difference between the blackbody spectrum from Wise & Abel (2011) and the one used in RT06.

In the bottom-left panel of Figure 3.6 we show the growth of the radius of the Strömgren

sphere as a function of time. We see that the radius of the ray tracing model begins lagging behind the analytic model while the gas is being heated, and later moves beyond the analytic model once the gas is heated beyond 10^4 K. These results are in good agreement with the models presented in RT06. Finally, the bottom-right panel of Figure 3.6 shows the histogram of neutral fractions (left plot) and temperatures (right plot) at $t = 10$, 100, and 500 Myr (solid lines) against the C^2-Ray results (dashed lines). Again, we see the neutral fractions are in excellent agreement, while our model heats the lower temperature gas slightly more, due to the adoption of a harder spectrum that was chosen to match the one adopted in Wise & Abel (2011).

3.3.4 Test 3 - I-front Trapping and Formation of a Shadow

Test 3 in RT06 is designed to test the diffusivity and angular resolution of the radiative transfer code. In this test, a field of uniform radiation is projected towards a dense sphere of uniform hydrogen surrounded by a very thin medium. We plot the results of our model in Figure 3.7. The I-front propagates at a constant velocity towards the clump until it reaches the surface, at which point the optically thick clump begins slowly absorbing the radiation. The lines of sight which pass through the sphere are trapped in the clump, causing a shadow to form behind the clump. The sharpness of the edge of the shadow is a measure of the diffusivity of the method. The edges of the sphere become ionized before the rest of the clump, causing the shadow to shrink; this allows to visually assess the angular resolution of the code. The test also tracks the rate at which the I-front progresses through the clump.

The original test presented in RT06 is contained within a 6.6 kpc box with resolution 128^3. The ambient medium has a density $n_{\rm out} = 2 \times 10^{-4}$ cm^{-3} and temperature $T_{\rm out,init} = 8000$

K, while the clump has a density $n_{\rm clump} = 0.04$ cm^{-3} and temperature $T_{\rm clump,init} = 40$ K. The clump has a radius 0.8 kpc and is centered at $(x_c, y_c, z_c) = (97, 64, 64)$ in grid units. The original test assumes plane-parallel radiation with flux 10^6 s^{-1} cm^{-3}. However, as in Test 0, we follow Wise & Abel (2011) and replace the plane parallel radiation with a single point source of radiation opposite the clump, with $(x_c, y_c, z_c) = (0, 64, 64)$. The luminosity of the source is set at $S_0 = 3 \times 10^{51}$ photons s^{-1}, so that the flux at the center of the clump matches the plane parallel flux of the original test. The simulation is evolved for 15 Myr, at which point RT06 find that the I-front is just past the center of the clump.

The top and middle left panels in Figure 3.7 show the neutral fraction in a slice through the center of the volume at $t = 1$ Myr and $t = 15$ Myr, respectively. The corresponding panels in the top and middle right show the temperature for the same slice and times. We notice immediately that the shadows in our simulation are opening with distance, a result of the relatively small distance of the point source from the clump, as opposed to the plane parallel radiation in the original test. We also see that the diffuse gas outside the clump is immediately ionized and photo-heated to the point of becoming optically thin. We also see that edge of the shadow is very sharp, with the neutral fraction going from ~ 1 to $\sim 10^{-4}$ in the space of $\sim 1 - 2$ cells.

The bottom panels in Figure 3.7 show the neutral fraction (bottom-left) and temperature (bottom-right) within the clump along the ray passing trough the center of the clump. The clump is centered at $r/L_{\rm box} = 0.75$ and extends between $r/L_{\rm box} \sim 0.6$ and $r/L_{\rm box} \sim 0.90$. We see that the photo-heated gas outside the clump reaches values above 3×10^4 K, while the photo-heated gas inside the clump has a temperature below 2×10^4 K, a result of the increased cooling at higher gas density. We see that the neutral fraction rises and the temperature falls as we move through the I-front. The neutral fraction reaches 50% at $r/L_{\rm box} \sim 0.8$, a result consistent with RT06.

For comparison we also plot the results of C^2-Ray as similarly colored dashed lines. While the neutral fractions agree well, there is more of a discrepancy in the temperatures. We believe this is a result of the ray divergence present with a point source that is absent in the case of plane parallel radiation. The diverging rays overheat the gas behind the front and underheat the gas in front of the front, an effect we see in our plot.

Finally, in Figure 3.8 we plot the average neutral fraction and temperature inside the clump for our model (solid lines) against the results from C^2-Ray (dashed lines). While we see that our model both heats and ionizes the clump less effectively than C^2-Ray, we again believe this is a result of the diverging nature of our radiation, which exposes a smaller portion of the sphere to the full radiation, at least initially. Despite this discrepancy, our results are still well within the spread of models shown in RT06.

3.3.5 Test 4 - Multiple Sources in a Cosmic Density Field

The final test in RT06 is a simple simulation of a cosmological density field. The simulation volume is cube with side $0.5\ h^{-1}$ cMpc at redshift $z = 9$ and resolution 128^3 (here $h = 0.7$). The source of ionization is 16 point sources centered within the 16 most massive halos, and emit $f_\gamma = 250$ ionizing photons per baryon with the same $T = 10^5$ K blackbody spectrum used in tests 2 and 3. The sources are assumed to live for $t_s = 3$ Myr, which is longer than the length of the simulation, so that they remain on for the entire simulation. The luminosity of each source is thus:

$$\dot{N}_\gamma = f_\gamma \frac{M\Omega_b}{\Omega_m m_H t_s}.$$

Here M is the mass of the halo, $\Omega_b = 0.043$, and $\Omega_m = 0.27$. It is assumed that radiation leaving the box is lost, and the volume is tracked for a total period of 0.4 Myr. The density grid and halo position/luminosities are currently available from the Cosmological Radiative Transfer Comparison Project website. In Figure 3.9 we show a selection of plots from this simulation. The top four figures are slices through the center of the simulation to allow for visual comparison between our results and the RT06 results. The ML and MR plots show slices through the neutral fraction grid at the $z = z_{\rm sim}/2$ plane of the simulation at $t = 0.05$ Myr and $t = 0.2$ Myr, respectively. The LL and LR plots show slices through the temperature grid at the $z = z_{\rm sim}/2$ plane of the simulation at $t = 0.05$ Myr and $t = 0.2$ Myr, respectively. The results are in good agreement, although the difference in colormap between our plots and those of RT06 may make visual comparison difficult. Our plots may also be visually compared to those of Wise & Abel (2011), which use a similar colormap and thus closely resemble our results.

The bottom-left panel in Figure 3.9 shows histogram of neutral fraction (left) and temperature (right) at times of 50, 200, and 400 kyr. These results are in good agreement with the codes presented in RT06. The bottom-right panel of Figure 3.9 shows the volume averaged (χ_v, solid black line) and mass averaged (χ_m, dashed red line) ionized fraction within the volume as a function of time. We see that the mass averaged ionized fraction is larger at early times, in agreement with the expectation of inside-out reionization within the RT06 simulation (Gnedin, 2000b; Miralda-Escudé et al., 2000; Sokasian et al., 2004). These results are again in good agreement with those presented in RT06.

3.3.6 Methodology Test: Spherical Correction Factor

The spherical correction factor defined in Section 3.2.5 represents an attempt to approximate the true overlap between the cubic grid cells and the conic ray intersecting at arbitrary angles. The true correction factor depends on too many factors to calculate exactly, so any simple approximation will necessarily be somewhat flawed. In Figure 3.10 we plot a collection of slice plots from Test 1 using different choices for the power used in the correction factor formula. The top and bottom rows of this figure represent the same slices with different choices of ray splitting rate ($\Phi = 1$ and $\Phi = 3$, respectively), while each column represents a different choice for the correction factor. The first column shows the case with no correction, and the artifacts are clearly visible. The second column shows the linear correction factor, and the second plot includes the secondary correction (which moves the excess radiation from a given cell to the adjacent cell which the ray is closest to). Finally, the last column shows the squared correction, as described in Wise & Abel (2011). We see that the linear power correction with the secondary correction results is the closest approximation of a true sphere. The inclusion of this secondary correction also serves the purpose of compensating for the lost radiation due to the primary correction factor. In Figure 3.11 we plot the comparison of the model radius to the analytic Strömgrem radius for the same four models in Figure 3.10. We see that the linear and square models reduce the radius of the model by factors of 5% and 10%, respectively. The correction factor, however, restores the accuracy of the model radius quite well. We thus chose to use in our simulations the secondary correction in combination to the f_c factor as it preserves photon conservation and removes azimuthal artifacts.

3.4 Benchmarks

The way in which our code processes individual rays with GPUs is inherently parallel (CUDA kernel). For a single GPU, the number of cores over which this work is distributed is fixed by the specifications of the individual GPU. We thus choose to benchmark our code by measuring its performance on a simple test problem varying the number of source of radiation and the dimension of the computational grid. Our test problem is based on a test in Wise & Abel (2011). We place the ionizing source(s) at the center of a cubic 64^3 grid of size 15 kpc. The medium is pure hydrogen of density $10^{-3}\,\mathrm{cm}^{-3}$, and the ionizing source(s) have a monochromatic 17 eV spectrum with luminosity 5×10^48 photons/s. The simulation is run with the radiation-base adaptive time step (see Section) for 250 Myr. For each simulations, we plot the time it takes for the entire ray tracing algorithm to run (Total Rad) as well as the time it takes for the algorithm to run on a single sub-volume (we divide the volume in 8 sub-volumes).

Dimension Scaling: We measure how our code performs with varying grid size by performing the test simulation on grids by running the same simulation with grid resolutions of 64^3, 128^3, 256^3, 512^3, 1024^3. The red line shows the time it takes for the full radiative transfer calculation, while the blue line shows the line shows the time it takes for a single sub-volume to complete its ray tracing calculation (the blue dashed line is the time it takes for the CUDA kernel to execute in isolation; this time is measured locally on the GPU, while the rest of the times are measured on the host node). The difference between these lines is the time it takes for the rays to be shared between the nodes and for the rays to propagate through the rest of the volume. We plot the difference between the total radiation calculation and the kernel execution times with the green line, as a measure of the MPI overhead during the radiative transfer calculation. Finally,

we include the time taken for the ionization calculation with the cyan line. We see that the MPI overhead varies little with dimension, while the overall execution time scales with grid dimension as a power-law with slope between $N_{\rm dim}^2$ and $N_{\rm dim}^3$. The problem should theoretically scale with $N_{\rm dim}^3$: two powers of $N_{\rm dim}$ for the number of rays, and one power of $N_{\rm dim}$ for the number of steps along a ray of the same physical length. However, our GPU code doesn't perform as well when the number of rays is small, as the occupancy of the GPU processors is lower and some computational power is wasted.

Source Scaling: We measure how our code performs with a varying number of sources by measuring how long the code takes to perform a single step of the calculation with N sources at the same location, with the luminosity of each reduced by a factor of N, so that the final result should be the same. In the left plot of Figure 3.12 we plot the resulting times for 1, 10, 100, and 1000 sources. We see that the code's overall time scales linearly with the number of sources, as anticipated.

Cosmic Simulation Benchmark: We also include a benchmark based on a full cosmic simulation run using our code on a 128^3 grid divided evenly into 8 sub volumes, each 64^3. In Figure 3.13 we plot the time taken for a complete ray tracing step divided by the total number of sources (blue) and by the number of sources in the most populated sub-volume (green). In such cosmic simulations, the bottlenecked is the sub-volume which takes the longest time to process, which is why we plot the time divided by the number of sources in the most populated volume. However, the since the GPU's process the rays independently, the time divided by the total number of sources is a better proxy for the time taken per step per source, as long as the sources are relatively evenly distributed between the sub volumes. In this simulation, when $x_e \sim 0.2$, the simulation is able to fully process rays from $\sim 10^4$ sources on a 256^3 grid in ~ 20

seconds.

3.5 Summary

We have described our implementation of the spatially adaptive ray tracing radiative transfer code ARC which is designed to take advantage of the extremely parallel processing power of the GPUs available in today's supercomputers. Our algorithm is based on the well known method presented in Abel & Wandelt (2002). We have presented the methodology of our code, as well as a novel approach to a correction method for ray tracing in a Cartesian grid. Our code is able to split a computational volume into sub-volumes, each of which is contained on an independent GPU linked by MPI, allowing our code to tackle very large problems avoiding the limitations related to the available memory on the GPU. We verified the accuracy of our method by performing a selection of tests presented in Iliev et al. (2006a). Finally, we discussed a selection of benchmarks to demonstrate the speed and scaling of our code.

We believe that the unique characteristics of GPUs make them ideal for the computational problem of ray tracing. ARC already takes advantage relatively recent innovation such as CUDA aware MPI (only becoming available in 2013), but the optimization and speed up of our code is an ongoing process that will take advantage of new capabilities of GPUs as they become available. The speed, number of cores and the memory of GPUs has been growing rapidly over the years allowing us to tackle problems which were computationally unfeasible only a few years ago.

The first application of ARC will be to simulate the reionization epoch using pre-computed dark matter simulations in a $(10\ \mathrm{cMpc})^3$ volume, with sufficient resolution to capture minihalos with masses $> 10^6\ \mathrm{M}_\odot$. These simulations can resolve the sites of formation of Population III

stars, which we will be able to model adopting the model in (Ricotti, 2016), and have a sufficiently large volume to capture the formation of galaxies observed in the Hubble ultra-deep fields. We will take advantage of the non-diffusive nature of the ray-tracing method to simulate realistic emission of ionizing radiation from the minihalos, including the short duration of the bursts of radiation and anisotropic emission expected if the dominant star formation mode is in compact star clusters (Ricotti, 2002a; Katz & Ricotti, 2013, 2014; Ricotti et al., 2016). Based on our previous analytical study (Hartley & Ricotti, 2016), we believe that properly accounting for these effects will have a major impact on both the topology of reionization and the budget of ionizing photons necessary to reionize by redshift $z \sim 6.2$.

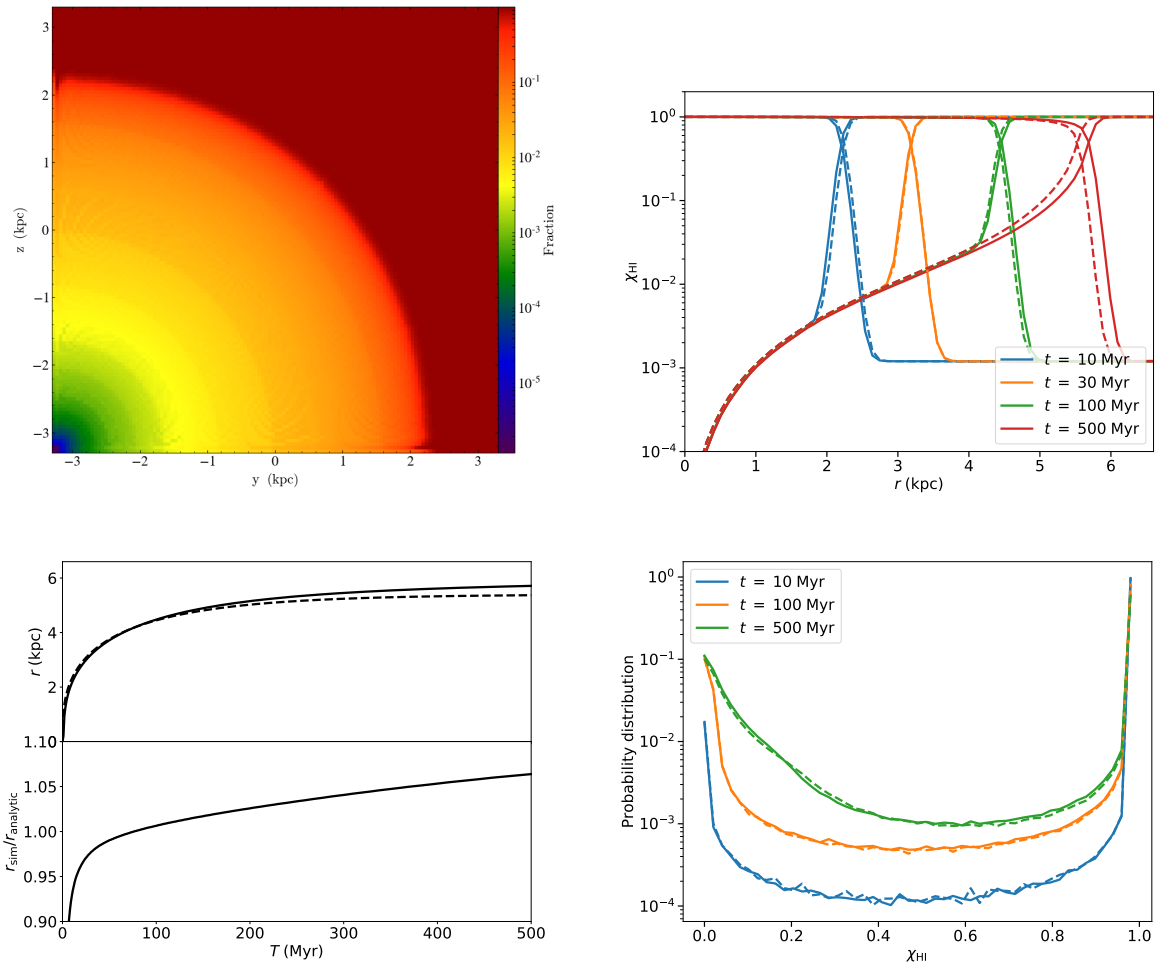

Figure 3.4: Test 1 - Expansion of an isothermal H II region in a gas of pure hydrogen. (*Top-left.*) Slice plots of the neutral fraction at 500 Myr into the simulation. The size and shape of the H II region are in good agreement with the results in RT06. (*Top-right.*) Radially averaged neutral fraction as a function of distance from the source, in kpc. The dashed lines are the same plots reproduced from RT06 using C^2-Ray. (*Bottom-left.*) Ionization front plotted along with the analytical model (top) and the ratio of these models (bottom). We note that the simulation deviates from the analytic solution at the beginning and end of the simulation. This is a result of the soft 13.6 eV spectrum being the worst case scenario for our optically thin approximation for non-equilibrium chemistry calculation. (*Bottom-right.*) Histogram of neutral fractions at $t = 10,\ 100,\ 500$ Myr. The dashed lines are the same plots reproduced from RT06 using C^2-Ray. We see that our model matches C^2-Ray very well.

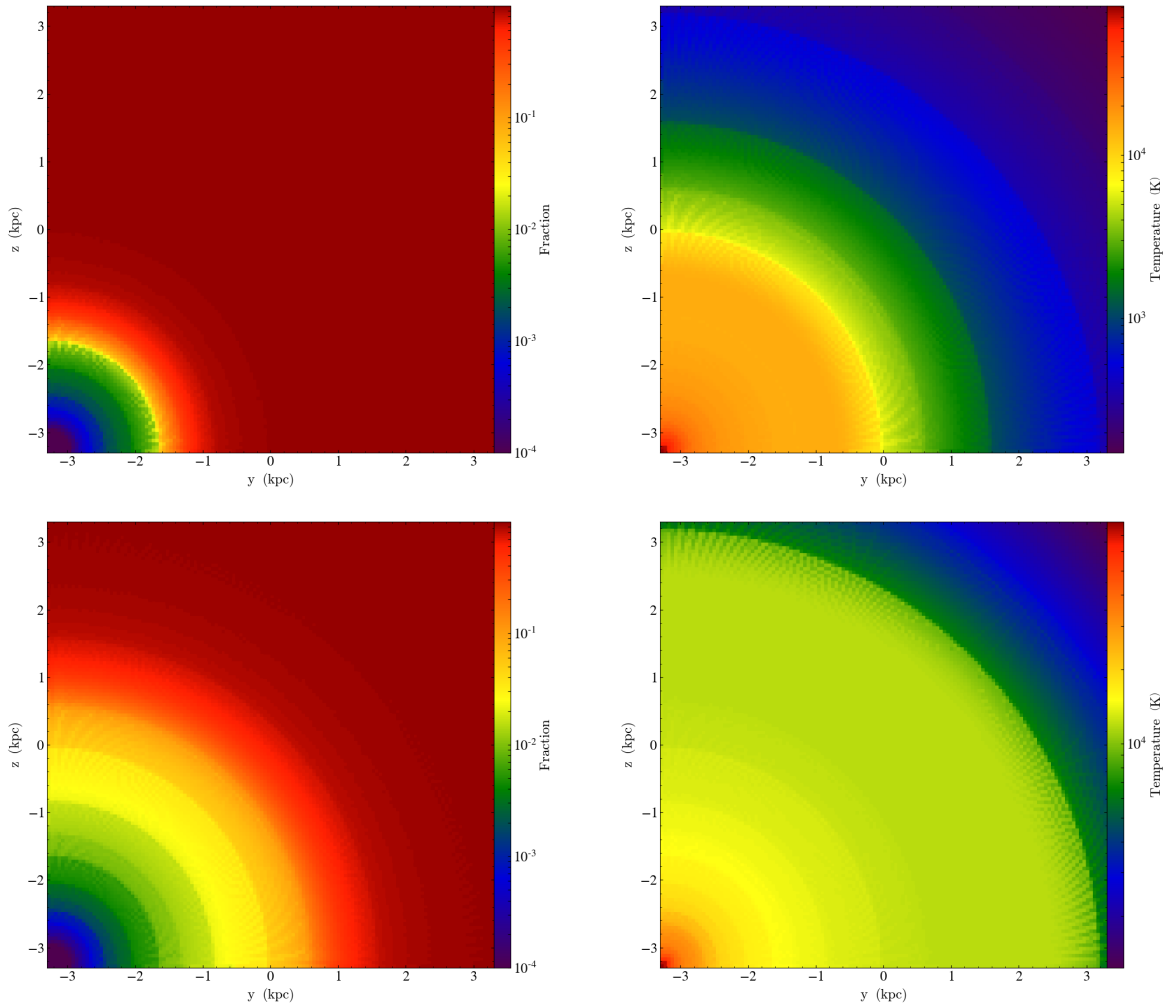

Figure 3.5: Test 2 - Expansion of H II region in a gas of pure hydrogen with evolving temperature. (*Top and bottom left.*) Slice plots of the neutral fraction at 10 Myr and 500 Myr into the simulation, respectively. (*Top and bottom right.*) Slice plots of the temperature at 10 Myr and 500 Myr into the simulation. The adopted 10^5 K blackbody spectrum introduces harder photons which heat the gas to a much larger radius than the size of the Strömgren sphere, in agreement with the results presented in RT06.

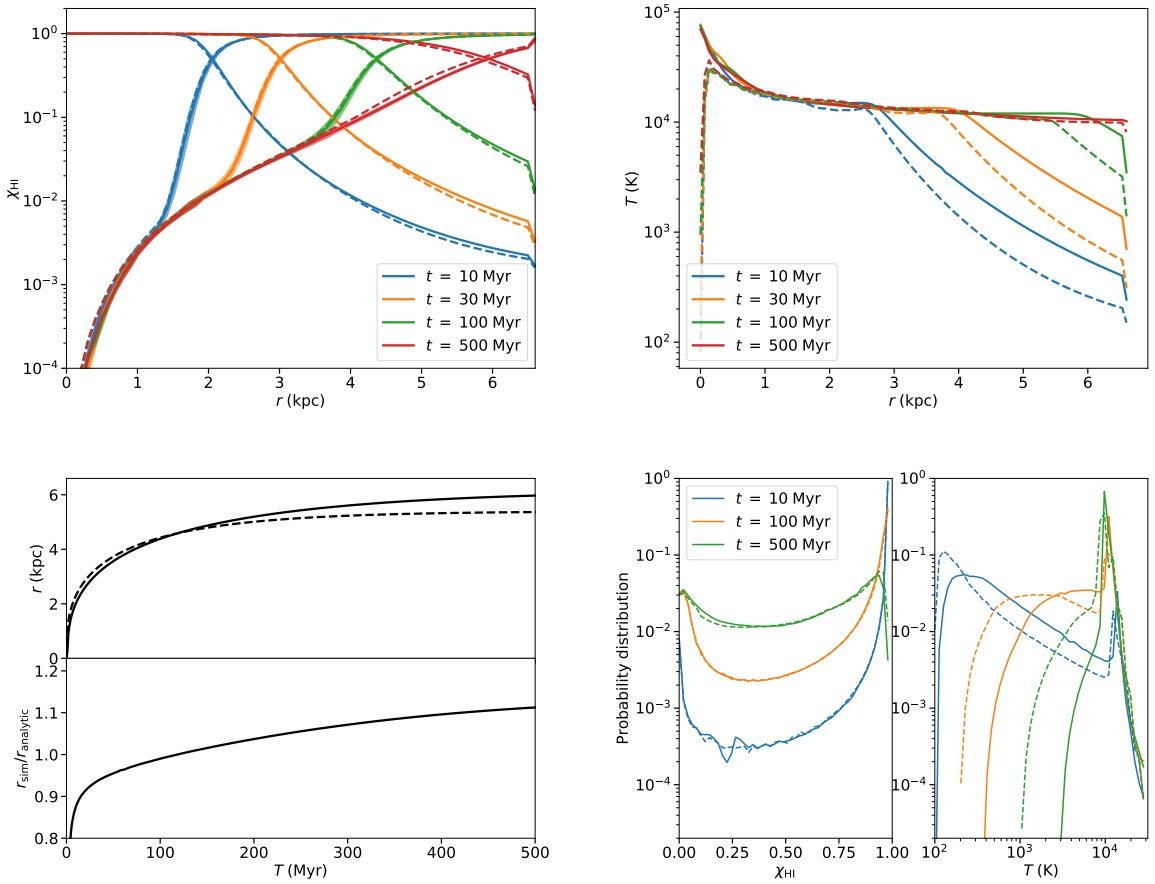

Figure 3.6: Test 2 - Expansion of H II region in a gas of pure hydrogen with evolving temperature. Radially averaged neutral fraction (*top-left*) and temperature (*top-right*) as a function of distance from the source, in kpc. The shaded thickness of the solid neutral lines represents the variance of the radial average. The dashed lines in both plots represent the same plots reproduced from RT06 using C^2-Ray. The slight disagreements are a result of a different assumption on the spectrum of the source (monocromatic vs blackbody spectrum) between our code and the C^2-Ray run in RT06. (*Bottom-left.*) Ionization front plotted along with the analytical isothermal-model (top) and the ratio of these models (bottom). This model deviates more than in test 1, a result of the non-isothermal nature of the simulation. (*Bottom-right.*) Histograms of the neutral fraction and temperature within the simulation, respectively. The dashed lines represent the same results calculated from C^2-Ray data from RT06. We find that our code produces slightly higher temperatures than C^2-Ray, a result of the difference assumption on the source spectrum between our code and the C^2-Ray run used in RT06.

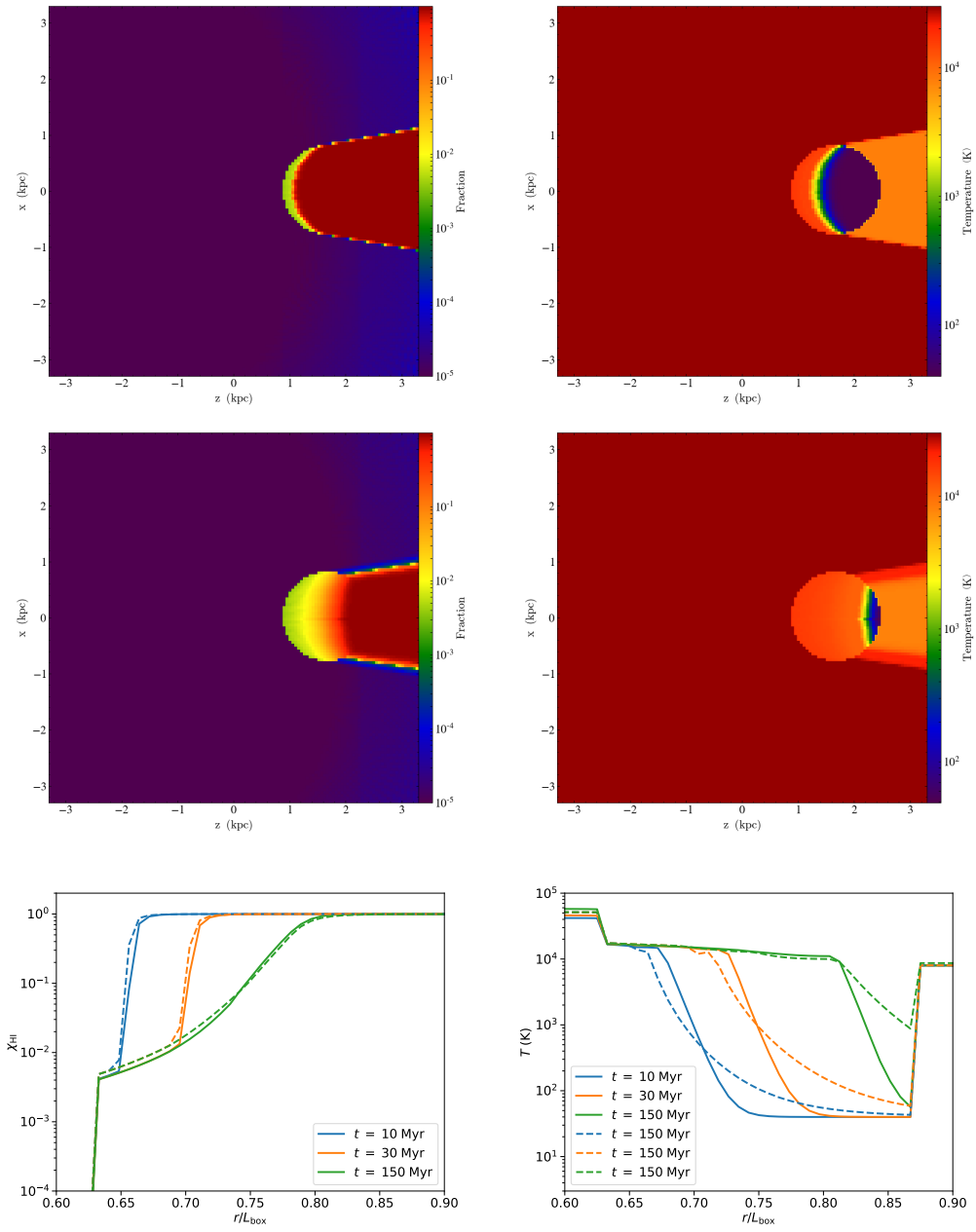

Figure 3.7: Test 3 - I-front trapping and formation of a shadow. (*Top and middle left.*) Slice plots of the neutral fraction at 1 Myr and 15 Myr into the simulation, respectively. (*Top and middle right.*) Slice plots of the temperature at 1 Myr and 15 Myr into the simulation. The shape of our shadow region expands with distance, a result of the point-like nature of the source of ionizing photons. Our results agree well with Wise & Abel (2011), a simulation which uses the same point-like configuration. Neutral fraction (*bottom-left*) and temperature (*bottom-right*) along a ray going through the center of the dense clump. The dashed lines represent the same plots calculated using C^2-Ray using data from RT06. We interpret the discrepancy in the right figure as the result of using a point source, as in our simulation, rather than a plan-parallel front. The diverging rays heat more effectively closer to the source (low r) and less effectively further from the source (high r).

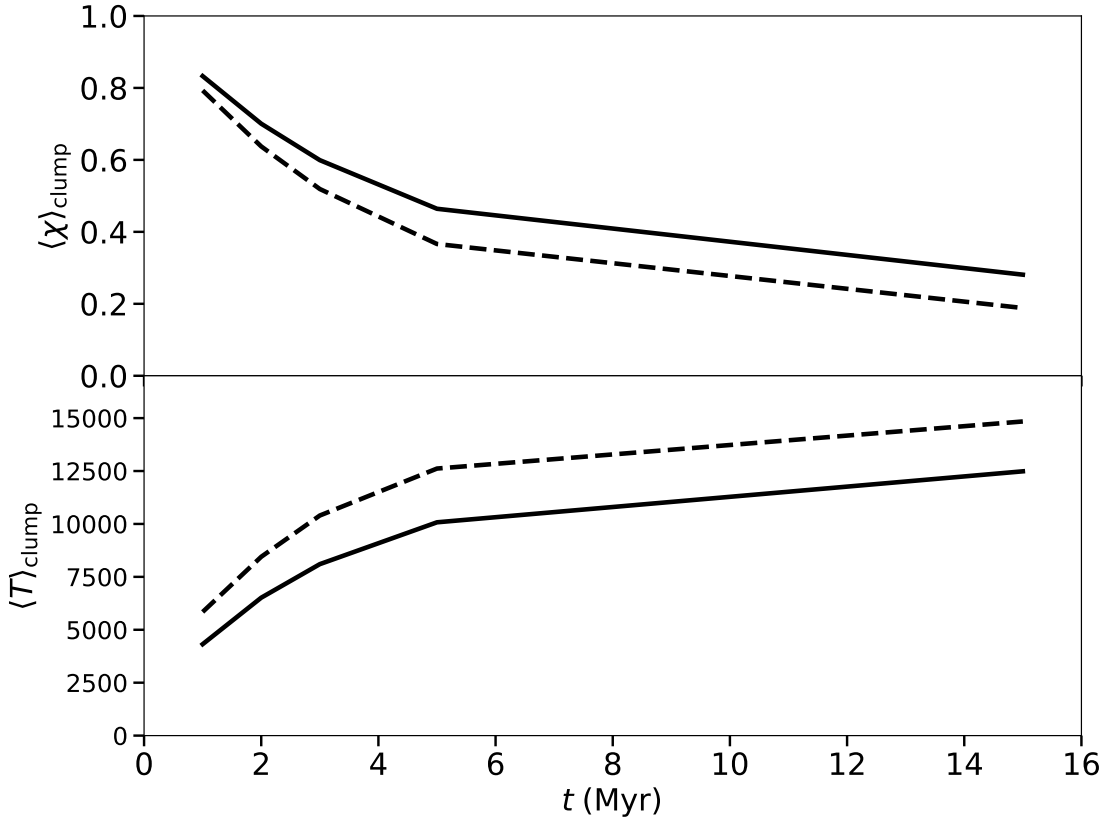

Figure 3.8: Test 3 - I-front trapping and formation of a shadow. Average neutral fraction (*top*) and temperature (*bottom*) within the dense clump as a function of time during the simulation. The dashed lines represent the same results plotted using C^2-Ray data from RT06. We see that our ionization and heating are lower; we interpret this discrepancy as the result of adopting a point source of radiation rather than a plane-parallel front.

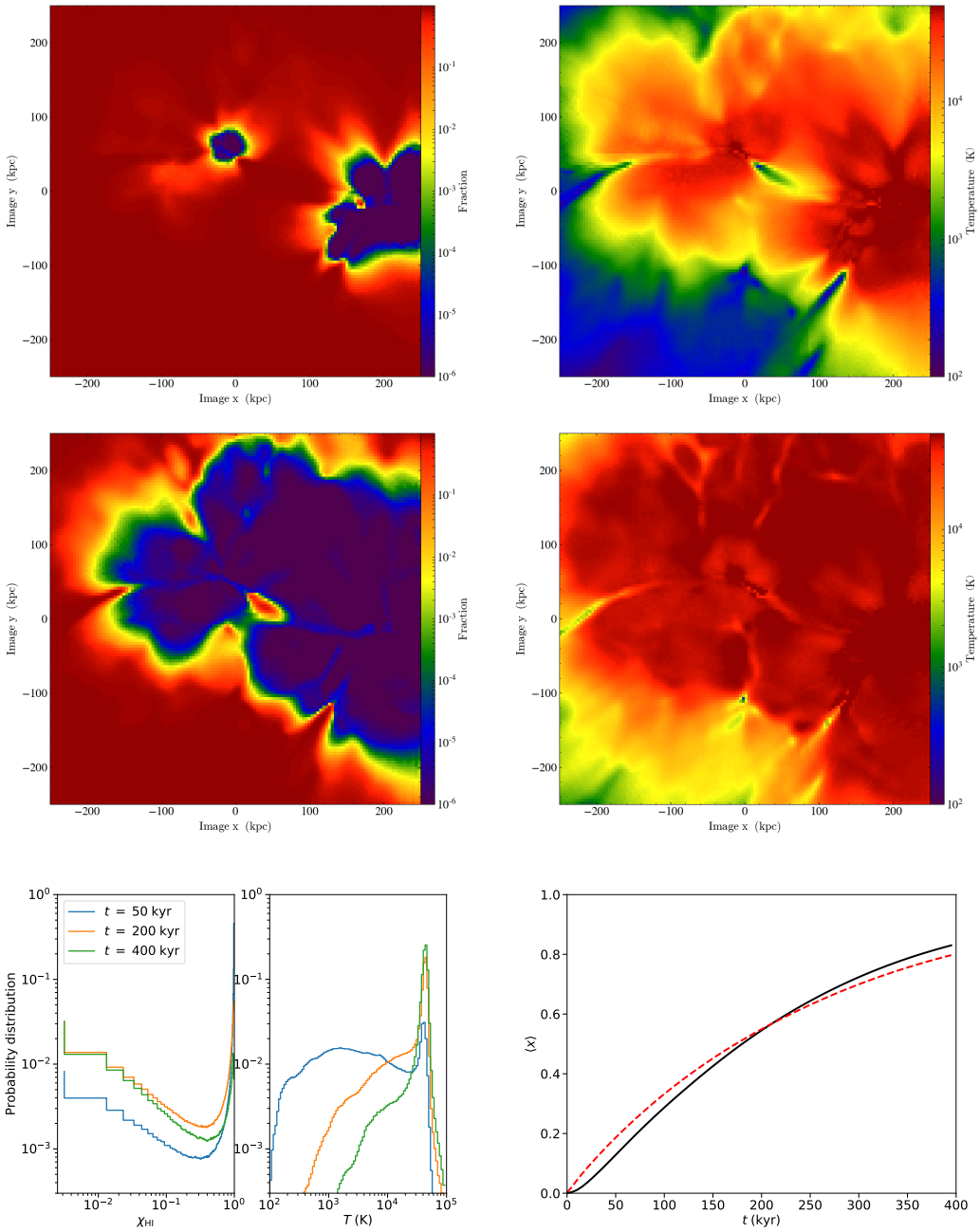

Figure 3.9: Test 4 - Multiple sources in a cosmic density field. (*Top and middle left.*) Slice plots of the neutral fraction at 0.05 Myr and 0.2 Myr into the simulation, respectively. (*Top and middle right.*) Slice plots of the temperature at 0.05 Myr and 0.2 Myr into the simulation, respectively. These results are almost identical to those presented in Wise & Abel (2011), which uses a similar adaptive ray tracing scheme to ours. (*Bottom left.*) Histograms of neutral fraction (left) and temperature (right) at fixed times in the simulation. (*Bottom right*) Mass averaged (dashed line) and volume averaged (solid line) ionized fraction as a function of time throughout the simulation box. These results are in good agreement with the results presented in RT06.

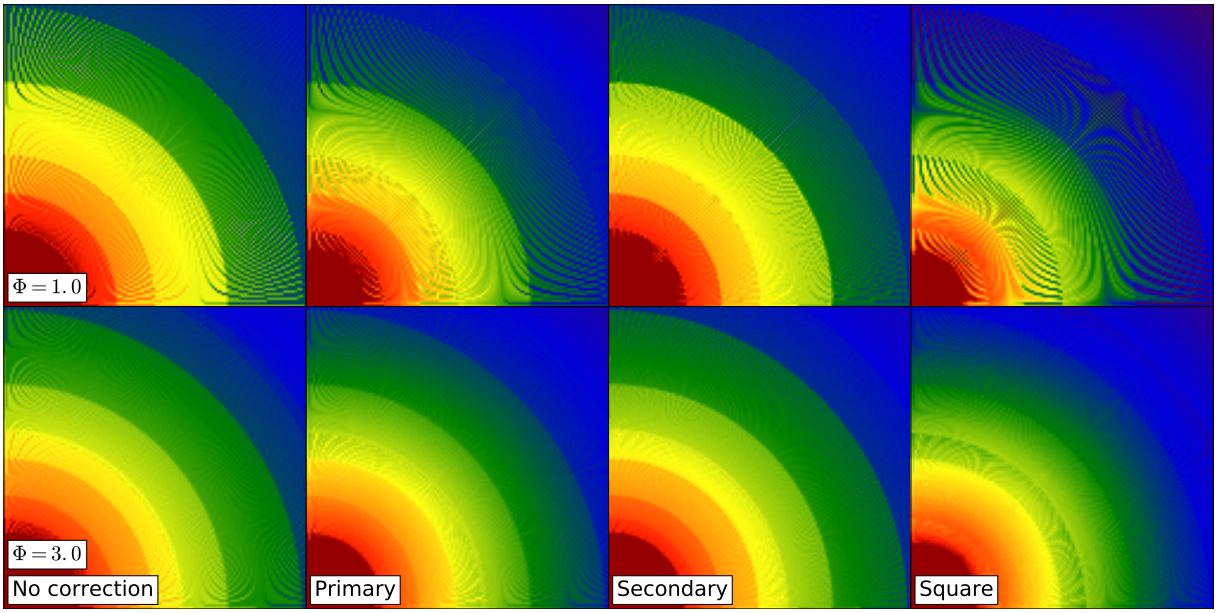

Figure 3.10: This figure shows a selection of correction methods for two different ray splitting rates. The parameter Φ represents the number of rays per cell area, so that the bottom row have three times the splitting rate of those in the top row. The first column represents the method without correction factors. The second column illustrates the case with a primary correction, or the application of Equation 3.2 to the rate of ionization in every cell. The third column represents the case with primary and secondary correction, or the application of Equation 3.3 to the cell which is closets to the cell face has the greatest overlap. The fourth column represents the case with only the primary correction with f_c squared, as presented in Wise & Abel (2011). The low $\Phi = 1$ plots were included to emphasize the difference between these configurations; clearly a higher value of Φ improves the accuracy for any configuration, but also increases the computational work for the simulation.

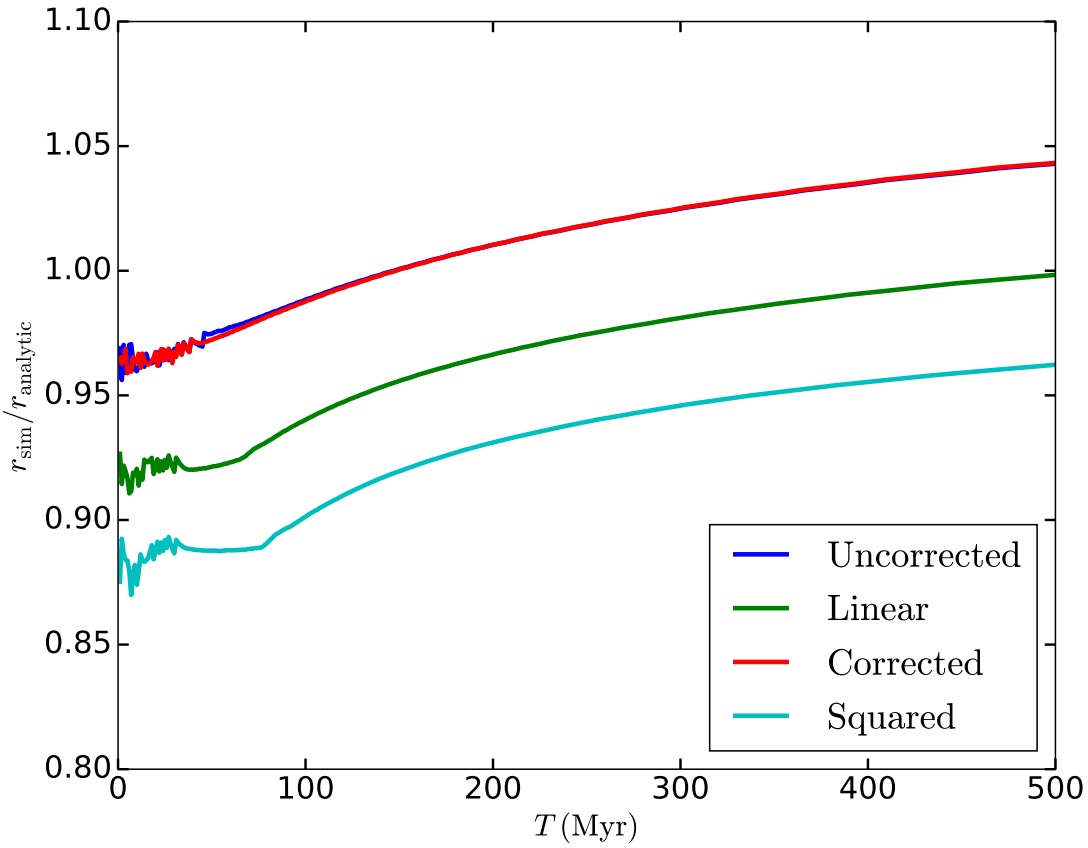

Figure 3.11: Same as the bottom panel in Fig. 4 (left) but using different spherical correction factors. We see that the f_c and f_c^2 corrections shrink the size of the ionized region by roughly 5% and 10%, respectively, while including a secondary correction to the f_c correction restores the size of the region to the uncorrected size.

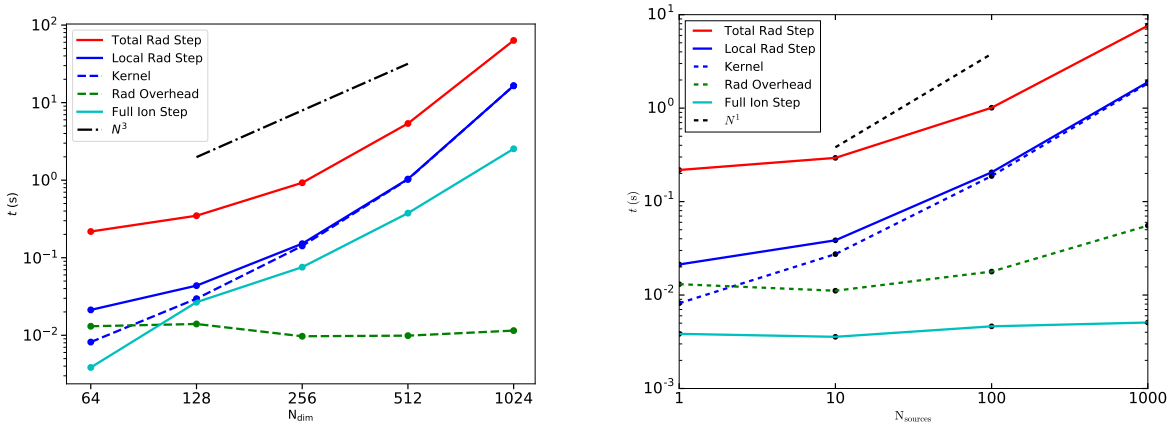

Figure 3.12: Benchmarks for a Strömgren sphere in a cubic grid. (*Left.*) The time taken to trace rays from a single source through the entire grid for grid sizes 64^3 through 1024^3. The time taken scales between N^2 and N^3. This is expected for this problem, as the number of rays scales with N^2 and the number of steps along each ray scales with N. (*Right.*) The time taken for a single radiative transfer step at grid size 64^3 for a number of sources between $n = 1$ and $n = 1000$. For small source counts the per-step overhead can be seen, and as the number of sources increase, the time taken scales linearly, as expected.

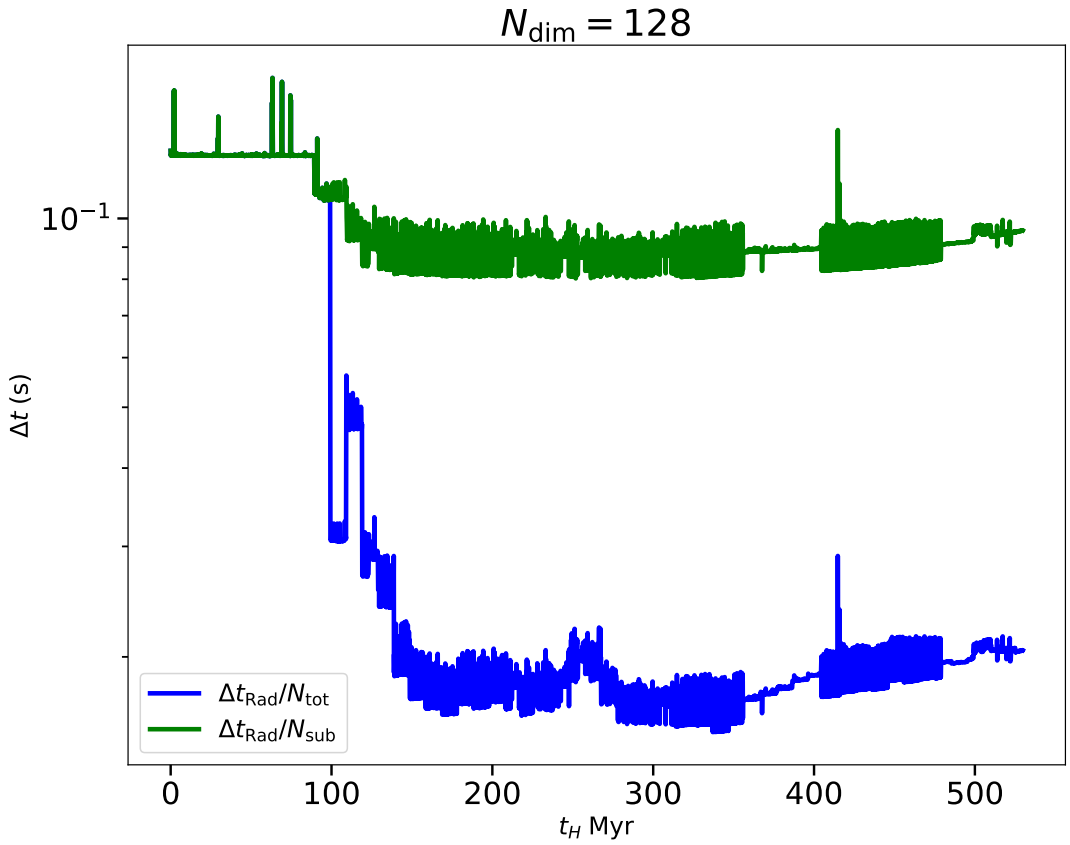

Figure 3.13: Execution times (in seconds) for a 128^2 simulation of cosmic reionization plotted as a function of simulation time (on 8 GPUs). The plot shows the time taken by a full radiative transfer calculation divided by the total number of sources (blue line) and divided by the maximum number of sources in any sub-volume (green line). The eight sub-volumes are processed in parallel, so as long as the sources are well distributed within the volume, the blue line should represent the mean time to execute the radiative transfer step per source at this resolution. We note that the times are unusually large at the beginning of the simulation, as the constant overhead dominates when the number of radiation sources is very small (*i.e.*, zero or one source)

Chapter 4: Simulating Bursty and Continuous Reionization with ARC

4.1 Introduction

The goal of this thesis is to demonstrate that bursty star formation is able to drive reionization more effectively than continuous star formation. The semi-analytic models presented in Chapter 2 gave us enough confidence that this effect is real and substantive to work towards creating a more robust model which is able to accurately evaluate how these modes of star formation affect the course of reionization. In Chapter 3, we presented ARC, the library we developed to accomplish this task. In this chapter, we present the results of a suite of simulations run using ARCand our conclusions from the results of these simulations.

In Section 4.2, we present several improvements and adjustments that we made to ARC while working towards a satisfactory comparison of bursty and continuous models. In Section 4.3, we describe all of the details and parameters that we chose for our simulations. In Section 4.4, we report the results of these simulations. In Section 4.5, we discuss the implications of these results. Finally, in Section 4.6 we conclude our findings and discuss what the next steps going forward will be.

4.2 Updates

The version of ARC that we presented in Chapter 2 was fully operational, and accurate to the standards presented in (Iliev et al., 2014). However, early attempts to simulate reionization showed us that the library and the results it produces would benefit greatly from a collection of relatively minor tuning changes and upgrades. Here we motivate and describe these updates.

4.2.1 Enforcing and Tracking Photon Conservation

Ray tracing methods are inherently photon conserving, meaning that all photons that are emitted by sources are absorbed by the gas in the grid in a one-to-one fashion. Thus, the overall rate of photon emission, $\Sigma_{i=1}^{N} S_0$ is equal to the density-weighted sum of all the photon ionization rates over the grid, The caveat is that when the universe becomes transparent to ionizing photons – near the epoch of reionization – the photons can escape from the simulation box without being absorbed. These photon ionization rates are then used to calculate the change in ionization of each cell, and it was at this step that we encountered difficulties in maintaining photon conservation. This difficulty can be attributed to two systematic sources of error, both of which are the result of using finite instead of infinitesimal time steps:

- Local: The rate at which the ionization of a cell changes is the ionization rate of the cell minus the recombination rate of the cell. The recombination rate (which depends on the square of the number of ionized particles) tends to increase, with the overall result being that a cell approaches complete ionization asymptotically. The semi-implicit calculation of Chapter 3 approximates this process very accurately; however, using different recombina-

tion rates over the course of a time step means breaking photon conservation. Each cell is also only able to absorb as many photons as there are absorbers within the cell, meaning that photon conservation will drift away from equilibrium as time steps get longer, and break down for time steps on the order of or larger than the ionization time of the cell.

- Inter-cell: As a cell becomes more ionized, its opacity decreases, which means that the intensity of a ray which has passed through the cell increases. Because the intensity remains constant between ray-tracing calculations, the overall result is an underestimation of the number of photons absorbed, which is qualitatively seen as the propagation of ionization fronts at slower than accurate rates. The error becomes relevant when the ionization fronts move through the grid at rates comparable to a single grid cell per time step.

While it is fundamentally impossible to eliminate these errors, they can be managed due to the fact that both are correlated to the length of the time step between ionization calculations. We added a system to the library for keeping track of the all quantities associated with photon conservation at the global level, which will allow us to adjust time steps in accordance with error management on the fly. The tracked quantities include:

- S_0: The cumulative rate at which photons are emitted by all sources within the simulation during a given time step.

- Γ_i: The cumulative ionization rate for species i in a given time step. This number represents the sum of the photon ionization rates after they have been distributed within the grid of species i by the ray tracing section module of the library.

- A_i: The cumulative recombination rate for species i in a given time step. This number is

calculated within the chemistry module of the library.

- $\Delta x_i/\Delta t$: The cumulative rate of change of the ionization fraction for species i in a given time step. This number is calculated within the chemistry module of the library, and unlike the previous quantities, the change in ionization fraction for a given cell is capped by the initial ionization of that cell, as the ionization fraction must be within the interval $[0, 1]$.

In Figure 4.1 we plot a sample of these quantities from a short section of one of our trial simulations. In theory, photon conservation is enforced when $S_0 = \Gamma$ and $\Delta x/\Delta t = S_0 - A$, where Γ, Δx, and A are the appropriately weighted averages of Γ_i, $\Delta x/\Delta t$ and A_i, respectively. In this example, all of the sources are steadily increasing in brightness, and when we read pre-calculated density files of cosmological volumes ever 10 Myr, the number of source and their locations is updated. The result of this is a jump in the recombination rate, as sources appearing in neutral gas quickly form H II regions that contribute to the overall recombination rate, and a resulting decrease in the overall ionization rate. The difference in S_0 and Γ_h is due to the fact that helium is also absorbing photons. Finally, the transient oscillatory behavior is due to the wave front traversing individual high density cells close to the sources themselves.

The tracking of these quantities allows us to evaluate to what degree and at what point in the calculation photon conservation is violated, and adjust parameters as necessary. To this end, there are two comparisons that are of particular interest to us:

- S_0 and Γ: The difference between the weighted sum of the Γ_i's and S_0 represents total error in the ray tracing process. Processes that can contribute to this include floating point error, rays leaving the computational grid, or rays that are terminated because their flux is below the threshold for tracking. This quantity does not depend on the length of time step.

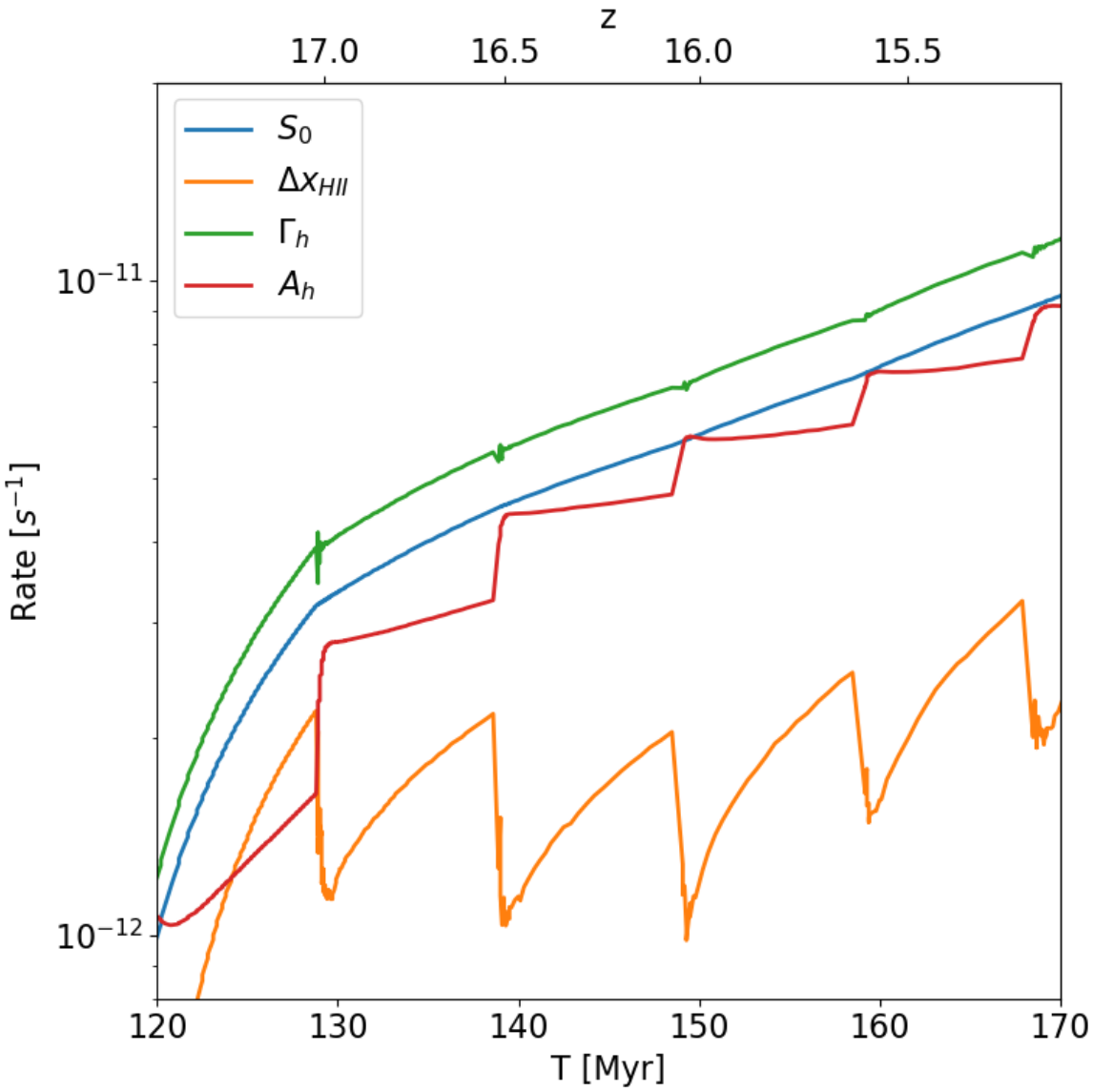

Figure 4.1: Example rates of photon production, hydrogen ionization, hydrogen recombination, and ionization change taken from a 50 Myr segment of one of our simulations.

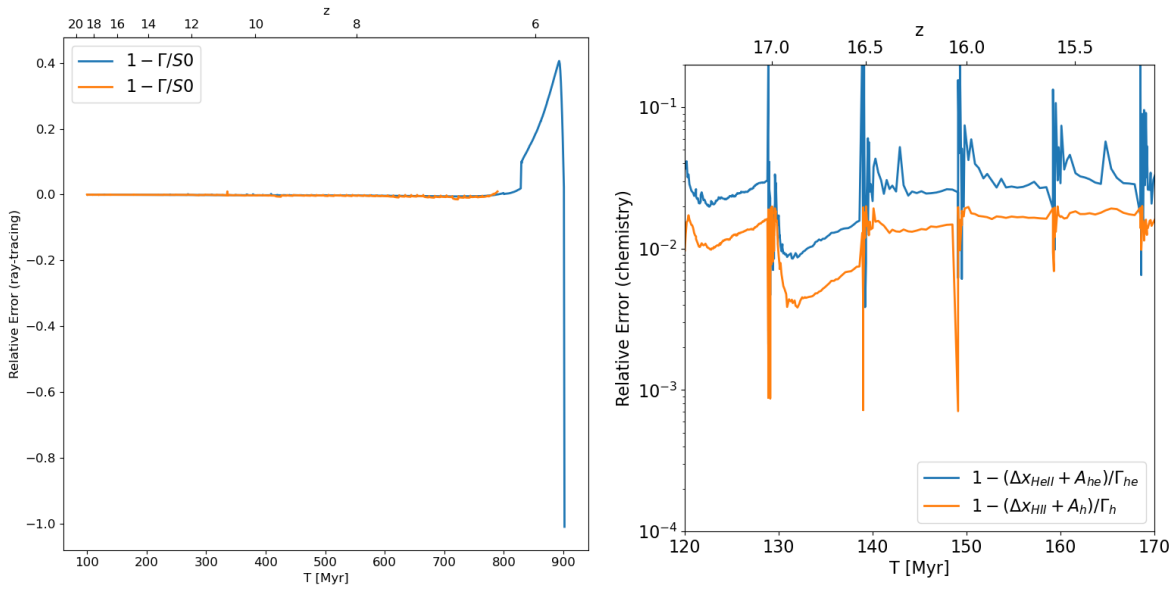

Figure 4.2: An example of the errors taken from a 50 Myr segment from on of our simulations (*Left*). The S_0 and Γ error plotted as the deviation of S_0/Γ from 1. A positive error indicates that photons are not being absorbed. (*Right*). The $\Gamma_i - A_i$ and $\Delta x_i/\Delta t$ error shown for hydrogen and helium separately.

- $\Gamma_i - A_i$ and $\Delta x_i/\Delta t$: The difference between the target rate of change in ionization ($\Gamma_i - A_i$) and the actual rate of change in ionization ($\Delta x_i/\Delta t$). The actual rate of change in ionization is capped, meaning that this error is an indication that the ionization rate or recombination rate for a set of cells is larger than the amount of ionization or recombination that they can support in a given time step. In practice, this is error is generally the result of cells very close to newly appearing sources ionizing on time scales shorter than the time step of the simulation.

In Figure 4.2 we plot a sample of these quantities. The left plot shows the S_0 and Γ error, which in general does not depend on the length of the time step, and is negligible when all of the photons are being absorbed. The right panels shows the conservation error comparing $\Gamma - A$ to $\Delta x/\Delta t$ for both hydrogen and helium (we show $[(\Gamma - A) - \Delta x/\Delta t]/\Gamma$). This error is mostly

due to cells which have ionization timescales shorter than the time step of the simulation.

Because photon conservation is critical to this thesis (*i.e.*, we want to compare the number of photons needed to reionize the universe at $z \sim 6$ in "bursty" versus "continuous" models of star formation in galaxies), we created a new adaptive time step paradigm focused on managing this error. The library calculates these error quantities before performing the ionization calculation, and reduces the time step as necessary to keep the photon conservation error below a supplied threshold level. The distributed nature of ARC made this computationally non-trivial, and we accomplish it with the following iterative process:

1. Calculate the local ionization rate from the ray tracing module.

2. Calculate the local recombination rate and ionization rate from a simplified version of the ionization module.

3. Average these rates over all nodes to determine the global error rate. This step introduces a round of message sharing between all computational nodes, making it the major complicating factor of this process.

4. Compute a new time step reduced by the ratio of the calculated error and the target error. This error depends sensitively on the details of the cells which violate conservation, making the calculation of an acceptable timestep impossible with a single guess.

5. Repeat until the computed error rate is below the target error.

In the right plot of 4.2, we enforce a 2% error limit on the photon conservation error for hydrogen. Helium generally has a higher error than hydrogen because it is more exposed to the hard photons which are able to reach greater distances into the neutral medium and affect more

cells. However, helium contributes much less to the cumulative quantities because: it is roughly 25% of the intergalactic medium by mass, each absorber corresponds to four baryons, and it does not absorb only from the harder photon bins.

In Figure 4.3 we plot a sample of time steps corresponding to the plots in Figure 4.2. The time step decreases dramatically when new sources appear in a previously dark halo (especially common in bursty models) as the error spikes when the intensity is very high in dense neutral regions. As the H II regions grow, the time step constraint is relaxed and the simulation proceeds more quickly. We can also see an overall trend towards longer time steps later, which is a result of the overall density of the simulation decreasing due to cosmological expansion. However, gas near or within dark matter halos continues to maintain high densities, so the initial drop in time step persists.

4.2.2 Periodic Boundary Conditions

The version of ARC presented in Chapter 3 treats rays which leave the simulation volume as immediately entering the ionizing background. While computationally accurate, we found that this was unphysical at early times, as the optically thick IGM would begin receiving ionizing radiation from a "background" that shouldn't able to reach it. Our solution to this was to upgrade the library to allow for periodic boundary conditions, which are defined such that rays leaving the computational volume in the $x = x_0$, $y = x_1$, or $z = x_2$ direction are transported to the opposite side of the axis they escaped on, with all other properties of the rays maintained. Explicitly, we

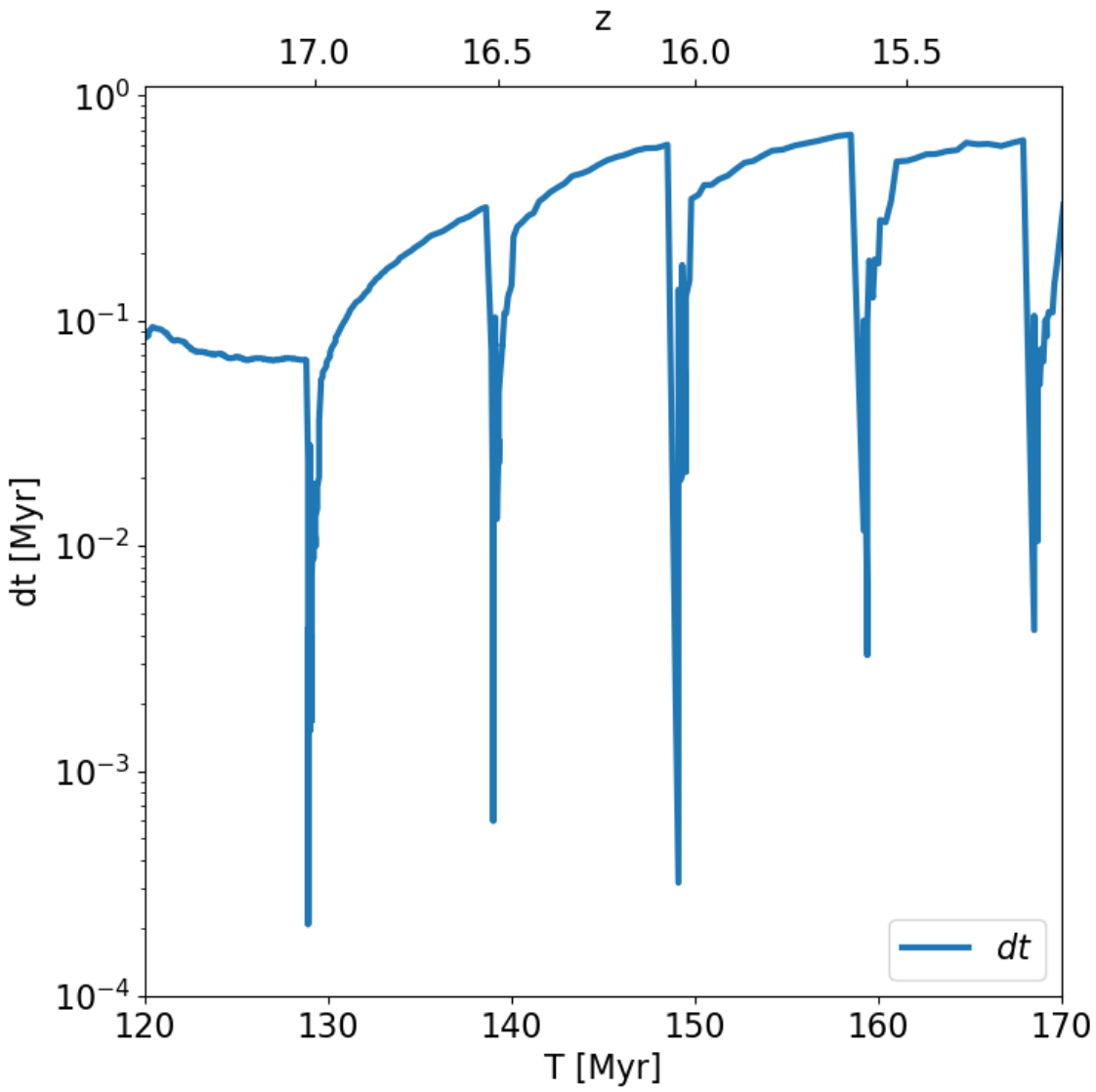

Figure 4.3: Example time steps used to enforce photon conservation at a 2% level, taken from the continuous model of Set 1.

write this as:

$$x_i = \begin{cases} x_i - L, & \text{if } x_i \geq L \\ x_i + L, & \text{if } x_i < L \\ x_i, & \text{otherwise} \end{cases} \quad (\text{for } i = 1, 2, 3).$$

Under periodic boundary conditions, rays do not have any fundamental limit to how far they can be tracked. However, the resolution constraints we described in Chapter 3 imply that the number of rays tracked by ARC is generally proportional to the distance the rays travel squared, which means the computational work will increase rapidly as the rays travel distances larger than the size of the box. We thus implement a hard distance limit for how long to track rays, thereby also limiting the maximum possible computational load. We fix this limit to be $\sqrt{3}$ times the length of the side of the simulation, which represents a single diagonal crossing the of cubic simulation volume. This choice means that rays will only stop being tracked once they've been able to travel at least once through the computational volume, which will only be possible when the simulation approaches reionization and the background we described in Chapter 3 is an accurate physical approximation.

4.2.3 Smooth Emission Adjustment

Our simulation allows us set the location, luminosity, and time of appearance of sources as desired. This process is inherently discrete, which means that the cumulative luminosity will jump discontinuously when new sources appear. To alleviate this effect, we added two paradigms for smoothing the cumulative flux:

- Flat luminosity boosting: in this paradigm, we calculate the cumulative flux for two con-

secutive source adjustments, S_i at t_i and S_{i+1} at t_{i+1}. We then apply an interpolation factor to the luminosity of every halo in the simulation:

$$\frac{t_{i+1} - t}{t_{i+1} - t_i} + \frac{S_{i+1}}{S_i} \frac{t - t_i}{t_{i+1} - t_i}$$

This factor transitions linearly from 1 to S_{i+1}/S_i, which allows a smooth transition in the cumulative flux. However, there is still some discontinuous behavior: the factor drops from S_{i+1}/S_i to 1 at a transition, and new sources appear with their full luminosity. This is the paradigm we chose for the simulations presented in this chapter.

- New source suppression: in this paradigm, new sources appear with a luminosity of 0 and ramp linearly to full luminosity over a time period of Δt, which is manually set.

4.2.4 Memory Management

We built ARC from the ground up, learning MPI and CUDA techniques as we worked. The method we used for storing rays relied on the creation of a new memory allocation for each level of HEALPix split. This meant that the library would repeatedly allocate and free memory for the array of rays through the course of a single ray-tracing step. This process is relatively slow, so we overhauled the library to use fixed portions of memory allocated at runtime, thus saving significant computational time. The library actively monitors the amount of this memory being used and is capable of processing more or fewer rays to saturate the memory and CUDA cores more effectively.

4.2.5 Source Clustering

As reionization simulations approach completion, the number of sources increases rapidly while the distance that rays need to be tracked from each source also increases. Both of these factors contribute to a workload which increases very rapidly as the ionization fraction approaches 1. Because of this, we implemented a simple procedure for merging halos within close proximity of each other in a way that minimizes the change in the simulation's output while significantly reducing the workload required to complete the computation, which we describe in this section.

Once the position and luminosity of all halos has been assigned at a given time step, we apply an out-of-the-box K-means clustering algorithm to all of the sources with a K value which is smaller than the total number of sources within the simulation. The K centroids produced by this algorithm are then each uniquely associated with a single source within the volume. The library then calculates the total luminosity of all of the sources associated with a given centroid and assigns that total to the centroid. These K centroids are then used as alternative sources within the simulation, reducing the amount of computation required for each ray-tracing step by a factor of roughly the ratio of the total number of original sources to K.

In Figure 4.4 we present an example of what this algorithm looks like. The left plot shows a 3D scatter plot of the 1171 sources taken from a snapshot at $T = 700Myr$ from one of the continuous models presented later in this Chapter. The right plot shows the result of applying the algorithm with $K = 500$, effectively reducing the number of sources by more than 50%. The algorithm preserves the spatial distribution of sources, including the overall mass distribution and location of isolated groups of sources.

The spatial accuracy of ARC's ray-tracing module is the most important factor in our choice

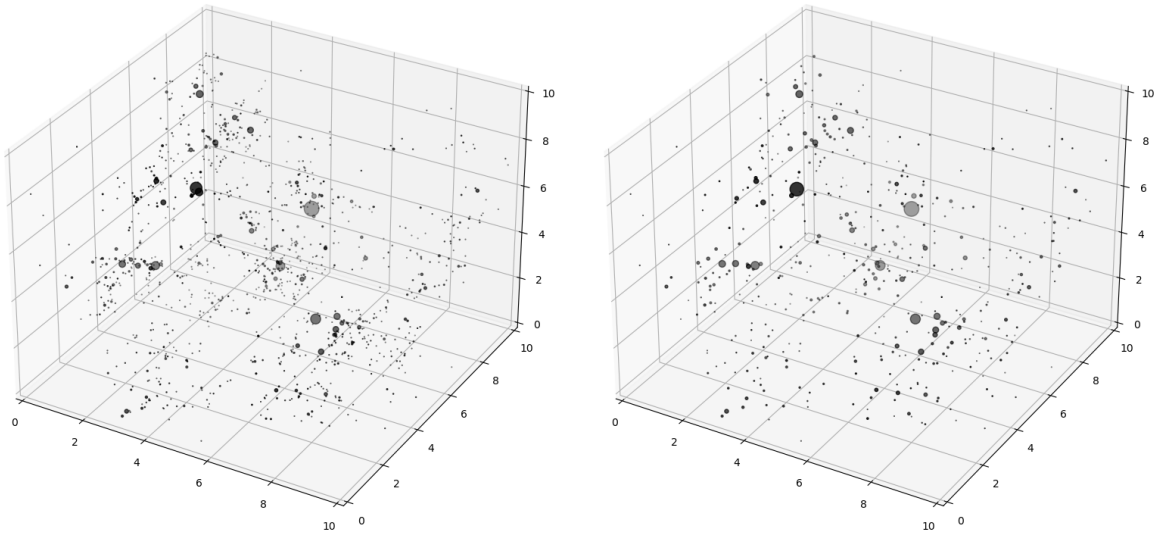

Figure 4.4: Example of the effect of source clustering from a snapshot of sources at $T = 700$ Myr. The left plot is all 1171 sources, and the right plot is the K-Means algorithm applied with $K = 500$. The size of the dots represents the stellar mass/luminosity of the sources. We see that the algorithm maintains the overall distribution as well as the isolated and outlier sources.

of this library, yet clustering sources sacrifices some of this accuracy by effectively smearing the resulting HII regions. Thus, great care is taken in the time and K values for which this method is applied to ensure that little accuracy is sacrificed. As the ionization fraction approaches 1, all H II regions within the simulated volume are relatively large and amorphous, meaning that it becomes easier and easier to find K that satisfy the necessary conditions for application. The simulations we present in this Chapter do not necessitate the use of this method; early tests suggested that the large number of sources present at late times would make the computations prohibitively expensive, however we settled on a luminosity cutoff that solved this issue.

4.3 Simulation Parameters

The goal of our simulations is to study the difference between bursty and continuous star formation models on the time evolution of the cosmic reionization process. To begin our simulations, we need to make a few specific choices for the parameters of the simulations:

4.3.1 Shared Properties

- Volume: The volume of our simulations is a cube with a side length of 10 cMpc. The simulations begin at $z = 30$, or roughly 100 Myr after the big bang. This is generally thought to be before any significant amount of Pop III stars had formed, making it a safe choice for a reionization study. There is no fundamental limit to how long we can follow a simulation, but the rest of our input data spans a length of 910 Myr, or up to a redshift of $z \sim 5.8$.

- Baryon Density: ARC uses user supplied density grids and halo data. The choices we made for these in our simulations are the results generated by a dark matter only simulation run by Emil Polisensky using Gadget-2 (Springel, 2005). These simulations produced outputs every 10 Myr, and we update the density information with this frequency. The outputs consist of overdensity grids, which represented the density of the dark matter at every point in the grid relative to the mean density of the entire volume. We further cap this overdensity factor at 200, to prevent the gas near the largest bursts of star formation from absorbing unrealistic amounts of photons.

- Temperature: We assume all of the ionized gas within the simulation is at a fixed temper-

ature of 10^4 K, which is a reasonable approximation for the temperature that gas reaches when ionized by a burst of star formation. Though this is unphysical, the temperature in our simulations only enters in the calculation of the recombination rate ($A(T)$). The rationale for assuming a fixed temperature is to more easily compare differences between continuous and bursty models, as a different temperature structure in these models would change the recombination rate. This choice should marginalize a possible confounding factor in our attempt to compare the two models. In future simulations we will use the already implemented calculation of the temperature evolution in the ionized and neutral IGM.

- Halo locations: The output of the DM-only simulation includes a halo merger tree produced using AHF (Gill et al., 2004; Knollmann & Knebe, 2009), which identifies and tracks dark matter halos that are more massive than $10^6\ M_\odot$ at every 10 Myr interval of the simulation. These merger trees include the positions, masses, and unique identifiers for every halo within the simulation volume, allowing us to tag halos which have hosted star formation and restart star formation in host halos that grow significantly by mergers or accretion between two consecutive snapshots.

- Halo luminosity: We assume a luminosity function of galaxies (*i.e.*, number of galaxies per unit comoving volume as a function of galaxy luminosity at a given redshift) consistent with measurements of candidate high-redshift galaxies in Hubble and JWST deep fields. The parameters of the luminosity functions in the rest-frame UV $L_{\nu 1500}$ (at 1500 Å, modeled as a Schechter function) are taken from the FIT model presented in Kuhlen & Faucher-Giguère (2012). The rate of ionizing photon emission S_0 in each frequency bin is what we require to advance our simulations of reionization. Depending on the spectral energy

distribution of the galaxies, we use a conversion parameter between the UV luminosity and S_0 as $S_0 = 2 \times 10^{25}(L_{\nu 1500})\zeta_{\rm ion}$, where $\zeta_{\rm ion}$ is a parameter related to the hardness of the spectrum (see, Kuhlen & Faucher-Giguère, 2012). The integrated luminosities of these models, assuming a luminosity cutoff of $M < -14$, is shown in Figure 4.5.

We then determine the luminosity of individual halos through the halo-matching procedure described in Chapter 2. We assume a luminosity cutoff of $M < -14$ in the continuous case, and thus generate a list L_i of length N to match the list M_i of the N most massive halos. We then apply the same luminosity list to the N most massive halos which have not hosted star formation in the last 100 Myr, thus achieving the bursty halo matching procedure described in Chapter 2. This choice of 100 Myr periodicity in 10 Myr bursts gives us an effective duty cycle of 10%. We also assume that halos host a burst of star formation when there is a merger of two halos that are large enough to host star formation, regardless of whether they are currently tagged as hosting star formation. We made this choice to emulate the idea that colliding galaxies tend to host star formation.

- Escape fraction of ionizing photons: Our suite of simulations include two different escape fraction assumptions: a constant $f_{esc} = 0.2$, and a variable f_{esc} which is 0.5 until $z = 10$, then decreases as a power law according to $f_{esc} = 0.5((1+z)/11)^2$. We chose this variable escape fraction model to attempt to represent the fact that more of the baryonic mass in a DM halo goes into star formation at higher redshift, making the star formation more efficient and allowing for a higher escape fraction.

- Spectrum: We assume that all of the halos emit the same spectrum with two frequency

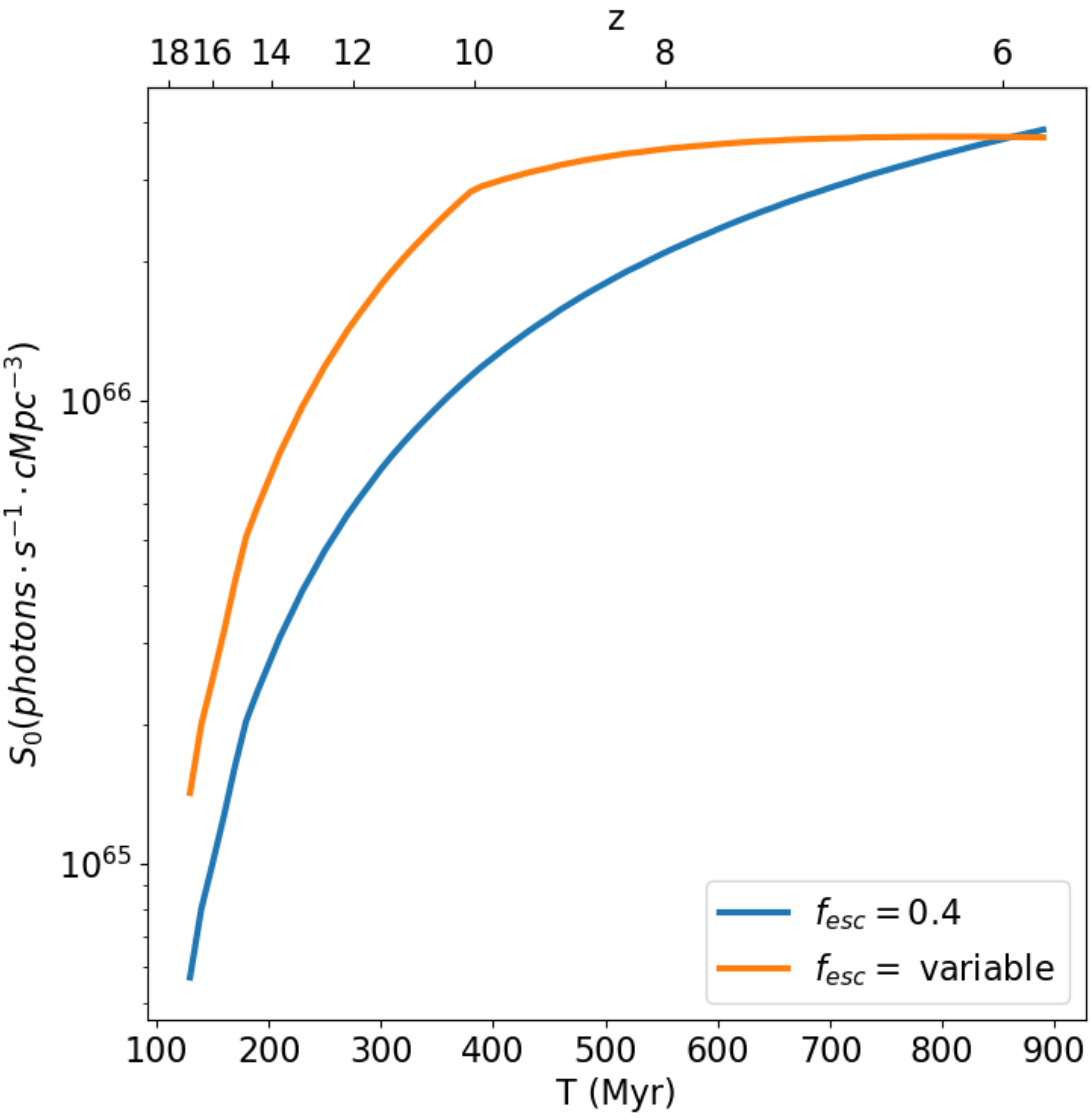

Figure 4.5: Cumulative photon count per comoving cubic Mpc for our models. The blue line represents a flat escape fraction $f_{esc} = 0.20$, while the orange line represents an escape fraction $f_{esc} = 0.50$ until $z = 10$, at which point the escape fraction drops with the inverse square of redshift.

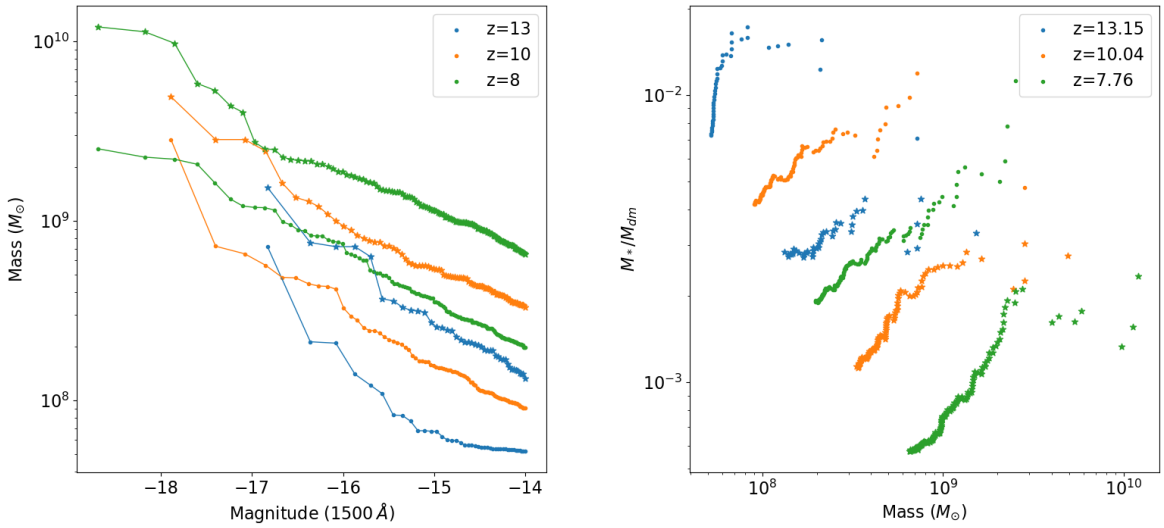

Figure 4.6: (*Left*). Typical halo mass hosting a galaxy of UV magnitude M_{UV} produced by the halo matching method. The stars represent the halos hosting continuous star formation, while the dots represent the halos hosting bursty star formation with $f_{\rm duty} = 10\%$. (*Right*). Star formation efficiency $f_* \equiv M_*/M_{\rm dm}$ as a function of halo mass associated with the our halos. Halo matching in the bursty star formation models places brighter stellar populations in less massive halos.

bins, defined by:

$$\text{Bin 0} = \begin{cases} E_0 = 13.6\,\text{eV} \\ g_0 = 0.414 \\ \sigma_0(\text{h}) = 3.007 \times 10^{-18}\,\text{cm}^2 \\ \sigma_0(\text{he}) = 0\,\text{cm}^2 \end{cases} \qquad \text{Bin 1} = \begin{cases} E_1 = 24.6\,\text{eV} \\ g_1 = 0.586 \\ \sigma_1(\text{h}) = 5.687 \times 10^{-19}\,\text{cm}^2 \\ \sigma_1(\text{he}) = 4.478 \times 10^{-18}\,\text{cm}^2 \end{cases}$$

This spectrum is based on one used in RAMSES simulations (Rosdahl et al., 2013).

- Time step: We use an adaptive time step which prioritizes keeping the error in hydrogen ionization fraction for a given time step below 2%. This results in shorter time steps at earlier times in the simulation (when density is high) and shortly after new halos first appear,

when ionization fronts are moving rapidly. Towards the end of some of our simulations, we relax this constraint to save on computational work, but also keep track of the moving average of the errors, which tend to be lower than the enforced limit.

4.3.2 Parameters Space Exploration of Simulation's Models

Our suite of simulations include three sets of runs – with two runs per set assuming either bursty and continuum galaxy formation models – with distinguishing properties as follows:

- **Set 1: High Resolution, Constant f_{esc}.** Our primary run is a 512^3 grid points cube with an escape fraction $f_{esc} = 0.2$. This escape fraction, paired with a luminosity cutoff of absolute magnitude $M_{UV} = -14$, allows for the continuous model to reionize at a realistic redshift of $z \sim 6$. This set includes a continuous and a bursty model.

- **Set 2: Low Resolution, Constant f_{esc}.** This set is the same as the first, but with a resolution of 256^3. This model was actually run first, as we tested various parameters until the time at which the models reached reionization was consistent with observations. The higher resolution simulation is more accurate, so we label it as the primary run. However, the differences between the low and high resolution simulations allow us to evaluate which aspects of the simulation are converging with increasing resolution. This set also includes a second bursty model with a tuned f_{esc}, with the goal being to reach reionization at the same time as the continuous model.

- **Set 3: Low Resolution, Variable f_{esc}.** This batch is the same as the second, but with a redshift-dependent mean escape fraction described in Section 4.3.1. Our goal with this set was to test how a higher photon emissivity at early times would affect the course of

reionization and the differences in our models. In addition, observations of the number of ionizing photons at $z < 6$ suggest that reionization was completed in a regime where the number of ionizing photons emitted is nearly balanced by the mean recombination rate, or a "photon-starving" regime (Bolton et al., 2011). Our choice of $f_{\rm esc}$ will also create these conditions at the end of the simulations.

4.4 Results

4.4.1 Set 1 - High Resolution Simulations

We begin our analysis of the results of our simulations with a set of 3D visualizations from Set 1. In Figures 4.7, 4.8, and 4.9 we plot 3D volume renderings of the continuous (left) and bursty (right) models at 6 different times throughout the simulations. The color map, which assigns a color to each ionization level, is a combination of Gaussian distributions centered at $x_{\rm H\,I} = 0.1$ (blue, high ionization), $x_{\rm H\,I} = 0.5$ (orange, medium ionization), and $x_{\rm H\,I} = 0.9$ (red, low ionization) on a log scale. This color map is provided in the bottom row of Figure 4.9 for reference. The rows correspond to times of 150, 250, and 350 Myr (Figure 4.7), 450, 550, and 650 Myr (Figure 4.8), and 750, 850 Myr (Figure 4.9) into the simulation, which gives a diverse sample of the course of reionization in both models. We also place a white dot at the locations where star formation is currently happening in every rendering.

Initially, both models are very similar. In the continuous model, H II regions generally grow monotonically as time moves on, with new bubbles appearing and overlapping as sources continue to appear. The presence of a blue shell within the red outer region is indicative of the presence of active star formation. In the bursty model, galaxies turn off after 10 Myr of activity,

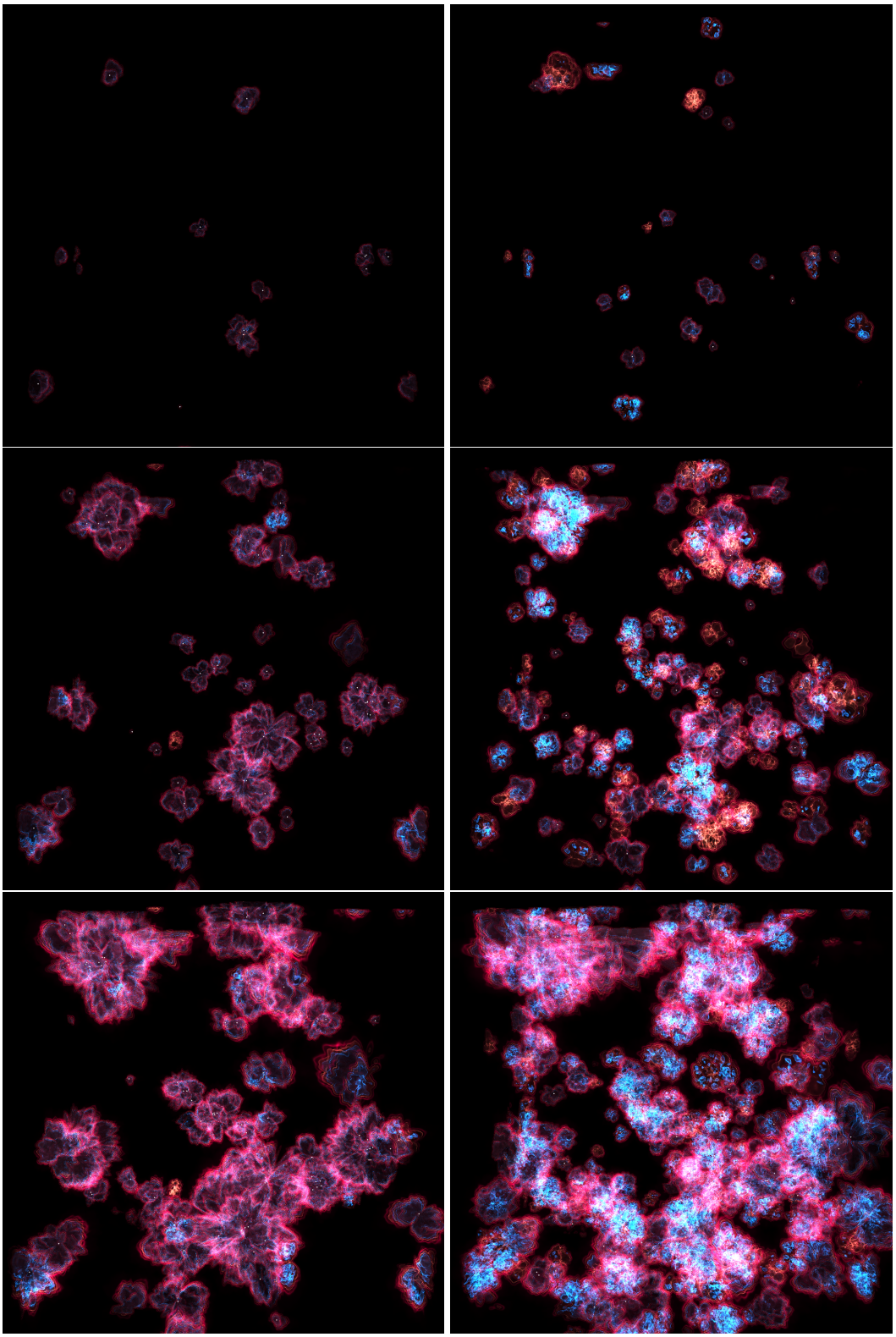

Figure 4.7: 3D visualizations from the high resolution Set 1 simulations at various times. The left and right columns correspond to the continuous and bursty models, respectively. The rows correspond to times of 150 Myr, 250 Myr, and 350 Myr into the simulations.

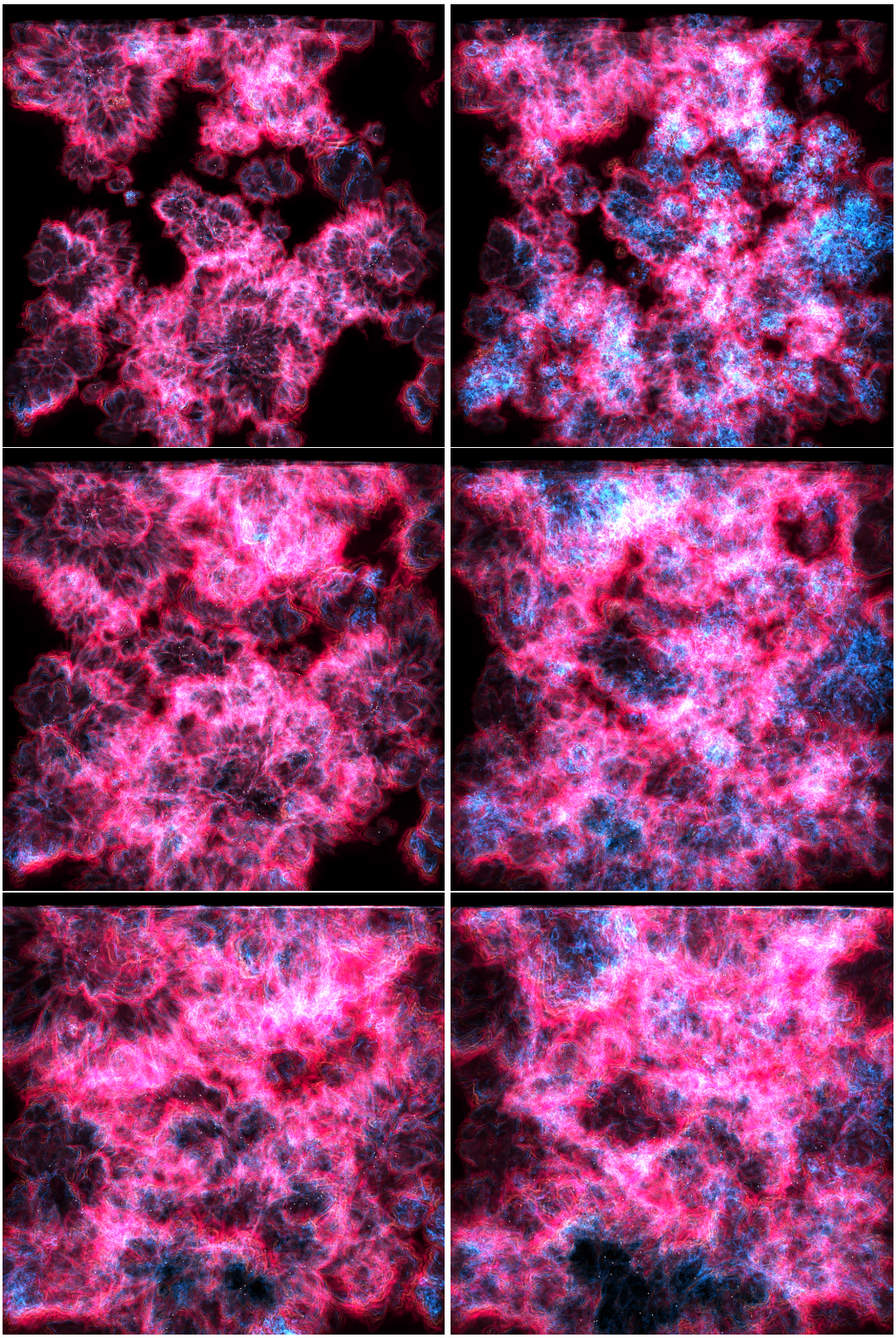

Figure 4.8: 3D visualizations from the high resolution Set 1 simulations at various times. The left and right columns correspond to the continuous and bursty models, respectively. The rows correspond to times of 450 Myr, 550 Myr, and 650 Myr into the simulations.

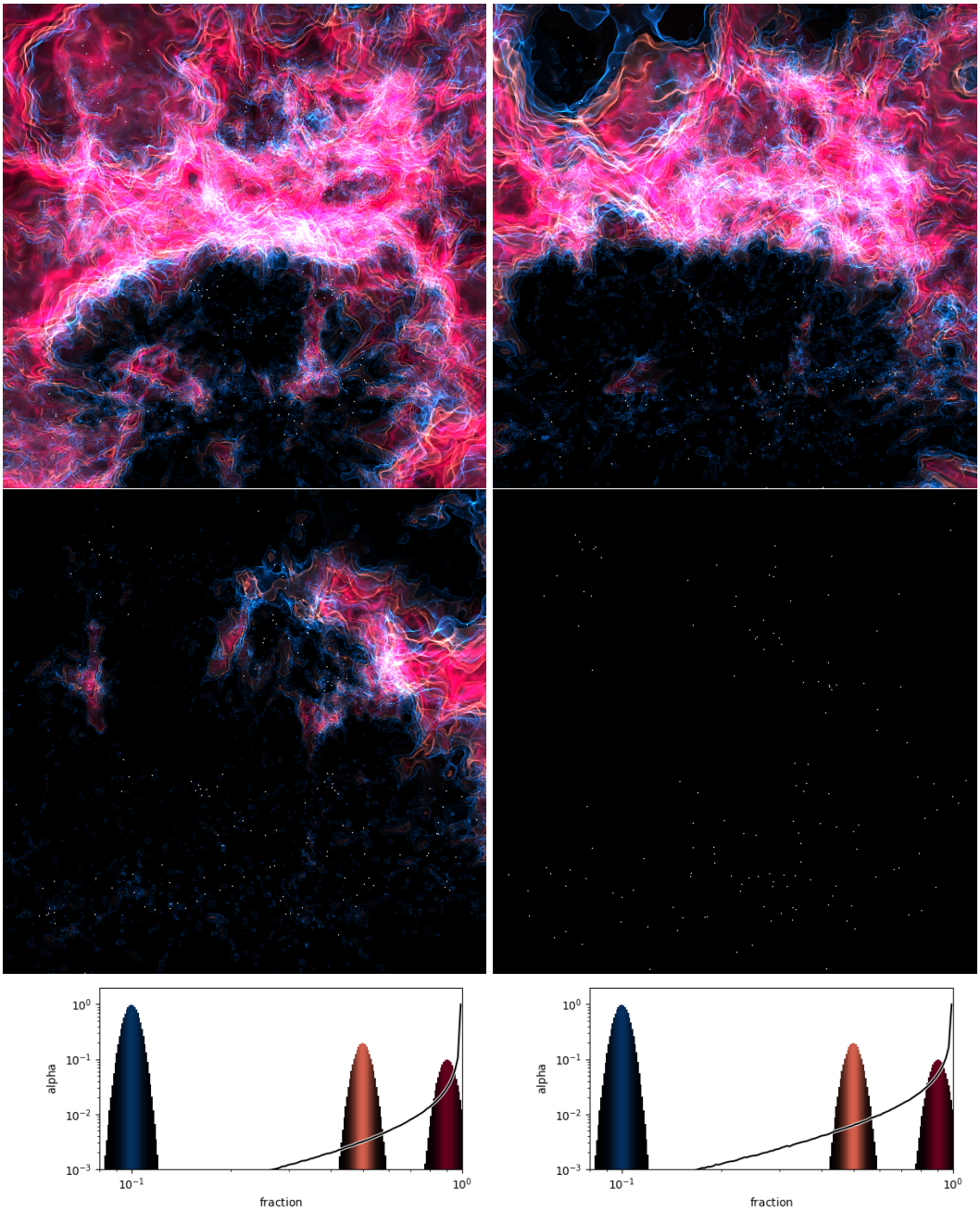

Figure 4.9: 3D visualizations from the high resolution Set 1 simulations at various times. The left and right columns correspond to the continuous and bursty models, respectively. The rows correspond to times of 750 Myr and 850 Myr into the simulations. The bottom row represents the color map transfer functions and associated histograms of neutral fraction for the 750 Myr plots.

which drives star formation into regions that aren't active in the continuous model. After 250 Myr, more small H II regions are forming, and relic H II regions that are recombining can be seen as regions pervaded by medium/orange ionization. This trend continues in the 350 Myr plots.

At 450 Myr (Figure 4.8), the continuous model is continuing to grow, while the bursty model is able to connect nearby H II regions more easily due to their abundance. The intense blue color can be more difficult to see in the continuous model because the blue, orange, and red shades typically appear together at the edge of the monotonically growing H II regions characteristic of continuous star formation. At 550 and 650 Myr, it is clear that the bursty model is ahead in overall coverage of the ionization, though this does not necessarily indicate that it will reach ionization first, as the partially ionized regions characteristic of the bursty model may take longer to fully ionize.

Finally, at 750 Myr, we can see the translucent fully ionized regions growing in both models, but more advanced in the bursty model. At 850 Myrs, the bursty model has reached full ionization, while the continuous model still has a few small regions to fill out.

Next, we examine the behavior of the quantities associated with photon conservation, namely ionized fraction rate of change ($\Delta x_i/\Delta t$), ionization rate (Γ_i), and recombination rate (A_i). In Figure 4.10, we plot these rates over the course of the simulation for both hydrogen and helium, as well as the combined rates. We notice immediately that both species, and helium in particular, exhibit transient oscillations on the 10 Myr time scale over which we update the halo properties. Much of this transient effect, particularly in the $\Delta x_i/\Delta t$ and Γ_i plots, is the result of the ionization fronts crossing individual cells on timescales on the same order as the time step, with the differences exaggerated by the difference in energy and travel distance of the different photon frequency bins. We also reiterate that a deviation in the helium fraction is of relatively

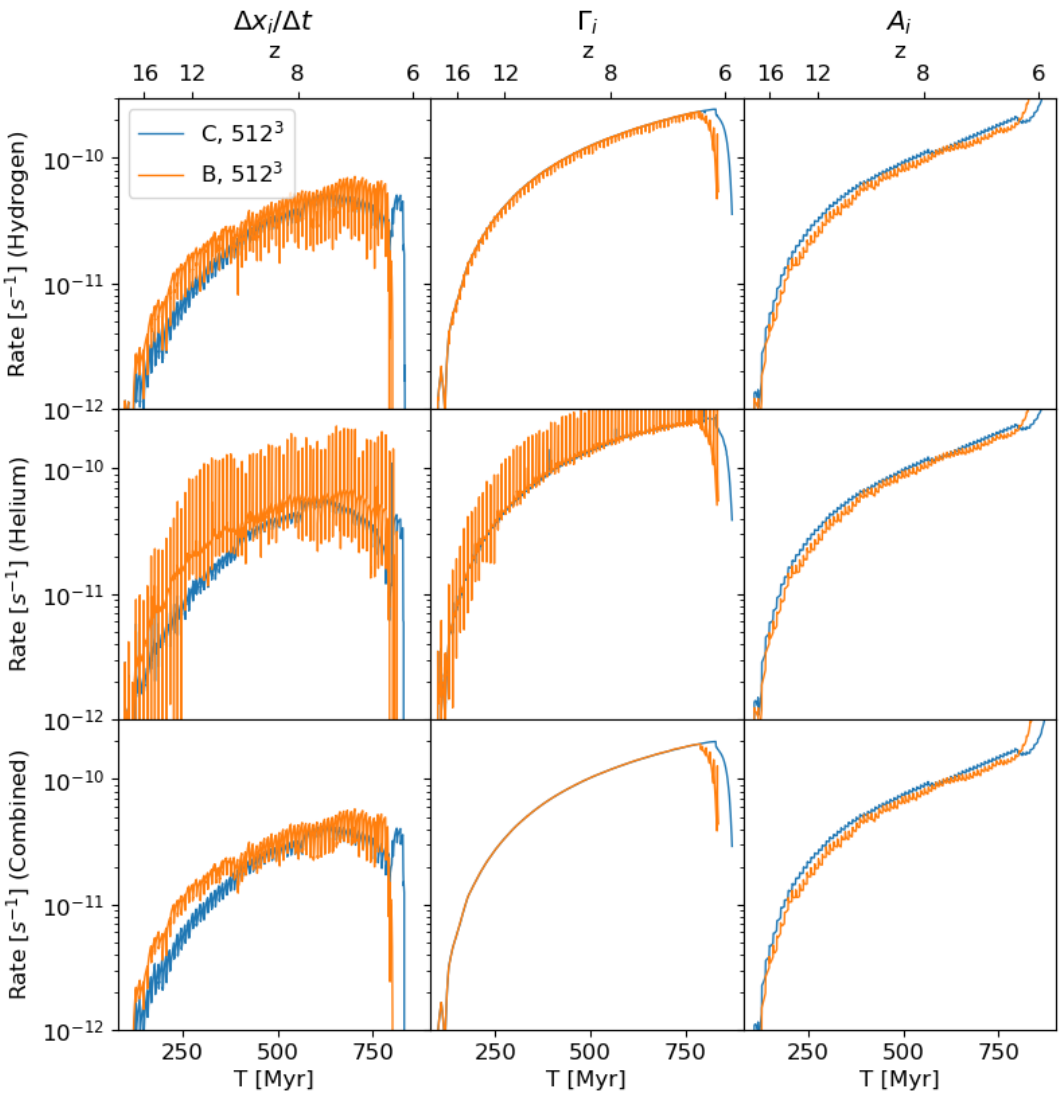

Figure 4.10: Ionization rates of change, ionization rates, and recombination rates for hydrogen, helium, and combined over the course of the simulation for both simulations.

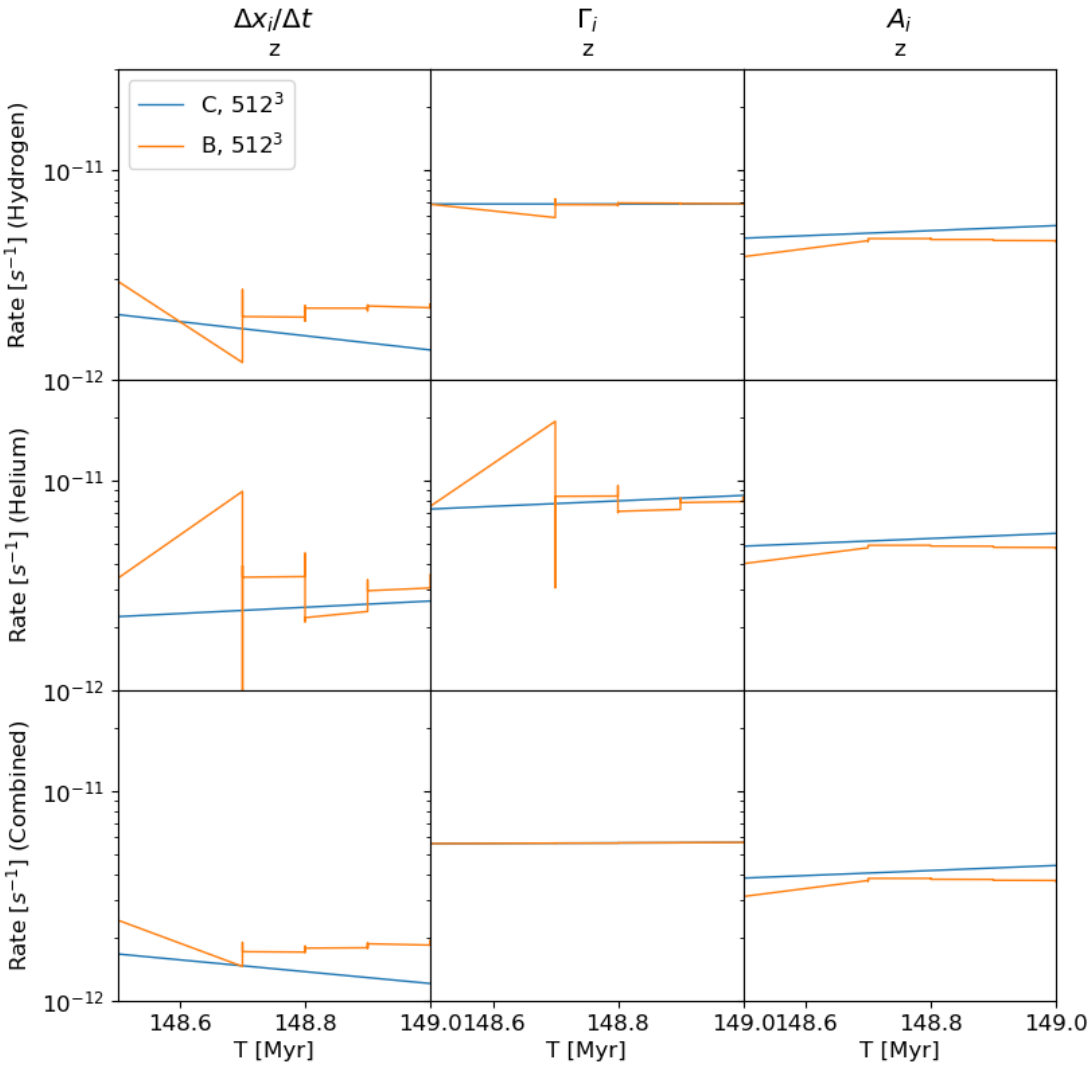

Figure 4.11: Ionization, absorption, and recombination rates for hydrogen, helium, and combined from 120 Myr to 170 Myr

small significance due to the properties and relative scarcity of helium. The combined results show that these transients are mostly an oscillation between ionization of hydrogen and helium as the front travels. The remaining oscillation is the result of the discontinuous behavior of our halo position and luminosity assignment scheme. Aside from these numerical issues, the main result shown in these plots is that the bursty and continuous models share the same ionization rate (as expected by construction), while the bursty model has a lower average recombination rate, resulting in a higher rate of increase of the gas ionized fraction. The rate of change of the ionized fraction and the ionization rate both drop by several orders of magnitude when reionization is completed, and it can be easily seen from these plots that the bursty model completes reionization first.

We also plot the same quantities over a smaller interval time (between 120 Myr to 170 Myr) in Figure 4.11. In addition to showing all of the same local properties discussed for the plots for the entire reionization history, these plots show more clearly that the transient behavior persists for only brief periods.

Next, we examine the behavior of the adaptive time step used in both simulations. The left plot of Figure 4.12 shows the (adaptive) time step in Myr, that was required to maintain the 2% error in hydrogen ionization fraction change rate throughout both simulations. We see that the introduction and redistribution of sources among halos every 10 Myr dominates the overall behavior of this quantity. The dips themselves tend to be less deep at later times, as the overall density of the universe decreases, the optical depth of a given cell decreases, and the error becomes easier to manage. The bursty model generally requires shorter time steps, which is in line with our expectation that the bursty model will place new halos in neutral regions much more frequently than the continuous model. The continuous model also reaches the step length cap of

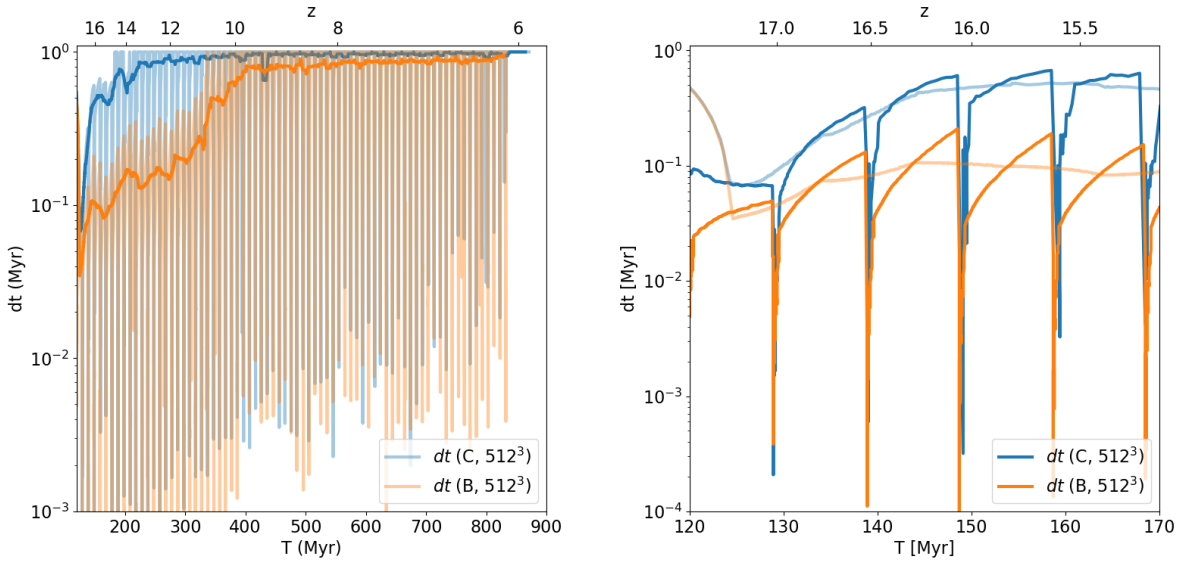

Figure 4.12: (*Left*). The adaptive time step used for the entire time span of both simulation runs. General trends such as longer time steps at later times and the drop in time step every 10 Myr can be seen. (*Right*). A smaller slice of the left plot. We can see that the required time step to maintain a fixed error drops dramatically when halos move or are introduced, but rises gradually as the H II regions grow. In both plots, we also show the moving averages over a 10 Myr interval to smooth out the noise.

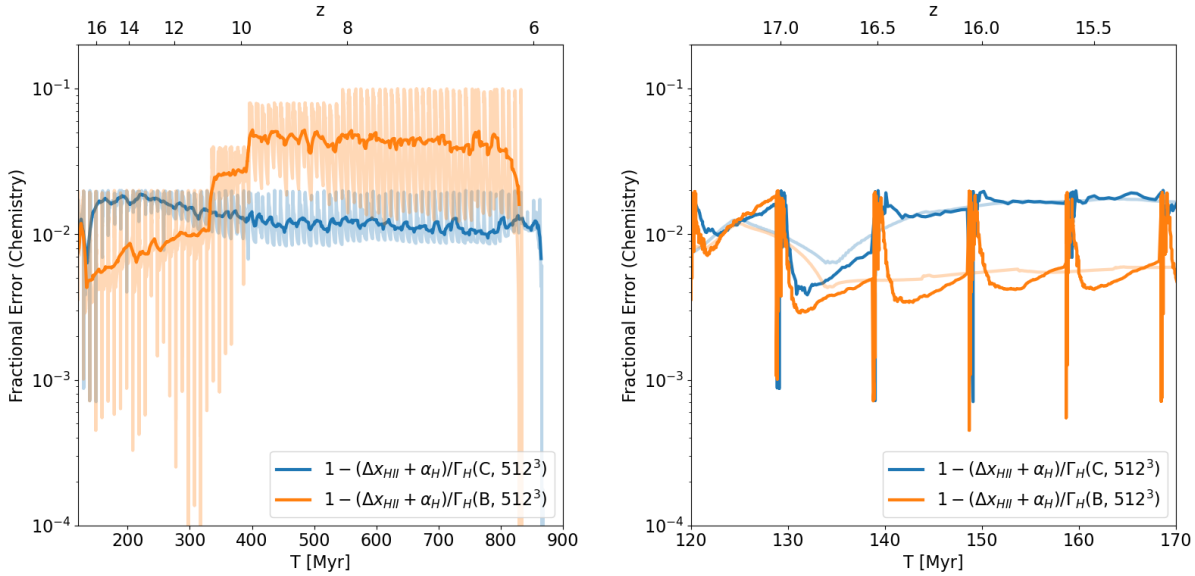

Figure 4.13: (*Left*). The error in H II associated with the time steps taken in both simulations (which are shown in Figure 4.12). At late times we relax the constraint on the bursty model. (*Right*). A smaller slice of the left plot. In both plots, we also show the moving averages over a 10 Myr interval to smooth out the noise.

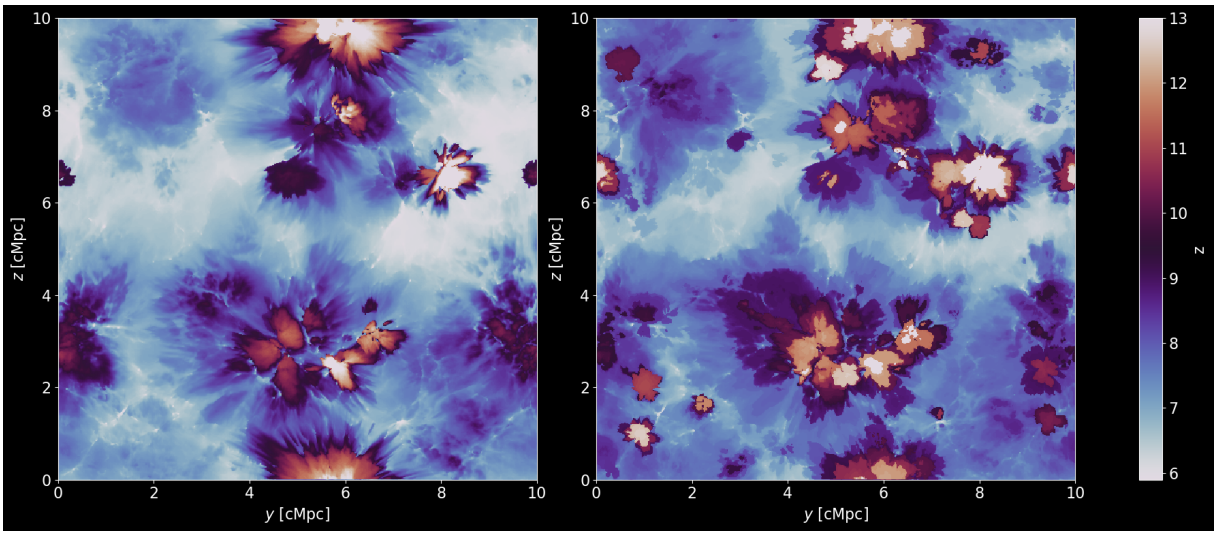

Figure 4.14: A visualization of the redshift of reionization (specifically $x_{\rm H\,II} > 0.9$) for a slice through the continuous (*left*) and bursty (*right*) models, respectively. In the continuous model, gas generally reionizes radially outwards from the brightest sources, while in the bursty model there smaller and more spread out regions of ionization.

1 Myr earlier and more frequently than the bursty model. In the right plot of Figure 4.12, we plot a zoom in to a smaller segment of the left plot, which shows that both models exhibit similar behavior.

In the left plot of Figure 4.13, we plot the fractional error for the hydrogen ionization rate associated with the time steps shown in Figure 4.12. While there are significant oscillations over the thousands of time steps, we see that the hard limit of 2% is being maintained. In the right plot of Figure 4.13, we plot a zoom in to the same time interval as the right plot of Figure 4.12. The smaller error in the bursty model is an artifact of the adaptive time step algorithm.

With a detailed understanding of how the models evolve, we can now examine how and when both models reach reionization. In Figure 4.14 we plot the redshift at which each point in a slice through both simulations reaches an ionization of level of $x_{\rm H\,II} > 0.9$, which allows us to visualize the entire course of reionization for that slice. In the plot of the continuous model (*left*), we see that ionization propagates outwards from a handful of locations, extending outwards into

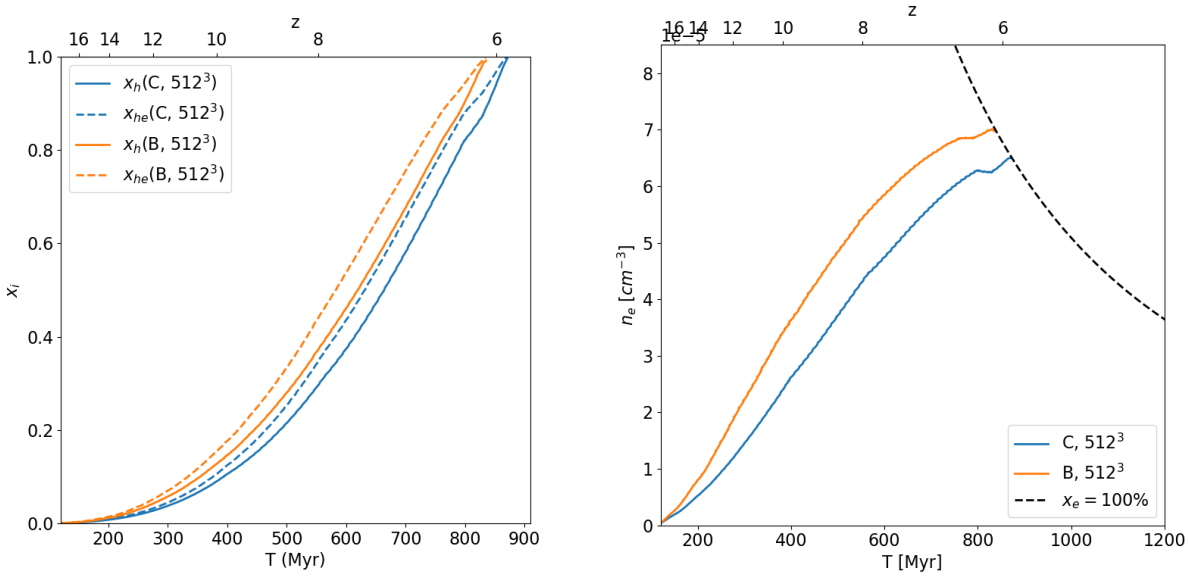

Figure 4.15: (*Left*). Overall mass-weighted ionization rates of hydrogen and helium. (*Right*). Physical density of free electrons.

the voids until the entire volume is filled. In the plot of the bursty model ((*right*), we see that the reionization begins in roughly the same places, but the behavior is very different away from these areas. Smaller pockets of reionization from bursts of star formation appear at various times, which drives the overall course of reionization towards earlier completion.

Finally, we plot the mean ionization fractions (for hydrogen and helium) as a function of time in the left plot of Figure 4.15. We see that the helium ionization fraction leads over the hydrogen ionization fraction, which is a result of harder photons with longer mean free paths ionizing helium more effectively. We then combine the free electron density from both hydrogen and helium ionization and convert them to physical units, as opposed to comoving units, to determine the physical electron number density as a function of time (and redshift), which we plot in the right panel of Figure 4.15. Later, we will use this result to calculate and compare the optical depth to Thompson scattering $\tau_{\rm CMB}$ (*i.e.*, scattering between CMB photons and free electrons) between different models.

4.4.2 Set 2 - Low Resolution Simulations

Our second set of runs is the same as the first in all ways except for the resolution, which is halved to become 256^3. In Figure 4.16 we plot the combined conservation quantities for these models against the same plots from Set 1, comparing the continuous and bursty models of both sets independently. We see that the ionization rates are nearly identical between the simulations, which we expect due to the halo luminosity function being identical between the two models. On the other hand, the recombination rate is higher in the higher resolution simulation, a difference which is more pronounced in the continuous model. This is the result of the method we used to produce the density field, which is essentially a moving average over the DM particles. The higher resolution density field has a smaller effective window size for the moving average, which allows for higher densities in the vicinity of DM haloes. This results in a higher average recombination rate near larger halos. This effect is present in both models, and is more pronounced in the continuous model, where star formation happens exclusively in the most massive halos.

With this set of lower resolution simulations, we also ran a third simulation with a lower escape fraction. Our goal with this run was to see if we could complete reionization using a bursty model at the same time as a continuous model with similar parameters but a higher relative escape fraction. We used analytic models of ongoing bursty models to predict a value of $f_{esc} = 0.15$ would roughly compensate for the difference. In Figure 4.17, we plot the results of this simulation along with the other simulations from Set 2. In the top plot, we plot the photon conservation rates, and we see that the ionization rate is lower, as expected from lowering the escape fraction. The recombination rate also decreases to a level lower than the other two simulations. However, the ionization rate of change keeps pace with the continuous model, which is a result of the more

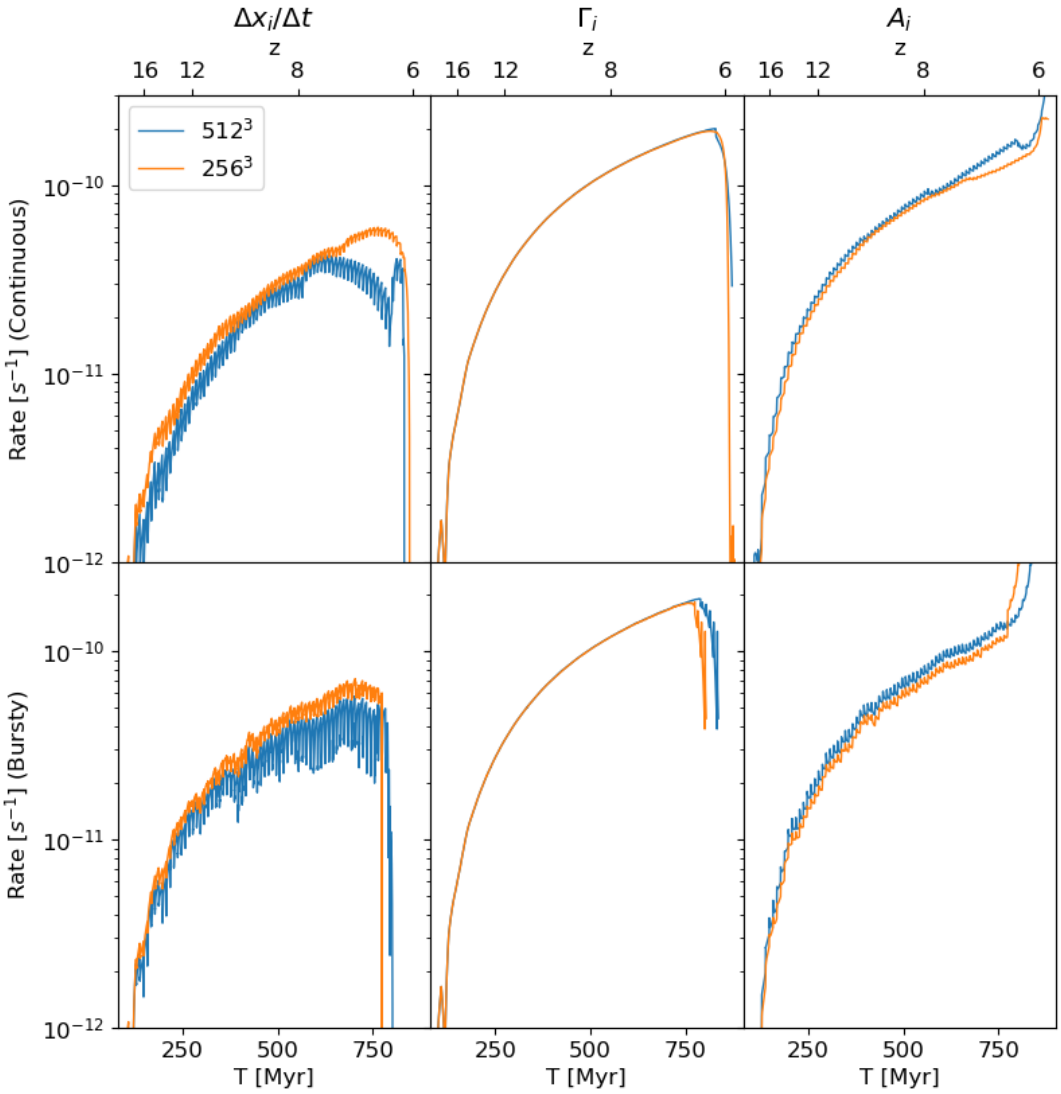

Figure 4.16: Ionization rates of change, ionization rates, and recombination rates for hydrogen, helium, and combined over the course of the simulation for both simulations plotted alongside the same quantities for the higher resolution simulations.

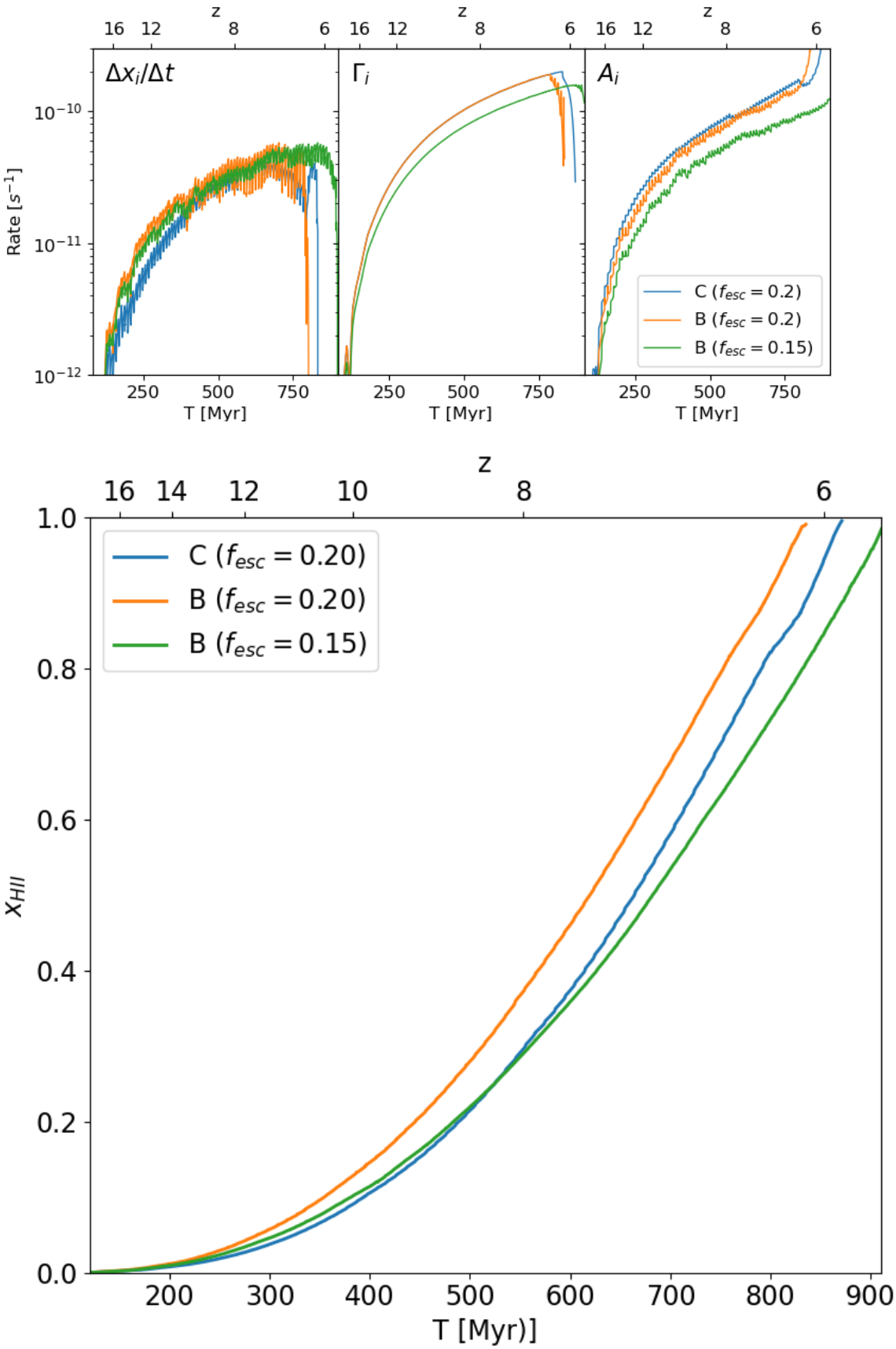

Figure 4.17: (*Top*). Ionization rates of change, ionization rates, and recombination rates for the Set 2 models. The bursty model has a lower escape fraction, which results in lower ionization and recombination rates, but a similar ionization rate of change. (*Bottom*). Hydrogen ionization over the course of the simulation for these models.

efficient use of photons in the bursty model. In the bottom plot of Figure 4.17, we see that the bursty model keeps pace with the higher emissivity continuous model, thus demonstrating the effect we hoped to see.

4.4.3 Set 3 - Variable escape fraction

Our third set of runs is identical to the second, with a variable escape fraction. The motivation of this choice in parameters is twofold.

1. The time evolution of the luminosity function is different from that of the mass distribution of DM halos. Over the course of our simulations, the mass of DM halos grows faster than the mass of stars in those haloes, meaning that the ratio of stellar mass to dark matter mass $M_{\odot}/M_{DM}$ tends to decrease with time. This can be seen clearly in Figure 4.6, where the halos of a given star formation mode at higher redshift are higher on the y-axis than those at lower redshift. Star formation in lower mass halos may then have less neutral gas to penetrate before entering the IGM, which would result in a higher escape fraction.

2. At higher redshift the density of the IGM is higher, which means that the recombination rate is correspondingly higher. As discussed in § 2, higher recombination rates may accentuate the differences between a bursty and continuous model of star formation.

In Figure 4.18 we plot the combined rates associated with this model alongside those from Set 2. The ionization rate Γ_i ramps up faster at early times before plateauing at $z \sim 10$. The recombination rate increases monotonically as the gas is ionized, and we see that its growth also slows at $z \sim 10$. The rate of change of the ionized fraction actually begins to decrease at the same time, which is expected when the ionization rate plateaus while the recombination rate continues

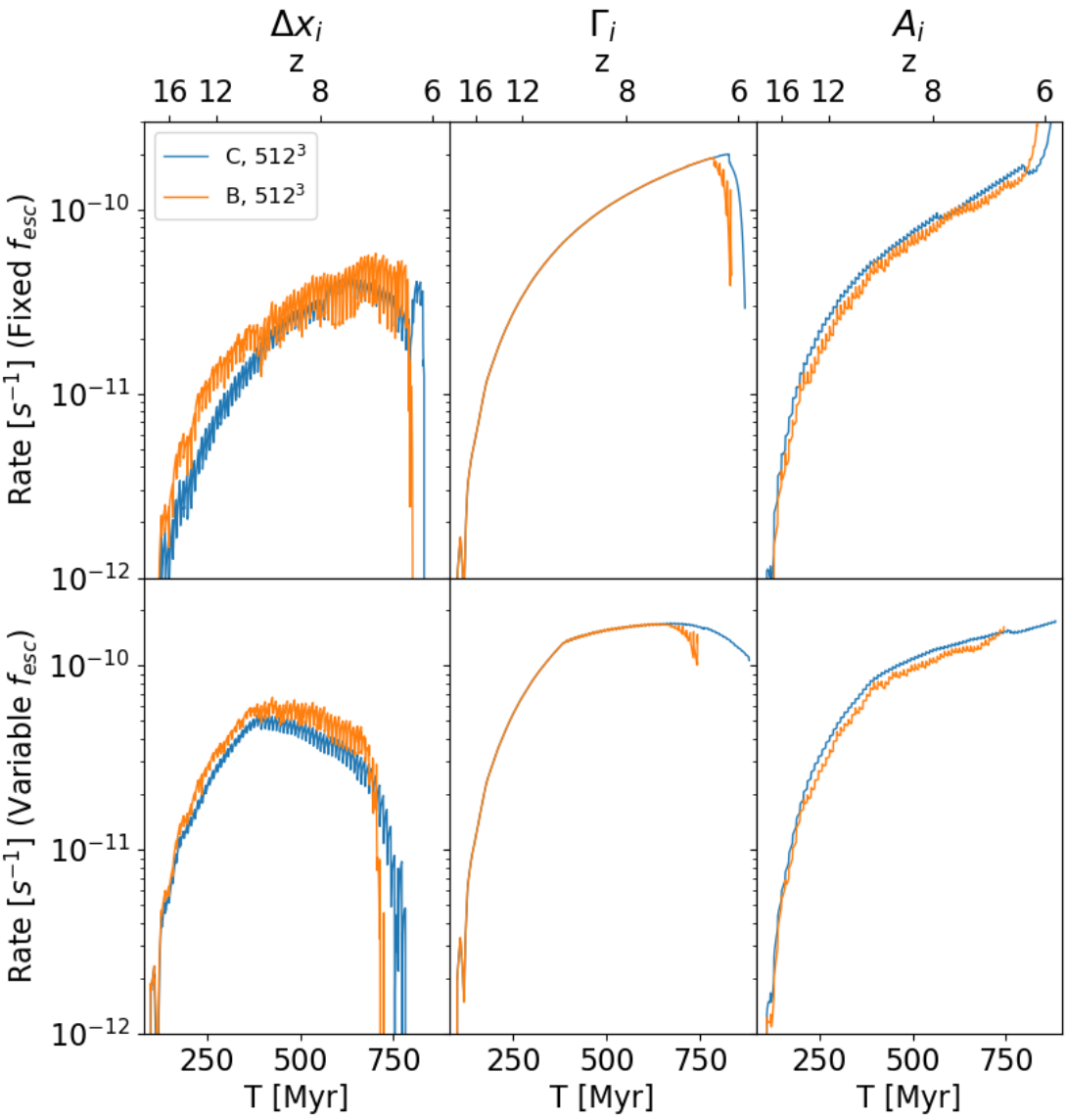

Figure 4.18: Ionization rates of change, ionization rates, and recombination rates for hydrogen, helium, and combined over the course of the simulation for both simulations.

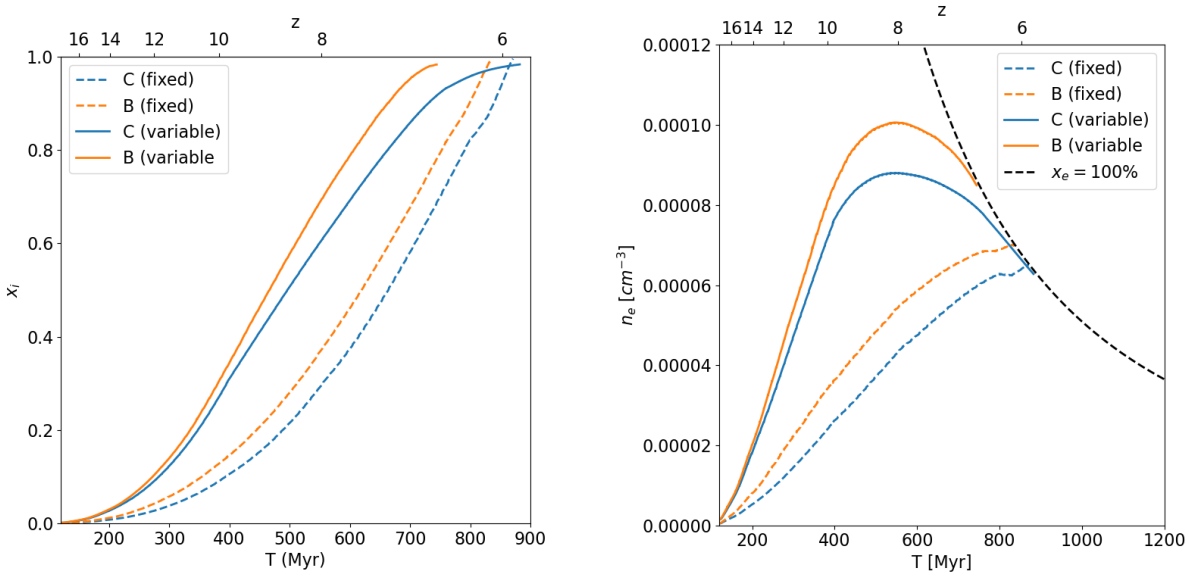

Figure 4.19: (*Left*). Overall mass-weighted ionization rates of hydrogen and helium. (*Right*). Physical density of free electrons.

increasing.

We show the overall course of reionization for these models in the left plot of Figure 4.19. We see that the fixed and variable f_{esc} models diverge early, with the variable models growing quickly before an inflection point at $z \sim 10$. Both continuous models achieve reionization at roughly $z \sim 5.9$, while bursty variable f_{esc} reaches reionization even earlier than the bursty model with fixed f_{esc}. In the right plot, we see that the density of free electrons is significantly higher than in the variable f_{esc} model, which means that these models will have correspondingly larger $\tau_{\rm CMB}$ values.

4.4.4 Physical Results

Although we have full information about the state of the IGM at all locations and all times in our simulations, most of this information is impossible to compare to current astrophysical observations. However, there are two observational quantities that we can use to evaluate the results

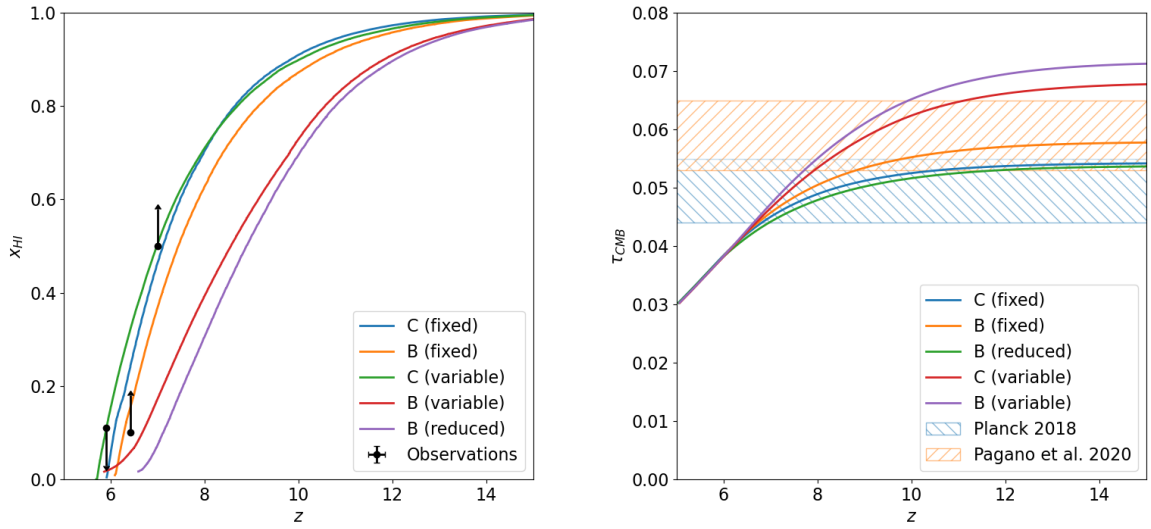

Figure 4.20: (*Left*). Overall mass-weighted ionization rates of hydrogen and helium. (*Right*). Optical depth of reionization.

of our simulations: high redshift hydrogen neutral fraction and the optical depth of Thompson Scattering for the CMB.

In the left panel of Figure 4.20, we plot the neutral fraction from our simulations beginning at $z = 15$ until reionization is complete, along with the same observational data presented in Table 2.1 of Chapter 2. Here, we clearly see the effect of bursty star formation, which for a given model pushes the neutral fraction down and the completion of reionization back to earlier times.

In the right plot of Figure 4.20, we plot the $\tau_{\rm CMB}$ for our models against two different observational estimates for this quantity. $\tau_{\rm CMB}$ is essentially the integral of the free electron density over time since recombination. Before reionization there are few free electrons, and the total ionization of the IGM after reionization makes the contribution from after $z = 5$ well known to be $\tau \sim 0.03$. We then use the results of our simulations to complete this calculation, which we show as the integral beginning at $z = 5$. The final result is the value that these integrals reach at $z = 15$. The hatched regions represent two different observational estimates of this quantity.

Here again we see that the bursty model pushes this quantity to higher values, with the region from $z \sim 7$ to $z \sim 12$ being most relevant.

4.5 Conclusion

In this chapter, we presented the distillation of a vast amount data produced over months of computation requiring hundreds of thousands of SU (Service Units, equivalent to an hour of wall time for a single CPU) by our simulations. These samples represent what we believe to be the most useful for exemplifying the important features of our exploration of continuous and bursty star formation. These features include:

- Bursty star formation spreads ionizing sources more evenly throughout the computational volume, allowing for a more even distribution of ionized gas. By forcing the stars to form away from places where there has been past star formation, sources with a given stellar mass tend to be able to form in less massive DM halos.

- Recombination happens less rapidly in lower mass DM halos, which means that fewer photons are wasted maintaining ionization in high density regions, and ionization happens more efficiently. The result of this is a neutral fraction which drops more quickly, an earlier reionization completion time, and a higher optical depth of the CMB to Thompson Scattering. This also means that a bursty model can complete reionization at the same time as a continuous model with a higher escape fraction.

- Our higher resolution simulations had higher fidelity in the density of DM, which meant higher recombination rates near the center of DM halos. This effect is more pronounced

near more massive halos, meaning it would tend to emphasize the effects we wanted to study.

- Increasing the ionizing emissivity at early times speeds up the process of reionization, but doesn't necessarily result in a larger difference between the continuous and bursty models.

Our simulations have only scratched the surface of what is possible with ARC. As we continue our work, we will further explore simulations with different parameters to understand how the effects we have described here behave at different scales, in different environments, and with different underlying assumptions. We will also work to make ARC more publicly accessible, with the goal being to allow other scientists to explore ideas relating to radiative transfer with high accuracy and speed.

Chapter 5: Conclusion and Future Work

5.1 Summary and Conclusions

This thesis presented our work towards demonstrating the idea that bursty star formation is a more efficient driver of reionization than continuous star formation. In Chapter 2, we discussed the motivation for beginning this study, which was based on a semi-analytic model built on Professor Ricotti's 1D radiative transfer simulation. These simulations showed us that, in theory, the reduced recombination rate in relic H II regions would lead to an overall increase in the speed of reionization. This observation provided the motivation for us to develop a more robust approach for studying these effects.

In Chapter 3 we presented ARC, the library we developed to test these ideas to a much higher accuracy and fidelity. ARC uses multiple levels of parallel computing: the computational volume is split between CPU nodes, and the ray tracing and chemistry calculation on individual nodes are split between the thousands of cores available on each independent GPU. Both levels of parallelization required careful handling of memory and messages between nodes and cores, but the overall effect is a fast and efficient ray tracing algorithm. We also demonstrate that this library is accurate to modern standards, and that its speed scales as desired with the use of GPU computing.

In Chapter 4 we presented results of our explorations into the ideas motivated by Chap-

ter 2. First, we discussed the additions and improvements made to ARC which lead to improved simulations. Next, we presented the suite of simulations we ran on ARC, which tested bursty and continuous star formation for different choices of $f_{\rm esc}$ and simulation resolution. We found that, for our choice of a 10% duty cycle of bursty star formation, reionization would complete roughly 60 Myr sooner, or could complete reionization at the same time with an escape fraction that is roughly 75% lower. We also found that bursty star formation leads to a higher optical depth of reionization to Thompson scattering, and a more even distribution of H II regions in the IGM during the reionization process.

5.2 Future Work

The work presented in this thesis represents the motivation, development, testing, and demonstration of a flexible and fast library for performing cosmic simulations at a variety of scales. Our next goal for this work is to package and annotate ARC in such a way that the scientific would be able to compile, configure, and run their own simulations without assistance. We believe that there are many questions that ARC is uniquely situated to investigate, as mentioned below, and handing it off to the scientific community is the first step towards making this a reality.

ARC is already available for access on GitHub at ARC with a README that would allow someone familiar with these types of simulations and accustomed to working on a supercomputer to run the library without assistance. Moving forward, we plan to maintain this database and update it as we continue running simulations and improving the base library.

We also find that it would not be difficult to continue the work towards examining star formation mode on the EoR using our library. Though there are endless possibilities, we present

a few that would likely be next steps along our current trajectory:

- Non-uniform emission/jets: all of the simulations we presented in Chapter 4 used spherically symmetric emission from all sources. However, the HEALPix foundation of ARC allows users to control the spherical profile of emission to any desired accuracy within memory constraints. Non-uniform/non-symmetric emission is something that astronomers believe to be likely.

- Updated simulation parameters: the simulations presented in this thesis were constrained by time and computer allocation considerations. We were continuously iterating on the library and performing runs until eleventh hour. Moving forward, we would like to run more simulations with tighter accuracy constraints, different values of $f_{\rm esc}$, different light spectra, more thermal physics, and different duty cycles of bursty star formation.

- Simulations on different scales: our simulations were also at a fixed scale of 10 cMpc. This volume, while large, is not necessarily large enough to represent a truly definitive statistical sample of a cosmic volume during the EoR. Thus, we would like to run simulations at larger and smaller scales, to see how a larger representative/more accurate simulations, respectively, would behave.

- Simulations with more physics: The simulations presented in Chapter 4 are essentially radiative transfer post-processing simulations on the DM-only simulations we used as inputs, so we would also like to incorporate more physics into future simulations.

A.1 Analytic Approximations

A.1.1 Ionization Profile During Star Formation

A simple model for the shape of the ionization profile may be found by solving the following formula iteratively:

$$x_e^2(u) = -\frac{1}{3u^2}\frac{d}{du}\exp\left(-\tau_0\int_0^u (1-x_e(u'))du'\right), \tag{1}$$

where $u = r/R_s$ is the dimensionless distance from the source and R_s is a scale length of the Strömgren sphere and $\tau_0 = n_{\rm HI}\sigma_\nu({\rm H\,I})R_s$ is the optical depth of neutral hydrogen for a column density of $N_{HI} = n_{\rm HI}R_s$. From this formula we see that τ_0 represents a scale length in the u domain for the drop in electron fraction. While this model is fairly precise, we require a simple analytic formula to construct our statistical model. We may, however, use the above formula to note that both since $\tau_0 \propto N_{HI}$, the width of the ionization profile in comoving coordinates should scale with N_{HI}.

A broad analysis of our suite of simulations showed that the ionization profile around a test halo during star formation was well fit by the following class of functions:

$$x_e(R) = \frac{1}{1+\left(\frac{R-A}{B}\right)^{1/C}} \tag{2}$$

$$R(x_e) = A + B\left(\frac{1}{x_e}-1\right)^C. \tag{3}$$

The fit of Equation (2) approximates the electron function as a function of radius for a fixed

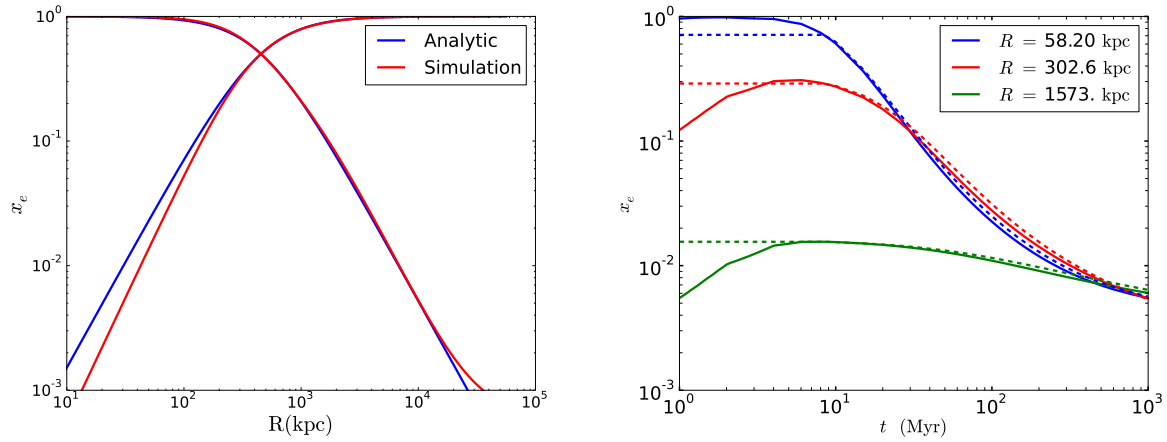

Figure 1: (Left) An example of the simple analytic fit to the ionization profile outside a halo. The red lines represent the neutral and ionized fractions produced by the numerical simulation. The blue line represents the best fit of the analytic model. (Right) An example of the recombination model compared to the numerical simulation. The solid lines represent the electron fraction at four fixed radii as a function of time from the numerical simulations. The dashed lines represent the behavior of the electron fraction as described by (7).

moment in time. The parameters A, B, and C may be interpreted as a distance offset, a scale factor, and a power law, respectively (this sentence needs work). These parameters are, in principle, functions of the Hubble time, time since star formation began, and luminosity of ionizing photons. An example of one such fit is given in Figure 1(top). We have included the inverse of the model fit because our simulation calls for the radius at which the ionization reaches a given level, as given by Equation (3).

An analytic model for the time dependence of the scale radius R_s of a Strömgren sphere in an expanding universe is presented in Donahue & Shull (1987) (see also Shapiro & Giroux

(1987)). Explicitly:

$$R_s(t) = (21\ \text{kpc}) \left[\left(\frac{S_0}{10^{49}\ \text{s}^{-1}} \right) \cdot f_1(t) \right]^{1/3} \tag{4}$$

$$f_1(t) = \frac{(8.95 \times 10^5)\lambda}{(1+z_0)^3} \exp\left(\frac{\lambda}{t_c}\right) \times \left[t_c E_2\left(\frac{\lambda}{t_c}\right) - E_2(\lambda) \right] \ \text{Myr} \tag{5}$$

$$\Delta R_s(t) = (3.2\ \text{Mpc}) \left(\frac{1+z}{10}\right)^{0.7} \times \left(\frac{S_0}{10^{49}\ \text{s}^{-1}} t \right)^{0.18} t_c^{-0.466} \tag{6}$$

Here, $\lambda = t_H/t_{\text{rec}}$, $t_c = 1 + t/t_{\text{rec}}$, and $E_n(x)$ is the exponential integral of the nth order. This more complicated model is needed to replace the simple Strömgren sphere model because of the Hubble expansion freezes recombinations when the recombination time becomes longer than the Hubble time.

We found that the parameters (A, B, C) of Eq. (3) are well fit using the following formula:

$$A(t, S_0) \propto S_0 \cdot R_s(t)$$

$$B(t, S_0) \propto S_0 \cdot \Delta R_s(t)$$

$$C(t, S_0) = 2.0.$$

A.1.2 Ionization Profile After Star Formation

When star formation ends, the ionized gas begins to recombine. This process is modeled by the differential equation $\dot{n_{\text{HI}}} = (\alpha n_e) n_{\text{HII}}$. Assuming a purely hydrogen medium, this equation

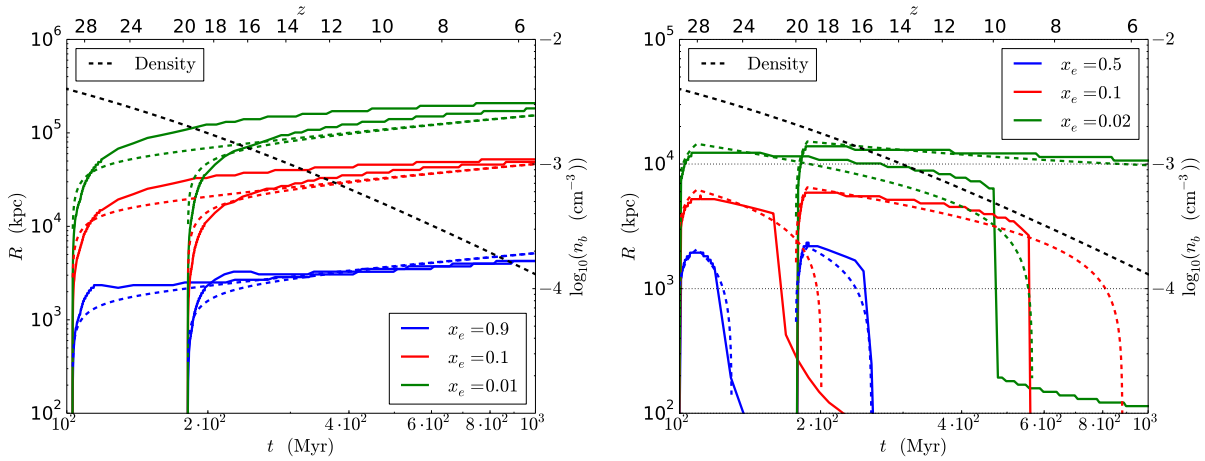

Figure 2: Full analytic model for ionization profile for (left) continuous and (right) instantaneous SF against the numerical model for at initial redshifts $z_0 = 30, 20$. The solid lines represent the radii at which a fixed electron fraction is found for the numerical model. The dotted lines represent the corresponding analytic fits. (left) For the continuous case, deviations observed are consistent with imperfections in the model as documented in Donahue & Shull (1987). (right) For the instantaneous case, deviations occur due to the spatial dependence of temperature.

is solved in an expanding universe by the following equation:

$$x_e(R,t) = \left(\int_{t_0}^{t} \alpha^{(2)} n(t) dt + \frac{1}{x_e(R,t_0)} \right)^{-1}. \tag{7}$$

The correction for this equation for the more general case of helium and hydrogen is well approximated by an multiplicative constant outside the integral that represents the extra electrons from ionized helium. The integral, in general, is impossible to solve without an analytic expression for the temperature. However, the recombination coefficient α depends weakly on temperature ($\alpha \propto T^{-0.7}$), we may assume α is constant as an approximation. The integral may then be solved explicitly:

$$\int_{t_0}^{t} n(t')dt' = \frac{2n_0\Omega_b}{3H_0\Omega_m^{1/2}} \left[(1+z_0)^{3/2} - (1+z)^{3/2} \right]. \tag{8}$$

This formula provides an excellent description of the time domain behavior of the electron fraction. A comparison between this model and a radiative transfer simulation is shown in Figure 1(bottom).

In the case of continuous star formation, the initial profile expands according to the time dependent scale in equation (4). An example of the resulting fit is plotted in Figure 2 against two examples of continuous star formation produced from the suite of numerical simulations. The analytic model captures both the transient behavior as the sphere initial expands and the long term steady expansion of sphere as recombinations freeze out with the expansion of the universe.

In the case of instantaneous star formation, the initial profile expands according to equation (4) while the brightest stars remain alive. After the luminosity begins to plummet as massive stars die, the ionized region begins decaying according to equation (7) and equation (8). An example of the resulting fit is plotted in Figure 2 against two examples of instantaneous star formation produced from the suite of numerical simulations. The analytic model captures both the growth and decay phases of the region. The largest deviation is the time taken by the outer regions to decay as compared to the numerical model; this is due to the dependence of temperature on radius.

A.1.3 Time-dependent Spectral Energy Distribution

A numerical model for the evolution of the ionization in a region around a source of ionization is presented in Ricotti et al. (2001). This numerical model takes a spectral energy distribution (SED) and outputs the density profile of neutral and ionized hydrogen (as well as several other elements) as a function of time.

In our statistical simulation, we take a star forming halo to be a point source of ionization. We make use of the code from Ricotti et al. (2001) to simulate the behavior of the IGM around a model point source for a suite of (S_0, z_0) values, where S_0 is the rate of ionizing photon production and z_0 is the redshift of initial star formation. The SED used for these simulations are produced using the Starburst99 code Leitherer et al. (1999) with the lowest possible metallicity ($Z = 0.001$).

The suite of (S_0, z_0) values was chosen to be large enough to cover the range of relevant luminosities and redshifts in our statistical simulation. The number of values in each range was increased until the change in behavior between adjacent luminosity and redshift points changed continuously and predictably.

A.1.4 Simulation Details

The simulation is divided into 72 intervals with $\Delta T = 12.5$ Myr. The start of each time interval is denoted T_i. At T_0, a number k of halos with mass M_{added} are added following a Poisson distribution with average masses extracted from a Sheth-Tormen distribution Press & Schechter (1974); Sheth & Tormen (2002):

$$f(k;\lambda) = P(M < M_{\text{added}} < M + dM) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{9}$$

$$\lambda dM = N(M_{\text{added}}, T_0)dM \quad \text{(Sheth-Tormen)}. \tag{10}$$

For later times, halo counts are drawn by the same Poisson process with $\lambda = N(M_{\text{added}}, T_i) - N(M_{\text{added}}, T_{i-1})$. Thus, at any time T_i, the total population is drawn from Poisson distribution with $\lambda = N(M_{\text{added}}, T_i)$, consistent with the Sheth-Tormen model.

A.1.5 Convergence Tests

The statistical simulation requires the discretization of several continuous variables. We decreased the spacing of these variables systematically until convergence of the simulation results was observed. Here we discuss in depth the relevant details of this process for all of the relevant variables:

- Electron fraction spacing (Δx): The minimum electron fraction x_0 was chosen to be 0.001. At this electron fraction, $\Delta\tau_e$ between $Q = 1$ and $Q = 0$ integrated across the redshift range of our simulation is negligible, so that smaller electron fractions are irrelevant. The maximum electron fraction x_{m-1} was chosen to be 0.9. Our analytic model for the profile produces $r(x)$, where $r \to 0$ as $x \to 0$. The radius at which $x = 0.9$ in the analytic model is a good representation of the area of total ionization within the simulation. The number of x_i points was increased until the results were found to converge beyond $m = 20$.

- Halo addition time step (ΔT): Halos are added to the simulation at constant time intervals. It was found that the results converged beyond 50 intervals, but computational complexity does not increase significantly with the number of time intervals, so a value of 72 was settled on.

- Output time step (Δt): The time interval between output reports does not have any effect on the results of the simulation, so convergence is not relevant.

- Comoving volume (V): Computational complexity scales linearly with the volume of the box. We found that a volume of $V = 10^6$ Mpc^3 produced sufficiently large halos to ensure complete reionization by a redshift of $z \sim 5.8$. Larger simulations are possible, but require

significant time resources and will be performed moving forward.

Bibliography

Abel T., Norman M. L., Madau P., 1999, Astrophysical Journal , 523, 66

Abel T., Wandelt B. D., 2002, MNRAS , 330, L53

Ahn K., Iliev I. T., Shapiro P. R., Mellema G., Koda J., Mao Y., 2012, Astrophysical Journal Letters, 756, L16

Altay G., Croft R. A. C., Pelupessy I., 2008, MNRAS , 386, 1931

Alvarez M. A., Abel T., 2007, MNRAS , 380, L30

Alvarez M. A., Bromm V., Shapiro P. R., 2006, Astrophysical Journal , 639, 621

Anninos P., Zhang Y., Abel T., Norman M. L., 1997, New Astronomy, 2, 209

Aubert D., Teyssier R., 2008, MNRAS , 387, 295

Aubert D., Teyssier R., 2010, Astrophysical Journal , 724, 244

Auer L. H., Mihalas D., 1970, MNRAS , 149, 65

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, Monthly Notices of the Royal Astronomical Society, 447, 3402

Becker G. D., Rauch M., Sargent W. L. W., 2007, Astrophysical Journal , 662, 72

Bolton J. S., Haehnelt M. G., 2007, MNRAS , 382, 325

Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS , 416, L70

Boutsia K. et al., 2011, Astrophysical Journal , 736, 41

Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, Astrophysical Journal , 811, 140

Bouwens R. J. et al., 2013, Astrophysical Journal Letters, 765, L16

Bouwens R. J. et al., 2022, Astrophysical Journal , 931, 160

Boylan-Kolchin M., Weisz D. R., Johnson B. D., Bullock J. S., Conroy C., Fitts A., 2015, MNRAS , 453, 1503

Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, MNRAS , 443, 2831

Choudhury T. R., Ferrara A., 2005, MNRAS , 361, 577

Ciardi B., Ferrara A., Marri S., Raimondo G., 2001, MNRAS , 324, 381

Ciardi B., Ferrara A., White S. D. M., 2003, MNRAS , 344, L7

Croft R. A. C., Altay G., 2008, MNRAS , 388, 1501

Davies F. B. et al., 2018, Astrophysical Journal , 864, 142

Deparis N., Aubert D., Ocvirk P., Chardin J., Lewis J., 2018, ArXiv e-prints

Donahue M., Shull J. M., 1987, Astrophysical Journal Letters, 323, L13

Donnan C. T. et al., 2022, Monthly Notices of the Royal Astronomical Society, 518, 6011

Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H.-W., 2002, Astronomical Journal , 123, 1247

Fan X. et al., 2006, Astronomical Journal , 132, 117

Faucher-Giguère C.-A., Lidz A., Hernquist L., Zaldarriaga M., 2008, Astrophysical Journal , 688, 85

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, Astrophysical Journal , 703, 1416

Fernández-Soto A., Lanzetta K. M., Chen H.-W., 2003, MNRAS , 342, 1215

Fè vre O. L. et al., 2020, Astronomy & Astrophysics, 643, A1

Finlator K., Özel F., Davé R., 2009, MNRAS , 393, 1090

Furlanetto S. R., 2016, in Understanding the Epoch of Cosmic Reionization, Springer International Publishing, pp. 247–280

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, Astrophysical Journal , 613, 1

Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS , 351, 399

Gnedin N. Y., 2000a, Astrophysical Journal , 535, 530

Gnedin N. Y., 2000b, Astrophysical Journal , 542, 535

Gnedin N. Y., 2008, Astrophysical Journal Letters, 673, L1

Gnedin N. Y., 2014, Astrophysical Journal , 793, 29

Gnedin N. Y., Abel T., 2001, New Astronomy, 6, 437

Gnedin N. Y., Kaurov A. A., 2014, Astrophysical Journal , 793, 30

Gnedin N. Y., Kravtsov A. V., Chen H.-W., 2008, Astrophysical Journal , 672, 765

González M., Audit E., Huynh P., 2007, Astronomy & Astrophysics, 464, 429

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, Astrophysical Journal , 622, 759

Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS , 404, 1111

Hansen S. H., Haiman Z., 2004, Astrophysical Journal , 600, 26

Hartley B., Ricotti M., 2016, MNRAS , 462, 1164

Hasegawa K., Umemura M., Susa H., 2009, MNRAS , 395, 1280

Hsiao T. Y.-Y. et al., 2023, Jwst nirspec spectroscopy of the triply-lensed $z = 10.17$ galaxy macs0647−jd

Hui L., Gnedin N. Y., 1997, MNRAS , 292, 27

Hurwitz M., Jelinsky P., Dixon W. V. D., 1997, Astrophysical Journal Letters, 481, L31

Iliev I. T. et al., 2006a, MNRAS , 371, 1057

Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS , 439, 725

Iliev I. T., Mellema G., Pen U.-L., Merz H., Shapiro P. R., Alvarez M. A., 2006b, MNRAS , 369, 1625

Inoue A. K., Iwata I., Deharveng J.-M., 2006, MNRAS , 371, L1

Iwata I. et al., 2009, Astrophysical Journal , 692, 1287

Johnson J. L., Greif T. H., Bromm V., 2007, Astrophysical Journal , 665, 85

Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2021, Monthly Notices of the Royal Astronomical Society, 511, 4005

Kannan R., Vogelsberger M., Marinacci F., McKinnon R., Pakmor R., Springel V., 2019, Monthly Notices of the Royal Astronomical Society, 485, 117

Kashikawa N. et al., 2011, Astrophysical Journal , 734, 119

Katz H., Ricotti M., 2013, MNRAS , 432, 3250

Katz H., Ricotti M., 2014, MNRAS , 444, 2377

Kimm T., Cen R., 2014, Astrophysical Journal , 788, 121

Knollmann S. R., Knebe A., 2009, ApJS , 182, 608

Krumholz M. R., Stone J. M., Gardiner T. A., 2007, Astrophysical Journal , 671, 518

Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS , 423, 862

Leitherer C. et al., 1999, ApJS , 123, 3

Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, MNRAS , 453, 960

Madau P., Haardt F., Rees M. J., 1999, Astrophysical Journal , 514, 648

Madau P., Rees M. J., Volonteri M., Haardt F., Oh S. P., 2004, Astrophysical Journal , 604, 484

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, Astrophysical Journal , 856, 2

McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS , 447, 499

McLure R. J., Dunlop J. S., Cirasuolo M., Koekemoer A. M., Sabbi E., Stark D. P., Targett T. A., Ellis R. S., 2010, MNRAS , 403, 960

McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS , 377, 1043

Mellema G., Iliev I. T., Alvarez M. A., Shapiro P. R., 2006, New Astronomy, 11, 374

Message Passing Interface Forum, 2021, MPI: A Message-Passing Interface Standard Version 4.0

Mihalas D., Mihalas B. W., 1984, Foundations of radiation hydrodynamics

Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, Astrophysical Journal , 530, 1

Mitra S., Choudhury T. R., Ferrara A., 2015, MNRAS , 454, L76

Mortlock D. J. et al., 2011, Nature, 474, 616

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, Astrophysical Journal , 710, 903

Nestor D. B., Shapley A. E., Kornei K. A., Steidel C. C., Siana B., 2013, Astrophysical Journal , 765, 47

Norman M. L., Paschos P., Abel T., 1998, MmSAI , 69, 455

Norman M. L., Reynolds D. R., So G. C., Harkness R. P., Wise J. H., 2015, ApJS , 216, 16

NVIDIA, Vingelmann P., Fitzek F. H., 2020, Cuda, release: 10.2.89

Ocvirk P. et al., 2020, Monthly Notices of the Royal Astronomical Society, 496, 4087

Oesch P. A. et al., 2014, Astrophysical Journal , 786, 108

O'Shea B. W., Wise J. H., Xu H., Norman M. L., 2015, Astrophysical Journal Letters, 807, L12

Ota K., 2008, Reionization and Galaxy Evolution Probed by z=7 Lyalpha Emitters in SXDS. Subaru Proposal

Ouchi M. et al., 2010, Astrophysical Journal , 723, 869

Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2013, MNRAS , 429, L94

Paardekooper J.-P., Kruip C. J. H., Icke V., 2010, Astronomy & Astrophysics, 515, A79

Pawlik A. H., Schaye J., 2008, MNRAS , 389, 651

Pawlik A. H., Schaye J., 2011, MNRAS , 412, 1943

Pawlik A. H., Schaye J., Dalla Vecchia C., 2015, MNRAS , 451, 1586

Pentericci L. et al., 2011, Astrophysical Journal , 743, 132

Petkova M., Springel V., 2009, MNRAS , 396, 1383

Petkova M., Springel V., 2011, MNRAS , 415, 3731

Planck Collaboration et al., 2015, ArXiv e-prints

Planck Collaboration et al., 2020, Astronomy & Astrophysics, 641, A6

Press W. H., Schechter P., 1974, Astrophysical Journal , 187, 425

Prochaska J. X., Worseck G., O'Meara J. M., 2009, Astrophysical Journal Letters, 705, L113

Razoumov A. O., Scott D., 1999, MNRAS , 309, 287

Ricotti M., 2002a, MNRAS , 336, L33

Ricotti M., 2002b, MNRAS , 336, L33

Ricotti M., 2016, MNRAS , 462, 601

Ricotti M., Gnedin N. Y., Shull J. M., 2001, Astrophysical Journal , 560, 580

Ricotti M., Gnedin N. Y., Shull J. M., 2002a, Astrophysical Journal , 575, 33

Ricotti M., Gnedin N. Y., Shull J. M., 2002b, Astrophysical Journal , 575, 33

Ricotti M., Gnedin N. Y., Shull J. M., 2002c, Astrophysical Journal , 575, 49

Ricotti M., Gnedin N. Y., Shull J. M., 2002d, Astrophysical Journal , 575, 49

Ricotti M., Ostriker J. P., 2004a, MNRAS , 350, 539

Ricotti M., Ostriker J. P., 2004b, MNRAS , 352, 547

Ricotti M., Ostriker J. P., Gnedin N. Y., 2005, MNRAS , 357, 207

Ricotti M., Ostriker J. P., Mack K. J., 2008, Astrophysical Journal , 680, 829

Ricotti M., Parry O. H., Gnedin N. Y., 2016, Astrophysical Journal , 831, 204

Ricotti M., Shull J. M., 2000, Astrophysical Journal , 542, 548

Rijkhorst E.-J., Plewa T., Dubey A., Mellema G., 2006, Astronomy & Astrophysics, 452, 907

Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, Monthly Notices of the Royal Astronomical Society, 436, 2188

Schaerer D., 2003, Astronomy & Astrophysics, 397, 527

Shapiro P. R., Giroux M. L., 1987, Astrophysical Journal Letters, 321, L107

Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., Erb D. K., 2006, Astrophysical Journal , 651, 688

Sheth R. K., Tormen G., 2002, MNRAS , 329, 61

Siana B. et al., 2007, Astrophysical Journal , 668, 62

So G. C., Norman M. L., Reynolds D. R., Wise J. H., 2014, Astrophysical Journal , 789, 149

Sokasian A., Abel T., Hernquist L. E., 2001, New Astronomy, 6, 359

Sokasian A., Yoshida N., Abel T., Hernquist L., Springel V., 2004, MNRAS , 350, 47

Songaila A., Cowie L. L., 2010, Astrophysical Journal , 721, 1448

Springel V., 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105

Steidel C. C., Pettini M., Adelberger K. L., 2001, Astrophysical Journal , 546, 665

Stone J. M., Mihalas D., Norman M. L., 1992, ApJS , 80, 819

Susa H., 2006, PASJ , 58, 445

Tilvi V. et al., 2014, Astrophysical Journal , 794, 5

Trac H., Cen R., 2007, Astrophysical Journal , 671, 1

Trac H., Cen R., Loeb A., 2008, Astrophysical Journal Letters, 689, L81

Trebitsch M. et al., 2021, Astronomy & Astrophysics, 653, A154

Trenti M., Stiavelli M., Bouwens R. J., Oesch P., Shull J. M., Illingworth G. D., Bradley L. D., Carollo C. M., 2010, Astrophysical Journal Letters, 714, L202

Vale A., Ostriker J. P., 2004, MNRAS , 353, 189

Vale A., Ostriker J. P., 2006, MNRAS , 371, 1173

Vanzella E. et al., 2010, Astrophysical Journal , 725, 1011

Venkatesan A., Giroux M. L., Shull J. M., 2001, Astrophysical Journal , 563, 1

Volonteri M., Gnedin N. Y., 2009, Astrophysical Journal , 703, 2113

Whalen D., Norman M. L., 2006, ApJS , 162, 281

Wise J. H., Abel T., 2011, MNRAS , 414, 3458

Wise J. H., Cen R., 2009, Astrophysical Journal , 693, 984

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS , 442, 2560

Wyithe J. S. B., Loeb A., 2003, Astrophysical Journal , 586, 693

Yajima H., Choi J.-H., Nagamine K., 2011, MNRAS , 412, 411

Yajima H., Li Y., Zhu Q., Abel T., Gronwall C., Ciardullo R., 2014, MNRAS , 440, 776

Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS , 414, 727