BRIEF REPORT

Setting a comparison between two reionization models through the understanding of a statistical bias

H. Baraer[a,1]
Supervised by : Prof. A. Liu[a,2]

[a]McGill University 845 Sherbrooke Street W Montréal, Quebec H3A 0G4

This manuscript was compiled on December 11, 2022

The study of the Epoch of Reionization (EoR) (z $\sim$ 20 to z$\sim$ 6) provides capital insights into the development of galaxy structure in our Universe and helps us to understand the process behind the galaxies' structures and their evolution. Modelling this reionization history of neutral hydrogen composing the intergalactic medium can be done through the redshift of reionization field. Unfortunately, the redshift of reionization field models does not concur, especially when modelling large volumes ($\gtrsim Gpc^3$). This study compares the redshift of reionization fields of the computationally fast *z-reion* to the widely used *21cmFAST*. The presented results compare the fields through their ionization history and their ionization maps at the mean redshift of reionization. The z-reion statistical term variational range when varying 21cmFAST inputs was also computed. Results show discrepancies in the fine structures (high frequencies) of the ionization maps and a disagreement in the width of reionization. The variational range indicates that optimizing the models' inputs to minimize the differences in the models' outputs is possible and is proposed as a future project.

Epoch of Reionization | Model Comparison | Simulations | Variational Range | Bayesian Inference

Introduction

About 380 000 years after the Big Bang, the early Universe's expanding "soup" of fundamental particles cooled enough for the protons and electrons to combine, forming neutral hydrogen in all of the Universe (1). In this process, the photons decoupled from matter, and the frozen imprints of the early Universe's structure radiated through the Universe, visible as blackbody radiation at a redshift $z \sim 1100$. This radiation is known as the Cosmic Microwave Background (CMB). Following the CMB, observations were limited as the absorbing effect of the neutral hydrogen composing the Universe made it opaque to most of the electromagnetic spectrum.

Over the next 250 million years, regions of the Universe with higher gas density slowly collapsed by gravity eventually, at a redshift $z \sim 20$, forming the first generation stars and galaxies (2). The photons released by the proto-stars of this cosmic dawn ionized the neutral hydrogen of their surrounding, carving out "bubbles" of ionization in the intergalactic medium (IGM) around them. These ionization bubbles characterized the Epoch of Reionization (EoR) that lasted until $z \sim 6$. Following that reionization process, the Universe became transparent and was observable. Much more complex structures from what was observed in the early Universe by the CMB were present. The EoR, therefore, provides capital insights into the development of galaxy structure in our Universe and helps to understand the process behind the galaxies' structures and their evolution (3). The EoR history can be studied through the evolution of the "bubbles" of reionization. These bubbles occurred in an inhomogeneous process, forming at different redshifts. Studying this reionization history helps us understand this structural evolution structure evolution of our Universe, and can be modelled through the redshift of reionization field, a three-dimensional field of our Universe showing the redshift at which each of these bubbles reionized.

Unfortunately, models of the redshift of reionization field do not concur. There are discrepancies between known models, especially when modelling large volumes of reionization ($\gtrsim Gpc^3$) (4). Since models use different methods to compute, input parameters change from code to code, making model comparisons on the same basis difficult. I.e., it is difficult to compare models with the same given physical and cosmological parameters.

A promising model called *z-reion* offers a computationally light simulation (5).To model the redshift of reionization field, it uses a density field: a three-dimensional field representing the inhomogeneous density regions in our Universe. A statistical term linked the two fields on a short scale ($\lesssim Mpc^3$). A density field was taken from particle-particle-particle-mesh (P3M) simulation of 2048^3 dark-matter particles, 2048 gas cells and 17 billion adaptive rays. The redshift of reionization field was taken from a model combining a radiative transfer algorithm and a cosmological hydrodynamic code (6). The fields were converted into fluctuation fields. Computed with the average of their fields, they show the fractional differences from the mean rather than the values themselves. Therefore, the values of voxels in the fields are unitless, making a comparison possible. The linear bias factor relates the fluctuations fields in momentum space such that the following equation is satisfied:

$$\tilde{\delta}_{\mathrm{z}}(k) = b_{mz}(k)\tilde{\delta}_{\mathrm{m}}(k), \qquad [1]$$

where $\tilde{\delta}_{\mathrm{z}}(k)$ and $\tilde{\delta}_{\mathrm{z}}(k)$ are the over-density field and the over-

The following people are involved

[1]The main author of this report is Hugo Baraer. E-mail: hugo.baraer@mail.Mcgill.ca

[2]This project is done under the supervision of Prof. Adrian Liu. E-mail: Adrian.liu2@mcgill.ca

redshift of reionization field respectively. This factor can be fitted by a thee-parameter function. Its simplest functional form can be written as:

$$b_{\rm mz}(k) = \frac{b_{\rm o}}{(1 + k/k_{\rm o})^{\alpha}}, \qquad [2]$$

where α, b_0 and k_0 are free parameters fitted for.

On the other hand, a much more used model: 21cmFAST (7), computes the redshift of reionization field independently from the density field. The model computes the ionization status of all the voxels at each redshift by identifying if the voxel's number of ionizing photons is higher than the number of hydrogen atoms it contains. I.e. A voxel is considered ionized at a given redshift if it meets the following criteria:

$$f_{\rm coll}(\mathbf{x}, z, R) \geq \zeta^{-1}, \qquad [3]$$

where ζ is the ionization efficiency parameter, and $f_{\rm coll}(\mathbf{x}, z, R)$ is the collapse fraction. A direct comparison f the outputs of *z-reion* and *21cmFAST* has never been done before, and would allow a better accuracy in the large scale modeling of the redshift of reionization field, as well as an easier usage of the computationally quicker *z-reion*.

Following that need, last semester's project was to construct an algorithm computing the linear bias factor (equation 1) from *21cmFAST*'s data. Using the density field and redshift of reionization field from *21cmFAST*, it follows the steps above described taken by *z-reion* to compute its linear bias. Once computed, the free parameters of the linear bias (equation 2) were fitted using Bayesian inference and Markov Chain Monte Carlo (MCMC). MCMC is a class of methods for sampling a probability density function using a Markov chain whose equilibrium distribution is the desired distribution (8). In other words, the MCMC evaluates the probability distribution for the value of a best-fit parameter. Posterior distributions, best-fit values and confidence intervals of the linear bias' free parameters were computed for *21cmFAST* as shown in Figure 1. The p parameter represents the standard error, a new feature added, further explained in section 2.

A new opening was created for model comparison with these best-fitted value parameters. Values for the linear bias parameters of 21cmFAST can be plugged in *z-reion*, and a comparison between the fields is possible. Moreover, the algorithm can be used to study the variational range of *21cmFAST* inputs (affecting equation 3) on the best-fit values.

This semester's project aims to use last semester's algorithm to set up a comparison between *21cmFAST* and *z-reion*. This report presents the work done during the fall semester in PHYS 479: Physics Research Project.

Methods

A series of steps is presented used to set up a comparison between 21cmFAST and z-reion using last semester's algorithm. It is achieved in two parts:

1. A code is created to compare the models' redshift of reionization field through ionization fractions and ionization maps.

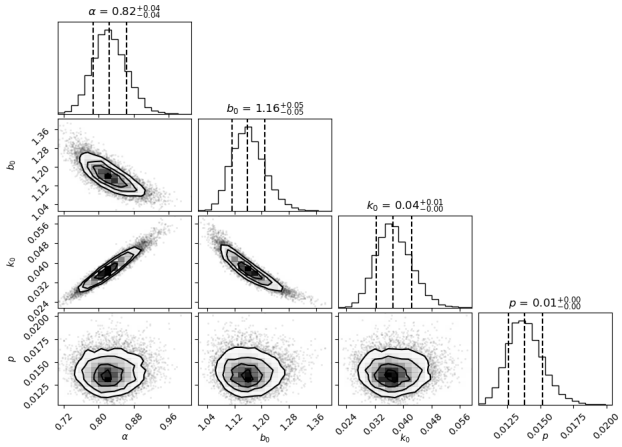


Fig. 1. Corner plots of the posterior distributions for the best-fit parameters of equation 2: α, β_0 and k_0, and the standard error p. In all four plots, the middle dotted line represents the distribution's median value, attributed to the best fit value. The first and last dotted lines are the 16% and 84% percentiles, showing the middle 68% of the posterior distribution region. The best fit value and the confidence interval are written above each posterior distribution. A grid of six 2-D histograms comparing each pair of variables is also presented. All the plots are normalized.

2. The parameters' variational range are looked at by computing the effects of 21cmFAST inputs on the linear bias parameters.

By doing this, models will be compared on a similar basis, and all the code constructed this term will be able to be used to find the corrected input parameters of *21cmFAST* to use to get similar outputs than *z-reion*.

In Cosmology, some parameters do not have a consensus value and can vary in the literature. These parameters includes the Hubble parameter $h = H_0/100 kms^{-1}Mpc^{-1}$, the density proportion of matter Ω_m, and the density proportion of dark energy Ω_Λ. In this work, these values are set in both models to: $\Omega_{\rm m} = 0.27, \Omega_\Lambda = 0.73, h = 0.7$

Two new modules were added to last semester's Python-based algorithm, each serving one step. *Plot_params.py* deals with the comparison between ionization histories and ionization maps, further detailed. *Variational_params.py* launches simulations with different inputs parameters. Hereby is presented a series of substeps leading to the completion of the main objectives.

Comparing redshift of reionization fields

To start this project, the found linear bias' free parameters from *21cmFAST* : α, b_0 and k_0 were plugged into the z-reion code. The values of the parameter plugged are presented in Figure 1. The density field computed by 21cmFAST was also used in *z-reion*, since as shown by equation 1, *z-reion* uses the density field and the linear bias term to generate the redshift of reionization field. Since the model is not publicly available, the parameters and the density field were given to another member of the *Cosmic Dawn Group* at *McGill University*, Ph.D. candidate Elizabeth McBride, to compute the redshift of reionization field. Two observables were then used to compare the redshift of reionization field.

Ionization histories represent a great way of interpreting the redshift of reionization field, as they can compare the evolution of reionization (5). Ionization histories are plots showing the evolution of the ionization fraction as a function of the redshift, where the ionization fraction is the number of ionized hydrogen voxels over the total number of voxels in the redshift of reionization field. To compute this, a function in *Plot_params.py* iterate through the redshift of reionization field and computes the ionization fraction for each redshift. Ionization maps were then visually compared, and their intersection, estimated to be at the mean redshift of reionization, was observed.

Ionization maps are the equivalent of the redshift of reionization field at a given redshift: indicating if a voxel was ionized at a given redshift or not. Comparing ionization maps is the best way to compare the structures present in the redshift of reionization field. A function in *Plot_params.py* computes the ionization maps at a given redshift, and the maps were compared visually. In addition, the map difference (ionization map of 21cmFAST - ionization map of z-reion) was computed and visually analyzed. Finally, the power spectrum of the difference between the maps was taken. Power spectrums are the equivalent of correlation functions in momentum space, designed to quantify the correlation of the differences in two maps. I.e., this evaluates how a voxel is related to its neighbour. It quantifies if the map differences are randomly dispersed in the field or if they are grouped and is defined as :

$$\left\langle \left| \widetilde{\chi_{21cmFast} - \chi_{z-reion}}(\mathbf{k}) \right|^2 \right\rangle \propto P_{diff}(\mathbf{k}), \qquad [4]$$

where $\widetilde{\chi_{21cmFast} - \chi_{z-reion}}$ is the Fourier transform of the difference between the ionization maps (denoted χ), and the brackets $\langle\rangle$ denotes an ensemble average. To compute the ensemble average, the three-dimensional ionization maps were converted into a one-dimensional array where the independent variable is the frequency k: the inverse of the voxel's distance from the center of the field. To do so, the fields are converted in spherical coordinates so that $\delta(k_x, k_y, k_z) \rightarrow \delta(k, \theta, \phi)$ where $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$. The fields are then averaged over phi and theta. I.e., the field is separated into spherical shells varying by a defined radius, and each of the shells is averaged to one value. However, since equation 4 is a proportionality relation, the percent difference of their power spectrum was taken to look at how the map differences are correlated compared to their own correlation. It is more of an absolute measure and can be expressed as :

$$\frac{\left\langle \left| \widetilde{\chi_{21cmFast} - \chi_{z-reion}}(\mathbf{k}) \right|^2 \right\rangle}{\left\langle \left| \widetilde{\chi_{21cmFast}}(\mathbf{k}) \right|^2 \right\rangle} = P_{abs\ diff}(k), \qquad [5]$$

This absolute power spectrum was then compared with the observations made to establish a comparison.

21cmFAST's inputs variability

The goal of this part was to observe the variability of the free parameters over the range of possible *21cmFAST* inputs. Results of the 21cmFAST's inputs variability were then used in the comparison analysis. The two primary inputs affecting 21cmFAST's redshift of reionization field (see equation 3) were studied: the ionization efficiency and the turnover mass. The ionization efficiency can range from 0 to 100, and the variability of the free parameters was tested over the whole range. Physically, it represents the ionizing efficiency of high-z galaxies, and higher values tend to speed up reionization. The turnover mass is the mass required for the quenching of star formation in halos, and a reasonable physical range of usage has been defined from 7.5 to 9 (in $log_{10}(M_\odot)$ units).

Last semester's algorithm was slightly modified and automated to allow multiples simulations with varying inputs to compute this variational range. At each simulation, the values of the mean and the values of the confidence interval (16th and 64th percentile) of posterior distributions were stored for all the free parameters.

The error estimate was revised to allow for a bigger range and a more accurate representation of the systematic error. Following a procedure done in *z-reion*, the errors were weighted by the cross-correlation term between the density field and the redshift of reionization field, and the function is of the form:

$$f_{weight}(k) = -e^{(0.8x+0.1)} + 2.2. \qquad [6]$$

A uniform standard error was also fitted for (noted as the p parameters in Figure 1). This is a widely used tool to estimate the measure of uncertainty. The error parameters σ^2 used in the likelihood of the MCMC

$$L_n = \sum_n \left[\frac{(y_n - y_{model\ n})^2}{\sigma_n^2} + \ln\left(\sigma_n^2\right) \right] \qquad [7]$$

can be defined as:

$$\sigma = \frac{p}{-e^{(0.8x+0.1)} + 2.2}, \qquad [8]$$

which is the fitted standard error p over the weighting function. Figure 2 visualize the above described σ term as exponentially growing error bars fitted for.

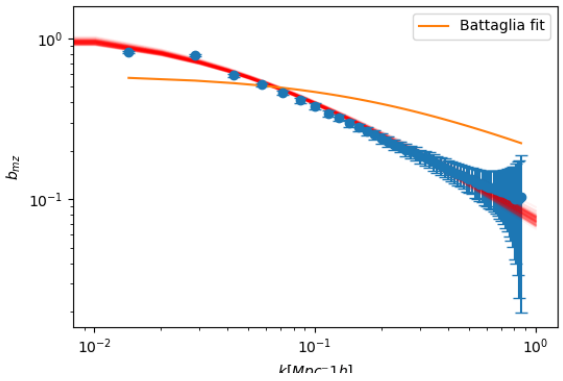


Fig. 2. The linear bias' data points and a 100 best-fitted linear bias functions (red) including α and k_0 parameters randomly selected in the parameter space. Error bars are shown in blue, describing the σ term presented in equation 8. A linear bias function with parameters proposed by z-reion is shown in orange.

Results

In this section, results from the different steps of the project are presented.

Ionization map comparison

Figure 3 presents the ionization history from redshift 5.5 to 13, with a resolution of 0.0075. The ionization history represents the evolution of the proportion of the ionized IGM, one being fully ionized and 0 not ionized. The significantly steeper ionization history curve from the *z-reion* simulation indicated that the reionization is occurring in a much quicker process. The values for the best fit parameters of 21cmFAST plugged in the Battaglia model were: $\alpha = 0.82$, $b_0 = 1.16$ and $k_0 = 0.04$. A Figure presented in the Battaglia et. al paper (the paper presenting z-reion) (5), suggest that for those parameter values, the width of reionization should be $\Delta z \approx 0.5$. The width of reionization is the redshift range in which 50% of the reionization occurred and can be defined by the following equation:

$$\Delta_z \equiv z\,(x_e = 25\%) - z\,(x_e = 75\%)\,. \qquad [9]$$

The observed reionization width from *z-reion* in Figure 3 is in concordance with Battaglia et al. observations. Although the curve appears like an unphysical step-function, the redshift of reionization field is physical, with only an abnormally short width of reionization. This plot does not suggest an agreement in the models' reionziation temporal evolution. The ionization history curves intersect at the mean redshift of reionization, and their ionization maps were observed at that point.

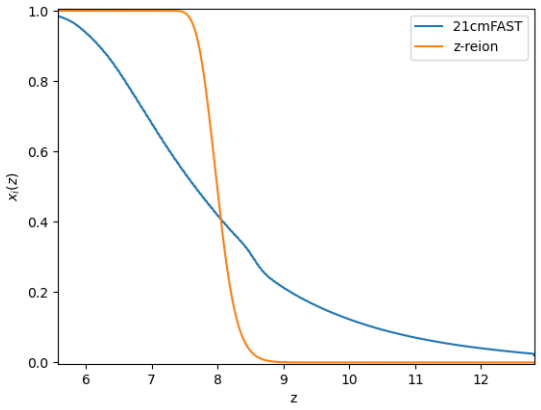


Fig. 3. The ionization fraction as a function of redshift for two different reionization models: *21cmFAST* and *z-reion* (blue, and orange lines respectively)

Figure 4 presents a slice of the ionization map of both simulations at the mean redshift of reionization $z = 8.011$, and their differences. The ionization map of 21cmFASt is shown on the blue scale, and the ionization map of z-reion on the red scale to ease visualization. Observations of the ionization maps show similar large-scale structures, but the ionization map from *z-reion* lacks in small scale structure, looking more "bulk." This is confirmed by looking at the ionization map difference (Figure 4 c)), where fine structures of red are observed. However, main ionization bubbles appear to be more extended in the *21cmFAST*, as the ionization maps difference shows grouped regions of *21cmFAST*'s ionized region around the main ionized structures.

Figure 5 shows the ratio of the power spectrum of the difference map over the power spectrum of 21cmFAST's ionization map. Therefore, this absolute measure of the power spectrum evaluates the correlations of the differences between the maps. The correlation in momentum space increases with k in a proportional manner, where the correlation in the differences peaks at high k. The power spectrum peaking at high frequencies means that the field is more correlated with small than large structures. This confirms the visual interpretation of the ionization maps. The analysis will be further developed in the Discussion section.

Parameter variation

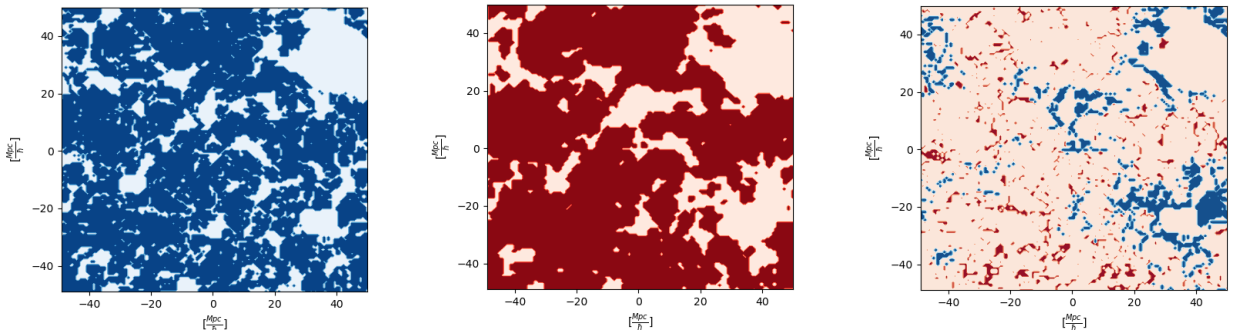
Finally, Figure 6 presents the variational range of the free parameters α, b_0 and k_0 with *21cmFAST* two major inputs : the ionization efficiency and the turnover mass. Results over the whole physically significant range of values are presented. Simple linear relations are observed, with a smooth and slowly varying trend in the best-fit values. Slow variation suggests a weak correlation between *21cmFAST*'s input and the best fit values of the free parameters.

Discussion

Results grants confidence that the comparative code is working. *z-reion* curve in Figure 3 width of reionization is in agreement with expected value from the Battaglia et al. paper. The power spectrum shows a high correlation between the ionization maps at higher k values, which is in agreement with the observations made in Figure 4 c).

Results indicates discrepancies in between the redshift of reionization fields from *z-reion* and *21cmFAST*. Firstly, the ionization histories from Figure 3 show a significant difference in the width of reionization, with the reionization in *21cmFAST* occurring in a much slower process. The Battaglia et al. paper presented an inversely proportional relationship between the α and k_0 parameters and the width of reionization. Figure 6 shows that these free parameters varies, although slightly, with *21cmFAST*'s inputs, indicating that the ionization width could be closer than for the default inputs values of *21cmFAST*. This suggests that a possible combination of inputs in *21cmFAST* can give ionization history curves closer to the one produced by *z-reion*.

Results from Figure 4 and 5 indicates a disagreement between the ionization maps at the mean redshift of reionization. The disagreement is occurring especially at low scales (high k values). These results are in agreement with expected behaviour: Since the *z-reion* model is a semi-numerical model computing the redshift of reionization through only a statistical term, it omits important physical laws. As the complexity and number of implemented physical laws in numerical simulations are directly proportional to the low-scale precisions of a model, it was expected that the *21cmFAST* model would produce an ionization map with more refined structures. Results confirmed that hypothesis



(a) 21cmFAST model's ionization map. Ionized regions are shown in dark blue.

(b) z-reion model's ionization map. Ionized regions are shown in dark red.

(c) The difference of the two models' ionization maps (21cmFAST- z-reion). Regions concurring in boths models are shown in neutral white. Regions ionized by 21cmFAST and not z-reion are shown in dark blue, while regions ionized by z-reion and not 21cmFAST are shown in dark red.

Fig. 4. Ionization maps at the mean redshift of reionization $\bar{z} = 8.011$ for two different models and their difference. All three maps are two dimensional slices of their corresponding 3-dimensional field, sliced in the middle of the third dimension. All plotted maps are $100\frac{Mpc}{h}$ by $100\frac{Mpc}{h}$, with a pixel resolution of $(1\frac{Mpc}{h})^2$

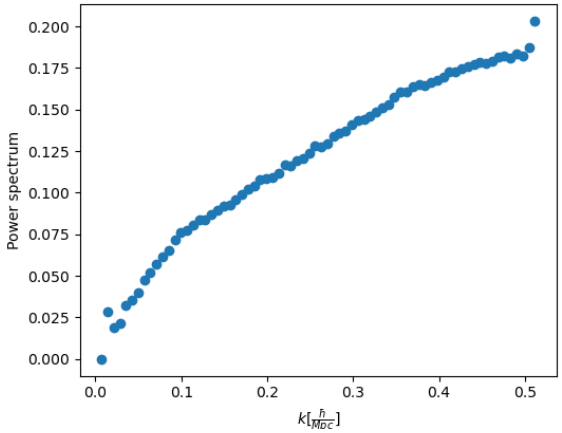


Fig. 5. The power spectrum as a function of the wavenumber k. Each data point represents the ratio of the differences in the ionization map and the 21cmFAST ionization map at a given k in Fourier space.

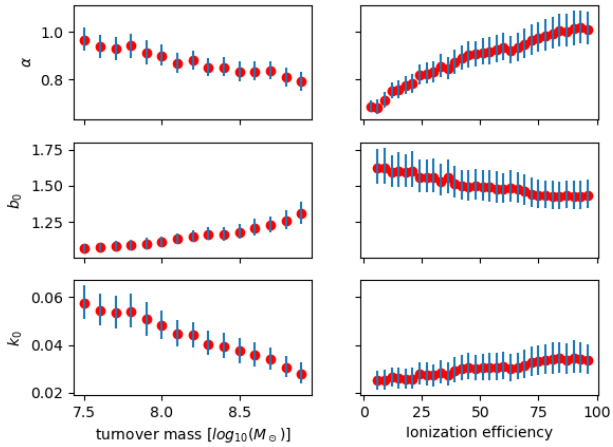


Fig. 6. The evolution of the three best-fitted parameters of the linear bias (equation 2) as a function of the turnover mass and the ionization efficiency. The red dots represent the median of the posterior distribution given by the MCMC presented in Figure 1. The error bars represent the confidence interval, defined as 68% of the posterior distribution. All plots are results of simulations with constant cosmological and astrophysical parameters, with a variation of the ionization efficiency, going from 3 to 99 by three increments.

and showed that these fine-structured discrepancies were significant. Significantly computationally faster, *z-reion* neglect of physical law is visible in the ionization map comparison.

In the near future, the project will be pushed forward during an internship at the *McGill Space Institute* in the *Cosmic Dawn Group* in the summer of 2022. Promising leads includes the optimization of *21cmFAST* inputs to obtain the closest possible ionization histories in between *21cmFAST* and *z-reion*, and to obtain ionization maps with the lowest power spectrum of differences at the mean redshift of reionization. In addition, convenient fitting functions will be created to figure out what equivalent parameters settings must be used in one code to produce a similar redshift of reionization fields. Further work includes a machine learning project to transfer knowledge between the redshift of the reionization field.

Conclusion

This project aimed to set up a comparison between 21cmFAST and z-reion, comparing their redshift of reionization field through their ionization history and their ionization maps at the mean redshift of reionization, while also studying the variational range of the free parameters when varying *21cmFAST*'s inputs to evaluate a possible optimization of parameters in both models to produce similar redshift of reionization field.

Creating a comparison code to compute the ionization histories and the ionization maps at the mean redshift of reionization $\bar{z} = 8.011$, the code comparison of the redshift

of reionization field showed that the models have significant discrepancies on finer scales. However, the free parameters' variational range showed that optimizing the models' inputs to minimize the differences in the models' outputs is possible, and the project's comparison goal has been reached.

ACKNOWLEDGMENTS. A huge thanks to my supervisor: prof. Adrian Liu for his precious guidance, time, confidence, tips and for allowing me to continue working on this project. I would also like to thank the Cosmic Dawn Group at McGill University, in particular Elizabeth McBride. She provided me with the necessary redshift of reionization field and has given me countless amount of help, support and time. Finally, a huge thank you to and loving family for their emotional support.

References

1. M Tanabashi, et al., Review of Particle Physics. *Phys. Rev. D* **98**, 030001 (2018).
2. Particle Data Group, et al., Review of Particle Physics. *Phys. Rev. D* **98**, 030001 (2018) Publisher: American Physical Society.
3. PA Carroll, et al., A high reliability survey of discrete Epoch of Reionization foreground sources in the MWA EoR0 field. *Mon. Notices Royal Astron. Soc.* **461**, 4151–4175 (2016).
4. JA Weber, AWA Pauldrach, JS Knogl, TL Hoffmann, Three-dimensional modeling of ionized gas - I. Did very massive stars of different metallicities drive the second cosmic reionization? *Astron. & Astrophys.* **555**, A35 (2013) Publisher: EDP Sciences.
5. N Battaglia, H Trac, R Cen, A Loeb, Reionization on Large Scales I: A Parametric Model Constructed from Radiation-Hydrodynamic Simulations. *The Astrophys. J.* **776**, 81 (2013) arXiv: 1211.2821.
6. H Trac, R Cen, A Loeb, Imprint of Inhomogeneous Hydrogen Reionization on the Temperature Distribution of the Intergalactic Medium. *The Astrophys. J.* **689**, L81 (2008) Publisher: IOP Publishing.
7. A Mesinger, S Furlanetto, R Cen, 21cmfast: a fast, seminumerical simulation of the high-redshift 21-cm signal: 21cmfast. *Mon. Notices Royal Astron. Soc.* **411**, 955–972 (2011).
8. S Sharma, Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. (2017).