# Recovering the Physics of the Epoch of Reionization using Semi-Analytic 21CMMC Code

Michelle Miller

May 9, 2018

**Abstract**

After the Big Bang the universe was filled with neutral hydrogen that began to cool and collapse into the first structures. These first stars and galaxies began to emit radiation that eventually ionized all of the neutral hydrogen in the universe. This is known as the Epoch of Reionization (EoR). 21cmFAST is a simulation that recreates the EoR by modeling the dark matter and baryonic distributions of the universe and then incorporating the ionization fields. 21CMMC is a semi-numerical code that takes mock measurements from the resulting boxes of 21cmFAST or comparable simulations. Those measurements are given to 21CMMC and help us determine key properties of this simulated universe: the mean free path of photons through the intergalactic medium (IGM), the efficiency of photons leaving each dark matter halo, and the minimum virial temperature of those dark matter halos. My project tests the robustness of 21CMMC on universe simulations other than 21cmFAST to see whether 21CMMC can properly reconstruct early universe parameters given a mock "measurement".

# Contents

# Acknowledgement

I would like to thank my advisor, Jonathan Pober, who has patiently responded to my plethora of late night emails over the course of the past couple of years. He is consistently encouraging and reassuring that research is not always a smooth process and can often be extremely frustrating. I'm grateful for all of his guidance.

I would also like to thank the creators of 21cmFAST and 21CMMC Andrei Mesinger and Brad Greig and their collaborations. I am grateful to Adam Lidz who provided the power spectra from the radiative transfer simulations by the creators of the Radiative Transfer simulations by Matthew McQuinn, Larz Hernquist, Matais Zalderriaga, and Suvendra Dutta.

Thank you to my family for getting me to Brown and standing behind me even when there were days we did not know if I would continue my time at Brown. I would not be here with them and their continued love and encouragement is a solid pillar in my life that I have come to trust and rely on more and more.

Lastly, thank you to all of my amazing friends who have supported me through the late nights and low days of struggle, while also providing emotional support and inspiration. They have brought light to my life in a way that I did not know existed before coming to Brown.

# 1 Introduction

After the Big Bang, most baryonic matter formed and recombined in the form of neutral hydrogen in a period called the "Dark Ages". As the universe continued to expand, it was mostly filled with hydrogen and some helium in a sea of dark matter. Because of this, gravitational collapse of matter came to dominate the outward pressure of radiation. This allowed the first structures such as population III (metal free stars made of hydrogen) stars and galaxies to form via gravitational collapse. They became the first sources of ionizing radiation, which began to ionize the hydrogen in the intergalactic medium (IGM). “Bubbles” of ionized hydrogen grew around the luminous sources and eventually overlapped. At around redshift, z=6, these bubbles expanded to permeate and fully ionize the universe [6]. This is known as the Epoch of Reionization (EoR) or also the “Cosmic Dawn”.

The EoR is a significant point of transition for the evolution of the early universe because it marks the formation of these first structures. In the early 20th century, physicists thought the galaxy itself was our universe. Now in a Big Bang model universe we know that there has been a much more dynamic evolution of structures over time. Thus, the questions arose:

- What were the first structures in the universe? When did the first generation of galaxies form?
- What were the properties of these first structures?
- How did these first structures evolve over time and subsequently change the surrounding intergalactic medium?
- What is the thermal and ionization history of baryons in the universe? [25]

Today’s instruments are not sensitive enough to directly detect the extremely faint first objects, so indirect observation is necessary. As a result of this, there is a two-fold problem in studying the EoR. First, the scale range being probed over Reionization is extremely wide from the size of single stars to spanning gigaparsecs. It is also inhomogeneous on these scales. Second, because very little is known about the first structures, there is an extremely wide astrophysical parameter space. A large goal of studying Reionization is to better constrain the properties of these objects like their masses, photon efficiencies, and distributions within the universe itself [25].

Simulations give physicists agency in their ability to tweak the properties of the early universe and then to "observe" the evolution of a fake universe. These fake universes simulate structures like galaxies and stars that emit ionizing radiation into the surrounding neutral IGM, and they track when the hydrogen is ionized. This is significant for tracking the 21cm spin-flip neutral Hydrogen signal, which is a signal emitted due to the change in spin alignment for the electron in neutral hydrogen. Though it is a rare occurrence, neutral hydrogen was ubiquitous enough that the universe is expected to be “bright” with this signal. It is an important probe because it can be measured (in theory) by telescopes and in simulations, but it is plagued by foreground and noise.

To combat this, N body or hydrodynamic simulations and their corresponding ionization fields can be extremely powerful in probing the universe. The initial models were based on tracking the expansion of ionized hydrogen around individual galaxies [3]. Such simulations can be smaller volume

but higher resolution boxes that can resolve objects like dwarf galaxies though not individual stars. The largest volume simulations have lower resolution [40]; though resolution and dimensions of the boxes are subject to vary from simulation to simulation. The goal of these simulations is to investigate the growth of the first structures "directly" since we are unable to take measurements.

Such simulations often take the form of N body codes, which can be extremely computationally expensive. Fully numerical hydrodynamic N body codes usually aim to solve the gravitational interactions of particles, while also tracking the fluid behavior baryon gas physics within a box that simulates sizes $> (100Mpc)^3$. As structures form and release photons, a radiative transfer simulation would track the photon propagation and ionization. Such simulations track the path of each individual particle through absorption and emission throughout the entire simulation. It is extremely computationally expensive to do this.

Additionally, computational cost is also high given the wide dynamic range in scale needed to simulate an entire volume of the universe. For example, star formation occurs on the scale of about $R_\odot \approx 2.3 * 10^{-14} Mpc$, while (as mentioned above) some of the boxes we are trying to simulate have dimensions of Mpc. This means there is a scale range of at least 18 orders of magnitude to probe. if we want to resolve star formation when modeling the universe itself [19]. Not even the largest super computers are capable of such computation. For example, it is only in the past couple of years that the first coupled radiation-hydrodynamic simulation, called CoDa, was developed. This ran on over 8,192 cores and 8,192 GPUs over 10 days, while using about 60 million core hours [32].

Despite this, there are avenues we can take to approximate and thus shorten computation time, which we see in the semi-analytic N body code, 21cmFAST. It is a modeling tool that creates 3D realizations of the evolved density and ionization fields of the simulated boxes. It uses approximations to evolve the density fields and to ultimately determine the collapsed dark matter halos and the resulting ionization bubbles. Taking these algorithmic short cuts allows for 21cmFAST to create realizations in a matter of minutes or hours as opposed to weeks or months as seen in traditional radiative transfer fully numeric simulations, while still producing the same results within a factor of 10% [24].

21cmFAST has been extremely useful to collaborations like the Murchinson Widefield Array (MWA) and Lunar University Network for Astrophysics Research (LUNAR), which have begun to use DexM and 21cmFAST to model the large-scale cosmological 21cm signal. These groups have fields of view on the order of several gigaparsecs, which is an unprecedented dynamic scale range, and are unattainable scales to investigate with fully numerical simulations at least at present.

Despite the benefit of having an efficient Reionization simulation, when making such simulations, there is still a need to deduce the original physics of the universe (i.e. in practice, particular parameter values) when the first real 21cm signal of the EoR is detected. 21CMMC is a semi-numerical Monte Carlo Markov Chain code that takes data or simulated data boxes of this ionized universe from 21cmFAST. These data or mock measurements are thrown into 21CMMC which will step through parameter space, making boxes using 21cmFAST, until it creates a box that can recreate the initial measurement. This final box that matches the initial measurement has values of fundamental parameters (in this case they correspond to galaxy size, photon efficiency, and photon mean free path) that influence the ionization

of the IGM (Greig and Mesinger, 2015). The key is that in determining the final parameters (and thus recreating the initial measurement), 21CMMC also produces full likelihoods around this best fit. This means we have a set of statistics to rely on that tell us how good our results are. Thus, the two simulations, 21cmFAST and 21CMMC, provide the potential to reduce the wide computational dynamic range (via 21cmFAST's approximations) while also tackling the large parameter space (via 21CMMC) of the early universe.

The purpose of this thesis is to verify how robust 21CMMC is in recreating the parameter values of the early universe of mock measurements from 21cmFAST and from a full radiative transfer simulation as seen in McQuinn et al. (2007).

# 2 Analyzing Reionization

## 2.1 The Struggle

Study of the EoR has evolved in time with growing technologies. Often, large radio interferometers are used to probe the 21cm signal of the EoR (which will be further explained below) as it is expected to be red shifted into the radio frequency. There are large collaborations building and refining radio interferometers globally such as the Murchison Wide Field (MWA) Array in Australia [20], the Hydrogen Epoch of Reionization Array (HERA) in South Africa [9], and the Square Kilometer Array (SKA) in Australia and South Africa [8]. The biggest struggle of such collaborations is that the 21cm signal from Reionization is too faint for detection due to loud foregrounds and small signal-to-noise ratios.

There are two major pieces of information needed in order to have an accurate understanding of the EoR: a reliable model of galaxy formation and an accurate treatment of the radiative transfer of ionizing photons. The problems in these constraints result from uncertainties in the modeling of several physical processes: properties of first stars and quasars, ionizing photon production and radiative transfer. Uncertainties in the photon escape fraction, population numbers of each possible source, and properties of the IGM itself make it very difficult to properly understand the evolution of the EoR. While the evolution of dark matter is understood, we do not understand the governing processes of luminous object formation enough to be able to have exact constraints on the Reionization parameters.

## 2.2 21cm Signal

How we probe the EoR is highly dependent on studying the 21cm transition of the hyperfine spin-flip transition of neutral hydrogen. During this transition, the hydrogen emits photons that correspond to a wavelength of 21cm. Even in a presence of a small amount of neutral hydrogen, the 21cm signal is not a common phenomenon for any given hydrogen molecule. However, there is a ton of neutral hydrogen in the universe, so it is likely that given such a large amount of neutral hydrogen then we will see the 21cm signal. Throughout the duration of Reionization, the intergalactic medium becomes increasingly ionized as the first light sources are emitting more and more photons. When the signal dies off, this means that the region of the universe we are observing is extremely ionized.

This is because ionized hydrogen will appear as "holes" in the temperature of the 21cm line, thus providing physicists with a probe to track when there is neutral hydrogen vs. ionized hydrogen [12]. Not only can the 21cm line provide a map of Reionization but also the Dark Ages in which the universe was filled with neutral hydrogen. As Hubble expansion occurs, gravitational instabilities expand resulting in more ionized or heated regions, which add to fluctuations in the 21cm line signals. As galaxies further reionize the IGM, these fluctuations fade, and then increase as bubbles of ionized gas around galaxies expands. Thus by studying the 21cm line signals, the larger scale relationship between the IGM and galaxies will be better understood (Backer et. Al).

We will quantify the 21cm signal in the form of the 21cm power spectrum. The power spectrum measures the fluctuations from a mean value as a function of scale quantified through Fourier space k-modes. This will be further discussed below.

### 2.2.1 Brightness Temperature

In order to take a power spectrum, we must understand the concept of **brightness temperature**, which is a measure of intensity. We measure the differential brightness temperature, which is the difference in brightness temperature between the 21 cm signal and the cosmic microwave background (CMB). This is equivalent to considering differential brightness temperature of the 21cm line relative to that of the CMB photons is given by:

$$\delta T_b(\nu) = \frac{T_S - T_\gamma}{1+z}(1 - e^{-\tau_{\nu_0}}) \tag{1}$$

$$\approx 27 x_{HI}(1+\delta_{nl}\frac{H}{dv_r/dr + H})(1 - \frac{T_\gamma}{T_S}) \times (\frac{1+z}{10}\frac{0.15}{\Omega_m h^2})^{1/2}(\frac{\Omega_b h^2}{0.023})mK \tag{2}$$

Where $T_S$ is the gas spin temperature, $\tau_{\nu_0}$ is the optical depth of the 21cm frequency, $\nu_0, \delta_{nl}(x,z)$ is the evolved (eularian) over density, H(z) is the Hubble parameter, $dv_r/dr$ is the gradient of the line of sight component of the velocity. A higher spin temperature is indicative of more neutral hydrogen (i.e. more spin flip transitions). The spin temperature determines the intensity of the 21cm signal [43].

It can be used to learn about the density of neutral hydrogen and the gas spin temperature [24]. Usually, the standard assumption made by physicists to simplify the above computation is that we are running Reionization in a post-heating regime ($T_S >> T_\gamma$). Violating the above assumption will lead to fluctuations in $\delta T_b$ that are driven neither by density fluctuations nor by the ionized regions regions.

The 21cm line for neutral hydrogen is seen in absorption when the spin temperature is less than the CMB temperature, while it is in emission when it is greater than the CMB temperature.

## 2.3 The Power Spectrum

In order to understand simulation of the EoR, we must understand how we try to study it. This begins with the growth of structure in itself. The EoR is inherently ingrained in the formation of the first structures, which also mark the non-linear growth of density in the early inhomogeneous universe. Given some energy density that is a function of radius, we can separate it into the mean value and the

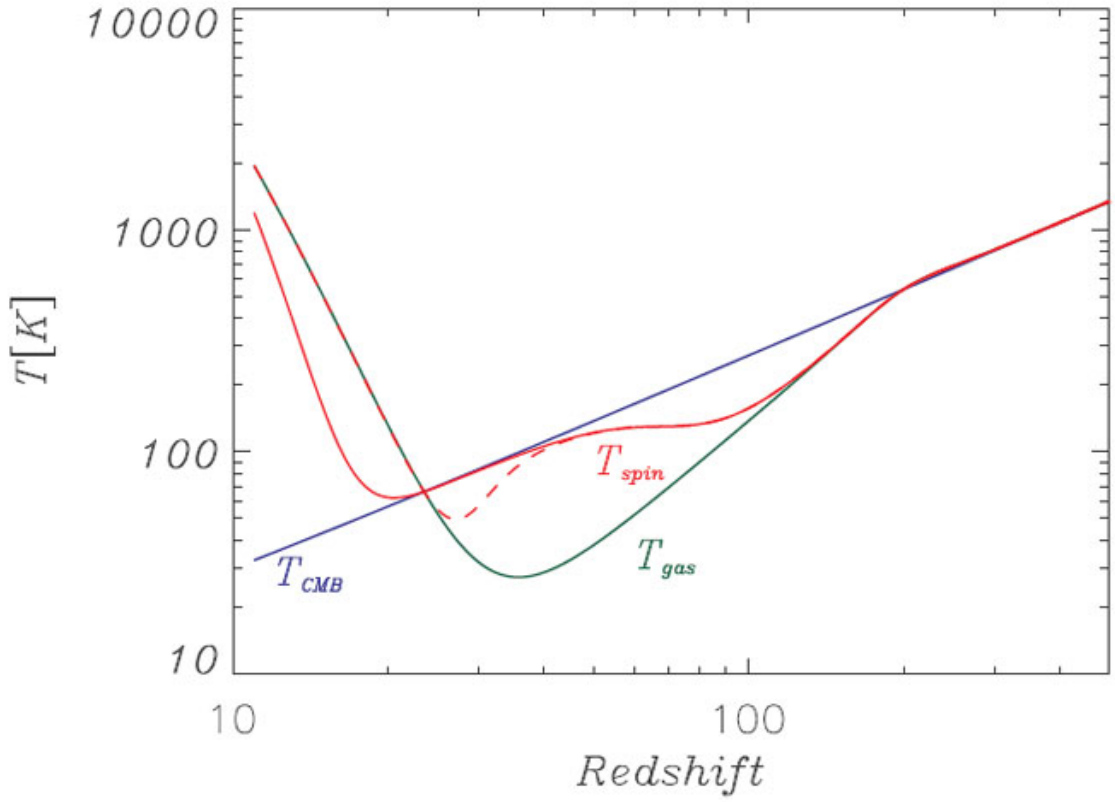



Figure 1: Pictured above are the spin, gas, and CMB temperature as a function of redshift and when they are coupled vs. independent of one another. At early redshifts (before z $\approx$ 200), the spin temperature is coupled to the CMB temperature and to the kinetic temperature of the hydrogen gas in the IGM. This occurs because the gas temperature is coupled to the CMB temperature via Compton scattering of the gas and also coupled to the spin temperature via electron collisions. At a redshift of z $\approx$ 200, the spin and kinetic temperatures briefly decouple from the CMB. Collisions drop off due to Hubble expansion, so the spin temperature begins to decouple from kinetic temperature, and recouple with the CMB temperature between a redshift of 100 and 30. At this time, the 21cm signal will mostly will mostly probe the cosmological density field. However, the first ionizing sources begin to turn on, the spin temperature decouples from the CMB and couples back with the gas temperature. The gas may get heated above the CMB temperature before Reionization, suggesting a separate "Epoch of X-ray Heating" that takes place prior to the EoR [26]. It also grows large relative to the background CMB radiation during this time, which is around when Reionization is beginning (z $\approx$ 20). It is around here when the spin temperature is larger than the CMB temperature that we see the 21cm signal largely in emission. Thus Reionization provides a natural physical reason to say the CMB temperature can be ignored [43]

perturbation due to some fractional over density of matter.

$$\rho(r) = \bar{\rho}(r) + \delta\rho(r) = \bar{\rho}(r)(1+\delta) \tag{3}$$

where $\delta$ is the perturbation to the energy density as defined $\delta = \frac{\rho(r)-\bar{\rho}(r)}{\bar{\rho}(r)}$. Using this, we can multiply together two perturbations from any two points in space and take the ensemble average of them over the volume. This is called the correlation function:

$$\xi(r_1 - r_2) =< \delta(r_1)\delta(r_2) > \tag{4}$$

The correlation function gives information on the excess in clustering of matter over densities for a given separation or scale. If the universe itself were totally random in its matter distribution, then the correlation function wouldn't tell us anything. If we take the Fourier transform of the correlation function, then we will get the **Power Spectrum**, which is a very powerful tool that is used to understand the clumping of matter at different scales:

$$P(k) = \int d^3r < \delta(r_1)\delta(r_2) > e^{-ikr} =< |\delta(k)|^2 >= \int d^3r\xi(r)e^{-ikr} \tag{5}$$

The fact that the Power spectrum is a function of the scalar k and not the vector k tells us that it is direction independent (isotropic). Thus it will reveal information about any given scale $2\pi$/k. Assuming the structure of matter is Gaussian (not always a fair assumption), the two-point statistic Power Spectrum fully describes the universe. When working with simulations of the EoR, we will be taking power spectra of the 21cm signal (via the ionized hydrogen distribution) within the boxes. Additionally, the power spectrum itself can be a powerful metric for any kind of varying density field such as ionization fields of the 21cm signal.

We can further write the power spectrum P(k) as a dimensionless power spectrum by taking out the $k^3$ dependence:

$$\Delta^2(k) = \frac{4\pi k^3 P(k)}{(2\pi)^3} = \frac{k^3 P(k)}{2\pi^2} \tag{6}$$

Up to this point we have discussed the power spectrum as a measure of over densities, which is often in the context of dark matter. However, we obviously want the 21cm power spectrum. We can obtain the 21cm power spectrum brightness fluctuations in Fourier space using the brightness temperature itself:

$$(2\pi)^3\delta^3(\mathbf{k\text{-}k'})P_{21}(\mathbf{k}) \equiv \langle \delta T_b(\mathbf{k}) \times \delta T_b(\mathbf{k'}) \rangle \tag{7}$$

where k is the wave-vector for a particular Fourier mode (inversely related to scale).

The dimensionless power spectrum of the 21cm signal is as follows:

$$\Delta^2_{21}(k,z) \equiv k^3/(2\pi^2 V)\delta\bar{T}_b(z)\langle \mid \delta_{21}(\mathbf{k},z) \mid^2 \rangle_k \tag{8}$$

where $\delta_{21}(\mathbf{x},z) = \delta T_b(\mathbf{x},z)/\delta\bar{T}_b(z) - 1$. This is often the value that is considered because the normal power spectrum P(k) often has a $k^n$ dependence, so the quantity $k^3P(k)$ is a scale invariant quantity in a gravitational potential [19].

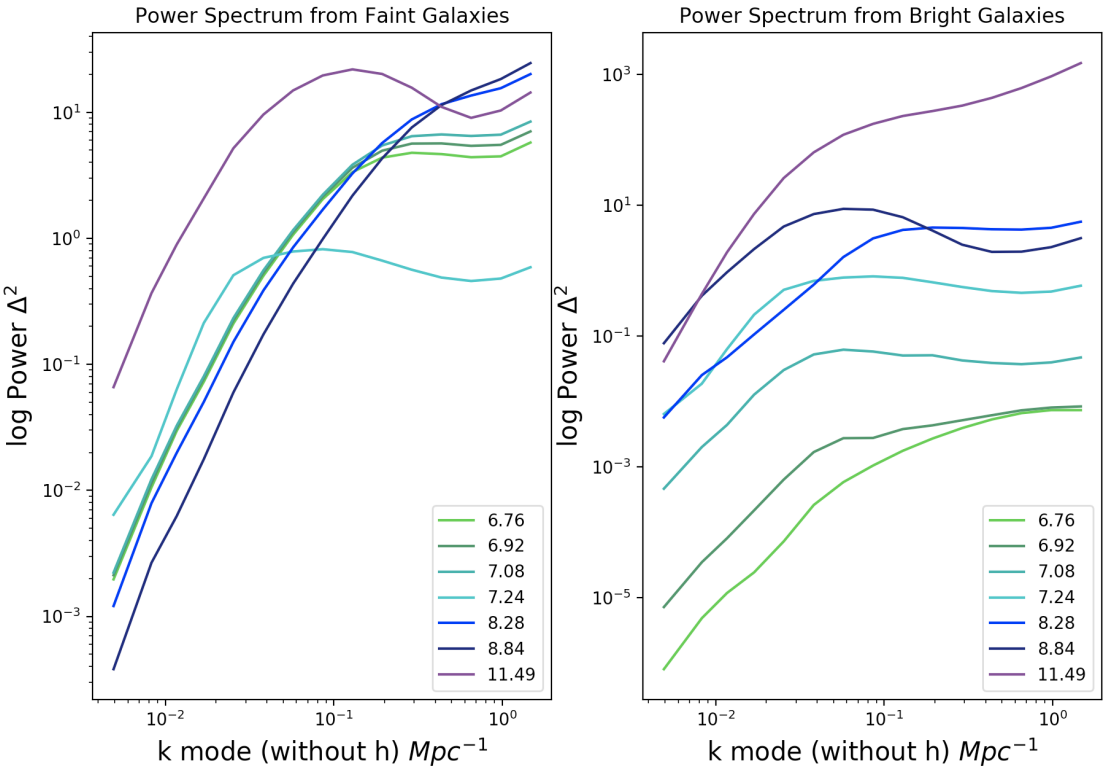



Figure 2: In particular, the way the power spectrum measures fluctuations on different scales can be especially powerful because this means the nature of our sources will be reflected in the power spectrum itself. For example, if very rare bright quasars were the prominant sources of Reionization, then we would expect to see a bump in the smaller k modes of the dark matter/ionization power spectra as rare objects will generally sit in large scale over densities. In contrast, if numerous, faint, small dwarf galaxies or even population III stars powered Reionization, then we would see a flattening at the lower k modes and the bump in the higher k modes. These objects would dominate the smaller scale over densities. We need a mass function so we can properly populate our simulated universe. This is because having the respective mass functions for these objects is important because it would define the scales in which objects are ionizing the surrounding IGM. Thus the power spectrum is a powerful tool in detecting the nature of the sources.

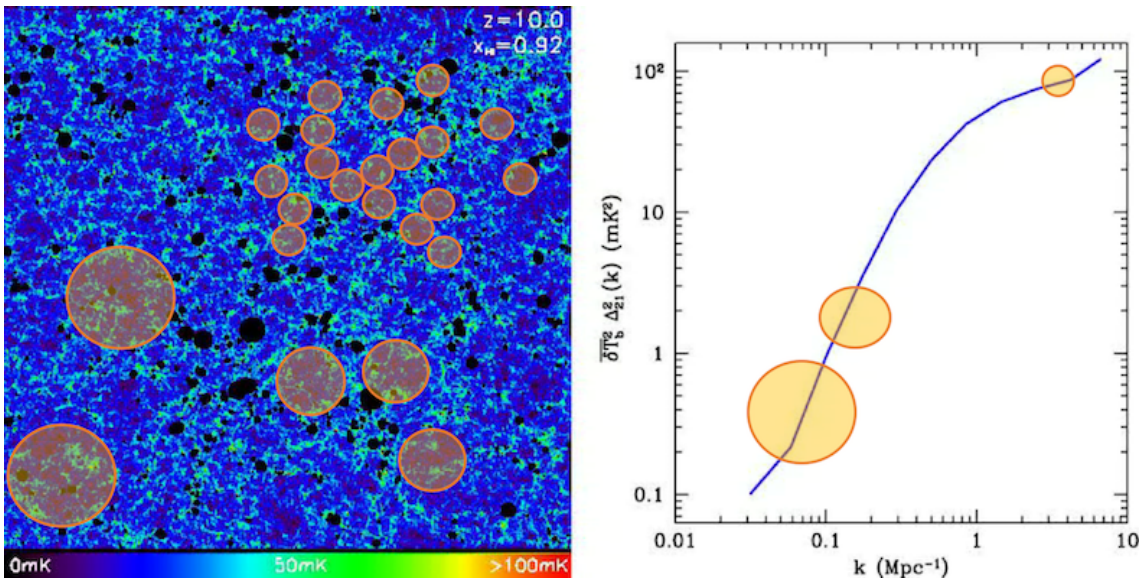



Figure 3: This graph demonstrates what the dimensionless power spectrum is trying to convey. It is a measure of how many over densities of neutral hydrogen exist relative to a mean density for a given scale. The orange circles show examples of the scale sizes being considered. Note that the x axis on the right plot is inverted. Small k values correspond to larger scales $Mpc^{-1}$ and larger k modes correspond to smaller scales. When we have a large power value at large k modes, then we know there are a lot of small scale fluctuations in the neutral hydrogen density. If there is a lot of power on small k modes, then the fluctuations and over densities in neutral hydrogen would be larger and "clumpier". The right plot is a bit misleading because it is the dimensionless power spectrum, which has an extra $k^3$ term in $\Delta^2 = k^3 P(k)$ that makes the power spectrum values for large k even larger.

# 3 Theory of Modeling the Universe

This section will discuss the theory behind the 21cmFAST simulation used in this work. I will derive the power spectrum given a perturbation in the matter distribution. This perturbation will also be reflected in applied Zeldovich approximation to build the dark matter halos themselves. I will then explain the dark matter halo collapse model, which will be discussed in the context of the spherical Top-Hat model and Press Schecter equation. The Excursion Set formalism will be shown to identify dark matter halos and the ionized hydrogen bubbles during Reionization itself.

## 3.1 Simulations

### 3.1.1 History of Simulation

The first structures formed from the growth of linear density fluctuations in the matter density field. Given that this means they have a well-defined set of cosmological initial conditions, simulations can be used to track this formation. It was in the 1960's that physicists began to realize the importance of molecular hydrogen cooling in the formation of proto-galactic objects [35] [22] [33]. Some of the first hydrodynamic simulations in the late 90's began to connect the simple initial conditions of the universe to the complex structure today using baryonic physics. These simulations began to incorporate baryonic physics in molecular cloud collapse (for population III stars) [1] [30] and the evolution of gas in collapsing dark matter mini-halos [7]. These simulations were also the first to incorporate multi-dimensional simulation [1]. More information is provided in this review [6] by Barkana and Loeb in 2001.

It was during this period that a diversification of cosmological simulation tools came about that vary their treatments of (i) source/sink fields and (ii) radiative transfer/ionization fields [25]. Coupled N body and hydrodynamic codes answer the former the best, while ray-tracing simulations handle the latter the best. Simulations like the *Renaissance simulations* by Xu et al. [42] are extremely powerful examples of this in which they probe the stellar and gaseous properties of the first generation galaxies ($10^{6\smile}10^{8} M_{\odot}$) as well as radiative feedback, metal pollution, stochastic star formation and recombination clumps in the IGM [25]. However, if these simulations wish to determine these early galaxy populations, then they are limited to scales of $\approx$ 1-10Mpc. Because of this, such cosmological simulations cannot resolve sub-structure of stellar environments.

Some later radiative tracing simulations came to incorporate approximate treatments of the EoR by either not resolving the interstellar medium (ISM) of galaxies and/or removing the hydrodynamics. This is allowed because large scale gravity dynamics generally dictate baryonic physics. They still use halo-finding algorithms, and from their use other statistical models like the Navarro-Frenk density profile [31] to populate the star-forming (and thus photon emitting) regions [40] [2] [18] [4]. By removing the baryonic physics, galaxy properties are dependent on the dark matter distributions. Despite limiting these simulations in their ability to make predictions, such approximations allow an extension of scale to $\approx$ 100Mpc [25]. Such scales allow for statistical sampling of the ionized boxes being produced. This approach was used in the radiative transfer simulations by Trac and Cen [40], whose power spectra we

tested 21CMMC on.

Other methods such as Monte Carlo methods, which sample probability density functions of source spectra and angular distributions to determine photon propagation, further quickened radiative transfer codes and maintained medium hydrodynamic resolution [41] [21] [4]. There are also moment-based ray tracing methods that have produced marginal improvement [25]. In the past decade, the more developed "semi-numerical" simulations have arisen, which apply an analog excursion set formalism by creating the ionization fields directly from the density fields [11]. Furlanetto, Hernquist, and Zaldarriaga develop an ionization model (from here on referred to as the FZH model) of Reionization that is dependent on key parameters. 21cmFAST is an implementation of this formalism [28] [36]. Now they no longer use ray-tracing algorithms, which creates a nearly unlimited dynamic range in scale and allows the simulations to probe similar fields of view to those of upcoming interferometers [25].

Further simulations that try to model Reionization in particular, generally track ionized hydrogen bubbles around individual galaxies [3]. Such models could use semi-numerical techniques to track the ionization fraction itself, but were unable to describe the morphology of Reionization [6]. There are many factors that influence the morphology of Reionization including but not limited to: locations of individual sources, the clumpiness of the IGM, the underlying density distribution, recombinations, and radiative transfer effects. Some popular simulations that arose to address this through semi-analytic methods included those which were dominated by recombinations (thus lower density regions ionized first). However, as simulation improved to include radiative transfer physics, it became apparent that Reionization seems to occur from high to lower density regions. This implies that recombinations play a secondary role to Reionization, and it is more likely that there are less numerous bright sources driving Reionization rather than numerous small sources.

### 3.1.2 General Setup

Generally, a piece of code will initialize a set of density and velocity "fields" that are within massive boxes that are meant to simulate some size of the universe (ranging from 100Mpc-2Gpc). The particles within these density fields have some initial velocities and positions that can be treated as Gaussian distributions. Particles will be allowed to evolve forward in time according to a gravitational collapse via a first order perturbation evolution. However, typically if we want to model baryons beyond dark matter, then we need to consider hydrodynamics. Hydrodynamic code will treat baryons as a fluid that can converge and thus can form sharp discontinuities, which the code maintains by using the continuity equations within the box. In particular, this includes the conservation of mass, momentum, and energy. Much of this is often according to the fiducial model of Lambda/CDM.

If we have a density field, simulation algorithms can determine what masses of particles may become dark matter "halos" where possible galaxies may live. "Halo finding" algorithms will often determine that some region that is dense enough will have some "radius" that we can say is now a halo that has collapsed. There are many different algorithms out there that help us in finding halos. This opens up the possibility for potential parameters that can be varied like the size of a halo that may correspond to

a quasar or to a dwarf galaxy (i.e. the virial mass or temperature). If we can recreate the ionization field we need, then we can further constrain the mass populations or luminosity populations of these objects.

Once we have the halo fields determined, the halos need to be populated with ionizing sources themselves. This will allow for the creation of the ionization field, which reveals information on where the neutral and ionized hydrogen is. From there, the resulting ionized/neutral boxes for a given redshift can be analyzed to "measure" a power spectrum.

## 3.2 Semi-Analytic Formalism

21cmFAST and 21CMMC are not full hydro-dynamical codes nor radiative transfer, but rather semi-numerical simulations. Semi-numerical simulations formally use approximate physics to still model the 3-dimensional universe, while providing faster computation[14].

Such approximations include using perturbation theory to step forward in coordinate space and the Excursion Set formalism, which allows the code to bypass halo-finding algorithms. The key here is that the code is acting directly on the dark matter distribution rather than evolving baryons hydrodynamically in the boxes themselves. This quickens computation time, while simultaneously allowing for the dark matter haloes to then be populated with galaxies for the production of the ionization field.

Additionally, nearly all cosmological simulations apply what is called *subgrid modeling* in which approximations are applied on scales smaller than the resolved voxels in a simulation. Other codes that simulate these smaller scales can produce statistical models of particular processes. This may include a statistical volume average of stellar, galactic, or IGM properties to be included per voxel. There are a plethora of subgrid models ranging from star formation to black hole growth and supernovae feedback [19]. In this case, 21cmFAST's algorithm determines the partial ionized fraction of a cell by assuming sub-grid sources are slowly ionizing the host cell [28].

## 3.3 21cmFAST Code

21cmFAST is a semi-numerical code that is designed to efficiently simulate the 21cm signal during Reionization. The use of these semi-numerical approximations allow 21cmFAST to run a single realization on one core in about 10 minutes as opposed to thousands of cores over many weeks or months as seen in fully numerical simulations. 21cmFAST in particular uses only N body evolution, while relying on approximate physics to implement semi-numerical code. It does this by generating 3 dimensional realizations of evolved density, ionization, peculiar velocity, and spin temperature fields. 21cmFAST differs slightly from typical N body simulations in that it does not treat baryons and dark matter separately. In other words, it acts on a dark matter field and does not capture baryonic physics, but uses the dark matter field to create the ionization field which is usually made from the baryonic distribution [28]. It uses the Zeldovich approximation, which is a order perturbation approximation that tracks the growth of large scale structure in many semi-analytic N body codes[19], and then it uses that dark matter density field to build an ionization field [28]. It bypasses typical dark matter halo finding algorithms and directly

acts on the dark matter field to produce this ionization field. It does this by finding regions that cross a certain ionization density to determine if it is ionized. Once it creates this ionization field, it calculates the brightness temperature to get the 21cm power spectrum.

21cmFAST was produced out of the development of its predecessor, DexM, which is also a semi-numerical simulation of the cosmological 21cm signal [27]. DexM differs from 21cmFAST in that it does not bypass the halo finding algorithm, which entails larger memory requirements. 21cmFAST also has some additional calculations incorporated with it including spin temperature calculations. It uses approximations (see Appendix for more information) in order to shorten the computation time of the code. Similar simulations to 21cmFAST that are fully numerical can often take many weeks to months to run. Thanks to the approximations and computational short cuts taken in the code, a single realization of 21cmFAST can be done in a space of 10 minutes on a single core processor. It does this while remaining within 10% accuracy of recreating the fully numerical simulation results [28]. A key to the model of 21cmFAST is its use of the FZH formalism for the ionization field. It applies an analogous form of what's known as the Press Schecter model, which is traditionally used to determine when dark matter halos have formed. In the FZH formalism, this is applied to the ionization field to determine when there are fully ionized hydrogen bubbles. There are the main parameters that will subsequently affect the geometry of the ionization field: the mean free path of photons ($R_{mfp}$), the ionization efficiency ($\zeta$), and the minimum virial temperature ($T_{virmin}$) of the halos. These parameters will be further explained below.

Each of these parameters will influence how the box is populated by galaxies and how far the corresponding photons will travel into the IGM. This will influence how much neutral hydrogen is left, which we can quantify with the final power spectrum.

### 3.4 Parameters of the 21cmFAST Ionization Field

Up to this point we have now described how the code builds density fields that then collapse into dark matter halos, but this hasn't described how these presumably galaxy containing halos ionize the nearby intergalactic medium. In order to do this we need to "populate" the Halo field with galaxies for the halos identified with Press Schecter [44]. For example, Potential Quasar locations could be found from analyzing dense clumps of gas within these boxes and implementing a possible luminosity function that would populate these regions with quasars. Given the placement of sources, we need to track how they come to ionize the surrounding IGM. As mentioned, to determine whether a particular voxel is ionized or not, in 2004, Furlanetto, Zalderriaga, and Hernquist (FZH) defined the parametrization for this. They built their model off of a few initial assumptions. The first assumption was that the main ionizing photons were UV or "hard" photons in which these high energy photons ionize the gas efficiently. This means that the hydrogen bubbles have "hard" edges rather than "fuzzy" edges, which allowed them to designate voxels as either ionized or not. This model allows them to associate large scale fluctuations in the resulting ionization field and to apply the excursion set formalism in determining the resulting size distribution of the ionized bubbles. 21cmFAST and 21CMMC utilize this three parameter model in the code itself. This parametrization has become common because 21cmFAST is so popular. It is designed to give power spectra that looked similar to radiative transfer and gave bubbles that looked right but it

can't seem to fit a better simulation. The three main parameters are:

- $\zeta$ Ionizing Efficiency
- $T_{vir-min}$ Minimum Virial Temperature (K)
- $R_{MFP}$ Max Mean Free Path (Mpc)

### 3.4.1 Ionization Efficiency

The ionization efficiency is the number of photons per baryon per dark matter halo. It is an important parameter for considering how luminous a light source is. That is, large galaxies with a lot of stars and low opacity will have a high efficiency because they will be emitting a large number of photons into the surrounding IGM. Thus, a high ionization efficiency would mean that Reionization would occur more quickly. It is defined by the fully degenerate equation:

$$\zeta = A_{He}[\frac{f_*}{.1}][\frac{f_{esc}}{.1}][\frac{N_{ion}}{4000}] \tag{9}$$

where $A_{He}$ is a correction factor for a universe that is not purely hydrogen. $f_*$ is the fraction of baryons in a galaxy that turn into stars, also known as the star formation efficiency. $f_{esc}$ is the escape fraction, which is the fraction of ionizing photons that escape into the intergalactic medium from a given galaxy. The escape fraction is a poorly constrained parameter as we do not know too much about the nature of the first galaxies and whether they easily allowed photons to escape. $N_{ion}$ is the number of ionizing photons per stellar baryon. We know this number for population I and II stars, but it is far less constrained from population III stars. This equation is sometimes called the Drake equation of Reionization as there are many different parameters that influence the efficiency but we do not have strong constraints on any of them.

There are two main difficulties in trying to determine the ionization efficiency, the first of which is the recombinations of hydrogen that can occur after ionized. If a photon traveling outward has sufficient energy to hit a hydrogen atom and ionize it, there is still the possibility of the hydrogen atom recombining with the electron. This means that not all of the hydrogen that gets ionized will stay ionized. These recombinations have to be considered because it will influence the ionizing photons per baryon around galaxies. Additionally, ionization substructure needs to be considered.

### 3.4.2 Minimum Virial temperature

Virialization is the point in which a dynamically relaxed self-gravitating object collapses into a stable state. Thus, the minimum virial temperature is the minimum temperature of the gas in a halo at which it virializes and forms a stable structure of a star-forming galaxy. In practice, this is the minimum halo mass for star formation to occur, so as a parameter it tells us about the population of photon emitting galaxies.

The larger the minimum virial temperature, the larger the halos need to be to virialize, which would result in larger and more sparsely populated galaxies. Typically a value of $10^4 K$ is chosen because

this is when atomic cooling becomes efficient [34]. If the virial temperature is larger, while the efficiency and mean free path are kept constant, Reionization will slow down. This is because there are fewer, more clustered sources. They will be more clustered and reside in more concentrated regions, while producing bigger bubbles. The larger bubbles will increase large scale power. Though they are larger, they would need a larger efficiency to have the same impact on the IGM as a more numerous set of smaller galaxies might have. In other words, if massive galaxies drive Reionization then they have to have high efficiencies so you need more photons per galaxies to get ionized IGM but if you keep the efficiency the same then there are not enough photons leaving the galaxies.

We can relate the minimum virial temperature to a corresponding mass of these galaxies:

$$M_{min} = 10^8 h^{-1} (\frac{\mu}{0.6})^{-3/2} (\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2})^{-1/2} \times (\frac{T_{vir}}{1.98 \times 10^4 K})^{3/2} (\frac{1+z}{10})^{-3/2} M_\odot \tag{10}$$

where $\mu$ is the mean molecular weight, $\Omega_m^z = \Omega_m(1+z)^3/[\Omega_m(1+z)^3 + \Omega_\Lambda]$ is the cosmology-dependent evolved matter energy density, and $\Delta_c = 18\pi^2 + 82d - 39d^2$ where $d = \Omega_m^z - 1$. We will use the above equation to calculate the minimum virial temperature from the S1 radiative transfer simulations given the virial mass. This minimum mass can be subject to numerous forms of feedback within galaxies, which often cannot be modeled in cosmological Reionization simulations because it occurs on scales below resolution. However, some subgrid physics may include negative supernova feedback on star formation, photoionization to prevent gas infall in small halos, and UV radiation influencing the cooling rate [19]. The above relationship between the virial mass and the virial temperature is also still important because we tried to run 21CMMC to converge on the virial mass parameter rather than the virial temperature parameter since the former is closer to the actual radiative transfer simulation setup.

### 3.4.3 Mean Free Path

The mean free path of a photon through the IGM is the average distance a photon will travel before hitting a neutral hydrogen and subsequently ionizing it. Typically, it can tell us something about the characteristic bubble size within any given simulation, and in essence decides the effective ionized bubble size. Within the 21cmFAST code, it is a key parameter set as a maximum distance for the photons to travel and thus is an effective "maximum" bubble size. The mean free path is set this way more out of convenience rather than out of a physically motivated purpose. It is somewhat fair to use this assumption because it limits the bubbles from growing too large. Ionizing photons are not "allowed" to travel distances farther than $R_{mfp}$. However, this is not as physically motivated given that a real mean free path of a photon can vary and grow over time. A real mean free path also is determined by a photon eventually hitting neutral hydrogen, whereas in the code the photon would just arbitrarily disappear after going a certain distance. This use of the mean free path in the code differs from typical radiative transfer simulations, which use ray tracing for individual photons, so there is no true mean free path. Rather than having a typical ionized bubble size being set, there is a fuzzier and more complicated fractal-like geometry as each individual photon has a different distance it will travel. In a full radiative transfer simulation, the mean free path is additionally difficult to model because of self-shielding regions that influence how far the photons can travel. Later versions of 21cmFAST and 21CMMC also try to

take recombinations of hydrogen into consideration. As hydrogen is ionized, it could also recombine with surrounding electrons, which would subsequently increase the neutral fraction and inhibit the efficacy of photons leaving any given galaxy.

### 3.5 Intermission

Thus up to this point we have described how simulations that model the early universe work. The 21cmFAST code begins with boxes of particles with a given resolution. Using the Zeldovich approximation, every particle in each voxel is given initial position and velocity conditions to evolve forward under the force of gravity. We evolve these boxes over redshift as the velocities are scaled according to time moving forward in the universe. These particles, which are often on the order of a size of a galaxy because of resolution and computational time purposes, eventually begin to form structures that can be recognized as dark matter halos. A halo finding algorithm under the Press-Schecter and Excursion Set formalisms can identify the halos on decreasing scales as it tries to find what will eventually become virialized objects. This assumes each halo has a galaxy, which generates the ionization bubbles given the FZH parameters of a maximum mean free path (radius) of photons, the minimum virial temperature of halos, and the efficiency of photons escaping these galaxies to ionize the nearby intergalactic medium [11].

## 4 Methods

Let's recall that my project tests the robustness of 21CMMC on these mock measurements from other simulations as opposed to that in 21cmFAST. This will help us determine whether 21CMMC can properly reconstruct early universe parameters. I originally pulled from a set of power spectra created by Adam Lidz at the University of Pennsylvania using the radiative transfer S1 simulation by Matthew McQuinn [23]. It uses some approximations in building the ionization field directly off of the dark matter distribution, then using ray tracing algorithms. This simulation was an improved version of those in [38] [37].

A single 21cmFAST run can take about 10 minutes on a single core processor. Using Oscar, which is a research computing cluster with several hundred multi-core nodes at Brown, we ran the computing jobs on a single node. Future work entails parallelizing this computation.

### 4.1 21CMMC Code

Thus we have described how the 21cmFAST code in 21CMMC is a semi-numerical Monte Carlo Markov Chain code that takes data or simulated data boxes of this ionized universe from another code called 21cmFAST. Mock measurements from 21cmFAST or other simulations are made in the form of the power spectrum. Those measurements are thrown into 21CMMC which will go through the same process as 21cmFAST. The difference here is that the ionization efficiency, photon mean free path, and the minimum virial temperature are not set to a particular value. Rather, 21CMMC will put in random values and try to run 21cmFAST over and over until it creates an output that resembles the original

input. This is useful because when given these mock power spectra, we do not actually know the value of the parameters (i.e. conditions of the universe). The power spectra merely tells us where the neutral hydrogen is clumping where. Thus, 21CMMC provides a framework that can recover constraints on the above parameters, and accommodate many different EoR models coupled with priors from other chosen model parameters. 21CMMC could be a powerful tool for when telescopes like HERA or MWA have the sensitivity to measure the 21cm signal in the early universe parameters. 21CMMC can be used to optimise: (i) interferometer designs; (ii) foreground cleaning algorithms; (iii) observing strategies; (iv) alternate statistics characterising the 21cm signal; and (v) synergies with other observational programs.

My project tests the robustness of 21CMMC on these mock measurements from other simulations as opposed to those from 21cmFAST. This will help us determine whether 21CMMC can properly reconstruct early universe parameter.

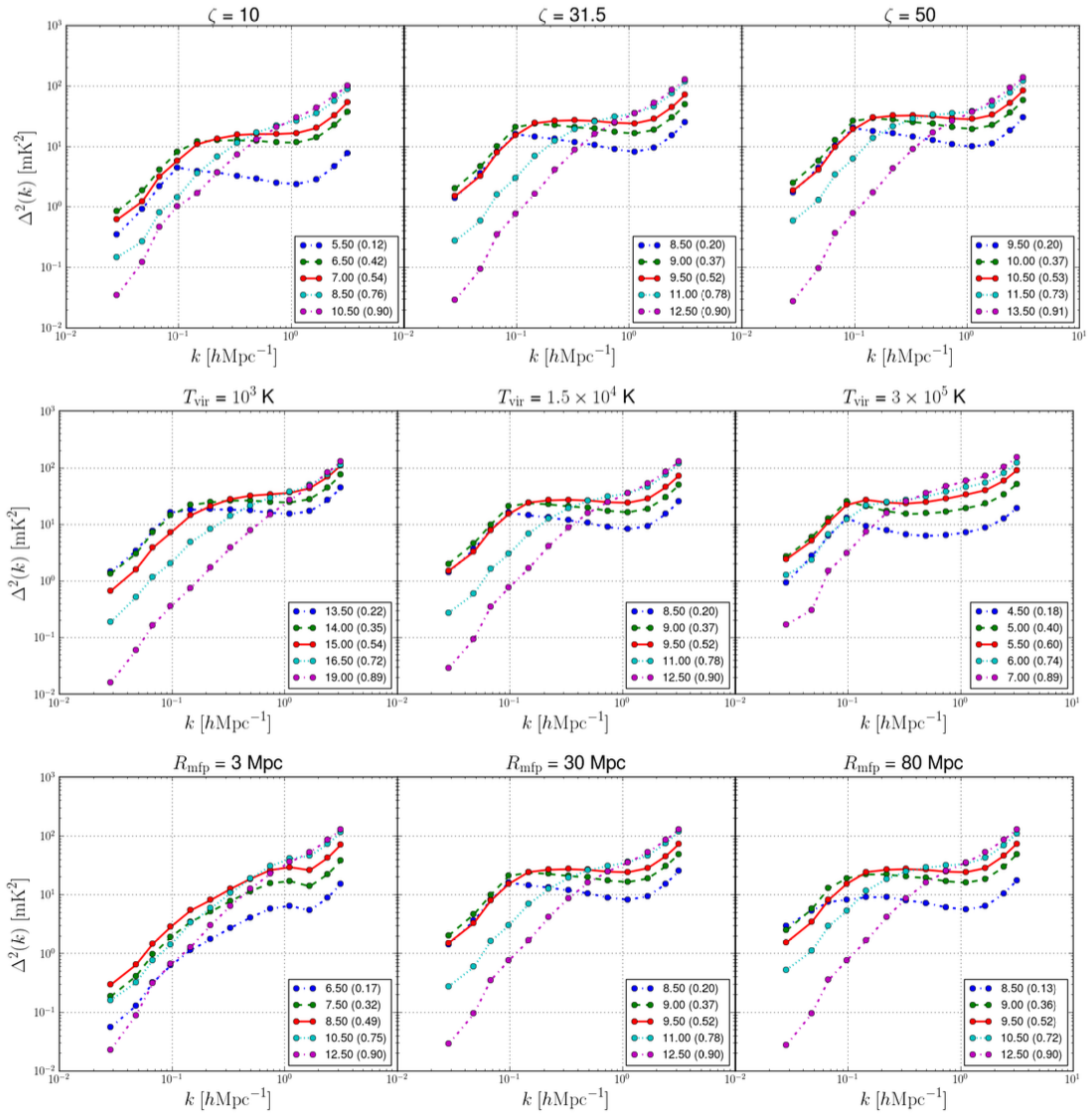
## 4.2 Monte Carlo Sampling in 21CMMC

Monte Carlo (MC) algorithms are a means to simulate complex stochastic processes. They rely on a set of random sampling within a large parameter space with some probability distribution. MCs are generally used for: sampling a parameter space, using that sampling with more statistics to optimize a model, and numerical integration. 21CMMC utilizes an affine invariant Monte Carlo Markov Chain algorithm (MCMC), which falls into the second use of MC. More typical versions of Monte-Carlo include Metropolis-Hastings algorithm, which requires several long computation chains, and can thus be extremely inefficient. The Markov Chain aspect of the Monte Carlo implies that the next step in the sampling is only reliant on the present step and no steps prior to that. The Monte Carlo Markov Chain applied in 21CMMC is a particular subset of Monte Carlo algorithms that samples from a *posterior distribution*.

21CMMC uses a goodness of fit statistic, i.e. the spherically-averaged 21cm power spectrum in order to quantitatively compare an 21cmFAST realization with the mock observation [14]. This mathematically appears when comparing the dimensionless power spectrum: $\Delta^2_{21}(k,z) \equiv k^3/(2\pi^2 V)\delta\bar{T}_b(z)\langle|\ \delta_{21}(\mathbf{k},z)\ |^2\rangle_k$ where $\delta_{21}(\mathbf{x},z) = \delta T_b(\mathbf{x},z)/\delta\bar{T}_b(z) - 1$

To put it simply, 21CMMC takes in a mock power spectrum from some outside source. It then uses the MCMC to tweak the three key ionization parameters (efficiency, mean free path, and minimum virial temperature) to run 21cmFAST. Then it takes the resulting 21cmFAST power spectrum and runs a goodness of fit between the new 21cmFAST power spectrum and the initial mock power spectrum. 21CMMC continues to repeat this process, each time comparing the new 21cmFAST power spectrum that results. If the new goodness of fit for a realization of a given set of sample parameters is better than what it was before, it will use the new 21cmFAST power spectrum instead of the last one for further comparison. 21CMMC does this several thousand times until it converges on a 21cmFAST power spectrum that results in the best goodness of fit. The software it uses is the python module EMCEE by Foreman-Mackey et al. [10], which utilizes an affine-invariant ensemble Monte Carlo sampler (as opposed to typical Metropolis-Hastings) developed by Goodman and Weare [17].

### 4.3 Understanding Shape of Power Spectra




(11)

[34] The efficiency $\zeta$ does not change the shape of the power spectrum, but it will change how quickly Reionization occurs and when it occurs. This is not surprising since the sources driving Reionization are not changing for a given efficiency. If we have a larger minimum virial temperature then we would have fewer but larger halos that can be stable enough for star formation, which would result in fewer sources of photons. However, these few galaxies would be brighter. This has some influence on shape of the power spectrum itself as well. This is because when the virial temperature is large, then there is a sparse population of large galaxies driving Reionization, which would cause large scale (small k) power to jump. Conversely, if the virial temperature is small, then their are less massive more numerous galaxies driving Reionization, so there will be a jump in small scale (large k) power. The mean free path influences the shape of the power spectrum significantly. The smaller the mean free path, the more larger scale (small k) power is reduced. This is particularly noticeable for very small mean free paths.

## 4.4 Radiative Transfer Power Spectra

As stated above, the radiative transfer power spectra came from an N-body radiative transfer code by Matthew McQuinn et al. [23]. It uses the N body code Gadget2 [39] to model the density field in boxes with $1024^3$ particles and box dimensions of $65.6h^{-1}Mpc$ . After identifying halos using a friends-of-friends algorithm [29], it uses modified algorithms from a radiative transfer hydrodynamic code by [37] [38], to cast rays outwards from a list of sources to calculate the ionization field. These are the same approximations discussed in history of simulations section 3.1.1 in which the ionization field is directly built off of the dark matter distribution. McQuinn et al. did a series of simulations to test how the changes in the algorithms and initial parameter conditions influenced Reionization. In this case, we used power spectra from the S1 simulation [23], which statistically added in smaller halos using subgrid physics. We initially included the full range of k values that the S1 Radiative Transfer power spectra provided (up to k $\approx$ 10). However, this is not realistic given instrument sensitivities, so we eventually came to limit the k values to 2.

## 4.5 21cmFAST2 Power Spectra

An updated version of 21cmFAST has been developed to contain new physics improvements including inhomogeneous, sub-grid modeling of recombinations, as well as calibrated photo-heating feedback on star formation during Reionization[**?**]. The impact of these sub-grid recombinations is to produce a lot of small scale power, equally resulting in a large number of small HII regions. Attempting to fit these with an $R_{mfp}$ should therefore result in low values being strongly preferred.

## 4.6 Switching From CosmoHAMMER to Polychord

The original set of runs were done using the CosmoHAMMER is a python package that embeds emcee [**?**], which is an implementation of an affine invariant ensemble sampler by Goodman and Weare [13]. It is an MCMC sampler that we give a set number of burn ins and samples for 21CMMC to use. For the sake of efficiency and better quality computation, we switched to a different sampler called PolyChord, which is a nested sampling algorithm tailored for high-dimensional parameter spaces [16]. Polychord doesn't have a maximum number of samples parameter, it just keeps running until it reaches its own convergence criterion.

## 4.7 21CMMC with Recombinations

Up to now, we have been using a slightly older version of 21CMMC. Recently, a newer version has been released, which includes inhomogeneous sub-grid recombinations. Additional parameters are included such as the X-ray energy threshold for self-absorption by the host galaxies $E_0$, the integrated soft-band luminosity $L_X$, and the X-ray spectra index $\alpha_X$. This version can also support simulated 21cm signal in co-eval and light cone formats and light-of-sight redshift distortions as well. For the purpose of this work, these new parameters are not important, but more information can be found about them at [15].

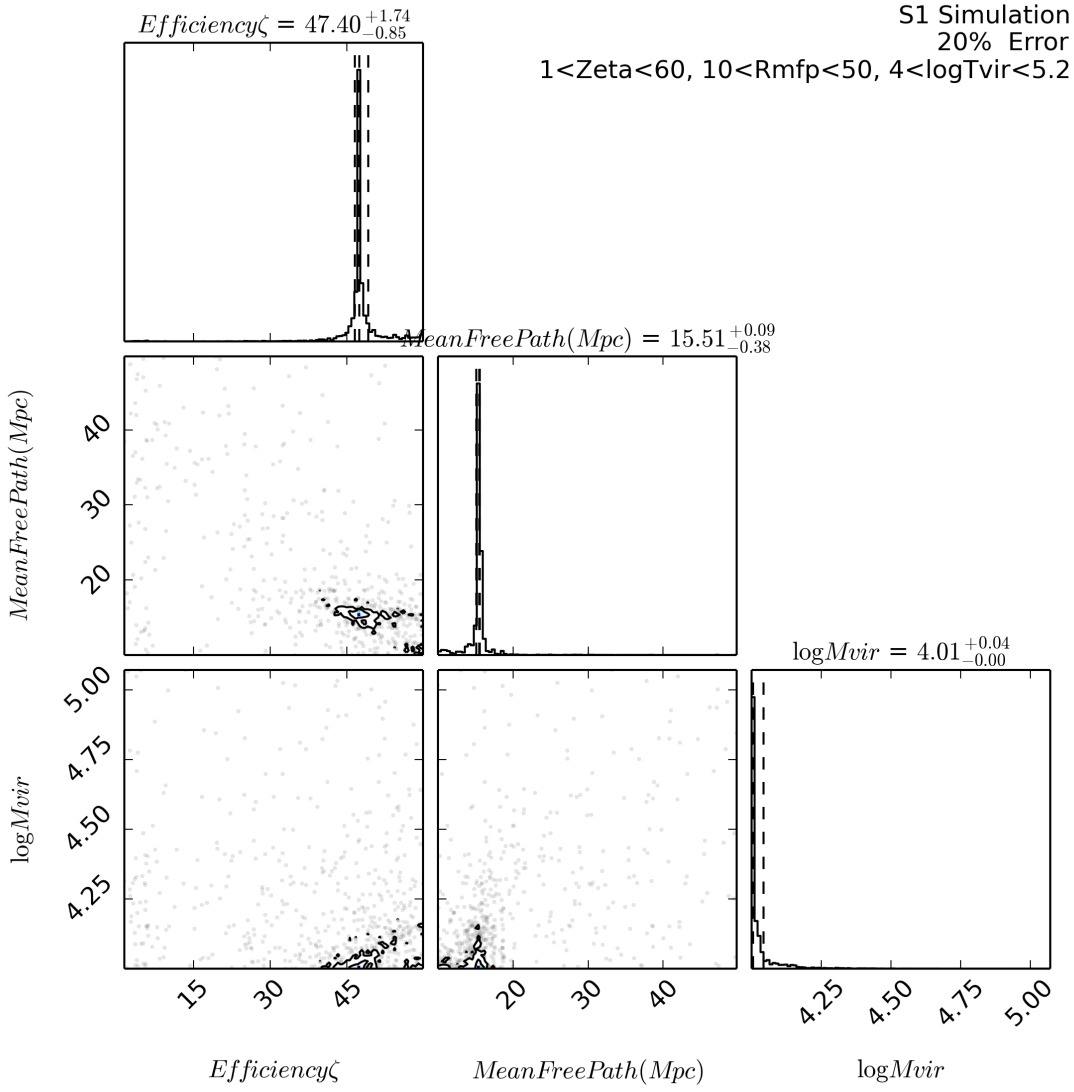
Rather, our main concern will be inhomogeneous recombinations, which can slow down Reionization as the ionized hydrogen recombines. This occurs as the ionized regions approach typical separation of the photon sinks. This means more photons are needed over time to fully ionize the IGM, which causes a decrease in power on large scales [15]. This sub-grid application was motivated by the work done in [36], and allows the code to adopt a flat prior over $R_{mfp}$. Typically they denote the effective horizon set by sub-grid recombinations. Before, the $R_{mfp}$ was never actually a mean free path, but rather an "effective [max] horizon" for photons. In the new 21CMMC, which uses the subgrid prescription of recombinations, there is no single "effective [max] horizon" or $R_{mfp}$ because it will evolve over redshift. Though, if one were to try to calculate an actual mean free path of photons, it would typically be larger then the $R_{mfp}$ parameter since the photons would be free to propagate indefinitely until they hit a hydrogen.

The motivation to incorporate this more detailed 21CMMC is that will get rid of the traditional mean free path parameters as we have been working with it (as a maximum horizon size). Additionally, by including new sets of parameters, it will be trying to account for more physics in the version of 21cmFAST that it uses, which will provide opportunity to have a better fit for future power spectra.

# 5 Results

I have been working on using power spectra from Adam Lidz at the University of Pennsylvania. These power spectra come from a full radiative transfer code in which fewer short cuts are taken in the computation of the simulation itself. This means that certain parameters in the code may not directly translate to those from 21cmFAST. The most apparent of this is the mean free path parameter, which is not a true mean free path in the 21cmFAST code. Using these power spectra, I was trying to determine what parameters 21CMMC would pop out. I subsequently took these parameters and threw them into 21cmFAST to get power spectra (in the hopes that they would look like the S1 spectra).

## 5.1 Power Spectra with 20 Percent Error




Above we have 1<Zeta<60, 10<Rmfp<50, 4<logTVir<5.2. I recreated a set of error files that would print out errors that were 10% and 1% (10-20% total error of the signal) the value of the power spectrum itself. Above we see the results from a run of a total 20% error. The peaks have converged very cleanly on what also seem to be pretty reasonable values. Though this plot is the cleanest plot with regards to convergence of parameters, it is misleadingly so. The fact that it has the 20% error is not in line with true error values. True error values have a scale and redshift dependence such that more importance is given to larger scales (smaller scales are harder to see with telescopes) and later redshifts (earlier redshifts are harder to see with telescopes). It is encouraging that the efficiency and mean free path are able to converge, but it was exceedingly difficult to get the mean free path to do so. It is the least physically motivated parameter and thus the most difficult to properly capture. The build up of the virial temperature at 1e4 also is not the most encouraging because this is a hard coded limit in the simulation, and it is difficult to know if this is a true convergence or an artifact of how the code is written.

## 5.2 Renormalized Error Files

Considering different modes of adding error is significantly important when considering the reality of the measurements. There is going to be a redshift dependence as values at higher redshifts will have a smaller signal to noise ratio. Additionally, larger k (smaller scale) power spectrum values will also have larger errors such that it scales as $k^{-3}$. 21CMMC uses a set of error files that correspond to the mock power spectra for each job that we run. Rather than having a fractional value artificially added onto the power spectra when using 21CMMC, we began creating errors files that had a redshift and scale dependence.

$$\sigma \approx \sigma_o \times (\frac{1+z}{1+z_0})^2 \times (\frac{k}{k_o})^3 \tag{12}$$

where $k_o \approx .2, z_o \approx 7.32, \sigma_o \approx 1mk^2$, We trust the lowest Redshifts the most ; we want to trust the high k's and the high redshifts less Below we have the results of a 21CMMC job after having included these normalized error files.

Above we clearly see that the mean free path does not want to converge. The log of the virial temperature converges on a value of about 4.35, which is not too far off from the fiducial value of 4.0. Meanwhile, the efficiency in this case in converging at a reasonable value of about 17.7. This is around a typical number on the order of 20, which is the typical order of efficiency for Reionization [34].

## 5.3 Virial Mass Runs

Before trying to do the calculations for the virial temperature given the virial mass from the S1 simulation, I tried to change the settings in 21CMMC so that it would converge on the virial mass as a parameter instead. Unfortunately, when we tried to do this, it would not converge on values that had much meaning.

## 5.4 21cmFAST2 Results

We ran several jobs in 21cmFAST2, but the mean free path kept converging on something very small. It was not uncommon for 21CMMC to try to converge on very small mean free path values. In order to determine whether this is a legitimate result or some other computational residual, we made larger higher resolution boxes using 21cmFAST to be used in 21CMMC. This would mean the pixels could actually be smaller than 5Mpc, so that if the mean free path does want to converge smaller than 10Mpc, we would see what this value is. This is also a much more computationally intensive practice because your increasing the computation 10-fold.

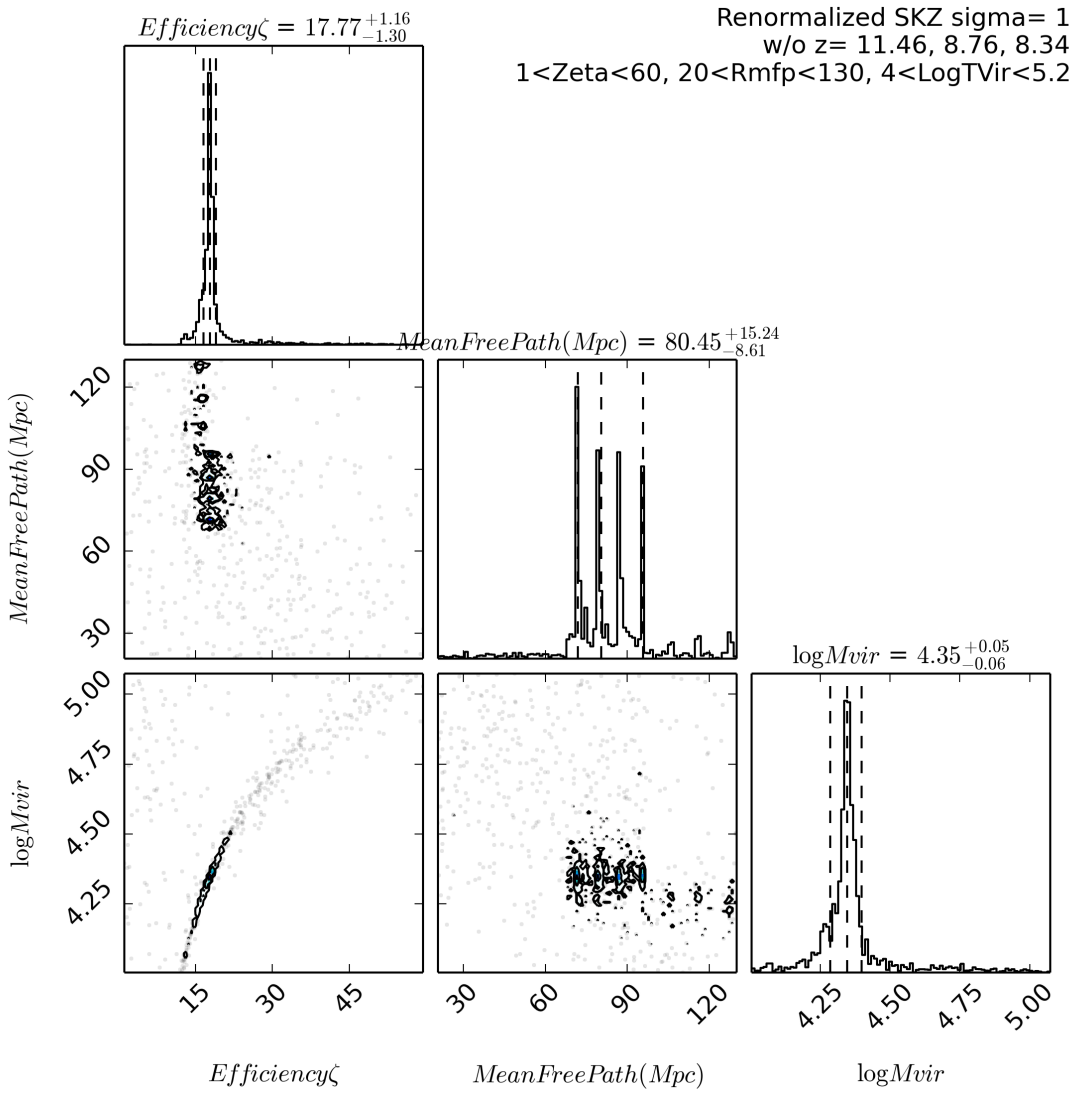



Figure 4: Above we see a run of 21CMMC on the S1 power spectra using the renormalized error files for a few select later redshifts. The earlier redshifts z = 11.46, 8.76, and 8.34 were removed to see if the mean free path can properly converge. There were many times in which the mean free path was simply unable to converge. Even above, it is clear the mean free path is giving a multi-model answer, which is likely due to the changing bubble size as a function of redshift in the full radiative transfer simulation. In 21cmFAST, the mean free path corresponds to a single maximum ionized bubble size. In the S1 simulations, there is no true mean free path, but rather the bubbles are "fuzzier" and have more irregular structures. The photons do not have a maximum distance to propagate as they do in 21cmFAST, but rather will travel outward until they hit an actual hydrogen molecule. Even then, as the ionized bubbles are forming, they will be growing as a function of redshift, which would mean there is no single answer for a maximum bubble size to converge on. Rather the mean free path would grow over time. Additionally, even within the S1 power spectra for each redshift, there are pointy oscillations, especially for larger scales (i.e. the scales that have smaller errors and thus larger importance). The mean free path produces a knee like bend in the power spectrum, so there are many of these "knees" in the power spectra. This is likely due to the existence of multiple sized bubbles in the full radiative transfer simulation. 21CMMC will try to match to each of these knees, which would give more than one mean free path answer.

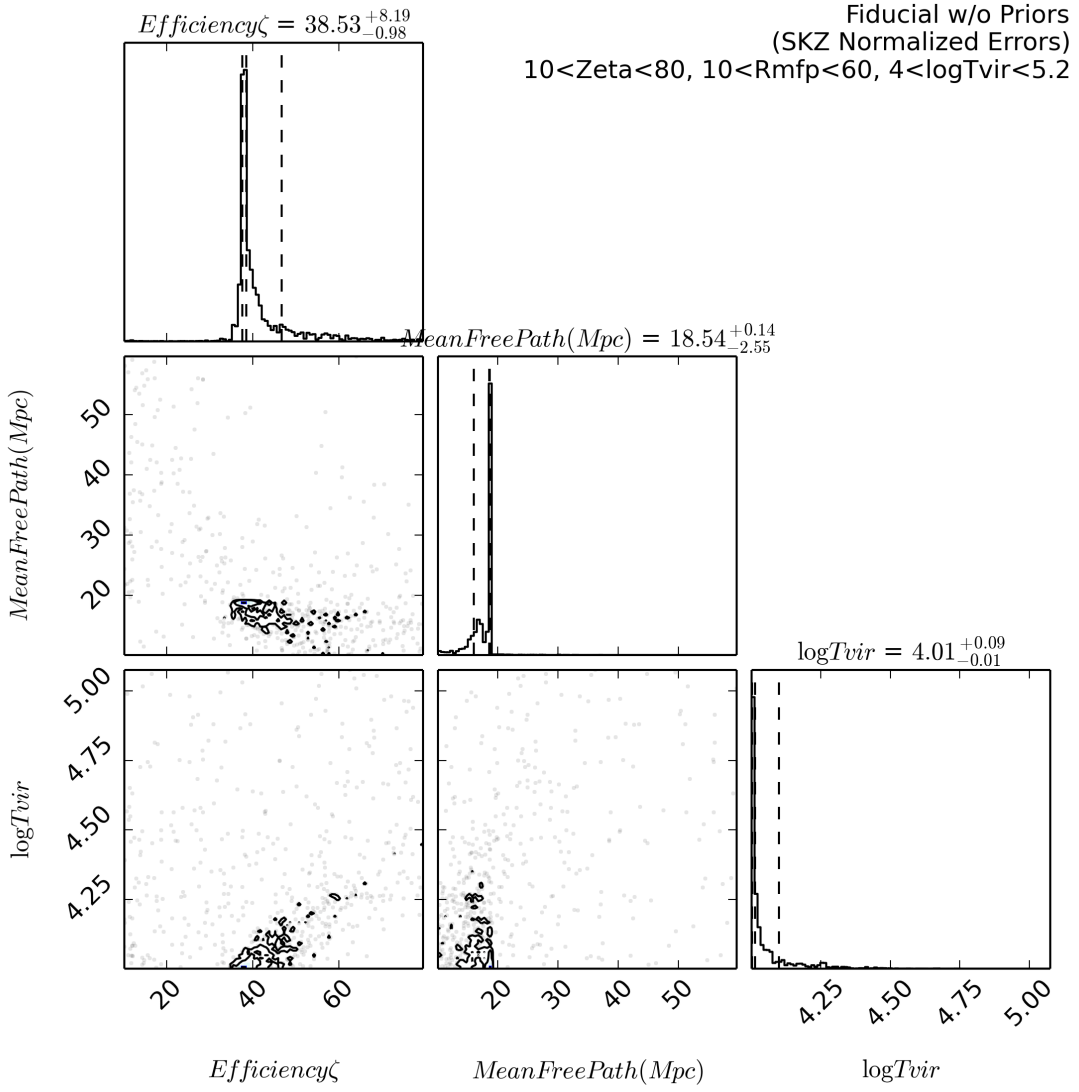



Figure 5: Within the above plot, we see that 21CMMC has converged onto a value for the mean free path, but contains a discontinuity next to this peak. It is likely this discontinuity is completely unphysical and a remnant of the code failing to converge. This could be due to small errors that cause the noise to dominate and result in these discontinuities away from the actual result. In particular, there is a characteristic knee present in power spectra during reionization. Within the S1 Radiative transfer simulations, this knee is highly pronounced. If you recall, the mean free path will influence the scale that could influence how the code is able to converge on a single mean free path. This makes picking the error files more important than one believe. Rather than having a 10% modification error of the signal added on top of another 10% error, this may mean a more complicated set of error files would prove useful.

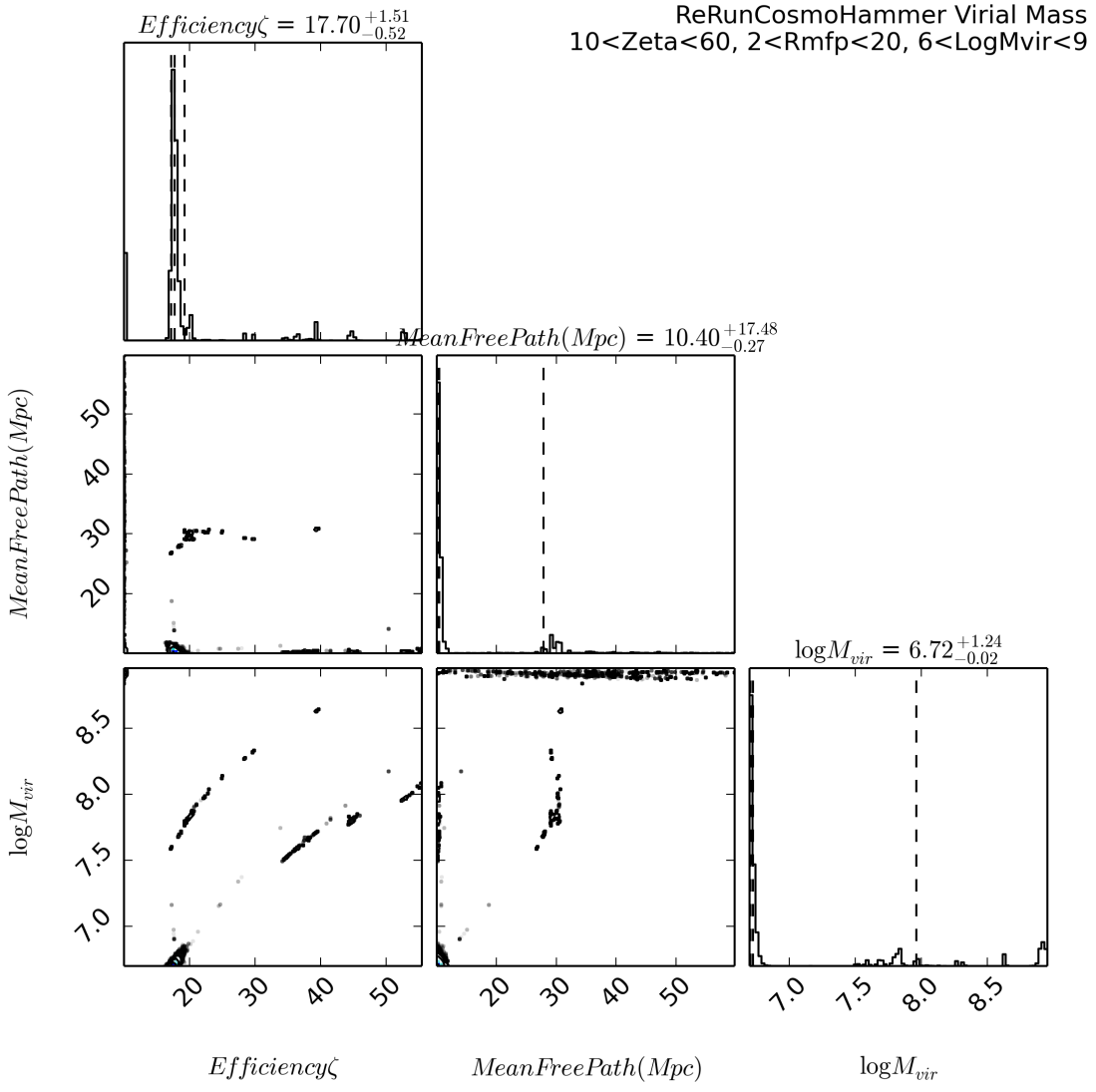



Figure 6: Here we used original sampler CosmoHammer to test the virial mass as a parameter in 21CMMC. All of the values converged at the lower limits of the histograms, so we could not determine whether these were an artifact of 21CMMC not working or actual parameter values. After testing this a few times with no luck, we moved forward to see whether 21CMMC could recreate itself with the virial mass setting on from 21cmFAST

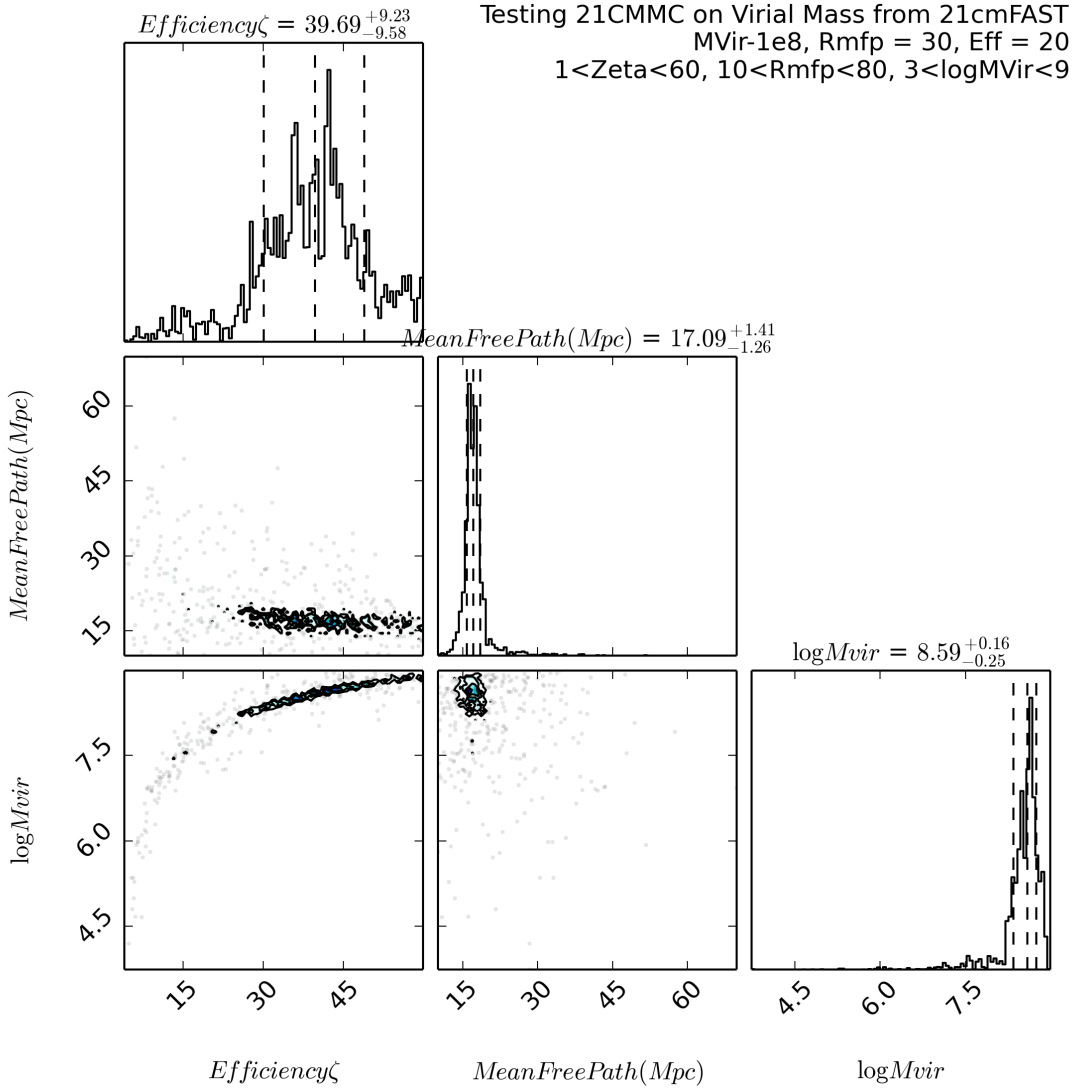



Figure 7: Above is a plot showing the 21CMMC input after having done a run on 21cmFAST with the virial mass job. This was an attempt to see if 21CMMC could recreate itself given power spectra from 21cmFAST. This is because we had been unsuccessfully trying to run 21CMMC on the S1 power spectra with the virial mass turned on. In those runs, it was clear the virial mass was giving the wrong answer. To do the switch from determining the virial temperature to determining the virial mass within 21cmFAST is a simple change in one of the header files of the code that turns the option on. Additionally, I had to add an extra few lines of code in one of the driver files to set the virial mass itself when it is non-zero.

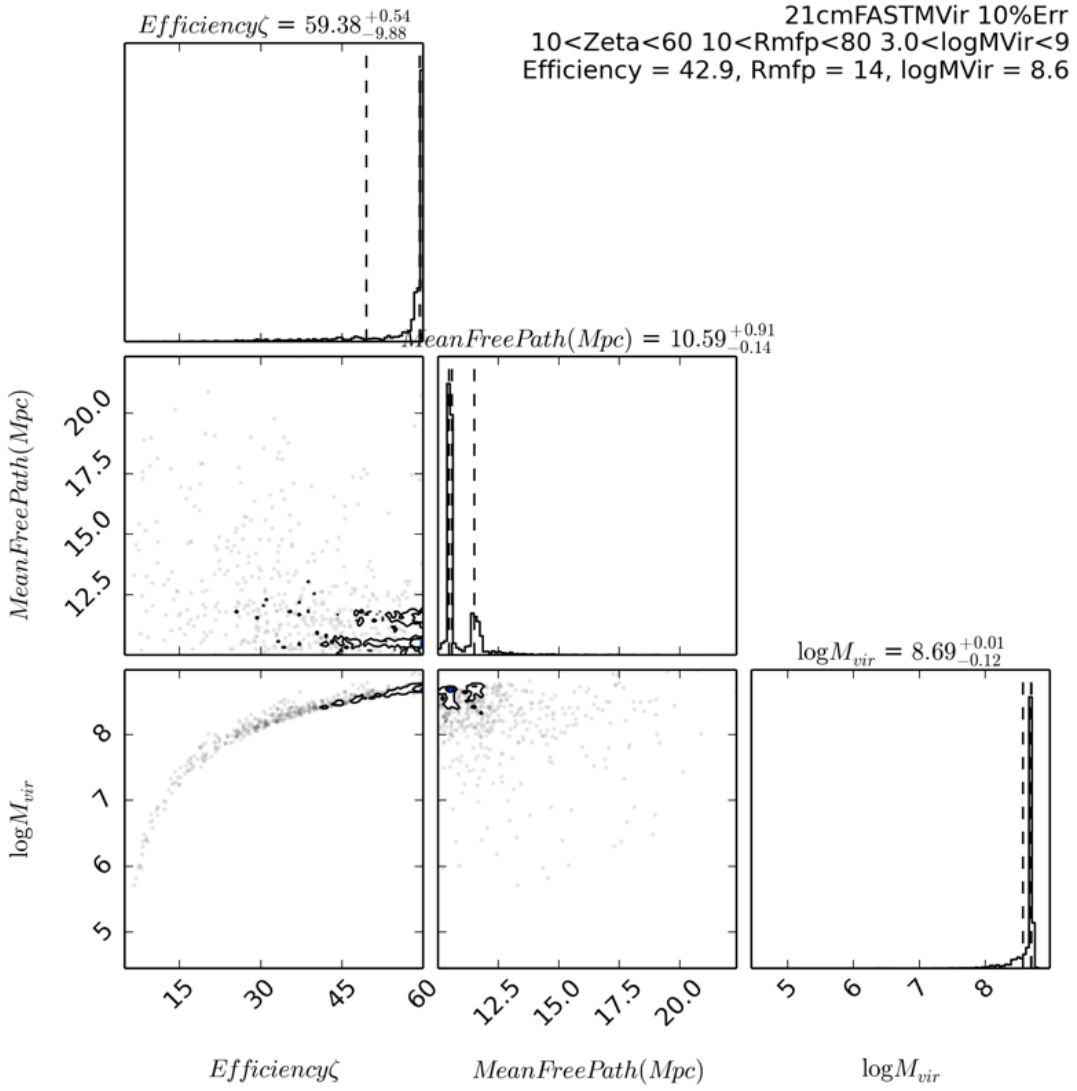



Figure 8: Above is a plot showing a run of 21CMMC with the switch to virial mass from virial temperature being made. In this case, I ran 21cmFAST with the virial mass on for values of the $\log M_{vir}$=8.6, the $R_{mfp}$=14, and $\zeta$=42.9. We chose these values because they were the outputs of a job very similar to the above corner plot in 7. Thus, this job was an attempt to "close the loop" to see if 21CMMC could converge on a similar set of values after doing this a couple of times. Unfortunately, this is not the case. The virial mass is not too inaccurate, but the mean free path is gets even smaller with a major discontinuity, and the efficiency pushes up on the upper edge of the efficiency sampling range. After seeing these results, we decided to try the 21cmFAST2 power spectra. The motivation of this was to take a step back to try a power spectra from a simulation different from 21cmFAST, but more similar than the full radiative transfer simulations.

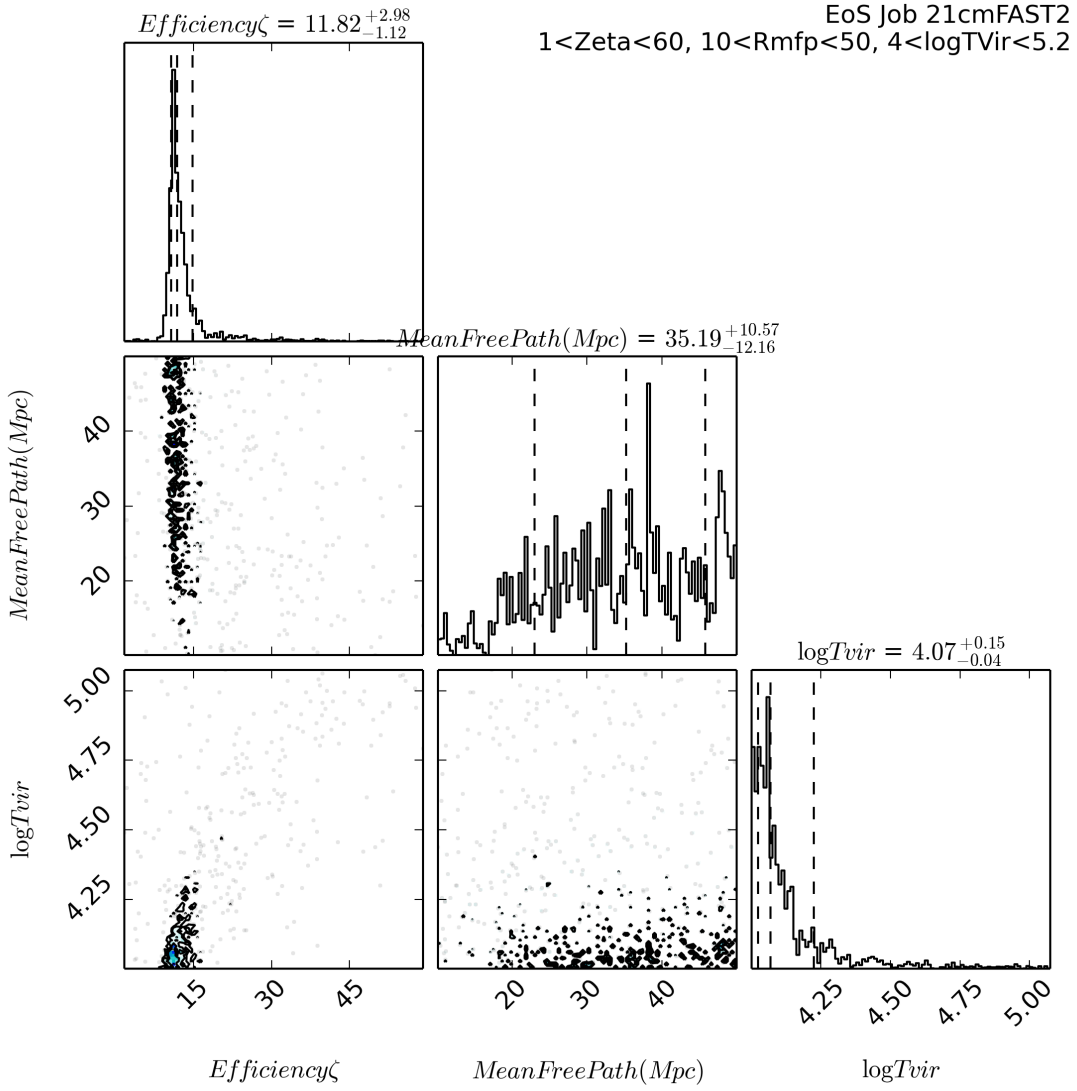



Figure 9: Above is a 21CMMC run on the 'faint galaxies' power spectra from 21cmFAST2. The efficiency to produce the power spectra was 20, the $\log T_{vir}$ is 4.3, and there is no mean free path. The correct virial temperature value is just out of the first sigma value of the mean. This is probably influenced by the mean free path not converging. The efficiency is also converging at 11.8 rather than 20.

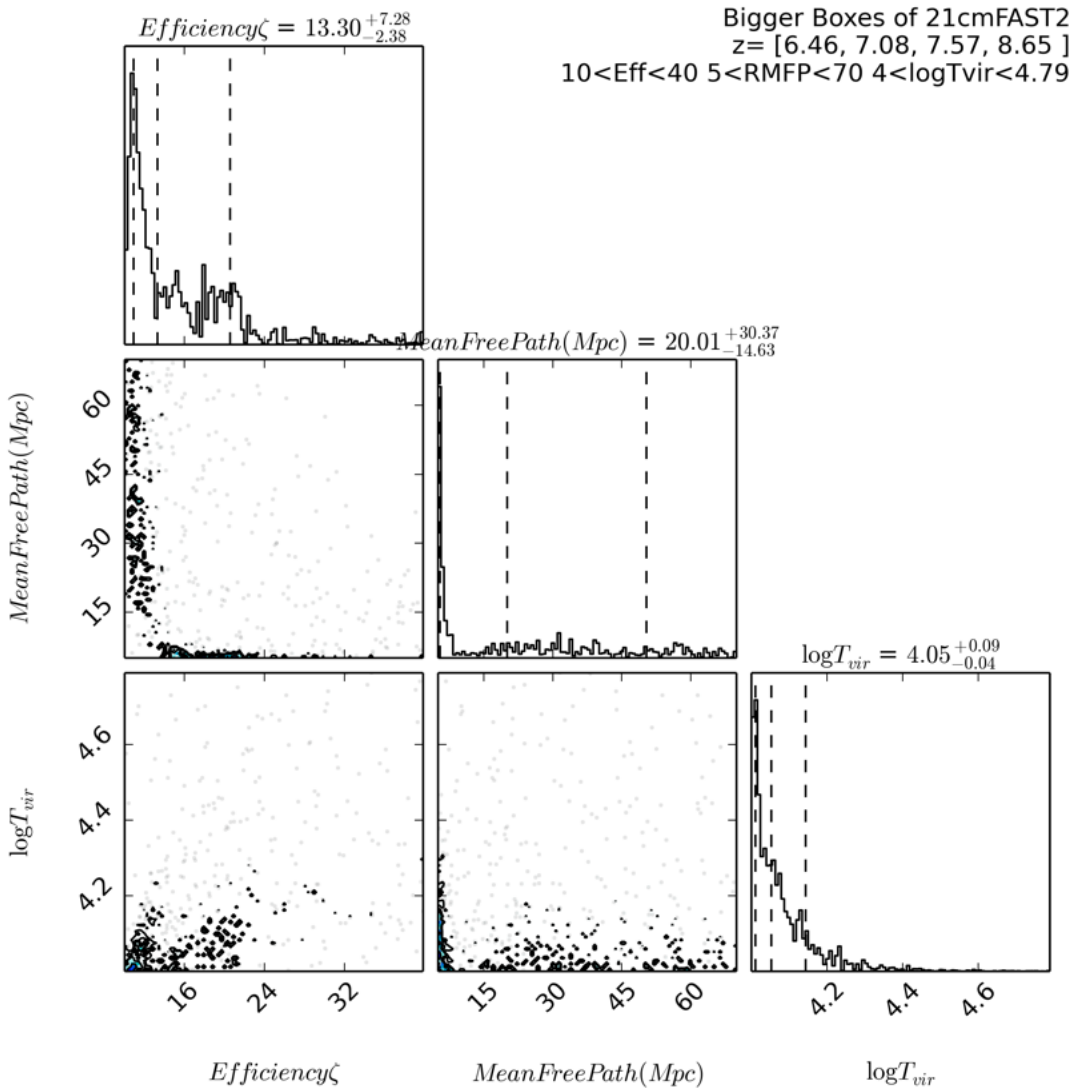



Figure 10: Above are the results for 21cmFAST2 after going back to increase the size and resolution of the boxes that we were using. I doubled the number of cells on each axis per box from 256 to 512 units and from 256Mpc to 512Mpc. A single pixel would be on the order of about 5Mpc, which allowed us to try and probe smaller mean free path values. This was motivated by the fact that we had yet to be able to get 21CMMC to converge on the right values, which were 20 for the efficiency and 4.3 for the log of the virial temperature. The virial temperature would bunch up at the lower end of 1e4K. The mean free path would either not converge or it would also bunch up at the lower end of the interval.

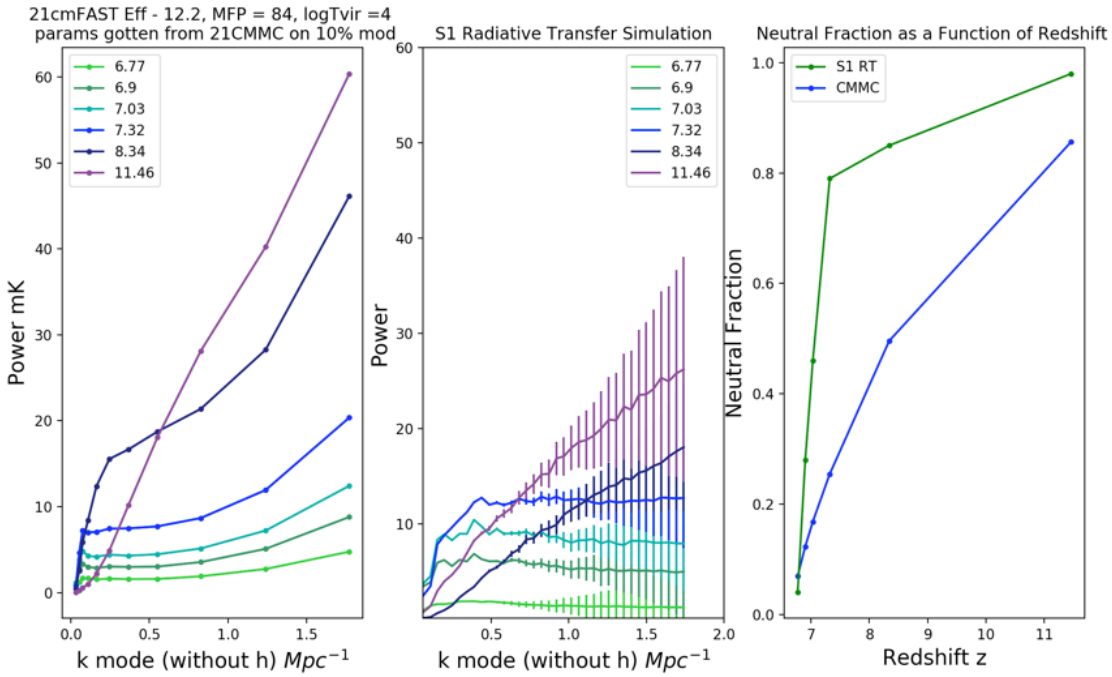



Figure 11: Above we are comparing the power spectra coming from 21cmFAST after giving it an input of Efficiency = 12.2, MFP = 84Mpc and logTvir=4. These numbers match the output from a 21CMMC job that converged on these values. This is to follow up testing whether the 21cmFAST will be able to recreate the S1 simulation's ionization history. The left plot is the output from 21cmFAST, the middle plot is the original set of power spectra with a set of redshift and scale-dependent error bars, and the right plot is the neutral fraction as a function of redshift. As seen in the right plot, 21cmFAST does a good job of tracking the neutral fraction for lower redshifts, but at higher redshifts it diverges from the actual neutral fraction from the radiative transfer simulations. The 21CMMC neutral fraction drops much quicker than the S1 neutral fractions until the later redshifts as they both become fully ionized.

# 6 Discussion

## 6.1 Comparing 21CMMC with Actual Data

### 6.1.1 Mean Free Path $M_{mfp}$

Until now, we ran 21CMMC on power spectra from Matt McQuinn's S1 radiative transfer simulations from[23], and we saw that 21CMMC was unable to produce a single value for the mean free path. There is no true mean free path in McQuinn's simulations as defined by 21cmFAST. The photons in the former are free to propagate until hitting a hydrogen atom, whereas they have a maximum distance they can propagate in 21cmFAST. This means that there is no corresponding "max" effective horizon for the photons in the McQuinn simulations.

### 6.1.2 Minimum Virial Temperature $T_{virmin}$

The other two parameters are values that we can determine from the original simulation run itself. McQuinn quotes the minimum virial mass of $M_{vir} < 10^8 M_\odot$. These N-body simulations do not resolve masses less than $10^9 M_\odot$ , but they use an analytical prescription for a minimum atomic cooling mass for the haloes to be $10^8 M_\odot$. This is the mass we will use when calculating the luminosity of haloes of mass m in ionizing photons per second per halo. Recalling the relationship between the virial temperature and the virial mass in equation 8 [6]:

$$M_{min} = 10^8 h^{-1} (\frac{\mu}{0.6})^{-3/2} (\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2})^{-1/2} \times (\frac{T_{vir}}{1.98 \times 10^4 K})^{3/2} (\frac{1+z}{10})^{-3/2} M_\odot$$

We can invert this expression to obtain:

$$T_{vir} = 1.98 \times 10^4 (\frac{\mu}{0.6}) (\frac{M_{vir}}{10^8 h^{-1}})^{2/3} (\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2})^{1/3} (\frac{1+z}{10}) K \tag{13}$$

where $\mu = 1.22$ is the mean molecular weight of hydrogen, h=.7 is the dimensionless Hubble parameter, $\Omega_m = .3$ is the current matter energy density, $\Omega_\Lambda = .7$ is the current dark energy density, $\Omega_m^z = \Omega_m (1+z)^3 / [\Omega_m (1+z)^3 + \Omega_\Lambda]$ is the cosmology-dependent evolved matter-energy density, and $\Delta_c = 18\pi^2 + 82d - 39d^2$ where $d = \Omega_m^z - 1$. Using the above expression, we can obtain a set of virial temperatures for the different redshifts: As indicated by the equation and the table, the virial temperature

Table 1: Virial Temperature Per Redshift

| Redshift | log $T_{vir}$ |
|---|---|
| 6.77 | 3.6 |
| 6.9 | 3.64 |
| 7.32 | 3.83 |
| 8.34 | 4.31 |
| 8.76 | 4.5 |
| 11.46 | 5.74 |

is actually a redshift evolving value. For the S1 simulations, the IGM is 50% ionized between redshifts 7

and 8. Because the virial temperature grows linear in redshift as seen above in equation 12, the average virial temperature the ionization history would correspond to about 1e4K.

### 6.1.3 Efficiency $\zeta$

The efficiency parameter does not directly translate to any value given from the radiative transfer simulations, and is therefore valuable to obtain. We can calculate the efficiency given the collapse fraction and the rate of photons emitted per baryon.

$$\zeta = \frac{\dot{n}_b}{\frac{df_{coll}}{dt}} \tag{14}$$

We can use the mean values of the derivative of the collapse fraction taken from the statistics of the 21cmFAST job. I did this by having the derivative of the collapse fraction for each individual voxel saved into an array. Then I calculated the statistics:

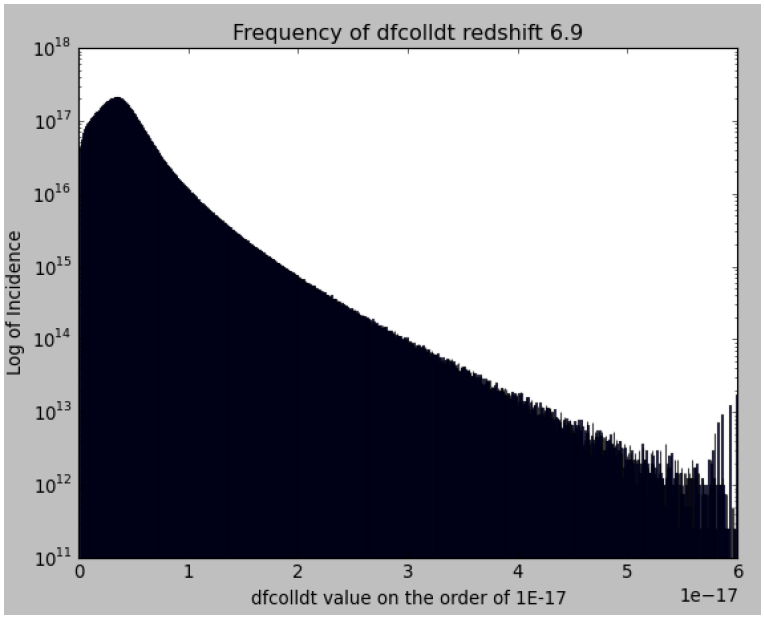



Figure 12: This is a histogram of the collapse fraction for a redshift of 6.9. To obtain the range of values given above, the histograms for redshifts 6.9-8.34 were calculated and the mean values for each histogram were used above

When considering the statistics of the above plots, we obtained a range of mean values for the derivative of the collapse fraction.

$$3.5 * 10^{-18} s^{-1} < \frac{df_{coll}}{dt} < 4 * 10^{-18} s^{-1}$$

Additionally, we needed to calculate the number of photons leaving per baryon per second. This is necessary for the collapse fraction calculation. We obtained this using dimensional analysis given the number of baryons per halo and the photons per halo per second:

$$\dot{n}_b = \dot{N}(m)/N_b \tag{15}$$

Recall that the number of photons per baryon, $N_b$, which is the total number of baryons per halo, could also be calculated using the mass the of halo and the baryonic fraction of the halos:

$$N_b = \frac{f_b * M}{m_p} = 2.38 * 10^{63} \tag{16}$$

using: 1) the baryonic fraction for a typical halo being set to $f_b = .02$ 2) $m_p = 1.67 * 10^{-27} kg$ the Planck mass and 3) $M = 10^8 M_\odot = 1.99 * 10^{38} kg$ the minimum halo mass resolved in the simulation.

Within McQuinn's simulations, the typical luminosity of haloes of mass m follows a linear relationship as $\dot{N}(m) = 2 \times 10^{49} \times m/(10^8 M_\odot) \frac{\text{photons}}{\text{halo}*\text{sec}}$. For the S1 simulations, m in this case will correspond to the minimum virial mass of the stars $10^8 M_\odot = 1.99 \times 10^{38} kg$ so $\dot{N}(10^8) = 2 \times 10^{49} \frac{\text{photons}}{\text{halo}*\text{sec}}$.

$$\dot{n}_b = \dot{N}(m)/N_b \tag{17}$$

$$= \frac{2 \times 10^{49}}{2.378 \times 10^{63}}$$
$$= 8.41 \times 10^{-15} \frac{\text{photons}}{\text{baryon} * \text{sec}}$$

We can then take what we have done up to now and plug these values into the expression for the efficiency (equation 12). Recall that the efficiencies that 21CMMC converged on were on the order of $\approx 40 - 50$. We obtained the derivative of the collapse fraction from the 21cmFAST code itself. Then we calculated the number of photons per baryon per second, given the number of baryons per halo and the number of photons per halo per second. The efficiency is the resulting quotient between the former and the latter.

$$\zeta = \frac{\dot{n}_b}{\frac{df_{coll}}{dt}} \approx \frac{\dot{N}(m)}{N_b \times \frac{df_{coll}}{dt}} \tag{18}$$

$$\propto \frac{\text{Photons}}{\text{Halo} * \text{sec}} \times \frac{\text{Halo}}{\text{Baryon}} \times \frac{1}{\frac{1}{\text{sec}}} \tag{19}$$

$\zeta \approx 2423 - 2036$ This is clearly a few orders of magnitude larger than the values 21CMMC obtained. Up to this point 21CMMC has been unable to produce a single mean free path. The virial temperature in most runs have been converging around 1e4K, though this is a lower limit potentially due to the fact that the 1e4K value is hard coded into 21cmFAST. Lastly, the efficiency calculated is several orders of magnitude off of what 21CMMC produces. Thus, 21CMMC has been unable to confidently reproduce any of the three ionization parameters.

# 7 Concluding Statements and Future Work

Above we have discussed how the 21CMMC code aims to use the 21cmFAST simulation of Reionization to determine the parameters of the early universe. This entails a Monte Carlo algorithm that steps through parameter space of the three key parameters in the 21cmFAST code: the mean free path of photons propagating in the IGM, the minimum virial temperature of the halos, and the efficiency

of photons leaving those halos. Up to now, we have used 21CMMC on a variety of power spectra including those from a full S1 radiative transfer simulation and 21cmFAST2 (utilizing both the virial temperature and the virial mass parameters). From the radiative transfer simulations, we have calculated that the accuracy in 21CMMC's ability to recreate the efficiency parameter is off by an order of magnitude.

The first set of power spectra came from Matt McQuinn's N-body S1 simulation (a cosmological simulation called Gadget2) and full radiative transfer algorithm (HealPIX). Such a simulation is significantly more detailed in the amount of physics that it is able to include, but also much more computationally expensive. Even this simulation was drastically simplified compared to other radiative transfer codes. Its algorithm runs the full N-body code using Gadget2, and then inputs the particle locations from Gadget2. Afterwards, the code incorporates a list of ionizing sources, which casts individual rays of photons from the sources. This differs from 21cmFAST in that the code tracks the individual path of each photon rather than giving the photons a maximum distance that they can propagate from each source before disappearing. Thus, the 21cmFAST approach is less physically involved. 21cmFAST also determines the number density of sources by computing the conditional collapse fraction $f_{coll}$ from the dark matter field, thus bypassing the more involved halo finder.

Given all of this, the power spectra from McQuinn's simulations are expected to contain much more physics, which our tests with 21CMMC struggled to capture. We found that 21CMMC had the most difficulty with converging on the mean free path parameter, which makes sense due to the fact that it is the least physical parameter in 21cmFAST. The $R_{mfp}$ is an ad-hoc parameter that produces a "knee" in the power spectrum on the characteristic scales of the actual mean free path. It is only a maximum bubble size for 21cmFAST to produce this "knee", so there is no surprise that the mean free path is the most troublesome of the three parameters. However, S1 power spectra were not smooth, but rather had several sharp "knee"-like structures (not the traditional knee but certainly sharp peaks at several scales, which is indicative of the array of bubble sizes being allowed). The presence of these sharp knees could mean that 21CMMC was trying to match power spectra from 21cmFAST to all of those knees for the mean free path, which would explain the multi-model results that we occasionally obtained. This is because the photons in the radiative transfer simulation will propagate to some arbitrary distance until they hit a hydrogen, which could mean there are many characteristic bubble sizes.

This discontinuous feature became more prominent in the case of a redshift and scale-dependent set of error bars. This gave stronger preference to larger scale power and lower redshift. At these scales and redshifts, the peaks in the radiative transfer code are being included more as the smoother parts are almost negligible because of large error values. This is consistent with the above argument, but discouraging because it means we cannot truly obtain a single value nor do we expect to.

Additionally, there is a hard-coded discontinuity for the virial temperature in 21cmFAST because at 1e4K, the physics of the emitting galaxies changes. More specifically, the mean molecular mass for the gas changes in an ionized vs neutral IGM (where an ionized IGM has a $T_{vir}$ of $10^4$ K). Given the fact that the approximate answer in McQuinn's simulation is near 1e4K, it is difficult to determine whether 21CMMC is truly converging on the right value or if it is running up against a computational wall. Additionally, the redshift dependence on this parameter would also influence the value we obtain. Lower

redshifts give lower virial temperatures near $10^{3.6} - 10^{3.8}$ near the end of Reionization. When using the redshift and scale dependent error files, the lower redshifts would thus favor lower virial temperatures, which we would not be able to probe properly given the hard-coded discontinuity in 21cmFAST for the change in mean molecular mass. Thus, we tried to switch 21CMMC to run with the virial mass as a parameter rather than the virial temperature. We knew from McQuinn's paper [23] that there was a minimum virial mass of 1e8, which does not face the same computational discontinuity as the virial temperature. We attempted to initially try to get 21CMMC to run with the virial mass turned on. First we had to show that it could recreate itself with the virial mass turned on, which it did with limited success.

Further, given that it is a radiative transfer code, there is no true mean free path. McQuinn's paper also provided the number of photons leaving each halo per second and the number of photons per baryon [23]. We used these values, coupled with the collapse fraction calculations pulled from 21cmFAST itself, to calculate the efficiency that the simulations should give. The value we calculated was on the order of ≈2000, whereas the result that 21CMMC gives is on the order of ≈40. This is likely due to the degeneracies between all of these parameters, so getting one wrong (i.e. mean free path) means getting more than one wrong. Thus, attempting to fit such 21cm power spectrum with the 3-parameter model will therefore never work very well. The correct parameter values were outside of the error bars in our plots, so we decided it was not worth continuing down this particular path.

After having worked with the power spectra from the full radiative transfer simulations, I tried running 21CMMC on something a bit less ambitious: 21cmFAST2. This is an updated version of 21cmFAST, but it does not include the mean free path and it does incorporate inhomogeneous recombinations. We saw that despite the fact that 21cmFAST2 is really not that different from 21cmFAST, 21CMMC was still not able to determine the correct parameter values for the mean free path, efficiency, and minimum virial temperature.

We saw this discrepancy by using the 'faint galaxies' model from the 21cmFAST2, which is an extreme case that allows inhomogeneous recombinations rather than the mean free path to have maximum impact because (i) Reionization is very slow; and (ii) the dominant galaxies are very susceptible to photo-heating feedback that couples with the recombinations to dramatically suppress large-scale power (Sobacchi & Mesinger 2014). $R_{mfp}$ should be a more effective parameter at higher values of $T_{vir}$ ('bright galaxies' 21cmFAST2 model) because recombinations don't play as much of a role in those models. However, because it is not truly a physical parameter in 21cmFAST2, we used the "faint galaxies" model. Using the "faint galaxies" model was a good test not only to take a step back given the troubles we faced with the full radiative transfer simulations, but the model also works in anticipation for the newest version of 21CMMC, incorporating recombinations. Thus, this model also allows us to get closer to accurately capturing the physics of Reionization.

In 21cmFAST2, there exists correct values for the efficiency ($\zeta = 20$) and the minimum virial temperature $T_{virmin} = (log(2E4))$, but we were not able to achieve them. Rather, we obtained values $\approx 10$ for the efficiency. $T_{vir-min}$ doesn't have a single correct value from the "faint galaxies" runs. Rather, it is a function of redshift. At high redshifts, the virial temperature is 2e4 K and then continues to rise

further at low z. Given the fact that our power spectra do not incorporate very low redshifts, we decided that a good approximation for the virial temperature was 2e4K. We see that when we run 21CMMC on the 21cmFAST2 power spectra, it converges at about 1e4 K (4.0 on the log axis).

As discussed above, the $R_{mfp}$ is an ad-hoc parameter that produces a "knee" in the power spectrum on the characteristic scales of the mean free path, which is not present in the 21cmFAST2 power spectra. 21CMMC with the simple 3-parameter model will therefore fit the "knee" to the 21cmFAST2 power spectrum at large-scales, but it will then drastically differ at intermediate to small scales. With 21cmFAST2, the sub-grid physics absorbs the mean free path in a redshift evolving parameter, and in 21CMMC the mean free path is not a redshift evolving parameter. Averaging that mean free path over redshift will bias it for the other parameters along the way. Due to the fact that 21CMMC is trying to match power spectra that are subject to change because of a single parameter, getting one area of physics wrong (i.e. the mean free path) will bias other physics to be wrong (virial temperature or efficiency).

Moving forward, we hope to incorporate the newest version of 21CMMC, which utilizes inhomogeneous recombinations, to see whether a newer version of 21CMMC works. As of now, the older version is unable to capture the physics of power spectra properly. Given how advanced radiative transfer simulations have become, it is likely using a more advanced version (and thus "better" power spectra) would not greatly improve the results of 21CMMC. Thus, it is probably best to turn to developing a more elaborate 21CMMC that incorporates more physics into the underlying simulation (i.e. 21cmFAST). The most up-to-date version includes inhomogeneous recombinations and no longer attempts to determine the mean free path parameter.

It is still an acute challenge of EoR cosmology to get meaningful constraints from 21cm power measurements. Instruments need to provide accurate power spectra and combining this all into one will make this even harder. There is hope for potential cosmological 21cm signals given the recent EDGES results in March 2018 [5]. By studying the resulting impact on the EoR astrophysical constraints, 21CMMC can be used to optimize: (i) interferometer designs; (ii) foreground cleaning algorithms; (iii) observing strategies; (iv) alternate statistics characterizing the 21cm signal; and (v) synergies with other observational programs.

# Appendices

## A Semi-Analytic Formalism

21cmFAST and 21CMMC are not a full hydro-dynamical code nor radiative transfer, but rather called semi-numerical simulations. Semi-numerical simulations formally use approximate physics to still model the 3-dimensional universe, while providing faster computation[14]. First of all, nearly all cosmological simulation apply what is called *subgrid modeling* in which approximations on scales smaller than the resolved voxels in a simulation are applied. Other codes that simulate these smaller scales can produce statistical models of particular processes. This may include a statistical volume average of stellar,

galactic, or IGM properties to be included per voxel. There are a plethora of subgrid models ranging from star formation to black hole growth and supernovae feedback. Such formalisms used in the code will be described below. In particular for 21cmFAST uses perturbation theory, Press Schecter model of dark matter halos, the excursion set formalism for collapse to obtain the non-linear density and ionization fields, and efficiency and virial temperature of galaxies.

In addition to subgrid modeling, other approximations that quicken the computation of the code are also applied. In our case this includes using perturbation theory and the Excursion Set formalism which allows the code to bypass halo-finding algorithms. The key here is that the code is acting directly on the dark matter distribution rather than evolving baryons hydrodynamically in the boxes themselves. This quickens computation time, while simultaneously allowing for the dark matter haloes to then be populated with galaxies for the production of the ionization field.

## A.1 Building The Density Field

Within a box of the simulated universe, each individual particle is given a set of initial positions and linear velocities. We can write two differential equations of motions that reflect the chosen cosmology of any given box [19]:

$$\frac{dx_i}{dt} = u_i \tag{20}$$

$$\frac{du_i}{dt} + 2H(t)u_i = -a^{-2}\nabla\phi \tag{21}$$

where the gravitational potential is determined by the Poisson equation $\nabla\phi = 4\pi G\bar{\rho}a^2\delta$. These equations allow for the simulation to tell the particles that they experience gravitational attraction.

## A.2 Using Perturbation Theory

Within a box of the simulated universe, each individual particle is given a set of initial positions and linear velocities. After writing the equations of motion from Poisson's equation for the gravitational potential, we can consider the initial conditions $[x_i(t_o), u_i(t_o)]$ to then numerically integrate the equations for future configurations. In order to create this linear density field, we rely on the standard **Zel-Dovich approximation**. This entails using a Lagrangian approach to cosmological fluid dynamics that tracks the path of individual particles through time (rather than the current as used in Eulerian code). In practice, this involves writing the new position as a function of a prior position x plus some new deformation $\psi(x,t) \approx \psi^1(x,t) + \psi^2(x,t) + \ldots$ where $\psi(x,t_o) = 0$. For the Zeldovich approximation we keep first order terms. This displacement field fully describes our cosmological fluid.

$$x_1 = x + \psi(x,t) \tag{22}$$

$$v(x,t) \equiv \dot{x_1} = \dot{\psi}(x) \tag{23}$$

Units are comoving. Provided a continuity condition for our fluid of particles, we can obtain an expression to relate the density perturbation to the current flowing in or out of a voxel:

$$\delta(x) = -\nabla \cdot [(1+\delta(x))\psi(x)]$$

$$\approx -\nabla \cdot \psi(x)]$$

Given that the perturbation $\delta(x)$ can be expressed in k space via a Fourier transform:

$$\delta(k) = \frac{V}{N} \sum \delta(x) \exp^{-k \cdot x}$$

Eventually, moving forward in time, particles will collapse under the force of gravity, which allows us to obtain a linear density field that contains dark matter halos. The details of how these equations are applied in the code itself are elucidated in the Appendix.

## A.3 Halo Collapse

As matter perturbations continue to evolve, they will go from linearly to nonlinearly evolving in the scale factor as structures themselves begin to form. This means that over densities are growing much more quickly in time such that $\delta > 1$. Assuming a cold dark matter universe, which is not unreasonable for the matter dominated era, we can solely rely on gravitational collapse (no outward pressure) for non-linear evolution. From here, we can use the Friedmann equations, which describe the expansion/contraction of the homogeneous, isotropic universe as governed by the composition (various energy densities) of that universe. Analytically, the spherical collapse model treats each halo as a "mini universe" where a Friedmann-like equation can be applied to include matter perturbations to then describe how dark matter halos collapse. This equation will tell us the cosmology of a dark matter halo collapsing given a perturbation $\delta$ in the matter density that evolves linearly with the scale factor a. When considering a particle at radius r from the center of the mass, we can ignore any gravitational influence external to the spherical matter and essentially assume a mini Einstein DeSitter (flat-matter-dominated) universe [19].

## A.4 Halo Formation and Identification

At a critical over density $\delta_c$ (which is calculated in the Appendices) we assume the halo is virialized in which a dynamically relaxed self-gravitating object collapses into a stable state. Once perturbations begin to collapse into dark matter halos, we need to determine two pieces of information to identify the collapsed dark matter halos:

- Mass Function - how many halos as a function of mass are there?
- Clustering - Is there a position correlation between the dark matter halos?

To answer these questions, we enter the quasi-linear regime that is known as the **Excursion Set Formalism**. Another way of describing this is a random walk/diffusion process of picking random points in space of a given $\delta_m$. Here we use a smoothing kernel to smear out $\delta_m$ for a given scale R (our smoothing scale). Though it is more common to talk about a mass scale $M = \frac{4\pi}{3} \bar{\rho}_m R^3$. We use a Fourier space Top hat to pick different cut off modes $k_c$ (which correspond to different smoothing scales in coordinate space). For a given cut off mode, we reject any modes larger than $k_c$ (i.e. smaller physical scales are washed out). The way we implement this cut off is with window function $W_k$, which in its

most basic form looks like a backward step function. We begin with a $k_c$ value and then make it larger by adding in more and more Fourier modes.

If we recall from our definition of the variance, $\sigma^2(k) = \int \frac{dk}{k}\Delta^2(k)W_k$, we see that as we add in more modes, we will add more contributions to the variance. As the variance grows, the perturbations themselves are growing. Thus, we want to know when it hits the proper scale that corresponds to $\delta_c$ itself. This can be plotted on a graph of the perturbation as a function of the variance itself. As the variance increases, there is a random walk in an upward trend until $\delta_m$ reaches the critical over density. This is like a point of no return because any regions that surpass the critical over density and cross the boundary will collapse. This is ultimately to find a sink over density and then putting the result into the spherical top hat. For a given k mode of collapse, we have a corresponding scale that is the virial radius of the collapsing halo. Given the virial radius, we can then calculate a virial mass $M = \frac{4}{3}\pi r^3 \rho$.
Now given a variance that is mass dependent and the initial assumption that the random density field is Gaussian, we can write an expression for the probability of a halo of mass m>M collapsing.

$$P = \frac{1}{2\pi\sigma^2}\int_{\delta_c(z)}^{\infty} d\delta e^{\frac{-\delta^2}{2\sigma^2(m)}}$$

If we take the derivative of this probability with respect to the mass, then we can come up with the **Press Schecter Mass function** [44], which describes the number of collapsed halos for a given mass. In other words, we have a population count of halos with different masses. This is incorporated into the code to determine the distribution of the ionized bubbles in addition to the dark matter distribution [19]..

# References


[1] T. Abel, G. L. Bryan, and M. L. Norman. The Formation and Fragmentation of Primordial Molecular Clouds, September 2000.

[2] K. Ahn, I. T. Iliev, P. R. Shapiro, and C. Srisawat. Non-linear bias of cosmological halo formation in the early universe, June 2015.

[3] J. Arons and D. W. Wingert. Theoretical Models of Photoionized Intergalactic Hydrogen, October 1972.

[4] S. Baek, B. Semelin, P. Di Matteo, Y. Revaz, and F. Combes. Reionization by UV or X-ray sources, November 2010.

[5] R. Barkana, N. J. Outmezguine, D. Redigolo, and T. Volansky. Signs of Dark Matter at 21-cm?, March 2018.

[6] Rennan Barkana and Abraham Loeb. In the beginning: the first sources of light and the reionization of the universe, 2001.

[7] V. Bromm, A. Ferrara, P. S. Coppi, and R. B. Larson. The fragmentation of pre-enriched primordial objects, December 2001.

[8] C. L. Carilli and S. Rawlings. Motivation, key science projects, standards and assumptions, December 2004.

[9] David R. DeBoer, Aaron R. Parsons, James E. Aguirre, Paul Alexander, Zaki S. Ali, Adam P. Beardsley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Chris L. Carilli, Carina Cheng, Eloy de Lera Acedo, Joshua S. Dillon, Aaron Ewall-Wice, Gcobisa Fadana, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Bradley Greig, Jasper Grobbelaar, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, MacCalvin Kariseb, Saul A. Kohn, Telalo Lekalake, Adrian Liu, Anita Loots, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary Martinot, Nathan Mathison, Eunice Matsetela, Andrei Mesinger, Miguel F. Morales, Abraham R. Neben, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Raddwine Sell, Craig Smith, Angelo Syce, Max Tegmark, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Hydrogen epoch of reionization array (hera), 2017.

[10] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer, March 2013.

[11] S. R. Furlanetto, L. Hernquist, and M. Zaldarriaga. Constraining the topology of reionization through Ly$\alpha$ absorption, November 2004.

[12] S. R. Furlanetto, A. Lidz, A. Loeb, M. McQuinn, J. R. Pritchard, M. A. Alvarez, D. C. Backer, J. D. Bowman, J. O. Burns, C. L. Carilli, R. Cen, A. Cooray, N. Gnedin, L. J. Greenhill, Z. Haiman, J. N. Hewitt, C. M. Hirata, J. Lazio, A. Mesinger, P. Madau, M. F. Morales, S. P. Oh, J. B. Peterson, Y. M. Pihlström, P. R. Shapiro, M. Tegmark, H. Trac, O. Zahn, and M. Zaldarriaga. Astrophysics from the Highly-Redshifted 21 cm Line, 2009.

[13] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance, 2010.

[14] B. Greig and A. Mesinger. 21CMMC: an MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21 cm signal, June 2015.

[15] B. Greig and A. Mesinger. Simultaneously constraining the astrophysics of reionization and the epoch of heating with 21CMMC, December 2017.

[16] W. J. Handley, M. P. Hobson, and A. N. Lasenby. POLYCHORD: nested sampling for cosmology, June 2015.

[17] Fengji Hou, Jonathan Goodman, David W. Hogg, Jonathan Weare, and Christian Schwab. An affine-invariant sampler for exoplanet fitting and discovery in radial velocity data, 2012.

[18] K. Kakiichi, S. Majumdar, G. Mellema, B. Ciardi, K. L. Dixon, I. T. Iliev, V. Jelić, L. V. E. Koopmans, S. Zaroubi, and P. Busch. Recovering the H II region size statistics from 21-cm tomography, October 2017.

[19] Abraham Loeb and Steven R. Furlanetto. The first galaxies in the universe, 2013.

[20] C. J. Lonsdale, R. J. Cappallo, M. F. Morales, F. H. Briggs, L. Benkevitch, J. D. Bowman, J. D. Bunton, S. Burns, B. E. Corey, L. Desouza, S. S. Doeleman, M. Derome, A. Deshpande, M. R. Gopala, L. J. Greenhill, D. E. Herne, J. N. Hewitt, P. A. Kamini, J. C. Kasper, B. B. Kincaid, J. Kocz, E. Kowald, E. Kratzenberg, D. Kumar, M. J. Lynch, S. Madhavi, M. Matejek, D. A. Mitchell, E. Morgan, D. Oberoi, S. Ord, J. Pathikulangara, T. Prabu, A. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. U. Shankar, K. S. Srivani, J. Stevens, S. Tingay, A. Vaccarella, M. Waterson, R. B. Wayth, R. L. Webster, A. R. Whitney, A. Williams, and C. Williams. The Murchison Widefield Array: Design Overview, August 2009.

[21] A. Maselli, A. Ferrara, and B. Ciardi. CRASH: a radiative transfer scheme, October 2003.

[22] Takuya Matsuda, Humitaka Satō, and Hidenori Takeda. Cooling of pre-galactic gas clouds by hydrogen molecule, 1969.

[23] Matthew McQuinn, Adam Lidz, Oliver Zahn, Suvendra Dutta, Lars Hernquist, and Matias Zaldarriaga. The morphology of h ii regions during reionization, 2007.

[24] A. Mesinger. How the first generations of luminous baryons established the X-ray and UV backgrounds, January 2011.

[25] A. Mesinger. Reionization and Cosmic Dawn: theory and simulations, January 2018.

[26] A. Mesinger, A. Ewall-Wice, and J. Hewitt. Reionization and beyond: detecting the peaks of the cosmological 21 cm signal, April 2014.

[27] A. Mesinger and S. Furlanetto. Efficient Simulations of Early Structure Formation and Reionization, November 2007.

[28] A. Mesinger, S. Furlanetto, and R. Cen. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal, February 2011.

[29] S. More, A. V. Kravtsov, N. Dalal, and S. Gottlöber. The Overdensity and Masses of the Friends-of-friends Halos and Universality of Halo Mass Function, July 2011.

[30] Fumitaka Nakamura and Masayuki Umemura. On the mass of population iii stars, 1999.

[31] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering, December 1997.

[32] P. Ocvirk, N. Gillet, P. R. Shapiro, D. Aubert, I. T. Iliev, R. Teyssier, G. Yepes, J.-H. Choi, D. Sullivan, A. Knebe, S. Gottlöber, A. D'Aloisio, H. Park, Y. Hoffman, and T. Stranex. Cosmic Dawn (CoDa): the First Radiation-Hydrodynamics Simulation of Reionization and Galaxy Formation in the Local Universe, December 2016.

[33] P. J. E. Peebles. Recombination of the Primeval Plasma, July 1968.

[34] J. C. Pober, A. Liu, J. S. Dillon, J. E. Aguirre, J. D. Bowman, R. F. Bradley, C. L. Carilli, D. R. DeBoer, J. N. Hewitt, D. C. Jacobs, M. McQuinn, M. F. Morales, A. R. Parsons, M. Tegmark, and D. J. Werthimer. What Next-generation 21 cm Power Spectrum Measurements can Teach us About the Epoch of Reionization, February 2014.

[35] W.C. Saslaw and D. Zipoy. Molecular hydrogen in pre-galactic gas clouds.

[36] E. Sobacchi and A. Mesinger. Inhomogeneous recombinations during cosmic reionization, May 2014.

[37] A. Sokasian, T. Abel, and L. Hernquist. The epoch of helium reionization, May 2002.

[38] Aaron Sokasian, Naoki Yoshida, Tom Abel, Lars Hernquist, and Volker Springel. Cosmic reionization by stellar sources: population iii stars, 2004.

[39] V. Springel. The cosmological simulation code GADGET-2, December 2005.

[40] Hy Trac and Renyue Cen. Radiative transfer simulations of cosmic reionization. i. methodology and initial results, 2007.

[41] B. A. Whitney. Monte Carlo radiative transfer, March 2011.

[42] Hao Xu, John H. Wise, Michael L. Norman, Kyungjin Ahn, and Brian W. O'Shea. Galaxy properties and uv escape fractions during theepoch of reionization: Results from the renaissance simulations, 2016.

[43] S. Zaroubi. The Epoch of Reionization. In T. Wiklind, B. Mobasher, and V. Bromm, editors, *The First Galaxies*, volume 396 of *Astrophysics and Space Science Library*, page 45, 2013.

[44] A. R. Zentner. The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth, 2007.