Effects of variable cosmological parameters on the hydrogen 21cm cosmic dawn signal

Supervised by Prof. Anastasia Fialkov and Jiten Dhandha

Kyle Wong[1]

[1] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*

(Dated: April 28, 2025)

I. INTRODUCTION

A. 21cm Cosmology

Understanding the formation and evolution of cosmic structure remains one of the central goals of modern cosmology. While observations of the cosmic microwave background (CMB) and large-scale structure surveys have provided invaluable insights into the early and late-time universe, there exists a significant observational gap between the release of the CMB (∼380,000 years after the Big Bang) and the emergence of the first luminous structures several hundred million years later. This intermediate period, encompassing the so-called Dark Ages, Cosmic Dawn, and the Epoch of Reionization (EoR), holds crucial information about the universe's thermal and ionization history, the formation of the first stars and galaxies, and the onset of feedback processes.

21cm cosmology offers a unique and powerful tool to probe this otherwise inaccessible era. The signal arises from the hyperfine transition of neutral hydrogen (HI), which occurs when the relative spin orientation of the proton and electron flips from parallel to antiparallel, emitting or absorbing a photon with a rest-frame wavelength of 21 centimeters (corresponding to 1.42 GHz). Because neutral hydrogen was the most abundant element in the early universe, the 21cm line provides a pervasive and potentially highly informative tracer of matter distribution over cosmic time.

As the universe expands, the 21cm signal is redshifted, allowing observations at different frequencies to correspond to different epochs. By mapping the sky across frequency channels, it is possible to construct a three-dimensional tomographic view of the intergalactic medium (IGM). This makes the 21cm line a particularly sensitive probe for the thermal history of the IGM, the timing and topology of reionization, the formation of the first stars and black holes, and potentially, physics beyond the standard cosmological model, such as dark matter interactions or exotic energy injection.

The brightness temperature of the 21cm signal, measured relative to the CMB, depends on the spin temperature of hydrogen, the neutral fraction, and the local density field. The differential brightness temperature can be written as probably remove this equation from the introduction and move it to theory

$$\delta T_b(\nu) = 27\ x_{HI}\ (1+\delta_b)\left(1-\frac{T_\gamma}{T_s}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_m h^2}\right)^{1/2}\left(\frac{\Omega_b h^2}{0.023}\right)\mathrm{mK} \tag{1}$$

where x_{HI} is the neutral hydrogen fraction, δ_b is the baryon overdensity, T_s is the spin temperature, T_γ is the CMB temperature at redshift z, and Ω_m, Ω_b are the matter and baryon density parameters respectively.

Detecting this signal presents substantial technical challenges. The cosmological 21cm signal is typically five orders of magnitude fainter than galactic and extragalactic foregrounds, including synchrotron emission

from our Galaxy. Additionally, instrumental systematics, ionospheric effects, and radio frequency interference (RFI) must be mitigated with extreme precision.

Despite these obstacles, a growing number of dedicated low-frequency radio interferometers—such as LOFAR, MWA, HERA, and the upcoming Square Kilometre Array (SKA)—are designed to detect and characterize the 21cm signal from the early universe. These instruments aim to measure the power spectrum of 21cm fluctuations, and eventually, perform direct imaging of the neutral IGM.

21cm cosmology is poised to become a cornerstone of observational cosmology, potentially offering a detailed timeline of the universe's first billion years and enabling precision tests of fundamental physics in a previously uncharted epoch.

II. 21CMSPACE

For any physical system, simulations are an invaluable asset for a multitude of reasons, including generating mock data from mathematical theory, which can then used for pipeline development and validation, as well as for foreground and instrumental modeling, which can then be used for experimental forecasting to inform instrument design.

However, evolving the early universe on a machine is no easy task. Therefore, there does not exist one optimal way to simulate the period of time between recombination and reionization, but rather a multitude of different methods each with their own advantages and disadvantages. These methods lie on a spectrum with a trade off between accuracy and runtime. On one end lies numerical simulations, which hold accuracy as its foremost priority. This is achieved through explicit evolution of structure formation using hydrodynamic theory cite thomas?, which can either be freshly written with the intent of simulating the 21-cm signal from nativity, or taken from a generic library upon which processes such as radiative transfer and chemical evolution can then be attached for the specific purpose of evolving the 21-cm observables. This approach, due to its theoretical ability to include a comprehensive suite of physical effects, advertises the best possible control over the processes at each evolution step. However, numerical simulations come with the heavy downside of computational cost. example? While it is true that computer performance is exponentially increasing alongside decreasing costs, the need and expectation for improvements in simulation accuracy, size, and resolution have continued to render purely numerical simulations too expensive both in computational power and time to perform large-scale explorations of parameter spaces.

On the other end of the spectrum are analytical simulations, which, in contrast to numerical simulations, do not explicitly evolve spatial volumes through time. Instead, they solve mathematical equations, which, using the plethora of numerical solution libraries available, can take less than a second using commercial hardware cite?. However, this comes at the heavy trade-off of losing significant physical detail, with analytical equations having only the capability to model fields through averaged quantities, requiring significant approximations. Still, these calculations are of incredible value since these approximations are often subdominant compared to the inherently large observational error of the field. Despite their practical merit, though, the shortcomings of analytical simulations with regards to the lack of explicit evolution often prohibits output of full 21-cm signal maps or 21-cm power spectra.

Through the advantages and disadvantages of both numerical and analytical models, the need for a third class of simulations is hopefully clear: semi-numerical simulations. As the name suggests, semi-numerical simulations numerically resolve and evolve a spatial expanse, but rather than incorporating full hydrodynamic calculations of all processes at every step and every point in space, semi-numerical simulations invoke analytic calculations to deal with approximated quantities. Semi-numerical simulations therefore offer a compromise between accuracy and low computational cost. It is to this class of semi-numerical simulations

that 21cmSPACE belongs.

A. Design principles

21cmSPACE (21-cm Semi-numerical Predictions Across Cosmic Epochs) evolves large-scale structure using analytic or perturbative solutions, while key astrophysical processes are included via parametric, sub-grid models and numerical integration where necessary. The core aim of 21cmSPACE is to propagate the 21-cm brightness temperature field forward in time, in order to compute observable quantities such as the global (sky-averaged) 21-cm signal and its power spectrum. To achieve this, the code self-consistently tracks the evolution of all relevant fields – e.g., the hydrogen spin temperature T_S, the background radiation temperature T_γ, the hydrogen neutral fraction $x_{\rm HI}$, as well as derived quantities like star formation rates and radiation intensities.

By design, 21cmSPACE divides the problem according to scale: large-scale intergalactic fields (such as density, velocity, radiation backgrounds) are evolved on a coarsely resolved simulation grid, while small-scale phenomena (halo collapse, star formation, feedback) are handled by sub-grid prescriptions. include illustration as in thomas thesis? This separation of scales is the key to computational efficiency, without needing to sacrifice essential physical processes.

Several design philosophies underlie 21cmSPACE. First, it emphasises flexibility in exploring astrophysical scenarios – a wide array of input parameters include table like in thomas thesis? control star formation efficiencies, feedback strengths, spectral emissivities, etc., enabling the user to test different models of early-Universe astrophysics. For example, the efficiencies of Pop II and Pop III star formation ($f_{*,II}$, $f_{*,III}$) and the delay time between Pop III and Pop II episodes ($t_{\rm delay}$) are all tunable inputs. Likewise, the relative X-ray luminosity of high-z X-ray binaries (f_X) and any additional radio background strength (f_r) can be specified. Second, the code is structured for performance: any components of the calculation that do not depend on the specific astrophysical parameters are precomputed once and stored for reuse. This includes, for instance, cosmological tables, linear perturbation growth factors, and radiation window functions (used for fast radiative transfer, discussed later). Third, 21cmSPACE assumes a fixed cosmological model during a run (by default the Planck 2013 ΛCDM parameters) and is optimized under that assumption. In principle the cosmology can be changed by updating internal lookup tables and regenerating precomputed grids, but the code is primarily designed to vary astrophysical inputs rather than cosmological ones, since uncertainties in high-z astrophysics are much larger than current cosmology uncertainties. Overall, the architecture of 21cmSPACE prioritizes physical fidelity (by including all major known 21-cm relevant processes) while maintaining speed through analytical treatments and precomputation. This makes it well-suited to produce rapid predictions of 21-cm observables across cosmic dawn and reionization for a range of scenarios.

B. Temporal, spatial, and cosmological framework

Before any simulation, the time, space, and cosmology domains must be configured. 21cmSPACE uses a fixed temporal stepping scheme in redshift space, spanning redshift $z \approx 50$ down to $z = 6$. The simulation begins at $z = 50$, a time when structure formation is in its infancy and linear theory is valid, and ends at $z = 6$, by which point reionization is essentially complete and the cosmological 21-cm signal is vanishing. The time steps are not uniform: from $z = 50$ to 15 the code advances in unit redshift increments $\Delta z = 1$, and then switches to finer steps of $\Delta z = 0.1$ from $z = 15$ to 6. This refinement at lower redshifts is to resolve rapid changes during reionization, while larger steps at high redshift suffice when the evolution is

slow. Starting at $z = 50$ is deliberate – it ensures the initial conditions can be taken directly from linear perturbation theory (with negligible nonlinear structure or star formation by that point). Accordingly, 21cmSPACE initialises each simulation with density, velocity, temperature, and ionisation fields drawn from standard linear codes: matter overdensity and baryon–dark matter relative velocity fields are generated from the linear power spectra (e.g., using CAMB outputs), and initial gas temperature and ionised fraction come from primordial recombination calculations (e.g. RECFAST). These initial conditions at $z = 50$ establish the proper starting state of the intergalactic medium (IGM) before the first stars.

In space, 21cmSPACE uses a three-dimensional Cartesian grid of cells to discretise the simulation volume. The cell size and number are chosen to balance two competing needs: capturing a large volume for statistical representativeness, and maintaining sufficient resolution to model the relevant physical scales. Throughout the project work, a grid of 128^3 cubical cells was used, with each cell being 3 comoving Mpc (cMpc) on each side. This yields a simulation box of side length $128 \times 3 = 384$ cMpc and volume $(384\ \mathrm{cMpc})^3$. Such a volume is large enough to provide meaningful predictions for global signals and power spectra (covering wavenumbers $k \sim 0.05$–$1.0\ \mathrm{cMpc}^{-1}$, matching the range probed by current interferometers like HERA). At the same time, a 3 cMpc cell size is fine enough to resolve features like the characteristic scale of baryon–dark matter streaming motions (~ 10 cMpc coherence length), and yet coarse enough that individual cell regions remain roughly in the linear regime until $z \sim 6$. (In fact, if the cells were much smaller, dense regions would collapse and deviate from linear evolution earlier, violating assumptions in the code's analytic updates.) The chosen 3 cMpc resolution thus ensures the sub-grid astrophysical models (for star formation, radiative sources, etc.) remain valid on the cell scale up to the end of the simulation. It is worth noting that 21cmSPACE is not limited to 128^3 – as of 2023, the code supports arbitrary grid sizes N_{cell}, allowing much larger volumes (e.g. to cover the full SKA-Low field of view) if computational resources permit.

As mentioned, the simulation assumes a ΛCDM cosmology (Planck 2013 parameters by default) for computing expansion history and linear perturbations. Within each time step, cosmic expansion is accounted for in all equations (e.g. redshifting of radiation, Hubble cooling of gas). The code updates the background CMB temperature $T_\gamma(z)$ according to $T_\gamma = 2.725(1+z)$ K (the CMB is taken as the default radio background) unless an additional radio background is specified. While alternative cosmologies could be input by manually editing the cosmological parameter table and recomputing certain grids, the architecture is optimized for a fixed cosmology, leveraging the fact that astrophysical unknowns dominate over cosmological ones for the 21-cm signal. In summary, the simulation grid and timeline are configured to reliably capture the full 21-cm relevant epoch ($50 > z > 6$) with adequate resolution and physical initial conditions, providing the stage on which the astrophysical simulation unfolds.

C. Main Simulation Loop and Sequence of Operations

Once initialized, 21cmSPACE enters its main simulation loop, iterating over each time step (redshift decrement) from $z = 50$ down to $z = 6$. At each step, the code updates all relevant fields in a prescribed sequence. The following outlines the major steps executed in one simulation cycle (Gessey-Jones):

1. Update large-scale cosmological fields: The cosmological density fluctuation $\delta(\mathbf{x})$ in each cell and the baryon–dark matter relative velocity $v_{\mathrm{bc}}(\mathbf{x})$ are advanced from the previous redshift to the new redshift. While density fluctuations eventually grow non-linear in overdense cells, in practice 21cmSPACE applies linear growth (scaling by the linear growth factor) at each small redshift step, which is a good approximation on the chosen grid scale until late times. Similarly, the streaming velocity decays with time according to linear theory and is updated analytically (Gessey-Jones). These updates account for

the cosmological expansion and structure growth between time steps.

2. Calculate halo abundance in each cell: Using the updated density (and residual streaming velocity), the code computes the halo mass function locally in each cell. 21cmSPACE employs an analytic Press–Schechter-like formalism (specifically a hybrid of the Press–Schechter and Sheth–Tormen prescriptions, following the method of Barkana & Loeb) that has been modified to include the effects of the local overdensity and streaming velocity on halo formation (Gessey-Jones). In essence, overdense regions and regions with low relative velocity form more halos, while underdense or high-$v_{\rm bc}$ regions form fewer. This yields a spatially varying halo mass function across the grid (each cell gets a distribution or count of halos of different masses). The method is calibrated against N-body simulations and is more accurate at high redshift than a simple global Press–Schechter recipe (Gessey-Jones).

3. Compute star formation rates (Pop III and Pop II): Given the halo population in a cell, 21cmSPACE next determines how many stars form in those halos. It uses a sub-grid star formation prescription adapted from Magg et al. (2018) to produce both Population III (metal-free) and Population II (normal metal-enriched) star formation in tandem (Gessey-Jones). In summary, when a halo first reaches the minimum mass to allow cooling and star formation, it is assumed to host a burst of Pop III star formation with an efficiency $f_{*,\rm III}$ (a fraction of the halo's baryons turn into Pop III stars). After this initial burst, the halo undergoes a recovery period of duration $t_{\rm delay}$ (to account for feedback from Pop III supernovae blowing out gas). Once this delay time has passed, the halo can then sustain continuous Pop II star formation with efficiency $f_{*,\rm II}$ going forward.Thus, each halo transitions from a Pop III-producing phase to a Pop II-producing phase. The parameters $f_{*,\rm III}$, $f_{*,\rm II}$, and $t_{\rm delay}$ are inputs to the code that can be adjusted to explore different astrophysical scenarios (for instance, to test how a more or less efficient first-star formation would affect the 21-cm signal) (Gessey-Jones). In the implementation, a set of fitting formulas (based on high-resolution "merger-tree" simulations by A-SLOTH) is used to compute the star formation rate in each cell from the halo mass distribution, taking into account these efficiencies and the timing of halo formation (Gessey-Jones). The outcome of this step is a cell-averaged star formation rate density (SFRD) for Pop III and Pop II in every cell at the current time.

4. Apply star formation feedback effects: As part of the star formation module, 21cmSPACE incorporates several key feedback mechanisms that can suppress star formation in low-mass halos. These include:

   - Lyman-Werner (LW) feedback: Dissociating UV photons (in the LW band) can destroy molecular hydrogen, which metal-free (Pop III) star formation relies on. A strong LW background thus raises the minimum halo mass $M_{\rm crit}$ needed for Pop III star formation.
   - Streaming velocity feedback: A large baryon–DM streaming velocity (a remnant from pre-reionization structure formation) inhibits gas collapse into small halos. This effect also effectively increases the minimum halo mass for star formation in regions with high $v_{\rm bc}$.
   - Photoheating feedback: Once reionization begins, ionizing photons heat the IGM and can evaporate gas out of small halos, preventing them from forming stars. This mainly affects the later stages (Pop II in low-mass halos during reionization).

   In 21cmSPACE, these feedbacks are represented by raising the local star-formation threshold mass $M_{\rm crit}$ above a baseline value (set by a virial temperature or circular velocity criterion) in cells where the LW intensity, streaming velocity, or ionizing background are significant. The code includes formulas for how each feedback boosts $M_{\rm crit}$ based on physical models or simulations (Gessey-Jones). By increasing

$M_{\rm crit}$, the star formation in that cell (especially Pop III) is suppressed accordingly. Additionally, 21cmSPACE offers an optional suppression of Pop II star formation efficiency in halos between the molecular-cooling threshold and the atomic-cooling threshold. This phenomenological tweak gradually reduces $f_{,\rm II}$ in small halos that are just above $M_{\rm crit}$, reflecting the idea that prior star formation in the halo can deplete or heat some of the gas (Gessey-Jones). When activated, this causes the Pop II star formation rate (SFR) to ramp up from zero at $M_{\rm crit}$ to the nominal $f_{,\rm II}$ by the time a halo reaches the atomic cooling mass scale. In the default scenarios studied, this extra suppression was not crucial and can be turned off. Another optional feature is to introduce stochasticity in star formation: instead of using the average SFR in each cell, the code can randomly sample the number of halos and their star formation outcomes (Poisson sampling) to mimic shot noise when halos are very few (Gessey-Jones). This stochastic mode captures additional fluctuations (important at very early times $z \gtrsim 25$), but it makes the simulation nondeterministic and was turned off in favor of reproducibility and easier statistical analysis (Gessey-Jones). After applying all feedback effects, the result of steps 2–4 is a finalized Pop III and Pop II SFR for each cell at the current time step.

5. Convert star formation to radiative emissivities: The newly computed star formation rates are then converted into emissivities of various radiation species. 21cmSPACE tracks several radiation fields that are critical for 21-cm physics:

   - Lyman-series (Lyα) photons: These are UV photons capable of scattering in the Lyα transition of hydrogen. They are responsible for the Wouthuysen–Field effect, which couples T_S to the gas temperature.
   - Lyman-Werner (LW) photons: UV photons in the 11.2–13.6 eV range that dissociate H_2, as mentioned in feedback.
   - Ionizing UV photons: Above 13.6 eV, these ionize hydrogen and drive reionization (discussed separately below).
   - X-ray photons: These can travel far through the neutral IGM, heating it and also producing some secondary ionizations. Fialkov et al.
   - Cosmic ray particles: High-energy particles (if included) that can propagate and heat the IGM.
   - Radio emission: Any additional radio background (e.g., from early radio galaxies or dark matter decay) that would effectively raise T_γ.

   For each of these, the code uses the star formation rate of Pop III and Pop II in a cell to calculate how many photons (or what luminosity) that cell produces. For example, Pop II star-forming halos are assumed to have an X-ray luminosity proportional to their star formation rate,

$$L_X = (3 \times 10^{40} \text{erg s}^{-1} M_\odot^{-1} \text{yr}) f_{\rm x} \text{SFR} \tag{2}$$

   where f_X is the X-ray emission efficiency parameter (Gessey-Jones). A similar relation (with potentially a different f_X) can be used for Pop III halos. The spectra of emitted X-rays can be chosen (21cmSPACE allows using a template Pop II X-ray binary spectrum, a power-law, or a Pop III spectrum from detailed models) (Gessey-Jones). Lyman-band photons from stars are computed by assuming stellar population spectra: Pop II stars use a standard stellar population spectrum (e.g., from Leitherer et al.), while Pop III stellar spectra are derived from the Pop III initial mass function in the model (Gessey-Jones). In essence, the code multiplies the star formation rate by an appropriate luminosity or photon production yield to get the emissivity (photons per second per comoving volume) for each radiation field of interest in each cell.

6. Propagate radiation fields through space: Once the emissivity (sources) are known, 21cmSPACE computes the radiation intensity filling each cell by propagating photons from all sources. A key simplification enabling fast computation is the use of a Fourier-space convolution (window function) method for radiative transfer (Gessey-Jones). Instead of tracing rays for millions of sources, the code uses precomputed spherical window functions that describe the average intensity profile around a source for each type of radiation. For photons that travel without absorption (like LW or Lyα beyond a certain wavelength), the window function is essentially a spherical top-hat: a photon travels until redshifted out of the band. For photons that undergo absorption (X-rays being absorbed by neutral gas, Lyα between Lyα and Lyβ undergoing scatterings), the window functions incorporate those effects (e.g. an exponential attenuation with distance for X-rays due to absorption cross-sections, or a diffusion kernel for Lyα scatterings). By convolving the emissivity field with these window functions, the code obtains the radiation energy density or intensity field in every cell. This convolution is done efficiently via Fast Fourier Transforms, treating the window function as a filter. The radiative transfer also accounts for redshift (cosmological) attenuation and light-cone effects (photon travel time across the simulation volume), ensuring that the finite speed of light and cosmic expansion are included (Gessey-Jones). The outcome is that for each cell we now have, for example, the local Lyα intensity $J_\alpha(\mathbf{x})$, the LW intensity, the X-ray flux spectrum $J_X(\mathbf{x}, E)$, etc., at the current time. These fields are crucial for the next step. (Ionizing UV radiation is handled a bit differently via an excursion-set approach in the reionization step below, because ionizing photons create sharp ionized/non-ionized regions rather than a smoothly decaying intensity field.)

7. Update gas temperature and ionization: Given the radiation fields and other local quantities, 21cmSPACE then advances the state of the intergalactic gas in each cell. This is done by solving a set of coupled differential equations that govern the heating/cooling of the gas and the growth of ionization. The code uses a numerical integrator (Runge–Kutta method) to solve these equations simultaneously at each time step (Gessey-Jones). One equation is the thermal evolution equation, which equates the change in gas kinetic temperature T_K to various heating and cooling terms. Heating terms include X-ray heating (from the X-ray intensity computed earlier), Compton heating by CMB photons, Lyα heating, cosmic ray heating (if enabled), etc., all summed into a total heating rate per baryon for the cell (Gessey-Jones). Cooling terms include adiabatic cooling due to the expansion of the Universe and additional cooling from the increased number of particles if the gas is being ionized (since energy gets distributed into new free electrons). There is also a small heating term from structure formation (shock heating from collapse, which 21cmSPACE approximates based on the change in baryon density). At the same time, an equation for the evolution of the ionized fraction x_e is solved. For ionization, 21cmSPACE adopts an excursion-set formalism to determine which regions become fully ionized by UV radiation (this handles reionization), coupled with a differential equation for partial ionization by X-rays and other minor sources (Gessey-Jones). In practice, the code first checks the collapse fraction (fraction of mass in halos) in a region against a threshold condition (governed by an ionizing efficiency parameter ζ) using an excursion-set approach similar to Mesinger et al. – if a cell (or a larger region containing that cell, scanned over various scales) has $\zeta f_{\rm coll}$ above a certain threshold, it is declared fully ionized by stellar UV photons. Those cells are set to $x_{\rm HI} = 0$. Cells that do not meet the criterion are not fully ionized by UV, but they can still have partial ionization from X-rays or cosmic rays. 21cmSPACE models those partially ionized cells as a two-phase medium: a fraction of the cell is treated as fully ionized (H II) and the remainder as mostly neutral with a small ionized fraction $x_{e,\rm oth}$ due to X-rays etc. (Gessey-Jones). The effectively neutral fraction in a cell is then $x_{\rm HI} = 1 - \zeta f_{\rm coll} - x_{e,\rm oth}$ (ensuring that when UV ionizations $\zeta f_{\rm coll}$ and other ionizations $x_{e,\rm oth}$

sum to 1, the cell is fully ionized). The parameter $x_{e,\mathrm{oth}}$ is computed from the X-ray ionization rate and any other non-UV contributions in the cell. By solving the temperature equation and updating ionization in this way, 21cmSPACE advances the spin temperature T_S toward the kinetic temperature (accounting for Lyα coupling strength) and updates the neutral fraction, completing the evolution of the 21-cm brightness temperature field for that time step.

8. Record outputs and reiteration: After updating the fields, any quantities that are needed later or for output are stored. For example, 21cmSPACE will save the current values of T_γ (effective background radiation temperature), the star formation rates, the ionization fraction, etc., as well as intermediate fields it might need for post-processing (Gessey-Jones). The simulation then moves to the next redshift step and repeats the loop (steps 1–8) for that new time. This process continues until the final redshift ($z = 6$) is reached.

Through this loop, 21cmSPACE self-consistently evolves the cosmic gas and radiation fields from the initial conditions to the end of reionization. The result is a time-series of 3D fields (density, T_K, T_S, x_{HI}, etc.) or equivalently the 21-cm brightness temperature field $\delta T_{21}(\mathbf{x}, z)$ at each redshift. These rich data contain the information needed to derive observable signatures of the 21-cm signal.

It should be noted that 21cmSPACE is optimized by precomputing any components of the calculations that do not depend on the particular astrophysical parameters of a run. For instance, the cosmological initial power spectra or the window functions for radiative transfer can be generated once and reused for many simulations. These are stored on disk and loaded as needed, saving runtime (Gessey-Jones). The code also allows for checkpointing: a simulation can be paused and resumed, which is useful if one wants to stop at an intermediate redshift and perhaps explore a branch of parameter space from there (Gessey-Jones). In normal operation, however, it runs straight through the loop described above.

D. Post-Processing and Outputs

After the main simulation loop concludes at $z = 6$, 21cmSPACE enters a post-processing stage to produce high-level outputs for analysis. The two primary outputs are the global 21-cm signal and the 21-cm power spectrum, but the code can also derive other quantities of interest. First, the code compiles the results of all time steps to construct the full 21-cm brightness temperature field as a function of redshift. This involves combining the evolved spin temperature, density, and ionization fraction to calculate the differential brightness temperature $\delta T_{21}(\mathbf{x}, z)$ in each cell at each epoch. If any additional effects need to be applied (for example, line-of-sight peculiar velocity gradients cause what are known as redshift-space distortions in the observed 21-cm field), those are accounted for at this stage to ensure the final signal is in the form an observer would see (Gessey-Jones). In particular, 21cmSPACE can adjust the 21-cm field for redshift-space distortions before computing summary statistics.

From this spatio-temporal data, the sky-averaged (global) 21-cm signal is obtained by averaging the 21-cm brightness over the entire simulation volume at each redshift. The result is a curve of the mean 21-cm brightness temperature as a function of redshift (or cosmic time). This global signal is a key target for experiments like EDGES and others, and it encapsulates the overall thermal and ionization history of the cosmic gas. 21cmSPACE's global signal prediction includes the impact of all the modeled physics (e.g. it will show the deep absorption feature when the IGM is cold and strongly coupled to T_S, and then a rise toward emission as X-ray heating dominates, etc., followed by a decline to zero as reionization completes). The 21-cm power spectrum is computed by taking the Fourier transform of the 3D 21-cm fluctuation field at various redshifts and calculating the variance as a function of scale (wavenumber k). Specifically, 21cmSPACE

will typically compute the dimensionless power spectrum $\Delta^2_{21}(k)$ from the simulation volume at a given redshift, which can be compared to interferometric observations (e.g. from HERA or the SKA). Because the simulation volume is finite, the code can reliably compute modes in a certain k range (for the 128^3 volume of side 384 cMpc, roughly $k \sim 0.05$ to 1.0 cMpc^{-1}, covering the range where current 21-cm instruments are most sensitive) (Gessey-Jones). The inclusion of redshift-space distortion effects in the previous step ensures that the power spectrum is computed in a way that corresponds to observations (Gessey-Jones). The power spectrum captures the scale-dependent fluctuations induced by, for example, patchy reionization (which boosts large-scale power when large ionized bubbles form) or the clustering of early galaxies. In addition to these, 21cmSPACE can output other diagnostic information. For instance, it can evaluate the contribution of early sources to the present-day unresolved X-ray background in a manner consistent with the simulation (by integrating the X-ray emissivity over redshift) (Gessey-Jones). It also tracks the progress of reionization (e.g. the volume-averaged ionized fraction as a function of z) and could output the size distribution of ionized regions if needed for analysis. The code is flexible in storing any intermediate fields; in practice one could extract, say, the evolution of the LW intensity or the heating rate in the simulation for further study. However, the thesis work primarily focuses on the global signal and power spectrum as the summary observables. Finally, the results are typically interpreted in the context of astrophysical parameters. Because 21cmSPACE runs quickly compared to fully numerical simulations, it can be run for many different parameter combinations (e.g. different $f_{*,\mathrm{III}}$, f_X, initial mass function assumptions, etc.) to see how the 21-cm outputs change. This makes it a powerful tool for comparing with data or forecasting the constraints that observations could place on the first stars. The comprehensive design of 21cmSPACE – incorporating time evolution, three-dimensional space, and detailed cosmological and astrophysical processes – enables it to serve as the engine behind theoretical studies of the 21-cm signal across cosmic epochs (Gessey-Jones8). Its outputs, such as the global signal and power spectrum, are the bridge to observations, allowing researchers to test models of the first stars and galaxies against the forthcoming 21-cm measurements in a computationally tractable yet physically robust way.

E. Project motivation and impact

21cmSPACE, while being a powerful tool for simulation of the 21-cm signal, still holds some limitations. Particularly, as outlined in the previous section, despite there offering the option to include or exclude effects from a plethora of physical evolutionary processes, the instantiation of the simulation from its initial conditions remains confined to the use of its default Planck 2013 best-guess cosmology, and propagation thereof. This is an extremely strong assumption potentially affecting most of the physical processes throughout the evolution of the 21-cm signal through, as a prominent example, the growth rate of the universe.

However, there is great value to be gained from the ability to evolve the 21-cm signal for variable cosmologies. As the amount of experimental data measuring various observable imprints of the 21-cm signal continues to increase, so too does the potential to use 21-cm data as a probe for constraining cosmological parameters. Specifically, the 21-cm signal depends on five cosmological parameters, as summarized in Table I.

These parameters propagate through 21cmSPACE through two main avenues: the initial conditions, which is an umbrella term comprising of the cosmological density fluctuation $\delta(\mathbf{x})$ and the baryon-dark matter relative velocity $v_{\mathrm{bc}}(\mathbf{x})$, as well as the halo mass function dn/dM.

The underlying cosmology in 21cmSPACE can therefore be changed by altering these functions. As a result, the effects of these varied cosmological parameters will be propagated through the evolution of the 21-cm signal, and finally imprinted on the observables of the 21-cm global signal and the 21-cm power

Parameter	Definition
h	Dimensionless Hubble constant
$\Omega_{\mathrm{b},0}$	Baryonic matter density/critical density at $z = 0$
$\Omega_{\mathrm{dm},0}$	Dark matter density/critical density at $z = 0$
$\Omega_{\mathrm{k}}, 0$	Effective curvature density/critical density at $z = 0$
$T_{\mathrm{CMB},0}$	Temperature of the Cosmic Microwave Background (CMB) at $z = 0$

TABLE I: Cosmological parameters impacting the 21-cm signal, along with their definitions.

spectrum.

In principle, this will enable exploration of the cosmological parameter space, relating each point in the five-dimensional space to a particular shape of the global signal and a particular shape of the power spectrum. Along with experimental data, the parameter space may then be constrained to narrow down possible locations for the true cosmology of the Universe. As well as this, the comparison of simulated signals with physical measurements will enable the verification or rejection of cosmological theories.

Do I need to show preliminary differences in the CAMB IC power spectra here, illustrating that the project is expected to make a measurable impact in simulated wavenumbers?

III. THEORETICAL BACKGROUND

A. 21-cm Signal Fundamentals

1. *The hyperfine transition*

cite Liu & Shaw 2019 The 21-cm signal is the result of the hyperfine transition of atomic hydrogen, which, as both the most abundant and most basic element, is comprised of a single electron combined with a single proton. Due to the slight energy discrepancy between the spin-aligned state of this electron-proton pair and the spin-antialigned state, during this transition, atomic hydrogen can either emit or absorb a photon of wavelength 21 centimeters. The spin temperature T_s is useful for studying the 21-cm emission line, and is given by

$$\frac{n_1}{n_0} = 3\,\exp\left(-\frac{h\,\nu_{21}}{k_b\,T_s}\right) \tag{3}$$

where n_1/n_0 is the number of hydrogen atoms in the excited hyperfine (aligned) state over the number of hydrogen atoms in the ground hyperfine (anti-aligned) state, h is Planck's constant, k_b is Boltzmann's constant, and $\nu_{21} \approx 1420.406$ MHz is the frequency of the 21-cm emission line in the rest frame.

It is important to note that the spin temperature is not directly observed; rather, it is the difference between the Cosmic Microwave Background (CMB) temperature T_γ and the 21-cm spin temperature which is measured. Areas in which the 21-cm spin temperature is higher than the CMB temperature result in excess emission compared to what is expected from CMB emissions; on the contrary, when $T_s < T_\gamma$, a photon deficit is measured instead.

The most important quantity when studying the 21-cm emission line is its brightness temberature T_b, which is defined in terms of T_s by

$$T_b(\hat{\mathbf{r}}, \nu) = [1 - \exp(-\tau_{21}(\hat{\mathbf{r}}, z))]\,\frac{T_s(\hat{\mathbf{r}}, z) - T_\gamma(z)}{1+z} \tag{4}$$

where $\hat{\mathbf{r}}$ is a radial unit vector from the observer in the direction of observation, the doppler-shifted frequency of the observed signal ν is given by

$$1 + z = \frac{\nu_{21}}{\nu} \tag{5}$$

and τ_{21} is the 21-cm optical depth (i.e., how much 21-cm light is absorbed or scattered) of the interstellar medium, defined as

$$\tau_{21}(\hat{\mathbf{r}}, z) = \frac{3\hbar c^3 A_{10}}{16 k_b \nu_{21}^2}\,\frac{x_{\mathrm{HI}} n_{\mathrm{H}}}{(1+z)(dv_\parallel/dr_\parallel)T_s} \tag{6}$$

with $v_\parallel$ the proper velocity along the line of sight $r_\parallel$, x_{HI} and n_{H} are the fraction and number density of neutral hydrogen atoms respectively, $\hbar$ is Planck's constant divided by 2π, c is the speed of light, and $A_{10} = 2.85 \times 10^{-15}\mathrm{s}^{-1}$ is the spontaneous emission coefficient of the hyperfine transition, quantifying the probability that an atom in the aligned state will spontaneously decay into the anti-aligned state.

B. Cosmic history

1. *Dark Ages (No Starlight,* $\mathbf{z \sim 1100 - 30}$*)*

After recombination (the decoupling of CMB photons at $z \approx 1100$), the universe entered the Dark Ages, before any stars or galaxies existed. During this era, neutral hydrogen filled the intergalactic medium (IGM) and the 21 cm spin temperature (T_s) was governed by collisions with the cooling gas. Initially, Compton scattering off residual electrons kept the gas (and thus T_s) thermally coupled to the CMB down to $z \sim 300$. Once this coupling broke, the gas cooled adiabatically faster than the CMB ($T_{\rm gas} \propto (1+z)^2$). With collisions still effective at high densities, T_s followed the gas temperature, dropping below the CMB temperature T_γ and producing a 21 cm absorption signal against the CMB. By $z \sim 30$, however, the expanding gas became too diffuse for collisions to maintain coupling, so T_s drifted back toward T_γ, causing the 21 cm signal to vanish (zero contrast). Throughout the Dark Ages, in the absence of astrophysical sources, the 21 cm fluctuations directly trace primordial density perturbations in the neutral hydrogen (assuming HI traces the underlying matter). This makes the Dark Ages 21 cm signal a pristine probe of fundamental cosmology (e.g. the matter power spectrum on small scales), potentially constraining inflationary parameters or dark matter properties (Loeb & Zaldarriaga 2004; Masui & Pen 2010; Muñoz et al. 2015). In principle, observations of 21 cm from these ultra-high redshifts could shed light on new physics beyond the CMB (Scott & Rees 1990; Furlanetto et al. 2019). In practice, detecting the Dark Ages signal is extraordinarily challenging: the relevant frequencies $\nu \lesssim 50$ MHz are heavily contaminated by bright Galactic foregrounds and blocked by the ionosphere on Earth. As a result, this epoch remains unexplored observationally, reserved for futuristic instruments (perhaps a lunar radio array) capable of overcoming these hurdles (Tegmark & Zaldarriaga 2009).

2. *Cosmic Dawn (First Light,* $\mathbf{z \sim 30 - 15}$*)*

The Cosmic Dawn began once the first generation of stars and galaxies formed (likely in halos of mass $\gtrsim 10^5$–$10^6, M_\odot$). Lyman-α photons from these early luminous sources triggered the Wouthuysen-Field effect (Wouthuysen 1952; Field 1958) explain?: Lyα absorption and re-emission cycles flip the hydrogen spin, coupling T_s to the kinetic temperature of the cold IGM gas. As soon as a pervasive Lyα background developed, T_s was driven below T_γ again, inducing a deep 21 cm absorption signal. This expected global absorption trough is the first prominent feature of the 21 cm history (Madau et al. 1997; Pritchard & Loeb 2012). Its depth and timing are sensitive to the onset of star formation and the Lyα production efficiency of the earliest galaxies. As cosmic dawn progresses, new radiative processes come into play: X-rays from the first X-ray binaries, mini-quasars, or hot interstellar gas begin to heat the IGM. These high-energy photons penetrate the IGM and photo-ionize atoms, depositing energy as heat via fast photo-electrons colliding with the gas. Gradually, X-ray heating raises the gas temperature. When the gas (and hence T_s) is heated above the CMB temperature, the 21 cm signal transitions from absorption to emission. The precise redshift at which this turning point occurs depends on the total X-ray luminosity of early sources and the hardness of their spectra (Fialkov et al. 2014; Mirocha et al. 2017). Throughout cosmic dawn, the 21 cm brightness is highly inhomogeneous: regions near early galaxies see strong Lyα flux and early heating, while faraway regions remain colder and unheated. This patchiness encodes rich astrophysical information. Measuring the 21 cm signal (globally or via its power spectrum) during cosmic dawn would allow us to infer properties of the first sources – for example, the minimum halo mass able to host star formation, the stellar initial mass function, and the X-ray production efficiency (Pober et al. 2013a; Mesinger et al. 2014; Fialkov et al. 2017).

In essence, 21 cm observations during cosmic dawn directly probe the birth of the first stars and galaxies, opening a window on astrophysics at high redshift that was previously accessible only through theory.

3. *Epoch of Reionization (IGM Transformation,* $\mathbf{z \sim 15 - 6}$*)*

As star formation accelerated, the Epoch of Reionization (EoR) unfolded, overlapping with the late stages of cosmic dawn. UV photons from young galaxies (and possibly quasars) gradually ionized the surrounding hydrogen gas, carving out growing ionized (H II) regions in the neutral IGM. Initially these ionized bubbles were small and isolated, but over time they expanded and merged. The volume-averaged neutral fraction of the universe dropped from essentially unity to a few percent by the end of reionization. The 21 cm signal during the EoR became highly patchy. In neutral regions that were already heated ($T_s \gg T_\gamma$), the 21 cm line appeared in emission. In contrast, within ionized zones (or where gas was fully ionized), the 21 cm signal was absent entirely. Thus, 21 cm observations of the EoR can spatially map the distribution of neutral and ionized regions across the universe. The characteristic size and growth of these 21 cm "bright" (neutral) and "dark" (ionized) patches inform us about the nature of reionization sources and the timeline of this phase transition (McQuinn et al. 2007; Friedrich et al. 2011). For instance, a rapid reionization would result in large, mergeable ionized regions appearing over a short interval, whereas a more extended reionization would produce a mix of bubble sizes over a longer period. Current observations (e.g. Gunn-Peterson troughs in $z \sim 6$ quasar spectra and CMB polarization measurements) indicate reionization completed by $z \approx 6$–7 (Fan et al. 2006; Planck Collaboration 2016a). The 21 cm signal provides a direct probe of this process, in contrast to these indirect tracers. By measuring the 21 cm power spectrum or imaging the neutral hydrogen distribution, one can constrain the evolving ionizing photon budget and ionization topology – for example, determining the efficiency of galaxies in ionizing the IGM and the clumpiness of gas that absorbs these photons (Robertson et al. 2010; Greig & Mesinger 2017). In summary, the EoR 21 cm signal links cosmological structure formation with early galactic astrophysics, illuminating how the universe's diffuse gas was transformed from fully neutral to (almost) fully ionized.

4. *Post-Reionization Epoch (After* $z \sim 6$*)*

Once reionization completed, the vast majority of the IGM remained ionized up to the present day. Only a small residual fraction of hydrogen stayed neutral: on the order of $x_{\rm HI} \sim 10^{-2}$ (a few percent of all hydrogen). This remaining neutral gas resides in dense, self-shielded regions such as the interiors of galaxies and pockets of the circumgalactic medium that survived the ionizing UV background. Cosmological simulations indicate that HI is preferentially found in halos with masses roughly 10^{10}–$10^{13} M_\odot$; in lower-mass halos ionizing radiation cannot be fully countered, while in higher-mass clusters gas is stripped out of galaxies. Because the neutral gas is confined to compact objects by this epoch, the 21 cm signal is no longer a continuous background pervading the cosmos, but rather originates from discrete hydrogen-rich systems (the atomic gas in galaxies and protogalaxies). Nonetheless, the 21 cm line remains a valuable tracer of large-scale structure and cosmology after reionization. At very low redshifts ($z < 0.1$), traditional surveys like HIPASS and ALFALFA have directly detected extragalactic objects in 21 cm emission (Barnes et al. 2001; Jones et al. 2018), measuring the HI mass function in the local universe. At intermediate redshifts up to $z \sim 1$, the 21 cm signal from individual galaxies becomes too faint to detect individually in bulk. Instead, researchers employ 21 cm intensity mapping, which measures the aggregate HI emission from many unresolved galaxies across large cosmic volumes (Chang et al. 2010; Battye et al. 2013). Intensity mapping treats the 21 cm sky as a

diffuse background (much like CMB or radio continuum surveys), capturing fluctuations in the HI brightness on large scales without needing to resolve each galaxy. This technique has already achieved first success: by cross-correlating 21 cm maps with optical galaxy surveys, studies have statistically detected cosmic HI at $z \sim 0.8$ (Chang et al. 2010; Masui et al. 2013). These measurements confirm that neutral hydrogen traces the same large-scale structure as galaxies and can be used to probe cosmology in the post-reionization era. For example, 21 cm intensity maps can be used to measure baryon acoustic oscillations (BAO) in the matter distribution as a function of redshift, providing a handle on the expansion history and dark energy (Ansari et al. 2018; Anderson et al. 2018). They can also improve constraints on quantities like the growth rate of structure and the sum of neutrino masses by mapping the matter distribution over immense comoving volumes (Villaescusa-Navarro et al. 2018; Obuljen et al. 2018). In short, even after the universe became fully ionized, the 21 cm line continues to serve as a potent tool: it transitions from a probe of early-universe astrophysics to a tracer of large-scale structure and cosmological evolution in the later universe.

include a graphic of the 21cm global signal, with shading, similar to Jiten's first year report.

C. 21-cm Observables

In astronomy, imaging a signal over a solid angle of sky is often seen as the holy grail of an experimental field. Unfortunately, in the case of 21-cm astronomy, this is inaccessible for a number of physical reasons. First and foremost, the 21-cm signal is extremely faint, with signals from the Epoch of Reionization often on the order of 1mK to 10mK cite Furlanetto et al. 2006, Morales & Wyithe 2010. In contrast, a single radio-antenna system at $\sim$ 150MHz (within the 50 to 250 MHz frequency band of HERA) receives hundreds to thousands of kelvin of Galactic synchrotron emission and receiver noise. Because of this low signal-to-noise (SNR) ratio, there is not enough sensitivity to produce a high-fidelity image. In addition to this, recent experiments in 21-cm astronomy have heavily favored interferometric approaches over single-dish approaches, due to the advantage that variable and vastly long baselines are essential in providing high angular resolution, as well as the statistical simplicity (bias free) in data correlation. However, interferometers inherently miss the total power unless single-dish auto-correlations are included, which is impractical in current experiments due to dominant receiver systematics. cite

Therefore, experiments in the field of 21-cm astronomy rely on statistical observations to inform the history of the universe. The most important of these is arguably the *power spectrum*, whose measurement is the main focus for many current experiments (cite HERA, etc?). To define the power spectrum, first consider the three-dimensional Fourier transform from physical space to Fourier space, defined by

$$\tilde{T}(\mathbf{k}) \equiv \int_{-\infty}^{\infty} d^3 r e^{-i\mathbf{k}\cdot\mathbf{r}} T(\mathbf{r}) \tag{7}$$

where $\mathbf{r}, \mathbf{k}$ are comoving position and comoving wave- vectors respectively. The inverse transform is given by

$$T(\mathbf{r}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} d^3 k e^{i\mathbf{k}\cdot\mathbf{r}} \tilde{T}(\mathbf{k}). \tag{8}$$

The power spectrum can then be defined by the equation

$$\langle \tilde{T}(\mathbf{k}) \tilde{T}(\mathbf{k}')^* \rangle = (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}') P(\mathbf{k}) \tag{9}$$

with δ^D the Dirac delta function in D dimensions, and $\langle ... \rangle$ the ensemble average operation. Equivalently, and perhaps more intuitively, the power spectrum can be interpreted as the Fourier transform of the correlation function $\xi(\mathbf{x}) \equiv \langle T(\mathbf{r})T(\mathbf{r}-\mathbf{x})\rangle$, emphasizing the fact that the power spectrum measures correlations in configuration space, but simply is expressed in Fourier space:

$$\xi(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} P(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{10}$$

This definition of the power spectrum includes the necessary information to completely statistically characterise a Gaussian random field [1], which underlies inflationary models. In cosmological literature, though, it is the quantity

$$\Delta^2(k) \equiv \frac{k^3}{2\pi^2} P(k) \tag{11}$$

which is more commonly seen. The reason for plotting this quantity Δ^2 instead of $P(k)$ can be explained by considering the variance of a zero-mean random temperature field, as in, for example, the case of the mean-subtracted 21-cm brightness temperature field:

$$\begin{aligned}
\langle T^2(\mathbf{r})\rangle &= \left\langle \left(\int_{-\infty}^{\infty} \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \tilde{T}(\mathbf{k}) \right) \left(\int_{-\infty}^{\infty} \frac{d^3q}{(2\pi)^3} e^{i\mathbf{q}\cdot\mathbf{r}} \tilde{T}(\mathbf{q}) \right)^* \right\rangle \\
&= \int_{-\infty}^{\infty} \frac{d^3k}{(2\pi)^3} \frac{d^3q}{(2\pi)^3} e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{r}} \langle \tilde{T}(\mathbf{k}) \tilde{T}(\mathbf{q})^* \rangle \\
&= \int_{-\infty}^{\infty} \frac{dk^3}{2\pi^2} P(k) \\
&= \int_0^{\infty} d\ln k \Delta^2(k).
\end{aligned} \tag{12}$$

Therefore, $\Delta^2(k)$ can be interpreted as the contribution to variance in configuration space in each logarithmic k bin.

While the power spectrum holds information about spatial fluctuations, it is also informative to investigate the *global signal* $\overline{T}_b$, defined as

$$\overline{T}_b(\nu) = \int d\Omega T_b(\hat{\mathbf{r}}, \nu). \tag{13}$$

As the notation suggests, this is an averaged quantity; specifically, it is the average power of the 21-cm signal across all sky angles, as a function of frequency.

D. Linear Cosmological Perturbations

https://arxiv.org/pdf/astro-ph/0103017

---

[1] https://arxiv.org/pdf/astro-ph/0103017

E. Initial conditions

1. Mass overdensities

2. Streaming velocity effect

To define v_{bc}, the velocity divergence must first be defined as

$$\theta \equiv a^{-1} \nabla \cdot \mathbf{v} \tag{14}$$

where a is the dimensionless cosmological scale factor of the universe. what is this divergence physically? Then, the relative velocity is

$$\mathbf{v}_{bc} = \frac{\hat{k}}{ik} \left[\theta_b(\mathbf{k}) - \theta_c(\mathbf{k}) \right] \tag{15}$$

where the subscripts b and c denote the θ for baryons and CDM respectively. The power spectrum of v_{bc} is given by

$$\begin{aligned} \langle v_{bc}^2(\mathbf{x}) \rangle &= \int \frac{dk}{k} \Delta_\zeta^2(k) \left[\frac{\theta_b(k) - \theta_c(k)}{k} \right]^2 \\ &= \int \frac{dk}{k} \Delta_{\mathrm{vbc}}^2(k) \end{aligned} \tag{16}$$

F. Press-Schechter Halo Mass Function

IV. METHODOLOGY

A. Baseline 21cmSPACE Operation

21cmSPACE, originally written in MATLAB, takes as input pre-computed 3-dimensional initial condition grids: the matter overdensity field $\delta_m(\mathbf{x})$ and the relative velocity field $v_{bc}(\mathbf{x})$, and global values for the initial gas temperature and ionization fraction. These were previously hard-coded to follow the Planck 2013 cite cosmology, with parameters defined as the best estimates of the Planck collaboration's results https://www.aanda.org/articles/aa/pdf/2014/11/aa21591-13.pdf table 2, Planck+WP best fit, as shown in Table II.

Parameter needs units	Value
h	0.6704
$\Omega_{\mathrm{b},0}$	0.022032
$\Omega_{\mathrm{dm},0}$	0.12038
$\Omega_{\mathrm{k}}, 0$	-0.001(https://www.aanda.org/articles/aa/pdf/2014/11/aa21591-13.pdf eq67b)
$T_{\mathrm{CMB},0}$	???

TABLE II: Planck 2013 parameter values

B. Initial Conditions Generation

The generation of initial conditions was done primarily in the file `21cmSPACE/grids_and_ic/make_ic/get_IC_N.m`, which defines the function `get_IC_N`. The function takes the two inputs `Npix` and `seed` as input, which respectively represent the side-length of the total simulated spatial volume in number of pixels and the random seed used. The function then outputs two 3-dimensional scalar arrays representing the δ_m field and the v_{bc} field, where the first array holds the unitless deviation from average density, and the second array holds the relative velocity between baryons and CDM in units of kms^{-1}.

`get_IC_N` also further imports three precomputed files, each for a different redshift from $\{40, 970, 1020\}$ containing *transfer function* [2] data output by CAMB. These transfer functions mathematically encode how primordial density fluctuations are affected by processes such as radiation pressure, horizon entry, particle free-streaming cite. As a function of wavenumber, the processed power spectrum $P(k)$ relates to the primordial power spectrum $P_0(k)$ through the transfer function $T(k)$ by the relation

$$P(k) = P_0(k)T^2(k). \tag{17}$$

The primordial power spectrum itself is defined by

$$P_0(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}, \tag{18}$$

[2] https://arxiv.org/pdf/astro-ph/0103017

where k_* is the pivot scale 0.002Mpc^{-1} [3], A_s the amplitude of the spectrum at the pivot scale, and n_s the scalar spectral index. This scalar spectral index represents the tilt of the power spectrum, with $n_s = 1$ corresponding to the Harrison-Zeldovich spectrum, physically meaning that the power is scale invariant.

21cmSPACE currently uses hard-coded values of $A_s = 2.01664 \times 10^{-9}$ and $n_s = 0.9675$, with n_s consistent with the Planck2013 best-fit model (although it is from the Planck+lensing column of table 2, not Planck+WP like the other cosmological parameters?). These values are considered by some to be part of what defines a cosmology; however, this is not unanimous.

These transfer functions from each redshift are then used in `get_IC_N` to calculate the overdensity field power spectra (these are in between power spectra and transfer functions – what are they called?) δ_m, δ_b, and δ_c for each of total matter, baryons, and cold dark matter respectively. To do this, they are first passed through a smoothing window function defined by

$$W(k) = \exp\left(-\frac{k^2 R_w^2}{2}\right) \tag{19}$$

where R_w is a characteristic smoothing scale, chosen to be $R_w = 1.7/\pi$ for the smoothing of 3 MPc pixels. [4] This window function serves the purpose of smoothing scales smaller than R_w in Fourier space, which correspond to higher k values. The quantities are then defined as follows:

$$\begin{aligned}
\delta_m &= W(k)\left[\left(\frac{\Omega_b}{\Omega_m}T_{\text{baryon}}\right) + \left(\frac{\Omega_m - \Omega_b}{\Omega_m}T_{\text{CDM}}\right)\right] k^2 \left(P_0(k)\right)^{1/2} \\
\delta_b &= W(k)\ T_{\text{baryon}} k^2 \left(P_0(k)\right)^{1/2} \\
\delta_c &= W(k)\ T_{\text{CDM}} k^2 \left(P_0(k)\right)^{1/2}
\end{aligned} \tag{20}$$

These quantities are then used to define the relevant power spectra. The mass power spectrum is somewhat simple to calculate:

$$P_m(k) = 2\pi^2 \Delta_\zeta^2(k) \frac{\delta_m^2}{k^3} \tag{21}$$

where $\Delta_\zeta^2(k) = 2.42 \times 10^{-9}$ is the primordial curvature perturbation power spectrum cite https://arxiv.org/pdf/0803.0586 [5].

The v_{bc} power spectrum, however, is less straightforward. First, the velocity divergences must be calculated as in eq. 14. This is impractical to do analytically; rather, in practice, the baryon velocity divergence is calculated using

$$\theta_b = \frac{H(z_{rec})}{c}(z_{rec} + 1)\frac{\delta_{b,1020} - \delta_{b,970}}{50} \tag{22}$$

with $z_{rec} = 1020$ the redshift at which the initial conditions are generated (and its value chosen as the redshift at which recombination occurs), and the 1020 and 970 subscripts denoting the redshift at which δ is evaluated. The CDM velocity divergence is calculated similarly.

---

[3] sometimes also chosen as 0.005Mpc^{-1} in other conventions

[4] how was this chosen?

[5] should this also be changed with cosmology, since it relies on WMAP data?

The initial velocity transfer function (what is this called, and what symbol should i use? and why is there no square?) $I_{\rm vbc}$ is then calculated according to the scalar form of eq. 15, appropriately scaling by the redshift:

$$I_{\rm vbc} = \frac{\theta_b - \theta_c}{ik(z_{rec}+1)}. \tag{23}$$

Finally, the v_{bc} power spectrum is calculated explicitly:

$$P_{\rm vbc} = 2\pi^2 \Delta^2_\zeta(k) \frac{I_{\rm vbc}}{k^3} \tag{24}$$

Now that the matter power spectrum and relative velocity power spectrum have been explicitly calculated, the true generation of initial conditions may begin. First, a Gaussian random field is initialized in configuration space. Then, its Fourier transform is taken – since the power spectrum correlates in Fourier space, it may only be applied in Fourier space. Initializing in configuration space before taking the Fourier transform enforces Hermitian symmetry, which is essential for maintaining real (as opposed to complex) overdensity fields. Further, the Gaussianity of the overdensities is preserved due to the fact that the Gaussian distribution is an eigenfunction of the Fourier transform.

This method, though, requires some normalization to be performed. The Fourier space variables are multiplied by a normalization constant $1/\sqrt{L_{\rm pixel}}$ and divided by their magnitude, and combed through [6] sheet-by-sheet to impose phase changes such that the complex random variables in the simulation box experience one full revolution of phase. After this is done in all three axes, the power spectra can at last be imposed, scaling the random field in Fourier space to ensure that the perturbations match the cosmological model. This scaling is done using the same random variable base for the scalar δ_m, as well as the v_{bc} along each of the three spatial dimensions separately.

Finally, the Fourier space variables are inverse Fourier transformed back into configuration space, and the three v_{bc} fields for each of the spatial dimensions are converted into a scalar field by taking the total magnitude of the three axis velocities. It is these configuration space boxes that are the final output of `get_IC_N`.

In order to implement variable cosmological parameters, `get_IC_N` was translated into Python. This decision was made for two reasons: firstly, the transfer functions used in the evolution of the power spectrum are output by CAMB, which is a code written mostly in Python. Secondly, the final implementation of variable cosmology in 21cmSPACE strives to use the Python-based `astropy` module's cosmology classes as an organizational tool for holding parameter values, as well as its object methods for performing standard cosmological calculations (such as calculating h at a given redshift).

Since the conventions used between `CAMB` and `astropy` are not identical, a `set_cosmology_camb` wrapper was written for `CAMB`'s `pars.set_cosmology` method. In particular, astropy's value for $\Omega_b, 0$ and $\Omega_c, 0$ differ from `CAMB`'s input by a factor of h^2; this wrapper multiplies by the relevant factors before passing the values to `CAMB`.[7]

To receive information about cosmological parameters, the function `get_IC_N` was altered to take cosmology as an input parameter. Specifically, in addition to the two inputs of `Npix` and `seed`, `get_IC_N` takes an `astropy.cosmology` instance such as `FlatLambdaCDM` or `LambdaCDM` as a third parameter.

---

[6] this is what is done in `get_IC_N.generate_k_values`,but what is it doing physically? and why is there a discontinuity in the middle?

[7] There is also some issue with the Tcmb0's datatype, but documentation is hard to find for this.

Practically, the cosmology enters `get_IC_N` in two ways. Most obviously, the transfer functions encoding effects such as inflation and horizon interaction are sensitive to the cosmological parameters. While these were previously precomputed and saved in the interest of computational efficiency, they must be recalculated for every cosmology, and therefore it is logistically sensible for the calling of CAMB to be directly incorporated as a subroutine within `get_IC_N`. The execution of CAMB, from instantiation using parameters to the output of the transfer functions for each of the three redshifts required, takes on the order of seconds on commercial hardware when using CAMB's default settings such as those governing the k range probed and the resolution of k sampling. While these settings can be changed to improve accuracy, which was originally done in the precomputed transfer functions, in practice this was found to make no difference in the final initial conditions[8].

Those familiar with CAMB may be aware that CAMB has the built-in functionality to output the power spectra directly as shown in section **??**, thus avoiding the need to manually scale and process the power spectrum from its primordial form. However, the manual scaling of the power spectrum was implemented as a design choice because the CAMB output power spectrum does not include the smoothing required at high k values cite?. As well as this, manual scaling offers a level of control over the primordial power spectrum that sets up future work.

As well as this, the calculation of velocity divergence requires the value of H at a specific redshift. This was originally a function `get_H_z` defined in another MATLAB file in the same directory, implementing the FRW equation

$$(H(z))^2 = H_0^2(\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_R(1+z)^4) \tag{25}$$

with the radiation pressure Ω_R hard-coded to 8.5522×10^{-5} where is this from?. To keep consistent with the input cosmology, this `get_H_z` method is now instead replaced with the `astropy.cosmology` instance's `H` method.

Since the new code is all written in the same language, a given random seed should cause all calculations to be deterministically performed. Therefore, any discrepancy in the initial condition grids must be due to a difference in the propagation of the cosmological parameters. This is illustrated in fig. 1, fig. 2, and fig. 3, which show a slice of initial condition boxes generated using the same random seed: the upper plots show that the initial conditions are, by eye, very similar (due to the deterministic nature of the random draws with a fixed seed). The lower plot, however, shows the difference in initial condition boxes, with deviations of up to 5%.

[8] show evidence?

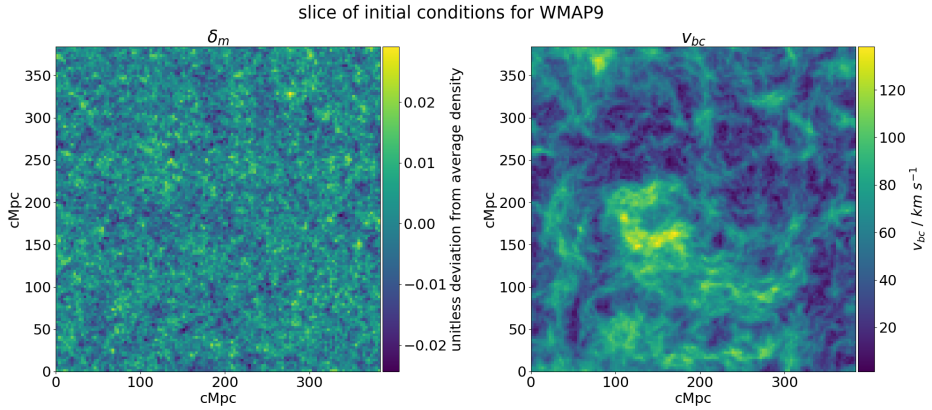

FIG. 1: CAPTION

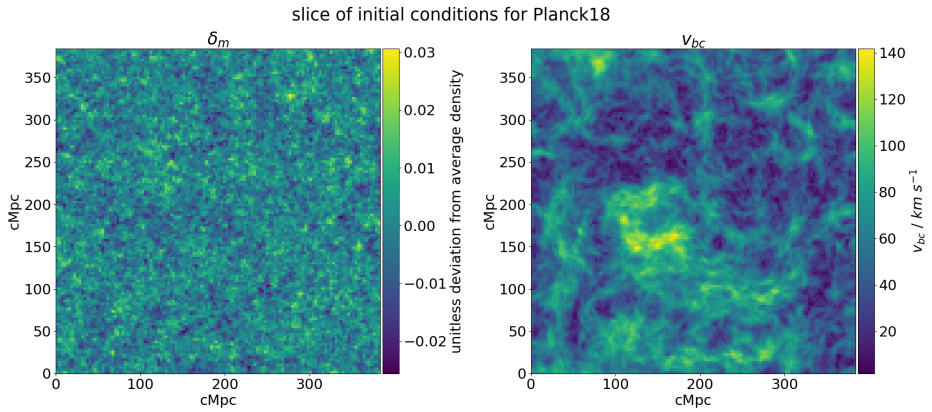

FIG. 2: CAPTION

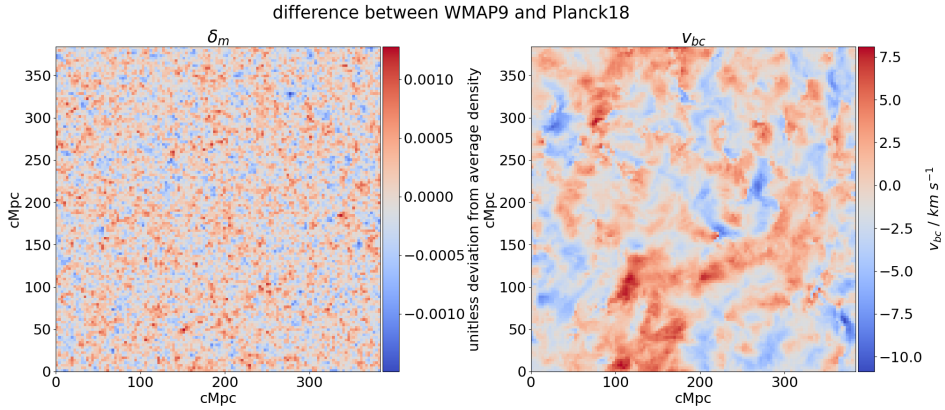

FIG. 3: CAPTION

C. Incorporation of Cosmology into the Halo Mass Function

D. Consistency Checks

Although the rewriting of code maintained the same mathematical algorithm, translating code from one language into another often introduces inconsistencies due to the nature of hardware interaction, or variations in implicitly-called library package subroutines. It is therefore important to perform consistency checks at every stage, in order to maintain confidence that the simulation is reflective of the underlying physics.

Due to the statistical nature of measurements, however, the stochacity in the initialization of the random Gaussian field propagates through the imposition of the power spectrum, imprinting inherent randomness onto the final result. To complicate matters further, the generation of random Gaussian fields is non-identical between MATLAB and Python, even when using the same random seed. Why? Saswata and Furen pointed this out in group meeting – usually it should be the same, but specifically in 3D or in random Gaussian field it is different. Where is the divergence?

It is therefore important to iterate the algorithm in both MATLAB and Python multiple times, and statistically compare the power spectra that these initial condition boxes produce. Specifically, after generating the initial condition boxes, the power spectrum can be computed independently from its generation. Taking each power spectrum from individual initial condition boxes, their average value at each k sample may be calculated as a best estimate of the power at that wavenumber; similarly, the standard deviation of the power value at each k sample may be used as an estimate of the uncertainty in the power, as a result of propogation of randomness from the initial Gaussian field. In principle, averaging an infinite number of iterations will return the original power spectrum which should perfectly agree with each other. However, in practice, each algorithm has an associated computational cost, limiting the number of samples; yet, as shown in fig. 4 and fig. 5, only as few as 10 runs is needed to clearly illustrate the agreement of the two code scripts with each other. The plots also include the fractional error, calculated by taking the absolute difference between the two methods, and then dividing by the legacy method's best estimate.

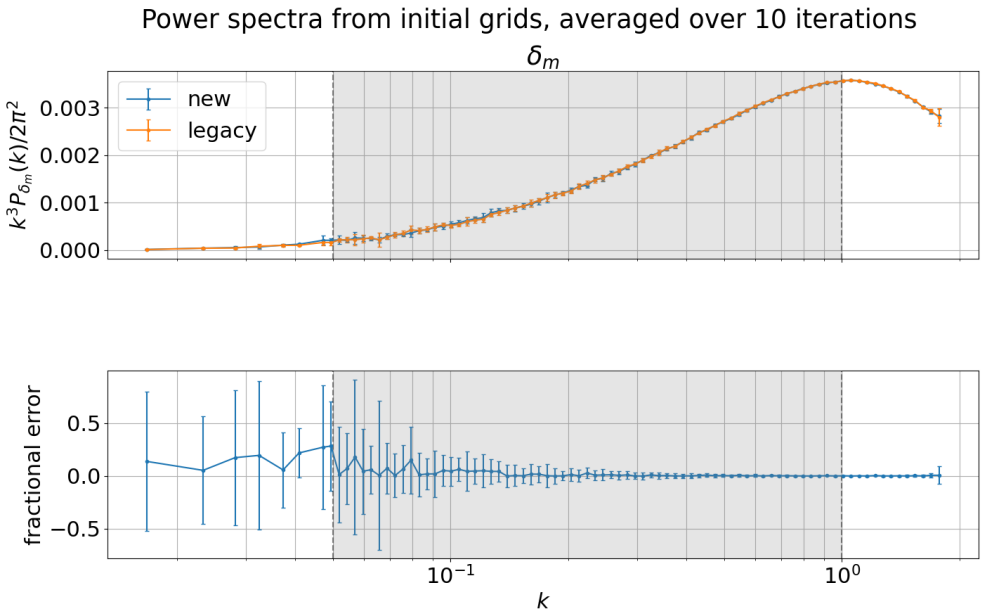

FIG. 4: The upper axes show a comparison of power spectra between legacy and new initial condition generation scripts, **before** the incorporation of CAMB. Best estimates are calculated using the average of power at each k sample over 10 iterations to smooth out effects from randomness in the initial Gaussian draws, and error bars are calculated by taking the standard deviation of the same quantity. The lower axes show the fractional error between the two scripts, treating the legacy script as truth.

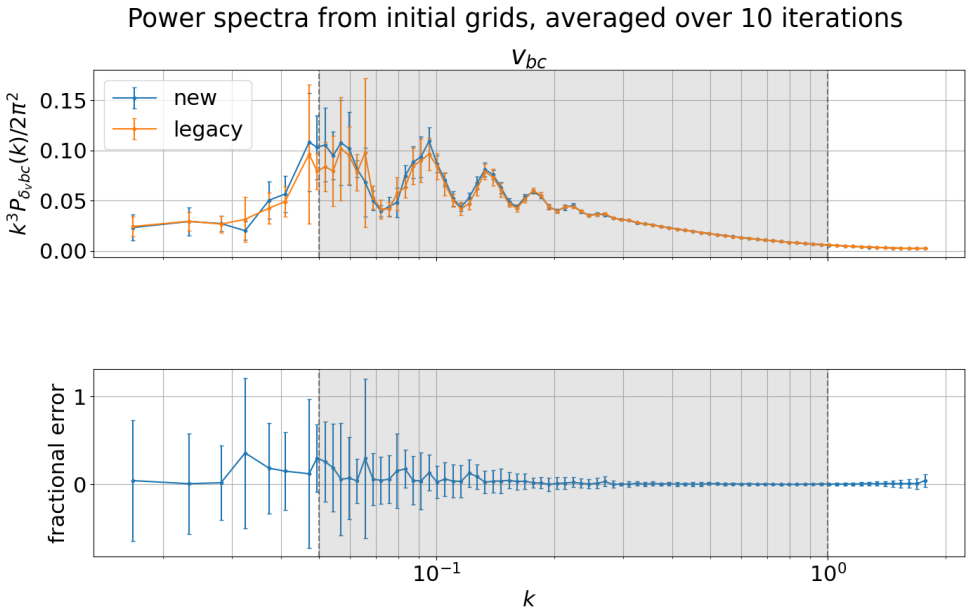

FIG. 5: CAPTION

As is illustrated the formatting needs to be cleared up; it is hard to see. in the fractional error plots, the deviation between the legacy and new scripts is negligible; furthermore, all error bars overlap with 0% error, signifying that the legacy and new scripts, despite written in different languages and having different subroutines for the initialisation of Gaussian random fields, perform the same calculations and, more importantly, preserve the physics behind the computation.

The incorporation of `CAMB` into each execution of `get_IC_N` is also a potential source of deviation. As `CAMB` is an independently maintained code package, it is possible that `CAMB` has undergone changes in its parameters, calculation, or post-processing over the decade since the original legacy precomputed transfer functions used in `get_IC_N` were written to file. Therefore, it is useful to check that `get_IC_N`'s output, when using the legacy code with hard-coded Planck 2013 parameters, is consistent with the new code when directly calling `CAMB`, also using Planck 2013 parameters. The comparisons are shown in fig. 6 and 7.

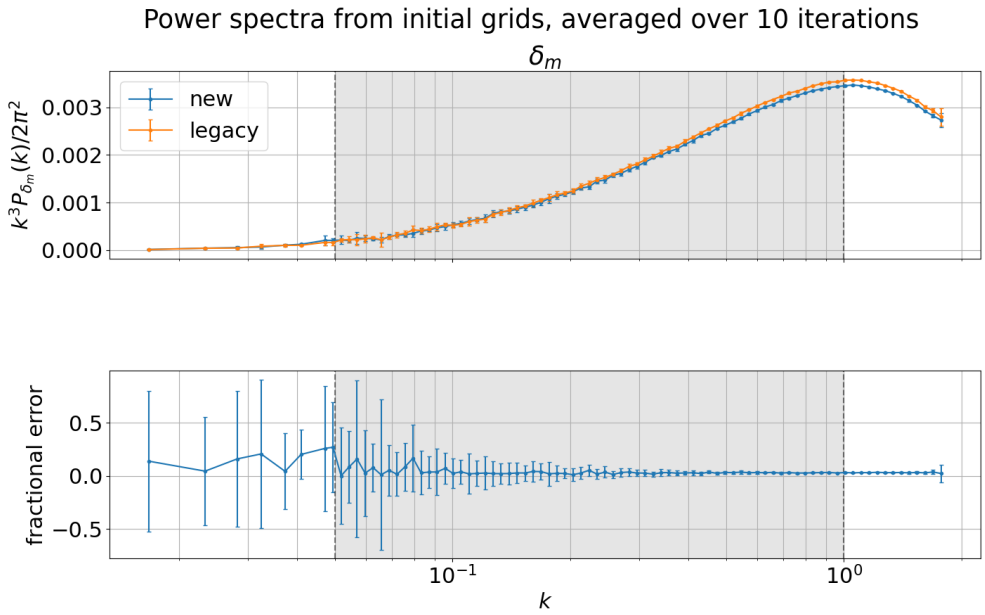

FIG. 6: The upper axes show a comparison of power spectra between legacy and new initial condition generation scripts, **after** the incorporation of CAMB. Best estimates are calculated using the average of power at each k sample over 10 iterations to smooth out effects from randomness in the initial Gaussian draws, and error bars are calculated by taking the standard deviation of the same quantity. The lower axes show the fractional error between the two scripts, treating the legacy script as truth.

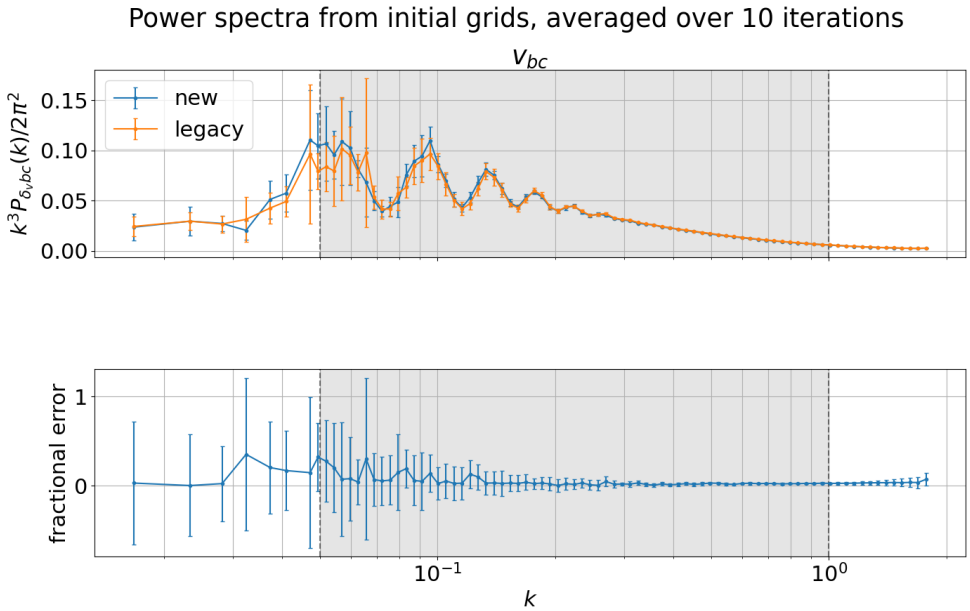

FIG. 7: CAPTION

The fractional error plot in v_{bc} shows all error bars overlapping with 0 error; therefore, the v_{bc} calculation when incorporating `CAMB` can be taken to be statistically identical. However, fractional error in the δ_m power spectrum shows an error of about 3%. This is perhaps unsurprising, since the legacy generation used parameters drawn from different sources, some updated over the years. As well as this, 3% error is negligible in the scheme of current instrumental errors in observational 21-cm astronomy. Together, these consistency checks are convincing enough to approve the use of the `CAMB`-incorporated Python version of `get_IC_N` into 21cmSPACE.

With the infrastructure in place, it is also important to check that different cosmologies indeed imprint differences in the power spectra. This is shown in fig. 8: in the range of wavenumbers simulated by 21cmSPACE, there is already measurable differences depending on the cosmology used.

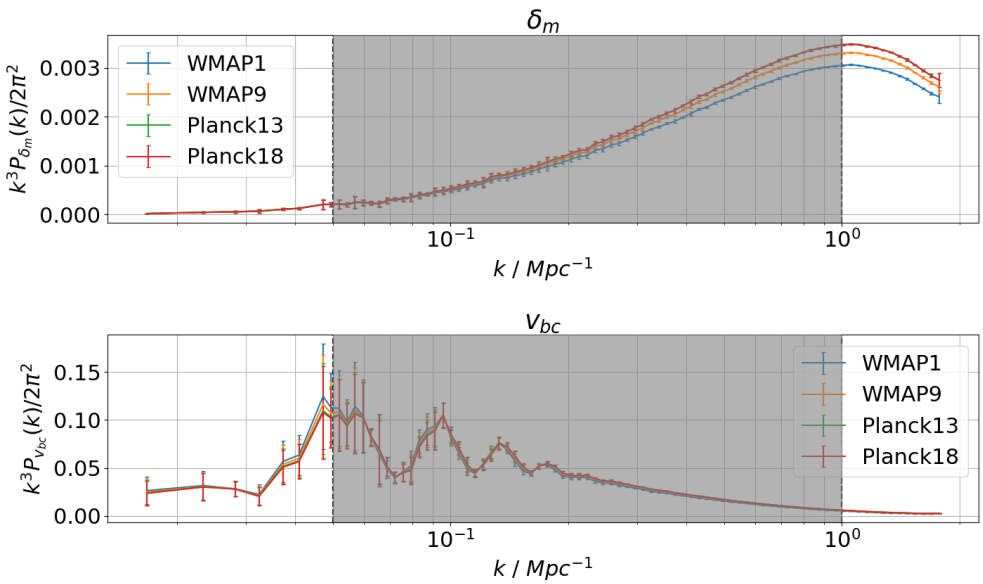

FIG. 8: CAPTION

V. RESULTS

A. Simulation Suite

So far, two (more to come) successful 21cmSPACE runs have been executed different initial conditions, each of which incorporates different cosmological parameters. These use cosmolgies included in `astropy`'s presets – specifically, the WMAP9 and Planck 2018 `FlatLambdaCDM` instances. Their parameter values are summarized in tab. III.

Parameter needs units	WMAP9	Planck 2018
h	0.6932	0.6766
$\Omega_{\mathrm{b},0}$	0.04628	0.04897
$\Omega_{\mathrm{dm},0}$	0.24022	026069
$\Omega_{\mathrm{k}}, 0$	0.0	0.0
$T_{\mathrm{CMB},0}$	2.725	2.7255

TABLE III: Included cosmology parameter values

B. 21cm Power Spectrum

Using these parameter values, 21cmSPACE was executed from beginning to end, and the results plotted. Fig. 9 shows the effect of the two different sets of cosmological parameters on the 21-cm power spectrum, with 10 and 11 the differences between the new cosmology and legacy code, before the incorporation of variable cosmology.

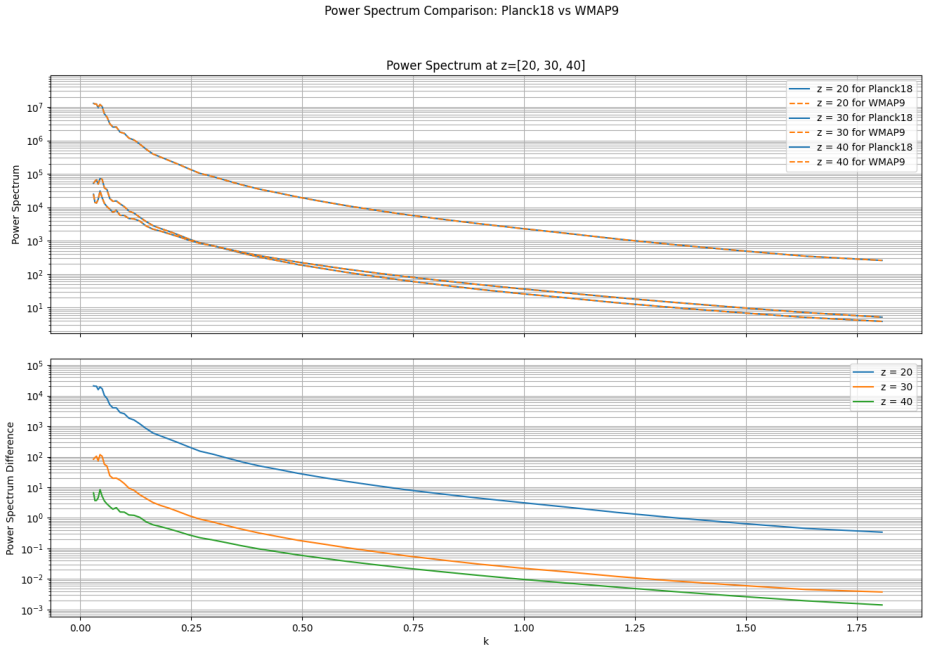

FIG. 9: CAPTION

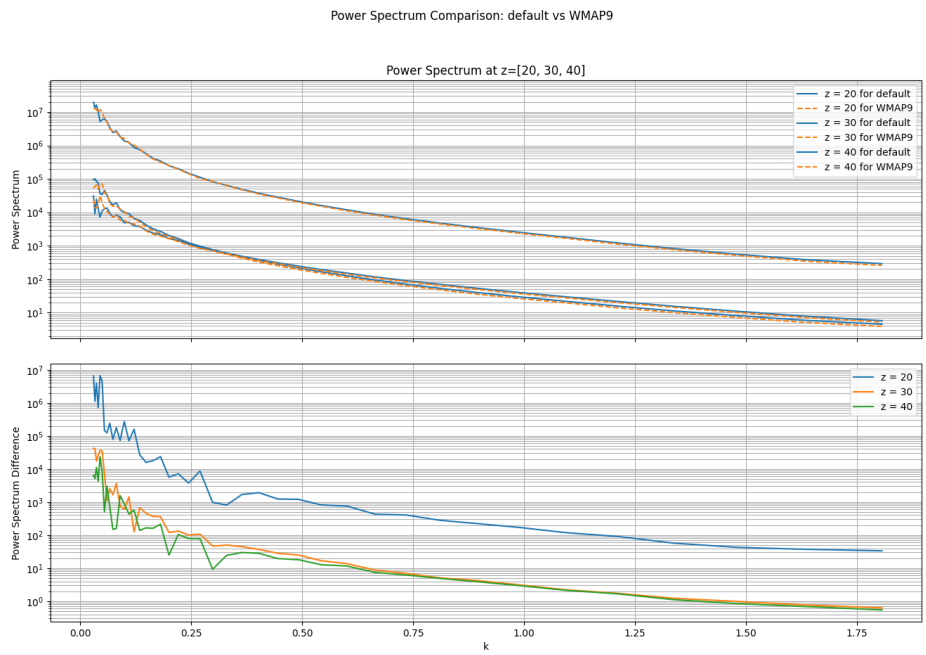

FIG. 10: CAPTION

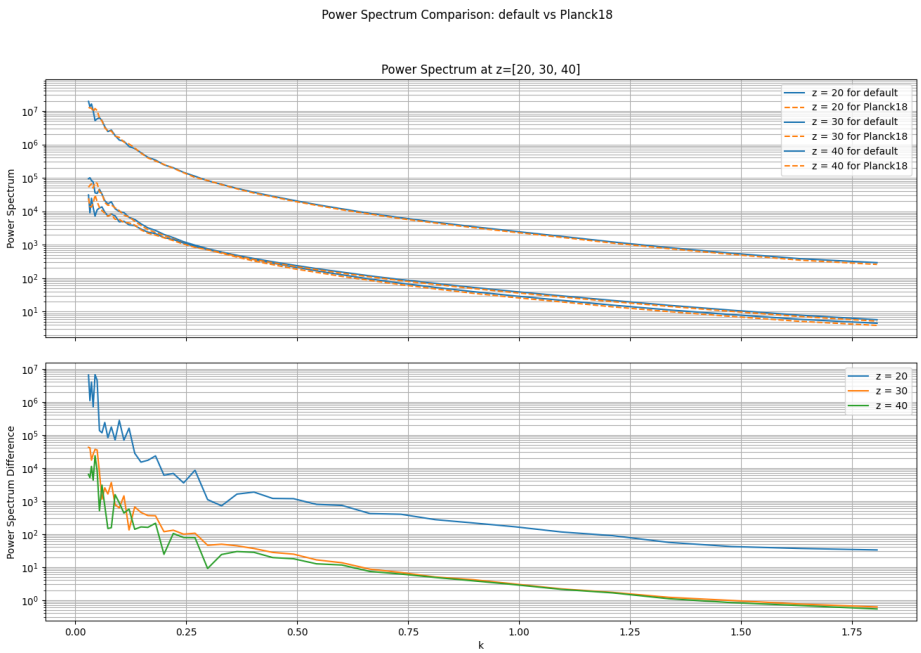

FIG. 11: CAPTION

C. Global 21cm Signal

Analogously to the 21-cm power spectrum, fig. 12 shows the effect of the two different sets of cosmological parameters on the 21-cm global signal, with 13 and 14 the differences between the new cosmology and legacy code, before the incorporation of variable cosmology.

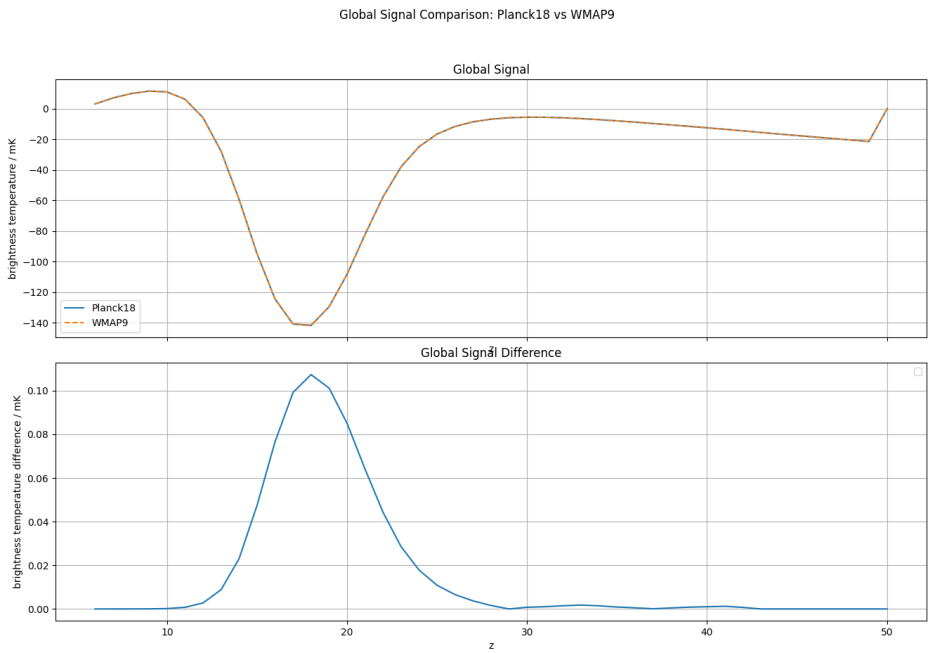

FIG. 12: CAPTION

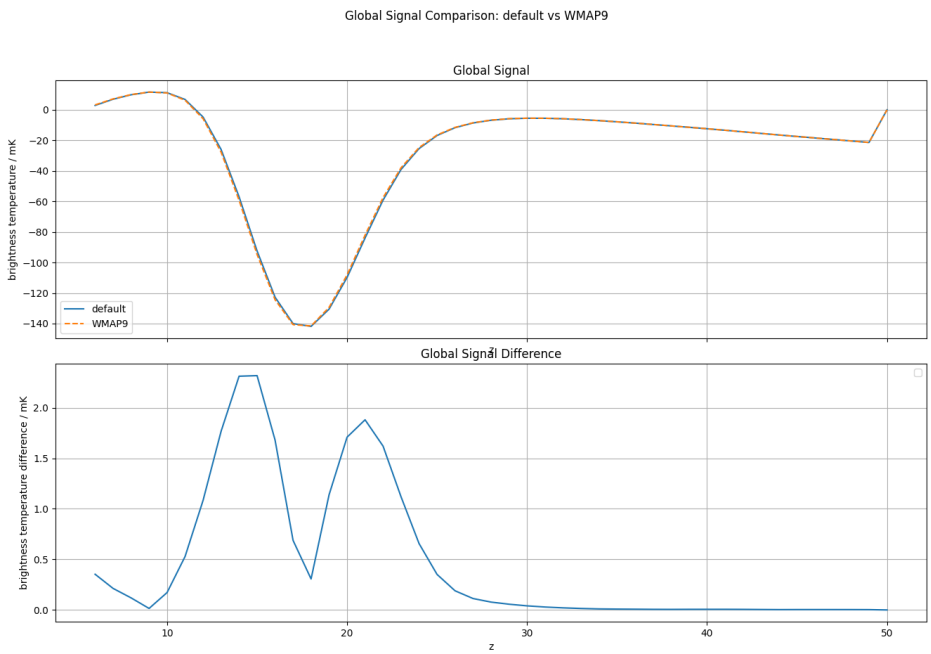

FIG. 13: CAPTION

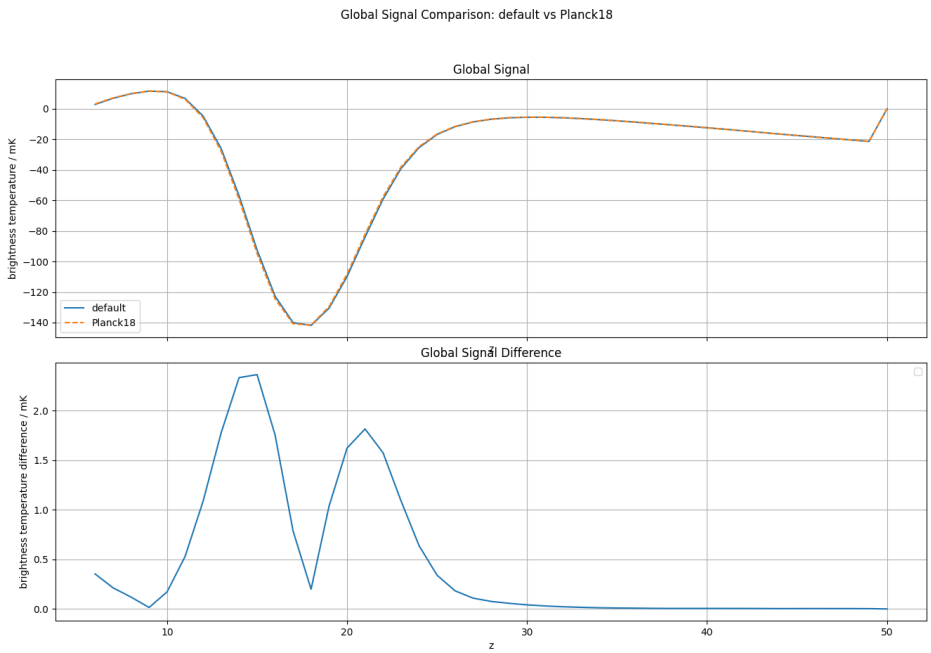

FIG. 14: CAPTION

VI. DISCUSSION

to be done after incorporating more results

A. Accuracy of Implementation

B. Comparison with Prior Work

C. Uncertainties and Error Bars

D. Relevance for Upcoming Experiments

VII. CONCLUSION AND FUTURE WORK

VIII. ACKNOWLEDGEMENTS

REFERENCES

[1] Jiten Dhandha. "21-cm cosmology and its complementarity with JWST observations". First Year PhD Report. University of Cambridge, 2022.

[2] Anastasia Fialkov, Rennan Barkana, and Eli Visbal. "The observable signature of late heating of the Universe during cosmic reionization". In: *Nature* 506.7487 (Feb. 2014), pp. 197–199. ISSN: 1476-4687. DOI: 10.1038/nature12999. URL: http://dx.doi.org/10.1038/nature12999.

[3] Anastasia Fialkov et al. "Impact of the relative motion between the dark matter and baryons on the first stars: semi-analytical modelling: Impact of relative motion on the first stars". In: *Monthly Notices of the Royal Astronomical Society* 424.2 (June 2012), pp. 1335–1345. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.21318.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2012.21318.x.

[4] Thomas Gessey-Jones. "Probing the First Stars with the 21-cm Signal: Theory, Methods, and Forecasts". PhD thesis. Apollo - University of Cambridge Repository, 2024. DOI: 10.17863/CAM.110892. URL: https://www.repository.cam.ac.uk/handle/1810/371844.

[5] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[6] Jonathan R Pritchard and Abraham Loeb. "21 cm cosmology in the 21st century". In: *Reports on Progress in Physics* 75.8 (July 2012), p. 086901. ISSN: 1361-6633. DOI: 10.1088/0034-4885/75/8/086901. URL: http://dx.doi.org/10.1088/0034-4885/75/8/086901.

[7] Eli Visbal et al. "The signature of the first stars in atomic hydrogen at redshift 20". In: *Nature (London)* 487.7405 (July 2012), pp. 70–73. DOI: 10.1038/nature11177. arXiv: 1201.1005 [astro-ph.CO].