Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation

Thomas A. J. Binnie
Astrophysics Group

Department of Physics
Imperial College London

Thesis submitted for the Degree of Doctor of Philosophy to
Imperial College London
· 2021 ·

Acknowledgements

Firstly I'd like to thank my supervisor Dr. Jonathan R. Pritchard. You have taught me everything I know about 21cm cosmology and my only published paper to date was your brain-child. Your support both academically and in a pastoral context is not something I will forget anytime soon. I look forward to seeing you when I haven't got thesis chapters to hand in and I hope one day you eventually have the time to sit down and listen to John Coltrane's 'A Love Supreme' in full, you deserve it.

I'd also like to thank Brad Greig, Steven Murray, Cath Trott, Alan Heavens, and Dave Clements for their collaborative support. This thesis would be a lot less rich if it wasn't for you.

A huge thanks to Adelie Gorce, Ian Hothi, Emma Chapman, Catherine Watkinson, Claude Schmit, and Suman Majumdar for valuable conversations in our research group meetings - you have all kept me on my toes when I've needed it. For generally making Imperial a nice place to be I'd like to thank the astrophysics group and ICIC - the atmosphere is friendly, there is always someone to ask questions, and I've learnt a lot.

Thanks to the outreach department for letting me use the inflatable planetarium, its been great fun. I'd also like to thank Roberto Trotta, I very much enjoyed being part of the dark matter experience.

Marija, Conor, Sanjay, and Selim your camaraderie during the last for years has been epic. I will be sure to create a 'bit' in every work place I go. Nick, Jonny, Sam, Ali, and Ali you've been there for me a lot and a large proportion of my fondest memories during the last 4 years are thanks to you guys. Thanks to everyone at Kay road, Global North Vibes Tribe, Freak Boutique, the IC Big Band and IC cross country and athletics club - I'm not sure you've sped up this process but you've certainly made it worth living. Alba, thank you for supporting me in the recent months. This list could go on forever, unfortunately not every hero can be sung for. If I've missed you out, there is probably a reference to you in the main text - let me know when you've found it ;)

Finally, the biggest thanks to Mum, Al and Dad. I hope I've been more support than annoyance since coming home. Finishing off this beast would have had me stressed at the best of times, let alone during a global pandemic. You know I'm not good with words but I mean it - thanks.

'There's a great beauty to having problems. That's one of the ways we learn.'

Herbie Hancock

Declaration and Copyright

Statement of Originality

Thomas Binnie, January 2021

Publications

• Paper I:

T. Binnie and J. R. Pritchard.

Bayesian Model Selection with Future 21cm Observations of The Epoch of Reionisation.

MNRAS, vol. 487, 1160-1177, July 2019.

Abstract

We are witnessing the evolution of cosmology into a precision science. When the universe cools after the big bang, hydrogen atoms eventually form releasing the cosmic microwave background. Measurements of this, combined with direct observations of fully formed galaxies through telescopes inform us about the seeding of structure across the cosmos. In-between these milestones there is little-to-no observational data, particularly regarding the first stars and their growth into mature galaxies. Primordial hydrogen must eventually progress into everything we see today, and we know from local observations of galaxies that the remaining hydrogen must also become ionised again. A new generation of radio telescopes aim to detect the cosmological 21cm line of these hydrogen atoms in order to observe these events directly. This will provide unmatched insight into the formation of the first stars and their progression into the early galaxies responsible for reionisation. As never seen before, these new instruments will provide tomographical maps of our universe that fill in a significant proportion of the unobserved universe. Scientists across the globe hope to witness the cosmic dawn within the next decade.

This work is focused on improving the statistical analyses techniques that probe the first star formation and when the early universe is being reionised - the cosmic dawn and epoch of reionisation respectively. The majority of this work looks at Bayesian model selection, which is a robust method for quantifying how well models fit observations. We apply it within the context of high redshift 21cm power spectrum observations, comparing a variety of scale and morphological implementations of the reionisation process. We also use it to look into distinguishing effects caused by the remnants of the first stars and the statistical benefits of including UV observations. We then turn to improving the likelihood statistic, required for Bayesian analyses. The Morlet transform is implemented as an alternative to the Fourier transform within the power spectrum to enable the whole observational light-cone to be analysed at once. Bayesian analyses are incredibly promising and with this work we are a step closer to statistical machinery that provides objective conclusions about the quality of models when presented with data. Synergy between observational methods is important to maximise the discerning power of any Bayesian method. Now that observations are becoming more precise within cosmology, we can begin to constrain primordial astrophysics.

Acronyms

EoR - Epoch of Reionisation

CMB - Cosmic Microwave Background

LF - Luminosity Function

FZH - Furlanetto, zaldarriaga, Hernquist

MHR - Miralde-Escude, Haenelt, Rees

IGM - Inter Galactic Medium

ISM - Inter Stellar Medium

GR - General Relativity

SED - Spectral Energy Distribution

$\Lambda - CDM$ - A Dark Energy and Cold Dark Matter dominated Cosmological model

rms - Root Mean Square

MCMC - Markov Chain Monté Carlo

pdf - Probability Density Function

CLT - Central Limit Theorem

MAP - Maxima a posteri

SDDR - Savage-Dickey Density Ratio

LCE - Light-cone Effect

FPS - Fourier Power Spectrum

MPS - Morlet Power Spectrum

$\langle x \rangle$ - spatial average of x

$\hat{x}$ - unit vector of x

$\tilde{x}$ - Fourier Transform of x

Cartesian coordinates (x, y)

Polar coordinates (r, θ)

Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Introduction

This chapter introduces the specific context necessary for applying Bayesian statistical analyses in the EoR. We begin with the cosmological influences on the 21cm signal in Section 1.2. This includes an introduction to the cosmic dawn and epoch of reionisation, including an overview of where they fit into the cosmological picture. Section 1.3 contains the astrophysical recipes for simulating observations of the 21cm brightness temperature and the galaxy ultra-violet luminosity functions that we use in later chapters. Simulation of a noise power spectrum for a given radio telescope when observing a specific mock 21cm power spectrum is also detailed here. We then turn to the fundamentals of Bayesian analyses in Section 1.4, describing the role of the likelihood statistic and the importance of prior considerations in the context of model selection. Finally, detailed descriptions of the MCMC algorithms implemented throughout this work are provided in Section 1.5. A summary of these topics is provided in Section 1.6. Introductory material for astrophysics, radio observations and statistical analyses are provided in Appendices A, B, and C respectively.

1.2 21cm application in cosmology

The field of cosmology has a reasonable agreement between observation and theory. In-between the CMB ($z = 1100$) and direct observations of QSOs at late times ($z < 7$) however, there is a

void of data. After recombination, the first stars must form and a reionisation of intergalactic hydrogen must occur. The formation of the first stars and galaxies is known as the cosmic dawn. Before QSOs can be observed in plenty, the phase of our universe changes from being predominantly neutral to ionised during what's defined as the epoch of reionisation or EoR. The cosmological 21cm differential brightness temperature is our best hope for filling this gap with observation. Here we discuss the physics underlying the cosmic dawn and EoR, which the cosmological 21cm signal will probe in the near future.

1.2.1 The cosmic dawn

The cosmic dawn occurs when the universe was only a few hundred million years old. These stars formed from gas composed mainly of molecular and atomic hydrogen. Large, short lived (so called Population III) stars are theorised to form because of this. Eventually they die in supernovae, enriching the surrounding gas within the dark matter halos in which they are embedded. Stellar processes are how the universe gains its complexity. Through various stellar feedback mechanisms the metallicity of the universe increases, leading to the rich variety of structure on all scales. The detail of how the first stars form is therefore important on many levels - it is even relevant in astro-biology, where the presence of specific elements enables the possibility for life to develop.

In our current construction, after recombination we have a roughly uniform distribution of hydrogen atoms and molecules (in ratio of approximately 1000:1). The universe can be described by the over and under densities of linear perturbation theory as well as the dark matter halo distribution from the P-S formalism as discussed in Section A.1.

Baryonic gas accretes onto the dark matter halo, heating the gas to a virial temperature. To quantify this we consider what are known as the Jeans mass, M_J, and the Toomre criteria, Q (equations 1.1 and 1.2). The Jeans mass is the criteria for a stationary cloud of gas (of density n and temperature T) to collapse and begin forming a proto-star,

$$M_J \approx 700 \left(\frac{T}{200\mathrm{K}}\right)^{\frac{3}{2}} \left(\frac{n}{10^4 \mathrm{cm}^{-3}}\right)^{-\frac{1}{2}} M_\odot. \tag{1.1}$$

A proto-star is young star that is growing star via accretion, prior to nucleosynthesis. For the

surrounding accretion disc that feeds the protostar we have the Toomre criteria,

$$Q \equiv \frac{c_s \kappa}{\pi G \Sigma} \leq 1, \tag{1.2}$$

describes the stability of a rotating gaseous accretion disc with sound speed, c_s, epicycle frequency, κ, and surface density, Σ. If the criteria is broken, the disc will fragment into multiple proto-stars leading to potential binary systems or clusters of stars.

To get from proto-stars to stars we need to consider the radiative feedback mechanisms that occur between the proto-stellar radiation and the surrounding gas. The accretion disk can be heated by the radiation produced by the proto-star, causing gas to evaporate from it. If the star is massive enough ($\sim 150 M_\odot$) it will be bright enough to remove the whole disc, however this mass is very close to the threshold required to collapse into a black hole - the precision of the first generation of stellar mass functions are vital. Multi-body dynamics combined with these sorts of subtle thresholds, quickly lead to complexity in the early universe.

The first generation of stars are known as population (Pop)-III and are defined by containing only hydrogen. This category can be subdivided into Pop-III.1 and Pop-III.2 by their formation process. Pop-III.1 are formed initially with mostly atomic hydrogen but crucially including enough molecules to benefit from molecular cooling. Hence a larger range of gas cloud masses are available since the gas can be cooled, allowing stars to form more efficiently ($m \sim [1, 1000]$ $M_\odot$, see Figure 1.1). Pop-III.2, on the other hand, are formed out of predominantly HII regions where the molecular cooling is negligible and there is limited atomic cooling, hence a much smaller range of gas cloud masses is available ($m \sim [10, 50]$ $M_\odot$). Other than the initial mass functions of these two categories, their properties are the same.

After the first stars, the cooling mechanisms available to next generation of stars is changed by the presence of ionising photons. The photo-dissociation of H_2 molecules will outweigh the ionisation of HI atoms exponentially, due to the drop in black body radiation intensity from emission at 4.5 eV (for H_2) to 13.6 eV (for HI). Hence the cooling mechanisms available to the first group of star formation will be different to the stars forming shortly afterwards (See Figure 1.1). The first generation of stars can be a result of cooling via the vibrational and rotational lines of the H_2 molecules (and the HI atoms), while the second generation gas will only have been cooled by atoms. Pop-III.1 stars born in the first generation will therefore have a larger

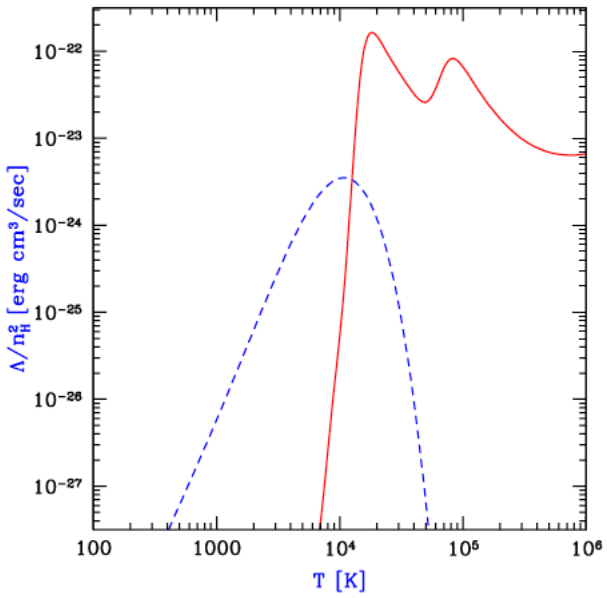


Figure 1.1: Cooling rate as a function of Temperature for atomic (red solid line) and molecular (blue dashed line) cooling. Peaks in the atomic cooling curve are characteristic to the collisional excitation of hydrogen and helium atoms. The temperature with which a proto-star can cool to influences the maximum mass a star can have. For reference, $n_{\rm HI} = 0.045{\rm cm}^{-3}$ and $n_{\rm H_2} \approx n_{\rm HI}/1000$ at $z = 10$. (Barkana & Loeb, 2001)

range of masses than the Pop-III.2 stars formed in the second generation.

The end state of Pop-III stars is particularly sensitive to their mass. Stellar remnants (and the processes by which they form) impact their surroundings and therefore the future generations of stars. Below we summarise the mass dependent end states of Pop-III stars (Loeb & Furlanetto, 2013):

- $m < 10$ M$_\odot$ - a white dwarf will form, any metallicity changes of the surroundings are well beyond the formation timescale of galaxies and are therefore unimportant.
- 10 M$_\odot$ $< m <$ 25 M$_\odot$ - Type II (core collapse) Supernova occur, leaving behind a neutron star. This provides a large contribution of surrounding metallicity.
- 25 M$_\odot$ $< m <$ 40 M$_\odot$ - Type II Supernova but resulting in a black hole, which the heavy elements will fall into meaning they will not be recycled into the next generation of star formation.
- 40 M$_\odot$ $< m <$ 100 M$_\odot$ - The star will collapse into a black hole without a supernova, this will have no direct impact on the surrounding metallicity.

- 100 $M_\odot < m <$ 140 $M_\odot$ - After helium burning, the star will become unstable due to pair production from the thermal pressure energy. This is not enough for the star to explode, so it will pulsate ejecting the outer layers until it proceeds as a lower mass star. Since only the outer layers are ejected it is mainly hydrogen and not metals that contribute to the surroundings.

- 140 $M_\odot < m <$ 260 $M_\odot$ - After helium burning, a pair-instability supernova will occur as the instability from the thermal pressure energy is enough to explode the star. This has never been observed but would leave a very specific metallicity signature - only elements with even atomic numbers and no elements heavier than Zinc (Heger & Woosley, 2002).

- $m >$ 260 $M_\odot$ - The star will collapse straight into a black hole (possibly with an accretion disc), meaning no metallicity input into the surroundings is performed.

The cosmic dawn is thought to occur between redshifts $\sim$ 18 and 35. Intricacies such as the mass of the first stars will dictate when precisely this occurs and have an influence on how reionisation progresses.

1.2.2 The epoch of reionisation

The reionisation of inter-galactic hydrogen likely progresses slowly at first with small faint galaxies. It is likely that it then speeds up, as ionised regions percolate, with the bulk of reionisation occurring between $z \sim 6 - 10$. By the end, a combination of large bright galaxies (that form slowly through mergers or accreting gas from the IGM) and multiple small faint galaxies will together ionise hydrogen in the IGM.

As phrased in Loeb & Furlanetto (2013), 'How was the primordial gas transformed to an ionised state by the first galaxies within merely hundreds of millions of years?' Galaxies are the fundamental objects that drive reionisation. To gain some insight into the above question, let us consider the simplest possible case: a single isolated galaxy that ionises its surroundings. By treating the ionisation front as a sharp cut off and assuming each photon ionises one hydrogen atom we can write down,

$$\langle x_{\rm HI} \rangle \, V = Q_i, \tag{1.3}$$

where $\langle x_{\rm HI} \rangle$ is the mean neutral hydrogen number density, V is the volume of the ionised sphere (of radius r) and Q_i is the total number of ionizing photons produced by the source.

The size of this region depends on several things, the total dark matter halo mas $M_{\rm h}$, the baryonic fraction of the structure $\Omega_b/\Omega_{\rm M}$; we also need to define an efficiency of the star formation $f_\star$, the fraction of ionizing photons that are actually released into the IGM, $f_{\rm esc}$, and $N_{\rm ion}$ is the number of ionising photons per stellar baryon. We are now in a position to consider an ionising efficiency, ζ, which combines these quantities as,

$$\zeta = A_{\rm He} f_\star f_{\rm esc} N_{\rm ion}. \tag{1.4}$$

Since the average ionising photon is also energetic enough to ionise helium atoms (which are non negligible in absorbing ionising radiation) we must attenuate our ionizing efficiency by a factor $A_{\rm He} \sim 1.22$. Typical fiducial estimates for the escape fraction of quasars are in the range $f_{\rm esc} \sim [0.08, 0.10]$ however it is still under debate. A recent observation as shown can be as high as $f_{\rm esc} \sim 0.21$ (Borthakur et al., 2014).

Equation 1.4 can be used to calculate a maximum comoving radius of the ionised region however we have not yet considered recombinations. Defining the recombination rate, α, allows us to write down the Strömgren sphere as,

$$\frac{dQ_i}{dt} = \alpha_{\rm rec} \langle x_{\rm HI} \rangle^2 V. \tag{1.5}$$

This is the steady state volume of an ionizing source embedded in a recombining gas (recombining at rate $\alpha_{\rm rec}$).

By considering a variable ionising source we can model the expansion of our HII region as,

$$\frac{dQ_f}{dt} \frac{1}{4\pi r^2} = \langle x_{\rm HI} \rangle^2 u_{\rm I}, \tag{1.6}$$

where the source produces an ionisation front from its peculiar velocity $u_{\rm I}$, and Q_f is the number of ionizing photons that reach the ionisation front. Combining these concepts together with the expansion of the universe leads to Equation 1.7,

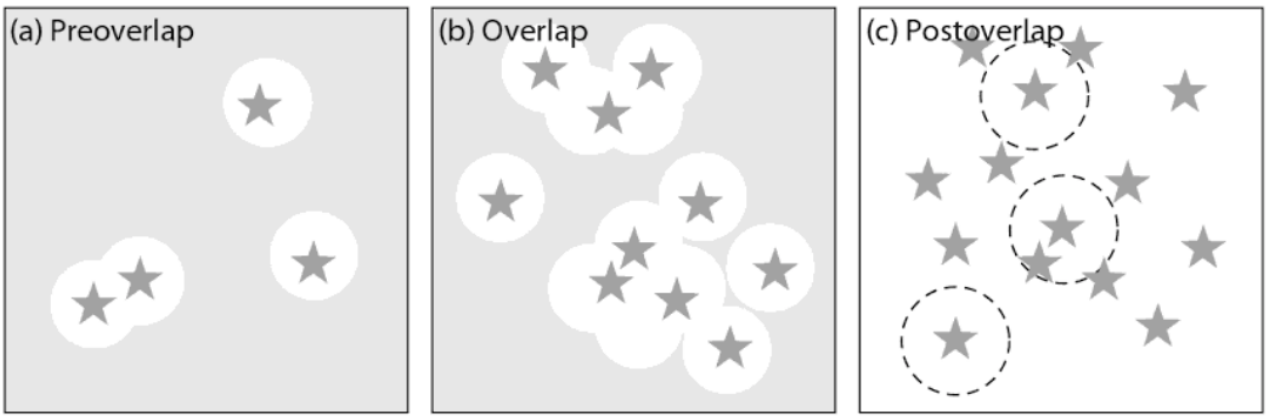


Figure 1.2: A toy model of the three phases of reionisation mentioned in the text. (a) *Preoverlap:* prior to galaxies, isolated HII regions grow slowly. (b) *Overlap:* galaxies form, bubbles begin to intersect rapidly. (c) *Postoverlap:* reionisation in the IGM is nearly over, photons can stream across intersected HII regions unless they are absorbed by sparsely populated dense HI clouds. The dashed lines remind us that the photon mean free path is finite. (Loeb & Furlanetto, 2013).

$$\frac{dQ_i}{dt} - \alpha \langle x_{\rm HI} x_e \rangle V = \langle x_{\rm HI} \rangle \left(\frac{dV}{dt} - 3HV \right). \tag{1.7}$$

Solving this equation gives an idea of the proportions of ionized hydrogen atoms surrounding ionizing sources. This very crude model can be taken further by producing a similar version of Equation 1.7 but averaged over the entire universe and by considering a clumping factor ($\langle x_{\rm HI}^2 \rangle / \langle x_{\rm HI} \rangle^2$). The clumping factor is necessary due to the non-uniformity of the IGM and is typically done numerically on full radiative transfer simulations.

In practice galaxies are only initially isolated like the example above. This leads to phases of overlap that are illustrated in Figure 1.2. Once overlap of the ionised regions has occurred the ionisation fronts propagate more easily and cumulate, accelerating the process once HII regions overlap. Despite this, dense pockets of HI will exist in-homogeneously throughout the largest ionised regions absorbing ionizing radiation in the post-overlap phase. Since this is entirely dependent on the clustering of structure formation this leads us to two possible reionisation models:

- *Inside-out* - where high density regions ionise first, because the galaxy population traces the density fields.
- *Outside-in* where low density regions ionise first, due to the high recombination rate of dense patches of hydrogen.

Clearly the three phases are not unique and are hard to distinguish at the cross overs. A mixture of both inside-out and outside-in reionisation will occur depending on the statistical distribution of matter throughout the universe. The 'patchiness' of reionisation, or morphology, therefore depends on the distributions of galaxies.

Reionisation is thought to progress slowly at first, starting around $z \sim 18$. Observations suggest that HI in the IGM is ionised by $z \sim 6$, and the bulk of it thought to occur after $z \sim 10$, however this is yet to be confirmed with data. Models of reionisation can also vary in the mechanisms of radiation production in galaxies, causing variation in the ionising efficiency and photon mean free path. The UV photons will typically depend on the stars, therefore their spectra will be distributed as a black body however this is not the only influence at play. Exotic reionisation models refer to including the radiation from quasars of various sizes, dark matter particles annihilating to X-rays, high mass X-ray binaries, and primordial black hole effects. The non-thermal ionizing spectra produced by the accretion onto the compact object will not lead to a smooth ionisation front, potentially altering the reionisation morphology. There is also helium reionisation to be considered. Here we have only scratched the surface of reionisation galaxies, please see e.g. (Loeb & Furlanetto, 2013; Dayal et al., 2014; Ellis, 2014), or Wise (2019) for more details.

1.2.3 The spin-flip background

With the physics of the 21-cm radiation and the astrophysical context we wish to probe in place, we now consider what actually drives these mechanisms and how they can be interpreted.

After $z < 70$, the universe becomes too sparse for collisional coupling to maintain thermal equilibrium between the spin temperature of HI and the CMB ($x_k \ll 1$). In order to prevent the coupling of the spin temperature and the CMB, we need another mechanism - the Wouthuysen-Field effect (Wouthuysen, 1952; Field, 1959). This is the resonant scattering of Ly-α photons that has the profound effect of driving $T_\alpha \approx T_k$. This is a subtle process which must be broken down into its steps to be understood. Firstly, we must pay attention to the electric dipole selection rules from quantum mechanics: $\Delta F = 0, 1$ and $F = 0 \not\to F = 0$ (see Figure 1.3). In the context of Ly-α photons, the hyper-fine splitting of the 1s and 2P states are bound by these selection rules. When the electron absorbs and re-emits the photon this will result in a mixing

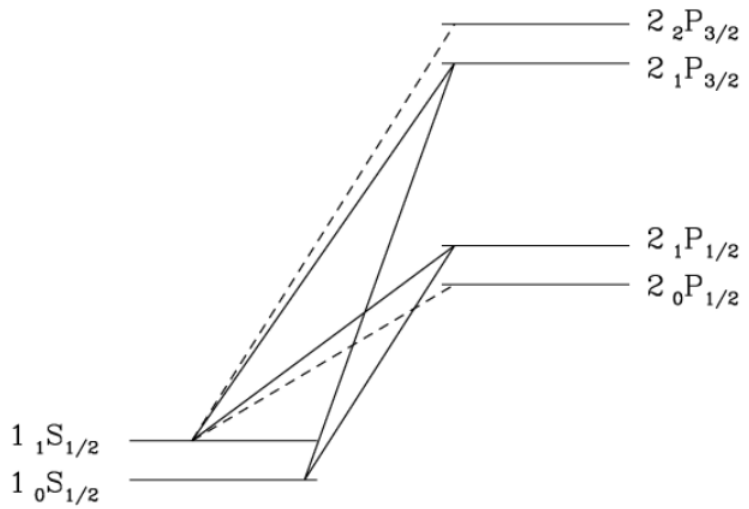


Figure 1.3: The hyper-fine splitting of the hydrogen 1s and 2p states. Solid lines represent the transitions that mix the 21cm excited and ground states. Dashed lines are the quantum mechanically allowed transitions that do not contribute to the mixing (notation of the energy levels is the form, $n\ {}_F L_J$). The relationship between these transitions leads to the WF effect. (Pritchard & Furlanetto (2006), Figure 1)

of the hyper-fine splittings of the ground states: $F = 0$ and $F = 1$, i.e. the mechanism increases the proportion of excited states of the 21cm emission. During this absorption and re-emission, in order to conserve momentum the kinetic energy of the hydrogen must decrease. This drives the colour temperature to be in equilibrium with the gas temperature $T_\alpha \approx T_k$, provided that the medium is optically thick. More simply, the scattering of the background field cools the spin temperature in order to find the equilibrium as,

$$\frac{\Delta E_{\text{recoil}}}{E} = \frac{h\nu}{m_p c^2}\left(1 - \frac{T_k}{T_\alpha}\right). \tag{1.8}$$

The full calculation of this coupling contributes to justifying the form of the UV coupling in Equation A.33, and therefore implies the form of the UV coupling coefficient x_α, used in the spin temperature calculation (Equation A.36).

Figure 1.4 shows the evolution of the different temperature scales that drive the 21cm brightness temperature. The CMB temperature (blue line) decreases with $(1+z)$ and is the reference point, which we will now discuss in detail. Since our observations will be related to the CMB black body radiation, the timing of thermodynamic events involving the HI gas become important. While the HI gas is colder than the CMB, CMB radiation will be absorbed; while HI is hotter, stimulated emission will boost the CMB intensity in comparison to the black body distribution.

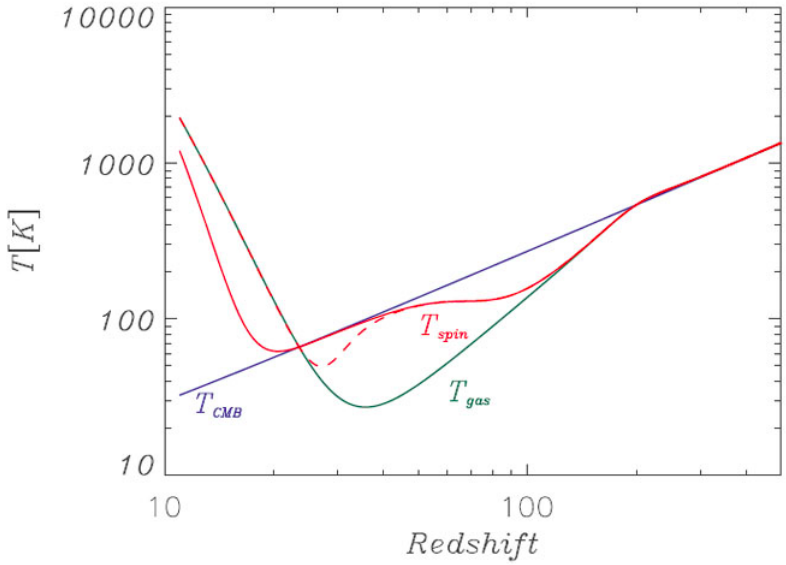


Figure 1.4: Evolution of the different temperature scales relevant to the cosmic dawn and EoR from redshifts 500 to 10, see text for more detail. $T_{gas} \equiv T_k, T_{spin} \equiv T_{\rm S}$ (Zaroubi, 2013).

At $z \sim 200$ the gas (green line) begins to cool adiabatically, decoupling from the CMB, until the first objects begin to form. The first objects will result in the heating the gas around $z \sim 30$ (Loeb & Zaldarriaga, 2004). Due to equilibrium thermodynamics, the spin temperature must be in between the gas and CMB temperatures at all times. Collisional coupling keeps the spin temperature in equilibrium with the gas during the initial adiabatic cooling. As the collisional coupling coefficient drops as the universe expands, the spin temperature begins to return to equilibrium with the CMB. Eventually the spin temperature will return to equilibrium with the gas temperature, but the timing of this is crucial. Whether this occurs before or after the gas temperature rises above the CMB temperature, will lead to very different patterns in the globally averaged 21cm brightness temperature. The red solid line shows the 'after' case, which will lead to only an emission 21cm signal. On the other hand the dotted red line shows re-coupling 'before' the emission phase, and hence an absorption signal (before the emission signal) will be seen somewhere in $z \sim [10, 30]$ (Pritchard & Loeb, 2008; Baek et al., 2010; Thomas & Zaroubi, 2011).

The global signal of the 21cm brightness temperature contains a vast amount of information about cosmic history (see figure 1.5). The regimes that are expected to be particularly important are summarised below (Furlanetto, 2006; Pritchard & Loeb, 2012). Events with unknown redshifts are stated in brackets next to the respective symbols.

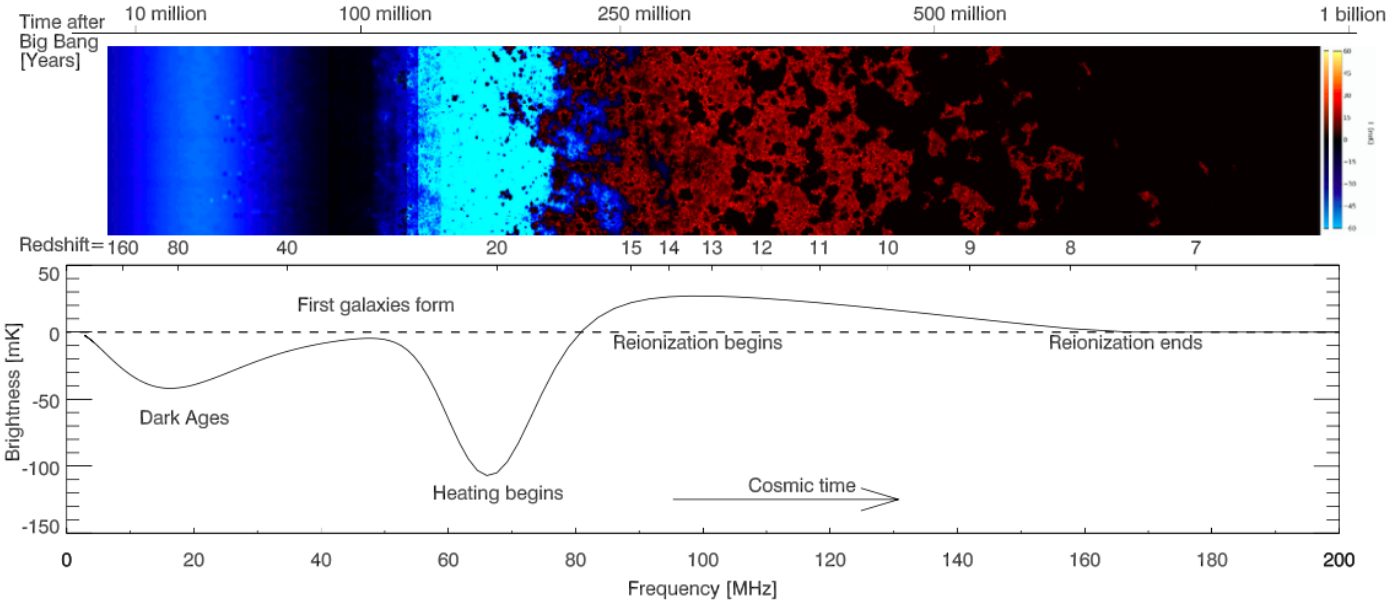


Figure 1.5: A possible brightness temperature history. (Top) This image is created from redshift slices of a simulated box, as it evolves with time. Red (blue) indicates 21cm emission (absorption). (Bottom) The dashed line is $\delta T_b = 0$ and the solid line is the sky averaged δT_b, corresponding to the specific history of the universe above. The SKA is sensitive to the redshift range $z \sim [7, 30]$, and will hopefully see from the formation of the first galaxies to the end of reionisation. (Pritchard & Loeb, 2012)

- $1100 > z > 200$ - After recombination, the high density of the universe drives an equilibrium of the CMB, spin temperature and gas temperature. $\delta T_b = 0$, therefore no signal.

- $200 > z > 40$ - As the universe expands, concentrations of particles decrease. Compton scattering decreases enough for the gas to adiabatically cool $\sim (1 + z)^2$. Collisional coupling keeps the spin temperature in equilibrium with the gas, an absorption signal is received.

- $40 > z > z_\star$ (when the first stars form) - Continued expansion of the universe leads to collisional coupling being ineffective. The spin temperature couples to the CMB, the absorption signal decreases.

- $z_\star > z > z_\alpha$ (UV coupling saturates) - First stars emit X-ray and UV radiation, heating the gas. The WF coupling drives $T_k \approx T_\alpha$, how quickly this happens defines whether the signal is absorption of emission at this stage (the example in figure 1.5 is absorption).

- $z_\alpha > z > z_h$ (the moment when $T_k > T_{\rm CMB}$) - Brightness temperature fluctuations are characterised by the gas temperature, which is being heated by UV, X-rays and possible

exotic sources. The brightness temperature is still negative (meaning absorption), and z_h is defined when the brightness temperature crosses $\delta T_b = 0$.

- $z_h > z > z_T$ (Fluctuations in the T_k become unimportant) - 21cm signal is seen in emission for the first time. The brightness temperature is driven by the fluctuations of the neutral fraction, baryon density perturbation and gas temperature.

- $z_T > z > z_r$ (reionisation ends) - Heating continues, driving the gas temperature significantly higher than the CMB. This implies $T_{\rm S} \approx T_k \gg T_{\rm CMB}$, meaning the brightness temperatures dependence on the spin temperature is removed. The signal is dominated by fluctuations in the neutral fraction.

- $z_r > z$ - Reionisation is complete, any HI left is negligible or in isolated small ($m < M_J$) clouds.

Current experiments aim to observe the 21cm signal for $z \sim [6, 12]$, where the limitations are mainly due to background noise (human and astrophysical) and the ionosphere, which becomes opaque to frequencies $\nu \leq 30$ MHz. This redshift range will be particularly insightful into the mechanisms driving reionisation and high redshift astrophysics. The observational intricacies are introduced at the end of the next section and discussed further in Appendix B.

1.3 Astrophysical simulation

Here we detail the recipes that are used in all the chapters for simulating the 21cm brightness temperature. Simulations are performed with a wide variety of techniques. Those providing the best detail involve hydrodynamic, N-body, and radiative transfer simulations combined, however they come at great computational cost (Zahn et al., 2011). Some examples of these comprehensive codes include C2RAY (Mellema et al., 2006), ATON (Aubert & Teyssier, 2008), EMMA (Aubert et al., 2015), and many more (Wise et al., 2012b,a, 2014; Feng et al., 2016; Finlator et al., 2018) or Ocvirk et al. (2016, 2018). These are designed to run on N-body simulation such as PMFast (Merz et al., 2011) or CUBEP3M (Harnois-Déraps et al., 2013), however they take at least an order of weeks on today's supercomputers. This is far from practical for use with a statistical sampler. As discussed in Appendix C.2.1, MCMC (Markov Chain Monté

Carlo) algorithms require $\sim 10^4$ simulated samples necessary for convergence. Because of this, simulations that are computationally expensive and have long time constraints cannot be used. Alternatives to MCMC (e.g. a grid-search) are significantly more expensive. We are in need of a balance of accurate and efficient simulations. So-called hybrid simulations (Ocvirk et al. (2016, 2018) such as CoDa), try to speed this up by combining multiple computational techniques. Hydrodynamic and gravitational steps are run on CPU's while multiple radiative transfer and ionisation rate calculations are done in parallel on GPU's. These are promising in the detail they provide, but the speed increase is still not suitable for statistical analyses. Semi-numerical simulations provide a better alternative for exploring the statistical properties of the signal. Here, Gaussian initial conditions are computed and the results of the simplified models are set to match the full N-body simulations within $\sim 10\%$. Examples include 21cmFast (Mesinger & Furlanetto, 2007; Mesinger et al., 2011), SimFast21 (Santos et al., 2010), or the simulations used in e.g. Fialkov et al. (2014b). Emulating simulations with neural networks is becoming increasingly fashionable with an impressive increase in speed once the network has been trained e.g. (Kern et al., 2017; Schmit & Pritchard, 2018; Hassan et al., 2019; Cohen et al., 2019) (and Ghara et al. (2020) for a recent application that interprets LOFAR data).

In the context of 21cm EoR observations, a state of the art parameter estimation code which we use extensively is 21cmMC. This is built a pon 21cmFast and the following subsections detail them both.

1.3.1 21cmFast: semi-numerical cosmic 21 cm signal

21cmFast (Mesinger et al., 2011) is a semi-numerical modelling tool that combines spatial field realisations of spin temperature, peculiar velocity, ionisation and evolved density via a combination of linear perturbation theory and the excursion-set formalism. The evolved density calculation is taken from the predecessor[1] code DexM (Mesinger & Furlanetto, 2007). 21cmFast numerically implements the EoR model developed by Furlanetto, Zaldarraiga and Hernquist (FZH) (Furlanetto et al., 2004) which uses the excursion set formalism to identify HII bubbles (rather than for identifying virialised dark matter halos as in Appendix A.1). Doing this allows a simultaneous treatment of baryons and dark matter to calculate a 3D coeval box of 21cm brightness temperature field. The 'semi-numerical' part refers to using approximate

[1] http://homepage.sns.it/mesinger/DexM___21cmFAST.html

physics within the code in comparison to a hydrodynamic N-body simulation of reionisation (referred to as numerical). 21CMFAST is specifically used instead of other EoR simulation codes for two main reasons. Firstly for speed, the full radiative transfer codes can take up to months to run. Above scales of 1 Mpc, 21CMFAST is in agreement with radiative transfer simulations for $z = [6, 250]$ (Zahn et al., 2011). For use in preparing statistical analyses for radio interferometers that aim to detect the 21cm signal, these scales are adequate. Secondly, the path for implementing Bayesian parameter estimation has already been paved in Greig & Mesinger (2015) with 21CMMC which will be discussed in the next section.

The initialisation of 21CMFAST begins with a linear density field, $\delta(\mathbf{x})$, which is defined in Fourier space so that the discrete simulation cube (of length L) contains a finite set of wavenumbers in each spatial dimension,

$$\tilde{\delta}(\mathbf{k}) = \sqrt{\frac{\sigma^2(k)}{2}}[a_k + ib_k], \tag{1.9}$$

where the tildes indicates Fourier space, the σ^2 is approximated by the matter power spectrum of Eisenstein & Hu (1999) corresponding to a matter dominated universe (CDM cosmology), and the density field Fourier coefficients a_k and b_k are drawn from a Gaussian distribution (mean zero and standard deviation 1 in the range $[-L/2, L/2]$). Note that since we are in the linear regime (Section A.1), the cosmic variance σ^2 can be replaced by a power spectrum without losing information about the statistical properties of the field. A corresponding velocity field is then calculated from the density field as,

$$\tilde{\mathbf{v}}(k, z) = \frac{i\mathbf{k}}{k^2}\dot{D}(z)\tilde{\delta}(\mathbf{k}), \tag{1.10}$$

where $D(z)$ is the growth factor which is proportional to $a(t)$ for $z = [1, 1000]$. Once initialised, the simulation starts by using linear perturbation theory, where the Zel'dovich approximation for gravitational collapse is applied (Zel'Dovich, 1970; Mesinger & Furlanetto, 2007). In other words, the linear density field is evolved using the field velocities but no deviation to the particles' initial trajectories are allowed. This means the fields evolve with separable solutions in Lagrangian space (reducing the required computing power significantly). The initialisation is taken at $z = 300$ on a 1536 Mpc^3 co-moving cube, which needs to be large to account for the mean properties of reionisation later on (Iliev et al., 2006). Our resulting field is then smoothed

onto a lower resolution grid.

The ionisation field is identified by selecting regions (groups of pixels) for which the number of ionizing photons is larger than that of HI atoms. The excursion-set formalism is now applied to identify HII regions (Furlanetto & Piran, 2006). Pixels are labelled as fully ionised regions if,

$$\zeta f_{\rm coll}(\mathbf{x}, z, R) \geq 1, \tag{1.11}$$

where ζ is the ionising efficiency of UV radiation and $f_{\rm coll}$ is the collapsed mass fraction. The scale R moves from large to small values (in cMpc) smoothing the above threshold and ranging from the ionizing photon horizon (mean free path) down to the cell size. A minimum virial mass, $M_{\rm vir}$, for a dark matter halo to host an ionising galaxy in the simulation is used to define $f_{\rm coll}$ as,

$$f_{\rm coll} = \int_{M_{\rm vir}}^{\infty} M \frac{dn}{dM} dM, \tag{1.12}$$

where dn/dM, the halo mass function (see Section A.1), is a Press-Schechter function multiplied by a Sheth-Tormen correction factor (Bond et al., 1991). The neutral fraction per redshift is calculated by counting the ionised pixels within the volume of the cube i.e. by setting $x_{\rm HI} = f_{\rm coll\_per\_voxel}$ to regions as the excursion set formalism iterates down scales. Initially, the UV ionising efficiency ζ is defined,

$$\zeta = 30 \left(\frac{f_{\rm esc}}{0.12}\right) \left(\frac{f_\star}{0.05}\right) \left(\frac{N_\gamma}{4000}\right) \left(\frac{1.5}{1+n_{\rm rec}}\right), \tag{1.13}$$

where f_*, is the fraction of galactic gas in stars; $f_{\rm esc}$, the fraction of ionising photons escaping form the host galaxy into the IGM; N_γ, the number of ionising photon per baryon within pop III stars; $n_{\rm rec}$, the typical average number of recombinations per hydrogen atom. f_* is taken to be 0.05 however both f_* and $f_{\rm esc}$ are observationally uncertain (Gnedin et al., 2008; Wise & Cen, 2009; Ferrara & Loeb, 2013). The variability in $f_{\rm esc}$ is primarily what varies ζ. Since pop II stars are assumed to be the dominant stellar type during reionisation, we take $N_\gamma \approx 4000$ (Barkana & Loeb, 2005). In the case of a *photon-starved* end point to reionisation we assume $n_{\rm rec} \approx 1$ (Sobacchi & Mesinger, 2014). The virial mass (implemented in Equation 1.12) is linked to virial temperature via,

$$M_{\rm vir} = 10^8 h^{-1} \left(\frac{\mu}{0.6}\right)^{-\frac{3}{2}} \left(\frac{T_{\rm vir}}{1.98 \times 10^4 K}\right)^{\frac{3}{2}} \mathcal{F}(z) M_\odot, \tag{1.14}$$

where μ is the mean molecular weight i.e. $\mu = 1.2$ for HI and 0.6 for HII, and the atomic cooling threshold of 10^4K corresponds to a halo mass of 10^8 $M_\odot$ (Barkana & Loeb, 2001). The cosmological dependency has been packaged together in $\mathcal{F}(z)$, defined,

$$\mathcal{F}(z) \equiv \left[18\pi^2 + 82\frac{\Omega_\Lambda H_0^2}{H^2} - 39\left(\frac{\Omega_\Lambda H_0^2}{H^2}\right)^2\right]\left(\frac{1+z}{10}\right)^{-\frac{3}{2}}.$$

Work by Barkana & Loeb (2005) claims that halos need to be calculated to $10^8 M_\odot > M$ in order to include the atomic cooling processes allowing structures to achieve temperatures below their virial temperatures.

Combining all of the above via Equation A.29 provides us with our desired result - the 21cm brightness temperature field. Initially (in Chapters 2 and 3) we implement 21CMFAST under the assumption that $T_S \gg T_{CMB}$, known as the 'post-heating regime'. During the latter stages of the EoR the post-heating approximation is believed to be a good approximation (Furlanetto, 2006; Baek et al., 2010; Chen & Miralda-Escudé, 2004) (this is relaxed in Section 1.3.3). The most dominant factors in driving the signal are therefore the ionised fraction and the underlying density field (first and second terms of Equation A.29).

Since we are in the linear regime, the redshift space distortions in the 21cm signal are treated analogously to the Kaiser effect (Kaiser, 1987), i.e. all scales of clustering in redshift space are distorted by peculiar velocities (the 3rd bracketed term in equation A.29). The assumption that $dv_r/dr \ll H(z)$ is implicit in the form of this term however in order to avoid divergence, 21CMFAST imposes a maximum value of $|dv_r/dr| = 0.5H(z)$ which is necessary due to the pixel re-scaling that happens as each co-moving box evolves. Gaussian and non-Gaussian contributions to large-scale flow can theoretically alter the Kaiser terms (and pairwise velocity distributions) outside of this constraint, but the imposed limit has a negligible effect on the results (Scoccimarro, 2004). The peculiar velocity gradient field is most prominent in small scale over-densities, which are the first HI regions to be ionised in an the *inside-out* reionisation implemented here. Because of this, they are often ignored and smoothed over.

21CMFAST aims to be predictable from $z = [6,\ 250]$ to incorporate the long term effect of the gas coupling to the CMB temperature field. This in an impressive range of predictability, which extends well beyond the cosmic dawn and is larger than most competing simulations. For interpreting the signal the tool of choice is the spherically averaged version of the power

spectrum defined in Appendix C.4 (detailed before its application in Chapters 3 and 4). The results produced agree closely with numerical simulations for an impressive range of relevant scales, but with significant improvements in duration and memory restrictions.

This is the common grounding used in every subsequent chapter and the starting point for the developments in the next sections. An example of a competing simulation is Baek et al. (2010) - where the EoR 21cm signal uses a different range of virial masses ($10^{10}\ M_{\odot} > M$) without the assumption $T_{\rm S} \gg T_{\rm CMB}$ and the chemistry of Helium is also partially included. The results of this simulation show that no level of X-ray heating will prevent the 21-cm absorption phase, however QSO contribution ranging from $[0.0, 10.0]\%$ can change the skewness measure between the brightness temperature signal and HI fraction around the middle stages of reionisation. The addition of HeI reionisation is negligible on the behaviour of HI reionisation, however it should occur at lower redshift around around $z \sim 3$, since a higher energy radiation background is needed for helium's ionisation threshold. Detecting HeI reionisation directly would be a useful independent probe of the re-ionising universe and might be easier to detect than for HI because it occurs later (Furlanetto & Oh, 2008).

Unfortunately, all of the simulations we have referenced in this section miss detail in some form or another. For example, no numerical simulation can include the radiative transfer of X-rays and Lyman-α photons in parallel with the cooling from atomic and molecular hydrogen that influences star formation within the cosmic dawn (Loeb & Furlanetto, 2013). It is also unknown how much contribution QSOs provide (Garaldi et al., 2019) - and this would certainly vary the range in virialised halos that need to be simulated. Ciardi et al. (2006) argues that simulations will need to be done down to $M \sim 10^4 M_{\odot}$ in order to include circum-galactic regions of high hydrogen recombination rate, referred to as 'photon sinks'. In order to probe these objects radio observations deep into the cosmic dawn are likely necessary. Photon sinks play a large role earlier in the cosmic dawn, and so does modelling of the spin temperature fluctuations due to X-ray heating. X-ray heating effects are included in detail in Section 1.3.3, while the effects of photon sinks are subtle and will likely only be measured in relation to VAO scales alluded to in Section A.1. Simulations and mock observations of these processes in the context of 21cmFast are underway (Gordon & Pritchard, 2009; Muñoz, 2019; Cain et al., 2020; Hotinli et al., prep) but detailed discussion of this is beyond the scope of this work. For analytical calculations of comparable ionisation fields, please see Furlanetto et al. (2004); Barkana & Loeb

(2004); and for more details on the specifics of ionisation fields and an in-depth comparison of simulating reionisation models, please see Zahn et al. (2011).

We have outlined of the basic implementation in 21cmFast that is streamlined into 21cmMC. The capabilities of 21cmFast do go beyond what is described here. In particular each step is modular and applicable at the will of the user. Since we are interested in its use in the context of Bayesian statistics we have focused on explaining what is used in the closely related code 21cmMC in the next sections. The following sections follow the chronology of 21cmMC publications. Each includes more of the physics that 21cmFast is capable of simulating (with further detail provided before their application in Chapters 3 and 4).

1.3.2 21cmMC: parameter estimation in the EoR

21cmMC (Greig & Mesinger, 2015) combines a streamlined version of 21cmFast with the statistical analysis code CosmoHammer (detailed in Section 1.5). The latest version of 21cmMC has emerged from a progression of works, which we will build up to in the following sections. Here we outline the basis of 21cmMC referring ahead to Section 1.4.1 for an introduction to parameter estimation.

The spherically averaged 21cm brightness temperature power spectrum is used in the likelihood statistic. A mock observation must be simulated with 21cmFast using pre-selected fiducial parameters that are to be retrieved via the methodology. A larger box is used for the mock data in order to minimise cosmic variance issues, we need to be confident that the mock data is not victim to sample limitations. The sampled box is also simulated with 21cmFast but with sampled parameters and can be reduced a little to increase speed and memory in the computation - as long as the parameters recovered do not suffer as a consequence. In reality the observed patch of sky from the real telescope will be larger than the simulated box, we therefore include this effect in the testing phase. 250Mpc boxes with 128 pixels are used for the simulated samples and the mock observation is taken from a 500Mpc box with the same resolution. To speed the process further, a matter density (δ) and velocity field ($\frac{dv_r}{dr}$) has been pre-calculated and is read in during each sample. Importantly these initial conditions must be calculated with a different random seed to the mock observation.

The selected likelihood statistic is a χ^2 between the power spectra from the mock data and

the streamlined model dependent simulation. Noise estimates from the radio telescope on measuring the 21cm power spectrum can be calculated by 21cmSense (Pober, 2016), taking into account the layout of the desired telescope and specific assumptions about foreground noise as described in Section 1.3.7. Figure 1.6 summarised the process in calculating the likelihood within 21cmMC. The construction of the machinery means any model of the 21cm brightness temperature power spectrum can be swapped in and out, we take advantage of this in later chapters. This structure includes the input of three observational measurements as priors. The calculation of the neutral fraction in Section 1.3.1 is used in three checks against observations within the likelihood construction:

- The CMB optical depth observed by Planck is: $\tau = 0.058 \pm 0.012$ (Planck Collaboration et al., 2016b). 21cmFast estimates τ by interpolating the neutral hydrogen fraction across the coeval cubes. This allows testing between the observed and simulated values of τ at each MCMC call.

- The Gunn-Peterson trough location : McGreer et al. (2015) shows that the EoR must be $> 90\%$ complete by $z = 5.9$. Therefore $x_{\rm HI} = 0$ is set for all $z \leq 5.9$. For $z > 5.9$ the estimated neutral HI fraction is compared with an half-Gaussian of mean $\bar{x}_{\rm HI} = 0.06$ and variance $\sigma^2 = 0.05$.

- The Red Ly-α damping wing: Greig et al. (2017) estimates the $x_{\rm HI}$ surrounding QSO ULASJ1120+0641. At $z = 7.08$ the neutral fraction is checked against the observational measurement of $\bar{x}_{\rm HI} = 0.4^{+0.41}_{-0.32}$ at 2σ.

A χ^2 for each of these is combined linearly into the likelihood (see 'Observational Checks' in Figure 1.6).

The three free parameters are initially $R_{\rm mfp}$, $T_{\rm vir}$, and ζ. The photon mean free path, $R_{\rm mfp}$ is implemented as the largest scale at which the excursion set formalism is performed (a maximum R in Equation 1.11). This is not strictly a mean free path, this is the maximum photon path due to the methodology. Although there is known to be some spherical bias of the ionised regions we would expect $R_{\rm mfp}$ to gravitate around the mean bubble size. The minimum virial temperature $T_{\rm vir}$, is implemented in two places. Firstly it dictates the size of galaxies that start

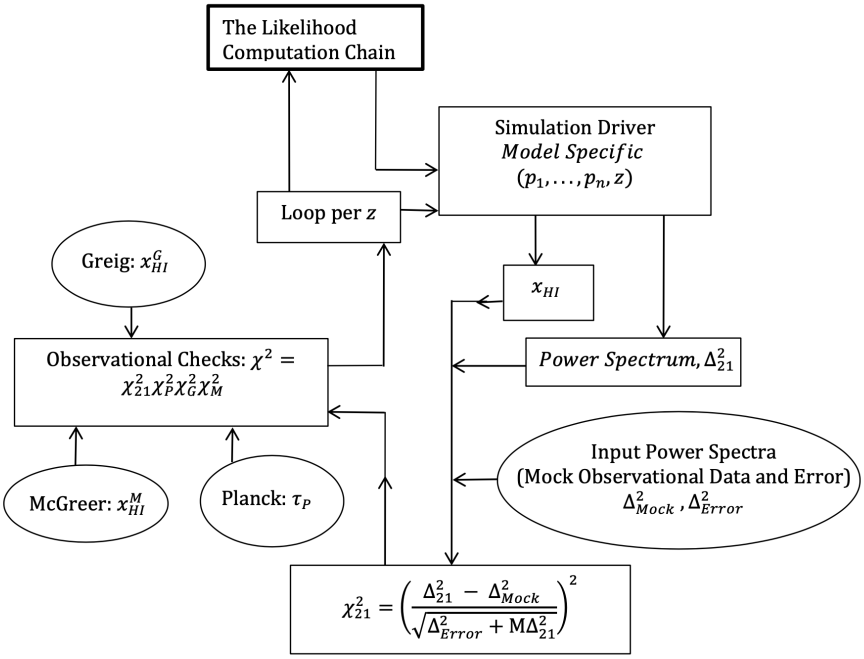


Figure 1.6: A summary of the likelihood computation chain used in 21CMMC, where M is a user defined model uncertainty, and Δ_{21}^2 is the spherically averaged 21cm power spectrum introduced in Chapter 3. In order to streamline 21CMFAST for use in 21CMMC the modular nature (described in Section 1.3.1) is reduced in order to have one callable function. The arrangement of a callable simulation driver means different elements of the 21CMFAST algorithm modules can be selected with the use of global parameters in the MCMC setup. This is particularly handy for comparing models with different physics.

ionising the IGM (with efficiency ζ) as,

$$\zeta = \begin{cases} \zeta_0, & \text{if } T'_{\rm vir} \geq T_{\rm vir} \\ 0, & \text{otherwise} \end{cases} . \tag{1.15}$$

Secondly, $T_{\rm vir}$ is used as minimum limit on the integral in Equation 1.12, used to calculate $f_{\rm coll}$ which is then used in the ionisation criteria defined in the previous subsection. The galaxy ionising efficiency, ζ, uses Equation 1.4 but we must relabel $\zeta \to \zeta_0$, as this parameter value is directly input into Equation 1.15. This form of ζ implies the assumption of a constant mass to light ratio for our ionising sources. This is the simpler of two models that are implemented in Greig & Mesinger (2015).

On top this 3 parameter model, a 4th parameter is introduced so that ζ can also be a power-law in halo mass. The model is adapted so that, $\zeta \propto (\frac{M'_{\rm vir}}{M_{\rm vir}})^{\alpha}$, where α is included as an optional 4th parameter that dictates the strength of the power-law scaling. It is worth emphasising that the 3 parameter reionisation model is the simplest, while this 4 parameter model (including α) is the simplest model that includes a model for stellar feedback. This will clearly effect the $f_{\rm coll}$ (Equation 1.12) as ζ is now a function of $M_{\rm vir}$. More detail (including a similar step-function in place of Equation 1.15) is provided in Chapter 3 (before this model is implemented). Both of these similar parameterisations assume $T_{\rm S} \gg T_{\rm CMB}$ for $x_{\rm HI} \leq 0.8$. Implementing the post-heating approximation in this way limits the applicable redshift range used, motivating the choice of evaluating coeval cubes for redshifts ($z \leq 10$).

These three and four parameter models of 21CMFAST are referred to as 3pFZH and 4pFZH respectively in Chapters 2 and 3. We will return to the discussion of prior range motivations for each parameter in Chapters 3 and 4, as well as details about how the likelihood statistic is implemented in each case.

1.3.3 21CMMC: the epoch of heating

Here we detail the background for relaxing the post-heating approximation, in order to model spin temperature fluctuations. The 21CMMC implementation (Greig & Mesinger, 2017b) introduces the concepts that lead into Model B in Chapter 4. Not including X-ray heating

in parameter estimation can lead to a $\sim 10\ \sigma$ skewing of the parameters depending on the magnitude of the temperature changes and intensity of the primordial X-ray sources.

Up until now we have assumed that the spin temperature overwhelms the CMB background ($T_{\rm s} \gg T_{\rm CMB}$). This is thought to be true during the latter stages of reionisation i.e. $z < 9$. (McQuinn et al., 2011). Prior to this, $T_{\rm S}$ offers a lot of insight into the formation, growth, and evolution of structure in our universe. The production methods of X-rays are varied, but they consist mainly of supernova remnants from Population III stars. These predominantly contribute via high mass X-ray binaries or dust content in the inter-stellar medium. Bremsstrahlung from supernova radio remnants can heat the ISM, absorbing the remnant emission spectrum and causing sparse patches of the ISM itself to emit in a different portion of the spectrum (Eide et al., 2018). Mini-quasars will also contribute (Qin et al., 2020) but we do not include mini-haloes in this work. The recipes used for the soft-band X-ray emission is robust to observations of nearby galaxies (Pacucci et al., 2014).

Here we look to calculate the spin temperature, ($T_{\rm S}$ in Equation A.35) from Section A.3.3. To implement this calculation of the spin temperature the approximation, $T_\alpha \approx T_{\rm K}$ is used. In other words the colour temperature is coupled closely to the kinetic gas temperature due to the Wouthuysen-Field effect mentioned earlier (Section 1.2). The collisional coupling coefficient used here is calculated in Furlanetto & Furlanetto (2007) and Zygelman (2005). T_α is calculated, via $T_{\rm K}$, following the kinetic theory analysis developed by Hirata & Sigurdson (2007). The primary source for this heating of the gas is theorised to be high redshift X-rays (Furlanetto & Pritchard, 2006; Pritchard & Furlanetto, 2007). The ionisation, temperature and $Ly-\alpha$ background can be considered by looking directly at, $\epsilon_{\rm X}$, the X-ray specific emissivity (detailed when the model is used in Equation 4.4). For now we can simply state $\epsilon_{\rm X} \sim L_{\rm X} \times$ [light cone volume], where $L_{\rm X}$ represents X-ray luminosity. The absorption cross section for this radiation scales with $\sim E^{-3}$ as redshift decreases, hence there are specific signatures in the heating of the gas that relate to the contents of the first galaxies.

In order to compute our angle-averaged specific UV intensity, $J(\mathbf{x}, E, z)$, we must add up the contributions from earlier times (z'). This is done with respect to the probability, $e^{-\tau_{\rm X}}$, that an X-ray photon will survive from z' until z. Namely the angle-averaged specific X-ray intensity is calculated by integrating the co-moving X-ray emissivity, $\epsilon_{\rm X}$, throughout the observed light-

cone to produce,

$$J(\mathbf{x}, E, z) = \frac{(1+z)^3}{4\pi} \int_z^\infty \frac{cdt}{dz'} \epsilon_{\rm X} e^{-\tau_{\rm X}} dz', \tag{1.16}$$

where $\tau_{\rm X}$,

$$\tau_{\rm X}(E, z, z') = \int_z^\infty c \frac{dt}{dz'} \bar{x}_{\rm HI}(z') \bar{n}_{\rm b}(z') \sigma_{\rm c}(z', E) dz', \tag{1.17}$$

is the scattering optical depth for X-rays (Mesinger et al., 2011). This relates to the attenuation of hydrogen (and helium) due to reionisation as, $e^{-\tau_{\rm X}}$. The species weighted photo-ionisation cross-section used is given by, $\sigma_{\rm c}$ as,

$$\sigma_{\rm c} = f_{\rm H}(1 - \bar{x}_{\rm e})\sigma_{\rm H} + f_{\rm He}(1 - \bar{x}_{\rm e})\sigma_{\rm HeI} + f_{\rm He}\bar{x}_{\rm e})\sigma_{\rm HeII}, \tag{1.18}$$

where $f_{\rm X}$ and $\sigma_{\rm X}$ represent the fraction and cross sections for the respective species. The conversion of X-ray photon energy from the emitted ($E_{\rm e}$) frame to the observed frame (E) needs to be included, and is given as,

$$E = E_{\rm e} \frac{1+z}{1+z'}. \tag{1.19}$$

We now need the photo-ionisation rate, $\Gamma_{\rm X}(\mathbf{x}, z)$ which is calculated by integrating J across all frequencies that can reach nearby gas clouds and for all species that interact with the radiation (HI, HeI, HeII). Once calculated we then infer the evolution of the IGM free electron fraction as,

$$\frac{dx_e(\mathbf{x}, z)}{dz} = \frac{dt}{dz}(\Gamma_X - \alpha_{\rm A} C x_e n_{\rm HI}), \tag{1.20}$$

where C is the clumping factor for the pixel; α_A is the recombination for optically thin gas (recombining radiation can escape and is ignored); $n_{\rm HI}$ is the number density of neutral Hydrogen; and as before, the Helium contribution can be ignored (Zahn et al., 2011). $T_{\rm K}$ is then solved for in tandem with Equation 1.20 by using the comoving UV energy density and electron number density.

Here, Greig & Mesinger (2017b) expand upon the previous 3 parameter model to include X-rays produced in the first galaxies because they result in an inhomogeneous heating of the gas at EoR redshifts. The free parameters introduced are: $\alpha_{\rm X}$, the X-ray luminosity power law index which describes the hardness of the radiation; $L_{\rm X<2keV}$/SFR, the normalisation of the X-ray luminosity for photons of $E < 2$keV normalised to the star formation rate; and E_0, the

minimum escape energy for an X-ray photon. The mean free path of X-ray photons is a strong function of photon energy, as the spectrum hardens and resultant heating through the IGM becomes more uniform. In galaxies the ISM HI column density and metallicity will effect the release energy of X-rays into the IGM. It is certainly true that low redshift (local) galaxies have higher metallicities than the simulated high redshift analogues. The X-ray SEDs of these simulated galaxies are dominated by the HI column density, motivating a step function cut off for the soft band X-ray luminosity. Typically during most of the EoR it is a good assumption to say that these high redshift galaxies are metal free. All assumptions used in the modelling here agree with observations of the soft-band X-ray spectrum from local star-forming galaxies (McQuinn et al., 2011; Das et al., 2017). More parameter discussion is included in Chapter 4 under the context of Model B. For full detail, please see Mesinger & Furlanetto (2007) and Mesinger et al. (2011).

1.3.4 21CMMC: introducing the light-cone

Up until now we have been dealing solely with coeval cubes. As discussed earlier (Section B.1.2) a radio telescope will take images of the night sky with a depth dictated by its frequency bands. For an accurate tomographical image the resolution of these discrete measurements will have to be better than the evolution rate of the comoving gas in the observed redshift range. At mid EoR redshifts (e.g. $z = 8$) our 8 MHz bandwidth stretches $\sim$ 125Mpc, motivating a bare minimum of box sizes that should be simulated for access to a volume averaged statistic, such as the power spectrum. In Chapter 4 we move from statistical analysis on coeval cubes (as in Chapters 2 and 3) to measuring the 21cm power spectrum from a light-cone. The coeval cubes of the 21cm brightness temperature are combined into a light-cone as detailed below. When compared to using coeval cubes, using the light-cone in a parameter estimation context can cause some parameter posterior distributions to change by $\sim 10\ \sigma$ in cases where the 21cm signal evolves rapidly (Greig & Mesinger, 2018).

Initially, coeval cubes are simulated as before. These cubes are then stitched together to emulate the evolution of structure along the line-of-sight. Flexibility on which redshifts are sampled as coeval boxes is dictated by the user. The inputted step size ($\Delta z_{\rm step}$) relates to a linear interpolation in cosmic time and the light-cone is constructed by interpolating properties between two of these coeval boxes at every point. These coeval boxes all have the same random

seed but are evolved to specific points in time dictated by the redshift step size. Each coeval box is evolved and stitched to itself along the light-cone in-between a user defined minimum and maximum redshifts. This provides N coeval pieces (where N is the number of redshifts simulated, defined by $\Delta z_{\rm step}$). Each piece represents a periodic sample of the overall (simulated) evolution. Now for each pixel in the light-cone, its properties are dictated by a weighted mean between the properties of the two coeval pieces it is between. If $z_{\rm cell}$ is the redshift of the pixel and $z_{\rm piece}$ is the redshift at which that strip was evaluated, then the weightings are determined as $|z_{\rm cell} - z_{\rm piece}|$. This means that for a pixel that is specifically at one of the simulated redshifts, the original properties are used; and for anything in-between, a weighted mixture is obtained depending on the location of the pixel relative to the two simulated coeval pieces. If the end of the coeval box is reached before the next box starts, the properties of the nearest box wrap around. Since the random seed is the same for the coeval boxes at all redshifts, the same essential structure is repeated with a period relating to original coeval box length. Major structures in the light-cone will appear visibly periodic because of this. One should note that the periodicity of the simulated structure (caused by the wrapping) is also a function of box length, even if an infinitesimally fine grid of redshifts are sampled. The result is a brightness temperature light-cone that provides statistical properties that are similar to those which a real radio telescope might observe. Deviation from this should be inversely proportional to $\Delta z_{\rm step}$.

For implementation in the 21CMMC likelihood, equal co-moving distance chunks of the light-cone are separated out. Effectively this is running a hard edged (non-overlapping) top-hat filter accross the light-cone, not dissimilar to observing with the discrete frequency bands dictated by a radio telescope's bandwidth. Each chunk then has its power spectrum calculated separately in order to capture the signal's evolution. The use of chunks also enables easy comparison between the statistics from the coeval cubes produced earlier. Since the power spectrum is now valid to larger scales, the 21cm χ^2 is implemented within an increased range of $k \in [0.1, 1.]$ Mpc^{-1} when using light-cones.

Calculating a summary statistic from a light-cone is a tall order. Applying the box-car sampling technique as described here is not void of the light-cone effect, where the evolution of the universe accross the size of the observation may effect the results (Datta et al., 2012, 2014). Ignoring the light-cone effect (as is done here) should not bias the parameter posteriors obtained, but it will worsen the associated uncertainty (La Plante et al., 2014; Mondal et al., 2017).

Currently, the field contains no unanimous way to remedy this. A novel method known as the wavelet transform (Trott, 2016) is thought to be able to decrease this uncertainty by taking advantage of the full radio interferometer's bandwidth at once rather than using individual chunks. In Chapter 5, we look at furthering this analysis.

1.3.5 Inhomogeneous recombinations

So far we have considered a reionisation with *inside-out* morphology. In Chapter 3 we introduce a toy *outside-in* model based on Miralda-Escudé et al. (2000) for comparison, but in reality the EoR will progress via a combination of both the *inside-out* and *outside-in* morphologies as discussed in section 1.2.2. Here we include a mechanism for including inhomogeneously the recombination rate of HI, that will dictate the morphology of the IGM depending subtly on the astrophysical parameters. Simulations dominated by bright galaxies will remain entirely *inside-out* and simulations with mostly dim galaxies will be mainly *outside-in*. Sobacchi & Mesinger (2014) is implemented in all of the models used in and after Chapter 4. Prior to this change, recombinations are homogeneously included as a constant ($n_{\rm rec}$) within ζ in Equation 1.13.

Assuming the density distribution of the gas responds instantaneously to heating, the fraction of volume in which the IGM gas is at over-density, $P_{\rm V}$ is expressed,

$$P_{\rm V}(\delta, z) = Ae^{-\frac{(\delta^{-2/3}-c)^2}{2(2J_{\rm L}/3)^2}}\delta^{-\beta}, \tag{1.21}$$

where c and $J_{\rm L}$ are fitted for a CDM universe (the latter loosely scaling as the Jeans length in the ionised IGM); and, β is obtained from an iso-thermal sphere profile for a high-density absorber.

Our ionisation threshold is then adapted to,

$$f_{\rm coll}\zeta \geq 1 + \bar{n}_{\rm rec}, \tag{1.22}$$

where the inhomogeneous recombinations are calculated as,

$$\bar{n}_{\rm rec} = \left\langle \int_{z_{\rm ion}}^{z} \frac{\bar{n}_H \alpha_{\rm B}}{1+\delta} \left[\int_0^{180} (1-x_{\rm HI})^2 P_{\rm V} \delta^2 d\delta \right] \frac{dt}{dz} dz \right\rangle_{\rm R}. \tag{1.23}$$

The angular brackets denote a spherical average within a bubble of radius R; δ is the non linear over-density; proper time is denoted, t; $x_{\rm HI}$ is the fraction of Hydrogen that remains neutral; $z_{\rm ion}$, is the redshift at which the given cell was first ionised; and, $\Delta \equiv n/\bar{n}$ represents the sub-grid over-density of gas. The upper limit of the integral, 180 is from the mean over-density within the spherical top hat collapse model, realistic values during reionisation should not challenge this. $P_{\rm V}$ is the volume averaged PDF for gas in a given volume to be an over-density. Its in a near-Gaussian form (Miralda-Escudé et al., 2000), including necessary corrections for self-shielding (Furlanetto & Oh, 2005; Sobacchi & Mesinger, 2014). $\alpha_{\rm B}$ is the recombination-coefficient calculated at 10^4K assuming any subsequent ionising recombination radiation is re-absorbed by the same region. In other words an effective recombination rate is approximated by excluding transitions that emit ionising photons (Osterbrock & Ferland, 2006). This is eligible in regions of high gas density (Rahmati et al., 2013) e.g. in McQuinn et al. (2011) when the IGM is optically thick.

The neutral fraction produced, $x_{\rm HI}(\delta, \Gamma_X, z)$ depends on the over-density, the reduction of gas from the local inhomogeneous ionisation rates, as well as redshift respectively. The uncertainty in $\Gamma_{\rm X}$ increases at lower redshifts due to the patchy end to reionisation. There exists evidence that it could scale as a power law with emissivity but local areas can easily be changed by the density of Lyman limit systems (McQuinn et al., 2011). Here Γ_X is quantitatively obtained from smoothed-particle hydrodynamic simulations, please see Rahmati et al. (2013) for more detail. Hutter (2018) show that semi-numerical techniques that vary the ionisation rate show similar results on global statistics like the power spectrum, despite photon non-conservation being an issue for localised detail (leading to patches of accelerated reionisation). When a photo-ionisation equilibrium is assumed, the form of Equation 1.23 contains the self-shielding contributions of the gas within the neutral fraction calculated as,

$$x_{\rm HI}\Gamma_{\rm X} = 1.08 \times n_{\rm HI}(1 - x_{\rm HI})^2\alpha_{\rm B}, \tag{1.24}$$

where the gas column density is taken to agree with observations (Furlanetto & Oh, 2005). Omitting this effect can suppress the 21cm power spectrum by factors of 2-3 at large ($k < 0.2$ Mpc^{-1}) scales at the tail end of reionisation ($z < 8$). Since this naturally influences the size of the ionised regions, a photon mean free path is redundant. This means $R_{\rm mfp}$ is no longer a variable, however a maximum scale for identifying bubbles is necessary in the

excursion set formalism. A constant $R_{\rm mfp} = 50\ h^{-1}$Mpc is used by default when inhomogeneous recombinations are included.

1.3.6 21CMMC: including the UV luminosity function

This section introduces three main improvements developed in Park et al. (2020) that are beyond the work found in 21CMFAST. Firstly, the parameterisation of ζ in Equation 1.13 as a constant does not agree with observations of high redshift galactic data. Here this is accounted for by expanding ζ into two power-laws: one for the UV escape fraction ($f_{\rm esc}$) and one for the baryonic mass fraction that has collapsed into stars (f_*).

Secondly, we move away from using $f_{\rm coll}$ (Equation 1.12) for the collapsed fraction of baryonic mass stars in favour of $M_{\rm turn}$, a turnover mass that dictates the abundance active star forming galaxies. This update moves away from the parameterisation of galaxy populated halos via $T_{\rm vir}$. Instead more physics is encapsulated via a suppression motivated by the galactic duty cycle. Small galaxies (below the threshold of $M_{\rm turn}$) have their star formation stalled by feedback mechanism that are dependent on the duty cycle of galactic emission. Functionally this provides a smoother output from the simulated galaxy clusters which is comparable to hydrodynamic simulations e.g. (Paardekooper et al., 2015; Ocvirk et al., 2016). Typically the star formation in small galaxies is suppressed by various mechanisms such as supernova feedback, photo-heating feedback, or inefficient gas accretion - however the mechanism of suppression is less important for EoR modelling than the initial scale and shape of the suppression (Wu et al., 2019). This parameter increases the complexity of the relationship between halo mass and star formation based processes accessible to semi-numerical EoR modelling.

Finally, a star formation timescale, t_* which is used to calculate a UV luminosity function for the high redshift galaxy clusters. Up until now 21cm literature has included three observational checks. As shown in Binnie & Pritchard (2019), these add little to no inference to the constraining power of the 21cm power spectrum. Luminosity functions (LFs) contain a lot of information about the nature of galaxies at different redshifts. This parameterisation allows a prescription for the star formation rate (SFR), estimated as a the total stellar mass in stars divided by the new critical timescale.

In Chapter 4 we detail these mechanism (as implemented in 21CMMC), including the UV LF

calculation and how it is combined into the likelihood function. The detailed descriptions for producing and implementing t_*, $M_{\rm turn}$, and the double power law in halo mass for ζ are also included in Chapter 4.

1.3.7 21cmSense: simulating interferometer noise with foreground avoidance for the EoR

Chapters 3 and 4 make use of 21cmSense (Parsons et al., 2012; Pober et al., 2013, 2014; Pober, 2016) - a code that simulates radio interferometer noise specifically for observing the 21cm power spectrum in the EoR and cosmic dawn. Here we apply our knowledge of noise theory (Section B.1.2) to 21cm power spectrum simulation (Section 1.3). 21cmSense is a foreground avoidance algorithm and operates in two halves: Firstly, it generates the properties of the telescope array - known as the array file (Alg 1); secondly, this is then used in parallel with the clean 21cm brightness temperature PS (from the simulation box) to calculate the observed signal - this is the sensitivity calculation (Alg 2).

The array file receives telescope details: name, central location[2] [WGS], antenna locations [ENU], antenna collecting area, beam type, dish size [λ] and receiver temperature [mK]. 21cmSense then performs a fast Fourier transform resulting in the U-V-W distribution which is flattened for the U-V distribution of the desired telescope array.

In the sensitivity calculation, the produced array file is then inputted into the sensitivity calculation along with the frequency of the observation [GHz] and the length of the observation [hours per day, number of days]. Equation B.48 is used to calculate the visibility measurements for observing the field, including the noise fluctuation from the telescope. Finally, we are left with an array of the power spectrum errors[3] for each $[|k|, T_{\rm errs}]$.

Radio telescopes can operate in two observational scanning modes: track or drift. Track is performed by recording a time difference between each end of the interferometer (on top of Equation B.46) which has the same effect as tilting a physical dish. Tracked scans will measure fewer unique baselines, and therefore have a poorer quality U-V space to interpolate with, but

[2]WGS (World Geodetic System, sometimes WGS84) is a spherical polar coordinate system based form the centre of the Earth; ENU (East-North-Up) is a Cartesian flat Earth equivalent of WGS.

[3]21cmSense contains the capability to include a 2D array for errors where only the parallel k modes are averaged, in this case the output is $[|k_\perp|, k_\parallel, T_{\rm errs}]$ and algorithm 2 has two k loops.

the signal to noise of the desired source will be reduced as the observation of the source is longer. Drift scans allow the source to move accross the telescope beam as the Earth rotates. The side lobes of a radio telescope beam are hard to pin down, both methods can be unpredictable in cases where the noise is dominant over the desired source. Using multiple drift scans can reduced this effect (since only the sky should change), leading to a more precise measurement of the beam response. Within old 'dome-style' radio telescopes drift scanning is favoured (Barvainis, 1997) since its own structure scatters radiation into the detector. For new technology, ground based antenna should immediately remove this effect meaning a drift scan should measure the foreground sources better while tracking is vital for optimising the signal to noise for a specific target. Choosing a drift or a tracked scan is a trade off between the speed of measurement and the precision of measuring the non-desired objects in the sky.

For avoiding the foreground wedge, 21cmSense contains three levels of increasing severity:

- Pessimistic: all baselines are added incoherently and no k modes are included in the horizon or the buffer zone.

- Moderate: all baselines are added coherently but no k modes from the horizon or buffer zone are included.

- Optimistic: All k modes in the primary field of view are used.

The buffer parameter can be changed to cater for larger horizon effects along the foreground wedge but we set this to the default value of 0.1 hMpc^{-1} throughout our analyses. Other possible parameter changes are the cosmological bandwidth (default 8Mhz), which captures the redshift range that can be considered coeval; and the number of channels (default 82), giving 1024 channels per 100 MHz of bandwidth - both are left unchanged throughout this work.

Algorithm 1: 21cmSense: The array file

Convert antennae distances to wavelength;

Calculate Baseline - i.e. vector distances between each antenna;

Fast Fourier transform the baseline distribution;

Flatten the resulting u,v,w, producing a grid of U-V bins;

(bin size is bandwidth dependent);

if *Track* **then**

| Time for U-V coverage is specified by the observation parameters;

end

if *Drift* **then**

| Time set by drifting sky accross the estimated beam - typically this is $0.15/\nu$ for the full width half maximum of a 2D Gaussian;

end

The receiver temperature, $T_{\rm RX}$, and distribution of the U-V bins are then passed to the sensitivity calculation.

Before proceeding with the sensitivity calculation, we need to get to grips with a few quantities. To convert between observational and cosmological units using the redshift dependent scalar, $X2Y(z)$, is defined,

$$X2Y(z) = \left(\frac{1.9}{2.91\times10^{-4}}\frac{1+z}{10}\right)^{\frac{1}{2}} \times \left[\frac{1.7}{0.1}\left(\frac{1+z}{10}\right)^{5} \times 1000\left(\frac{\Omega_{\rm m}}{0.15}\right)^{-\frac{1}{2}}\right], \tag{1.25}$$

where the first term (left bracket) converts the angle on the sky to proper distance [h^{-1}Mpc], and the second (square bracket) represents, $dD_c/d\nu$, changing a bandwidth [GHz] to a line-of-sight distance in h^{-1}Mpc (both for a given redshift) (Furlanetto et al., 2006b,a; Parsons et al., 2012). Equation 1.25 contains, $dk/d\eta(z) = 2\pi/\left(dD_c/d\nu\right)$, which converts a baseline length in λ to k mode to transverse k-mode [hMpc^{-1}] at a given redshift. The effective beam solid angle, $\Omega_{\rm eff}$, is defined in reference to the amplitudes received by the telescope (Equation B.48) as,

$$\Omega_{\rm eff} = \frac{\left(\int |B_\nu(l,m)| dldm\right)^2}{\int |B_\nu(l,m)|^2 dldm}, \tag{1.26}$$

i.e. it is the ratio of the power-square beam to the square of the beam's integrated power (Parsons et al., 2012, 2014) and we refer to Equation 1.26 as, $\Omega_{\rm eff} = \Omega_{\rm p}^2/\Omega_{\rm pp}$, in Chapter 5.

This is specific to use in the power spectrum as the numerator comes from measuring two visibilities and the denominator from the volume in the desired integral. In the limit of a top-hat beam model the power-square beam and the beam's integrated power squared are the same. The nature of using units of λ and having two discrete halves easily allows changing the observing wavelength since this will effect the relative size of the dish, and therefore the nature of Ω_{eff}. Continuous galactic HI emission dominates the foreground and is modelled from observation (Jacobs et al., 2013) to be,

$$T_{\text{sky}} = 60 \times 10^3 \left(\frac{\nu}{0.3\text{GHz}}\right)^{-2.55} \text{mK}, \tag{1.27}$$

before being combined into the system temperature as,

$$T_{\text{sys}} = T_{\text{sky}} + T_{\text{RX}}. \tag{1.28}$$

Algorithm 2: 21cmSense: The sensitivity calculation

$|u| = u^2 + v^2$ **for** u, v **do**

  if *pess, mod* **then**

    horizon_limit $= \frac{dk}{d\eta}(z)\frac{|u|}{\nu}$ + buffer ;

  end

  if *opt* **then**

    horizon_limit $= \frac{\mathrm{dk}}{\mathrm{d}\eta}(\mathrm{z})\sin\left(\frac{1}{2} \times \mathrm{first\_null}\right)$;

  end

  for k **do**

    $T_{\mathrm{rms}} = T_{\mathrm{sys}}/\sqrt{2B \times 10^9 \times t_{\mathrm{obs}}}$;

    Noise $= \mathrm{X2Y(z)} \times \Omega_{\mathrm{eff}} \times \mathrm{k}^3/2\pi^2$;

    Sense $= [\mathrm{Noise} \times T_{\mathrm{rms}}^2 + \Delta_{21}^2]^{-2}$;

  end

end

where sense is the visibility sensitivity, which must be normalised by the observing scale to become the telescope sensitivity, Tsense, which we desire.

The errors are added in inverse quadrature, and we still need to normalise for the number of independent fields observed:

for $k_\parallel$ **do**

  sense$(k) =$ sense$(k)^{-\frac{1}{2}}$;

  for $k_\perp$ **do**

    Tsense $=$ Tsense $+ 1/$sense$(k_\parallel)^2$;

    $k = \sqrt{k_\perp^2 + k_\parallel^2}$;

    Tsense$(k) =$ Tsense$(k_\perp)^{-\frac{1}{2}}$;

  end

end

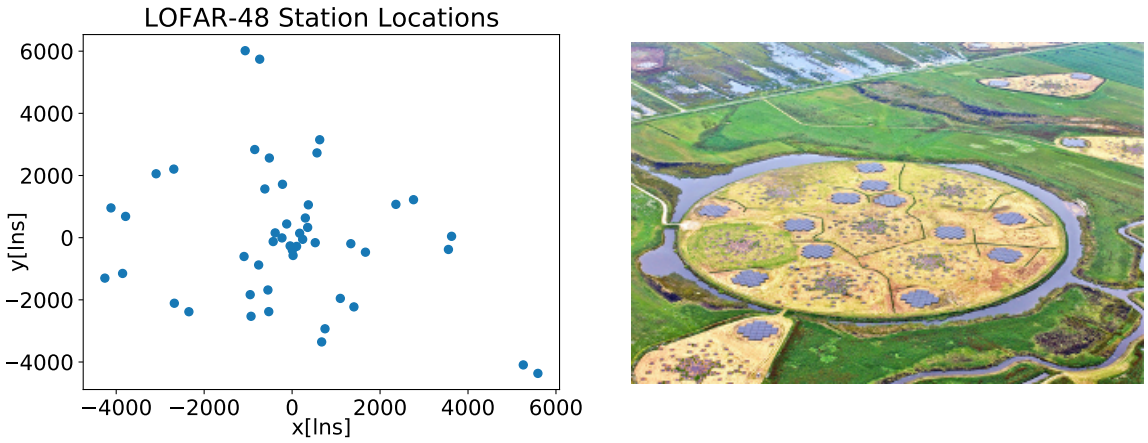
Undoing the inverse quadrature leaves us with the final result:

for k **do**

  Tsense$(k) =$ Tsense$(k)^{-\frac{1}{2}}$;

end

We now have a telescope's measurement sensitivities per $k_\perp$ for a given 21cm power spectrum and observing frequency. The noise sensitivity of a radio telescope power spectrum measurement roughly scales as $(Bt_{\mathrm{obs}})^{-\frac{1}{2}}$, where B originates in the limits of the integral in Equation B.48



(a) (b)

Figure 1.7: 1.7(a) Each blue point represents a station location for LOFAR-48 (photographed in 1.7(b) by ASTRON, https://www.astron.nl/telescopes/lofar/). These locations are fed into 21cmSense, which estimates the sensitivity of the interferometer by Fourier transforming the vector baselines between each station.

and $t_{\rm obs}$ from the thermodynamics of signal from the observed source dominating over the instrumental noise (which will saturate, see Section B.2). By assuming Gaussian errors on cosmic variance, we have expressed the total uncertainty with an inversely weighted summation across all the k modes as,

$$\delta\Delta^2_{21}(k) = \left\{ \sum_i \frac{1}{[\Delta^2_{N,i}(k) + \Delta^2_{21}(k)]^2} \right\}^{-\frac{1}{2}}, \tag{1.29}$$

where $\Delta^2_{N,i}(k)$ is represented by Noise $\times$ $\mathrm{T}^2_{\rm rms}$ product within Alg 2.

For accurate implementation of 21cmSense a balance between overcoming both cosmic variance and thermal noise from the telescope must be done. Medium-deep observations are the best i.e. 100hr observations of 10 independent fields favours 1000hr (1hr) observations of 1 (100) field(s) (Greig et al., 2020c), however cuts are conventionally made to ensure the results are accurate. Within Chapters 3 and 4 we utilise the observed power spectrum between $[0.15, 1]$ and $[0.1, 1]h\mathrm{Mpc}^{-1}$ to account for observing coeval cubes or the light-cones respectively.

We implement 21cmSense with three up and coming radio interferometers, which show promise in using the 21cm power spectrum a parameter estimation setting (McQuinn et al., 2006). These are LOFAR-48 (The LOw Frequency ARray[4]), HERA, and the SKA-512, where the number

[4] http://www.lofar.org

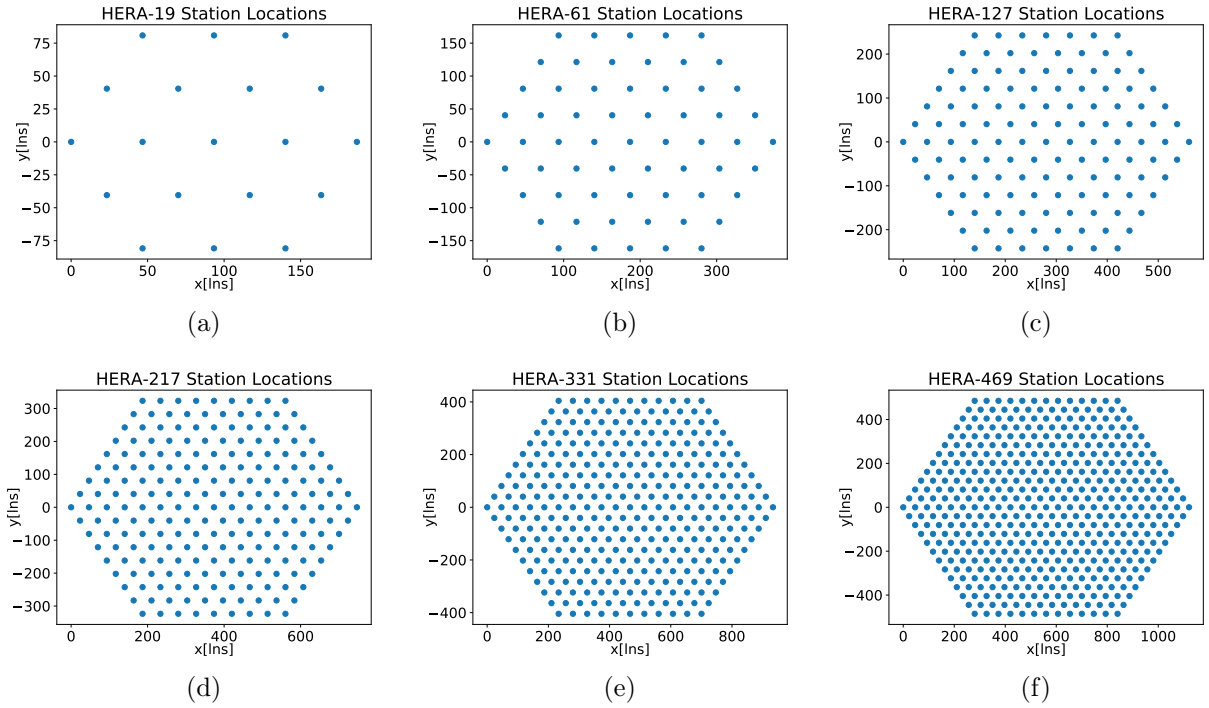
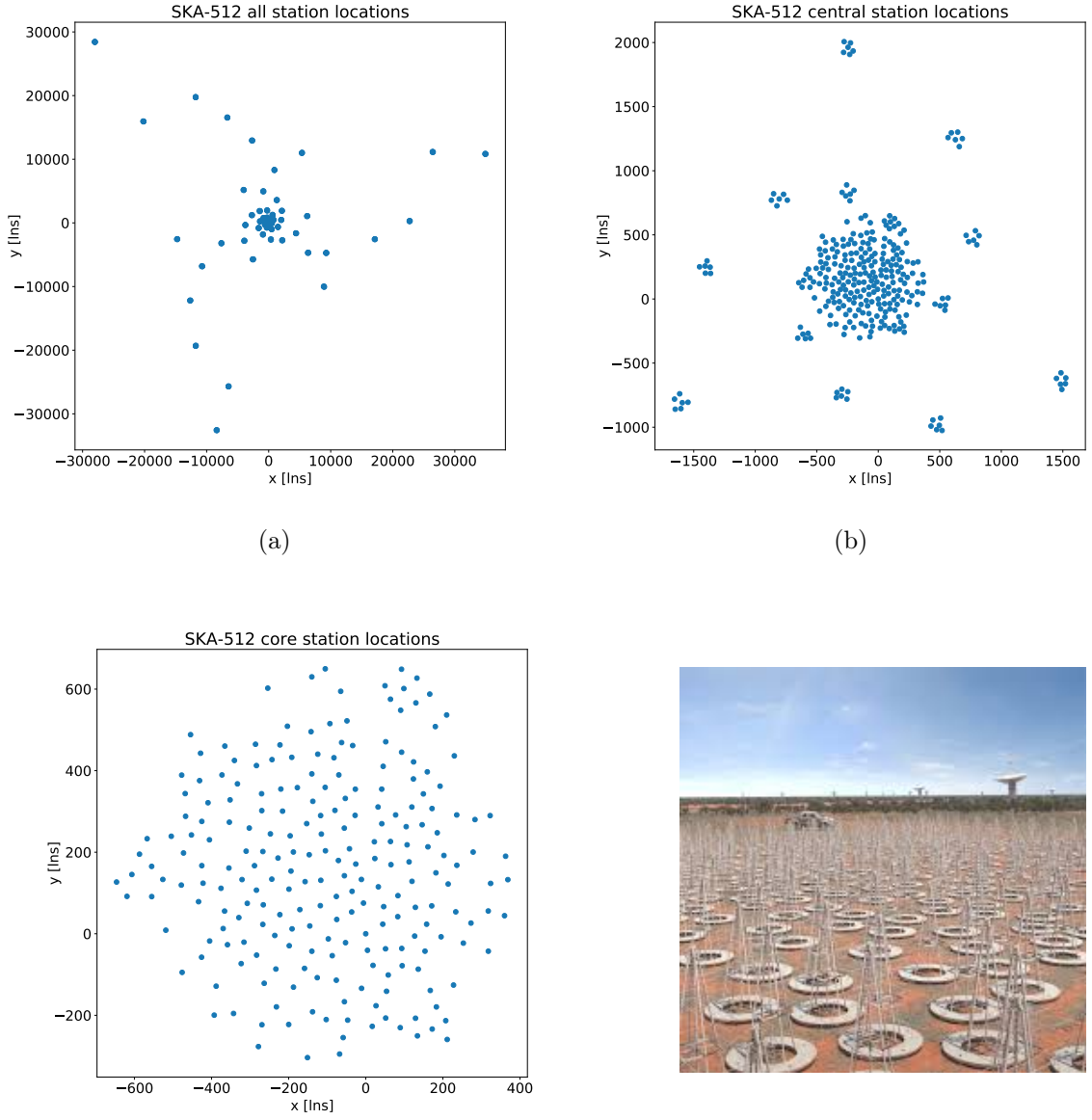


Figure 1.8: Configurations of HERA dipoles. Figures 1.8(a), 1.8(b), 1.8(c), 1.8(d), 1.8(e), and 1.8(f) represent the array with 19, 61, 127, 331, and 469 dipoles respectively.

represents how many dipole stations exist in the array. HERA is built in a hexagonal structure which is easily extendable and so we vary the number of antennae used in its observation as HERA-N where $N \in [19, 61, 127, 217, 331, 469]$ - each with $[3, 5, 7, 9, 11, 13]$ dipole stations along each side. Figures 1.7, 1.8 and, 1.9 plots the locations of the antenna for LOFAR-48, HERA-N, and SKA-512. The three telescopes are located in the Netherlands (North of Exloo '52:54:32', '6:52:8'), South Africa (North of Capetown '38:25:59.24', '-79:51:02.1'), and Australia (Miluera '-26:42:11', '116:40:14') respectively. LOFAR does tracked scans, HERA drift and SKA can perform both. The receiver temperatures used are 140K, 100K, 30K respectively. We have assumed a 2D Gaussian fit for each of the three telescope beams.

Of these three telescopes LOFAR is the only currently operational one and so it serves as a modern day comparison against the future developments that are HERA and the SKA. HERA and the SKA both test two different methods for noise reduction which boil down to how well the delay spectrum can be obtained from interpolating U-V space. The HERA station layout has multiple repeated measurements of the same baselines, and so each U-V measurement will increase with accuracy as the telescope increases in size. This enables the thermal noise on each measurement to reduce $1/\sqrt{N}$. SKA on the other hand has a much larger variety of independently sized baselines, providing a well covered interpolation into U-V space. The

argument here is that the better established the Fourier transform from visibility space, the more accurately the measurement of the wedge will be and as a result the fidelity of the 21cm signal estimation will be increased. This roughly corresponds to being able to subtract a more precise sky map from the image created by the instrument, if the Fourier dual of the image is better measured the resulting subtraction will be more accurate. It is an active area of research as to which of these techniques will be favoured (Dillon & Parsons, 2016; Byrne et al., 2019; Dillon et al., 2019). HERA is optimised for the power spectrum as its sparse measurement of the U-V space will not likely allow statistics of higher order to be measured. However there is also an active area of research into what order poly-spectra can be accurately reached, given the presence of such large foregrounds in the first place (Watkinson et al., 2020). A balance between the two methods is likely required.



(a) (b)

(c) (d)

Figure 1.9: The SKA will be the world's largest FM radio consisting of 10^6 antennae by 2030, after its secondary upgrade. Figures 1.9(a), 1.9(b), 1.9(c) represent the full, central and core stations. The spiral arms in the full array stretch $\sim$ 40km accross and the core is $\sim$ 400m accross. 1.9(d) shows an artist's impression of the finished construction (SKA Organisation/Swinburne Astronomy Productions, https://www.skatelescope.org/multimedia/image/ska-low-frequency-aperture-array-close-up-australia/).

1.4 Bayesian analysis

The statistical analyses in this work are Bayesian. Bayesian inference comes in three forms:

- Parameter Inference (Section 1.4.1), where a model is assumed true and we desire that model's best fit to the data.
- Model Selection (Section 1.4.3), where we have (at least) two competing models. Odds are calculated for which model best fits the data.
- Model Averaging (briefly discussed in C.2.3), where we accept that no single model is as descriptive as the data. We then look to see what can be learnt by combining the posteriors from all the models for each parameter. We are left with a total uncertainty on the parameter under question.

1.4.1 Parameter estimation

The key to answering any problem in a Bayesian framework is summed up by the following bullet points:

- What are the data?
- What is the model?
- What are the model's parameters?
- What is the likelihood function?
- What prior should be used?

To approach this we write, D, as data and M, the model which has a parameter vector, $\boldsymbol{\theta}$. We desire the distribution of our parameters (the posterior), given the data set at hand and the model we are using. We can then rewrite Bayes' theorem (Equation C.1) as,

$$P(\boldsymbol{\theta}|D, M) = \frac{P(D|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M)}{P(D|M)}, \tag{1.30}$$

where each of these four terms plays a fundamental part of the Bayesian process. $\mathcal{P} = P(\boldsymbol{\theta} \mid D, M)$ are the parameter posterior distributions; $\mathcal{L} = P(D \mid \boldsymbol{\theta}, M_i)$ is the likelihood; the prior distribution $\Pi = P(\boldsymbol{\theta} \mid M_i)$ is our knowledge from before this experiment; and the Bayesian Evidence, $\mathcal{Z} = P(D \mid M_i)$ (also known as the marginal or model likelihood). Giving $\mathcal{P} = \mathcal{L}\Pi/\mathcal{Z}$, for short. With this laid out we can decide what we really want from parameter estimation - i.e. the MAP (maximum a posteriori) parameters and their variance.

To briefly touch on the prior distribution, before discussing it in depth in Section 1.4.4, we can differentiate Bayes theorem with the chain rule as,

$$\frac{\partial \mathcal{P}}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\frac{\mathcal{L}\Pi}{\mathcal{Z}}\right] = \frac{\Pi}{\mathcal{Z}}\frac{\partial \mathcal{L}}{\partial \theta} + \frac{\mathcal{L}}{\mathcal{Z}}\frac{\partial \Pi}{\partial \theta} = 0, \tag{1.31}$$

since the Evidence has had its parameter dependence marginalised out (discussed further in Section 1.4.3). Rearranging, and inputting a uniform prior distribution gives,

$$\frac{\partial \mathcal{L}}{\partial \theta} = -\frac{\mathcal{L}}{\Pi}\frac{\partial \Pi}{\partial \theta} = 0, \tag{1.32}$$

hence motivating the peak finding exercise in Section C.1.5. For uniform priors the peak of the posterior is the peak of the likelihood. This shows that in a parameter estimation context, the likelihood takes the fundamental role.

1.4.2 Likelihoods

Before diving in to discuss the intricacies, it is worth mentioning a few conventional traits that are taken advantage of regularly. Namely, since parameter estimation is a peak finding exercise of the likelihood (given uniform priors, see Section 1.4.4), we lose no information dealing the log of $\mathcal{L}$ rather than $\mathcal{L}$. In a lot of scenarios this is significantly more convenient.

Second is the form of the likelihood, which is taken to be Gaussian since we are often dealing with Gaussian errors due to the CLT (Section C.1.3). Conventionally a χ^2 is used to implement this. For application in astronomy, x is a measure of the simulated data set, (simulated with parameter θ), compared with the same statistical measure applied to our data set x_D, which

has an observation with error σ_D. We then have,

$$\ln\mathcal{L} \sim -\frac{(x - x_D)^2}{2\sigma_{\rm D}^2}. \tag{1.33}$$

The statistic used for x is the choice of the user and this is often a non-trivial choice. It may not be obvious how best to interpret measurements depending on the nature of the observing technique, the desired observed quantity, or both. One must be particularly careful not to confuse the likelihood statistic as the probability of hypothesis validity. We need to be able to interpret the chosen statistic in a way that enables proper interpretation of the desired hypothesis. For example with a Gaussian likelihood, the mean (and associate variance), may be all that is necessary. Any deviations from perfect Gaussianity however would require higher order statistical moments. Leaving these out could limit the analysis, even if the information exists within the data set.

The Anderson-Darling test (Anderson & Darling, 1952) would be another potential likelihood (and is used in Section 1.5.2 within the MultiNest algorithm later on). It is useful for testing a distribution, f, by comparing its cumulative distribution functions against a set of points x. Since the integration of f is required it is less useful for observational science as data sets cannot always be written analytically. In this case one would look to minimise A^2 in the form,

$$A^2 = -N - \sum_i^N \frac{2i-1}{N}\left\{\ln\left[\int f(x_i)\right] + \ln\left[1 - \int f(x_{n+1-i})\right]\right\}, \tag{1.34}$$

where the value for success is defined by a user defined confidence interval.

Statistical functions are often defined by moments of the pdf or poly-spectra, if more detail is required. Single point moments are defined as,

$$\langle x^{\rm n}\rangle = \int x^{\rm n}{\rm pdf}(x)dx, \tag{1.35}$$

where $n = 1,\ 2,\ 3$, and 4 represent measurements of the mean, variance, skew, and kurtosis (tailed-ness) of the distribution. In some real scenarios these moments will lose information (similar to cutting off a Taylor expansion at the first order), and hence the use of poly-spectra becomes necessary.

The more complex (typically the more non-Gaussian) the distribution the higher order moment is necessary to adequate describe the underlying features of the data set. The Fourier transform of the n-point correlation function are known as poly-spectra (discussed in detail in Section C.4). The power spectrum, bi-spectrum and tri-spectrum are respectively defined for $n = 2, 3$, and 4, and it follows from the moments that each is specifically aimed at looking towards contributions within the target distribution's variance, skew, and tail respectively. Chapters 2, 3, 4 and 5 all make use of the power spectrum directly.

Now not every situation can have these statistical moments formulated, (the example in Section C.1.5 shows this via the Cauchy distribution) but as we have established in Section C.1 - there is no one tool for answering statistical questions, familiarisation with the hypothesis under question is key.

1.4.3 Model selection

A full approach to model selection requires us to rewrite Bayes theorem (Equation C.1) again, but this time with a posterior for the model in mind,

$$P(M_i|D,T) = \frac{P(D|M_i,T)P(M_i|T)}{P(D|T)}. \tag{1.36}$$

Here the Bayesian Evidence $\mathcal{Z} = P(D|M_i,T)$ takes the roll of the likelihood (i.e. measuring how likely it is that model i has produced the data set at hand). T describes the theory space that the model is derived from. The details of T need not be a concern for implementation, it is included for completeness. The Bayesian Evidence,

$$\mathcal{Z} = \int \mathcal{L}\ \Pi\ d\theta, \tag{1.37}$$

takes the central role in model selection. Bayesian model selection can be summarised as the exercise of finding M_i so that $\mathcal{Z}$ is maximised.

Notice that, in comparison to parameter estimation, we must now be more careful with the use of the prior distribution since $\mathcal{Z}$ includes Π in its definition in Equation 1.37. The result in model selection is always dependent on the prior distribution. A particularly useful aspect of the Bayesian Evidence is that it rewards models for how well they predict data. The form of the

prior in this integral implies that similarly redundant parameters are inherently penalised. A concept known as Occam's Razor[5] (Ockham, 1318). Unless the more complex model constrains the posterior distribution noticeably better than the simpler model, the extra parameter will acquire a penalty for containing parameter space that is redundant. The presence of Occam's Razor is implicit in Bayesian model selection, due to the factor of $1/V_{\text{prior}}$ within Equation 1.37. If implemented correctly, this is one of the major benefits of testing models with this methodology.

1.4.4 Priors

Returning to the general picture, the prior distribution of each parameter provides the opportunity to input user knowledge into the hypothesis at hand. If the bounds of the parameter priors cannot be meaningfully set, the physics that has gone into the model is not detailed enough. One must be absolutely certain these are valid, or they are capable of skewing the results.

We have just shown that the Bayesian Evidence is the key to model selection, and in order to calculate the Evidence the prior-likelihood product must be integrated. It therefore follows that the choice of prior has more potential to skew results here than in parameter estimation scenarios.

Choosing an uninformative prior is usually the best way to go. This way it is comparatively easy to interpret the influence of the data, and the risk of the user influencing the results will be minimised. Conventionally this can be done in two forms:

- Uniform - $p(\theta) = \text{constant}$,
- Log - $p(\theta) \propto \frac{1}{\theta}$.

Where the latter is also known as the Jeffreys' prior, and is used when parameters scale across multiple orders of magnitude (when uniform distributions stretch orders of magnitude, significantly more parameter space exists towards the larger end).

[5]William of Ockham (1287-1347) was an English Franciscan friar, theologian and philosopher. His text is originally in Latin and can be translated as: 'More things should not be used than are necessary'.

In Section 1.4.1 we have showed that the use of a uniform prior distributions, $\Pi = \text{const}$, simplifies Equation C.1 to,

$$\mathcal{P} \propto \mathcal{L}. \tag{1.38}$$

The Evidence is a marginalisation of the likelihood times the prior, so it is also a constant with respect to the model's parameters. In this case the integration of the likelihood can simply be scaled by the parameter space volume V_{prior}, however this is not entirely fool proof in model selection as two similar competing models may have their results swayed if the priors are not chosen fairly.

If we briefly return to the context of our quantitative example in Section C.1.5, Equation C.11 has already made use of an uninformative uniform prior (via the proportionally signs, similar to Equation 1.38). Since we desired the MAP parameters and their variance, this problem was simplified to measuring the maximum of the likelihood and its variance due to our choice of prior. It is not always possible to use completely uninformative priors. In Section C.1.5, the lighthouse could have been anywhere along the coast, i.e. between $[-\infty, \infty]$ however this would nullify any influence of the likelihood as the pdf contribution from Π would be infinitesimal. Here, and in most cases, a sensible cutoff should be applied by the user. By sensible we mean absolutely justify-able i.e. in the case of the lighthouse, at some point the curvature of the Earth would get in the way or the luminosity of the source would diminish beyond detection. This further illustrates the point that user discretion must be used, but above all else the data must overwhelm the priors.

In practice, it is worth testing[6] how robust the results are with respect to a wider Π distribution and δ-function in order be aware of the potential skew a poorly chosen prior can inflict. For the case $\Pi = \delta(\theta - \theta_{\text{MAP}})$, we are taking the limit $d\theta \to 0$ around the MAP value,

$$\mathcal{Z} = \frac{1}{V_{\text{prior}}} \int_{\theta_{\text{MAP}} - \frac{d\theta}{2}}^{\theta_{\text{MAP}} + \frac{d\theta}{2}} \mathcal{L} d\theta. \tag{1.39}$$

As the the prior volume is reduced $V_{\text{prior}} \to d\theta$, resulting in,

$$\lim_{V_{\text{prior}} \to 0} \mathcal{Z} = \left[\mathcal{L}(\theta) \frac{d\theta}{V_{\text{prior}}} \right]_{\theta_{\text{MAP}}} = \mathcal{L}(\theta_{\text{MAP}}), \tag{1.40}$$

[6] We do this as a test to the methodology on the toy EoR models in Chapter 3.

i.e. the likelihood evaluated at the MAP is the maximum Bayesian Evidence obtainable.

One way of dealing with the possible variety in prior distributions is to define an Evidence with multiple different priors per model likelihood (Martin et al., 2014). Then the Prior is a characteristic of the model (in this context) and the Bayesian Evidence (per model) is dependent on integrating the prior-likelihood combination as before.

When experiments iterate it can be sensible to input the posterior from a previous experiment as the prior for the next iteration. For example, fits to the CMB Planck data can use the WMAP results as a prior since they follow similar methodologies. This is a simple subset of Hierarchical Bayesian modelling. In general, Bayes theorem is adapted to contain latent variables of the model at hand to avoid measurement of parameters that are hard to obtain,

$$p(\theta|s,t) \propto \int p(s,t,x,y|\theta)p(\theta)dxdy. \tag{1.41}$$

For unknown parameters s and t, where the known (latent variable) parameters are x and y. These must therefore be marginalised out during Equation 1.38. A typical example of this is obtaining photo-metric redshifts from multiple different colour plates. The distribution obtained by each colour must produce a joint flux distribution when the plates are marginalised over. This is a powerful tool, since Bayes theorem need only be marginalised to infer information about parameters that indirectly influence observed data sets. In most cases (e.g. Section C.1.5) the prior distinguishes the frequentist estimator distribution from the Bayesian posterior. Here it is not so simple, the MAP is not necessarily the same as the maximum likelihood (e.g. if each separate distribution has a uniform prior, the weighted sum may change this for the joint distribution). The inclusion of multiple distributions can be easily written down for the Bayesian case (as we have done, Equation 1.41), however for the frequentist more information is necessary to do so. A rigorous application of the frequentist approach in this context is beyond the scope of this work.

The interested reader should refer to Jaynes (2003) which contains quantitative examples of how priors can move posterior distributions. Examples of applying Bayesian methods within cosmological contexts are available in Hobson et al. (2009).

1.4.5 The Bayes factor

As we have established in Section 1.4.3, the Evidence, $\mathcal{Z} = p(D|M)$, takes the key role in the Bayesian model selection. However we have not discussed how it is used formally or how to interpret the results. In order to proceed we derive what's known as the Bayes factor from the ratio of two different model-posteriors (Equation 1.36) both fitting the same data set,

$$\mathcal{B}_{12} = \frac{P(M_1|D,T_1)}{P(M_2|D,T_2)} = \frac{P(D|M_1,T_1)P(M_1|T_1)}{P(D|M_2,T_2)P(M_2|T_2)}, \tag{1.42}$$

here the models are not necessarily derived from the same theoretical motivations, however the observed data should be independent of the theoretical motivations and will therefore cancel. We are left with the ratio of the Bayesian Evidences and a model-prior where the user can implement how likely they think a model is representative of a data set. Throughout this work we assume all models have an equally fair chance, i.e. they have access to similar physics. We therefore simplify Equation 1.42 and define the Bayes factor to be,

$$\mathcal{B}_{12} = \frac{\mathcal{Z}_1}{\mathcal{Z}_2}. \tag{1.43}$$

The Bayes factor is then simply interpreted as a betting odds. In the form of Equation 1.43, $\mathcal{B}_{12} > 1$ (< 1) represents model 1 (2) having a better chance at reproducing the data and is therefore considered favoured.

1.4.6 The Jeffreys' scale

For the non-trained statistician, interpreting the odds produced by the Bayes factor analysis can be non-trivial. To ease interpretation we adopt the Jeffreys' scale as prescribed in Hobson et al. (2009). Namely,

- *Strong* - $\mathcal{B}_{12} > 150$ then model 1 has objectively outperformed model 2 at describing the data at hand.
- *Moderate* - $10 < \mathcal{B}_{12} < 150$ care must be taken to assure no skew is added via redundancy in the prior ranges, however the two models are likely to be distinguishable.

- *Weak* - $\mathcal{B}_{12} < 10$ the two models are indistinguishable given the quality of data.

We emphasise that this is ad hoc and that the true results are the odds. Despite any user's best efforts, there will be some leeway in choosing the prior distribution. Therefore it is wise to require the results reach a threshold in order to draw justifiable conclusions.

1.4.7 The Savage-Dickey density ratio

The Savage-Dickey density ratio (SDDR) is defined as the ratio of a model's Bayesian Evidence to that of the same model but with a certain parameter fixed (Dickey, 1971). This sub-model (with the parameter fixed) is referred to as 'nesting'. By varying the nested parameter throughout its allowed prior range we can gauge that parameter's impact on the model under scrutiny. In chapters 3 and 4 this is used as an analysis of how effective a parameter, θ_*, is in a model selection setting. Thinking of the SDDR as a parameter dependent Bayes factor, we can write,

$$\mathcal{B}(\theta_*) = \frac{\mathcal{Z}(\theta_*)}{\mathcal{Z}}, \tag{1.44}$$

where the chosen nested (fixed) parameter is labelled, θ_*. Using our definition of the Bayesian Evidence in Equation 1.37 we can expand Equation 1.44 into,

$$\mathcal{B}(\theta_*) = \frac{1}{\mathcal{Z}} \int \mathcal{L}(\boldsymbol{\theta})\Pi(\boldsymbol{\theta}) d\theta^{N-1} \tag{1.45}$$

$$= \int \mathcal{P}(\boldsymbol{\theta}) d\theta^{N-1} \propto \mathcal{P}(\theta_*) \ .$$

Here $d^{N-1}\theta$ refers to the differential $d\theta^N$ without the $d\theta_*$ term. To reiterate $d\theta_1 ... d\theta_N \equiv d\theta^N \equiv d^{N-1}\theta d\theta_*$. As can be seen on the right hand side of Equation 1.45, the SDDR shares the shape of the parameter posterior for θ_*. If this is not the case, the sampler has either incorrectly integrated the parameter space or the incorrect posterior distribution has been found. Plotting the profile of the likelihood is significantly easier to do than calculating the Evidence (and therefore the SDDR), and for a single parameter the computational overhead of a grid-search isn't expensive compared to sampling. A simple cross check for the sampler is then easily performed to see if this profile and the posterior match (in the case of uniform priors). If they don't the sampler has incorrectly integrated the likelihood space.

When the priors and likelihoods are completely separable, the SDDR can also be used to reduce computation time, since only the Evidence of the models with the larger number of parameters needs evaluating[7].

A novel method developed by Mootoovaloo et al. (2016) attempts to use the SDDR more broadly for model selection. This can be done by creating what is referred to as a supermodel, where hyperparameters are used to select the constituent model within the supermodel. Unfortunately in our context computation is the main overhead, which worsens with increasing dimensionality. We do not consider it further as adding this would increase the overhead, out-weighing any added convenience.

1.5 Applying MCMC algorithms

Throughout this work the posteriors shown will be produced by one of the two following algorithms. Emcee (Foreman-Mackey et al., 2013) is the conventional algorithm of choice for parameter estimation. As MCMC algorithms go it is very robust for the majority of likelihood spaces, however it does not calculate the Bayesian Evidence. In cosmology, the well known CosmoHammer (Akeret et al., 2013) is a framework which implements Emcee for use on large computing clusters when combining multiple experiments into the likelihood. 21cmMC is the combination of CosmoHammer and 21cmFast (discussed in Section 1.3). MultiNest is the other algorithm, geared at model selection. The bulk of Chapters 3 and 4 is the application of MultiNest in place of CosmoHammer within 21cmMC. In brief, MultiNest efficiently integrates the likelihood space to obtain the Bayesian Evidence.

The following subsections detail these two algorithms. Since the use of implementing Emcee is well established and that MultiNest in the EoR is the bulk of chapters 3 and 4, there is significantly more discussion of this sampler and the variation of its possible implementations.

[7]For more information on this please see Hobson et al. (2009). This was not used as we found in Chapter 2 that the EoR models have correlation between the parameters and are therefore non-separable.

1.5.1 EMCEE

EMCEE (Foreman-Mackey et al., 2013) is an affine invariant ensemble sampler. The ensemble refers to a group of so called 'walkers', where each walker is a chain of samples similar to individual M-H algorithms (Section C.2.1). Every time a walker updates, it does so dependent on the other walkers within the ensemble. The walkers are allocated into pairings at random and it is along this bisector that the walker will move in order to create the next sample in that walker's chain.

Affine invariant means the method does not depend on the geometry of the likelihood space i.e. it is insensitive to linear transformations and unaffected by awkwardly shaped parameter covariance distributions (Goodman & Weare, 2010). To illustrate this, the example likelihood,

$$p(\theta) \propto \mathbf{e}^{\frac{-(\theta_1-\theta_2)^2}{2\epsilon} - \frac{(\theta_1+\theta_2)^2}{2}}, \tag{1.46}$$

is very ellipsoidal in shape. The narrow nature of this means the M-H algorithm's probabilistic acceptance step would become very inefficient and is likely not to explore the parameter space fully. In order for it to succeed, a transformation of the form $\phi_1 = (\theta_1 - \theta_2)/\sqrt{\epsilon}$ and $\phi_2 = \theta_1 + \theta_2$ is necessary to resolve this as the covariance of the new distribution is diagonalised and unitary. EMCEE implements this by updating the point under consideration ($X_k(t)$, assigned to walker k) along its bisector to the randomly selected point (X_j, assigned to a different walker, j, from the same ensemble). The authors refer to this as the 'stretch' move and it captures how EMCEE performs so well on troublesome likelihood spaces.

Algorithm 3 details the steps in the process, where the adjustable scale parameter, a, dictates the size of the step S. Walker positions are x, the updated step position is Y, t denotes the iteration (trial point), and d is the number of dimensions.

Algorithm 3: EMCEE

Distribute K walkers uniformly throughout the Prior;

while *Convergence unsatisfied* **do**

  Select half the walkers at random;

  for *k in K/2 walkers* **do**

    draw the step size $S \in [\sqrt{a}, 1/\sqrt{a}]$;

    draw random walker j (from the other half); $x_k(t) \rightarrow Y = x_j + S \times [x_k(t) - X_j]$;

    Define the acceptance criteria $q = S^{d-1}\Pi(Y)\mathcal{L}(Y)/\Pi(x_k(t))\mathcal{L}(x_k(t))$;

    Draw random number r, form [0,1];

    if $q \leq r$ **then**

      $x_k(t) = Y$;

    else

      $x_k(t) = x_k(t+1)$;

    end

  end

end

In essence the EMCEE algorithm is similar to M-H but with the stretch move implemented on each iteration for half the walkers. Half the walkers must be used to update the ensemble if performing the algorithm in parallel, otherwise detailed balance is violated and the algorithm's performance varies on its selection of walker j. Detail balance is insured by updating the walkers half at a time, and the use of the probabilistic acceptance value q. In other words as the ensemble updates the covariance distribution it also converges on the posterior. Detailed balance is only satisfied generally for the steps once the covariance is near unitary, and the walkers have populated the posterior peak.

For usage EMCEE recommend[8], $a = 2$, and the number of walkers must be increased alongside dimensionality for it to work well. Within 21CMMC $K = 16 \times d$ is recommended with 3000 iterations (and 250 discarded burn-in iterations) to satisfy the Gelmin-Rubin convergence check (Section C.3.2). The 21CMMC settings: walkersRatio=16, burninIterations=250, sampleIterations=3000, threadCount=20, reuseBurnin=False; prove to be suitable.

[8]Please see https://emcee.readthedocs.io/en/stable/ for the full API.

1.5.2 MultiNest

MultiNest is an MCMC implementation of nested sampling, an algorithm designed to calculate the Bayesian Evidence. The tricky part of this is evaluating iso-likelihood contours (discussed in Section C.2.1). Ellipsoidal rejection sampling is the tool of choice. But this alone would bias the likelihood space to the shape of an ellipsoid and so a number of ellipsoids are fit and sampled within. To choose the number of ellipsoids a variant on k-means clustering is implemented and is detailed below. How MultiNest deals with multimodal likelihood spaces and how the statistical errors are calculated are also detailed here. Extensive detail of MultiNest's usage and some preliminary tests on few toy models are shown in Section C.3.1. The appendices also detail some features within the MultiNest code that we do not implement.

Ellipsoidal rejection sampling

Within MultiNest the iso-likelihood contours are approximated with multiple N-dimensional ellipsoids. The optimum number of ellipsoids is fit via the dinosaur algorithm in Section 1.5.2. For now lets assume we are fitting one ellipsoid to a number of (live) points labelled, $\boldsymbol{x}_i$, which are part of the set, $S = \{\boldsymbol{x}_1, .., \boldsymbol{x}_n\}$ containing N points, please see Mukherjee et al. (2006) for an example implementation of nested sampling in this way. Their covariance is defined in the normal way with mean μ,

$$\mathcal{C} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T, \tag{1.47}$$

the ellipsoid is then approximated as,

$$E = \boldsymbol{x}^T(f\mathcal{C}^{-1})\boldsymbol{x}, \tag{1.48}$$

where f is a buffer factor used to increase the sampling efficiency in Section C.3.1. This is then the proposal distribution for the rejection sampling ($E \leq 1$), which proceeds as in Algorithm 4.

Algorithm 4: Rejection Sampling

Pick a point x within Π;

Sample a random number $r \in [0, max(\Pi\mathcal{L})]$

(where the maximum is dictated by the live points in the ellipsoid);

if $\Pi(x)\mathcal{L}(x) > r$ **then**
  accept x;

else
  reject x;

  repeat;

end

Since multiple ellipsoids are fit (see 1.5.2) the acceptance probability of each sample is penalised $1/n_e$ (n_e being the number of ellipsoids) in order to account for the sampling of overlapping ellipsoids.

The ellipsoidal fit of the isolikelihood contours is not perfect and this worsens in higher dimension. As the isolikelihood contours contract on the posterior the sampling efficiency lowers. Multiple sampling attempts are often required in the latter stages of the algorithm, and this effect scales geometrically with dimension. For $D > 30$ this methodology becomes too inefficient to converge and alternatives (such as those in Section C.3.3) are required. The authors recommend the switch to PolyChord for $D > 10$.

The dinosaur algorithm

MultiNest does not use one ellipsoid to fit the live-points. The number is chosen by a combination of algorithms to best fit the shape of the likelihood space. The basis of this is called k-means clustering (Hartigan & Wong (1979) detailed in Algorithm 5), which determines k subsets within the set of live points. These subsets of the live points each have an ellipsoid fit to them and the total volume of these is used as the rejection sampling proposal function in the previous section.

Determining how many subsets of data to fit (or what k to use) is a non-trivial question. Here a method called an expectation minimisation scheme (Shaw et al., 2007; Choi et al., 2006) is used.

Algorithm 5: K-means Clustering: fitting K distributions within a data set.

Pick **k** locations in the parameter space;
Assign each data point to its nearest neighbouring **k** location;
Move each **k** position to the mean of each assigned set of data points;
while $\Delta\mathbf{k} \neq 0$ **do**
    Reassign the data set to each nearest neighbouring k location;
    Recast each k location to the mean of each assigned data set;
end

Once an ellipsoid is fit we test to asses the quality of that fit, and in particular whether two ellipsoids would provide a better fit. First lets consider the true volume of the isolikelihood contour we are trying to replicate $V_{\rm true}$. We must define a metric, T to test how well we have replicated the truth with the ellipsoids,

$$T \equiv \frac{1}{V_{\rm true}} \sum_{j=1}^{N_s} V(E_j), \tag{1.49}$$

where the volume of the ellipsoidal fit in Equation 1.48 becomes $V(E_j) \propto \sqrt{f\mathcal{C}_j}$ for the jth ellipsoid. Now we seek ΔT between a k and k+1 iteration of the ellipsoidal fit. To approximate this we say $V(S) \propto\!\!\!\sim V(X)$ since the full likelihood space will exist inside the suitably chose prior distribution, where $V(X)$ is the volume related to the prior fraction X (assuming a uniform prior). We then weight the volume ratios to add some accountability for a poorly fit ellipsoid,

$$\Delta T_{k,k+1} \propto\!\!\!\sim \left(\frac{V(E_k) d_{\rm k}}{V(X)} - \frac{V(E_{k+1}) d_{k+1}}{V(X')} \right), \tag{1.50}$$

where the Mahalanobis distance metric, $d_k = (\boldsymbol{x} - \boldsymbol{\mu}_k)^T (\mathcal{C}_k)^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k)$, is the aforementioned weight, and X' is the prior volume fraction for the set of points where the k+1 ellipsoids have been fit. Equation 1.50 is particularly subtle and must be calculated for each combination of livepoints assigned within the E_k ellipsoid. The assignment of points to each half of the split ellipsoid is tested intrinsically by including the Mahalanobis distance in Equation 1.50. At each iteration, each half is tested against being split again, whilst the assignment of points within each half is optimised. The split of the kth ellipsoid is kept if the sum of the volumes of these two new ellipsoids becomes less than the volume of the previous ellipsoid or the separation of the two new ellipsoids is below a different user defined value (in order to prevent overlap of modes). This is a costly algorithm, and for optimisation only likelihood spaces that score $T \geq 1.1$ are partitioned using the above scheme.

To try and explain a simpler application, we have one ellipsoid, E (found within prior volume X) which is being tested against a split into two, E_1 and E_2. To test E_1 and E_2, points are assigned to each ellipsoid as to minimise Equation 1.50. If $V(E_1) + V(E_2) < V(E)$ or $V(E) > 2V(X)$ the ellipsoidal split is kept, and the test is repeated on each new ellipsoid.

The expectation minimisation scheme, k-means clustering, and ellipsoidal rejection sampling are packed together (along with the algorithms in the appendix) and labelled the dinosaur (or D-means) algorithm within MultiNest.

Multi-modality

Once the posterior is well sampled, so that each mode is approximately established, an algorithm created by Alfano & Greer (2003) is implemented to test whether the ellipsoids overlap. First we consider two neighbouring ellipsoids fit via Equation 1.48 which have separate sub sets of the total live points (labelled j, k). First we rewrite the ellipsoid equations with a shared matrix, M as,

$$0 = M^T(f\mathcal{C}_j^{-1})M, \qquad \text{and} \qquad 0 = M^T(\lambda f\mathcal{C}_k^{-1})M, \tag{1.51}$$

where λ scales the ellipsoid without a loss of generality. These are rearranged and factorised to give the eigenvector equation,

$$M\mathcal{C}_j^{-1}(\lambda I - \mathcal{C}_j\mathcal{C}_k^{-1})M^T = 0, \tag{1.52}$$

where λ is selected to produce a singular bracketed matrix. This recasts the problem as the intersection between two quadratic surfaces. If X exists in a way that produces real solutions then the two ellipsoids must share parameter space and therefore intersect.

When multimodal = $True$, this process is iterated in parallel with the sampler. The first live point fit (with 1 ellipsoid) is labelled the active group and as the sampler proceeds any splits from this are labelled as independent modes until there are no live points left in the active group. These are then integrated using the trapezium rule as (Equation C.23) and summed per mode, m, with an adjusted weighting, $w_i(X_{i-1} - X_i)$, so that the ratio of the prior volume used incorporates the split from the active group (i.e. G_B and G_A are the numbers of live points in the group before and after the split respectively), and each mode m will contain a different N_s.

These adjustments to the trapezium rule are expressed as,

$$\mathcal{Z} \approx \sum_m \frac{1}{2} \left[\sum_{i=1}^{N_s-1} (X_{i-1} - X_{i+1}) \mathcal{L}(X_i) \frac{G_A}{G_B} \right]_m . \tag{1.53}$$

Note that the sum will stack contributions in the order that the modes are found. Since this is done in parallel it incorporates the factor based on the prior volume in the active group from the iteration before the subdivision.

Uncertainty

The statistical uncertainty on the Evidence comes from stochasticity between the points and from the use of the trapezium rule. For the trapezium rule, the error scales as the square inverse of iterations and is therefore easily reduced. To proceed estimating the stochastic noise we calculate the negative relative entropy,

$$H = \int \ln \frac{d\mathcal{P}}{dX} dX, \tag{1.54}$$

allowing us to write $X \approx e^{-H}$. Inputting Equation C.23 and Bayes theorem gives,

$$H \approx \frac{1}{2} \sum_i^{N_s-1} \frac{\mathcal{L}(X_i)}{\mathcal{Z}} (X_{i-1} - X_{i+1}) \ln \frac{\mathcal{L}}{\mathcal{Z}}. \tag{1.55}$$

Now $\ln X_i$ depends on the iteration as $(-i \pm \sqrt{i})/N_s$, assuming Poisson variability. From our definition of relative entropy we directly have this as equivalent to $N_s H \pm \sqrt{N_s H}$. We can see that as N_s scales it will be the ratio of $\ln X_i$ and N that drives the relative entropy down, resulting in,

$$\Delta \ln \mathcal{Z} = \pm \sqrt{\frac{H}{N}}. \tag{1.56}$$

Therefore, as one should expect, the stochastic error decreases as the number of samples increases. Convergence is achieved when enough samples are gathered so that the stochastic error has become negligible.

1.6 Summary & Outlook

In Section 1.2, we covered astrophysical detail of the EoR and cosmic dawn. We defined two toy scenarios of reionisation morphology and discussed how the properties of the first galaxies will depend intricately on the masses of the first (Pop-III) star formation. The 21cm signal has now been introduced as a promising way to trace events throughout cosmological history. If supplementary material is required, we present an introduction to cosmological theory in Appendix A.1. Appendix A.2 shows observations relevant to cosmology and highlights when cosmological observations need improving. This appendix also contains more detail on observations of the red Ly-α damping wing, the Gunn-Peterson trough and the Planck reionisation optical depth - these are referred to as the observational priors in later chapters. For a derivation of the 21cm brightness temperature, detail about what astrophysical phenomena contribute towards the 21cm spin temperature, and the mechanism behind the hydrogen 21cm signal can be found in Appendix A.3.

Section 1.3 introduced the semi-numerical recipes used in 21CMFAST. By implementing the excursion set formalism to identify ionised regions of the IGM on a simulated density and velocity fields, we can simulate the 21cm power spectrum for different reionisation scenarios fast enough for statistical analyses. 21CMMC unites 21CMFAST with EMCEE to enable parameter estimation in the EoR. Detailed improvements of the basic 21CMFAST model have been described chronologically - including: simulating the light-cone by joining coeval cubes; including spin temperature fluctuations; and including inhomogeneous recombinations in the IGM. Further detail will be provided when relevant adaptations are made to the basic 21CMFAST model described here (particularly when varying morphology in Chapter 3, and when including X-ray heating and UV luminosity functions in Chapter 4). The foreground avoidance code 21CMSENSE is also introduced as well as a detailed description of three telescopes - LOFAR, HERA and SKA. For an introduction to radio observations and observing the 21cm power spectrum with radio interferometers, see Appendix B. We have introduced the methods used to simulate cosmological observations, we then turned to analysing them.

We have looked at how to construct a Bayesian analysis in Section 1.4, including an introduction to likelihoods and priors. Parameter estimation has been framed as finding the peak of the likelihood function, and we have use a χ^2 as the likelihood function (justified in Appendix C.1

by the derivation of the CLT). In model selection, the Bayesian Evidence $\mathcal{Z} = p(D|M)$ is key and must be maximised. We also discuss the importance of choosing carefully the parameter prior in order to avoid skew within model selection. To implement model selection we use the so called Bayes factor (Section 1.4.5), an odds ratio determining which of two competing models has higher odds at producing an observed data set. The Jeffreys' scale is conventionally used for interpreting these odds ratios. Even though the Jeffreys' scale is a bit ad hoc it is a good starting place. Ultimately familiarisation with the problem at hand is necessary to accurately assess the odds. The Bayes factor can also be used for analysing parameters with the SDDR. Please see Appendix C.2 for supplementary material on Bayesian analyses, including how to implement these techniques with MCMC algorithms. Statistics is a powerful tool in astrophysics and cosmology, the development of Bayesian statistics has allowed a robust analyses for model selection and parameter estimation when applicable. For an introduction to Bayesian techniques and specific information of the role of the power spectrum in cosmology, see Appendices C.1 and C.4 respectively.

Finally, we have introduced a statistical toolbox in Section 1.5. EMCEE (Section 1.5.1) and MULTINEST (Section 1.5.2) are our tools of choice for parameter estimation and model selection respectively. One, or both of these are used in each subsequent chapter. For supplementary material about convergence tests and the implementation of MULTINEST please see Appendix C.3.

In the next chapter we turn to applying Bayesian model selection the EoR.

Chapter 2

The Bayesian Evidence Calculated from MCMC chains

2.1 Introduction

Here we experiment with calculating the Bayesian Evidence directly from MCMC chains. In Chapter 1 we discussed the importance of the Bayesian Evidence for model selection and we presented MultiNest, a nested sampling algorithm designed for calculating the Evidence. We also introduced 21cmMC, an MCMC analysis tool for fitting parameters within 21cmFast by implementing Emcee. Since the literature already contains an MCMC parameter estimation framework for the EoR, calculating the Evidence directly from these chains would be a useful cross check for when we apply MultiNest in Chapters 3 and 4. We also know that MultiNest performs significantly worse in high dimensions than Emcee and it would therefore be useful for the future, when the number of parameters within models like 21cmFast are expanded to capture more complex astrophysics.

In this chapter we implement MCEvidence (Heavens et al., 2017a) which uses the number density of a converged MCMC chain to estimate integration of the target posterior. A properly converged MCMC chain will have a density of points that represents the target posterior. Given uniform priors this will match the likelihood and be the Bayesian Evidence we desire.

Heavens et al. (2017b) shows the success of the methodology in a cosmological context, favouring

$\Lambda-$CDM over other models when using the CMB data from Planck. We apply this methodology to the EoR.

Section 2.2 details the MCEvidence method; Section 2.3 compares analytic solutions of toy models to both MultiNest and MCEvidence; Sections 2.4 and 2.5 apply MultiNest and MCEvidence to toy EoR models before we conclude in Section 2.6. The toy EoR models implemented in this chapter are labelled 3pFZHf1 and 4pFZHf1 (Furlanetto et al., 2004; Mesinger et al., 2011) (as used in Chapter 3). Chains in this section are run with 32 walkers, producing 320000 samples in total. Long chains are needed here to cater for thinning, which we will discuss. The usable samples from each chain are slightly lower than this as some points in the Emcee posteriors form islands away from the posterior peak (described in more detail in Section 3.4.1). Before the MCEvidence calculation is performed, we removed these islands from the chains by only using points with likelihood, $\ln\mathcal{L} > -10$. Throughout this chapter, we compare the results from MCEvidence with the output of MultiNest. In Appendix C.3.1, we found that integration with MultiNest is more efficient and as accurate as integrating analytically. We include the analytic result also for the toy models in Section 2.3 but for efficiency reasons, we only compare with MultiNest when using the toy EoR models in Sections 2.4 and 2.5.

2.2 MCEvidence

Using the conventions from Section 1.4, we follow closely the derivation of $\mathcal{Z}$ from Heavens et al. (2017a).

We begin relating the number density of points in a selected MCMC chain, n, to our target posterior distribution as,

$$n(\mathbf{\Theta}|\mathrm{D},\mathrm{M}) \propto \mathcal{P}(\mathbf{\Theta}|\mathrm{D},\mathrm{M}) \propto \tilde{\mathcal{P}}(\mathbf{\Theta}|\mathrm{D},\mathrm{M}), \tag{2.1}$$

where the tilde represents an unnormalised distribution; $\mathbf{\Theta}$, D, and M represent the parameter vector, data, and model respectively. By representing the 'N' points in the chain with Dirac δ-functions we can express the Evidence as,

$$\mathcal{Z} = a \int n(\mathbf{\Theta}|D, M) d\Theta = a \int \sum_{i=1}^{N} \delta(\mathbf{\Theta} - \mathbf{\Theta_i}) = aN. \tag{2.2}$$

We need only be concerned with finding the proportionality constant, a - relating the number density to the unnormalised posterior. In particular we desire a MAP value for, a, given our distribution of points, n. To proceed we assume the points in the chain are not correlated. Due to the MCMC methodology, this is known to be false to various degrees (depending on the algorithm) but with adequate thinning of the chain, this is a satisfactory approximation.

Heavens et al. (2017a) use a kth nearest neighbour method to approximate the target distribution with Bayes theorem. To avoid a dimensional nearest neighbour distance we transform the nearest neighbour vector $\mathbf{x}_k$ from the original parameter space to r_k, the Mahalanobis distance (also known as a pre-whitening step),

$$r_k = \sqrt{(\mathbf{x_k} - \bar{\mathbf{x}})^T \mathcal{C}^{-1} (\mathbf{x_k} - \bar{\mathbf{x}})}, \tag{2.3}$$

where we calculate the covariance matrix $\mathcal{C}$ and mean $\bar{\mathbf{x}}_k$ from the chain. The pdf of the absolute distance, r_k to the nearest neighbour point is assumed to be Poisson distributed. Expressing this as a hyper-spherical infinitesimal shell of radius $[r_k, r_k + dr_k]$ gives,

$$p(r_k|n) = \frac{n^k V_d^{k-1} e^{-nV_d}}{(k-1)!} \frac{dV_d}{dr_k}, \tag{2.4}$$

where d is the dimensionality of the parameter space and V_d is the volume of said hyper-sphere given by $V_d(r_k) = \pi^{\frac{d}{2}} \frac{r_k^d}{\Gamma(1+d/2)}$. We then take advantage of using Poisson statistics allowing n (if sufficiently large) to be uniformly distributed along the desired scales. The next step is to combine this into a Bayesian likelihood from all the points in the chain. For a selected point indexed with α, we get the distribution,

$$p(a|\mathbf{r}) \propto \left[\prod_{\alpha=1}^{N} p\left(r_{k,\alpha}|n\right) \right] \Pi(a), \tag{2.5}$$

in terms of a vector, $\mathbf{r}$ containing each nearest neighbour arrangement within our sample set, n. $\Pi(a)$ is the prior on the proportionality constant which we want to find. Keeping in mind

we can rewrite $a = \mathcal{Z}/N$ (or $a = \tilde{\mathcal{P}}/n$), we can express the posterior for the Evidence as,

$$\ln\, p(\mathcal{Z}|\mathbf{r}, M) = \text{constant} - Nk\ln\mathcal{Z} - \frac{N}{\mathcal{Z}}\sum_{\alpha=1}^{N} V_d(r_{k,\alpha})\tilde{\mathcal{P}}_\alpha + \ln\Pi(a). \tag{2.6}$$

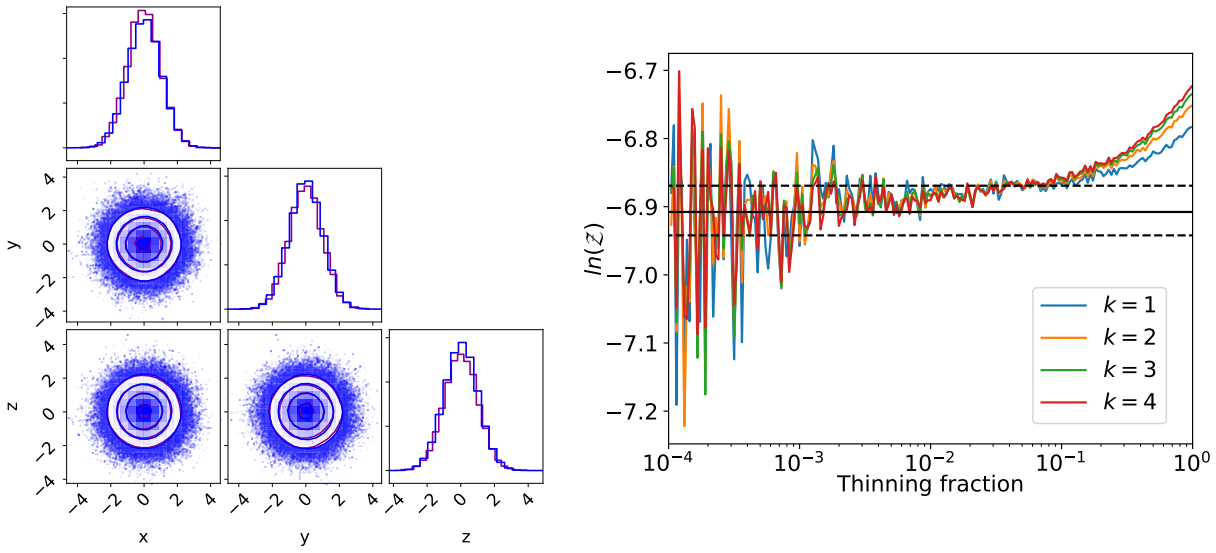
In the case of a uniform prior on a (and therefore $\mathcal{Z}$) we can finally write a MAP Evidence value as,

$$\mathcal{Z}_{\text{MAP}} = \frac{\sum_{\alpha=1}^{N} V_d(r_{k,\alpha})\tilde{\mathcal{P}}_\alpha}{k}. \tag{2.7}$$

The question is whether suitable thinning of the chain can reduce the correlation enough for our approximations to hold.

2.3 Application to a toy model

To test this methodology we apply MCEvidence to a simple toy model. We use a three-dimensional Gaussian likelihood about mean, $\boldsymbol{\mu} = (0, 0, 0)$ with variance, $\boldsymbol{\sigma}^2 = (1, 1, 1)$ across a uniform prior ranging $([-5, 5], [-5, 5], [-5, 5])$. Figure 2.1(a) shows the posteriors of this distribution as obtained by both Emcee (blue) and with MultiNest (magenta). Next to Figure 2.1(b) we have $\ln\mathcal{Z}$ as produced by MCEvidence for $k = 1, 2, 3, 4$ as a function of thinning fraction used on the chain. The thinning fraction is implemented as follows: a value of 1 implies the use of the full chain; 10^{-4} implies 1 in every 10^4 samples is used. Comparison $\ln\mathcal{Z}$ values are also plotted: the black solid line is the analytically integrated value; and the black dotted lines are the error bars produced by Multinest. Even here, with perfectly orthogonal parameters, a chain thinning fraction of 0.01 is required for MCEvidence to produce the correct result. Too little thinning and the correlation in the parameters from the MCMC chain causes the result to be incorrect - $\ln\mathcal{Z}$ is a function of the chain thinning. Too much chain thinning however and the number density is not representative of $\mathcal{P}$ (Equation 2.1 fails) - the posterior on $\ln\mathcal{Z}$ (Equation 2.6) is no longer tightly constrained. Notice that towards the left-hand side of 2.1(b), the variance on $\ln\mathcal{Z}$ increases sharply for every nearest neighbour evaluations. We recommend implementing a thinning fraction in the range $\in [0.003,\ 0.03]$.



(a) (b)

Figure 2.1: This illustrates the success of MCEvidence in the case of a model with orthogonal parameters. 2.1(a) shows agreement between the posteriors of MultiNest (Magenta) and Emcee (blue) on our toy Gaussian model (detailed in Section 2.3). 2.1(b) shows agreement between MCEvidence (coloured lines), MultiNest (upper and lower limits marked by the dotted black lines), and the analytically integrated results (black line). We recommend implementing the thinning fraction in the range $\in [0.003, 0.03]$ for three orthogonal parameters. See text for the discussion of how $\ln\mathcal{Z}$ varies with the thinning of the MCMC chain.

2.4 Application to toy EoR models

After showing the success of the methodology in the previous section, we now apply it to toy models of the EoR. We start with the three-parameter version of 21cmMC (Greig & Mesinger, 2015) (referred to as 3pFZHf1) which we will briefly summarise here.

The three parameters are:

- ζ - an ionising efficiency of the galaxies;
- $R_{\rm mfp}$ - a maximum photon mean free path; and
- Log$T_{\rm vir}$ - a minimum virial temperature.

For the fourth parameter (used in 4pFZHf1), the ionising efficiency is expanded to a power law in virial temperature as $\zeta \sim (T_{\rm vir})^{\alpha}$.

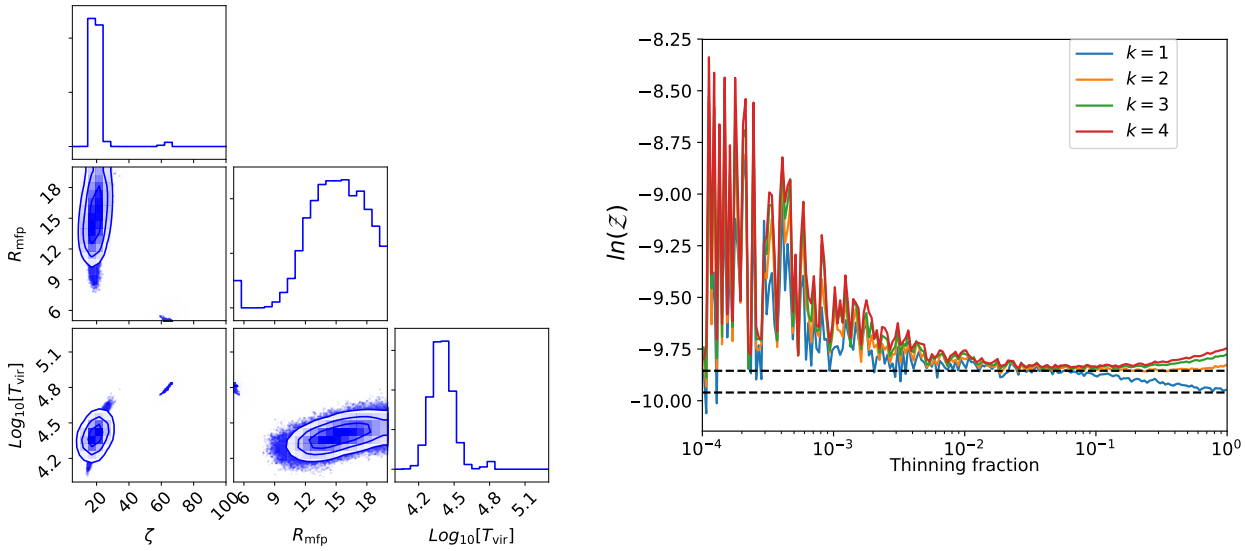
The Hydrogen ionising threshold is defined as,

$$\int_{M_{\rm vir}(T_{\rm vir})}^{\infty} \zeta \, \frac{dn(R)}{dm} m \, dm \geq 1, \tag{2.8}$$

where $R_{\rm mfp}$ is the maximum value of R, the scale at which the excursion set formalism is performed. Notice that they are not independent of each other (due to the nature of the greater than or equal sign). And in the case of 4pFZH they are not separable from each other, as ζ cannot be factored out of the integral.

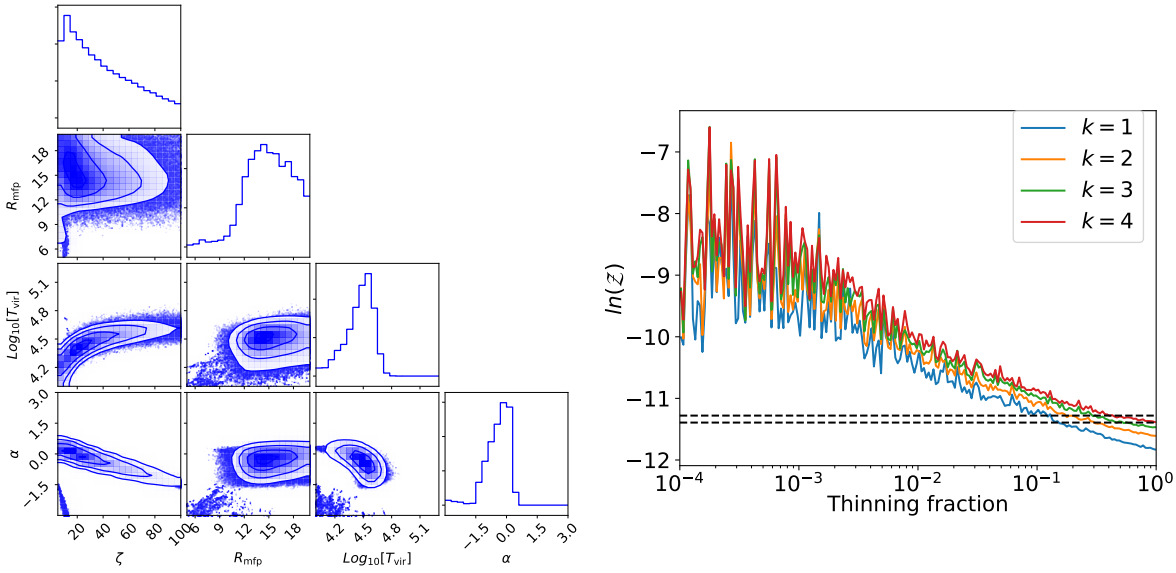
The posteriors produced in Figure 2.2(a) shows ellipsoidal contours between all three of the parameters - their relationship is more complicated than in the previous section, as indicated by Equation 2.8. Each of the parameters has a slightly ellipsoidal shape, hinting that they might be correlation. For both $R_{\rm mfp}$ posteriors of the ellipsoidal posteriors are mostly parallel to the plane of the $R_{\rm mfp}$, indicating correlation with this parameter is minimal. For $T_{\rm vir}$ and ζ the ellipsoidal posterior is diagonal, indicating a direct correlation. Luckily in this case the posterior is only slightly ellipsoidal, and not enough to be a problem for MCEvidence. The ζ parameter produces a near circular posterior (similar to the toy cases) because it can be factored out of the integral in Equation 2.8. As a result, the corresponding Evidence values converge for all k on the upper limit of MultiNest's error bar in Figure 2.2(b). Regardless of these pitfalls in calculating $\ln\mathcal{Z}$ for 3pFZHf1; we proceed in analysing 4pFZHf1 with the hope of ruling it out in a model selection context.

In the 4pFZHf1 case, pitfalls are significantly worsened. This can be seen both in the wider contours in the parameter posteriors of Figure 2.3(a); and in the dependence of $\ln\mathcal{Z}$ with thinning fraction in Figure 2.3(b). It is also worth noting that for no values of thinning fraction do the four nearest neighbour values converge on each other (unlike the 3pFZHf1 case).



(a) (b)

Figure 2.2: This shows a semi-successful attempt at applying MCEvidence to the 3pFZHf1 toy EoR model. The parameters are not orthogonal (See Equation 2.8) but the correlation is minimal, resulting in an approximately correct MCEvidence values for $\ln\mathcal{Z}$ in Figure 2.2(b)



(a) (b)

Figure 2.3: Here we illustrate the failure of MCEvidence when applied to 4pFZHf1. The parameters ζ and $T_{\rm vir}$ are correlated by the addition of the power-law index α resulting in large posterior contours (in comparison to Figures 2.1(a) and 2.2(a)). $\ln\mathcal{Z}$ and thinning fraction have become a function of each other. MCEvidence is not reliably applicable to this situation.

2.5 Attempted remedy

In order to prescribe the failure of this methodology as an impact of correlated parameters, we recreate the behaviour in the previous section using toy models. We then attempt to remove this correlation by selectively slicing the likelihood space before the pre-whitening step.

2.5.1 Replicating correlated behaviour

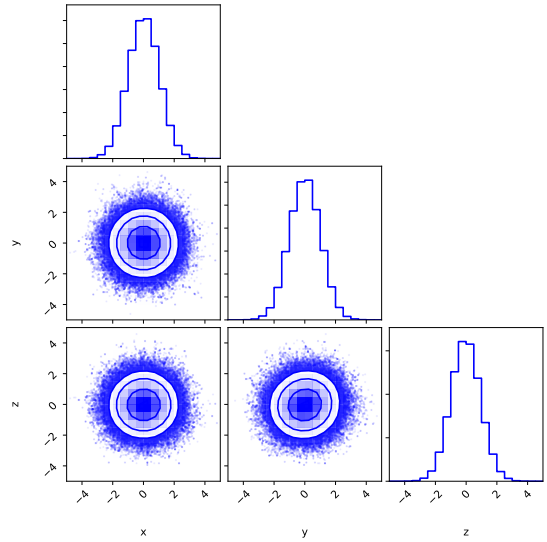
We endeavour to recreate the failure of MCEvidence by using the three dimensional toy Gaussian model of Section 2.3. We correlate the parameters of a toy model as follows,

$$z = A\ y + (1. - A)\ z. \tag{2.9}$$

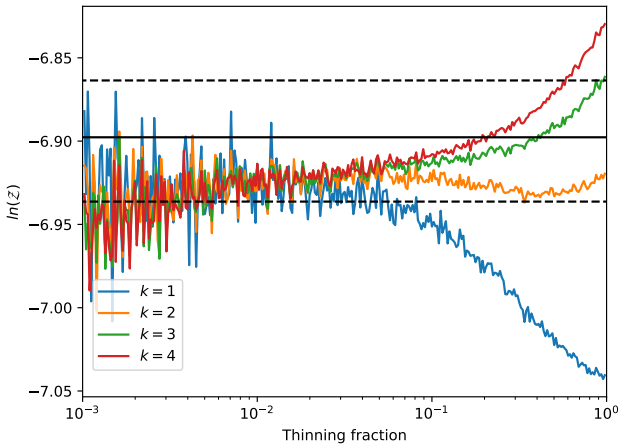
Here x remains the same as in Section 2.3, and A has followed values of 0.001, 0.01, 0.1, 0.12, 0.13, 0.14, 0.15, 1/5, 1/2, and 1. Figures 2.4, 2.5, and 2.6 show $A =$ 0.01, 0.1, 0.12, 0.13, 0.14, and 0.15 respectively in pairs. Each figure shows a parameter posterior produced by Emcee (left) and an $\ln\mathcal{Z}$ against thinning of the chain (right). In the case of $A = 0$, we retain the results in Section 2.3. For $A \geq 0.15$ the posteriors and Evidence plots are indistinguishable to $A = 1$.

As can be seen, MCEvidence agrees with MultiNest when the parameters are mildly correlated (Figures 2.4(a)). Figure 2.4(b) has a mild correlation with $A = 0.01$, and shows a result within MultiNest's error bars, however it is already on the cusp. By the time the correlation has reached $A = 0.1$ (Figure 2.4(d)) the approximations that the chain is uncorrelated no longer holds. As A increases (Figures 2.5(b), 2.5(d)) the $\ln\mathcal{Z}$ values of the nearest neighbour begin to switch order. It is clear that there is a turning point in this behaviour (Figure 2.6(b) - the last plot where all 4 nearest neighbour points agree). The steep slope in $\ln\mathcal{Z}$ against thinning fraction shown by the 4pFZHf1 EoR model (Figure 2.3(b)) is replicated when $A = 0.2$.

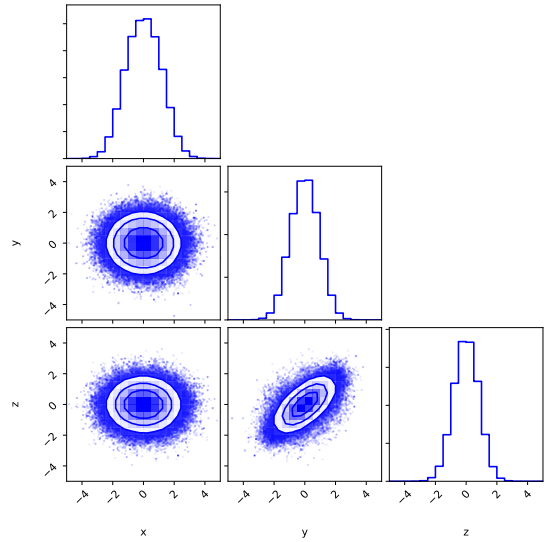
As A values are the only change between these models - correlation is responsible for violating the assumptions used in Section 2.2. But we have been unable to offer a deeper explanation as to why the Evidence value has become a function of thinning.



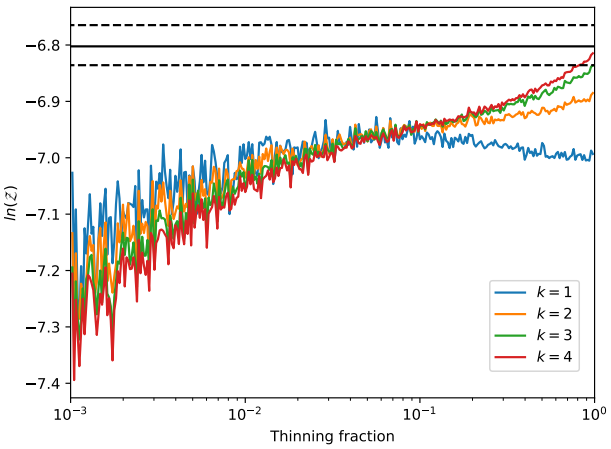
(a)



(b)

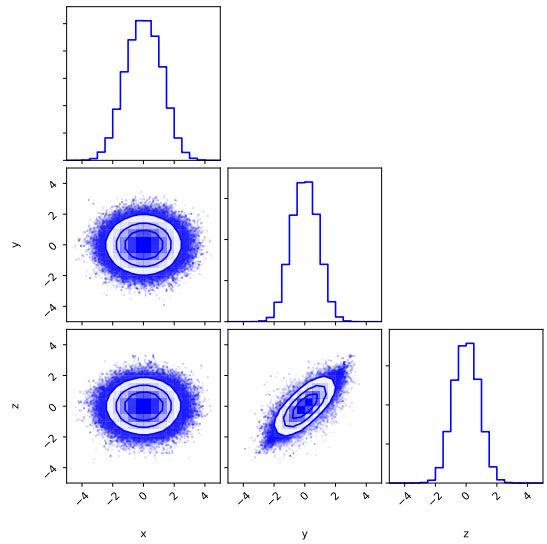


(c)

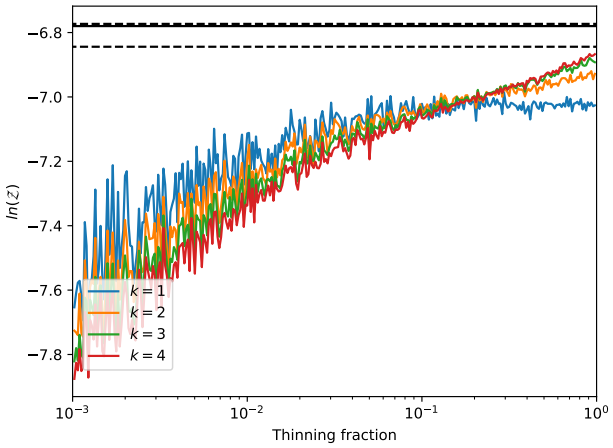


(d)

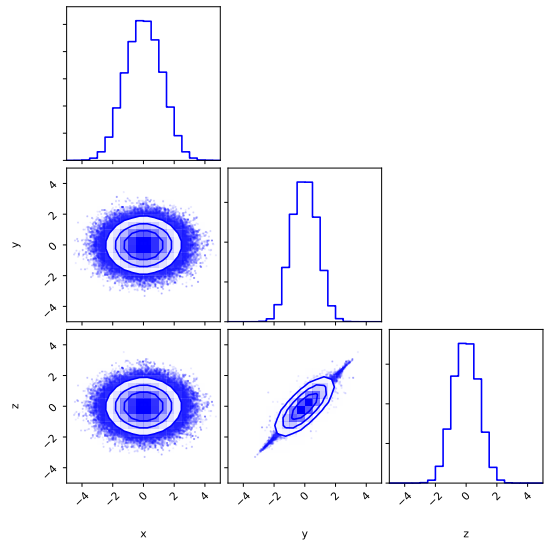
Figure 2.4: These plots refer to the correlation applied to the toy Gaussian model via Equation 2.9. Figures 2.4(a)/2.4(b) and 2.4(c)/2.4(d) are EMCEE's parameter posteriors and $\ln\mathcal{Z}$ against thinning fraction referring to $A = 0.01$, and 0.1, respectively. These Figures show the cross over from MCEVIDENCE behaving appropriately to when the algorithm breaks down due to correlation between the parameters. Figure 2.4(b) shows MCEVIDENCE producing Evidence values in agreement with MULTINEST; 2.4(d) shows that $\ln\mathcal{Z}$ has become a function of the chain's thinning fraction.



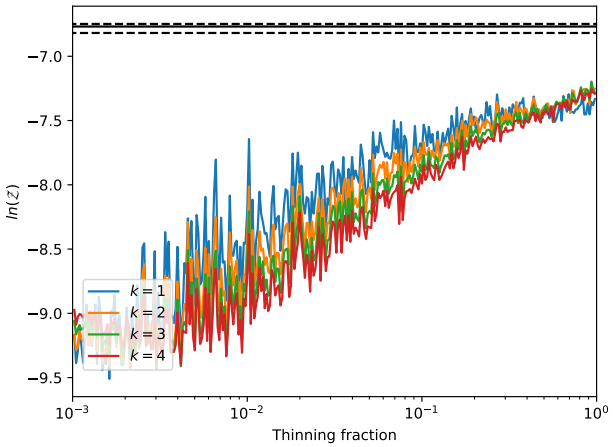
(a)



(b)

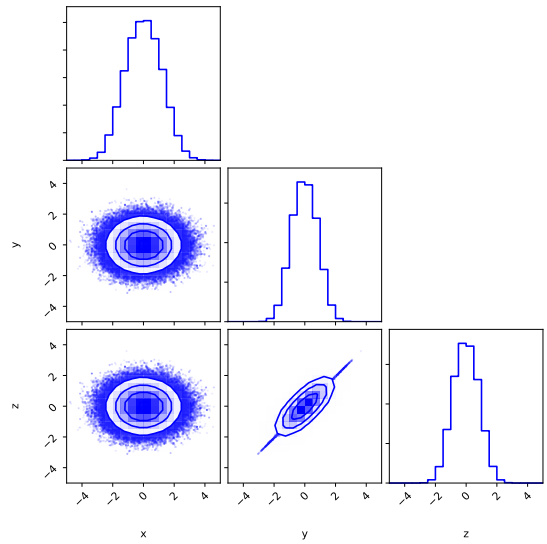


(c)

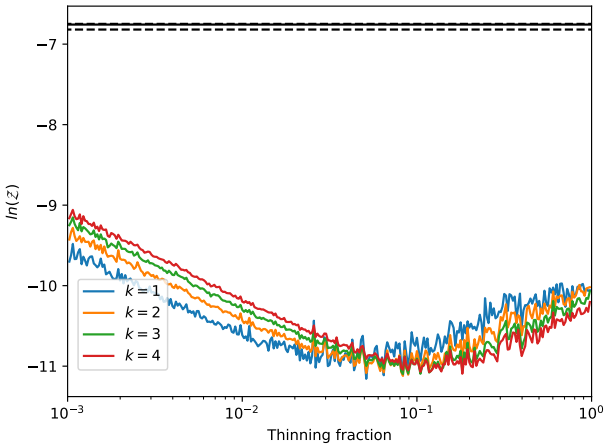


(d)

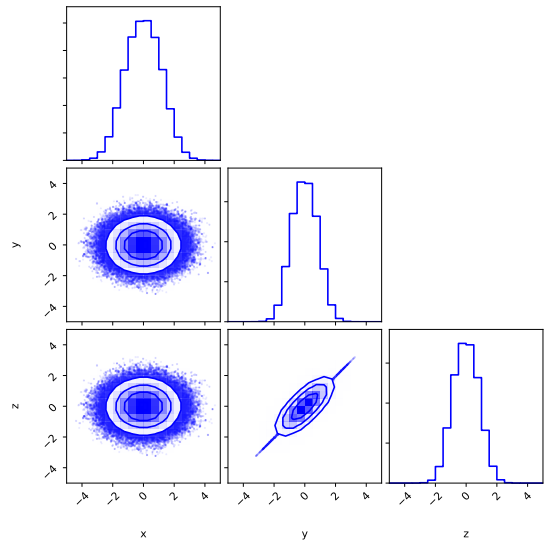
Figure 2.5: A continuation of Figures 2.4 and 2.6. These plots refer to the correlation applied to the toy Gaussian model via Equation 2.9. Figures 2.5(a)/2.5(b) and 2.5(c)/2.5(d) are EMCEE's parameter posteriors and $\ln\mathcal{Z}$ against thinning fraction referring to $A = 0.12$ and 0.13, respectively.



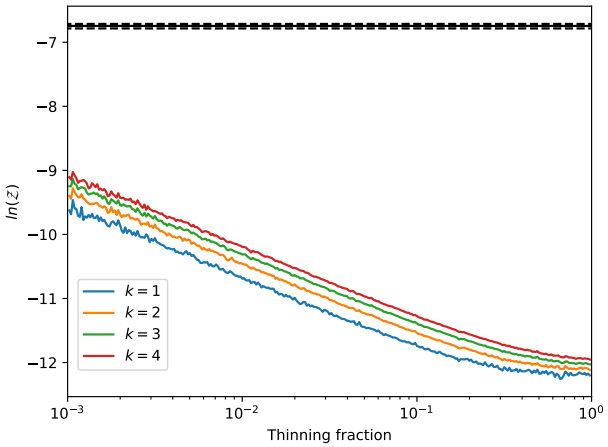
(a)



(b)

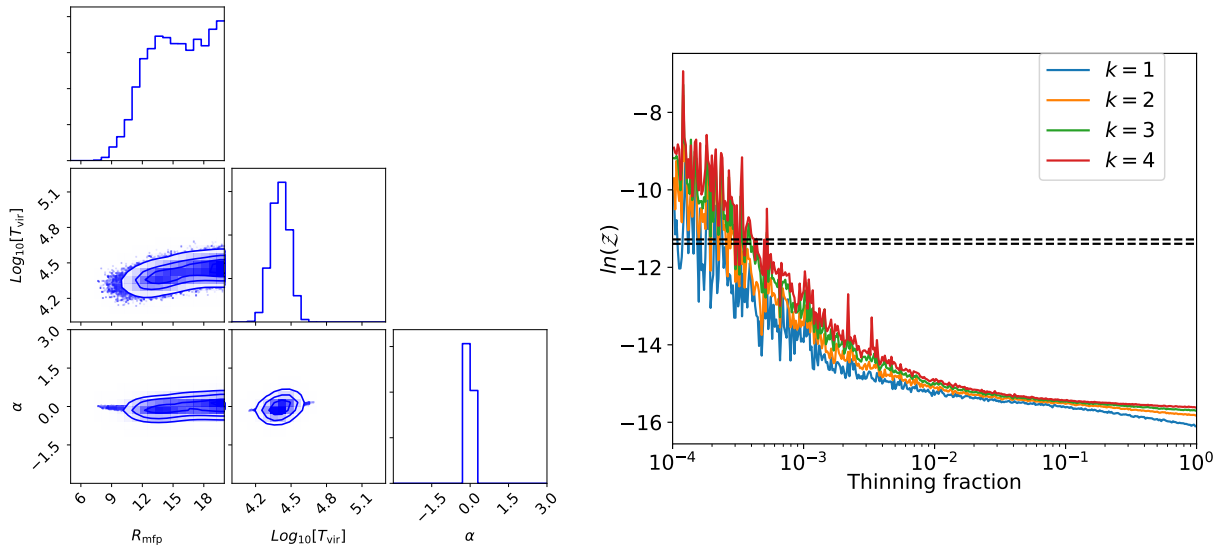


(c)



(d)

Figure 2.6: A continuation of Figures 2.4 and 2.5. These plots refer to the correlation applied to the toy Gaussian model via Equation 2.9. Figures 2.6(a)/2.6(b) and 2.6(c)/2.6(d) are EMCEE's parameter posteriors and $\ln\mathcal{Z}$ against thinning fraction referring to $A = 0.14$ and 0.15, respectively.



(a) (b)

Figure 2.7: This shows a failed attempt at remedying the results in Figure 2.3(a). We have applied a δ-function prior on the fiducial value of ζ (hence plotting the other three parameters only). The $\ln\mathcal{Z}$ values are better than in Figure 2.3(b) however this method remains unusable in the EoR context.

2.5.2 Likelihood cuts

To salvage some information on the Bayesian Evidence from the 4pFZHf1 chain, we attempt to perform MCEvidence on the 4pFZHf1 model with a Dirac δ-function in place of the prior distribution for the ζ parameter. The resulting three parameter posterior and corresponding Evidence plots are shown in Figure 2.7. We are able to obtain some convergence between the nearest neighbours value, but only at the cost of broadening the posterior on $\ln\mathcal{Z}$. As can be seen we are unable to remedy the situation consistently.

We also note that the posteriors produced by 21cmMC are truncated. The $R_{\rm mfp}$ in particularly cannot be untruncated by extending the prior. To probe this we truncated the toy Gaussian model from Section 2.3 but found no deviation to the MCEvidence's performance unless the MAP parameter was included in the truncation[1].

[1]We did not re-run the MultiNest or Emcee with truncated priors since the MAP parameters are captured in each 21cmMC posterior. We therefore cannot ensure convergence within this test.

2.6 Conclusion

MCEvidence can produce the Bayesian Evidence from MCMC chains if the parameters are orthogonal, as shown in Section 2.3. In the context of the EoR, we are unable to obtain reliable values of the Bayesian Evidence directly from MCMC chains (Section 2.4). Heavens et al. (2017a) state that correlations within the MCMC chain are problematic in the MCEvidence method (detailed in Section 2.2). We can recreate MCEvidence's behaviour when failing on toy EoR models by correlating the parameters of the toy Gaussian models used in Section 2.5.1. Unfortunately we are unable to produce sensible results by selectively slicing the parameter space in an attempt to limit the correlation (Section 2.5.2).

We conclude that we are unable to calculate the Bayesian Evidence directly from MCMC chains in the context of EoR analyses. Even with toy EoR models, the correlations between parameters are significant and prevent the methodology from being successful. In the next chapters we use MultiNest to calculate the Bayesian Evidence.

Chapter 3

Bayesian Model Selection with Future 21cm Observations of The EoR

3.1 Introduction

In this chapter we show the foundations for using Bayesian model selection in the context of the EoR. We have introduced an example semi-numerical simulation, 21CMFAST which is used for parameter estimation in 21CMMC in Chapter 1. Here we present an attachment to 21CMMC called 21CMNEST[1], which tests toy models of the EoR by implementing MULTINEST (detailed in Section 1.5.2) in place of COSMOHAMMER. The highlights of this Chapter are used in Binnie & Pritchard (2019)[2].

The models tested in this chapter are based on the approximate descriptions touched on in Section 1.3 and they are detailed in Section 3.2.1. To recap, we split these via the morphology of percolated ionised regions:

- *Inside-out* - High density gas is ionised first, due to star formation being most dominant in dense regions of gas. The density field and radiation field correlate.
- *Outside-in* - Low density gas is ionised first, due to dense gas recombining faster than it is ionised by the radiation. The density field and radiation field anti-correlate.

[1]Publicly available at https://binnietom@bitbucket.org/binnietom/21cmnest_1.0.git
[2]Available at https://arxiv.org/pdf/1903.09064

And by the scale that these behaviours are associated:

- *Local* - There is no dependency on the surrounding area. The ionisation threshold is implemented pixel by pixel.
- *Global* - The ionisation criteria is scale dependent. Each pixel has a sphere of influence limited by the mean free path of a photon.

These are approximate descriptions. In theory reionisation progresses with a variety of these qualities at various different times, as seen in detailed radiative transfer simulations (Finlator et al., 2018).

Data is starting to become available with instruments such as LOFAR (Patil et al., 2017), MWA (Dillon et al., 2015), PAPER (Ali et al., 2015), and HERA (DeBoer et al., 2017). With a little luck one of them will soon detect the 21cm power spectrum (Chapman et al., 2012). In the meantime we proceed with a mock observation simulated semi-numerically.

The main aim of this chapter is to look at how feasibly three selected radio interferometers can rule out toy models (representing the above properties). We summarise these with three questions:

- What length of observation is needed to distinguish between the toy models with LOFAR? (Section 3.4.3)
- For a fixed observing timescale (1080 hours), what number of dipoles are necessary to distinguish between the toy models with HERA? (Section 3.4.4)
- What is the minimum observation time required to distinguish between the toy models using SKA? (Section 3.4.5).

In order to effectively distinguish toy models all the results are compared with the Jeffreys' scale discussed in Section 1.4.6.

The structure of the chapter is as follows: firstly we detail the toy EoR models, continuing from Chapter 1, including the details of telescope noise implementation. We discuss the considerations of the parameter priors in Section 3.3. Then we show agreement between 21CMMC

and 21cmNest in parameter estimation (Section 3.4.1). We then proceed in distinguishing the toy models with various radio telescopes. For each telescope, we try and answer the specific questions as above. We would like to know how long a LOFAR observation would need to be distinguish between the toy models (Section 3.4.3). For HERA (Section 3.4.4) we set a 1080 hour observation and see how many dipoles in the array are necessary to answer the same question. Section 3.4.5 looks into how short the observation would need to be with SKA-512. We also analyse the contributions to the Evidence calculation from the various likelihood constituents in Section 3.4.6. Within the FZH, model we then use the SDDR to analyse the inference obtainable from each parameter (Section 3.4.7). Finally, we are unable to distinguish a power-law from constant in halo mass for the ionising efficiency ζ with just the 21cm power spectrum and observational prior checks (Section 3.4.8). We draw conclusions in Section 3.5.

3.2 21cmNest

In this Chapter we use the likelihood structure in Greig & Mesinger (2015). We calculate a χ^2 statistic between a mock 21cm observed power spectra from fiducial parameters and a simulated one dictated by the sample. This is combined with the observational priors introduced in Section 1.3.2 as,

$$\ln\mathcal{L} = \chi^2_{21} + \chi^2_{\rm Greig} + \chi^2_{\rm McGreer} + \chi^2_{\rm Planck}. \tag{3.1}$$

In this (and the previous) chapter χ^2_{21} is calculated for 8 k bins from a foreground corruption limit of $k = 0.15$ Mpc^{-1} to a shot noise limit of $k = 1.0$ Mpc^{-1} (discussed in more detail in Section 3.2.2). The spherically averaged power spectrum, Δ^2_{21}, is used in the likelihood statistic. It is defined,

$$\Delta_{21}(k)^2 \equiv \frac{k^3}{2\pi^2 V}\overline{\delta T_b}^2(z)\langle\delta_{21}(\mathbf{k},z)|^2\rangle \ \mathrm{mK}^2, \tag{3.2}$$

via deviations to the mean brightness temperature as,

$$\delta_{21}(\mathbf{x},z) \equiv \frac{\delta T_b(\mathbf{x},z)}{\overline{\delta T_b}(z)} - 1,$$

where the bar and angular brackets denote a spatial average and a k-space average respectively.

3.2.1 Toy EoR models

Initially, these models are taken from Watkinson & Pritchard (2014) which are motivated by Furlanetto et al. (2004) (FZH) and Miralda-Escudé et al. (2000) (MHR). As this is a proof of concept work, we follow the literature for selecting toy EoR models. These are very different models that provide a qualitative bracket on the range of possible morphologies. The physical behaviour of all EoR models should exist within these models and they are therefore used here. If our observations are unable to distinguish the physics of these morphological extremes, the concept has failed.

We also add a few of our own (namely 1p MHR, F MHR, and F invMHR) which change the number of parameters used in the basic MHR model without changing the 21cm power spectrum calculation. These permutations are included to test specifically how the statistical machinery scales in calculations with different numbers of parameters. Figure 3.1 illustrates the 4 differing ionisation thresholds used in this chapter. A summary of these models is provided in Table 3.1 and we stress that some of the permutations are unphysical and have been created to test the statistical methodology. Once models of definitive morphologies can be successfully ruled out, only then does it makes sense to consider more physical models containing more complicated morphologies. For example, Pagano & Liu (2020) discuss parameterising the correlation of ionisation and density field as a possible follow up to Binnie & Pritchard (2019).

Figure 3.2 shows a brightness temperature field slice given at $z = 8$ with $x_{\rm HI} \approx 0.5$. Before proceeding with the 21cm PS, all models were run with just the observational priors (i.e. $\ln\mathcal{L} = \chi^2_{\rm Greig} + \chi^2_{\rm McGreer} + \chi^2_{\rm Planck}$). The MAP values, MAP power spectra and reionisation histories are shown in Table 3.2 and Figure 3.3. In all models UV radiation will eventually dominate and is assumed to drive reionisation (Garaldi et al., 2019). In reality, it is likely that the reionisation morphology will begin *outside-in* and end *inside-out*.

FZH - the 21cmFast models

Furlanetto, Hernquist and Zaldarriagan (FZH), is summarised in Section 1.3. This implements the simplest scenario (Barkana & Loeb, 2001), i.e. that the mass of the collapsed object relates to the mass of the ionised region via the ionising efficiency ζ as,

$$m_{\rm ion} = \zeta m_{\rm gal}. \tag{3.3}$$

To implement this we use the collapse fraction (Equation 1.12), $f_{\rm coll}$, for a Press-Schechter mass function (Equation A.10) which allows us to write,

$$f_{\rm coll} = {\rm erfc}\left\{\frac{\delta_{\rm crit} - \delta}{\sqrt{2[\sigma^2(m_{\rm min}) - \sigma^2(m)]}}\right\}. \tag{3.4}$$

Rearranging Equation 3.4 allows us to express an ionisation criteria with respect to the density field as,

$$\delta \geq \delta_{\rm crit} - \sqrt{2[\sigma^2_{\rm min}(m_{\rm min}) - \sigma^2(m)]}\ {\rm erfc}^{-1}\left(1 - \frac{1}{\zeta}\right). \tag{3.5}$$

The factor of 2 in Equations 3.4 and 3.5 comes from the symmetry of random walk excursions discussed in Appendix A.1. For every random walk that reaches this critical density barrier, half will continue their walk above the barrier. Therefore half will collapse to form virialised objects.

FZH is applied numerically by smoothing from large to small scales, incorporating the global behaviour surrounding each pixel. Should Equation 3.5 be achieved before the filter scale reaches the pixel size, these central pixels are flagged as ionised. Some sub-grid regions will have a mixed combination of ionised and neutral hydrogen. To incorporate this, after the filtering scale has reached the pixel size the neutral fraction of $x_{\rm HI} = 1 - \zeta f_{\rm coll-per-pixel}$ to each pixel that is yet to be fully ionised[3]. Pixels are defined to be fully ionised if,

$$\zeta f_{\rm coll} \geq 1. \tag{3.6}$$

Throughout this chapter we refer to 3pFZH and 4pFZH. These refer to the number of parameters implemented in the FZH model. In particular it refers to the ionising efficiency. When ζ is a constant (3p), or when this option is relaxed in the form of a power-law dependency in halo-mass (4p). The 4th parameter, is α, enabling relationship between ζ and $T_{\rm vir}$ as,

[3]This includes rounding to 0 if $x_{\rm HI} < 0$ and 1 if $x_{\rm HI} > 1$.

$$\zeta = \begin{cases} \zeta_0 \left(\frac{T'_{\rm vir}}{T_{\rm vir}}\right)^{\alpha}, & \text{if } T'_{\rm vir} \geq T_{\rm vir} \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

The ionising efficiency ζ is effective as,

$$\zeta_{\rm eff} = \frac{\int_{M_{\rm vir}}^{\infty} dm \frac{dn}{dm} m\ \zeta(m)}{\int_{M_{\rm vir}}^{\infty}\ dm \frac{dn}{dm} m}, \tag{3.8}$$

where the mass function is Press-Schechter with a Sheth-tormen correction factor (Press & Schechter, 1974; Bond et al., 1991; Sheth & Tormen, 1999) and $T_{\rm vir}$ is integrated out. We make note of this here as it is implemented in various ways. For the constant form, the integrals cancel and we are left with $\zeta_{\rm eff} = \zeta$.

The mock observations labelled f1 and f2 are created with 3pFZH and 4pFZH respectively using the parameters in Table 3.5 - these are the faint and bright galaxy models respectively from Greig & Mesinger (2015). In the context of comparing all the toy models (rather than just the 21CMFAST models detailed here), FZH refers specifically to 3pFZH.

Inverted (Inv) FZH

Inverting the ionisation threshold for FZH (Equation 3.5) produces an *outside-in* model sensitive to *global* scales. Here underdense regions are ionised first to form bubbles, with the edge of these bubbles remaining in equilibrium[4]. The ionising radiation will increase as star formation progresses, until eventually the UV background will dominate over the recombination rate in overdense regions anticipating the completion of reionisation. To detail InvFZH we first redefine ζ as ζ' in order to keep the parameter framework within 21CMMC unchanged. ζ' is the *background* ionising radiation efficiency. This inverts the critical barrier as $\delta'_{\rm crit} = -\delta_{\rm crit}$. As is done for the FZH models, we assume Gaussian perturbations for our regions of over-density. Now we need to calculate the fraction of matter that satisfies $\delta < \delta'_{\rm crit}$, which requires a new collapse fraction $f'_{\rm coll}$. Below this HI structures are sparse enough to be ionised, and all that

[4]Similar to a Strömgren sphere but on galactic scales.

needs doing is to change the limits of the original $f_{\rm coll}$ integral as,

$$f'_{\rm coll} = \frac{1}{\rho_{\rm M}} \int_0^{M_{\rm vir}} m \, \frac{dn}{dm} \, dm \tag{3.9}$$

$$= \mathrm{erfc}\left[\frac{\delta_{\rm crit} + \delta}{\sqrt{2(\sigma^2(m_{\rm min}) - \sigma^2(m))}}\right] . \tag{3.10}$$

We now have an analogous ionisation threshold, $f'_{\rm coll}\zeta' \geq 1.$, that can be implemented easily within the 21CMMC framework. As before, we assign $x_{\rm HI} = 1 - \zeta' f'_{\rm coll}$ to any partially ionised pixels.

MHR

For a local scale implementation of reionisation we implement the work of Miralda-Escudé, Haenelt and Rees (Miralda-Escudé et al., 2000). This is a *local outside-in* model - underdense regions of the IGM ionise first due to a recombination rate that is dependent on the density of the gas. In contrast, the denser regions of gas recombine quickly and therefore ionising radiation struggles to permeate these regions meaning they are the last IGM regions to be ionised. Circumgalactic gas, within the halo, becomes ionised first before ionising photons infiltrate into the IGM. HII regions expand in the direction of low gas density and the EoR is defined to end when these regions overlap. Typically this is achieved by a background of X-ray photons which heat the gas due to their large mean free paths. Since we do not directly calculate the X-ray background here (introduced in Chapter 4), we equate this background to ζ. We therefore define the MHR ionisation threshold via a neutral fraction $\bar{x}_{\rm HI} = 1 - f_{\rm coll}\zeta$. To implement this we first order the pixels by HI density, ensuring consistency between the $x_{\rm HI}$, then we find the i'th pixel (of total $N_{\rm p}$) that satisfies the ionisation criteria as,

$$\frac{i}{N_{\rm p}} = 1 - \bar{x}_{\rm HI} \ , \qquad \delta < \delta_i \ . \tag{3.11}$$

The ionisation criteria is a function of density. In summary the only deviation from 21CMFAST is the definition of ionisation via this new density threshold. It is worth noticing that this no longer depends on $R_{\rm mfp}$, making MHR a 2 parameter model.

Prescribing the $\bar{x}_{\rm HI}$ directly is referred to as the 1-parameter (1p) MHR model. This is included

only for testing how using different parameter numbers can skew the Evidence.

Inverted (Inv) MHR

Here we produce a *local inside-out* model of reionisation by following the same procedure as above. The choice of the j'th pixel then follows by using,

$$j = N_{\rm p} - i \ , \qquad \delta > \delta_j \ . \tag{3.12}$$

Notice the greater than sign is reversed to produce an over-density threshold. The use of i (from Equation 3.11) ensures that the neutral fraction remains consistent between the MHR models.

Filtered (F) MHR

In order to test 21CMNEST more thoroughly, a third parameter is added to the basic MHR model. Due to the Occam's razor effect inherent in calculating the Bayesian Evidence (see Section 1.4.3), we want to be aware of any possible skewing effect that arises from a change in prior volume when comparing models with different numbers of parameters. Here, the added parameter is a Gaussian k-space filter performed across the density field. The integrated volume of this Gaussian is equivalent to a real space top-hat with radius $R_{\rm Filter}$. This smoothing reduces the range of values within the density field, which has an effect on the number and density of ionised structures. We expect this added flexibility to relax the *local* scenario, allowing some fitting flexibility on the reionisation scale from this new parameter. This means the model has a *global* scale when fitting for an FZH fiducial power spectrum, but has a redundant parameter when fitting for an MHR fiducial power spectrum (Section 3.4.2). In the F MHR model ζ, the background ionising efficiency, is pushed below that of the original MHR model, with the opposite being true for the F Inv MHR. For the filter scale itself, we found this tended roughly to the mean size of HII bubbles accross the boxes. The F Inv MHR shares morphology and scale with FZH, and is therefore the most challenging toy model to distinguish from f1.

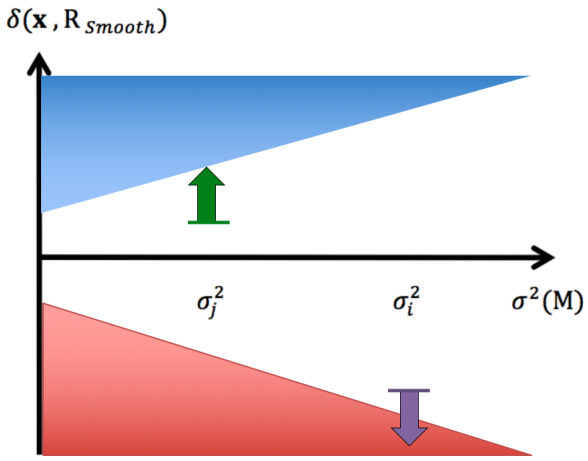


Figure 3.1: An illustration of the ionisation thresholds for the 4 models detailed Table 3.1. Blue (red) region are for the mass variance dependent criteria for FZH (Inv FZH) detailed by Equation 3.5 (Equation 3.10). Purple (green) represent the critical density threshold for the pixel-by-pixel implementation of MHR (Inv MHR) in Equation 3.11 (Equation 3.12). See Section 3.2.1 for the full detail. This Figure has been re-styled from Watkinson & Pritchard (2014).

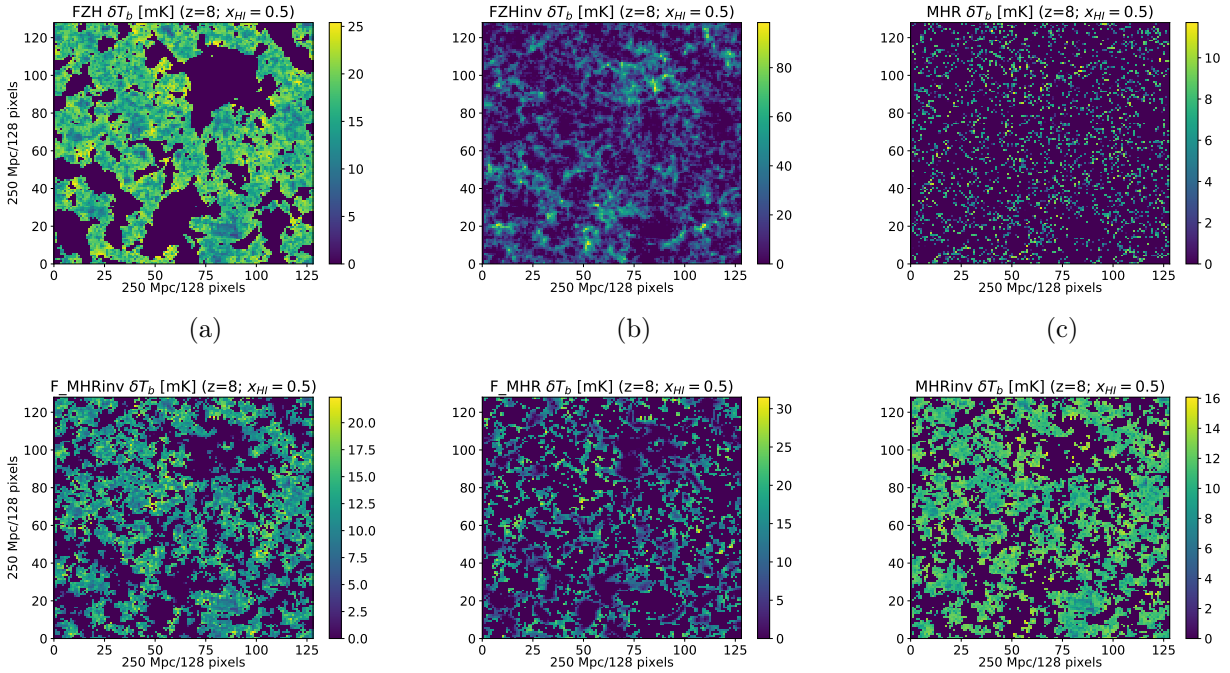


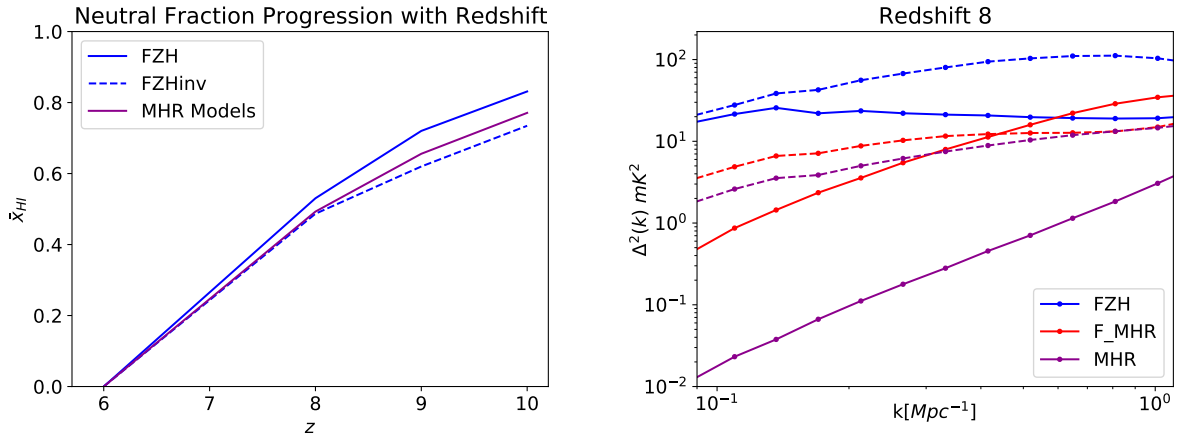
Figure 3.2: Slices of the redshift 8 coeval brightness temperature cube for the toy models. Figures 3.2(a), 3.2(b), 3.2(c), 3.2(d), 3.2(e), and 3.2(f) represent the FZH, Inv FZH, MHR, F Inv MHR, F MHR, and Inv MHR models respectively. Please see Table 3.1 for a summary of the models.

Model	*Physical motivation*	*Ionisation criteria*	*Resulting type of reionisation*
FZH	Over-dense regions collapse to form stars. As stars form UV radiation increases and is the driving force behind ionised bubbles. The size of these bubbles dictates the the brightness temperature and reionisation history.	Over-density increases until above the critical barrier ($\delta_{\rm crit}$).	*Global inside-out*
Inv FZH	Over-dense regions remain neutral due to a high recombination rate, therefore reionisation begins in underdense regions. As star formation progresses, under dense regions grow and dominate as hard UV radiation becomes the dominant radiation background and the IGM becomes ionised.	Over-density decreases until below the critical barrier.	*Global outside-in*
MHR	Star formation ionises circum-galactic gas. Ionisation proceeds through underdense regions as atoms in dense regions recombine faster than they are ionised. As background radiation increases it becomes dense enough to dominate.	An under density threshold is implemented pixel-by-pixel.	*Local outside-in*
Inv MHR	Star formation ionises the circum-galactic gas and over-dense regions become ionised. It is easy to implement rather than being physically viable since even dense regions ionise despite not hosting star formation or having a line of sight to a radiation source.	An over density threshold is implemented pixel-by-pixel.	*Local inside-out*

Table 3.1: A summary of the main four EoR models tested in this work (Watkinson & Pritchard, 2014). It is worth noting that filtering the density field (referring to F MHR) allows the pixel-by-pixel implementation of the ionisation threshold to change the scale of reionisation from *local* to *global*. See Figure 3.1 for an illustration of the ionisation criteria.

Model	ζ	$R_{\rm mfp}$	$\log_{10}[{\rm T_{vir}}]$
FZH	95.4 ± 61.9	12.3 ± 4.3	5.17 ± 0.29
Inv FZH	71.8 ± 68.5	14.9 ± 5.4	4.72 ± 0.34
F MHR	413 ± 289	9.3 ± 4.3	5.80 ± 0.31
F Inv MHR	268 ± 290	18.3 ± 4.4	5.61 ± 0.34
MHR	404 ± 293	-	5.72 ± 0.33
Inv MHR	799 ± 291	-	5.88 ± 0.33

Table 3.2: The MAP parameter values ($\pm 1\sigma$, standard deviation) for the toy models using the three observational prior checks (and no 21cm data). Each standard deviation is large (compared to each respective parameter prior) hinting that this data will not be constraining in the context of distinguishing toy EoR scenarios (discussed further in Section 3.4.6).



(a) (b)

Figure 3.3: A summary of the toy models used. 3.3(a) EoR histories for each toy model, produced with $\bar{x}_{\rm HI} \approx 0.5$ at $z = 8$. Blue and purple represent the 21CMFAST models (FZH) and MHR based models respectively, with dotted lines to indicate the inverse the models. Please note that all MHR based models are constructed to have the same EoR history, hence there is only one purple line. 3.3(b) 21cm PS for the 6 constructed toy models, corresponding to the ionisation histories. Red indicates adding the filter scale to the MHR models. The models are summarised in Table 3.1 (see Section 3.2.1 for more detail), the brightness temperature slices from the full coeval cubes for each of these models are plotted in Figure 3.2. Despite very similar EoR histories for all models, the 21cm power spectra are vastly different.

3.2.2 Telescope assumptions

In this chapter we make use of 21CMSENSE with *moderate* foreground settings for the instruments LOFAR, SKA, and HERA, as detailed in Section 1.3.7. A summary of the 21CMSENSE settings can be found in Table 3.2.2. For LOFAR and SKA we vary $t_{\rm int}$ and hence expect the noise to scale roughly as $t_{\rm int}^{-0.5}$. Figure 3.4 shows example error on a 21cm power spectrum (f1 to be precise).

The k-range for χ^2_{21} is dictated by cosmic variance, shot noise and foreground corruption. We use a coeval box size of 250 Mpc, which leads to limitation in the k-scale of the volume averaged statistic below $0.15{\rm Mpc}^{-1}$. For shot noise, our pixel resolution is 128 per box, or $\sim$ 2Mpc meaning we have little resolution on low scales, motivating the cut at $k = 1{\rm Mpc}^{-1}$. Given that we are not using the light-cone, we are effectively running a non-evolving box-car average along the light-cone, we would start to see deviations in our power spectrum around $k = 0.8{\rm Mpc}^{-1}$. Applying the light-cone in place of this is done in Chapter 4. Signal corruption due to foregrounds not prominent within these constraints, given we are using the moderate 21CMSENSE assumptions (which are ambitious in the authors opinion).

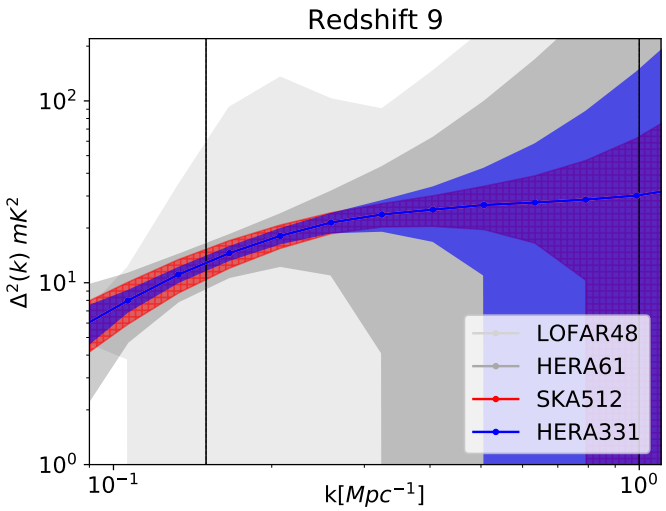


Figure 3.4: The 21cm PS error with LOFAR, HERA-61, SKA and HERA-331 on the f1 power spectra (defined in Section 3.2.1) calculated by 21CMSENSE (See Section 3.2.2. Only the values between the black vertical lines are used for the χ^2 calculation in the likelihood. The upper and lower limits (in k) are respectively defined by foreground contamination and shot noise limitations from the box size.

The noise power from HERA-331 with 1080 hours of observation is used throughout this Chapter unless specified otherwise. In Sections 3.4.3, 3.4.4, and 3.4.5, the results use LOFAR-48, multiple HERA configurations, and the SKA-512 respectively. A comparison of power spectrum error on the f1 mock observation is shown in Figure 3.4 for LOFAR, HERA-61, HERA-331 and the SKA at $z = 9$. The SKA has longer baselines than HERA and is therefore more sensitive at small scales. But due to HERA-331 having many more smaller baselines it has greater sensitivities to large scales, making it comparable to the SKA at model selection via the 21cm power spectrum.

3.3 Prior distribution considerations

As discussed in Section 1.4.4, a potential source of bias in Bayesian model selection is a poor choice of prior distribution for each parameter. Here we discuss the limitations and steps taken to ensure the results are robust. The results in Section 3.4 implement uniform prior distributions for all parameters. Similar assumptions to the original 21CMMC publication (Greig & Mesinger, 2015) are implemented.

Within the 21CMFAST models, we adapt the upper end of the ζ parameter to ensure the bright-

Telescope	LOFAR	HERA	SKA-Central
Number of dipoles Stations	48	331	296
Station Diameter [m]	31	14	35
Collecting Area [m^2]	35,762	50,953	492,602
$T_{\rm R}$[K]	140	100	$40 + \frac{T_{\rm sky}}{10}$
Observing Range [MHz]	[110,240]	[50,250]	[50,350]
Observation Time [hrs]	1080	1080	1080
Scan Type	1-hr Track	Drift	1-hr Track

Table 3.3: The different telescope specifications used with 21cmSense in this chapter. Of the SKA's 512 stations, we simulate using the 'central' 296 stations since the sensitivity to the EoR frequencies is dominant here and the longest baselines add significant computation time for negligible increase in sensitivity.

est galaxies are included: i.e. $\zeta \in [5., 100.]$ is changed to $[5., 250.]$. It is impossible to measure the ionisation flux directly but since the release of 21cmMC so called 'leaky' galaxies (with $\zeta > 100$) have been discovered (Borthakur et al., 2014). Due to the lack of observational data surrounding high redshift astrophysics, we err on the side of caution when deciding statistical limitation. To really understand how this limit is implemented we must break down ζ into its constituents. The variation typically comes from $f_{\rm esc}$ which is set within $[0.05, 1.0]$. $f_* = 0.05$ is used due to motivation of the theoretical modelling in (Dayal et al., 2014). This value has been confirmed to be adequate by the existing observational constraints in Gorce et al. (2018), which are similar to the observational prior checks used here. However developing a fully satisfactory model for ζ that can be pinned down is tricky due to it being related to so many astrophysical variables. This is motivation for leaving ζ behind as a parameter in favour of $f_{\rm esc}$ and f_* themselves (discussed in Park et al. (2019), and implemented in Chapter 4). As discussed in Section 1.3, N_γ and $n_{\rm rec}$ are taken to be 4000 and 1 respectively (Barkana & Loeb, 2001; Sobacchi & Mesinger, 2014).

For $\rm Log_{10}[T_{vir}]$, we implement within the range $[4.0, 5.3]$ corresponding to $T_{\rm vir} \in [10^4, 2 \times 10^5]$. The atomic cooling threshold defines the lower bound. The upper bound is set via Lyman break observations, these stem from the ionised gas threshold that the Lyman radiation must overcome to be observed. Feedback mechanisms are relevant here as above this threshold galaxies are too small to continue hosting star formation (Fialkov et al., 2014b). The exponential drop off of the halo-mass function used in calculating $f_{\rm coll}$ means that changes to the top end of this prior distribution cause insignificant changes to the Evidence.

The parameter $R_{\rm mfp}$ is motivated via simulation as the excursion-set formalism is scale dependent. In this context, a photon's mean free path should depend only on the recombination rate of the IGM HI at that instant. Therefore we must identify the neutral hydrogen fraction within these scales to accurately capture each galaxy cluster's sphere of influence. Since typical bubbles in the 21CMFAST simulation are $\sim 10h^{-1}$Mpc, we expect $R_{\rm mfp}$ to lie around this (discussed further in 3.4.7). To ensure this captures the typical variance of bubbles and to allow flexibility for the scales at which the percolation can proceed, we set $R_{\rm mfp} \in [5., 20.]$ as in Greig & Mesinger (2015). For the majority of this chapter α is fixed to zero for simplicity.

To test the impact of these choices, we explore the skewing results via tampering with the choice of prior distribution. We use both a widened prior and a δ-function (discussed further in Section 3.4.8, Figure 3.21). The widened prior extensions are: $R_{\rm mfp} \in [0., 30.]$, $\zeta \in [0., 1000.]$, $\mathrm{Log}_{10}[T_{\rm vir}]$, and $\alpha \in [-3.0, 3.0]$. For $R_{\rm mfp}$, nothing was gained beyond 20 Mpc - the posterior peak extended linearly to whatever value was set (not shown). Once the excursion set scale is large enough to capture all the bubble sizes in the box, the simulation is unchanged and produces similar results.

For the toy models, the priors are expanded so that the peak of the posteriors are included. This is not physically motivated and some bounds have been selected to prevent machine limitations. From a statistical standpoint, this provides enough information to estimate a minimum and maximum skew that can be introduced to each model's Evidence. If our results can be robust regardless of this skew, these results are a testament towards Bayesian model selection in this proof-of-concept work. We need to know the maximum likelihood point if we are to estimate the maximum Evidence for each model, and since the priors are uniform the constant volume can be used to estimate a lower limit. In Section 3.4.4 (Figure 3.12 in particular) the differences in Bayes factors are large and well beyond any possible skew obtained from altering the parameter priors. Any Occam's razor based penalty due to redundant parameter space from increasing the prior is overwhelmed by being able to integrate the full shape of the likelihood distribution.

Since we have actively tried to skew our results, and have been unsuccessful - we are confident that the conclusions drawn from any Bayes' factor results are valid and that we have chosen sensible prior distributions. Figure 3.21 does show points scoring *strong* on the Jeffreys' scale (in the case of the δ-function prior), however these can easily be disputed by physical motivation. Table 3.4 shows the prior distributions used to calculate the Evidence results used in

Model	ζ	R	Log[T_{vir}]	α
3pFZH	[5.,250.]	[5.,20.]	[4.0,5.3]	-
4pFZH	[5.,250.]	[5.,20.]	[4.0,5.3]	[-3.,3.]
InvFZH	[5.,5000.]	[0.1,10.]	[4.,7.]	-
MHR	[5.,4000.]	-	[4.,7.]	-
Inv MHR	[5.,1200.]	-	[4.,6.]	-
F MHR	[5.,1000.]	[0.1,10.]	[4.,6.]	-
F Inv MHR	[5.,1200.]	[0.1,10.]	[4.,6.]	-

Table 3.4: The uniform prior ranges used for the Evidence values calculated in this chapter. See Section 3.2.1 for a detailed description of these models and their parameters.

this chapter. In general, as long as the peak of the likelihood has been captured by the prior, the Bayes factor results are easily repeatable. For the MHR models we add in two reparameterisations (detailed in Section 3.2.1) in order to cross-check whether or not Occam's razor is implemented in a satisfactory way. These are: 1p MHR - the one parameter version, dependant only on the neutral fraction; and F MHR - the 3 parameter version, where a Gaussian filter is initially run accross the density field. These results are discussed in 3.4.2. Finally we calculate the Evidence for $\zeta_{\rm eff}(\zeta, \alpha)$ when comparing the 4pFZH and 3pFZH mock data sets in Section 3.4.8.

Priors are best when uninformative. Typically the choice is therefore between a uniform and a Log (Jeffreys') prior. To cross check our choice of uniform prior distribution the $\mathcal{Z}$ calculation for 3pFZHf1, 3pFZHf2 and 3pFZH(with $\alpha = 0.4$)f2 was repeated with a log (Jeffreys' prior) accross the same boundaries as discussed above. Identical posterior distributions and $\mathcal{Z}$ values within the computational error bars provided by MultiNest were obtained.

3.4 Results

All results (unless specified) are tested against a mock power spectrum observation simulated by FZH with one of two sets of fiducial parameters, namely: f1 [$\zeta = 20.$, $R_{\rm mfp} = 15.$ Mpc, $T_{\rm vir} =$ 30000 K]; and f2 [$\zeta = 15.$, $R_{\rm mfp} = 15.$, $T_{\rm vir} = 50000$, $\alpha = 0.4$]. These respectively represent a bright and faint galaxy population as is done in Greig & Mesinger (2015), the original 21cmMC. We emphasise that f1 is made by 3pFZH and f2 by 4pFZH.

In this chapter, the 21cmFast models collectively refer to the use of FZH simulations, with

3p referring to those with $\alpha = 0$ (therefore implying Equation 1.15) and 4p to any with $\alpha \neq 0$ (Equation 3.7). The toy models refer specifically to FZH, Inv FZH, MHR, Inv MHR, F MHR and F Inv MHR collectively, where FZH and Inv FZH refer to the 3p case (with a constant ionising efficiency). For example, the notation 3pFZHf1 (or FZHf1) refers to a 3 parameter FZH model fitting for the f1 mock power spectrum simulated with the first set of fiducial parameters.

3.4.1 Direct comparisons of 21cmNest and 21cmMC

We first reproduce the results of 21cmMC as the main check of our methodology, i.e. can 21cmNest produce the same posteriors. These are done in comparison to the two mock observations (f1 and f2) measured across three redshifts ($z = 8, 9, 10$). As is done in Greig & Mesinger (2015), we recover the faint and bright galaxy parameter sets successfully. Figure 3.5 shows the overlaid posterior distributions from 21cmMC and 21cmNest for all the 21cmFast model results. Table 3.5 shows the relevant parameter statistics. The small areas in blue away from the posteriors are called islands and are referred to in the Emcee API. They are regions where small groups of walkers have pooled together enabling detailed balance to be solved locally without finding the posterior peak. They typically cost Emcee a 2% reduction in efficiency. We include them here for transparency in our comparison of the posteriors produced by each algorithm. Importantly when these islands are ignored with a likelihood cut, the posteriors of Emcee and MultiNest are matched within statistical errors.

Given we have a small number of parameters, MultiNest is a more efficient sampling algorithm than Emcee. Less likelihood evaluations are necessary due to the weighting of points calculated by the change in prior volume in the algorithm. In comparison, Emcee has obtained the parameter posteriors with more likelihood evaluations and it does not directly provide the Evidence required for model selection. To implement model selection from MCMC chains directly an additional algorithm such as MCEvidence (Heavens et al., 2017a) must be used. Since MCMC chains are correlated by construction (discussed in Appendix C.2.1), this is already an uphill battle and it requires an order of magnitude more points than nested sampling. As we found in Chapter 2, correlation between the parameters worsens this and any Evidence values becomes too unreliable to use. At low dimensionality (up to 4 is considered in this chapter) MultiNest performs as well as Emcee for parameter estimation purposes also. On a 20 core machine, Emcee produced 39000 likelihood evaluations in 32 hours while MultiNest ob-

tained 50031 samples in 15 hours (both with 3pf1). For the problems at hand in this chapter, MULTINEST is more efficient and therefore faster. Beyond these toy models however this is no longer the case due to the increase in dimensionality penalising the methodology in nested sampling (discussed in Appendix C.2.1).

***Cross-check*: the effective ionising efficiency, $\zeta_{\rm eff}$**

Here we cross check the physics methodology by analysing $\zeta_{\rm eff}$ from the 21CMFAST models (Section 3.2.1). We recapitulate that the 3pFZH model ($\alpha = 0$) is identical to the relationship in Equation 1.15, while the 4pFZH model use Equation 3.7 ($\alpha \neq 0$).

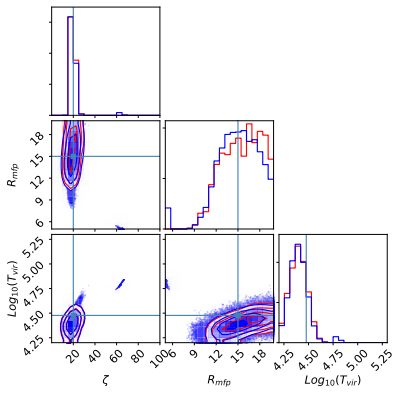
Figure 3.5 in the previous section shows agreement between the two statistical algorithms well. However in 3.5(b) both posteriors peak at a value that is different from the fiducial (shown by the pale blue lines). This is because the FZH model prescribes the aforementioned effective ionising efficiency $\zeta_{\rm eff}$. Since the f2 mock observation was created with $\alpha \neq 0$ the 3p model cannot retrieve the exact fiducial ζ value that satisfies this specifc power spectrum.

The consistency between the ζ of 3pf1 and the $\zeta_{\rm eff}$ of 4pf1 in Table 3.5 shows the reverse of this skew. Adding the α parameter allows a compensation between ζ and $\log_{10}[\mathrm{T_{vir}}]$. Therefore the standard deviations of ζ (σ_ζ) are much larger when α is included as a parameter in the 4p models. In particular we note that when the ζ posteriors do not match the fiducial, the $\zeta_{\rm eff}$ do (within σ_ζ).

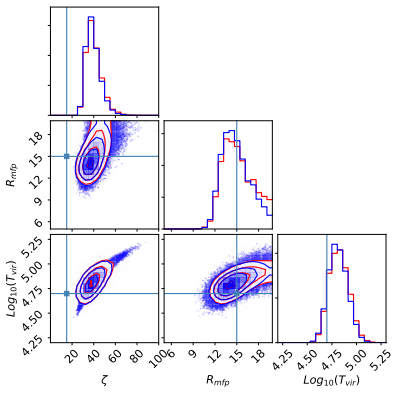
On top of this, any deviations from recovering the exact fiducial values are due to the observational priors inputted in the likelihood. However these do not sway the parameters much since all the 21cm power spectra used fit within the current EoR constraints. We look at this directly in Section 3.4.6.

3.4.2 *Cross-check*: varying parameter number with the MHR models

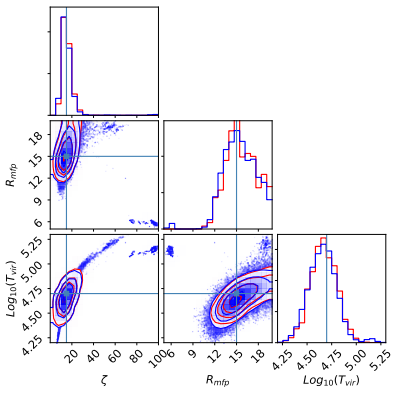
To cross-check our statistical methodology we created a third mock fiducial data set using Inv MHR with $[\zeta = 30, \mathrm{Log}[\mathrm{T_{vir}}] = 4.5]$. Figure 3.4.2 clearly shows the input model has been recovered with *strong* evidence. This is further proof of concept that Bayesian model selection is successful in distinguishing simple models of the EoR.



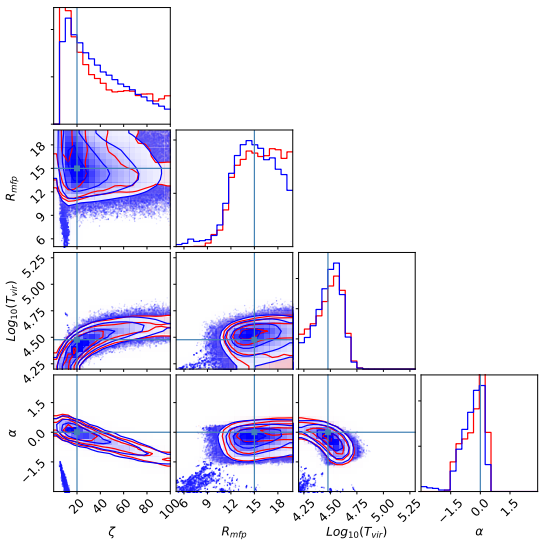
(a) 3pFZHf1



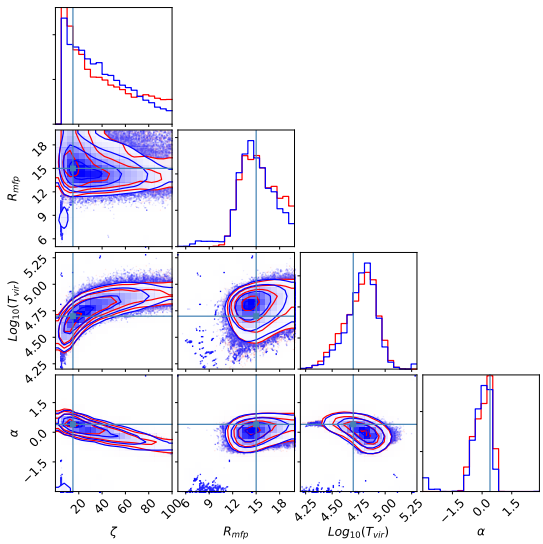
(b) 3pFZHf2



(c) 3pFZH($\alpha = 0.4$)f2



(d) 4pFZHf1



(e) 4pFZHf2

Figure 3.5: Here we show agreement between parameter posteriors from 21CMNEST (red) and 21CMMC (blue) for the 3 and 4 parameter 21CMFAST models. 3.5(a) The simple model can re-obtain its own fiducial parameters (shown by the pale blue lines). 3.5(b) When fitting for the f2 mock observation, the 3p model requires ζ to compensate. This is seen by an offset which between the blue fiducial value and the ζ posterior (and is quantified in Section 3.4.1). 3.5(c) Fixing $\alpha = 0.4$ allows the 3p model to easily recover the 4p fiducial parameters used for the f2 mock observation. As expected the 4p parameter can easily fit both mock observations, due to the flexibility provided by ζ's dependency on the halo mass (Figures 3.5(d) and 3.5(e). See Table 3.5 for MAP parameters corresponding to these plots. The 'islands' of points away from the modes (only in blue) are a pitfall of using EMCEE. Walkers have become stuck in regions of low likelihood (expected in the EMCEE API for more detail). Typically they contain $\sim 2\%$ of the total samples. Publications involving EMCEE typically have these removed with a likelihood cut.

	ζ	$\log_{10}[T_{vir}]$	α	ζ_{eff}
3pFZHf1	18.7 ± 2.0	4.366 ± 0.09	(0.0)	18.7
4pFZHf1	5.8 ± 27.6	4.148 ± 0.16	0.36 ± 0.52	6.6
f1	20. ± 5.	4.447 ± 1.1	(0.0)	20.
3pFZHf2	34.9 ± 3.2	4.747 ± 0.05	(0.0)	34.9
3pFZHf2	12.2 ± 4.2	4.576 ± 0.13	(0.4)	66.1
4pFZHf2	5.3 ± 27.3	4.473 ± 0.15	0.624 ± 0.40	7.8
f2	15. ± 3.8	4.699 ± 1.2	0.40 ± 0.10	39.6

Table 3.5: The MAP parameter values $\pm 1\sigma$ for the 5 simulations performed with permutations of 3pFZH and 4pFZH fitting for their mock observations f1 and f2 (Section 3.4.1). The posteriors are shown in Figure 3.5. ζ_{eff} is the actual ionisation efficiency parameter (that the ensemble of galaxies produces, via Equation 3.8). This inclusion of the power law in halo-mass causes this to deviate from ζ (the input parameter) when $\alpha \neq 0$. For the fiducial values, f1 and f2, σ_ζ has been approximated by using 25% of each value as suggested in Mesinger et al. (2011), motivated by the discrepancies between semi-numerical and radiative transfer simulations.

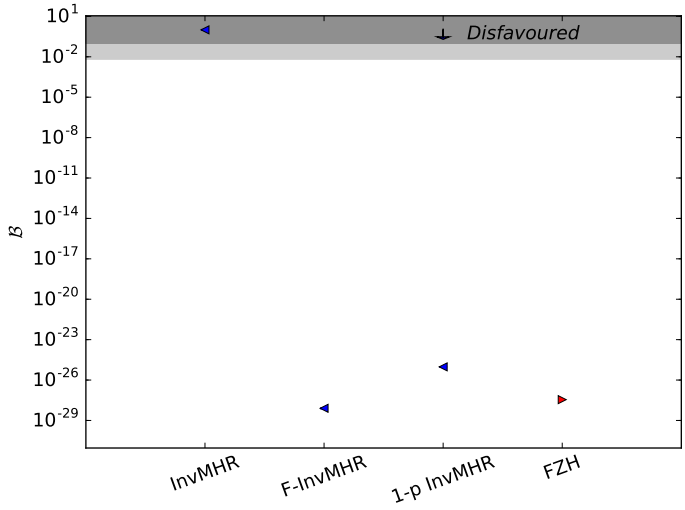


Figure 3.6: Bayes factors for the MHR based models. Only the input model is able to retrieve the mock power spectrum (created by Inv MHR). FZH has been included as this shares morphology with Inv MHR. See Section 3.4.2 for detail.

We also assess the choice of prior width with the reparameterisations of MHR here. The 1p Inv MHR is a simplified version of Inv MHR. Since it has a vastly reduced prior space (reduced from $\sim$ 2000 to 1) it seems a good place to check how heavily the Occam's razor has been implemented by the methodology. It is clear in Figure 3.6 that the reduction in predictive power is not dominated by the reduction in prior volume. Without the use of $f_{\rm coll}$, there is no intrinsic redshift dependence in 1p Inv MHR and it is comfortably rejected. The cosmological parameters required to calculate Equation 1.12 make redshift dependence implicit in the other MHR models. It is therefore comforting that we reject a model whose neutral fraction does not decrease with redshift as reionisation progresses.

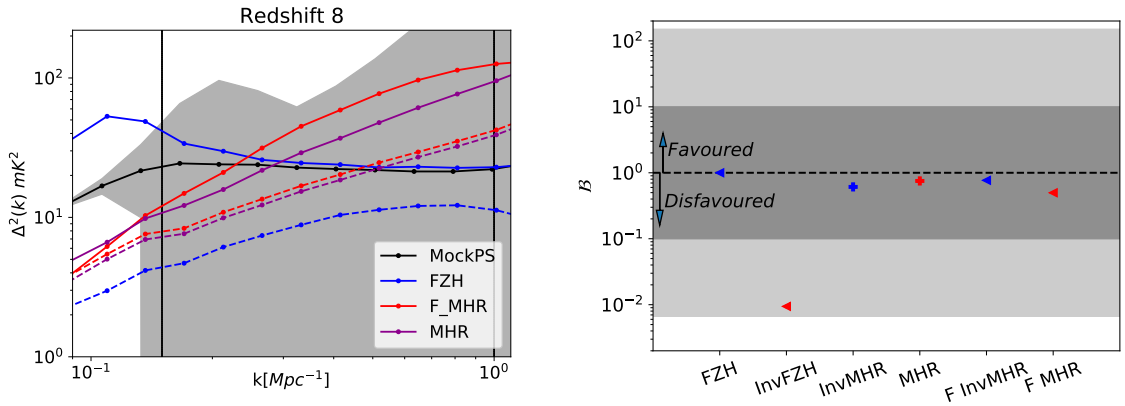
On the other hand, F MHR has an increased parameterisation and can smooth the density field. The addition of $R_{\rm Filter}$ allows F MHR to adapt to the scaling of the ionised bubbles but since the prior of $R_{\rm Filter}$ restricts it from becoming zero, the added parameterisation worsens the fit. When fitting for f1 this added parameter makes the F MHR and F Inv MHR harder to distinguish than the unfiltered MHR and Inv MHR. But in the context here the mock data from Inv MHR is made with no filter scale, therefore increasing the flexibility of this model increases the prior volume without any improvement to the fit. The added space is therefore redundant and F Inv MHR is appropriately penalised.

1p Inv MHR and F Inv MHR models achieve respective Bayes factors of 10^{-25} and 10^{-29}, illustrating the points above. For reference 3pFZH obtains 10^{-28} and Inv MHR has $\mathcal{B} = 1$ by definition.

3.4.3 EoR model selection with LOFAR-48

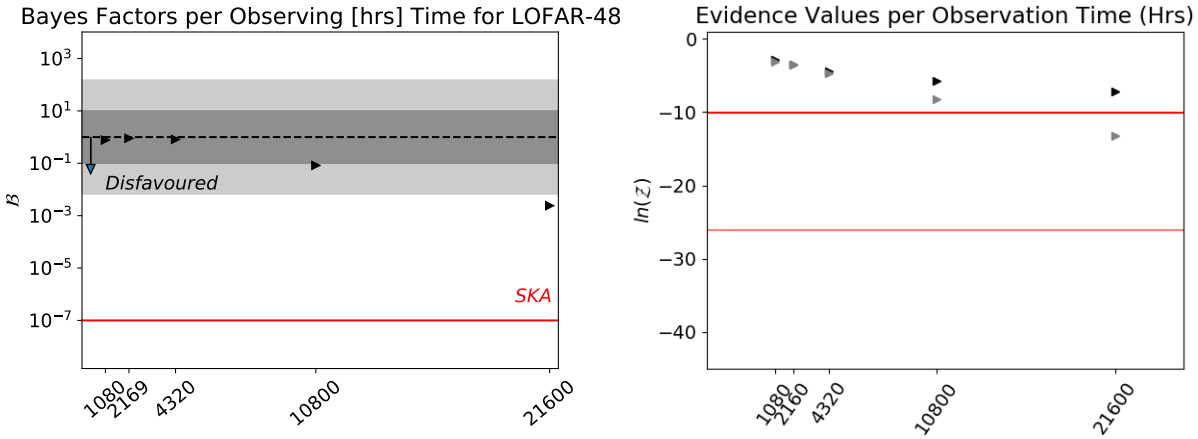
We now proceed in answering the main questions of this chapter. Namely: what length of LOFAR observations is needed to perform Bayesian model selection with the toy EoR models?

We initially use 1080 hours of observing time (Table 3.3). Each observation is a 1 hour tracked scan with 6 hours of observing time per day, corresponds to 6 different fields - which average together to produce the f1 power spectrum. The Bayes factors and MAP power spectra ($z = 8$) are shown for the toy models in Figure 3.7. Due to large error regions on each power spectrum, *weak* results are obtained for all except the InvFZH model.



(a) (b)

Figure 3.7: MAP power spectra (for $z = 8$) and Bayes factors are shown respectively in 3.7(a) and 3.7(b) when using LOFAR-48 with 1080 hours of observation. In 3.7(a) the dotted lines represent the inverse of the model stated with colours in the legend and the observational error is shown in grey. In 3.7(b) the white, light and dark grey regions represent *strong*, *moderate*, and *weak* respectively on the Jeffreys' scale; Blue/red points indicate *outside-in*/*inside-out* morphology, while </+ shapes indicate a global/local implementation. Note that the error power is much larger than the 21cm signal power spectrum. This is reflected in the Bayes factors as no models are *strongly* ruled out (all points exist within the grey regions). The fit of the mock observation can vary vastly due to the poorly constrained mock observation, except for the case of the Inv FZH model which is disfavoured *moderately* in 3.7(b).



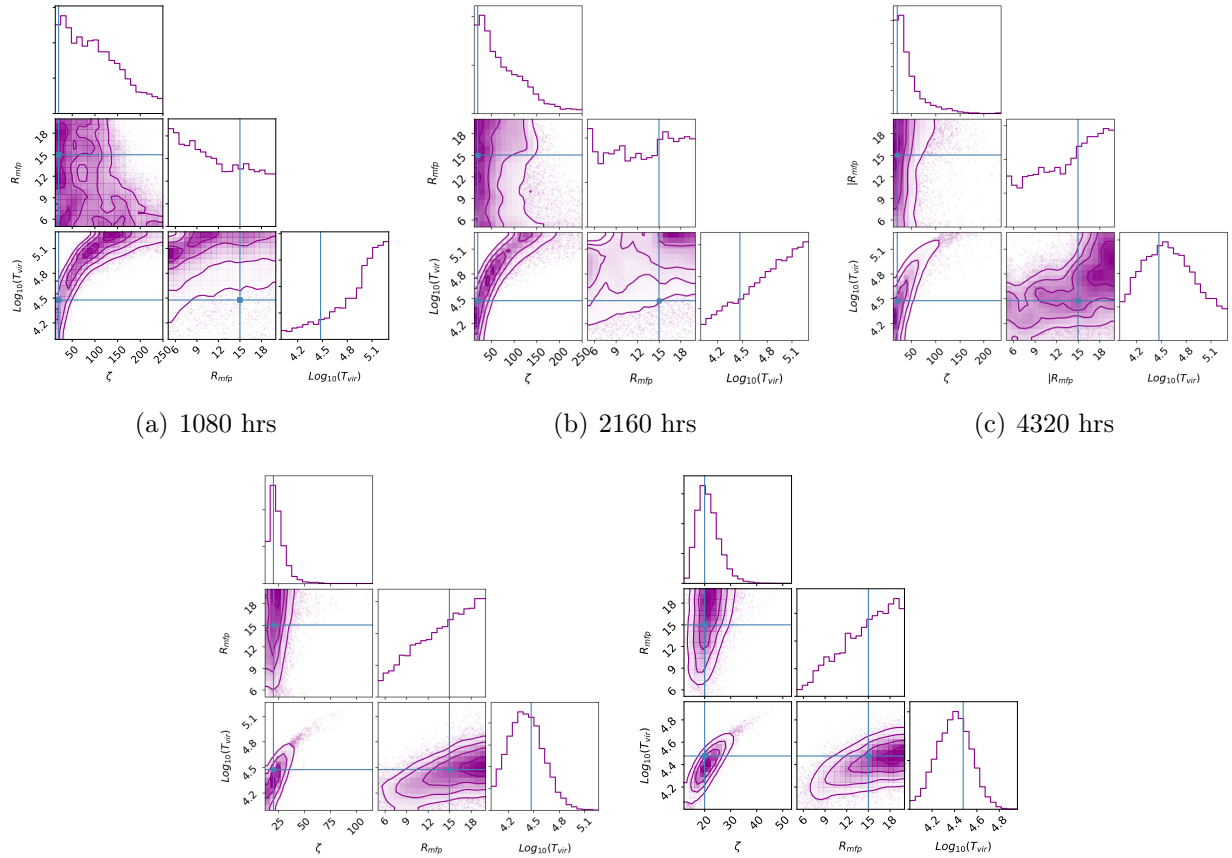
(a) (b)

Figure 3.8: 3.8(a) shows the Bayes factors of 3pFZHf1 against F Inv MHRf1 at with different LOFAR-48 observation lengths. Each Evidence value of 3pFZHf1 corresponds to integrating a posterior in Figure 3.9. The red line(s) shows the Bayes factor (Evidences) using the SKA-512 with 1080 observing hours. In order to score *strong* on the Jeffreys' scale, ruling out the toy models, LOFAR must observe for 21600 hours (indicated by the marker passing into the white disfavoured region). 3.8(b) shows the Evidence values obtained with 3pFZHf1 (F Inv MHR) in with black (grey) points. This comparison is chosen as these models share morphology and scale (and are the hardest toy models to distinguish). The red lines represent values obtained using SKA-512 (light and dark corresponding to 3pFZHf1 and F InvMHR Evidences respectively).

The InvFZH model obtains *moderate* results because of penalties from the observational priors (see Section 3.4.6, for more discussion). Many different power spectra are capable of fitting within the grey error region of the LOFAR-48 mock signal, and this is reflected in the width of the posteriors in plot 3.9(a). Note that with this 1080 hour observation, the posteriors for $R_{\rm mfp}$ and $\rm Log_{10}[T_{vir}]$ have not peaked at the fiducial parameter values.

To remedy this, we vary the observation timescale. LOFAR achieves signal to noise ratios (to the nearest integer) of 1, 2, 3, 7, and 11 with 1080, 2160, 4320, 10800, and 21600 hours of integration time respectively at $z = 8$. Posteriors for 3pFZHf1 are shown in Figure 3.9. We require a total of 21600 hours of observing time to obtain *strong* conclusions when distinguishing the toy models - Figure 3.8(a). The posterior shown in Figure 3.9(e) (for a 21600 hour LOFAR-48 observation) agree with the posteriors from the 1080 hour HERA-331 observation in Figure 3.5(a). These selected observing timescales increase along a very ambitious trajectory since LOFAR has observed a total of $\sim$ 1300 hours of the NCP field in its 7 years of EoR activity. To summarise, we think model selection is unlikely to be achievable with LOFAR because of the limited k-range in which it is sufficiently sensitive.

These calculations use 21cmSense's *moderate* foreground settings (Section 1.3.7) for LOFAR to ensure our analysis is fair when comparing these results with HERA and SKA. However, LOFAR aims to observe within the foreground wedge (i.e. in a k range that extends lower than $k = 0.15$ Mpc^{-1}) - the results published in Patil et al. (2017) are within $k = [0.05,\ 0.13]$ Mpc^{-1}. This implies these results are conservative estimates, and for the sake of argument we repeated our 1080 hour LOFAR analyses with an extended range of $k = [0.05,\ 1.]$ Mpc^{-1} (but still with *moderate* foreground removal). The obtained parameter posteriors are comparable to the 21600 hour observations in the previous k range ($[0.15,\ 1.]$ Mpc^{-1}). In the context of model selection, the results gain *moderate* but not *strong* inference in ruling out F Inv MHR ($\mathcal{B} = 148$) and so decisive model selection would still be challenging for LOFAR. Ideally LOFAR will be operational without any foreground contamination (the *optimum* 21cmSense setting) in the extended k range used above. If these criteria are achievable, LOFAR can obtain $\mathcal{B} \approx 120000$ and is likely able to perform model selection on the same timescales as HERA and SKA. Our conclusions stick within the conservative estimates in order to maintain a consistent analysis between the telescopes.



(a) 1080 hrs (b) 2160 hrs (c) 4320 hrs

(d) 10800 hrs (e) 21600 hrs

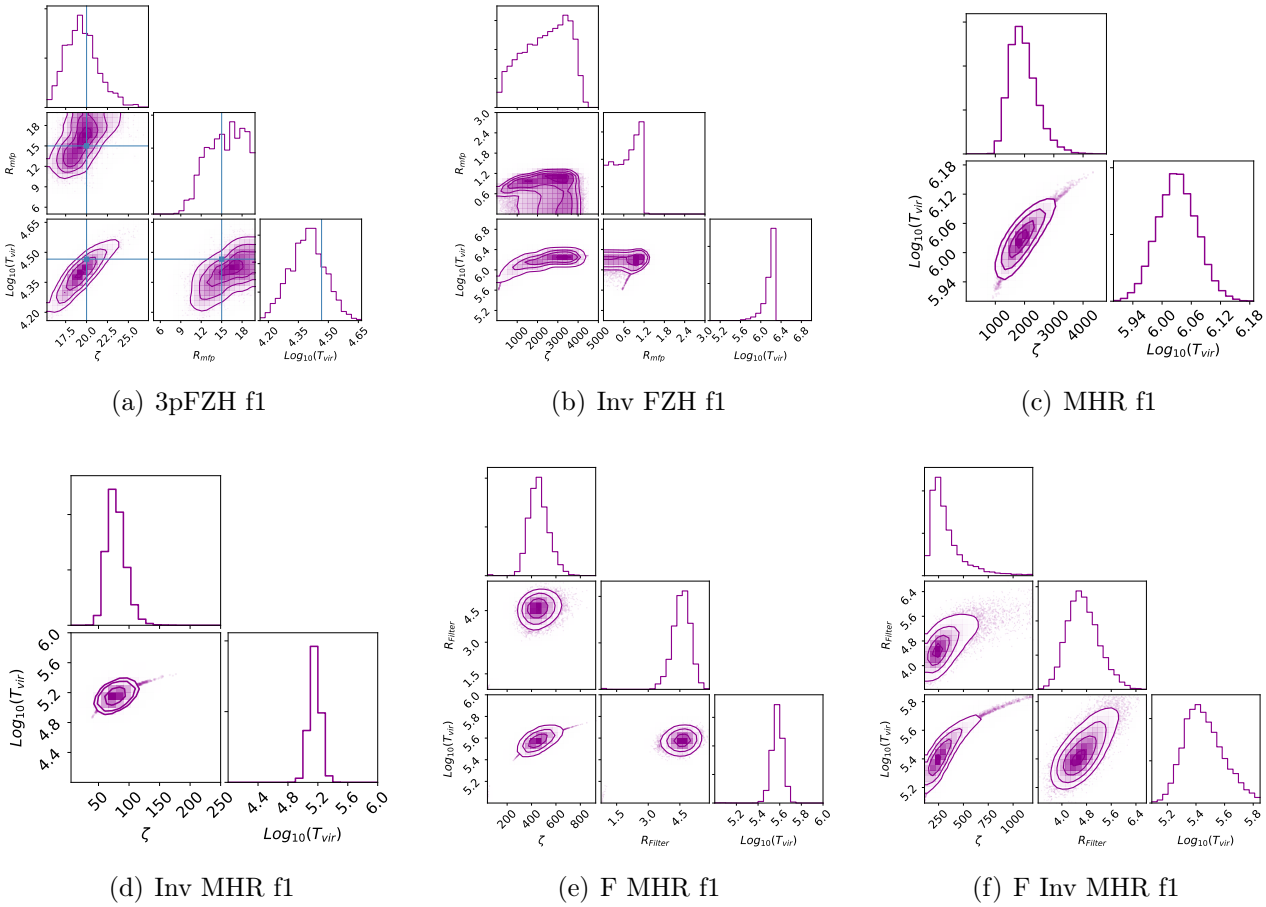
Figure 3.9: Posterior distributions for the 3pFZH parameters used to recover the f1 mock observed power spectrum with LOFAR-48 at with various observation lengths. Figures 3.9(a), 3.9(b), 3.9(c), 3.9(d), and 3.9(e) use 1080, 2160, 4320, 10800 and 21600 hours respectively. Only the 21600 hour observation provides the posterior distributions with the fiducial parameters comfortably within the 1σ contour. At 6 observing hours per night 21600 hours corresponds to $\sim$ 10 years of use with LOFAR-48 (assuming constant good weather and no other instrumental intricacies).

3.4.4 EoR model selection with HERA

Here we attempt to distinguish the toy models against the mock observed f1 power spectrum (generated with 3pFZH). We show the obtained parameter posteriors and MAP power spectra in Figures 3.10 and 3.11 respectively. As mentioned in Section 3.3 the prior ranges shown in Table 3.4 go beyond physical bounds to contain the peaks of the likelihood distributions. Referring to the posteriors, adding the filter scale to the MHR models smooths the density field, producing over-densities that are fewer in number but larger on all scales. Therefore the (background) ionising efficiency ζ must compensate by increasing in the F Inv MHR case (plot 3.10(f) has a larger ζ than 3.10(d)) so that over-densities can be ionised at a similar rate. The opposite is true for the MHR and its F MHR (plot 3.10(e) has a smaller ζ than 3.10(c)). In this case, the background ionising radiation is less efficient when ionising the smoother density field. A qualitative reflection can be interpreted from the sizes of the bubbles (signal-less regions, Figure 3.2), forming smaller and more numerous bubbles requires a surplus in faint small galaxies. By construction, all models contain some degree of degeneracy between ζ and $\mathrm{Log}_{10}[\mathrm{T}_{\mathrm{vir}}]$. This is observed by the arcs of varying length in the corresponding 2d posterior plots.

The toy models predict vastly different reionisation morphologies and therefore all can be distinguished with *strong* evidence on the Jeffreys' scale. This is shown by the Bayes factors in Figure 3.12. The hardest to distinguish model proved to be F Inv MHR. This is because FZH (creating the mock observation) and F Inv MHR share both morphology and scale of reionisation. Visibly, they are also the most similar in Figure 3.2. In contrast, Inv MHR has the *inside-out* morphology, but is *local* rather than *global* in scale - this explains why it is rejected more heavily than its filtered version in Figure 3.12. The addition of a filter scale within the MHR models clearly increases the flexibility of our *local* based models. To reiterate, the *local inside-out* Inv MHR model proves a challenging test when the *local* constraint is relaxed by smoothing the density field. We also add that when the morphology is different, the relative similarity of Inv FZH, MHR and F MHR (red points) implies little compensation is achievable by varying the scale of reionisation.

Next we varied the number of HERA dipoles for the 1080 hour observation (maintaining *moderate* 21cmSense foreground settings). In particular we looked to find what number of dipoles are required for HERA to attain the model selection capabilities of LOFAR or the SKA. As



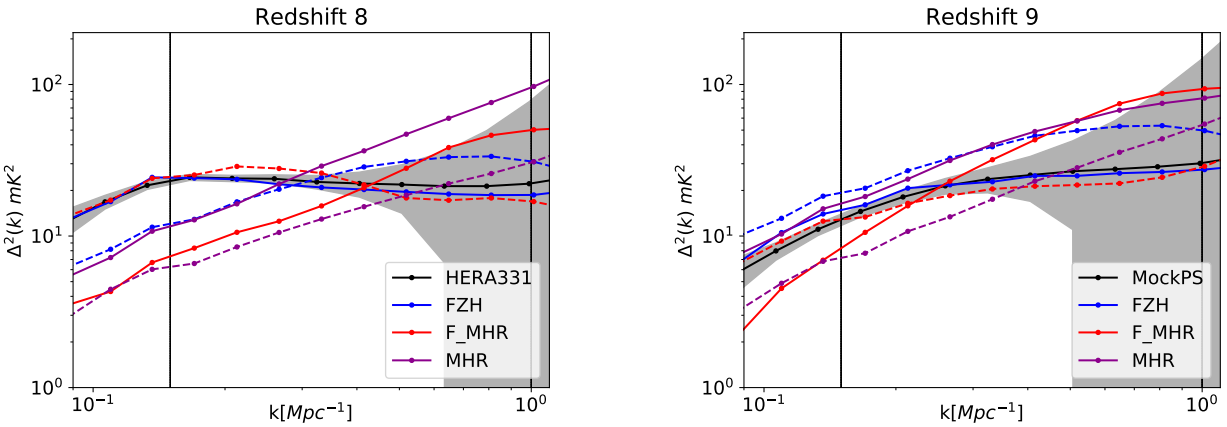
(a) 3pFZH f1 (b) Inv FZH f1 (c) MHR f1

(d) Inv MHR f1 (e) F MHR f1 (f) F Inv MHR f1

Figure 3.10: Parameter posteriors produced by 21CMNEST for each toy EoR model (as they fit for the f1 mock observation) using HERA-331 with 1080hrs of observation. The MAP values of these distributions are shown in Table 3.6, and the MAP power spectra are shown in Figure 3.11. The blue lines correspond to the parameter values used to produce the f1 data set, they are only comparable to the 3pFZH.

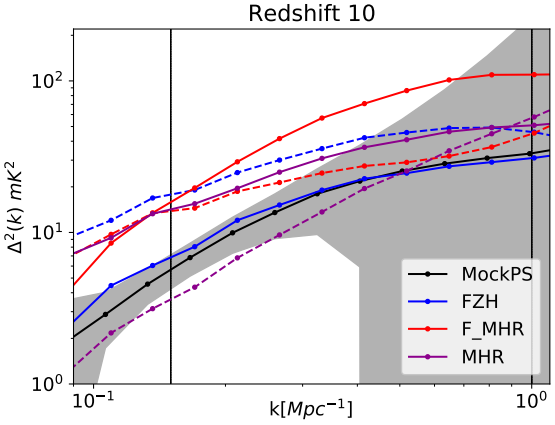
Model	ζ	R	Log[T_{vir}]
FZH	18.8 ± 2.0	14.4	4.38 ± 0.09
Inv FZH	1680 ± 1080	1.05	6.16 ± 0.1
MHR	1690 ± 470		6.02 ± 0.04
Inv MHR	71. ± 15		5.12 ± 0.07
F MHR	432 ± 80	4.65	5.56 ± 0.06
F Inv MHR	203 ± 170	4.35	5.3 ± 0.1

Table 3.6: The MAP values ($\pm 1\sigma$) for every toy model's fit to the FZH mock observation (f1). These results correspond to those plotted in Figures 3.10, 3.11, and 3.12 when observing for 1080 hours with HERA-331.



(a)

(b)



(c)

Figure 3.11: MAP 21cm power spectra at redshifts 8, 9 and 10 (3.11(a), 3.11(b) and 3.11(c) respectively) for the FZH (blue), Inv FZH (dashed blue), MHR (magenta), Inv MHR (dashed magenta), F MHR (red), and F Inv MHR (dashed red) models. The MAP parameter values are shown in Table 3.6 and their posteriors are shown in Figure 3.10. In the likelihood, the χ^2 combines these three measurements between $k = [0.15,\ 1.0]$ Mpc (the vertical lines). The f1 mock-observed power spectrum, created by 21CMFAST (3pFZH), are shown in black with the grey region identifying the addition of the 1080 hour HERA 331 noise PS shown in grey. Note that these power spectra have very different shapes i.e. only FZH model which created the mock observation is within the error bars for every redshift. Because of this, there are a vast range of Bayes factors shown in Figure 3.12.

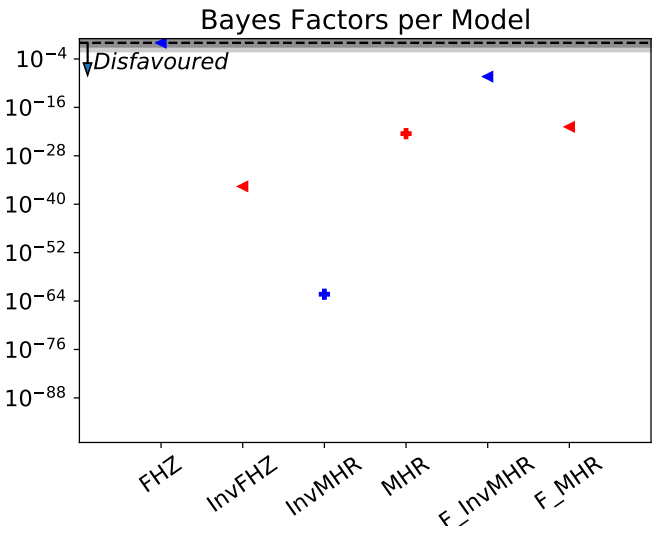
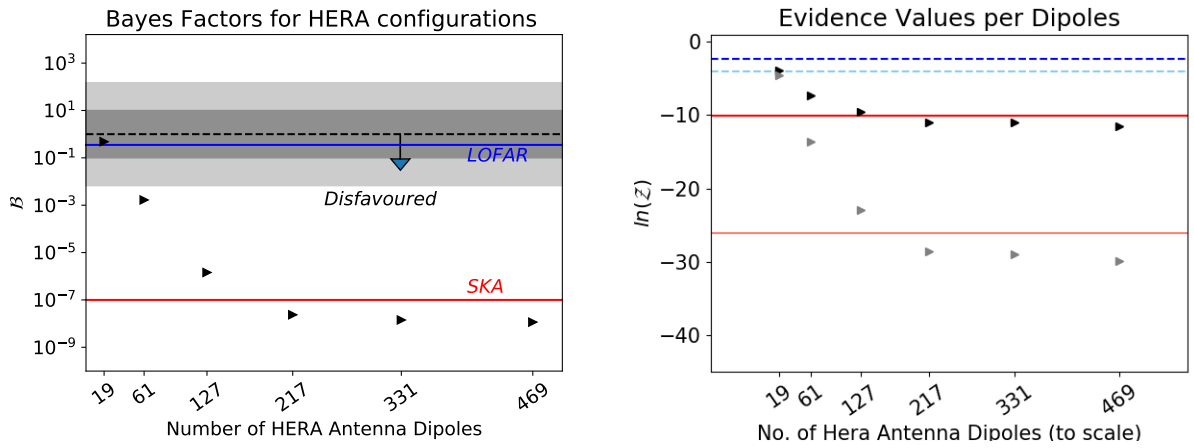


Figure 3.12: The Bayes factors for the toy models when using a 1080 hour observation with HERA-331. Blue/red colours indicate an *inside-out*/*outside-in* morphologies respectively. The points $<$ and $+$ represent implementation of the ionisation threshold on *global* and *local* scales respectively. All toy models are distinguished with *strong* evidence (all competing models are in the white region). Inv MHR (*local inside-out*) and the Inv FZH (*global outside-in*) are easy to distinguish because they differ in scale or morphology to the mock simulation (*global inside-out*). The corresponding power spectra and parameter posteriors are plotted in Figures 3.10 and 3.11 respectively.

mentioned in Section 1.3.7, HERA's structure is organised as hexagonal with redundant baselines aligned to take repeated measurements of each unique baseline in order to reduce the noise on the collective measurement (Dillon & Parsons, 2016). HERA achieves signal to noise ratios of 3, 9, 18, 28, 39, and 50 with 19, 61, 127, 217, 331, and 469 dipoles respectively at $z = 8$ on the f1 power spectrum.

Figure 3.13 shows that HERA-19 (the lowest dipole configuration) performs similarly to LOFAR at model selection. From the use of 61 dipoles, the observation is able to distinguish the toy models with *strong* evidence. HERA-217 becomes comparable to the SKA (see Section 3.4.5), while HERA-331 (and higher) are able to produce larger Evidence values than the SKA (when just using the power spectrum).



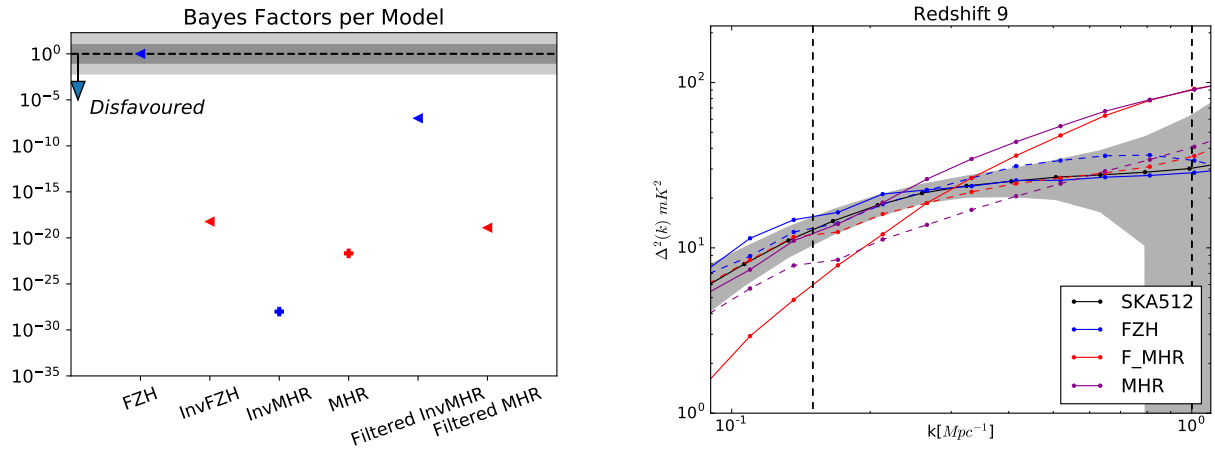
(a) (b)

Figure 3.13: This plot is similar to that for LOFAR in Figure 3.8 but with varying number of HERA dipoles. 3.13(a) and 3.13(b) show the Bayes factors and Evidence respectively for F Inv MHR against 3pFZH (both fit for the f1 mock observation produced by 3pFZH). The dotted and solid lines represent values obtained using LOFAR (blues, dotted) and SKA-512 (reds, solid) respectively; with dark and light colours representing the values obtained for 3pFZH f1 and F InvMHR f1 respectively. LOFAR and HERA-19 score *weak* on the Jeffreys' scale. All other telescope configurations (including HERA-61) score *strong* results. Above 217 dipoles HERA gains more sensitive to the 21cm PS over the selected redshift range than the SKA.

3.4.5 EoR model selection with SKA-512

The MAP PS and Bayes factors for the toy-models when observing with the SKA-512 are shown in Figures 3.14(b) and 3.14(a) respectively. As with HERA, these are a 1080 hour observation for a 1 hour tracked scan of the f1 power spectrum with 6 hours of observing time per day (across 6 fields). The results found are similar to the observation for HERA-331 (i.e. the posteriors obtained by SKA512 are also similar to the HERA-331 distributions in Figure 3.10).

We then repeat the analyses that were done for LOFAR-48 in Section 3.4.3 but with observing time decreasing instead of increasing. At $z = 8$ SKA achieves signal to noise ratio of 4, 7, 10, 14, and 22 with 108, 216, 324, 540, and 1080 hours of integration time. Figure 3.15(a) shows the Bayes factors and Evidences obtained for the different observation times used. As before, we have assumed the moderate foreground model for the wedge and buffer within 21cmSense and the Bayes factors are calculated between 3pFZHf1 and the F Inv MHRf1 since this toy model is the hardest to distinguish. We find that 324 hours of observation is enough to distinguish the toy models with *strong* evidences. The posterior distributions obtained are not shown (they behave in a similar way to those from LOFAR in Figure 3.9 but for the five selected observing

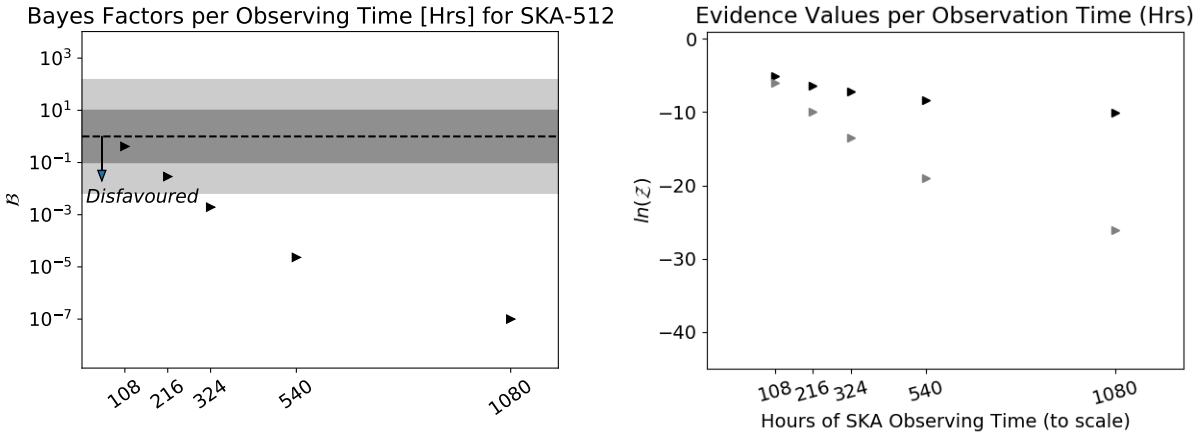


(a) (b)

Figure 3.14: A summary the results for SKA-512. 3.14(a) shows the Bayes factors obtained with SKA of each toy model against 3pFZH as they fit for the f1 power spectrum. All toy models are *strongly* ruled out (by margins similar to those obtained by HERA-331). For details of the plot style in 3.14(a) please see Figures 3.7(b) and 3.12. 3.14(b) shows the MAP parameter PS at $z = 9$. for the toy models fit against the f1 mock data set with 1080 hours of observation. This shows the SKA achieving similar error bars in comparison to HERA-331 in Figure 3.11.

times in decreasing order).

In comparison to HERA, SKA has significantly more independent baselines but they are each measured with a marginally lower signal to noise. It is still an active area of research as to whether a more richly populated U-V space or a precisely measure sparse U-V space will be better for practical measurement so the power spectrum. Despite this SKA does not achieve higher Evidence values than HERA-331, when considering the power spectrum however this may not remain the case when including higher order statistics such as the bi-spectrum (Watkinson et al., 2019). Using a well interpolated U-V space will certainly be necessary if statistics of higher order than the power spectrum are to be utilised fully. For a direct comparison of the error PS from HERA-331 and SKA's observation on the f1 mock power spectrum see Figure 3.4. In conclusion, SKA performs similarly to HERA-217 at EoR model selection with the 21cm PS (Figure 3.13(a)).



(a) (b)

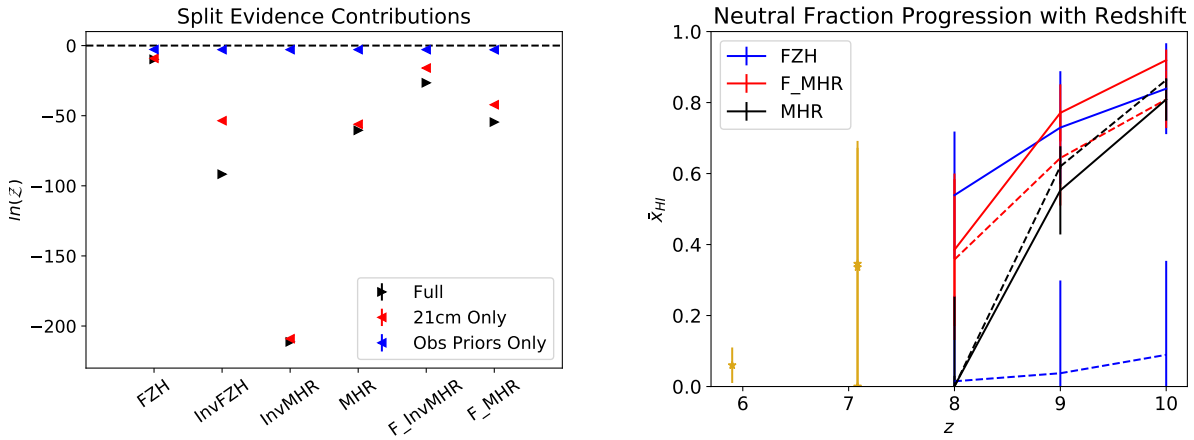
Figure 3.15: 3.15(a) Bayes factors for comparing 3pFZH and F Inv MHR as the produce the f1 mock observation (produced by 3pFZH). 3.15(b) the Evidence values obtained for 3pFZH/F Inv MHR in black/grey. Both follow the same conventions as Figures 3.8 and 3.13 but varying SKA-512 observing times. Note that in order to obtain a *strong* ruling out between the toy models, SKA must observe for at least $\sim$ 324 hours (the minimum observation for the most similar toy model to pass out of the grey Jeffreys' scale region).

3.4.6 The Evidence as an inference measure: the 21cm power spectrum vs alternate reionisation data

In this section we discuss the contributions to the Evidence from the observational priors. These are the Greig, McGreer and Planck contributions to the likelihood in Equation 3.1 and we compare them with the f1 21cm PS mock observation with 1080 hours of HERA-331 working at *moderate* foreground avoidance. Since the observational priors are linearly input as part of the likelihood they are easily separable. When the 21cm likelihood or the likelihood constructed from the observational prior checks are allowed to vary individually they can all fit the neutral fraction checks well. By separating out the 21cm power spectrum contribution, χ^2_{21}, from the Planck and QSO data, $\chi^2_{\rm Planck} + \chi^2_{\rm McGreer} + \chi^2_{\rm Greig}$, it becomes apparent how little quantitative impact these observations have on model selection in the EoR. This is observed as the lack of deviation of the blue points from the black dashed line in Figure 3.16(a). These blue points have been run separately from the simulation with only the 21cm power spectrum and can be fit easily by all of our toy models (Table 3.2), which follows from all of the toy models having similar reionisation histories in Figure 3.3(a). All of the toy models posses parameter space that agrees with the constraints of the observational priors. To break this model degeneracy for each model, more data or different data are necessary - the 21cm power spectrum being

a viable tool for the job. The Inv FZH is seen to be the least physical, as it has the most tension between the observational priors and the 21cm PS. The MAP parameters of the Inv FZH model differ significantly (by a few σ, see Figure 3.10) depending on what contributions to the likelihood are included. When recovering f1, we indirectly observe this via σ values that are of a comparable size in ratio to the MAP parameter value (see Table 3.6). Although the observational priors do not do much when χ^2_{21} is present (due to the fiducial parameters in f1 being a physical choice), they are capable of ruling out vastly wrong reionisation models. It is fair to say the observational prior checks definitely constrain a larger region than the 21cm power spectrum, so although they are capable of producing heavy penalties our sensibly chosen prior ranges prevent this from happening. The discrepancies of the black (full likelihood, Figure 3.16(a)) Evidence points compared with the red (only the 21cm Likelihood) show when models are un-physical as they are penalised by the measured data used in these observational prior checks. A better way of harnessing the non-21cm data is required if a fair cross check is to be performed against the 21cm PS. A visualisation of why our redshift dependent neutral fraction measurements from the QSO data are so poorly constraining is shown in Figure 3.16(b). In this Figure, the plotted points are the redshifts from the coeval simulations which are interpolated together to obtain each reionisation history. The error bars on the McGreer and Greig neutral fraction checks (yellow) are simply too large to separate out any of the simulated reionisation histories. It is therefore difficult for them to add any constraining power to our toy models of the EoR. Similarly small likelihood contributions come from the Planck prior (which uses the optical depth). This should be interpreted such that other data (including more numerous and precise neutral fraction measurements), as well as the 21cm power spectrum, is necessary to pin down the inference of model parameters in the EoR. For completeness, we found that the observing telescope has a negligible effect on the variance of neutral fractions obtained. Also, the 4p FZH model fit within the 3pFZH error bars (the blue solid line) i.e. a power law in ionising efficiency has a small impact on the neutral fraction.

We defer a model selection analysis including this improved framework to Chapter 4, where we implement galaxy UV luminosity functions to harness more information (Park et al., 2020). Equation 3.7 is no longer used due to its inadequacy at reproducing observed UV luminosity functions. In the UV LF case, the parameterisation of ζ_0 is expanded via power laws in $f_{\rm esc}$ and f_* instead. We conclude by emphasising that this quantitative approach is in agreement with the qualitative discussion in Greig & Mesinger (2017a).



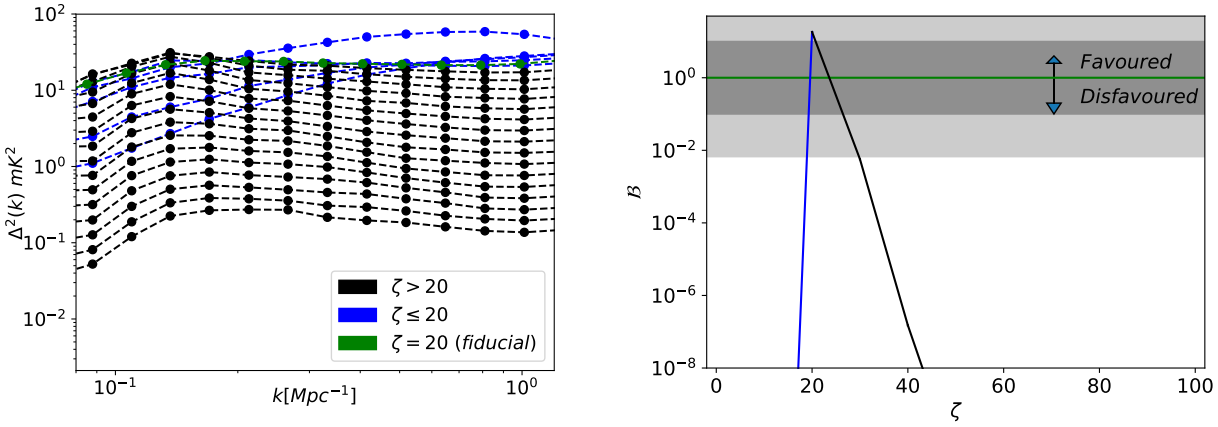
(a) (b)

Figure 3.16: An analysis of the contributions to the likelihood (Equation 3.1). In 3.16(a) we quantify the Bayesian Evidence for the 21cm power spectrum alone compared to the observational priors separately (with a 1080 hour observation of HERA-331). The Evidence from the observation priors (blue) are negligible since their scores lay close to zero compared with the 21cm PS only (red) and combined cases (black). This becomes apparent in 3.16(b) where the reionisation histories for all the toy models are plotted (dotted lines represent inverse models for the colours in the legend). The yellow points at $z = 5.9$ and 7.1 are the McGreer and Greig observations respectively. Since all the error bars on the neutral fractions obtained are large, they have a negligible effect on constraining the models in comparison to the 21cm PS.

3.4.7 Identifying redundant parameters with the SDDR

The Savage Dickey density ratio is detailed in Section 1.4.7. Since we have constructed a framework for calculating the Bayesian evidence for EoR models, nesting parameters within the models can be achieved easily with 21CMNEST. In the following sub-subsections we proceed to test each of the four parameters used in 4pFZH: ζ, $R_{\rm mfp}$, $Log_{10}[T_{vir}]$, and α. We explore whether the parameter under question is redundant in being able to reach meaningful conclusions from Bayesian inference.

The results are displayed in similar Figures (3.17-3.20) where the left shows the variation of the power spectrum while the right shows the parameter dependency of the nested Bayes factors. In all cases the colours are consistent with blue (black) representing below (above) the fiducial value (shown in green). The black and blue lines meet above the green line at $\mathcal{B} = 1$ due to the reduction in prior volume when a parameter is nested. This discrepancy is therefore related to the prior width of the parameter under question. In the ζ case (which has the widest prior Figure 3.17(b)) the Bayes factor rises into the *moderate* Jeffreys' scale threshold. This draws light on the implicit nature of Occam's razor when implementing Bayesian Model selection,



(a) (b)

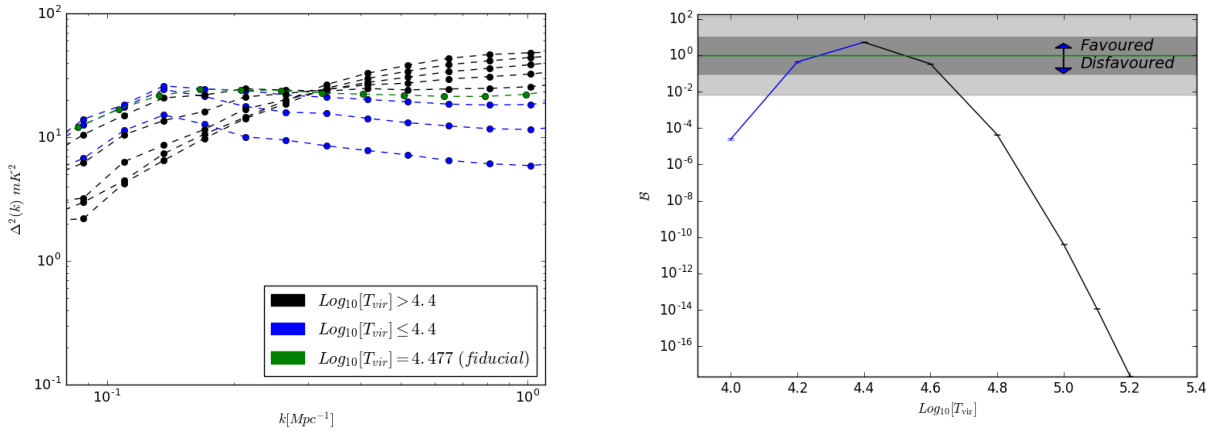
Figure 3.17: Using the Savage-Dickey density ratio - The impact on the Bayes factor when fixing ζ as a set of models nested from 3pf1. 3.17(a) and 3.17(b) show the variation of the 21cm PS and Bayes factors respectively. The colour coding is consistent across both plots in this and subsequent figures: the green indicates the fiducial simulation, black (blue) are values over (under) the fiducial. In 3.17(b) *strong*, *moderate* and *weak* inference scores are indicated by the white, light and dark grey regions respectively. 3.17(a) is shown for $z = 8$ ($z = 9$ and 10 are similar). ζ shows *strong* results rejecting all but the fiducial values.

as mentioned in Section 1.4.3. If in doubt, the variability of the power spectra quantitatively reveals the dynamic range of each parameter.

SDDR with ζ

The Bayes factor results plotted in Figure 3.17(b) show a distribution for $\mathcal{B}(\zeta)$ that has a sharp peak on the fiducial parameter which agrees with the fiducial parameter result shown in green. This is because ζ produces no degenerate power spectra throughout any of its prior distribution (Figure 3.17(a)). It is therefore a very useful parameter. When ζ values are small, the power spectra shows the simulation trying to compensate by increasing the number of small structures (the blue power spectra are largest at the small scale end of Figure 3.17(a)). On the other hand, when galaxies have a high ionising efficiency the simulation responds decreasing the size of structures on all scales (the shape of each black power spectrum is maintained as the amplitude decreases). Within all of ζ's prior range, there are dynamic changes to the simulation that are observable with the power spectrum.

SDDR with $Log_{10}[T_{vir}]$



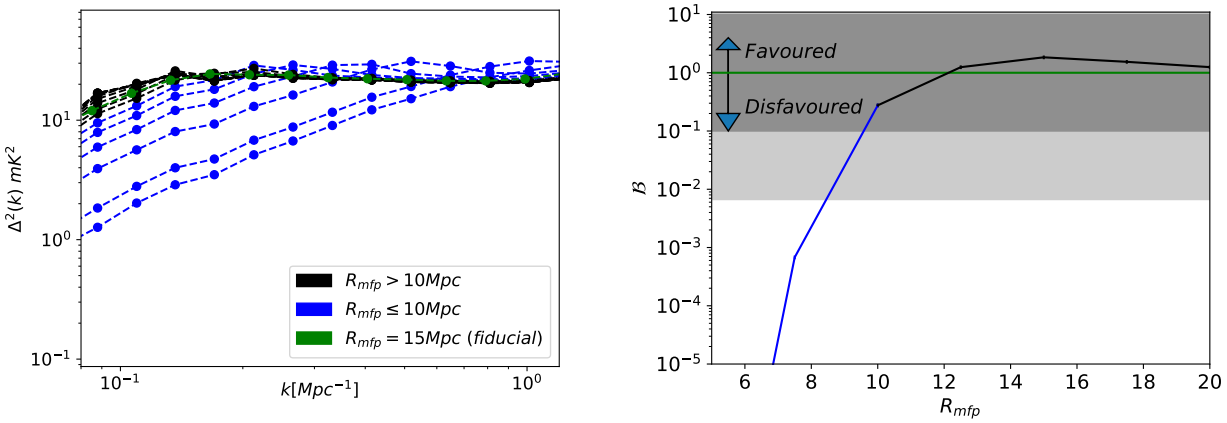
(a) (b)

Figure 3.18: Using the Savage-Dickey density ratio - The impact on the Bayes factor when fixing $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ as a set of models nested in 3pf1. Plots 3.18(a) and 3.18(b) show the ($z = 8$) power spectra and Bayes factors respectively (See Figure 3.17 for a full plot description). As can be seen, $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ varies the power spectral tilt, leading to *strong* inference against values outside $\mathrm{Log}_{10}[T_{\mathrm{vir}}] < 4.1$ and $4.7 < \mathrm{Log}_{10}[T_{\mathrm{vir}}]$.

Figure 3.18(b) shows a smooth peak in the distribution of $\mathcal{B}(\mathrm{Log}_{10}[T_{\mathrm{vir}}])$ that agrees with the fiducial results shown in green. Although the peak is not as sharp as for ζ, it is still a very useful parameter. $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ produces no degenerate power spectra throughout any of its prior distribution (Figure 3.17(a)), but its prior distribution is smaller than for ζ and adjacent to the fiducial parameter there are similarities in the power spectra. When $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ values are small, there are less small scale structures and the power spectrum is suppressed on larger k (blue 3.18(a)); the opposite is true for large $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ (black). In aesthetic terms the parameter describes the tilt of the power spectrum slope. The structures are formed from the same density field, so it makes sense that an increase in structure on a certain scale leads to a decrease elsewhere. Within all of the $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ prior range, there are dynamic changes to 21cm power spectrum.

SDDR with R_{mfp}

The R_{mfp} parameter is not really a physical parameter at all. It is in fact the maximum scale with which the excursion-set formalism can identify ionised bubbles (the maximum radius of the top-hat filter in k-space). Above a model dependent threshold, increasing the maximum



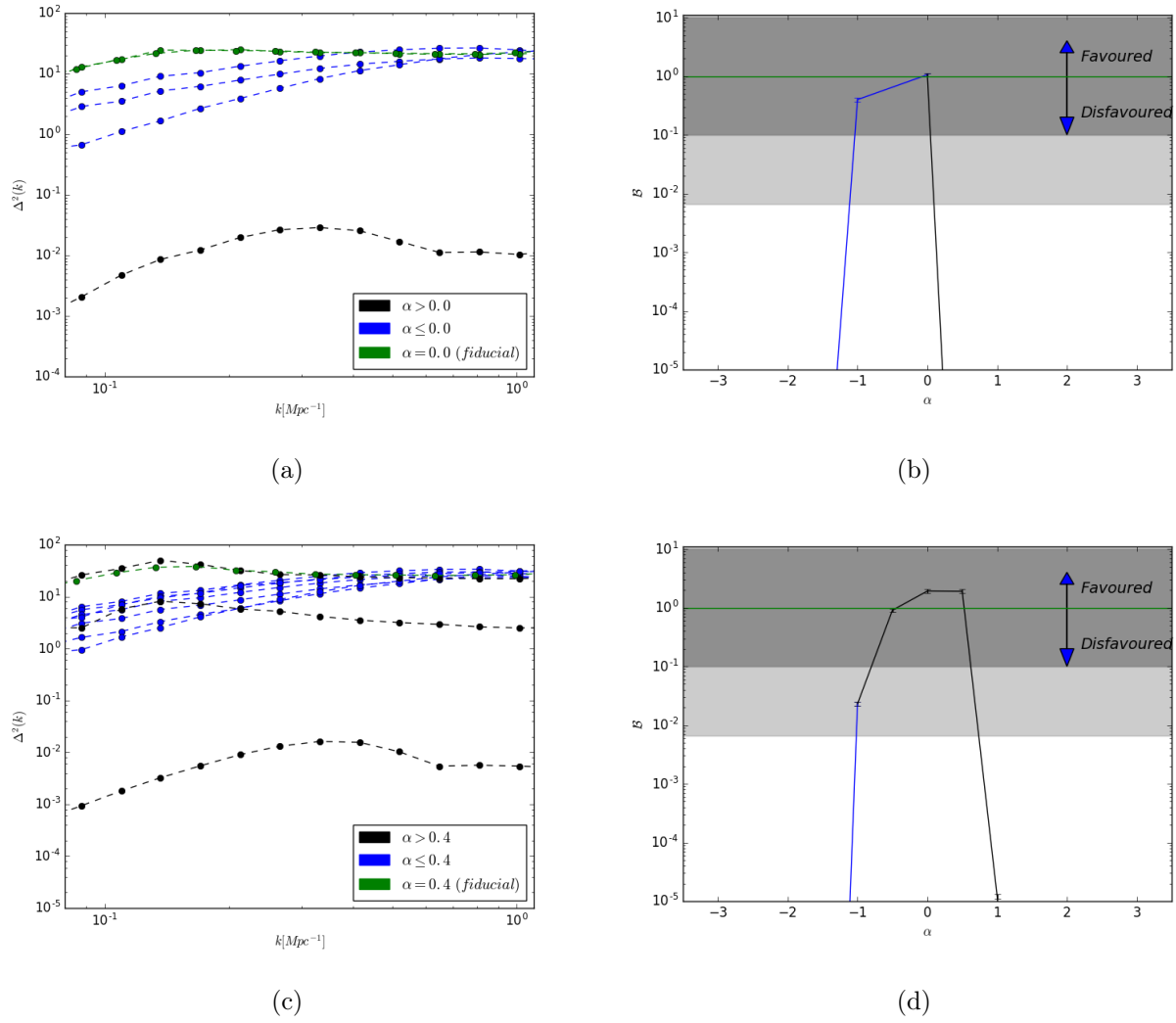
(a) (b)

Figure 3.19: Using the Savage-Dickey density ratio - The impact on the Bayes factor when fixing $R_{\rm mfp}$ as a set of models nested in 3pf1. Plots 3.19(a) and 3.19(b) show the ($z = 8$) power spectra and Bayes factors respectively (See Figure 3.17 for a full plot description). Notice that for $R_{\rm mfp} > 10$Mpc, the Bayes factor provides *weak* inference. Above this parameter value there is no additional inference to be gained.

smoothing scale will not capture any extra structural detail in the simulation. Below this threshold the model is unable to capture the relevant physics, leading to a poor performance with the SDDR against a larger $R_{\rm mfp}$ threshold. It is therefore an obvious target for scrutiny with the SDDR. Figure 3.19(a) clearly shows a loss of large scale (blue, small k) structure from the power spectrum as $R_{\rm mfp}$ decreases below the fiducial value, while the small scale structure remains intact (large k). The nested Bayes factor results for $\mathcal{B}(R_{\rm mfp})$ in Figure 3.19(b) follow suit. All values of $R_{\rm mfp} > 10$ agree with the fiducial results in green by scoring *weak* discrepancies on the Jeffreys' scale. Negligible inference is provided by varying $R_{\rm mfp}$ above this.

SDDR with α

As before, we plot the nested Bayes factors for $\mathcal{B}(\alpha)$ in Figures 3.20(b), 3.20(d) and the variation of the power spectrum with α in Figures 3.20(a) and 3.20(c). The top row uses f1 (with the fiducial values fixed to $\alpha = 0$ and the remaining parameters $[\zeta, \log_{10}[T_{\rm vir}], R_{\rm mfp}] = [20., 4.477, 15.]$) and the bottom row f2 (with the fiducials fixed to $\alpha = 0.4$ and the remaining parameters $[\zeta, \log_{10}[T_{\rm vir}], R_{\rm mfp}] = [15., 4.699, 15.]$). We firstly note α has narrower peak when fitting for f1 than f2. This is mainly because of $\zeta_{\rm eff}$ in Equation 3.8. The power law parameter allows



(a) (b)

(c) (d)

Figure 3.20: Using the Savage-Dickey density ratio - The impact on the Bayes factor when fixing α as a set of models nested in 4pFZH fitting for f1 (top row) and f2 (bottom row). Plots 3.20(a)/3.20(c) and 3.20(b)/3.20(d) show the ($z = 8$) power spectra and Bayes factors respectively (See Figure 3.17 for a full plot description). Without the use of the power-law in (f1, $\alpha = 0$), there are less allowed values for $\zeta_{\text{eff}}(\zeta, \alpha)$, hence a narrower peak in 3.20(b) compared to 3.20(d). The ability to pin down α depends on the fiducial parameter choice in the mock observation.

a widening in the applicable ζ values. To be specific, In f1 the degeneracy here arises from the behaviour of $\zeta_{\rm eff}(\zeta,\ \alpha = 0)$. In f2 the degenerate parameter space has increased since the ionising efficiency now also depends on halo mass, $\zeta_{\rm eff}(\zeta,\ \alpha,\ \log_{10}[T_{\rm vir}])$. Since the relationship between α and the 21cm power spectrum is clearly different depending on the values of ζ and $\log_{10}[T_{\rm vir}]$. We cannot be certain on the effective range of α because of this. Given the two sets of fiducial values, it is clear in both $\mathcal{B}(\alpha)$ figures that there is a fast drop off outside a total of $\alpha \in [-1.5, 1]$. Hence we have been lenient with the limitations of the prior (currently $\alpha \in [-3, 3]$) which result in $\zeta_{\rm eff} \in [2.15, 582]$ given the influence of the other parameter ranges. It is certainly possible to observe changes in the 21cm power spectrum from varying α but we emphasise that unless we include more than just the current data used in the statistic, we will not be able to distinguish the power-law parameters for the ionisation efficiency's dependency on halo mass.

Within the prior when α values are lower, the blue power spectra lose large scale structure until quickly all structure is lost, this is akin to reducing the input ζ. When increasing α, the behaviour of the black power spectra mimic an increase in ζ. However in both directions, an α of large magnitude causes the power spectra to quickly fall off the scales of the plot entirely as reionisation is either ended prematurely or too dim to progress.

To reiterate, since the degeneracy of our parameter's fit to the power spectrum depends on the choice of the fiducial parameters in the mock observation, we cannot decisively evaluate the influence of this parameter with the 21cm power spectrum. This conclusion follows into the next section where we are unable to distinguish an ionising efficiency that is a constant from one that is a power-law in halo-mass.

3.4.8 Can we distinguish a constant ζ from a power law in halo mass?

We have shown in previous sections that HERA & SKA can easily determine the most suitable reionisation scenario. We now explore refining the number of parameters, testing the ability to determine the use of the 3 or 4 parameter FZH model detailed in Section 3.2.1.

Figure 3.21 shows the Bayes factors for 3p and 4p FZH models tested with the priors used in Table 3.4. The black triangles show that the different parameterisations of 21cmFast that we consider scored *weakly* on the Jeffreys' scale when observed with HERA-331. In other words,

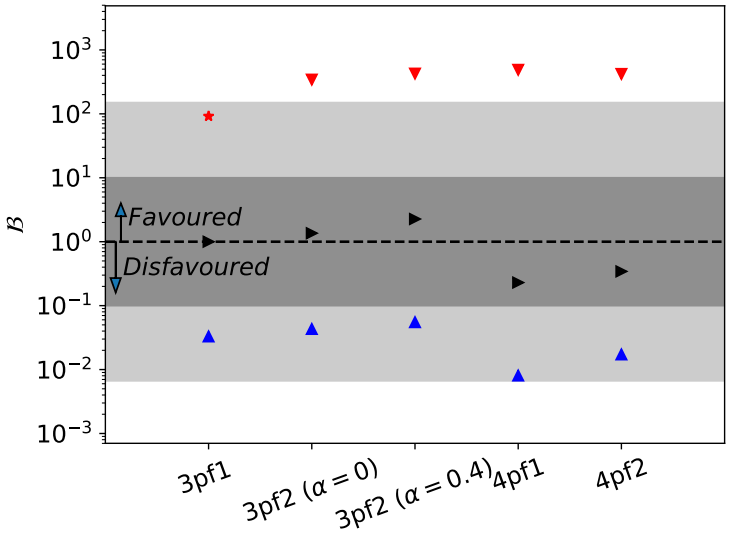


Figure 3.21: The Bayes factors for all of the 21CMFAST, FZH based models (in ratio against 3pFZHf1 with $\alpha = 0.$). *Weak*, *moderate* and *strong* inference scores on the Jeffreys' scale are indicated by the dark, light greys and white region respectively. The black points represent the prior volumes chosen (i.e. using sensible motivations, Section 3.3) - all models here score *weak* results on the Jeffreys' scale. Firstly this is evidence towards being unable to distinguish a power law in halo mass for the ionisation efficiency with the 21cm PS. The secondary aim of this plot shows the effect of using a prior to skew the Evidence values and therefore the Bayes factor results. Red indicates the use of a δ-function across either the fiducial values (starred point - 3pFZHf1 only) or the MAP values (red arrows). Blue points have widened priors to skew against the models. Note that it is possible to achieve for this skewing of Bayes factor to achieve *strong* evidence scores with narrow enough priors. We emphasise that the Jeffreys' scale is a guide only, see Section 3.4.8 for the discussion.

the power spectrum from this 1080 hour observation is not sensitive to the chosen changes in the fiducial parameters - motivated by the faint and bright galaxy models in Greig & Mesinger (2015). The change of fitting with a constant ζ to a power law in halo mass (Equations 1.15 and 3.7) is also negligible. Both SKA (not shown) and HERA obtain similar results that score *weak* Bayes factors when recovering the f1 and f2 power spectra with the 3p and 4p FZH models. Applying larger values of α makes these models more distinguishable, however this produces unphysical results. Evidences produced with larger α values are therefore dominated by contributions from the observational priors. Increasing α also increases the ionising efficiency (Equation 3.8) and therefore the EoR finishes too early if α is largely positive. This can be seen by the sharp cut on the right hand side of the α posterior in Figure 3.5(e), the bump in the power spectra as α changes in Figure 3.20, and is particularly challenged by the McGreer prior in Figure 3.16(b). In summary, the two versions of the basic 21cmFast model cannot be distinguished on the Jeffreys' scale. We cannot tell from the 21cm power spectrum which of 3pFZH or 4pFZH has produced the mock observation when observing for 1080 hours with HERA-331 or SKA-512.

We then use this context to inspect the effect of skewing these results with a naive choice of prior. Figure 3.21 also contains the information obtained when attempting to skew our own results with prior choice (discussed in Section 3.3) - δ-function and widened uniform priors are the red and blue points respectively. Expanding (reducing) the prior width will decrease (increase) the posterior density and therefore the Evidence. As we have established in Section 1.4, attempting to skew results is the only way one can be sure the results are not prior dependent. The blue results use prior ranges that include unphysical parameter values. These produce Evidences that still fall short of the *strong* Bayes factor score meaning we have been unable to skew the result - we consider this a success of the machinery. In other words if one naively chose too wide a uniform prior distribution (a sensible approach for a first attempt), the skewed result would be *moderate* at worse. This user, in their own prior analysis, should look closer at the physics at hand and try to reduce the wide priors to those resembling Table 3.4 - which were chosen via physical reasoning and familiarisation with the problem (black points). It is worth noting that each of the black points translates down to the blue points in a uniform manor as the parameter prior volume is factored out of the Evidence integral as a constant. For the red shapes (using δ function priors), the maximum possible Evidence is obtained by choosing the MAP parameters, $\mathcal{Z} = \mathcal{L}(\theta_{\rm MAP})$. This is observed by the alignment of the red

triangles in Figure 3.21. The starred red point (only above 3pf1) is calculated with the fiducial parameters (f1), this point deviates down from this line of maxima because no consideration of the observational priors were made when choosing the fiducial values (but this is not a concern here[5]). It is worth noting that a clear give-away of a prior being chosen to be too narrow is the posterior plots will show distributions that either are close to uniform distributions or have no clear peak. In these cases either the parameters dynamic range is not expressed by the chosen prior (the case here), or the parameter is redundant (the case with $R_{\rm mfp}$ in Section 3.4.7).

The choice of α and ζ_0 in the fiducial values impacts whether the power-law in halo mass for ζ can be distinguished from a constant ζ. Since this chapter is aimed at distinguishing morphology and scale we refrain from deviating too far off course and stop here. More physical models already exist and we will proceed with their model selection in the next chapter.

3.5 Conclusions

First we discuss comparisons of 21CMMC and 21CMNEST, in particular showing that MULTINEST can produce matching posterior distributions to those produced by COSMOHAMMER (Section 3.4.1). The 'moderate' foreground wedge model within 21CMSENSE is used to calculate the telescope noise in all cases.

When using Bayesian model selection, the literature discussed in Section 1.4 warns of two potential heffalump traps: these are a skew of results from dependency of the prior and an overcompensated penalty against redundant parameters via Occam's Razor. We carefully address these issues in Section 3.3 to show the validity of the Bayesian model selection in this chapter. For the non-statistician and to check sensible thresholds for the results, we adopt the Jeffreys' scale to distinguish *weak*, *moderate* and *strong* conclusions allowing a qualitative interpretation of our results.

This chapter considers toy EoR models as well as the models in the original 21CMMC and Watkinson & Pritchard (2014). Namely these are: an inverted version of the original 3 parameter FZH (*global inside-out*) model used in 21CMFAST (within which the excursion set formalism

[5]Mild disagreements between the fiducial power spectrum parameters and those satisfying the observational priors has been addressed when cross checking the methodology against the Inv MHR fiducial data set in Section 3.4.2 and in Section 3.4.6 when analysing the contributions to the likelihood separately.

solves for reionised bubbles); a simpler model MHR (*local outside-in*), in which ionisation is defined via a density criteria; then both of these models have their ionisation criterion mathematical inverted (Inv FZH, *global outside-in*; and Inv MHR, *local inside-out*); and finally we add a 3rd parameter to the MHR and Inv MHR models in the form of a Gaussian density field filter creating - F MHR, *global outside-in*; and F Inv MHR (*global inside-out*). We show that using the specifications of HERA-331 (or SKA-512, see Table 3.3 for the telescope details) 1080 hours of observation is easily enough to distinguish these toy models with *strong* evidence on the Jeffreys' scale (Section 3.4.4).

We calculate Evidences for the separate contributions to the likelihood in Equation 3.1 to quantitatively test evaluations of the neutral fraction (surrounding high z quasars) and the reionisation optical depth (Section 3.4.6). This shows that more reionisation data is required if observational checks are to provide comparable inference to the 21cm power spectrum. In the next chapter, we repeat this quantitative analysis of the inference capabilities of the UV luminosity function data as prescribed in Park et al. (2019).

On a subtler level the considered version of 21CMMC contains the addition of a 4th parameter which enables the ionising efficiency to have dependency on the halo mass. In other words this relaxes the assumption of a constant mass to light ratio for the simulated ensemble of galaxies. We show in Sections 3.4.7 and 3.4.8 that these cannot be unquestionably distinguished with 1080 hour observations from HERA-331. For the two mock observation choices (f1 and f2) that represent faint and bright populations of galaxies we find only *weak* scores on the Jeffreys' scale. This result becomes easier to conclude if the fiducial parameters used in the mock observation are chosen the more unphysical. We therefore cut our losses and progress to more physical EoR models in the next chapter.

Using the Savage-Dickey density ratio we proceeded to show the redundancy of $R_{\rm mfp} > 10h^{-1}$ Mpc in 21CMMC (Section 3.4.7) - hence justifying its omission from the newest parameterisations of 21CMMC and 21CMFAST. The inhomogeneous recombinations discussed in Section 1.3.5 (Sobacchi & Mesinger, 2014) are used instead in Chapters 4 and 5.

The most difficult to distinguish of the toy models is F Inv MHR because it shares scale and morphology with FZH. It is therefore used as a bench mark for answering our primary questions about feasibility:

- LOFAR-48 would struggle to perform model selection in the EoR. We require 21600 hours of observation to provide *strong* disfavouring of the toy models (with odds $\sim 400:1$).
- Using 1080 hours of observation, HERA would require at least 61 dipoles in its configuration to rule out the toy models with *strong* evidence on the Jeffreys' scale (scoring odds of $\sim 600:1$).
- Finally, with the SKA-512: the toy EoR models can be distinguished with *strong* evidence quickly, with only 324 observing hours needed to obtain odds of $\sim 500:1$.

With 217 dipoles HERA becomes comparable to the SKA at observing the 21cm EoR power spectrum. As can be seen in Figure 3.4, SKA has more observational clarity on small scales while HERA dominates the sensitivity on larger scales. Assuming the *moderate* noise settings within 21cmSense can be achieved, the redundant baseline instrumental method shows a lot of promise in performing model selection with the 21cm EoR power spectrum (Byrne et al., 2019; Dillon et al., 2019).

We have now set the scene for ruling out toy models of the EoR with Bayesian Model Selection. This work shows that modest 21cm experiments, such as HERA-61, are likely able to use their observations to pin down the correct morphology and scale of reionisation. In the next chapter we look more closely at the level of precision that can be obtained from more involved models of the EoR using the newer versions of 21cmMC. Once the desired reionisation scenario can be quantitatively chosen, parameter inference should be performed to infer the involved astrophysics.

Chapter 4

Bayesian Model Selection for Astrophysical Scenarios in the EoR

Simulations of the 21cm light-cone that include spin temperature fluctuations require larger boxes than those used in this chapter. The coeval cubes we stitch into the light-cones here are 125Mpc in each dimension - this is a mistake. Evidence calculations and MAP parameters presented in the results of this chapter are currently being redone with 250Mpc boxes. Any power spectra and light-cones presented in this chapter have been re-calculated with the 250Mpc boxes however the MAP parameters, Evidence values and Bayes factors are likely to change a little. Although a box length of 125Mpc is a suitable minimum box size for the observational bandwidth used here, it will not allow X-ray heating to be appropriately captured. The subsequent publication will not be released until the corrected simulations have finished.

4.1 Introduction

In the previous chapter we addressed the application of Bayesian model selection to toy scenarios of the EoR (Binnie & Pritchard, 2019). We addressed the two possibilities of morphology and scale within which reionisation can progress. In reality, nature is likely to use a combination of both and so in this chapter we turn to performing model selection between more realistic models of the EoR.

The EoR is driven by a variety of objects including quasars (Garaldi et al., 2019) and the

remnants of the first stars, which are likely to create to a diffuse background of X-rays ($\sim 0.1 - 2$ keV). These X-rays will heat the IGM causing a unique signature that influences the 21cm spin temperature and can therefore be observed.

The star forming galaxies that drive reionisation are commonly characterised by the UV luminosity function (LF). These have been targeted in high-redshift galaxy surveys, particularly by observations from the Hubble Space Telescope (HST[1]) (Bouwens et al., 2015; Oesch et al., 2018). The UV LF are determined by a galaxy's baryonic mass in stars, star formation history, and the surrounding dust in the inter-stellar medium (ISM) - they contain a lot of information about galaxy evolution. Only the brightest end has been observed in the EoR, but this enables constraints on the UV LF up to $z \leq 10$ (Rojas-Ruiz et al., 2020). We apply these in combination with more complete constraints of the faint end (e.g. Yue et al. (2018) from lensing) that reach to $z \leq 6$. With the onset of next generation galaxy surveys like JWST (James Webb Space Telescope[2]) (Gardner et al., 2006), and next generation 21cm experiments like HERA (Hydrogen Epoch of Reionization Array[3]) (DeBoer et al., 2017) and the SKA (Square Kilometer Array[4]) (Mellema et al., 2013; Koopmans et al., 2015), plentiful observations of the EoR seem likely in the near future.

We extend from the previous chapter by building on the 3pFZH model. We include parameterisations for X-ray heating and UV LFs in the models that are tested. Bayesian inference has proven to be a powerful tool for analysing the EoR so far and we look to test it on more detailed physical scenarios. Parameter estimation can already apply constraints to these models using data from MWA and LOFAR observations (Greig et al., 2020a,b), we are working towards achieving similar things with model selection. As before, we build upon 21CMMC, the parameter estimation for semi-numerical EoR simulations (Greig & Mesinger, 2015) - a state of the art code combining the affine-invariant EMCEE sampler (Foreman-Mackey et al., 2013) with 21CMFAST (Mesinger et al., 2011), the semi-numerical 21cm signal simulation within the COSMOHAMMER framework (Akeret et al., 2013). COSMOHAMMER is then replaced with the nested sampling algorithm MULTINEST (Feroz et al., 2009) to produce 21CMNEST[5] which applies Bayesian model selection to distinguish models of the EoR. With this technique we aim

[1] http://hubblesite.org
[2] https://www.jwst.nasa.gov
[3] https://reionization.org
[4] https://www.skatelescope.org
[5] An updated version of 21cmNest containing the models from this chapter will soon be available at https://binnietom@bitbucket.org/binnietom/21cmnest_2.0.git.

to objectively distinguish models using observations.

As we increase the variety of physics that is included in the selection of EoR models, we aim to give a realistic idea as to which of the phenomena can be distinguished by applying Bayesian model selection to upcoming observations. We proceed by investigating four models in 21CMMC:

- A - Reionisation with a constant ionising efficiency, ζ, a simple scenario (sharing properties with the 3pFZH model in the previous chapter).
- B - Including X-ray heating from stellar remnants (and therefore spin temperature, $T_{\rm S}$ fluctuations).
- C - Including a single power-law in halo mass for ζ and modelling of UV LFs (sharing properties in the 4pFZH model from the previous chapter).
- D - Including X-ray heating, $T_{\rm S}$ fluctuations, UV LFs, and a double power-law in halo mass for ζ - a model combining all of the new physics introduced in this chapter.

We use these four models in an attempt to answer three questions:

- Can SKA observations of the 21cm PS measure the fluctuations of the 21cm spin temperature from X-ray heating?
- Does the high redshift UV LF provide comparable inference to the EoR 21cm PS in a Bayesian context?
- Can the 21cm PS and UV LF distinguish models with a double power-law in halo mass for the ionising efficiency?

While these questions have been investigated previously, there are two main distinctions. Firstly, we answer them with Bayesian model selection rather than with parameter estimation, as in Greig & Mesinger (2017b) and Park et al. (2019). The other distinction is that our inference is specific to the backbone of the EoR only ($8 < z < 10$), since this is where the bulk of reionization occurs. It is also where observational endeavours currently aim to detect the 21cm PS. But wherever this focus is aimed, it is likely that the 21cm PS will be observed

in parts. We therefore deem it a useful and interesting proof of concept to restrict the full cosmological 21cm signal, just in-case an instrumental complication gets in the way at certain bandwidths. Finally, since saturation will likely occur at some point accross this specific period, distinguishing the properties of X-ray heating via the spin temperature will be a challenge for the methodology.

The chapter is structured as follows: Section 4.2 contains a description of 21CMNEST, and the recipes from 21CMMC used for the 21cm signal; Section 4.3 details each of the four EoR models as well as the observational error for the 21cm power spectrum; in Section 4.4 we discuss the physical motivations behind choosing the parameter prior distributions; Section 4.5 presents the results; Section 4.6 contains a full discussion of proposed questions before we summarise in Section 4.7.

4.2 21CMNEST

This chapter follows on from the previous one, here we will only highlight the differences. Please see the detail surrounding 21CMMC and 21CMFAST in Chapters 1 and 3 for the full recipes used for the 21cm PS. All the models in this chapter are based on FZH (Furlanetto et al., 2004), the most physical model from the previous chapter.

The first difference to Chapter 3 is the use of inhomogeneous recombinations (Sobacchi & Mesinger (2014), detailed in Section 1.3.5). This is incorporated in the ionisation threshold as,

$$n_{\rm ion}(x, z|R, \delta R) \geq (1 + \bar{n}_{\rm rec})(1 - \bar{x}_e), \tag{4.1}$$

where the maximum excursion set scale (maximum bubble size), R, is set to $R \leq 50$ Mpc; $n_{\rm ion}(\mathbf{x}, z|R, \delta R)$, is the number density of ionising photons at a position, $\mathbf{x}$, and redshift, z; $\bar{n}_{\rm rec}$, is the spatially averaged number density of inhomogeneous recombinations; $\bar{x}_e$ is the spatial average of the free electron fraction in the IGM - assumed to be the ionised fraction.

The other difference is calculating the 21cm PS from chunks of light-cone rather than from coeval cubes of brightness temperature (Greig & Mesinger (2018), detailed in Section 1.3.4). Equation 3.2 remains the same but with V representing the volume of the chunk rather than the cube. To produce the light-cone we stitch coeval cubes together with logarithmic steps of

Model	**A**	**B**	**C**	**D**
Parameters	$\mathrm{Log}_{10}[T_{\mathrm{vir}}],\ \zeta$	$\mathrm{Log}_{10}[T_{\mathrm{vir}}],\ \zeta,$ $E_0,\ \alpha_{\mathrm{X}},\ L_{\mathrm{X<2keV}}$	$\mathrm{Log}_{10}[M_{\mathrm{turn}}],\ f_{*,10},$ $\alpha_*,\ t_*$	$f_*,\ f_{\mathrm{esc},10},\ \alpha_*, \alpha_{\mathrm{esc}}$ $\mathrm{Log}_{10}[M_{\mathrm{turn}}],\ E_0,\ L_{\mathrm{X<2keV}},\ t_*$
$\delta T_b(T_{\mathrm{S}})$	No	Yes	No	Yes
$\mathcal{L}(\chi^2_{\mathrm{LF}})$	No	No	Yes	Yes
Fiducial values	$[4.7,\ 30]$	$[4.7,\ 30,$ $500,\ 1,\ 40.0]$	$[8.69897,\ -1.301,$ $0.5,\ 0.5]$	$[-1.301,\ -1,\ 0.5,\ -0.5,$ $8.69897,\ 500,\ 40.5,\ 0.5]$

Table 4.1: A summary of the four models analysed in this chapter. B and D relax the post-heating approximation; C and D produce UV LFs. See Section 4.3 for details of the parameters in the top row. The fiducial parameters used to simulate each model's mock data are listed in the bottom row. The fiducial parameter choices for C and D are motivated by consistency with the UV LF data. B and D have more freedom, their fiducial parameter choices are consistent with the observational priors (as well as the UV LF data for D).

$\Delta z_{\mathrm{step}} = 1.04$ between redshifts 6 and 35. We then calculate the 21cm PS from three chunks evenly spaced between redshifts 8-10.

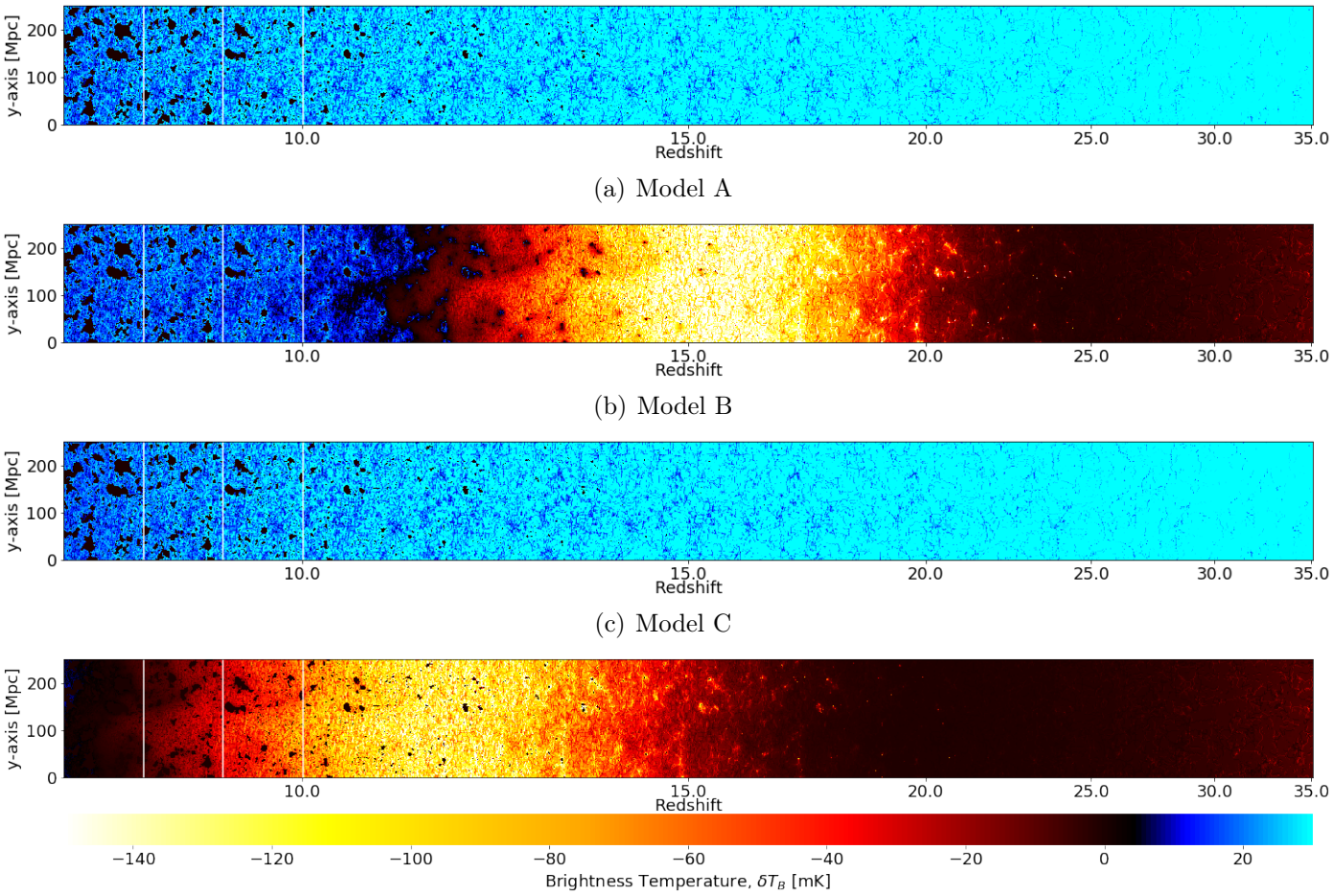
Binnie & Pritchard (2019) showed that the improvement in inference gained from the observational prior checks provide negligible inference in comparison to the mock 21cm PS observation (see Chapter 3 for detail). With the introduction of more realistic model prescriptions we can apply UV LF constraints at given redshifts via the prescription detailed in Section 4.3.3 (Park et al., 2019). The likelihood from the last chapter is used with the addition of a UV luminosity function term,

$$\ln\mathcal{L} = \chi^2_{21\mathrm{cm}} + \chi^2_{\mathrm{Planck}} + \chi^2_{\mathrm{McGreer}} + \chi^2_{\mathrm{Greig}} + \chi^2_{\mathrm{LF}}, \tag{4.2}$$

where, χ^2_{LF} measures the UV LF directly against data from Bouwens et al. (2015) for $z = 6, 7, 8$ and Oesch et al. (2018) for $z = 10$.

4.3 EoR models

We now detail the 4 models used in this chapter. For $\chi^2_{21\mathrm{cm}}$ in Equation 4.2, each model produced a mock observation by simulating the 21cm power spectrum data from a cube of side 500Mpc using the fiducial parameters in Table 4.1.



(a) Model A

(b) Model B

(c) Model C

(d) Model D

Figure 4.1: Slices through each model's mock observed 21cm brightness temperature signal light-cone. The fiducial parameters are given in Table 4.1. At $z = 8$, the left-hand end of each light-cone respectively, $\bar{x}_{\rm HI} = 0.61,\ 0.61,\ 0.55$, and 0.55 at $z = 8$. The power spectra in Figures 4.2 and 4.3 are calculated from these light-cones. The white lines indicate the chunks (~ 170 Mpc h^{-1} in size).

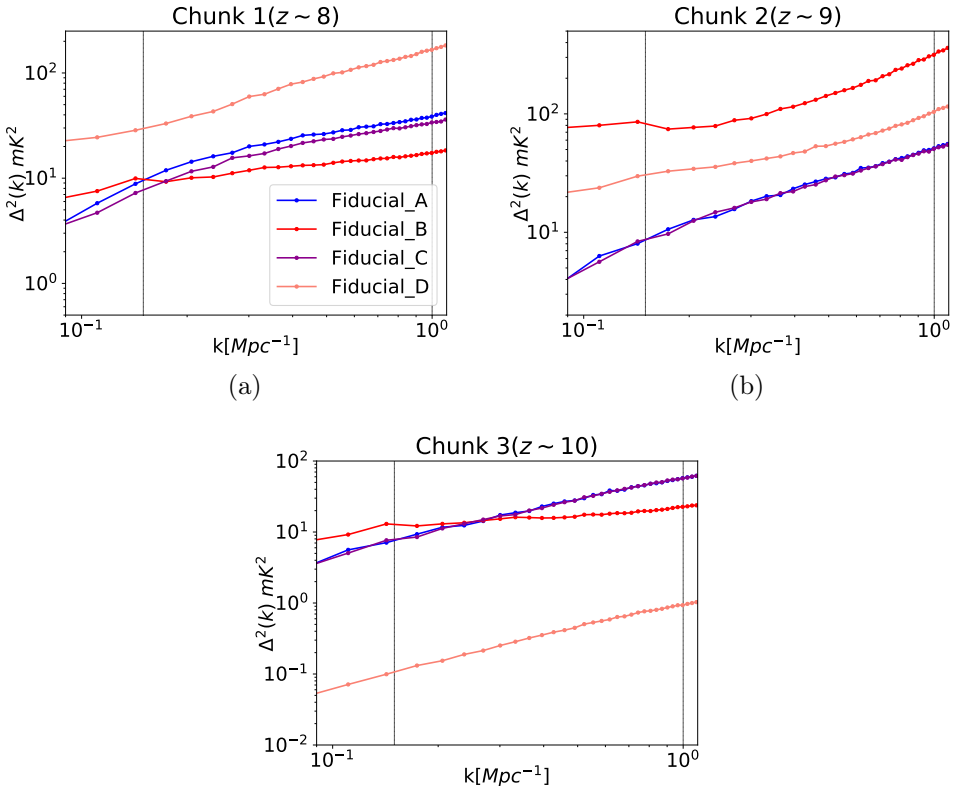


Figure 4.2: Mock observed 21cm PS data for the models in this chapter. They are simulated with the fiducial parameters given in Table 4.1. The observation is performed in between a foreground corruption limit (See Section 4.3.6) and shot noise cuts ($k = 0.1$ and 1.0 Mpc^{-1} respectively). The left and right grey vertical lines represent these, within which the 21cm PS χ^2 is calculated.

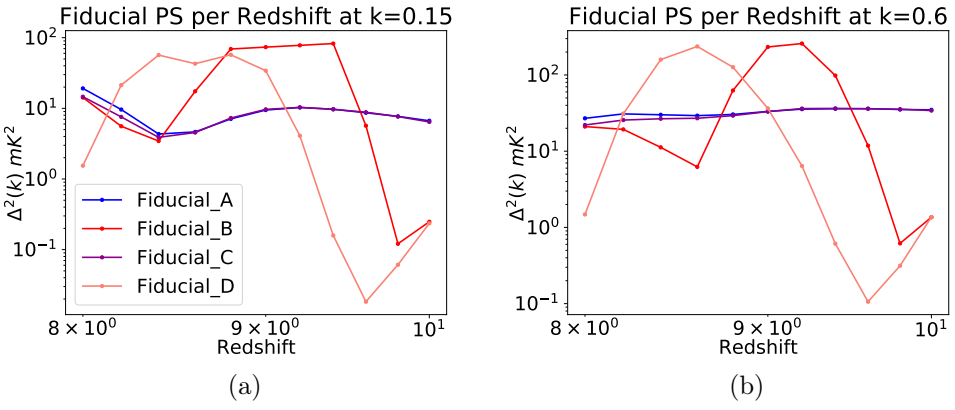


Figure 4.3: Fiducial mock 21cm PS data but plotted at a given k-scale [Mpc^{-1}] per redshift. As above, these are simulated with the fiducial values given in Table 4.1. The discretisation caused by chunking the light-cone is increased from 3 chunks to 10 (for $z \in [8, 10]$) to visibly capture the shape of the 21cm PS. Our χ^2 uses three chunks (as in Figure 4.2) from the light-cones in Figure 4.1.

4.3.1 Model A - simple scenario

Model A is closely related to the 3pFZH model in the previous chapter (Furlanetto et al., 2004; Mesinger et al., 2011). Ionised regions are identified using the excursion set formalism.

This basic model uses the two main parameters from before: ζ an ionisation efficiency of UV radiation, and $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ the minimum virial temperature for halos to host galaxy formation. Both parameters are implemented in,

$$n_{\mathrm{ion}} = \left[\frac{\Omega_b}{\rho_b \Omega_{\mathrm{M}}} \int_{M_{\mathrm{vir}}}^{\infty} m \frac{dn}{dm} dm\right] \zeta, \tag{4.3}$$

where $M_{\mathrm{vir}} \propto T_{\mathrm{vir}}^{2/3}$ (Barkana & Loeb, 2001). f_{coll}, the fraction of dark matter collapsed into halos from before is contained in the square brackets (Greig & Mesinger, 2015).

Inhomogeneous recombinations are accounted for by the ionisation threshold, instead of including it as a constant in the prescription for ζ (Equation 1.13). Therefore $\zeta = f_* f_{\mathrm{esc}} N_{\gamma/b}$, no longer includes a spatially-averaged factor to represent the contribution of IGM recombinations into the model. We have moved away from using a photon mean free path and instead a maximum bubble size is fixed to $R_{\mathrm{mfp}} = 50\mathrm{Mpc}$. This is the maximum scale with which the excursion set formalism can identify HII regions. Therefore the neutral fraction, $x_{\mathrm{HI}}(\delta, z)$, is expressed as a function of density and redshift.

4.3.2 Model B - the epoch of heating

Three new parameters are introduced in Model B to detail X-ray production early in the cosmic dawn. These X-rays cause inhomogeneous heating signatures in the IGM after they have been redshifted enough to interact with HI gas. This causes an observable change in the 21cm brightness temperature via fluctuations in the spin temperature. Hence Model B relaxes the post-heating approximation which was used in Model A.

We continue the use of parameters $\mathrm{Log}_{10}[T_{\mathrm{vir}}]$ and ζ for calculating n_{ion} (as with Model A). X-ray emission is given by,

$$\epsilon_{\mathrm{X}}(\mathbf{x}, E_{\mathrm{e}}, z') = \frac{L_{\mathrm{X}}}{\mathrm{SFR}} \left[(1 + \bar{\delta}) \rho_b f_* \frac{df_{\mathrm{coll}}}{dt}\right]. \tag{4.4}$$

E_0, is the minimum escape energy for an X-ray photon to penetrate into the IGM; and $L_{\rm X<2keV}/{\rm SFR}$, normalises the soft-band X-ray emission below 2keV. These are the first and second new X-ray parameters introduced from Greig & Mesinger (2017b). There is no free parameter that controls time-scale, hence we use $f_{\rm coll}$ (see model C, Section 4.3.3). The third free parameter in this model is implemented within the X-ray luminosity as a power-law index, α_X, where,

$$L_{\rm X} \propto E^{-\alpha_{\rm X}}. \tag{4.5}$$

This radiation is likely dominated by the properties of a single X-ray source prior to the EoR. Here the mechanism of production is less important than the X-ray flux itself. HMXB are expected to emit X-rays with power law $\alpha_{\rm X} = 1$. Other sources, such as mini-quasars, may have a steeper emission profile. The fluctuations in the 21cm PS from its spin temperature provides insight into the first X-ray sources. This parameter therefore provides insight into what species drives the inhomogeneous gas heating that precedes reionisation (Pritchard & Furlanetto, 2007; Greig & Mesinger, 2017b). Beyond 2keV the X-ray photon mean free paths are longer than the universe (Pritchard & Furlanetto, 2007). The normalised X-ray luminosity is,

$$L_{\rm X<2keV}/{\rm SFR} = \int_{E_0}^{2{\rm keV}} \frac{L_{\rm X}}{\rm SFR}\, dE_e, \tag{4.6}$$

where E_0 is dictated by the inter-stellar medium (ISM) column hydrogen density. Since $L_{\rm X}$ is X-ray luminosity is given per star formation rate (SFR). The square brackets in Equation 4.4 denote the SFR density along the light-cone, equivalent to expressing the X-ray emissivity per volume.

The spin temperature is calculated as Equation A.35. To elaborate, T_α corresponds to the colour temperature of the Ly-α background; x_α corresponds to the Wouthuysen-Field coupling strength; and $x_{\rm c}$ is the collisional coupling coefficient. The kinetic gas temperature, $T_{\rm K}$, is solved in tandem with the redshift dependency of free electrons (Mesinger & Furlanetto (2007); Mesinger et al. (2011), see Section 1.3.3 for details). $T_{\rm K}$ is evolved in each pixel as,

$$\frac{dT_{\rm K}(\mathbf{x}, z)}{dz} = \frac{2Q}{3k_B(1+x_e)}\frac{dt}{dz} + \frac{2T_{\rm K}}{3n_b}\frac{dn_b}{dz} - \frac{T_{\rm K}}{(1+x_e)}\frac{dx_e}{dz} \tag{4.7}$$

where the heating rate per baryon, Q, contains the sum of contributions from X-ray and Compton heating (see e.g. Baek & Ferrara (2013); Madau & Fragos (2017); Eide et al. (2018)

for the full calculation). The Ly-α background is produced by the cosmic dawn (between $z \sim 20$ and 35), filling the Lyman continuum. These are the photons that are absorbed by lower Lyman lines after being redshifted by the Hubble flow. Background radiation densities are constructed by HI X-ray excitation (from J, Equation 1.16) and stellar emission of photons in the Lyman bands. Composite stellar spectra are scaled with the SFR estimates from modelling (Barkana & Loeb, 2005).

4.3.3 Model C - UV luminosity functions

Model C implements a saturated spin temperature (similar to model A) as well as a halo mass dependency via a power law in ionising efficiency (specifically in the fraction of mass in stars), it therefore shares some similarity with 4pFZH in Chapter 3. A fourth parameter is also added to characterise a star formation timescale, enabling the simplest calculation of the UV LF that agrees with observed high redshift galaxies. The observed UV LF data we use combines 10,000 galaxy observations from the HST Legacy fields (Bouwens et al., 2015; Oesch et al., 2018). This enables a constraint on stellar mass at each of the redshifts that we have galactic UV LF data.

We now have a star formation timescale, hence model C introduces a more detailed star formation prescription. Rather than using ζ as before, we start with the halo mass function. We then calculate both $n_{\rm ion}$ and the galaxy SFR allowing both the 21cm and UV LF to be used as constraints simultaneously. Different mass halos have their star formation described by four free parameters: α_*, the power-law scaling of f_*; $f_{*,10}$, the normalisation of f_* for $10^{10} M_\odot$ halos; $M_{\rm turn}$, the turn-over mass; and t_*, the time-scale for star formation. This is the model used in Park et al. (2019) but with $f_{\rm esc}$ set constant.

The number density of ionising photons is calculated as,

$$n_{\rm ion} = \frac{1}{\rho_b} \int_0^\infty \frac{dn(M_{\rm h}, z|R, \delta R)}{dM_{\rm h}} dM_{\rm h} f_{\rm duty} M_* f_{\rm esc} N_{\gamma/b}, \tag{4.8}$$

where M_h is the mass of dark matter halos; the duty cycle, $f_{\rm duty}$, represents the suppression of ionising radiation from low mass halos due to their lower star formation rate; $f_{\rm esc}$ is the fraction of UV photons that escape from the galaxy into the IGM; and $N_{\gamma/b}$ is the number of ionising photons per baryon, set to 5000 for the case of the Salpeter initial mass function for stars (Salpeter, 1955; Barkana & Loeb, 2005). The stellar mass within a halo, M_*, is defined

as,

$$M_* = M_h f_* \frac{\Omega_\mathrm{b}}{\Omega_\mathrm{M}}. \tag{4.9}$$

Equation 4.9 is dependent on the fraction of baryonic mass in halos f_* as well as the halo mass.

We move away from ζ as a constant parameter to incorporate a power-law in halo mass for f_*,

$$\zeta \sim f_{*,10} \left(\frac{M_\mathrm{h}}{10^{10} M_\odot} \right)^{\alpha_*}, \tag{4.10}$$

where $N_{\gamma/b}$ and f_esc remain constant. We introduce t_* as a dimensionless time scale (as a fraction of the Hubble time) on which mass is converted into stars. Since we now have a timescale, n_ion is calculated fully (Equation 4.8), including a suppressed contribution from low mass halos via the duty cycle,

$$f_\mathrm{duty}(M_\mathrm{h}) = e^{-\frac{M_\mathrm{turn}}{M_\mathrm{h}}}. \tag{4.11}$$

Calculation of the SFR, $\dot{M}_*(M_h, z)$, is also dependent on t_* as,

$$\dot{M}_*(M_h, z) = \frac{M_* H(z)}{t_*}. \tag{4.12}$$

To calculate the UV LF (ϕ) we relate the UV luminosity to $\dot{M}_*(M_h, z)$ as,

$$L_\mathrm{UV} = \kappa \dot{M}_*(M_h, z), \tag{4.13}$$

where the constant, $\kappa = 8.70 \times 10^{27} M_\odot^{-1}$yr erg s^{-1} Hz^{-1}, is calculated for a Salpeter initial mass function while ignoring dust extinction (Sun & Furlanetto, 2016). This agrees with the data obtained during WFC3/IR field observations which are selected from Hubble to specifically represent star forming galaxies at redshifts 7 & 8.

The LF is given by,

$$\phi(M_\mathrm{UV}) = f_\mathrm{duty} \frac{dn}{dM_\mathrm{h}} \left| \frac{dM_\mathrm{h}}{dL_\mathrm{UV}} \frac{dL_\mathrm{UV}}{dM_\mathrm{UV}} \right|, \tag{4.14}$$

where L_UV is related to the UV magnitude (M_UV) using the AB magnitude relation (Oke & Gunn, 1983).

When calculating $\phi(M_\mathrm{UV})$, for the UV LF likelihood the bright end of these simulated LFs is known to underestimate the number of galaxies and is therefore cut to $M_\mathrm{UV} > -20$. The faint end is poorly constrained because of the difficulty in observing faint galaxies. Unfortunately

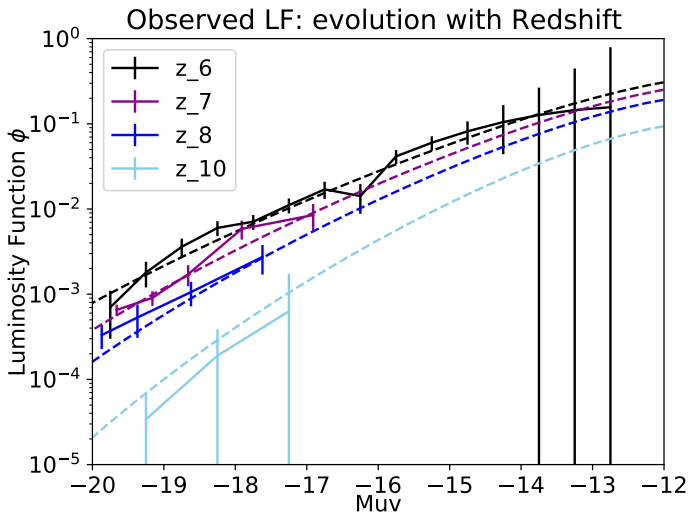


Figure 4.4: The observed UV LF for redshifts 6,7,8, and 10 as well as those simulated by Model C (dotted lines). Evaluation of the UV LF likelihood is done only at the observed data points (where the error bars are). See text for more detail.

.

for us faint galaxies are believed to dominate the photon budget for reionisation (Wise et al., 2014). Error and data values are taken from Bouwens et al. (2015) for $z = 6, 7, 8$ and Oesch et al. (2018) for $z = 10$. As with the observational prior checks, $\chi^2_{\rm LF}$ is linearly inputted to the likelihood.

The fit to the UV LF is used to produce the corresponding 21cm PS mock observation for model C (the fiducial parameters are given in Table 4.1). Using the above prescription allows the luminosity functions to obtain consistent UV emissivities, as shown in Figure D.2 for varying values of t_*. Bouwens et al. (2011) fit Schechter functions to obtain analytic forms of the observed UV LFs, which are comparable to those simulated here. Model C is the simplest implementation of an EoR model that is capable of fitting these data, as shown in Figure 4.4.

Gaussian random noise is added to all simulated UV LFs to approximate modelling uncertainties. The mean is taken along each implementation of Equation 4.14 with unit variance, the resulting value is used in the χ^2 calculation. Our statistical methodology had no trouble identifying the input parameters and increasing the variance above unity simply increased the number of samples required for convergence. All UV LFs shown in Figures have the variance reduced to 0.01 for clarity.

4.3.4 Model D - double power-law parameterisation

Model D is the most descriptive model used in this chapter. We relax the post heating regime, and makes use of the above UV LF calculation. The ionising efficiency is characterised by two power laws, one in the fraction of mass in stars, as before. And now there is a second power law in halo mass for the UV escape fraction. All the preceding physics is included as well as a double power-law in halo mass for the ionising efficiency,

$$\zeta = N_{\gamma/b} f_* f_{\rm esc} \sim f_{*,10} \left(\frac{M_{\rm h}}{10^{10} M_\odot}\right)^{\alpha_*} f_{\rm esc,10} \left(\frac{M_{\rm h}}{10^{10} M_\odot}\right)^{\alpha_{\rm esc}}, \tag{4.15}$$

where $N_{\gamma/b}$ is the number of ionising photons per stellar baryon. It differs from N_γ in previous chapters only so that our units are consistent with the new ionisation threshold presented in Equation 4.1. Model D is the full parameterisation used in Park et al. (2019). The 8 free[6] parameters are: f_*, $f_{\rm esc}$, α_*, $\alpha_{\rm esc}$, $\mathrm{Log}_{10}[M_{\rm turn}]$, E_0, $L_{X<2\rm keV}$, and t_*.

4.3.5 Model discussion

To understand the key features of the models in this chapter we plot the 21cm mock observations in a variety of ways (using the fiducial parameters in Table 4.1). First, Figure 4.1 shows the brightness temperature light-cone slices for the full calculation ($z < 35$). Ionised bubbles begin to percolate around $z \sim 10$, with all models heading for a reionisation ending around $z \sim 6$. Note that including the history of the light-cone (redshifts above those used in the likelihood) is only important for the models using the spin temperature. Each model uses the same underlying density field and hence contains structures (and the surrounding ionised bubbles in black) which form at the same spatial locations in each slice. The introduction of the X-ray heating is very apparent as the gas begins to cool and is then heated rapidly in of the middle of the Model B light-cone (red/yellow patch around $11 < z < 21$). This behaviour appears in the model D light-cone also, but is extended and delayed (roughly to $9 < z < 15$) due to the double power-law in halo mass adding more flexibility to ζ. Secondly, we look at the power spectra from each of these light-cones. The white vertical lines (from redshifts 8-10) on the light-cones indicate the chunks where each 21cm PS is calculated. Figure 4.2 shows these fiducial power spectra plotted per chunk, as used in the 21cm PS likelihood. The discrete dots indicate which

[6] A fixed $\alpha_{\rm X} = 1$ is implemented in the X-ray calculation within model D.

k values are used for the χ^2 calculation. The likelihood is calculated between the vertical lines in Figure 4.2, these are dictated by the telescope assumptions and simulation sizes (detailed in Section 4.3.6).

The fiducial parameters used for each mock observation in Table 4.1 are in agreement with the observational priors for all four models. For Model A and Model B, these parameters are selected from the literature (Greig & Mesinger, 2017b). Models C and D have their fiducial parameters selected in agreement with the UV LF data. While these data constrain the parameters in Model C well, the X-ray heating parameters in Model D are left to be sampled from within their uniform prior ranges. The brightness temperature realisations shown in Figure 4.1 show Model D to have significantly colder gas than in the other realisations. There is heating in Model D, but in this realisation it is delayed and reduced in comparison to the X-ray heating that occurs in Model B's realisation. Across the redshifts analysed ($z \in [8, 10]$): the brightness temperatures from Model D are progressing from absorption to emission; Model B's more extreme heating has almost saturated the spin temperature by $z = 10$ and has certainly done so by $z = 8$.

Most of the models (A, B, and C) have comparable brightness temperature amplitudes in the light-cone but the power spectra reveal that B has significantly brighter 21cm signal. However the inhomogeneous heating of the gas causes ionisation to progress faster on smaller scales. In comparison to B, model D has its heating epoch delayed. The model D 21cm PS begins (chunk 3) dimmer on all scales, before rising with a broader peak (Figure 4.3) that spans accross a large portion of both chunks 1 and 2. To help interpretation, Figure 4.3 shows the same fiducial 21cm PS evolving with redshift at a given scale. We must be careful with what inference we can be obtained when using chunks, their size will influence how well the different 21cm PS represent the evolution of the signal throughout the light-cone. This Figure has the number of chunks increased from 3 to 10 to illustrate the underlying physics, the points used appear significantly more jagged when interpolated (approximately the average of every three points plotted). When comparing Figures 4.2 and 4.3, the brightness temperature fluctuates more in the direction of redshift than in k. In Figure 4.3 Models A and C fluctuate a lot less than models B and D. The 21cm signal amplitude in model D has begun to diminish by $z = 8$ (chunk 1, lowest redshift), but it is still brighter than in models A, B, and C. The delayed impact of the heating epoch on model B compared with model D are particularly apparent here. Model D is

nearly an exact translation of model B on the scales used, except for an increase in absorption just below redshift 10. A, B, and C approach the same values at the $z = 8$ end, as the spin temperature saturates in model B - this is the post-heating approximation used in models A and C. The spin temperature in model D will also saturate, but this occurs a later ($z < 8$, not plotted here).

4.3.6 Telescopes

The telescope assumptions in this chapter use the *moderate* 21cmSense settings for a 1080 hour observation with SKA-512 as in the previous chapter (Pober, 2016).

The only difference here is that the 21cm PS is calculated from chunks of the light-cone rather than from coeval cubes. Since we are interested in performing our model selection tests in the backbone of reionisation, we are using the same redshift range ($z = 8 - 10$) as the previous chapter. Our 21cmSense simulations are performed with the frequency from the centre of each chunk.

The light-cone is constructed with coeval cubes calculated with a logarithmic step of $\Delta z_{\rm step} = 1.04$. For example, in our light-cone ($z = [8, 35]$) the redshifts sampled are 8, 9.05, 10.15, 11.31, 12.5 and so on. In the 21cmSense calculation the cosmological bandwidth is 8Mhz with 82 channels, all the parameters are shown in Table 4.2. This is the minimum size of signal that can be considered to have coeval cosmology and means we have 1024 bins per 100MHz. At redshift 8 this corresponds to a $\sim 210h^{-1}$ Mpc distance between redshifts 7.78, 8.23. We therefore are confident in the box size choice of $250h^{-1}$ Mpc being able to capture the appropriate physics in the likelihood (Deep Kaur et al., 2020). This is far from our telescopes' maximum frequency resolution, which would dictate a the minimum chunk size being is limited by the cosmological bandwidth.

There is a caveat from only using three light-cone chunks. Signal shape is potentially missed depending on how quickly reionisation evolves. Should our analysis fail to obtain the astrophysical parameters with sufficient precision, more chunks may be necessary to accurately pin down the evolution of the 21cm PS within the light-cone. This is partly a failure of the power spectrum as it is a non-ergodic representation of the entire light-cone. We return to this discussion in Chapter 5.

Parameter	SKA-Central
Number of dipole Stations	296
Station Diameter [m]	35
Collecting Area [m^2]	492,602
T_R[K]	$40 + \frac{T_{sky}}{10}$
Observing Width [MHz]	[50,350]
Observing Time [hrs]	1080
Scan Type	1-hr Track

Table 4.2: The telescope parameters used in the 21cmSense simulation for SKA-512 with (Pober, 2016). We only require the central 296 of SKA-512's planned stations as this is where sensitivity to the EoR lies. For more detail see Section 4.3.6.

We defer an analysis of which power-law in D is prevalent when including multiple realisations of the spin temperature fluctuations in models that include the UV LF. We will also be better equipped to answer this with the improved telescope assumptions as discussed above and in Section 4.5.

4.4 Considerations of prior sensitivity

All parameters prior distributions in this chapter are uniform accross the ranges shown in Table 4.3.

For model A, since our parameterisation is similar to the previous chapter so are our prior considerations (Binnie & Pritchard, 2019),

- ζ is bound within $[10, 250]$. The bottom end has been increased from 5 to 10 from our previous work, this eases computation without compromising the physics. we adopt a maximum of 250 to ensure this current observations are captured, values around 200 should be more than enough but we are cautious.

- $\text{Log}_{10}[T_{\text{vir}}] \in [4.0, 6.0]$. The atomic cooling threshold sets the lower end, we do not consider mini-halos in this chapter. 10^6K corresponds to the sizes of galaxy clusters, it is unlikely that galaxies larger than (or near to) this contribute to reionisation (as seen by forward modelling simulations - Wise et al. (2014); Feng et al. (2016) for example).

Next is Model B, introducing the affects of X-ray heating and spin temperature fluctuations (Greig & Mesinger, 2017b).

- $L_{\mathrm{X<2\ keV}}$/SFR The X-ray spectrum is normalised in Equation 4.6. It is implemented as a log prior in $[10^{38}, 10^{42}]$ erg s^{-1}M$_{\odot}^{-1}$yr, motivated by observations e.g. (Saitou et al., 2011).

- E_0 is varied in the range $[100, 1500]$ eV. In principle the top end could be extended to 2000 eV - which would imply no soft extra emission ($\nu_{\mathrm{X}} \propto E_{\mathrm{X}} < 2$keV) emerging from the host galaxy. The used limits are motivated by the ISM HI column density from hydrodynamic simulations, corresponding to $\log_{10}[N_{\mathrm{HI}}\ \mathrm{cm}^{-2}] \in [19.3, 23.0]$ (Das et al., 2017).

- $\alpha_{\mathrm{X}} \in [-1, 3]$, covers the range of power-laws for typical X-ray sources that populate galaxies during the EoR. This is discussed further in Section 4.5.4.

The α_{X} parameter is known from Greig & Mesinger (2017b) to be degenerate with E_0. To analyse this we ran 21CMNEST with a reduced model B parameterisation for its mock 21cm PS data, i.e. just α_{X}, just E_0, and both. We found that the posterior variance on this(these) parameter(s) in either individual option more than doubled compared to when using both (discussed further in Section 4.5.4). The changes in Δ^2_{21} are dominated by E_0 in the backbone of the EoR accross the prior ranges above. α_{X} is therefore left constant in model D ($\alpha_{\mathrm{X}} = 1$), after showing this with the SDDR.

To calculate the UV LF the following parameters (Park et al., 2019) are used:

- t_*, the fraction of Hubble time, which cannot exceed [0,1]. See Section 4.5.4 for further discussion.

- M_{turn} is implemented as a log prior across $[10^8, 10^{10}]M_{\odot}$. The upper limit is ruled out by the UV LF observations, and the lower limit is from the atomic cooling mass in halos[7].

- α_*, the strength of scaling between stellar collapse fraction and halo mass, arbitrarily set to $[-0.5, 1]$ (See text below).

[7]The numerical simulation containing spontaneous emission coefficients and stellar spectra have been performed with RECFAST (Seager et al., 1999). Pre-calculating these speeds up the simulation, allowing MCMC analysis, we do not explore the statistical influence of expanding this prior.

- $f_{*,10}$, we implement this with a log prior over $[10^{-3}, 1]$. This parameter is normalised at the $10^{10} M_{\odot}$ scale i.e. $f_* < 1$ means stellar mass cannot exceed the total baryonic mass in that region. The bottom end of $f_* > 0.001$ removes galaxies that have negligible[8] contributions to reionisation.

The number of ionizing photons per stellar baryon, $N_{\gamma/b}$, is degenerate with f_*. Using $N_{\gamma/b} = 5000$ influences the appropriate prior range for this parameter, this value is chosen to be consistent with observations of stellar variation (that the baryons can collapse into) (Sun & Furlanetto, 2016).

Finally the second power-law in the ionisation efficiency is implemented with:

- $\alpha_{\rm esc} \in [-1., 0.5]$ is used here, constraining the power-law scaling with halo mass (the same as α in Chapter 3).
- $f_{\rm esc,10}$, normalised on a $10^{10} M_{\odot}$, the escape fraction is between $[10^{-3}, 1.]$

The upper end should should not exceed 1 as it is a fraction. $f_{\rm esc}$ is limited to $> 10^{-3}$ in order to agree with the observational priors i.e. reionisation finishing by $z \sim 6$. This is similarly to the cut at the bottom end of f_*, if not enough ionising radiation can escape the galaxy it will not effect reionisation and it can be safely omitted irrespective of how efficiently it is producing stars and vice versa.

Table 4.3 contains a summary of the parameter prior distributions for reference. Section 4.5.1 discusses actively changing prior distribution ranges for specific analyses.

4.5 Results

Firstly we found agreement between the UV LFs compared with the observed data (Bouwens et al., 2015; Oesch et al., 2018); agreement is also found between our parameter posterior distributions and those produced by 21CMMC in most cases (Appendix D.1). Each model is able to retrieve its own fiducial parameter set (and the observed data for the UV LF calculations) confirming the basic methodology.

[8] Intentionally avoiding philosophical discussions of what defines the first galaxy.

Parameter	Model	Prior Range
$\text{Log}_{10}[T_{\text{vir}}/\text{K}]$	A, B	$[4.0, 6.0]$
ζ	A, B	$[10, 250]$
E_0	B, D	$[100, 1500]$ eV
α_{X}	B	$[-1, 3.0]$
$L_{X<2\text{keV}}$	B, D	$[38, 42]$ erg $\text{s}^{-1}\text{M}_{\odot}^{-1}\text{yr}$
$\text{Log}_{10}[M_{\text{turn}}/M_{\odot}]$	C, D	$[8, 10]$
$f_{*,10}$	C, D	$[-3, 0]$
α_*	C, D	$[-0.5, 1]$
t_*	C, D	$[0, 1]$
$f_{\text{esc},10}$	D	$[-3, 0]$
α_{esc}	D	$[-1, 0.5]$

Table 4.3: 21cmNest distributes each parameter across uniform prior distributions with the ranges shown here. The total parameter prior volume for each of the four models is 480, 10752000, 9, and 226800 respectively. For use in model D, α_{X} is fixed to 1 and t_* has its prior range reduced to $[0.05, 1.0]$ (see Section 4.5.4). When fitting models with a heated gas approximation to those that model the spin temperature fluctuations (A or C to D or B), the ζ implementation has its upper bound increased with no noticeable benefit (see Section 4.5.1 for detail).

Bayes factors are shown in Table 4.4 where the table axis (I, J) dictate the numerator I|J of the Bayes factor calculation in question. The corresponding denominator in each Bayes factor is accross the bottom. The notation I|J, where I, J $\in$ Model A, B, C, or D, represents an Evidence value obtained by a 21cmNest run simulating model I, when reproducing fiducial (mock) data produced by model J (with the fiducial parameters given in Table 4.1) e.g. I|J = P(Model I|MockData J). The MAP values for main analyses are shown in Table 4.5, with corresponding 21cm PS shown in Figure 4.5. Deviations from the four main models are written into the corresponding I (or J) e.g. NoLF or OnlyLF, detailed in Section 4.5.3. We now proceed to answer the main questions of the paper (bullet points in Section 4.1).

4.5.1 Prior range discussion

For reference the prior ranges are shown in Table 4.3 and the MAP parameters are shown in Table 4.5. To test the robustness of model A's results we extend the upper bound of the prior distribution for ζ from 250 to 10^3. The MAP values for show ζ against the upper bound of the prior ($\zeta \approx 250$). $\zeta = 10^3$ is well beyond the physical motivations discussed in the previous section, but here we are only testing the statistical methodology (this value was selected due to computational limits). The posteriors obtained while calculating A|B and A|D are inhibited by

Model	A	B	C	**Mock Fiducial** C-NoLF	ObsLF	D	D-NoLF
A	1.0	$\mathbf{e^{-6685}}$	(23000)	**0.027**		(e^{-910})	$\mathbf{e^{-920}}$
B	0.0028	**1.0**.	(670)	**0.00077**		(0.93)	**0.0066**
C	(10^{-5})	(e^{-6144})	**1.0**			$\mathbf{e^{-1105}}$	
C-NoLF	4.4	$\mathbf{e^{-6686}}$	(870000)	**1.0**		(e^{-1273})	$\mathbf{e^{-1278}}$
C-OnlyLF			(290)		**1.0**	(420)	
D	(10^{-31})	(e^{-32})	$\mathbf{10^{-6}}$			**1.0**	
D-NoLF	10^{-6}	**0.33**	(600)	**0.00069**		(140)	**1.0**
D-OnlyLF			(410)		**1.4**	(600)	
Comparison	A|A	B|B	C|C	C-NoLF|C	C-OnlyLF|ObsLF	D|D	D-NoLF|D

Table 4.4: The complete list of Bayes factors used in this work (2 s.f. or nearest integer power if large), key values are shown in Figures 4.6, 4.9, and 4.12 (identified in bold). We include the calculation for models C and D for calculations with and without the UV LF, the brackets to indicate unweighted comparisons (see Section 4.5.3) i.e. like-for-like likelihoods must be compared for sensible Bayes factor calculations. The notation (model I fit for fiducial mock data set J, as in Section 4.5) shows each Bayes factor calculated with the Evidence I|J from this Model|Fiducial Table in ratio with the Evidence of the Comparison row (J|J on the denominator). The unweighted calculations provide insight into the washing out of the UV LF depending on the realisation of the 21cm signal. From a theoretical perspective to accurately analyse these different methods of likelihood requires weighting the likelihood contributions differently, they are included mainly for completeness and pedagogy. Empty spaces are for calculations that cannot be performed (e.g. models A and B do not produce UV LFs). The notation (described under Section 4.5) includes 'NoLF' for using both the observations prior checks and the 21cm PS likelihood, while 'OnlyLF' have $\mathcal{Z}$ calculated with just $\mathcal{L}_{\rm LF}$. The symmetry of the table is not exact due to different models having different fitting flexibility (in terms of the shape of the UV LF, 21cm PS or both). See text for discussions of these results.

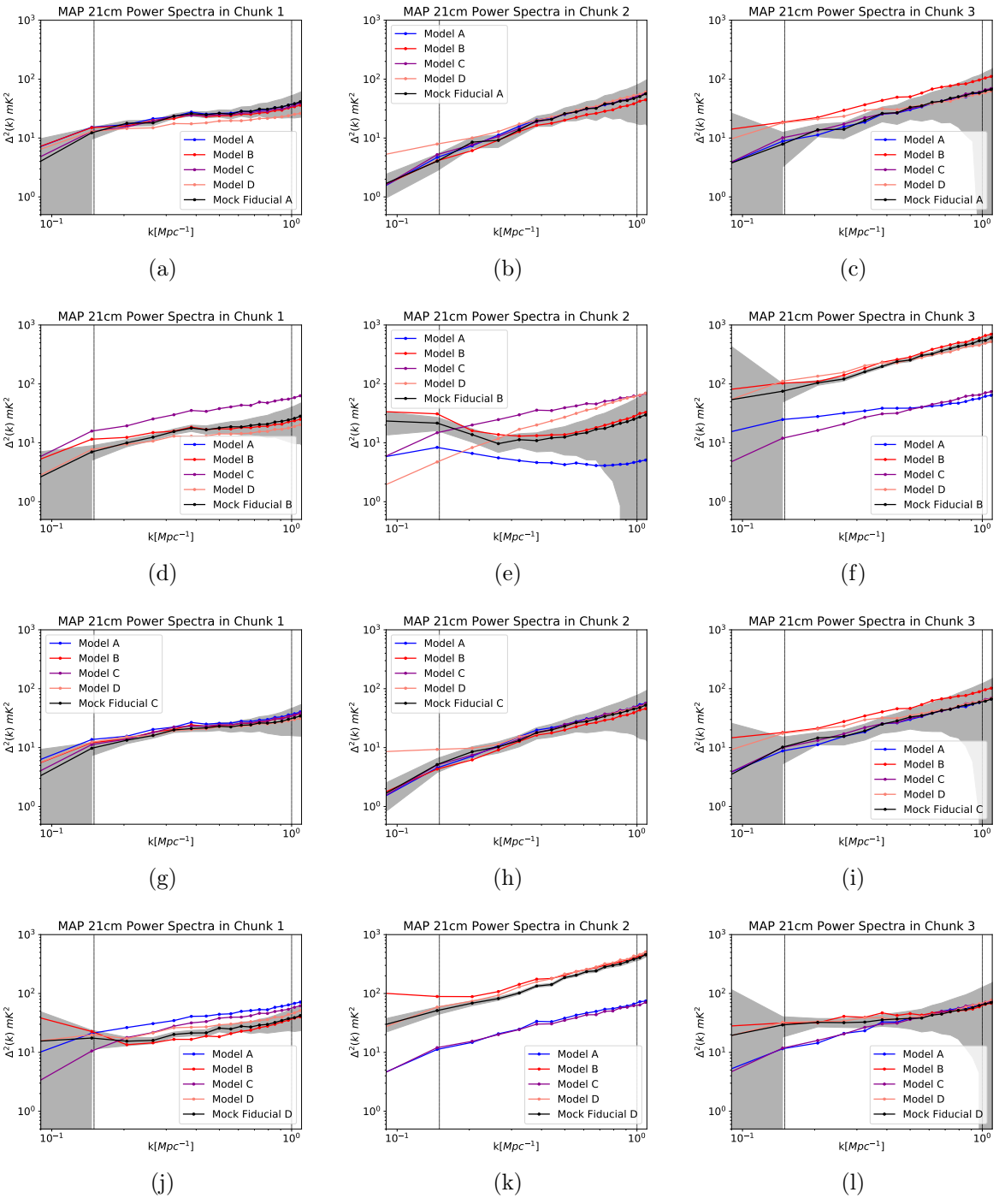


Figure 4.5: MAP 21cm PS fit for each model (A, B, C, and D are blue, red, magenta and orange respectively) with the mock fiducial data set (in black, one row for each respective model). The shaded grey region represents the 21cmSense error estimate for a $\sim$ 1000 hour SKA-512 observation (Section 4.3.6). The 21cm PS for models A and C, with saturated spin temperatures, are fit well by models B and D (rows 1 and 3), although some large scale structure is gained in removing post heating approximation at the highest redshift (chunk 3, low k rises out of the telescope error region). For the fiducial mock data B (2nd row) the 21cm PS of models A and C are unable to fit the mock 21cm PS without including X-ray heating: Model A's attempts fit to the heated gas at high redshift have caused reionisation to end by chunk 1, resulting in heavy penalties from the fiducial mock data and observational priors; Model C is able to fit closer but at the cost of heavy penalties from the observed UV LF. Model D is capable of fitting the mock B 21cm PS, when $\mathcal{L}_{\rm LF}$ is not included (not shown, but can be interpreted from the bottom row where B is fit to the 21cm PS mock data for D). When both likelihoods are used 4.5(e) shows that D must compensate its fit on the 21cm PS to correctly fit the UV LF data (see also Figure 4.13(a)).

Model	Parameters	Mock Fiducial Data A	B	C	D
A	ζ,	32.7 ± 4.5,	250 ± 0.07,	31.1 ± 4.1,	250 ± 0.6,
	$\mathrm{Log}_{10}[T_{\mathrm{vir}}]$	4.66 ± 0.08	4.20 ± 0.0002	4.63 ± 0.08	5.05 ± 0.06
B	$L_{\mathrm{X<2keV}}/\mathrm{SFR}$,	41.2 ± 0.14,	40.0 ± 0.09,	41.2 ± 1.5,	40.0 ± 0.07,
	E_0,	885 ± 240,	440 ± 136,	765 ± 254,	481 ± 110,
	α_{X},	-0.304 ± 1.2,	0.93 ± 1.2,	-0.737 ± 1.2,	2.92 ± 1.1,
	ζ,	34.7 ± 4.6,	33.6 ± 6.0,	31.3 ± 4.3,	183 ± 22.4,
	$\mathrm{Log}_{10}[T_{\mathrm{vir}}]$	4.66 ± 0.08	4.68 ± 0.9	4.69 ± 0.08	5.46 ± 0.01
C	α_*,	0.477 ± 0.071,	0.815 ± 0.013,	0.484 ± 0.07,	1.00 ± 0.03,
	$f_{*,10}$,	-1.19 ± 0.07,	-1.71 ± 0.06,	-1.2 ± 0.06,	-1.02 ± 0.03,
	M_{turn},	8.92 ± 0.17,	10.0 ± 0.04,	8.84 ± 0.18,	10.0 ± 0.02,
	t_*	0.586 ± 0.093	0.054 ± 0.011	0.523 ± 0.09	0.998 ± 0.019
D	α_{esc},	0.498 ± 0.10,	0.469 ± 0.04,	0.184 ± 0.178,	-0.588 ± 0.196,
	α_*,	0.538 ± 0.08,	0.147 ± 0.025,	0.545 ± 0.074,	0.508 ± 0.64,
	$f_{*,10}$,	-1.26 ± 0.29,	-1.29 ± 0.326,	-0.304 ± 0.315,	-1.20 ± 0.34,
	$f_{\mathrm{esc},10}$,	-1.19 ± 0.27,	-0.922 ± 0.322,	-1.94 ± 0.30,	-1.13 ± 0.27,
	$L_{\mathrm{X<2kev}}/\mathrm{SFR}$	41.0 ± 0.14,	40.3 ± 0.20,	40.9 ± 0.1,	40.0 ± 0.2,
	M_{turn}	8.62 ± 0.23,	8.00 ± 0.01,	8.58 ± 0.20,	8.69 ± 0.16,
	E_0	631 ± 143,	639 ± 270,	560 ± 170,	516 ± 340,
	t_*	0.638 ± 0.27	0.992 ± 0.350	0.112 ± 0.268	0.699 ± 0.265

Table 4.5: The MAP values found by 21CMNEST for each full model (3 s.f. $\pm 1\sigma$ with matched d.p.). The corresponding 21cm PS are shown in Figure 4.5. Example MAP UV LFs for C and D are shown in Figure 4.13.

the parameter priors, we want to be sure this has not influenced the Bayesian Evidence values. We found an increase in $\ln\mathcal{Z}$ to be ~ 700, resulting in a rejection on model A (in comparison to model B) of order $\ln\mathcal{B} \sim 6000$, i.e. the expanded prior range does not change the conclusions discussed in Section 4.5.2.

C|B and C|D also produced parameter posteriors pushing against the upper end of prior ranges. To explore exaggerating the ionising efficiency (as above does with ζ) we expand the parameter priors ranges for $f_{*,10}$ and α_* to $[-3., 3.]$ and $[-1., 2.]$. The MAP values for both parameters remained within the original prior distribution bounds. As a cross check the σ value on either parameter remained unchanged (as a percentage of the used prior range). The MAP parameter values for t_* and $M_{\rm turn}$ show similar behaviour. t_* is explored further in Section 4.5.4, and $M_{\rm turn}$ has its upper prior range limited by UV LF observations ($M_{\rm turn} \approx 10$). We interpret this further with the context of Section 4.5.2 i.e. Model B's mock 21cm PS data could be an unphysical X-ray heating realisation, and hence this needs to be looked into further.

4.5.2 Are models with spin temperature fluctuations distinguishable?

For the parameter set chosen for the fiducial mock data (model B, Table 4.1), the effect of the spin temperature is objectively distinguished. Models A and C do not have access to the spin temperature fluctuations, and score *strong* disfavouring in Figure 4.6. This figure plots the Bayes factors for models A, B, and C as they each fit a 21cm PS to the fiducial mock 21cm PS data produced by model B (using the parameters in Table 4.1). For ease of interpretation the grey shaded regions indicate the Jeffreys' scale described in Section 1.4.6.

The log Bayes factor for these three models under question the are -6685, 0, and -6144 for A|B, B|B, and C-NoLF|B respectively (in ratio with B|B). The simple model (A), and single power-law model (C, $\zeta(M_{\rm h})$) both score *strong* disfavouring. This indicates that the spin temperature fluctuations should be detectable, given the assumptions we have used for our ~ 1000 hour observation with SKA-512. The Log factors in the 1000's are large since they do not include important physics and can never reproduce the spin temperature models.

As the spin temperature is approaching saturation, the change in amplitude of the 21cm PS is different on multiple scales. In particular, C & A struggle with the increase in power on all scales at higher redshift ($z \sim 10$, 4.5(f)), and the lack of mid-scale power in the middle chunk

($k \sim 0.5\text{Mpc}^{-1}$, 4.5(e) - the visible dip in the red 21cm PS) introduced by spin temperature fluctuations. Most of the fitting attempts cause large penalties from the observational priors (particularly from $\mathcal{L}_{\text{McGreer}}$, e.g. A's fit shows reionisation ending in chunk 1, Figure 4.5(d)). Here, 21CMNEST has validated the results of 21CMMC by using Bayesian model selection instead of parameter estimation. Greig & Mesinger (2017b) show that including the detail in IGM heating should produce 21cm PS that differ noticeably in a parameter estimation framework. When model B tries to reproduce the simpler models' 21cm PS it can. The Bayes factors for B|A, B|B and B|C are 0.74, 1.0 and 2.01, both falling within the *weak* Evidence score on the Jeffreys' scale (not plotted, Table 4.4). This shows that the detail added by these extra parameters is not redundant[9], and not every redshift in reionisation need be observed to do so.

X-rays have a large mean-free path, which is longer depending on the hardness of the radiation (Pritchard & Furlanetto, 2006). They heat the hydrogen in the IGM after being redshifted by the Hubble flow, which enables their interaction with the Lyman series (Furlanetto & Pritchard, 2006). Approximating this heating mechanism with a saturated spin temperature ($T_{\text{S}} \gg T_{\text{CMB}}$) will not be valid until very late in reionisation. Because of the inhomogeneous nature of the heating, the percolation of ionisation is subtly different between the two models. Greig & Mesinger (2017b) show that in $z \in [8, 10]$, the variety in bright or faint realisations of EoR galaxies will have a different impact in how closely their 21cm PS follows the approximated equivalent. It is therefore comforting that Bayesian model selection is able to distinguish models which use the post-heating approximation (A & C) from models using full calculation (B) close to the saturation of the spin temperature. Figure 4.5 shows the 21cm PS for model B (in Red) as it fits for each of the four fiducial data sets (Each row representing each mock fiducial data set).

How distinguishable X-ray heating is from ionisation alone depends intricately on the 21cm PS amplitude, and the telescope observation. If the fiducial parameter set we have selected is the real observational truth, the 21cm PS will have a large variety of amplitudes (at multiple scales) enabling us to identify model B's parameterisation of X-ray heating. Should this not be the case, then a larger observational range will be necessary to break parameter degeneracies and the full observational capabilities of the SKA will be a necessity (making this hard to cross-

[9] apart from α_{X}, see Section 4.5.4.

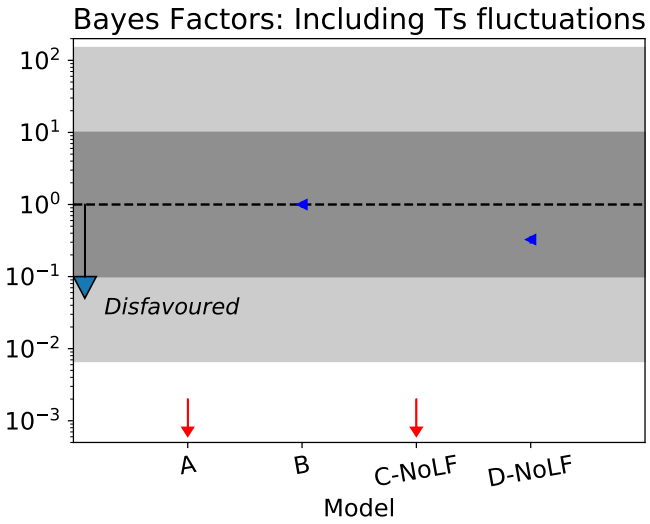


Figure 4.6: The Bayes factor models A, B, C and D (C, and D without the UV LF likelihood), as they recover the fiducial mock 21cm PS data produced by model B. Blue represents models that include spin temperature fluctuations; red represents models in the post-heating approximation (however both of these fall below the axis, represented by the arrows). A summary of the models and the fiducial parameters used can be found in Table 4.1. The grey regions represent the *weak* (dark), *moderate*(light), and *strong* (white) Evidence outcomes on the Jeffreys' scale.

check with smaller experiments). Should this have been a problem with the fiducial parameters used, we could extend the 21cm PS likelihood to include more redshifts (maximally [6-25]). These should be chose to provide a variety in 21cm PS amplitudes from the X-ray heating realisation which will be done in future work, when more precise models for telescope noise and foreground removal are applied.

4.5.3 Inference from the UV luminosity function

Model C is the simplest parameterisation that can produce UV LFs in agreement with the observations (Figure 4.4). C also contains a power-law in halo mass for the collapsed baryon fraction in stars (Equation 4.9). In our previous work (Binnie & Pritchard, 2019), we showed that the 21cm PS alone was not enough to distinguish a power-law in halo mass for the ionising efficiency in the post-heating approximation. This remains the case here with Bayes factors 4.4, and 1.0 for A|C and A|A (not plotted, see Table 4.4), i.e. *weak* favouring of the simpler model when only the 21cm PS is used. In contrast, when including the UV LF in the likelihood ($\mathcal{L}_{\rm LF}$) the power-law parameters ($f_{*,10}$, α_*) are constrained to $\sigma \sim 0.1$ (1 s.f.). Tightly constrained posteriors are obtained when the UV LF is included ($C|C$, Figure D.1(c)).

We calculated the Bayesian Evidence with alternate combinations of the likelihood (identified by different colours in Figure 4.7). $\ln\mathcal{Z}$ was calculated with: $\ln\mathcal{L}_{21} = \chi^2_{21\text{cm}} + \chi^2_{\text{Planck}} + \chi^2_{\text{McGreer}} + \chi^2_{\text{Greig}}$ (labelled noLF, blue); $\ln\mathcal{L}_{\text{LF}} = \chi^2_{\text{LF}}$ (onlyLF, black); and the full version from Equation 4.2 (red). Figure 4.7 corresponds to integrating the likelihood space of model C as it reproduces the observed UV LF data and fiducial 21cm PS mock data made by each model. In particular the deviations of the black points down from the dotted line in 4.7(a) comfortingly infer that the $\ln\mathcal{Z}$ is non-negligible in comparison with the 21cm PS (~ 3 orders of magnitude difference). To analyse the UV LF further we plot the C|C posteriors with only $\mathcal{L}_{\text{LF}}$ and without the use of the $\mathcal{L}_{\text{LF}}$ in Figure 4.8 (blue and purple respectively). Here it is clear that the pinning down of the power-law parameters has come from $\mathcal{L}_{\text{LF}}$ (as it should, Section 4.3.3), but the synergy of 21cm and UV observation is necessary to pin down every model C parameter. Bayes factors for these are plotted in Figure 4.9. Each Bayes factor is measured against the mock observation produced by model C's fiducial parameter set, including ratios between Evidence's of similar likelihood construction only.

Model C can be distinguished from model D with just the 21cm PS and both the UV LF and 21cm PS (Figure 4.9). *Strong* disfavouring of D is achieved with the 21cm PS and both the UV LF-21cm PS synergy. The fiducial parameters for model C's mock 21cm PS data is set by fitting C to the UV LF data. Without the UV LF data model C shows similar flexibility to the 4pFZH model fitting for f1 or f2 in the previous chapter. Penalties for D's extra parameters are exaggerated when both are used, D is rejected more heavily with both likelihoods than just the 21cm PS as a result. The change in parameter prior volume between models C to D (ln ratio of -10.13, See Table 4.3) influences the *strong* rejection of DnoLF|C. When only using $\mathcal{L}_{\text{LF}}$, both C and D fits produce UV LFs via the same mechanism, hence C and D are approximately the same in Figure 4.9. However, when reproducing the mock fiducial 21cm PS from C, the X-ray heating parameters are redundant causing a penalty against model D.

Including the spin temperature (given our parameterisation in Section 4.5.2), Figure 4.7(b) shows that the Evidences obtained with the X-ray heating models differ from those without by $\sim$ 1000s of orders of magnitude. Here the UV LF likelihood is providing negligible inference in context of the 21cm PS's input towards model selection (black points remain on the dotted line in Figure 4.7(b)). Since the clarity of the observation (signal to noise) depends on the amplitude of the 21cm PS, the X-ray heating realisation is important - particularly if we are

constrained to using only the backbone redshifts of the EoR. If the X-ray heating influence is small (smaller than in Model B's mock fiducial observation), it is likely that observations of the 21cm PS will need to be deeper into the cosmic dawn than just the EoR backbone if decisive statistical results are desired[10]. Furthering this point, the X-ray heating can produce 21cm PS amplitudes which are larger than physically viable from the constraints from including the UV LF data. The UV LFs respond mainly to the galaxy parameterisation ($M_{\rm turn}, \dot{M}_*$), while the X-ray parameters (E_0, $L_{\rm X<2keV}$/SFR, $\alpha_{\rm X}$) depend mainly on the 21cm PS. Since there is some degeneracy between these, their dependency on the choice of fiducial parameters must be looked into further. A closer analysis is needed relating to the behaviour of the UV LF depending on the realisation of the X-ray heated spin temperature fluctuations. We defer this to future work, when we are using the aforementioned improvements on telescope assumptions.

It is worth noting here that the differing likelihood constructions are not directly comparable. A weighting analysis[11] would be necessary to cross compare these likelihoods individually (hence multiple large bracketed values in Table 4.4). But since we are interested in their model selection synergy the conclusions in this work source from like-for-like comparisons of the likelihood only.

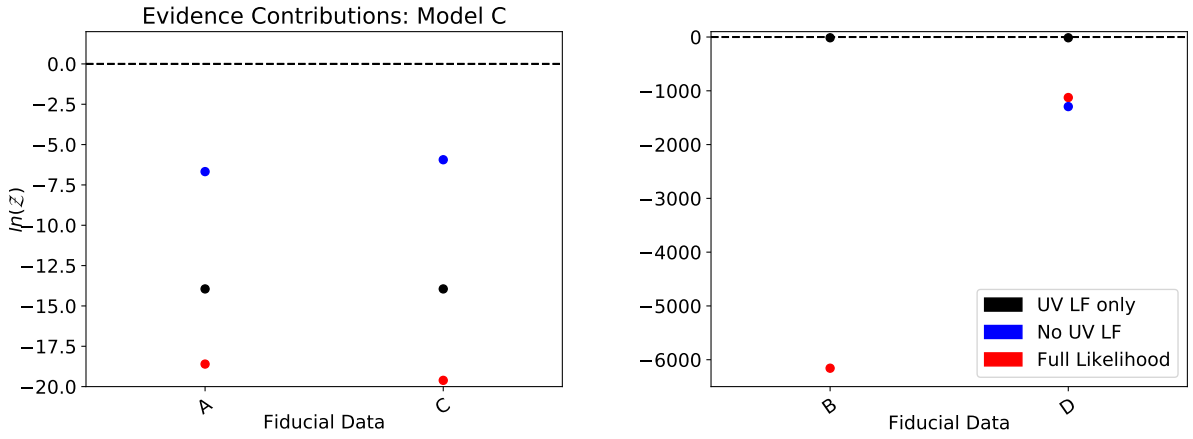
4.5.4 Parameter analysis with the SDDR

The SDDR is detailed in Section 1.4.7 and has already been implemented in Chapter 3 (Binnie & Pritchard, 2019).

We test $\alpha_{\rm X}$ and t_* since the posterior variance on these parameters are comparable to their respective to prior distribution ranges (the B|B and D|D posteriors are visibly flatter than for the other parameters in Appendix D.1). The posterior shape directly relates to the shape of the SDDR Bayes factor plots (in Figures 4.10 and 4.11).

[10]Ideally the full redshift range used in (Greig & Mesinger, 2018, 2017b) will be observed, but we would like to know how much of that portion is necessary to decisively select the most authentic astrophysical scenario.

[11]Each likelihood contribution would have to be weighted as $\ln\mathcal{L} = \sum \alpha_i \chi_i^2$ e.g. Trotta (2008), requiring a separate Bayesian analysis (beyond the scope of this work) to pin down each α_i. We have no reasons to believe the inherent weighting in each χ^2 due to observational noise is inadequate.



(a) (b)

Figure 4.7: 4.7(a) The UV LF likelihood calculation ($\ln\mathcal{Z}$ in blue) is providing more inference, (but importantly of comparable magnitude) to those obtained with the 21cm PS (and observational prior checks, red) with the post-heating approximation applied. 4.7(b) shows the UV LF contribution becoming negligible in comparison to that of the 21cm PS when spin temperature fluctuations are included (the black points are negligible in the lower plot). Models that include spin temperature fluctuations require the 21cm PS to distinguish them (the huge difference between the red points for models B and D). See Section 4.5.3 for more detail and discussion.

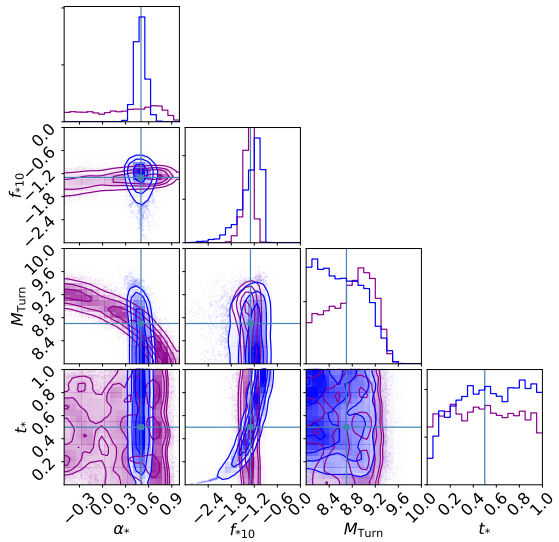


Figure 4.8: The posteriors for model C's parameters when reproducing only the 21cm PS (and observational prior checks, purple) and for reproducing only the observed UV LF data (blue). In the notation discussed under Section 4.5, these are the parameter posteriors for C-noLF|C (purple) and C-onlyLF|C (blue) respectively. The parameter posterior from the full likelihood (Figure D.1(c)), is approximately a convolution of these two distributions. $M_{\rm turn}$ and t_*, in particular, look to require a combination of both posteriors to obtain the fiducial model parameter. It is clear that the UV LF likelihood dominates the constraints on parameters that dictate the power-law in ζ. This can be seen by the sharp peaks for both α_* and $f_{*,10}$ in blue but not purple.

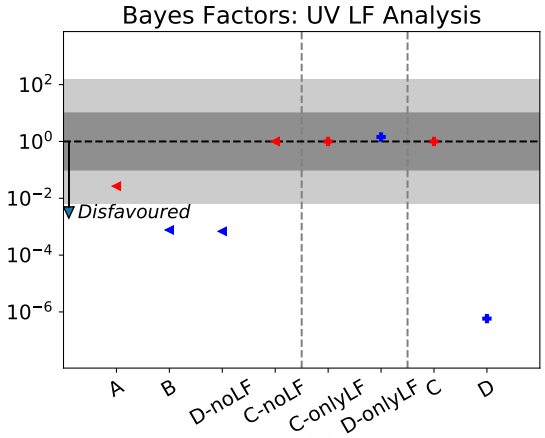


Figure 4.9: Bayes factors for each model for fits of the fiducial mock data set from C. This plot follows the colour conventions of Figure 4.6, the + sign indicates the use of the UV LF likelihood, with vertical dotted lines separating like-for-like analyses. Without the UV LF the simple model (A) shows *moderate* Evidence but since it cannot produce the UV LF there is no comparison for the full likelihood (due to no halo mass dependence in the UV LF). B is rejected by the $\mathcal{L}_{21}$ with *strong* Evidence however this is largely to do with the increase parameter space being redundant (particularly from E_0). Model D is able to fit the UV LF data better than C but again the extra parameter space is largely redundant resulting with a *weak* conclusion. The full model D (i.e. (D|C/C|C) on the far right) is rejected with *strong* Evidence due to the same reasons as B. See 4.5.3 for more detail.

How informative is α_X?

The X-ray spectral energy distribution is $L_X \propto E^{-\alpha_X}$, where α_X can vary as $[-1., 3.]$ (Greig & Mesinger, 2018). Soft X-rays are used, as this minimises the α_X and E_0 degeneracy (with respect to the 21cm PS). This also allows the X-ray constraints to be compared with nearby observations (e.g. Fragos et al. (2013); Lehmer et al. (2016), see Greig & Mesinger (2017b) for a fuller description and reference list).

The range of α_X represents different source types that produce X-ray radiation. A sensible range for this parameter could be $[-2, 3]$ encapsulating high/low mass X-ray binaries (HMXB/LMXB), mini-quasars, diffuse ISM emission, and supernovae remnants (Eide et al., 2018). We see in Figure 4.10(a) that below $\alpha_X \in [-1., 3.]$ this parameter scores *strong* disfavouring on the Jeffreys' scale. Figure 4.10(b) shows negligible changes to the 21cm PS for $\alpha_X \in [-2., 3.]$. For $\alpha_X < -1$, the rejection comes almost entirely from the observational prior checks ($\mathcal{L}_{\rm McGreer}$ to be specific). In Binnie & Pritchard (2019) we showed that these observational prior checks only have influence in particularly unphysical regimes. Therefore it is likely that the QSO constraints in the observational priors can rule out the particular sort of X-ray

source that correspond to $\alpha_X < -1$. For $\alpha_X > -1$, we are unlikely able to distinguish between X-ray sources.

Since we have been using redshifts $[8, 10]$, we are near the post-heating regime, and the effect of α_X fluctuating the 21cm PS is minimal. Between redshifts $[10, 18]$, α_X is likely to have more impact on the 21cm PS[12]. To test the degeneracy between E_0 & α_X, we simulated model B with only α_X, only E_0, and only both $[\alpha_X, E_0]$ as the free parameters (the other parameters were fixed to fiducial parameters for the mock fiducial data set from model B). We looked at the ratio of posterior variances (not shown) to uniform prior distribution width, obtaining 13%, 20%, and $[27\%, 40\%]$ respectively to the nearest integer. During the backbone of the EoR, the E_0 parameter has more influence on the 21cm PS.

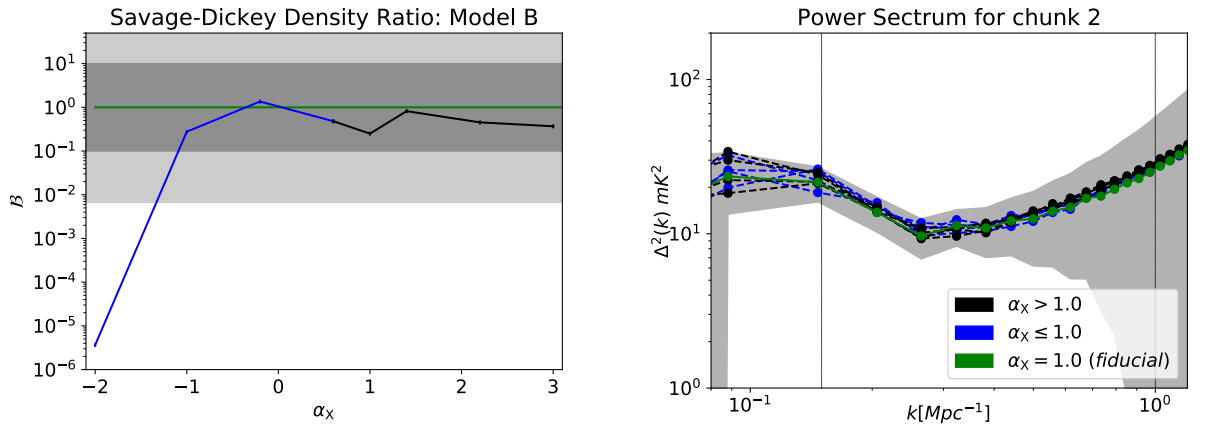
The SDDR results for α_X are shown in Figure 4.10. Figure 4.10(a) shows the Bayes factors and Figure 4.10(b) the variety in 21cm PS with respect to α_X. The constant behaviour of $\mathcal{B}(\alpha_X)$ in Figure 4.10(a) indicates that this parameter is not providing any additional inference given the observations. This is reflected in Figure 4.10(b) by all power spectra lying within the grey telescope error region. To conclude we omit α_X as a free parameter for the X-ray heating within model D. For simulation purposes we proceed with $\alpha_X = 1$, prescribing the dominant X-ray background producer as HMXBs (Das et al., 2017).

How informative is t_*?

The time scale for star formation t_*, is used to calculate the SFR (Equation 4.12), and is conventionally interpreted as however long is needed to form the stars[13] necessary for the produced UV LF. It is implemented assuming the dynamical timescale of our universe scales as $1/H(z)$; and that the fraction of mass collapsed in stars, does so linearly. t_* cannot exceed $[0, 1]$, as it is a fraction of the Hubble parameter. The minimum timescale for star formation to occur is $\sim$ Myr (crudely estimated with the Jeans time), which would correspond to $t_* \gtrsim 0.0001$. Estimates for the cosmic dawn happen around $z \sim 35$, implying $t_* \sim 0.5$. This agrees with the behaviour of Figure 4.11(a) which shows the SDDR analysis for t_*.

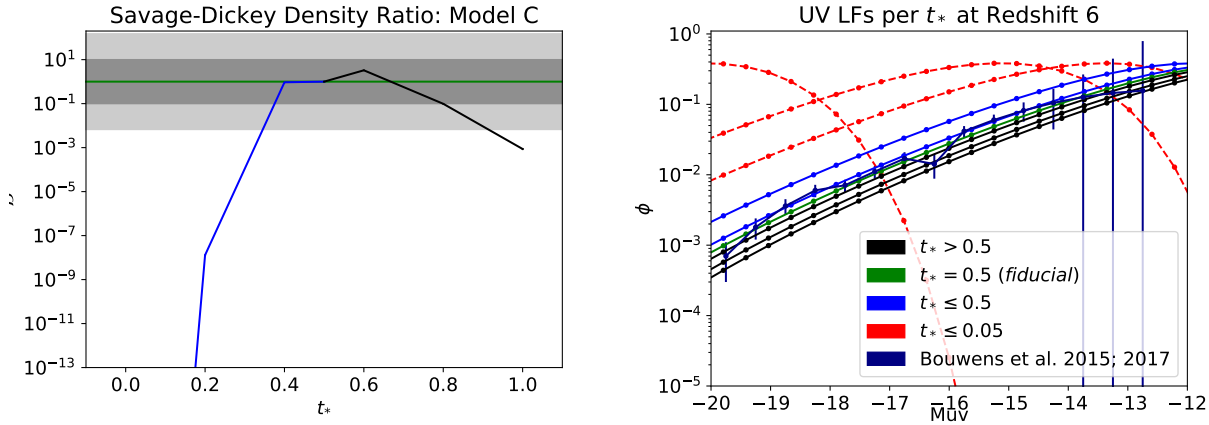
The $\mathcal{B}(t_*)$ with corresponding luminosity functions are in Figures 4.11(a) and 4.11(b) respectively. $\mathcal{B}(t_*)$ is broadly redundant for $0.8 > t_* > 0.3$, the model agrees with the theoretical

[12]Greig & Mesinger (2017b) shows this for redshifts $[6, 20]$ (their Figure 1, using $\alpha_X = -0.5, 1., 2.5$).
[13]Assuming one ensemble of galaxy in reionisation, made entirely from Salpeter IMF stars.



(a) (b)

Figure 4.10: Figure 4.10(a) shows the Bayes factors (Jeffreys' scale in grey) calculated with the SDDR for α_X, with the corresponding variety in 21cm PS plotted below (4.10(b), with telescope noise in grey for the SKA). The colour coding of the 21cm PS is consistent between plots, with the fiducial parameter for the mock data in green, and values of α_X above (below) the fiducial indicated by black (blue). The 21cm PS's in 4.10(b) all fit within our telescope error margin at any value of $\alpha_X \in [-2., 3.]$. Chunks 1 and 3 are similar (not shown). Figure 4.10(a) clearly shows rejection of $\alpha_X < -1$ via the McGreer likelihood. We therefore conclude with the SDDR that α_X is redundant within the range $[-1, 3]$, given our telescope assumptions in Section 4.3.6.



(a) (b)

Figure 4.11: The SDDR results for t_* (Section 4.5.4). 4.11(a) contains the Bayes factors $\mathcal{B}(t_*)$ for the following values of $t_* = [0.0001, 0.01, 0.05, 0.2, 0.4, 0.5, 0.6, 0.8, 1.0]$, with the corresponding UV LF shown in 4.11(b). Colours are consistent between the two plots, with the red dotted UV LFs falling well below the axis on 4.11(a). The corresponding 21cm power spectra cannot be distinguished, lying within the telescope error region (similarly to Figure 4.10(b), not plotted). The UV LF on the other hand, lies within or close to the observational error bars, resulting in (at most) *moderate* results for $0.8 > t_* > 0.3$. For low values of t_* however (red, dotted), the fit is rejected with *strong* Evidence. This motivates adjusting its uniform prior to $t_* \in [0.05, 1]$.

constraints mentioned above. Within this region t_* is incapable of providing further inference given the sensitivity of the data used. Looking closer at the separated posteriors in Figure 4.8, the lower end penalty for $t_* < 0.4$ is from the UV LF (and not from the 21cm PS at least when using $z < 10$). This can be observed in Figure 4.11, since the UV LF shift brighter exponentially as t_* decreases (blue solid, then red dotted). However at the very low t_* (< 0.05, red dotted) we face computational struggles as the likelihoods produced are extremely small.

To conclude, t_* captures an interesting question and has regions of influence. We would rather a seek better/additional data source than removing it completely. We therefore cut the lower end of the parameter prior to remove vastly wrong UV LFs. In conclusion to this section, we advise the adaptation of the t_* prior range from $[0, 1]$ to $[0.05, 1.]$. The details of SF timescales within reionisation galaxies are poorly constrained. We think there is more to be gained from this parameter in pinning down the behaviours of star formation timescales within the EoR galaxy ensemble. More precise modelling as well as other observational synergy is needed to do so.

4.5.5 Resolving the ionising efficiency double power-law in halo mass

The double power-law in halo mass for ζ provides a significant improvement in the fitting flexibility of the EoR model. A good fit of both a X-ray heated 21cm PS and the observed UV LFs is now obtainable.

The variances of each power-law related parameter are $\sim 10\%$ (as a % of the prior range) in D|D, (Tables 4.3 and 4.5). But t_* is no longer constrained (in comparison to the posterior variances for C|C. Within the posterior of C|C the variance is 10%, while for D|D it becomes 28%. This consolidates the final point in Section 4.5.4, we think more observational synergy, or improved statistical analyses of the 21cm signal will remove degeneracy in this parameter.

Variety in brightness temperature intensity is seen clearly, for model D compared to A, B, or C, at the high redshift end of the light-cone slices in Figure 4.1. The power spectra for D (containing the double power law, orange coloured) reflects this in Figures 4.2 and 4.3, by reaching both higher and lower amplitudes in the 21cm PS than any other model.

Although the majority of the flexibility in the 21 cm PS is from the X-ray parameterisation

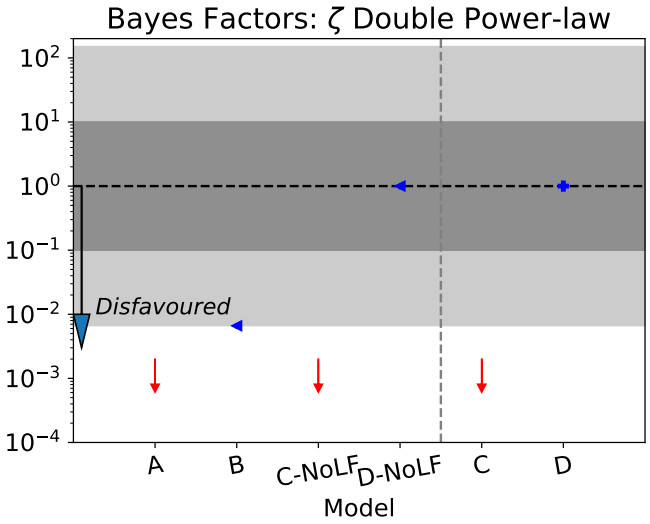
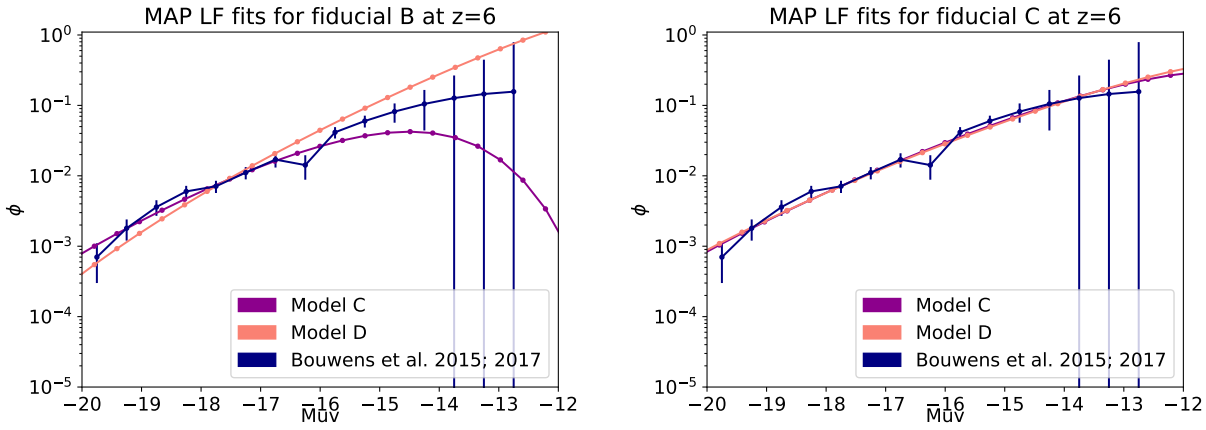


Figure 4.12: Bayes factors for models A, B, C, and D as they recover the mock fiducial data from model D. The conventions of this plot are described in Figure 4.6 and 4.9. Only model D, with the double power law for ζ, is able to reproduce both its own fiducial 21cm PS and the UV LF data (shown in Figures 4.5 and 4.13 respectively). The single power-law in ζ (A & C) is so heavily disfavoured that it is below the axis (red arrow). See Section 4.5.5 for the related discussion.

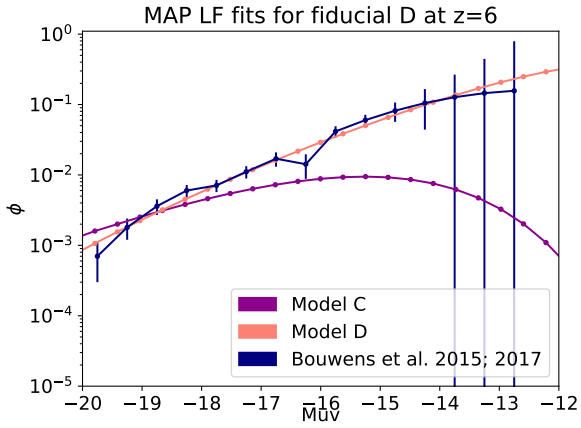
shared with model B, the second power law causes subtle deviation visible in the over-predicting large scale structure of chunk two ($z \sim 9$, small k, Figure 4.5(k)) and under-predicting the middle scales of chunk 1 ($z \sim 8$, Figure 4.5(j)). The Bayes factors calculated without the UV LF reflect this producing *strong* odds of 1:151 in recovering model D in favour of B (Figure 4.12). This is extremely borderline, highlighting the dangers of interpreting results strictly from the Jeffreys' scale. Firstly B contains the α_X which we have shown to be redundant in Section 4.5.4. Secondly, B has a larger prior volume than D due to α_X and its parameterisation of the ionisation efficiency being constant, see Table 4.3). Since B would not be able to fit any redshift dependency on the UV LF (because ζ is constant), we accept the Jeffreys' scale conclusion here.

Figure 4.13 shows the MAP UV LFs produced when C and D are simultaneously constrained by the observed UV LFs and the 21cm PS mock fiducial data (for A, B and C). Model D is capable of fitting the UV LF data regardless of which mock 21cm PS data is chosen, however the fit worsens for the realisation used in the mock 21cm fiducial set B. In contrast model C is heavily disfavoured by poor fits on both the UV LF and 21cm PS when fitting B or D (purple line in Figures 4.13(a) and 4.13(c)). In 4.13(b), both C and D fit the observed UV LFs well. However when model C is also trying to fit for the mock data 21cm PS from D it can no longer



(a)

(b)



(c)

Figure 4.13: UV LFs at redshift 6 produced by models C (purple) and D (orange) for their MAP fits against the mock fiducial data sets from B, C, and D, 4.13(a), 4.13(b) and 4.13(c) respectively. The corresponding 21cm PS are plotted in Figure 4.5, with the MAP parameters used in Table 4.5. Fits for fiducial mock data set from A are similar to 4.13(b), as are redshifts 7, 8, and 10 with their corresponding fiducial mock data set (not shown).

fit the observed UV LF data. Figure 4.13(c) shows this as the purple line is too low by an order of magnitude within $M_{\rm UV} \in [-16, -14]$.

The addition of the second power law in halo mass to the model is not fool proof. 4.13(a) shows the orange line for model D over and underestimates the dim and bright respective ends of the UV LF only in the case of fitting the X-ray heating prescribed in the mock observation created from fiducial set B. $M_{\rm turn}$ and t_* are close to the edges of their prior distributions in Table 4.3 (lower and upper end respectively). Because $M_{\rm turn}$ is bound by observational data, this likely hints that the mock 21cm PS fiducial set in B is an unphysical realisation. But we proceed with an open mind and emphasise that t_*, the star formation timescales within EoR galaxies, needs to be pinned down further.

4.6 Discussion

This work is a continuation of 21CMNEST (Binnie & Pritchard, 2019), applying Bayesian model selection to the epoch of reionisation (EoR). In brief, 21CMNEST is constructed by replacing the sampler, COSMOHAMMER (Akeret et al., 2013), from 21CMMC, with the nested sampling algorithm MULTINEST (Feroz et al., 2009).

21CMNEST uses four astrophysical models of the EoR (labelled A, B, C, and D, Section 4.3). Each is designed with a specific purpose in mind, specifically attempting to answer three questions stated in the bullet points below. To recapitulate the models: A is a simple scenario (Furlanetto et al., 2004; Greig & Mesinger, 2015); B contains an X-ray heating parameterisation, representing $T_{\rm S}$ fluctuations driven mostly by HMXBs (Greig & Mesinger, 2017b); C has no X-ray heating, a single power law in halo mass for the ionising efficiency, and is the simplest model capable of producing UV LFs that fit observational data (see Park et al. (2019), Figure 4.4). Finally, in D the ionising efficiency has a double power-law in halo mass as well as all physics from the previous models.

On top of using realistic EoR models, every model measures the 21cm PS from the light-cone (Greig & Mesinger, 2018) and include inhomogeneous recombinations in the IGM (Sobacchi & Mesinger, 2014). The 21cm observations are performed with 21CMSENSE (Pober, 2016) using *moderate* foreground assumptions for SKA-512's central stations (detailed in Section 4.3.6).

The full list of Bayes factors obtained are shown in Table 4.4. A MAP 21cm PS for each model's fit for each model's mock observational data is shown in Figure 4.5. MAP parameters are shown in Table 4.5.

- Can observations with SKA-512 observe the impact of fluctuations in the 21cm spin temperature from the 21cm power spectrum?

Spin temperature fluctuations in the 21cm brightness temperature power spectrum are distinguishable close to the post heating regime (Section 4.5.2) with the redshift range $z \in [8, 10]$, smaller than has been previously proposed (e.g. Ewall-Wice et al. (2016) suggest $z \in [5, 25]$). The large amplitudes of the 21cm PS in chunk 3 (Figure 4.5(f)) and on a variety of scales in chunk 2 (small and large k, Figure 4.5(e)) lead quickly to *strong* Evidences. 21CMNEST has produced odds of $\text{Log}\mathcal{B} \sim 6000$ in separating models with a saturated spin temperature (A and C) from model B (containing X-ray heating). The results are in agreement with (Greig & Mesinger, 2017b) - a large IGM temperature leads to a very different 21cm PS compared with a more detailed modelling of the heating. Large differences in Evidence values are related to similar changes in the magnitude of maximum likelihood points of posterior distributions. The assumptions we have used show that the SKA-512 will comfortably detect the presence of inhomogeneous heating with a ~ 1000 hour observation.

To improve the robustness of this result, more attention is needed in the precise modelling of our telescope observation. Nasirudin et al. (2020) for example show how telescope noise modelling can move parameter posteriors by $\sim 5\sigma$, subject to some movement upon convergence. Future work might better explore more detailed X-ray heating scenarios by using more detailed instrument simulations. Alternate foreground mitigation methods and various magnitudes of the X-ray heated mock fiducial 21cm PS data will be considered in future work.

In Section 4.5.4 we showed that the α_{X} parameter is redundant within $[-1, 3.]$ for use in Bayesian inference due to its lack of influence on the 21cm PS during the EoR backbone ($z \in [8, 10]$).

- Does the high redshift UV LF provide comparable inference to the EoR 21cm PS in a Bayesian context?

The UV LF likelihood is important for constraining reionisation parameters and does so comparably to the 21cm PS. However more work is necessary to better understand this synergy.

Binnie & Pritchard (2019) found that a halo mass power-law for ζ was indistinguishable from a constant using just 21cm PS data and the observational prior checks (described at the end of Chapter 3). Park et al. (2019) show that inclusion of the UV LF can break this degeneracy. We can confirm this, identifying the power-law parameters (α_* and f_{*10}) to $\sigma = 0.07$ and 0.06 respectively (1 s.f.). Section 4.5.3 discusses this and shows the posteriors obtained by separately using the UV LF and 21cm PS (and observational prior) likelihoods (Figure 4.8). 21CMNEST was used with, without, and with only the UV LF in the likelihood (for models C and D). Like-for-like comparisons of likelihood constructions are shown in the Bayes factor plots (Figures 4.6, 4.9, and 4.12).

As can be seen in Figure 4.7, the Evidence value from the UV LF (compared to the 21cm PS) change depending on the realisation. In the post heating regime, the UV LF inference dominates (there is a 10^3 difference in the Evidences in Figure 4.7(a) - comparing only/without using the UV LF likelihood). However, the increase of 21cm PS amplitude when including the IGM gas heating carries across to the Evidence values and washes out the contribution from the UV LF (Evidence values from the UV LF only likelihood 4.7(b)). This dependency on the mock fiducial 21cm PS makes it unrealistic to quantitatively measure one value for the inference of these likelihoods.

Reiterating from the results of the previous question, the spin temperature model can fit mostly down to the post-heating approximation power spectra (red line in the 1st and 3rd row, Figure 4.5). It follows here that a more intricate analysis of the spin temperature fluctuations is necessary to quantify the inference of the UV LF Evidence (since the magnitude of the χ^2_{21} depends on the telescope noise precision, which depends on the brightness temperature signal magnitude). Despite this intricacy, Figure 4.9 shows *strong* Evidence in telling apart a constant from single-power law ionisation efficiency with the 21cm PS, when spin temperature fluctuations are included (odds $\sim$ 1300 : 1). Section 4.5.3 contains the full model selection discussion.

In Section 4.5.4, we analysed the t_* parameter within model C. It has negligible influence on the 21cm PS, but is powerful in keeping the UV LF from the unphysical ends of its uniform prior range. We recommend a minor adjustment of its prior range to $t_* \in [0.05,\ 1.]$ based on estimates from the Jeans time for a minimum star formation timescale (as well as the computational issues within MULTINEST). It is a useful parameter that provides insight into the star formation time-scales of EoR galaxies. Separate observation will be necessary to pin

down this parameter further.

- Can the 21cm PS and UV LF distinguish a double power-law in halo mass for the ionising efficiency?

Yes - given the model choices and telescope assumptions we have used, however not every parameter's variance is reduced when comparing model D to models C or B. Figure 4.12 shows the double power-law being recovered with *strong* conclusions in all cases with and without the UV LF. The Bayes factors for B|D against D-noLF|D is rejected with odds of $\sim 151:1$, however (as discussed in Section 4.5.5) this result is marginal (e.g. the redundant parameter space from $\alpha_{\rm X}$ remains in Model B but not D). As a point of interest, it appears we can tell apart power-laws from a constant ζ with the 21cmPS only when IGM gas heating is modelled (as opposed to assuming a saturated spin-temperature). If B had the UV LF included (with a constant f_*) it would be heavily penalised by the observed data. This may be resolved by including more light-cone chunks for the 21cm PS likelihood, however precise telescope foreground mitigation is required in order for us to match these chunks with an instrument's frequency resolution - too small and line-of-sight effects may dominate the 21cm signal variation, too large and we are not optimising the inference gained from the 21cm PS likelihood.

The power of including the UV LF in the likelihood becomes apparent in Figure 4.13. The MAP UV LF fits for D|C and D|D show that the single power law in C is unable to fit the UV LF and the mock fiducial 21cm PS of D, respectively Figures 4.13(c) and 4.13(b). Figure 4.13(a) suggests that intricate study of different realisations of X-ray heating parameterisations can influence the model selection process because this will influence the model's fit to the UV LF data. There is more insight to be obtained here from the play off between various realisations of X-ray heating parameters before their influence on the UV LF violates the observational data (due to the mild parameter degeneracies within the 21cm PS). Future work will contain more redshift chunks related to the telescope assumptions, enabling us to pin this down. Our analysis of the galaxies that drive the EoR must include detail about the prevalence of each power-law and break down the ionising efficiency into further, more manageable, constituents.

Park et al. (2019) have shown that the observational synergy between the UV LF and 21cmPS is necessary to constrain the parameters used here. We consolidate this by showing that the power-law parameters in model D's posteriors are constrained with variances of $< 20\%$ (with

respect to the uniform prior range, Table 4.5, Figure D.1(b)). $M_{\rm turn}$ and t_* are significantly less peaked (and have larger variances) in the posteriors for model D than for C where they are tightly constrained (Figures D.1(b) and Figure D.1(c) respectively). Without the UV LF the double power-law parameter variances are broader $\sim 30\%$. However we lose precision (in comparison to model C) on t_* a result. It is likely that more avenues of synergy will be necessary if all the parameters in model D are to be more precisely pinned down.

4.7 Summary

21cmNest has validated the results shown in Greig & Mesinger (2017b) and Park et al. (2019) in a Bayesian model selection setting, using only the backbone of reionisation. This is a reliable test of Bayesian model selection in the context of realistic semi-numerical EoR models. But even for a single ensemble average for the species of EoR galaxy, model selection in the EoR is tricky. We have largely been able to answer the desired questions, although some are set by construction.

To conclude, we specify two main avenues for future work in improving the use of Bayesian model selection in the EoR. The first is improving our telescope modelling capabilities in line with Nasirudin et al. (2020). This would allow redshift chunks to be more accurately associated with the frequency resolution of the interferometers, and in turn capture the 21cmPS more reliably in the likelihood. Bayesian data analysis averaging, as Gillet et al. (2020) have done for combining UV LF observations, should be applied to 21cm experiments - potentially combining the variety in foreground mitigation techniques, maximising the data from each telescope, as well as producing a combined constraint from multiple telescopes. Secondly, further analysis of different X-ray heating scenarios is needed to better understand the influence these parameters have on the UV LF and 21cm PS throughout the EoR. This will provide insight into when the spin temperature will saturate, as well as how important the timing of this event actually is for EoR inference. Some inference may be possible on by the timing of the cosmic dawn, here taken to be fixed at $z = 35$. It may be interesting to better investigate the role of the timing of IGM heating relative to reionisation.

This ends the model selection portion of this thesis, however it is a work in progress. In the next chapter we turn to improving the 21cm power spectrum measurement.

Chapter 5

The Morlet Transform as a Power Spectrum

5.1 Introduction

In Chapter 1 we detailed treatment of 21cm brightness temperature light-cones in the 21CMMC likelihood. To recap, the power spectrum (detailed in Appendix C.4) is implemented on coeval cubes (as in Chapter 3) or on chunks of a light-cone created from stitching together coeval cubes (as in Chapter 4). Previously we needed the power spectra from each chunk to be ergodic within the light-cone - i.e. each chunk needs to be large enough that the power spectrum is a valid representative of the full light-cone, but small enough not to smooth over the signal evolution.

Here we present the Morlet power spectrum, a statistic based on the two point correlation function. We replace the Fourier transform in the power spectrum with a Morlet transform - built up of multiple Fourier bases each wrapped in a non-ergodic wavelet. In Chapters 3 and 4 we have used coeval cubes and light-cone chunks respectively. Greig & Mesinger (2018) show the pitfalls of calculating the power spectrum from the coeval cubes rather then the light-cone. Namely the power spectrum can contain bias of up to 10σ in the desired parameter posterior distributions. We aim to show the pitfalls of analysing that light-cone with Fourier power spectrum (FPS) compared to a Morlet power spectrum (MPS). If possible, we look to remove the light-cone effect from EoR parameter estimation entirely. This is motivated by Trott (2016), who implemented this wavelet transform on a toy wide-bandwidth observation to

show improved performances in 21cm signal estimation.

In this chapter, we describe the problems of the light-cone effect in Section 5.2, before going into the detail of the Morlet transform in Section 5.3. Section 5.4 contains the implementation of the Morlet power spectrum within the 21CMMC framework. Our preliminary results are shown in Section 5.5 and we emphasise that this Chapter is a work in progress. Appendices E.2 and E.1 respectively contain a comparison against real MWA data and a discussion on interpreting the properties of a brightness temperature light-cone from the MPS. We conclude and discuss future directions for the MPS in Section 5.6.

5.2 The light-cone effect

The light-cone effect (LCE) is the measurable difference between reference frames when the finite speed of light is accounted for. This typically arises in the context of comparing simulated and observed power spectra, since the simulation is typically in a coeval cube while the observed universe exists in comoving light-cones. At high redshift, the sizes of coeval boxes are necessarily large and the light-cone effect is prominent. It has been observed in galaxy surveys (Yamamoto et al., 1999), where the galaxy power spectrum from simulations requires a correction due to the non-linear evolution of space. For the 21cm power spectrum, the observed distribution of bubble sizes relates to the collapse fraction, $f_{\rm coll}$ (see Section 1.3.1) and is therefore different from the observed reality because it is evaluated instantaneously. This effect of the finite speed of light imposes a maximum observable bubble size which we can quantify by assuming large scales and early times. We consider a simple case that includes only the photons generated within a region of radius R. With small deviations this gives,

$$\frac{\delta z}{(1+z)} = \frac{\delta_R}{\delta_{\rm crit}(z)} - \left[1 - \sqrt{1 - \frac{\sigma_R^2}{\sigma_{\rm min}^2}}\right], \tag{5.1}$$

where, z, represents the observed redshift that occurs when the universe reaches the collapse fraction on average (different by, δz, to the redshift at which the the mean over-density in the region, δ_R, reaches the critical collapse fraction). All other symbols retain their previous definition. The timing of reionisation is not prominent here, but Equation 5.1 shows that the duration, δz, can dictate an upper limit to this effect. As the radius of bubbles (R) increases

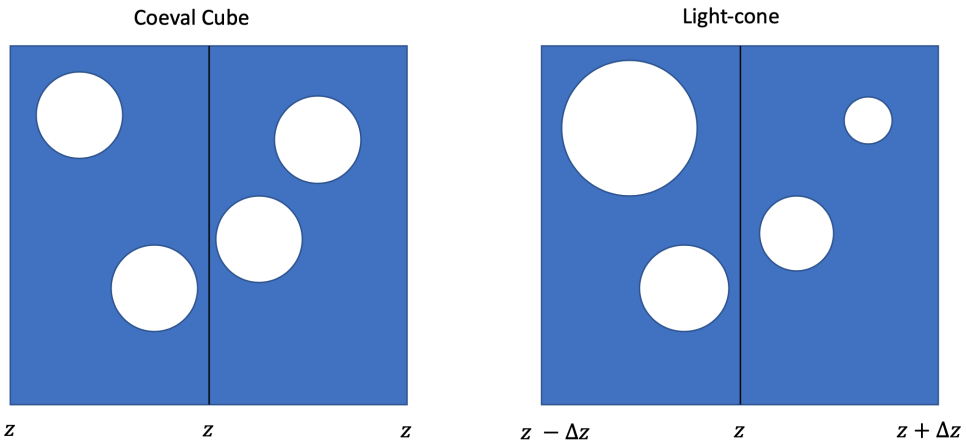


Figure 5.1: A cartoon illustration of the light-cone effect. The coeval cubes have the sizes of near side HII bubbles shrunk and further bubbles exaggerated while the light-cone represents the true evolution of the ionised regions.

the distribution of over-densities diminishes, but the time it takes ionising photons to traverse the bubble increases. We are left with an approximately constant maximum bubble size which is $\approx$ 10cMpc when evaluated with the cosmology defined in Appendix A.1.

During the majority of reionisation, signal from the far end of the chunk will be amplified, and signal from the near end reduced - an imbalance is created in the 21cm power spectrum within the light-cone due to the way it is measured. Figure 5.1 is a cartoon illustrating this effect. Note that larger bubbles will correspond to a lower power for larger k values. In the literature Datta et al. (2012) follow this effect through to the calculation of the 21cm power spectrum showing that using the spherically averaged 1d power spectrum reduces this effect (compared to a 2d power spectrum for example). In an attempt to mitigate this, (Datta et al., 2014) shows that there is an optimal bandwidth for each frequency that can balance the signal bias in the observation. Mondal et al. (2017, 2018) verify this and show that not properly accounting for the light-cone effect can cause a bias in the 21cm PS of up to a factor of 4 at $k \sim 0.1$ Mpc^{-1} for $B \sim 15$ Mhz across $z \in [7.5, 8.5]$.

Current state of the art analyses of the LCE do not include research into foreground and telescope noise effects. Experiments (e.g. LOFAR, HERA and the SKA discussed in Section 1.3.7) apply specific bandwidth sizes (discussed above), but with a taper function to account for the noise. Although this method should be accurate, it reduces the used quantity of raw data by a further $50 - 75\%$. It is likely that including the mitigation of telescope and foreground noise will cause losses on top of this. Splitting the data this way also loses sensitivity to large

scale modes which are essential for pinning down accurate astrophysics (Greig & Mesinger, 2017b). The Fourier analyses only has access to the longest modes that fit within the frequency channels available in the bandwidth, rather than the longest modes that exist. Hence we need new machinery in order to adapt the observational bandwidth along the light-cone at the same time as transforming the signal into a Fourier-like space.

5.3 What is a Morlet transform?

In order to avoid the pitfalls discussed in the previous section we turn to using Morlet wavelets (Goupillaud et al., 1984). Each wavelet is a non-ergodic statistic, the idea being that the Fourier basis adapts to the evolution to the light-cone as it evolves through redshift space. The wavelets are therefore an ergodic representation of the light-cone when combined.

The Morlet wavelet basis function, Υ, is described in terms of a line-of-sight Gaussian packaging, dictated by the scale η,

$$\Upsilon(\nu_i|\eta,\nu_c) = e^{-\eta^2(\nu_i-\nu_c)^2}e^{2\pi i\eta(\nu_i-\nu_c)}, \tag{5.2}$$

where the centre of the wavelet package is associate with a frequency at a given redshift $\nu_{\rm c}$ and η has Fourier space units $\tilde{\eta}$ [Mhz]$^{-1}$. These centres will be positioned evenly through the bandwidth ($B = N\Delta\nu$), for N spectral channels giving,

$$\nu_{\rm z} = [0, \Delta\nu, 2\Delta\nu, ..., (N-1)\Delta\nu]\,, \quad \text{and} \tag{5.3}$$

$$\eta = \left[\frac{1}{B}, \frac{2}{B}, ..., \frac{N}{2B}\right],$$

which is motivated by the Nyquist sampling frequency. For a signal $T(\nu)$ we can implement the Morlet transform as,

$$\Upsilon(\eta,\nu_z) = \frac{\sqrt{|\eta|\Delta\nu}}{\pi^{\frac{1}{4}}}\sum_{i=0}^{N-1} T(\nu_i)\Upsilon(\nu_i|\eta,\eta_c), \tag{5.4}$$

where the normalisation of the wavelet scales with the wavelet's energy content.

In the FPS case we measure brightness temperature signal, $\delta T_b(\nu|u,v)$, averaged across the scales given by the Fourier modes of the interferometer at a given frequency, $k_\perp \propto \sqrt{u^2+v^2}$.

Here we obtain an MPS for each angular mode, $k_\perp$. This disguises the evolution of the light-cone by including the warping of the square relationship of $\nu_z - \eta$ within a non-square $z - k_\parallel$ space,

$$z(\nu_z) = \frac{\nu_{21}}{\nu_z} - 1, \tag{5.5}$$

$$k_\parallel = \eta \frac{2\pi H_0 \nu_{21} \sqrt{\Omega_{\rm M}(1+z)^3 + \Omega_\Lambda}}{c(1+z)^2}. \tag{5.6}$$

Hence the evolution which causes the light-cone effect is contained within the wavelet's Fourier basis. To reiterate, the coordinate basis of the Morlet transform intrinsically caters for the distortion of the light-cone effect. Our treatment of preparing a MPS is otherwise similar to that of the FPS.

To recapitulate, the absolute square of the observed modes is used after being normalised by the bandwidth. The volume of the observed light-cone is,

$$V(z, k_\parallel) = B_\parallel X Y^2 \frac{c^2(1+z)^2}{\nu_{21}^2 A_{\rm eff}} \; h^{-3}{\rm Mpc}^3, \tag{5.7}$$

where the $k_\parallel$ bandwidth ($B_\parallel$) is,

$$B_\parallel(z, k_\parallel) = \frac{\sqrt{2\pi}}{k_\parallel} \; {\rm Mpc}, \tag{5.8}$$

and for completeness the $k_\perp$ bandwidth ($B_\perp$) is,

$$B_\perp(z, \eta) = \frac{\sqrt{2\pi}}{\eta} \; {\rm MHz}. \tag{5.9}$$

Finally we have the angular average of the Morlet Power spectrum (MPS) as,

$$P_{\rm MPS}(z, k_\parallel | k_\perp) = \frac{1}{V(k_\parallel, z)} |\Upsilon|^2 \; {\rm mK}^2 h^{-3} {\rm Mpc}^3. \tag{5.10}$$

The output of each transform is described in terms of an amplitude, and position in the wavelet's Fourier space. We therefore have an array of these separated by a distance in the light-cone, i.e. between each mean wave-packet. Figure 5.2 shows a toy construction of the FPS and MPS in tandem for comparison.

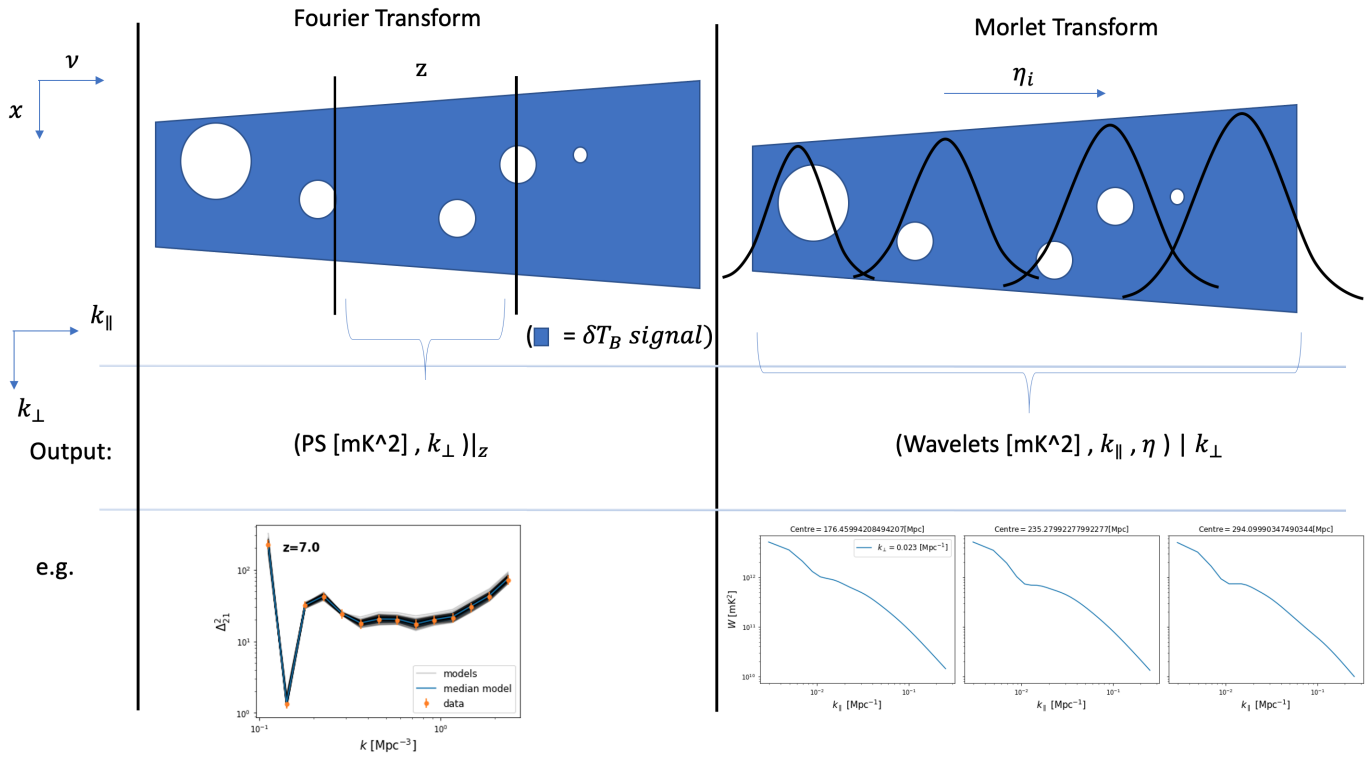


Figure 5.2: A toy comparison of treating the light-cone with the Fourier power spectrum (left) and the Morlet power spectrum (right). The MPS $k_\perp$ modes are averaged but the $k_\parallel$ are not, unlike the FPS in which a 3d angular average is performed across all k modes in each chunk of light-cone. The line-of-sight Gaussian packages (that are the Morlet wavelets) have their widths defined using $k_\parallel$ in Equation 5.6. After performing the FPS we are left with a power per averaged k-mode at a given redshift. Using the MPS provides a power per ($k_\parallel$, $k_\perp$) at various different locations within the light-cone, defined by η. If the MPS Gaussian envelope was replaced by a top-hat function and η was chosen to provide no overlap, the method would be the same as the FPS. Appendix E.1 shows a comparison of these on simulated 21cm brightness temperature light-cones.

5.4 Application within MCMC

As discussed in Section 1.4.2, we take the form of the likelihood to be Gaussian,

$$\ln\mathcal{L} = -\frac{1}{2}\left[\log 2\pi + \log|\mathcal{C}| + (\mathbf{x}-\mu)^T\mathcal{C}^{-1}(\mathbf{x}-\mu)\right], \tag{5.11}$$

however we have not had to worry about parameter dependence within the covariance, $\mathcal{C}$, until now. This term can be split up into three component covariance matrices,

$$\mathcal{C} = \mathcal{C}_{21} + \mathcal{C}_{\rm FG} + \mathcal{C}_{\rm TN}, \tag{5.12}$$

where the subscripts 21, FG, and TN represent the sample light-cone, foreground signal, and the noise from the telescope respectively. We are assuming for now that we are in the regime where foregrounds have been successfully cleaned to the wedge (including a buffer) and therefore we may omit $\mathcal{C}_{\rm FG}$ from Equation 5.12.

In earlier chapters (with the FPS), we have been able to add thermal noise and sample variance into the likelihood assuming each 3D k-mode is independent. Here it is not quite that simple, since each neighbouring Morlet wavelet will share some of the modes in the adjacent packet. We can assume that 2D modes are independent here (with the same assumption), but there is correlation between the line of sight modes. Assuming the foreground are accurately cleaned we can write,

$$\mathcal{C} = \left[\sigma^2_{\rm TN}\delta_{ij} + \mathcal{C}_{21}\right] e^{-\left(\frac{k_\parallel(x_{c,i}-x_{c,j})}{2}\right)^2}, \tag{5.13}$$

where x_c are distance representations of ν_c and $\sigma_{\rm TN}$ is the 21cmSense output for example. We have a block diagonal covariance form when applying this to the light-cone - diagonal terms are from thermal noise and each block depends on the overlap in neighbouring wavelet distributions. For analytic expressions of the full data covariance of toy signal models please see Trott (2016).

This calculation of the covariance is the only difference to the MCMC framework used in previous chapters. The following subsections detail this via calculation of the cosmic variance and thermal noise before we justify it with simulations in Section 5.5.2.

5.4.1 Obtaining the MPS

We are working in the flat-sky approximation and with the delay spectrum (discussed in Appendix B). To briefly reiterate, this means that we can obtain an unbiased estimate of the 21cm FPS simply by Fourier transforming the visibility measurements (and their temporal information) defined to be the cross product of temperature fluctuations as,

$$\left\langle \tilde{V}_i \tilde{V}_j^*(\mathbf{k}) \right\rangle = \frac{1}{\kappa^2} \int \frac{d^3\mathbf{k}'}{(2\pi)^3} \frac{d^3\mathbf{k}''}{(2\pi)^3} \left\langle \widetilde{T}^i(\mathbf{k}') \widetilde{T^{*j}}(\mathbf{k}'') \right\rangle \widetilde{\mathcal{K}}(\mathbf{k}-\mathbf{k}') \widetilde{\mathcal{K}}(\mathbf{k}-\mathbf{k}''). \tag{5.14}$$

where a measured visibility is defined,

$$\tilde{V}(\mathbf{k}) = \frac{1}{\kappa} \int \frac{d^3\mathbf{k}'}{(2\pi)^3} \widetilde{T}(\mathbf{k}') \widetilde{\mathcal{K}}(\mathbf{k}-\mathbf{k}'). \tag{5.15}$$

providing us with,

$$P_{21}(\mathbf{k}) = \frac{\kappa^2}{\int d^3\mathbf{r} \mathcal{K}^2(\mathbf{r})} \tilde{V}_i \tilde{V}_j^*. \tag{5.16}$$

Since we have assumed the foreground temperature contribution to the signal are successfully cleaned, the signal is therefore the sum of the telescope noise and brightness temperature: $T = \delta T_b + T_{\rm N}$. The constant, κ, represents the relationship $T \propto I_\nu/\nu^2$, providing units of [mk sr/Jy]. $\mathcal{K}$ is the convolution kernel,

$$\mathcal{K}(\mathbf{r}) = \left(X_\nu^2 Y_\nu\right)^{-1} \nu^2 A_\nu(\mathbf{r}_\perp) \phi(\nu), \tag{5.17}$$

dependent on the redshift to cosmological angular distance conversion, X^2Y and the effective aperture, A_ν (both discussed in Section 1.3.7). The taper function $\phi(\nu)$, is typically a top-hat function (returning to the chunk divisions of light-cone given a user defined bandwidth, as discussed in Section 5.2). Here it will be Gaussian with a width that varies depending on its location along light-cone's line of sight. For reference the auto-correlation of the 21cm brightness temperature dual defines the 21cm FPS as,

$$\langle \tilde{\delta T}_b \tilde{\delta T}_b'^* \rangle = (2\pi)^3 \delta(\mathbf{k}-\mathbf{k}') P_{21}(\mathbf{k}). \tag{5.18}$$

Equation 5.16 is only true if T is homogeneous (i.e. the expectation of its mean is constant with respect to spatial variation). We note that this is not strictly true of the $\delta T_{\rm b}$ field which

varies with frequency/redshift and location in the sky (and therefore so does this field's expected mean), however this is precisely the non-ergodicity which we are trying to solve. The wavelet-based estimator formalism if applied here will contain an inconsistency when using this covariance formalism in the case of the MPS. We assume this inconsistency is a second-order effect and proceed without it but we note that this calculation can only be applied to the 21cm temperature field and not to the noise temperature because of this assumption.

5.4.2 Estimate of the sample covariance

To detail the covariance we must expand the signal temperature contributions in full. Taking $\widetilde{T}_i' = \widetilde{\delta T}_{\rm b}' + \widetilde{T}_{N,i}'$ and $\widetilde{T}_j'' = \widetilde{\delta T}_{\rm b}'' + \widetilde{T}_{N,j}''$, we have,

$$\left\langle \widetilde{T}_i' \widetilde{T}_j''^* \right\rangle = \left\langle \widetilde{\delta T}_{\rm b}' \widetilde{\delta T}_{\rm b}''^* \right\rangle + \left\langle \widetilde{\delta T}_{\rm b}' \widetilde{T}_{N,j}''^* \right\rangle + \left\langle \widetilde{T}_{N,i}' \widetilde{\delta T}_{\rm b}''^* \right\rangle + \left\langle \widetilde{T}_{N,i}' \widetilde{T}_{N,j}''^* \right\rangle , \tag{5.19}$$

where the subscripts i and j represent antennae. All terms containing $\widetilde{T}_{N,i}$ are assumed negligible (as discussed in the previous section). The linearity of Fourier transforms, $\tilde{V}_i = \tilde{V}_{21} + \tilde{N}_i$, allows us to write,

$$\begin{aligned} \mathcal{C}\left[\hat{P}(\mathbf{k})\hat{P}(\mathbf{k}')\right] = \Phi^2 \mathcal{C}\Bigg[& \left(\left| \tilde{V}_{\delta T_{\rm b}} \right|^2 + \tilde{V}_{\delta T_{\rm b}} \tilde{V}_{{\rm N},j}^* + \tilde{V}_{\delta T_{\rm b}}^* \tilde{V}_{{\rm N},i} + \tilde{V}_{{\rm N},i} \tilde{V}_{{\rm N},j}^* \right) \\ & \times \left(\left| \tilde{V}_{\delta T_{\rm b}}' \right|^2 + \tilde{V}_{\delta T_{\rm b}}'^* \tilde{V}_{{\rm N},j}' + \tilde{V}_{\delta T_{\rm b}}' \tilde{V}_{{\rm N},i}'^* + \tilde{V}_{{\rm N},i}'^* \tilde{V}_{{\rm N},j}' \right) \Bigg] , \end{aligned} \tag{5.20}$$

where Φ is defined as, $\kappa^2 / \int d^3\mathbf{r} \mathcal{K}^2(\mathbf{r})$, from Equation 5.16. Since noise terms produce expected values of zero we obtain,

$$\mathcal{C}\left[\hat{P}(\mathbf{k})\hat{P}(\mathbf{k}')\right] = \Phi^2 \mathcal{C}\left\{ |\tilde{V}_{\delta T_{\rm b}}|^2 |\tilde{V}_{\delta T_{\rm b}}'|^2 + 2\,{\rm Re}\left[\langle \tilde{V}_{\delta T_{\rm b}} \tilde{V}_{\delta T_{\rm b}}'^* \rangle \langle \tilde{V}_{\rm N} \tilde{V}_{\rm N}'^* \rangle \right] + \langle \tilde{V}_{\rm N} \tilde{V}_{\rm N}'^* \rangle^2 \right\} . \tag{5.21}$$

This is the 21-cm signal covariance, regardless of the statistical analysis method. The application of the covariance within the FPS and MPS are detailed in the following subsections in reference to Equation 5.21.

Covariance for the FPS

We look at the covariance in detail here because conventionally cross-correlation terms are ignored in its calculation, however these are needed for accurate interpretation of the MPS in the Section 5.4.2.

First we deal with the signal covariance. Making use of Equation 5.16 the first term in Equation 5.21 becomes,

$$\Phi^2\mathcal{C}\left[|\tilde{V}_{\delta T_\mathrm{b}}|^2|\tilde{V}'_{\delta T_\mathrm{b}}|^2\right] = \Phi^2\langle|\tilde{V}_{\delta T_\mathrm{b}}|^2|\tilde{V}'_{\delta T_\mathrm{b}}|^2\rangle - P_{21}P'_{21}. \tag{5.22}$$

Fully expanding the remaining visibilities gives,

$$\langle|\tilde{V}_{\delta T_\mathrm{b}}|^2|\tilde{V}'_{\delta T_\mathrm{b}}|^2\rangle = \frac{1}{\kappa^4}\int d^3\mathbf{x}_1 d^3\mathbf{x}_2 d^3\mathbf{x}_3 d^3\mathbf{x}_4 \Upsilon(\mathbf{x}_1)\Upsilon(\mathbf{x}_2)\Upsilon(\mathbf{x}_3)\Upsilon(\mathbf{x}_4)$$

$$\times\langle T_1T_2T_3T_4\rangle e^{-i[\mathbf{k}_1\cdot(\mathbf{x}_1-\mathbf{x}_2)+\mathbf{k}'\cdot(\mathbf{x}_3-\mathbf{x}_4)]}, \tag{5.23}$$

where the antenna temperatures are labelled 1-4 for simplicity. From here, we will assume that the temperature field is Gaussian (up until now the treatment has been general). We note that the temperature field is definitely not Gaussian (Watkinson et al., 2019), however we expect that the Gaussian approximation is a good approximation as we are interested in only the covariance and the power spectrum. We can now apply Isserlis' theorm[1]: $\langle T_1T_2T_3T_4\rangle = \langle T_1T_2\rangle\langle T_3T_4\rangle + \langle T_1T_3\rangle\langle T_2T_4\rangle + \langle T_1T_4\rangle\langle T_2T_3\rangle$; knowledge that the PS must be symmetric ($P(-k) = P(k)$); as well as a homogeneous temperature field to obtain,

$$\langle|\tilde{V}_{\delta T_\mathrm{b}}|^2|\tilde{V}'_{\delta T_\mathrm{b}}|^2\rangle = \frac{P(\mathbf{k})P(\mathbf{k}')}{\kappa^4}\left[\frac{\kappa^4}{\Phi^2} + \left|\int\frac{d^3\mathbf{k}''}{(2\pi)^3}\widetilde{\Upsilon}(\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\widetilde{\Upsilon}(\mathbf{k}'')\right|^2\right.$$

$$\left. + \left|\int\frac{d^3\mathbf{k}''}{(2\pi)^3}\widetilde{\Upsilon}(\mathbf{k}+\mathbf{k}'-\mathbf{k}'')\widetilde{\Upsilon}(\mathbf{k}'')\right|^2\right]. \tag{5.24}$$

For the purposes of simplicity, we only consider correlations between $\mathbf{k}$ and $\mathbf{k}'$, which is valid for large values of $|\mathbf{k}|$. This implies a symmetry between $\mathbf{k}'$ and $\mathbf{k}''$ leaving the cosmic covariance as,

$$\Phi^2\mathcal{C}\left[|\tilde{V}_{\delta T_\mathrm{b}}|^2|\tilde{V}'_{\delta T_\mathrm{b}}|^2\right] \approx P(\mathbf{k})P(\mathbf{k}')\frac{\left|\int\frac{d^3\mathbf{k}''}{(2\pi)^3}\widetilde{\Upsilon}(\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\widetilde{\Upsilon}(\mathbf{k}'')\right|^2}{\left|\int d^3\mathbf{x}\Upsilon^2(\mathbf{x})\right|^2}. \tag{5.25}$$

[1]A zero mean Gaussian vector field can be expressed as the sum of the contributions to its covariance only. This is known as Wick's theorem in the context of quantum mechanics.

The numerator is a sharp peak at $\mathbf{k} = \mathbf{k}'$, and therefore contains very little correlation between distant k-modes in this limit. Applying this symmetry gives,

$$\Phi^2 \mathrm{Var}\left(|\tilde{V}_{\delta T_\mathrm{b}}|^2\right) \approx P_{21}^2(\mathbf{k}), \tag{5.26}$$

where the approximation is due to the assumption that the temperature field is Gaussian.

We note that the noise statistics belong to visibility space, but here we approximate the noise to be independent per baseline and per frequency. This justifies our assumption that the noise is a zero-mean complex Gaussian variable. The noise covariance is then defined from,

$$T_\mathrm{sys} = T_\mathrm{sky} + T_\mathrm{rec} = V_\mathrm{N,rms} \frac{\kappa}{\nu^2 \Omega_p(\nu)} \sqrt{B t_\mathrm{obs} N_\mathrm{pol}}, \tag{5.27}$$

with B the full bandwidth of the observation, t_obs the duration of the observation, Ω_p the solid angle of the effective telescope aperture, $V_\mathrm{N,rms}$ is the root-mean-square of the visibility voltage and N_pol the number of polarisation states recorded by the correlator. To ensure the frequency dependence of the visibility potential during the MPS is not confused with the FPS frequency dependence, γ_ν defines the conversion factor for running the Υ_ν kernel through the visibility space. This means our system temperature in the MPS case is subtly different, but both are represented as: $T_\mathrm{sys} = V_\mathrm{N,rms} \gamma_\nu \sqrt{B t_\mathrm{obs} N_\mathrm{pol}}$. Since, in practice, observations deal with discrete bins we write the discrete Fourier transform,

$$\langle \tilde{V}_N(\mathbf{u}, \eta) \tilde{V}_N^*(\mathbf{u}, \eta') \rangle = (\Delta\nu)^2 \sum_{j,k}^{N_f} \phi(\nu_j)\phi(\nu_k) \langle V_N(\mathbf{u}, \nu_j) V_N^*(\mathbf{u}, \nu_k') \rangle e^{-2\pi i(\eta\nu_j - \eta'\nu_k)}$$

$$= \frac{\Delta\nu}{N t_\mathrm{obs} N_\mathrm{pol}} \sum_{j}^{N_f} \phi^2(\nu_j) \frac{T_\mathrm{sys}^2}{\gamma_\nu^2} e^{-2\pi i \nu_j (\eta - \eta')}, \tag{5.28}$$

For use in the Morlet transform the number of filters, N_f is defined via Equation 5.3. T_sys is assumed independent of frequency allowing us to write,

$$\langle \tilde{V}_N(\mathbf{u}, \eta) \tilde{V}_N^*(\mathbf{u}, \eta') \rangle = \frac{T_\mathrm{sys}^2 \Delta\nu}{N t_\mathrm{obs} N_\mathrm{pol}} \sum_{j}^{N} \frac{\phi^2(\nu_j)}{\gamma_\nu^2} e^{-2\pi i \nu_j (\eta - \eta')}. \tag{5.29}$$

This shows that the visibilities are obtained via the Fourier transform of the taper function in each observational frequency band (or Fourier transform of Morlet kernel between the Gaussian

centres, $\eta - \eta'$ given the assumptions made in this chapter). We consider only the variance and we consider all functions of frequency to be much broader than the taper, ϕ, and therefore can be evaluated at the central frequency of the taper, ν_c, giving,

$$\Phi\langle\tilde{V}_N(\mathbf{u},\eta)\tilde{V}_N^*(\mathbf{u},\eta)\rangle \approx X^2Y\frac{T_{\rm sys}^2}{Nt_{\rm obs}N_{\rm pol}}\frac{\Omega_{\rm p}^2}{\Omega_{\rm pp}}\ \ {\rm mK^2Mpc^3}. \tag{5.30}$$

Where $\Omega_{\rm pp}$ and $\Omega_{\rm p}$ are the beam squared power and beam power squared respectively integrated from the antenna response. Since band width information remains in the form of this equation (compared to Pober et al. (2014) as shown in Section 1.3.7), this cancels with the form of the system temperature providing a result that is only depend on the taper function $\phi(\nu)$.

For the cross term we need only calculate the dual of visibilities using the Morlet wavelet kernel,

$$\begin{aligned}\langle\tilde{V}_{\delta T_{\rm b}}\tilde{V}'^*_{\delta T_{\rm b}}\rangle &= \frac{1}{\kappa^2}\int d^3\mathbf{x}_1 d^3\mathbf{x}_2\Upsilon(\mathbf{x}_1)\Upsilon(\mathbf{x}_2)\langle T_1T_2\rangle e^{-i(\mathbf{k}\cdot\mathbf{x}_1-\mathbf{k}'\cdot\mathbf{x}_2)}\\ &= \frac{1}{\kappa^2}\int d^3\mathbf{k}''\widetilde{\Upsilon}(\mathbf{k}-\mathbf{k}'')\widetilde{\Upsilon}(\mathbf{k}'-\mathbf{k}'')P(\mathbf{k}'').\end{aligned} \tag{5.31}$$

This takes advantage of the definition of the power spectrum as the Fourier transform of the correlation function as discussed in Appendix C.4.2. However Equation 5.31 is only applicable to the temperature domain and not the noise domain because we have assumed the temperature field to be homogeneous in Equation 5.24, which is not strictly true as the 21cm signal itself evolves with frequency as well.

Finally these are combined (along with the variance from Equation 5.26) as,

$${\rm Var}(\hat{P}) = \left[P_{21}(\mathbf{k}) + \frac{(T_{\rm sys})^2}{t_{\rm obs}N_{\rm pol}}\frac{\int d\nu\ \Omega_p^2(\nu)\nu^4\phi^2(\nu)}{\int d^3\mathbf{x}\ \Upsilon^2(\mathbf{x})}\right]^2. \tag{5.32}$$

X and Y are assumed to be frequency independent across each channel (or wavelet) allowing us to write,

$${\rm Var}(\hat{P}) = \left[P_{21}(\mathbf{k}) + X^2Y\frac{T_{\rm sys}^2}{t_{\rm obs}N_{\rm pol}}\frac{\Omega_p^2}{\Omega_{\rm pp}}\right]^2. \tag{5.33}$$

This allows us to match the form in Section 1.3.7 and the literature (Parsons et al., 2012; Pober et al., 2014). Estimates of the effective beam, $\Omega_{\rm eff}$ are given by $\Omega_{\rm p}^2/\Omega_{\rm pp}$ and are typically ~ 0.5 for an interferometer (Parsons et al., 2014).

Covariance for the MPS

As mentioned earlier, we now apply a similar methodology to the above section, but with the Morlet kernel (Equation 5.10) as the taper function (Section 5.4.1). This causes each of the terms in Equation 5.21 to include a covariance between wavelets, hence evaluation of this equation is different between the MPS and FPS. The resulting form of the signal covariance is block-diagonal, as shown in Section 5.5.2.

Since we are aiming to reproduce ergodic and isotropic power within each wavelet, we assume these to be true. Note that each individual wavelet will be only roughly ergodic and should combine to be exactly ergodic. If the sample is sufficiently large, we are able to interchange the wavelet ensemble average and signal spatial average. Therefore cosmic variance is catered for by assuming each $(k_\perp,\ k_\parallel)$ vector to be an independent realisation. If this assumption causes a bias in the estimation of the MPS, then more k modes are needed in the observation to ensure their distribution does not introduce discretisation effects.

Instrumental noise follows analogously to the previous section as,

$$\Phi\langle\tilde{V}_N(\mathbf{k},\nu_c)\tilde{V}_N^*(\mathbf{k},\nu_c')\rangle = \Phi(\Delta\nu)^2\sum_j^{N_f}\phi(\nu_j)\phi'(\nu_j)\langle V_N(\mathbf{k})\rangle V_N^*(\mathbf{k}) \tag{5.34}$$

$$\approx \frac{X^2YT_{\rm sys}^2\Omega_p^2}{N_f\Omega_{\rm pp}t_{\rm obs}N_{\rm pol}}\int_{B_\nu}d\nu\phi(\nu)\phi'(\nu) \approx \frac{X^2YT_{\rm sys}^2\Omega_{\rm eff}}{N_ft_{\rm obs}N_{\rm pol}}e^{-k_\parallel^2(x_c-x_c')^2/4}$$

$$= P_Ne^{-k_\parallel^2(x_c-x_c')^2/4},$$

note that N_f has dropped out of the integral unlike Equation 5.28. This is in anticipation that B is instead $\sim\Delta\eta$, i.e. the range between the wavelets.

Combining the noise and cosmic variance gives,

$$\mathcal{C}\left[\Upsilon,\Upsilon'\right] = \left\{\sqrt{\mathcal{C}\left[P_{21},P_{21}'\right]} + P_Ne^{-k_\parallel^2(x_c-x_c')^2/4}\right\}^2, \tag{5.35}$$

which we implemented throughout the rest of this Chapter. Note that P_{21} in this section is as before (Equation 5.16) but with ϕ as the Morlet kernel in Equation 5.17.

5.5 Results

In this Section we discuss the results obtained so far. Section 5.5.1 shows a direct comparison between MCMC with the FPS and the MPS on toy models before discussing its application to more realistic cases. We also check the assumptions used for Equation 5.35 in Section 5.5.2.

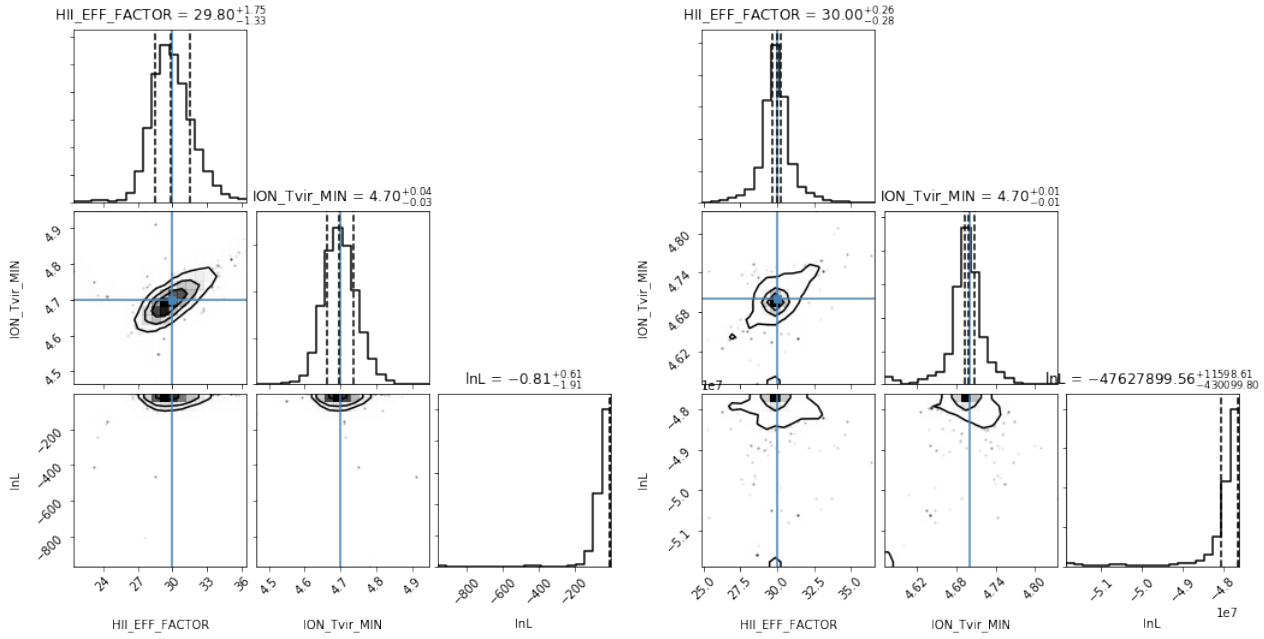
5.5.1 Comparison of the MPS and FPS

In these analyses, Model A from Chapter 4 is used for the 21cm light-cones throughout. To recap, this is a simple scenario with two parameters ($\zeta \in [5., 250]$, $\text{Log}_{10}[T_{\text{vir}}] \in [4, 6]$) as well as inhomogeneous recombinations. Both parameters use uniform priors across the ranges (as before).

We performed the MCMC with likelihoods using the MPS and FPS within $z \in [8, 10]$ on 250×250 Mpc^2 boxes with 128 pixels per box. Both MCMC runs were implemented with 8 walkers calculating 8000 samples. The resulting posteriors are shown in Figure 5.3 with the original 21CMMC on the left and 21CMMC with wavelets on the right. We cannot directly compare the values of the likelihood from the FPS and the MPS since they have fundamentally different methodologies (via the differing kernel in Equation 5.31). But what is important is the shape of the posterior distributions obtained through the likelihood. Using the MPS obtained a reduction in posterior variance by a factor of 5 for both ζ and $\text{Log}_{10}[T_{\text{vir}}]$ with changes of $\sigma^2_\zeta = 0.05 \to 0.01$ and $\sigma^2_{\text{T}} = 0.01 \to 0.002$.

Attempting to calculate the MPS on larger light-cones has proven difficult. In order to reproduce the posteriors used in Greig & Mesinger (2018) we need to use a light-cone that ranges $z \in [6, 25]$ and is of size 600×600 Mpc^2. Preliminary tests on light-cones of this size caused prompt machine crashes, with requests of RAM in the terabytes.

Interpretation of the MPS has proven more difficult in comparison to the power spectrum, but there is definitely promise for its use in the future. This is discussed in Appendix E.1 using Model A with a variety of parameterisations for sample models of reionisation.

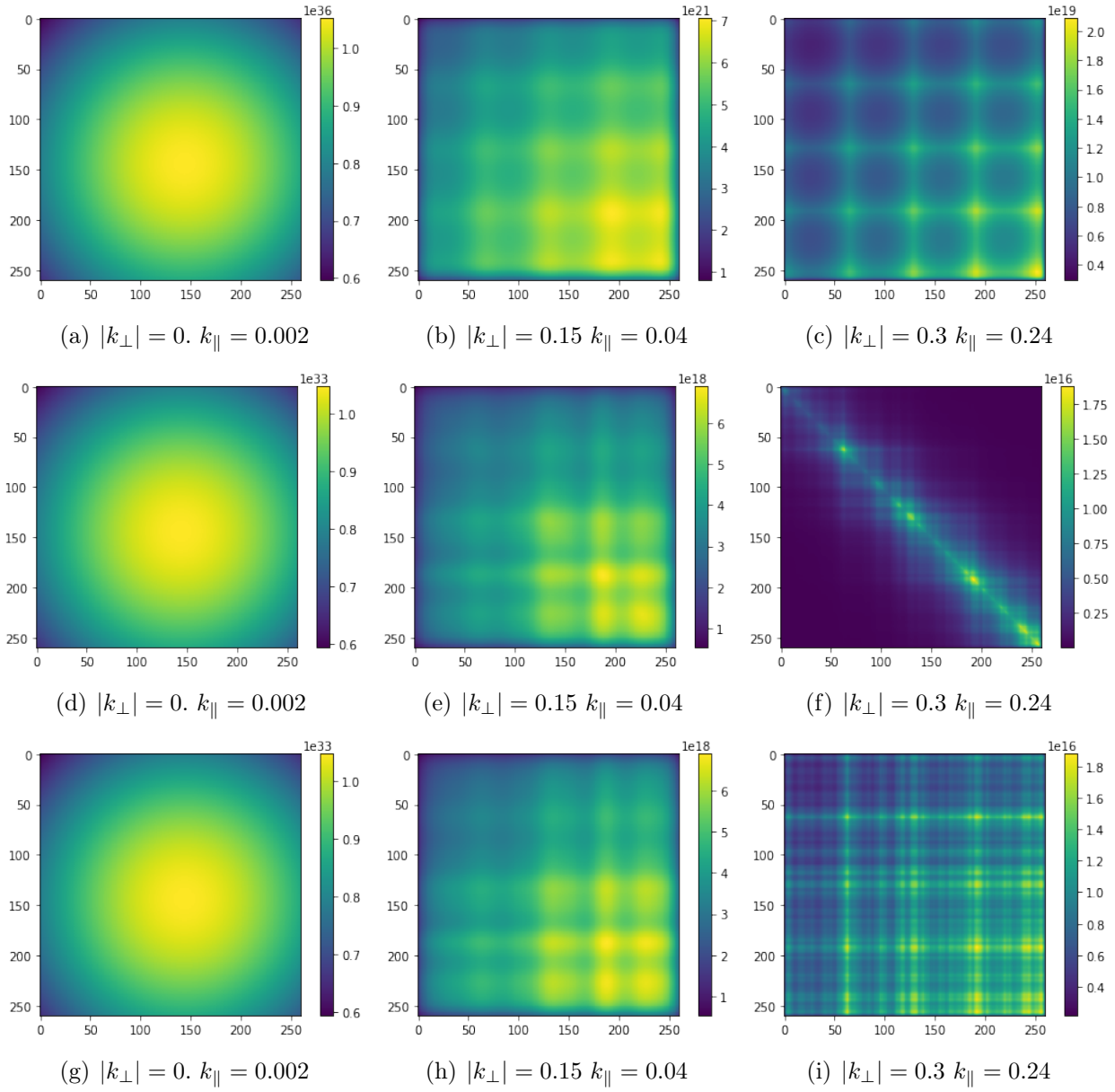


(a) FPS (b) MPS

Figure 5.3: Posteriors from using simple two parameter test model with reduced redshift range for testing. 5.3(a) and 5.3(b) show the obtained posteriors when using the FPS (as in 21CMMC) and MPS respectively. Notice the MPS produces tighter posteriors than those obtained by the FPS.

5.5.2 *Cross-check*: covariance

Writing down the form of the covariance for the Morlet wavelets on a light-cone in Section 5.4.2 has been a lengthy process. In particular, we have had to make the assumption that the temperature field is homogeneous within each wavelet, which is not strictly true. In this Section, we attempt to check the form of the streamlined MPS covariance estimate in Equation 5.35 to justify its use in the MCMC via Equation 5.11. We calculate 1000 light-cones with differing random seeds (1-1000 inclusive) and measure the covariance between the Morlet wavelets from these light-cones directly from the line-of-sight pixels for every value of $|k_\perp|$ and $k_\parallel$. Figure 5.4 shows a comparison of the measured covariance and the estimate we have produced in rows one and two respectively. The third row implements Equation 5.35 with, $k_\parallel = 0$, so as to show the behaviour of the covariance of a single light-cone for a Morlet wavelet without the Gaussian envelope. The $k_\perp$ and $k_\parallel$ modes are consistent across each column and are representative of the total behaviour seen within the larger set. The majority of the matrices look similar to the first two columns, with a good qualitative agreement between all three of the methods. The final column is an example of when the single light-cone method (in the third row) is no longer representative of the measured covariance form multiple light-cones (in the first row). In this



(a) $|k_\perp| = 0.$ $k_\parallel = 0.002$ (b) $|k_\perp| = 0.15$ $k_\parallel = 0.04$ (c) $|k_\perp| = 0.3$ $k_\parallel = 0.24$

(d) $|k_\perp| = 0.$ $k_\parallel = 0.002$ (e) $|k_\perp| = 0.15$ $k_\parallel = 0.04$ (f) $|k_\perp| = 0.3$ $k_\parallel = 0.24$

(g) $|k_\perp| = 0.$ $k_\parallel = 0.002$ (h) $|k_\perp| = 0.15$ $k_\parallel = 0.04$ (i) $|k_\perp| = 0.3$ $k_\parallel = 0.24$

Figure 5.4: Each row calculates the covariance along the light-cone for three differing methodologies. The first row calculates the covariance of the MPS from 1000 light-cones with differing randomly seeded density fields - this covariance has been measured from the Morlet wavelet of each light-cone. The second row utilised Equation 5.35, estimating the covariance between the Morlet wavelets. The third row uses $k_\parallel = 0$ for every implementation of Equation 5.35, regardless of the wavelet. The columns dictate different selected values of $(|k_\perp|, k_\parallel)$, the majority of matrices show qualitative agreement in between all three rows, similar to those shown in the first two columns. See text for discussion.

case the Gaussian envelope visibly suppresses the covariance off the diagonal of the matrix, removing the adverse effect. When the resolution of the box and the limitations of sample variance (from using one random seed) begin to impair the covariance of the Morlet wavelets, the MPS suppresses the covariance. This is seen by the apparent diagonal shape in Figure 5.4(f).

In the majority of the estimated covariances matrices, including the Gaussian envelope in Equation 5.35 produces smoother version estimates of the covariance, giving closer agreement to the measured covariance (comparing the second and third rows with the first row in Figure 5.4). Neither approximation is as smooth as calculating the covariance directly however the direct calculation is too cumbersome for statistical use. Correlation between the light-cone coordinates changes depending on the $k_{\parallel}$ used but Equation 5.35 is a qualitatively good comparison below $k_{\parallel} \sim 0.20$ Mpc^{-1}. We notice that the covariance estimate is significantly worse at large $k_{\parallel}$ (with the largest shown in the third column of Figure 5.4). We emphasise that the MCMC has performed well in the previous sub-section, even with the differences in the large $k_{\parallel}$ covariance. More work is needed to be done on larger light-cones to see if this effect persists.

The cross/checked pattern consistent across all covariance figures appears due to the stitching of the light-cone's constituent coeval cubes, as described in Section 1.3.4 (Greig & Mesinger, 2018).

5.6 Discussion

We detail the light-cone effect in Section 5.2 and present a solution in form of the Morlet transform in Section 5.3. We then look at applying the Morlet power spectrum to MCMC analyses in Section 5.4, paying particular detail to treatment of the covariance in Section 5.4.2, before proceeding to the results in Section 5.5, which we emphasise is a work in progress.

In Section 5.5.1 we showed that the application of the MPS in place of the FPS in the 21CMMC likelihood as prescribed in Section 5.4 reduces the variance on the parameter posteriors, improving the constraints on the toy reionisation model we have used. We have also cross checked our results in Section 5.5.2 by showing that our approximations surrounding the covariance hold. Appendix E.1 shows that the MPS can provide information on the neutral fraction evolution

as well as the brightness temperature in a similar way to the FPS. However we encountered computational issues when trying to include a fuller analysis since the MPS requires processing of the entire light-cone in one go - a light-cone of similar size to those calculated in Greig & Mesinger (2018) was unfeasible.

To try to remedy the computational limits encountered in Section 5.5.1 when calculating the MPS on larger, realistic, light-cones we introduced an integer 'stride' parameter. Rather than calculating a Morlet wavelet for every line of sight light-cone pixel, we calculate the wavelet at every stride value, i.e. stride=1 represents the MPS we have used, stride = 2 every other pixel and so on. Since there is correlation between the wavelets we initially thought that this would not be a large effect whilst reducing the memory requirements as a function of 1/stride. It changes the MPS posteriors obtained in Figure 5.3 negligibly below a value of 5 before a small increase in the variance and a shift in the MAP parameter occurs. Above stride values of 2, the MPS no longer represents the detail in light-cone as in Appendix E.1. We realise that having a constant stride is somewhat naive since the covariance between wavelets will vary along the line of sight, the wavelets are designed to do this. Future work consists of trying a logarithmic spacing to reduce the memory, while still capturing information about the light-cone quantities we wish to probe. Particularly for the far end of the light-cone the Gaussian wavelet envelope will be steep (the envelopes will have a smaller variance than due to the larger values of η) and sampling will need to be every pixel. Neighbouring wavelets at low η should share significant information, so we should be able to increase the stride to find a balance between reducing the memory requirements and maintaining the quality of the parameter estimation and analysis of the light-cone quantities. If this is unsuccessful we will also try calculating the MPS on chunks of the full light-cone, as is done with the FPS.

So far we have ignored the telescope noise term (we've implemented $P_N = 1$ in Equation 5.11) for simplicity but this must be improved in future work. Our first attempts will consist of running a Gaussian filter across the noise power spectrum output from 21cmSense as shown in Equation 5.35 and repeating the methodology developed in Section 5.4.

As discussed in Section 1.3.7, 21cmSense is a foreground avoidance algorithm. One advantage of using an ergodic statistic like the Morlet transform is that the entire light-cone can be treated at once. This opens up possibility for improving foreground suppression since techniques ideally need as much bandwidth as possible to accurately mitigate the foregrounds and noise. For ex-

ample Gaussian Process Regression (GPR) (Mertens et al., 2018) and Blind Source Separation (BSS) algorithms (Alonso et al., 2015) take advantage of the continuity in astrophysical foreground signal to separate it from imprints of the instrument, such as mode-mixing (Cunnington et al., 2020). However in GPR, the signal evolution across the observed bandwidth worsens the fit of an accurate coherency scale, limiting the noise structure that can be produced by the covariance kernel. This poses an interesting question as to which foreground mitigation method will perform best when the bandwidth can be increased to the entire light-cone. Currently GPR is the mitigation technique of choice for interferometers (Mertens et al., 2020) however the success of the method relies on the accuracy of the noise covariance (Offringa et al., 2019; Kern & Liu, 2020). GPRs current success may be mildly benefited by slicing the light-cone into chunks since this allows a different coherence scale to be fit per chunk. It is not obvious how one could implement a coherence length that varies along the light-cone without significant change to the method, and it is therefore unclear which of these two methods will be more effective over extended bandwidths. Once the MPS is developed further, it will be interesting to see how foreground mitigation can progress when one single statistic can ergodically represent the entire light-cone.

Finally, with the introduction of this framework, there is the possibility to try various different transforms within the power spectrum. We have tried Gaussian wavelet envelopes only[2], but other wavelets could certainly be experimented with. A possible alternate is the Daubechies wavelet (Daubechies & Bates, 1993). This includes a scaling function to ensure the maximum number of vanishing moments, ensuring the transform is done using the smallest (sparsest) possible set of wavelet coefficients in the chosen basis. But we must not lose sight that the aim of this is to retrieve unbiased astrophysics from the observation. We have furthered the concept of wavelets from toy power spectrum models (Trott, 2016) to simple models of the EoR, reducing the light-cone effect. Once the light-cone effect is removed, peculiar velocities (Chapman & Santos, 2019) and the Alcock-Paczyński effect[3] are the only observational sources of non-isotropy that remain.

[2]C. Trott tried a set of Blackman-Harris envelopes to find they exaggerate covariance between wavelets making it significantly worse than with the Gaussian envelopes used here (private communication).

[3]Distortion of the light-cone through the geometry of the universe observed through the ratio of angular distance to redshift (Alcock & Paczynski, 1979).

Chapter 6

Thesis Conclusion

The cosmic microwave background (CMB) is the earliest observable signal in our universe. It signifies the formation of the first hydrogen atoms as the universe enters the dark ages before stars form for the first time in the cosmic dawn. It also shines through these primordial hydrogen atoms to excite the 21cm emission line which scientists today are looking to measure before that hydrogen becomes ionised during the epoch of reionisation (EoR). These first stars hold the secrets to how the universe populates itself with such a rich variety of species - ranging from stellar remnants that emit X-rays to the galaxies that drive reionisation and even the mature galaxies we see today. The EoR and cosmic dawn are a vast subject and the 21cm signal is sure to be revolutionary in shedding light on the deepest depths of our universe. With the upcoming of huge collaborative efforts, revolutionary instruments like the SKA will be able to put our most beloved theoretical models to the test. A new generation of precision cosmology is becoming a reality.

Summary

This work aims at improving the statistical analyses techniques used to probe the epoch of reionisation and cosmic dawn. All of the necessary literature is covered in the first chapter, with extensive supplementary material contained in the appendices. The next three chapters apply model selection to reionisation and the remaining work is aimed at improving the likelihood statistic used to constrain reionisation models.

Chapter 1 began with an introduction to the cosmic dawn and cosmic reionisation. Particular attention is given to how observations of the 21cm signal can be used to infer astrophysical properties about these cosmological epochs. We also introduced the simulation techniques and the specific algorithms used in later chapters. These include: 21CMSENSE, a foreground avoidance code that we use to estimate errors on simulated 21cm power spectrum observations; and 21CMMC, a state of the art EoR parameter estimation code which we use heavily in this work. The latter of these combines the semi-numerical reionisation simulation from 21CMFAST with the MCMC sampling algorithm EMCEE. An introduction to Bayesian analysis is also provided, along with a detailed description of the MULTINEST algorithm. EMCEE and MULTINEST are used extensively in this work for parameter estimation and model selection analyses respectively.

The research began in Chapter 2 with attempts to utilise MCMC chains from parameter estimation for the purpose of model selection. We introduced MCEVIDENCE, an algorithm designed to estimate a posterior for the Bayesian evidence from the density of points in the chain. The method has been shown to work on cosmological models and we showed that it works well on a model made from toy Gaussian likelihoods. Unfortunately when applied to EoR models the method was unable to match the results of MULTINEST. By recreating the behaviour within our toy model we conclude that the correlation of parameters is what causes assumptions in the method to breakdown. This explains how the method works for cosmological models, but not in the context of reionisation. It would have been a nice shortcut to performing Bayesian model selection but this was not to be and we quickly moved on.

Successful model selection begins in Chapter 3 via the development of 21CMNEST. By using MULTINEST within the framework of 21CMMC we are able to apply Bayesian model selection to multiple reionisation models in a flexible way. We test models of reionisation varying in morphology, from inside-out to outside-in; as well as scale, from global to local implementations. In each case we are able to distinguish the fiducial model with decisive results on the Jeffreys' scale. Distinguishing the fiducial parameters of the models proved tricky, particularly in the case of distinguishing galaxies with a constant versus power law mass-to-light ratio parameterised through the ultra violet ionising efficiency. Once model selection was achieved, an analysis of the observational priors found that neutral fraction checks at specific redshifts add little to no inference in comparison to the 21cm power spectrum (PS). We are also able to identify the photon mean free path parameter ($R_{\rm mfp}$, a maximum scale for the excursion set formalism) as

redundant via the application of the Savage-Dickey density ratio.

Chapter 4 is a continuation of the work started in Chapter 3. We move from simulating the 21cm PS with the coeval cubes to slices of the light-cone and introduce inhomogeneous recombination in the inter galactic medium. The models we test range from a simple model as in the previous section; a model that includes X-ray heating via high mass X-ray binaries produced by the remnants of the first generation of stars, this causes fluctuations in the 21cm spin temperature; a model with a power law in halo mass for the UV ionising efficiency via the star forming efficiency, this model can also produce accurate ultra violet luminosity functions (UV LFs) at high redshift; finally a model that combines both these attributes as well as a 2nd power law in ionising efficiency representing the UV escape fraction's scaling with halo mass. Although there is a major caveat in the work in this chapter, it appears as though we will be able to decisively distinguish models that include X-ray heating using 21cm PS observations ranging from redshifts, $z \in [8, 10]$. The UV LF shows significantly more promise than the observational prior checks used in Chapter 3, and we can decisively distinguish models with a single power law in halo mass for the ionising efficiency. We then used the Savage-Dickey density ratio to analyse the influence of the X-ray spectrum power law and the star formation timescale via the parameters $\alpha_{\rm X}$ and t_* respectively. We conclude that $\alpha_{\rm X}$ has little influence over the back-bone of reionisation but can be pinned down with observations of the 21cm PS that reach into the cosmic dawn. On the other hand, t_* would require more precise observational data from the UV LF or another observational synergy. These conclusions will be reassessed when the simulations with more appropriate box sizes have finished.

Chapter 5 looks at improving the statistical analyses used for constraining models with observations of the 21cm signal. The Morlet power spectrum (MPS) is new to the field and shows an increase to the constraining capabilities of the 21cm light-cone. In the previous chapter we perform our likelihood calculation by slicing the lightcone into chunks before calculating the 21cm PS. However this ignores evolution accross the chunk and leads to bias within the statistic. Here we replace the Fourier transform with a Morlet transform, enabling wavelets to sample the entire light-cone in an ergodic manor. Our results show a vast improvement to the parameter posteriors for the simple model from Chapter 4, however these results are very preliminary. Using a redshift range greater than two resulted in computational problems and so we defer a more complete analysis to future work.

Appendix A introduces astrophysical material, bridging the gap in astrophysical theory between undergraduate cosmology to the application of research in this field. We briefly recap the $\Lambda-$CDM model, highlighting where the gaps in observations lie before introducing the hydrogen 21cm line as a mechanism for resolving these gaps. Emphasis is drawn on the Gunn-Peterson trough, red Lyman-α damping wing, and the CMB optical depth since they provide a constraint on the end and duration of reionisation. These are referred to as the observational priors later on. The fundamentals of radio astronomy are included in Appendix B as a preparation for observing the 21cm line with interferometers. Here we also discuss the foreground wedge as well as different foreground mitigation techniques for calculating observational errors on the spherically averaged 21cm power spectrum. In Appendix C, we introduce Bayesian statistical methods and their implementation via MCMC algorithms. A derivation of the central limit theorem is provided as motivation for using a χ^2 as the Bayesian likelihood. Material that is supplementary to the material in Chapter 1 is provided here. Definitions of the power spectrum as the Fourier transform of the two point correlation function are provided in Appendix C.4 as this is used for the 21cm likelihood statistic. Supplementary analysis material for the research in Chapters 4 and 5 are included in Appendices D and E respectively.

Outlook

This work contains multiple unfinished parts that provide obvious places for its continuation. The most obvious being with the Morlet spectrum - for the current capabilities of computers it needs to be implemented in a way that uses significantly less memory. Naturally the method requires holding a statistic in RAM for each pixel along the line-of-sight of the light-cone, and then calculating a covariance between these. Possible solutions include slicing the light-cone into smaller chunks, as for the Fourier power spectrum; or using a logarithmic stepping of the wavelet centres throughout the light-cone. Since the covariance calculation could be parameterised as matrix multiplications, this could be an interesting avenue to apply GPU programming.

Once the Morlet power spectrum is running it may be possible for the 21cm light-cone to constrain a vast array of reionisation models. With model selection there are lots of options for continuing this research. For example, 21cmFast now contains the capability to simulate mini-halos - adding a second species to the simulation's galactic ensemble. Observational synergy is

also an avenue that can be expanded. The constraints from neutral fraction checks at specific redshifts, the UV LF, and the Planck optical depth will only improve as more high redshift objects are discovered and more detailed CMB observations will be improved by the likes of the James Webb space telescope and the Simons observatory. The global 21cm signal has not been utilised here either, and soon instruments such as REACH will be able to verify the EDGES results. Creating synergy between different observational methods within the Bayesian likelihood is, in the author's opinion, the most effective way to improve statistical analyses. Once the appropriate astrophysical model has been quantitatively selected, precisely constraining that model's parameters is necessary to infer the relevant astrophysics.

The cosmological 21cm line is set to be unrivalled in the information content it will provide due to the sheer volume that it will be able to probe. However its detection should be not taken for granted since the foregrounds outweigh the signal by several orders of magnitude. Current research surrounding foreground removal methods take advantage of the smooth nature of the spectra from the foreground signal. Having access to the entire light-cone in one statistic (which the MPS enables) has the potential to revolutionise the precision of interpolated foregrounds throughout an instrument's bandwidth.

It has been a pleasure contributing, even a little, towards the unveiling of the first galaxies in the universe.

Bibliography

Akaike H., 1974, IEEE Transactions on Automatic Control, 19, 716

Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2013, Astron. Comput., 2, 27

Alcock C., Paczynski B., 1979, Nature, 281, 358

Alfano S., Greer M., 2003, J. Guid. Control. Dynam., 26, 106

Ali Z. S., et al., 2015, ApJ, 809, 61

Alonso D., Bull P., Ferreira P. G., Santos M. G., 2015, Mon. Not. Roy. Astron. Soc., 447, 400

Anderson T. W., Darling D. A., 1952, The Annals of Mathematical Statistics, 23, 193

Anderson C., et al., 2018, Galaxies, 6, 127

Aubert D., Teyssier R., 2008, Mon. Not. Roy. Astron. Soc., 387, 295

Aubert D., Deparis N., Ocvirk P., 2015, Mon. Not. Roy. Astron. Soc., 454, 1012

BICEP2 Collaboration et al., 2014, Phys. Rev. Lett., 112, 241101

Baek S., Ferrara A., 2013, Mon. Not. Roy. Astron. Soc., 432, L6

Baek S., Semelin B., Di Matteo P., Revaz Y., Combes F., 2010, A&A, 523, A4

Barkana R., 2018, Nature, 555, 71

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125

Barkana R., Loeb A., 2004, ApJ, 609, 474

Barkana R., Loeb A., 2005, ApJ, 626, 1

Barry N., et al., 2019, ApJ, 884, 1

Barvainis R., 1997, Publ. Astron. Soc. Pac., 109, 1167

Bayes T., Price n., 1763, Philos. T. R. Soc. Lon., 53, 370

Becker R. H., et al., 2001, Astron. J., 122, 2850

Benson B., 2018, in 42nd COSPAR Scientific Assembly. pp E1.2–26–18

Binnie T., Pritchard J. R., 2019, Mon. Not. Roy. Astron. Soc., 487, 1160

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bond J. R., Jaffe A. H., Knox L., 1998, Phys. Rev. D, 57, 2117

Borthakur S., Heckman T. M., Leitherer C., Overzier R. A., 2014, Science, 346, 216

Bouwens R. J., et al., 2011, ApJ, 737, 90

Bouwens R. J., et al., 2015, ApJ, 803, 34

Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67

Burke B. F., 1985, in Mendell W. W., ed., Lunar Bases and Space Activities of the 21st Century. pp 281–291

Burke B. F., Graham-Smith F., 2002, An Introduction to Radio Astronomy: Second Edition

Byrne R., et al., 2019, ApJ, 875, 70

Cain C., D'Aloisio A., Iršič V., McQuinn M., Trac H., 2020, ApJ, 898, 168

Chapman E., Santos M. G., 2019, Mon. Not. Roy. Astron. Soc., 490, 1255

Chapman E., et al., 2012, Mon. Not. Roy. Astron. Soc., 423, 2518

Chapman E., et al., 2013, Mon. Not. Roy. Astron. Soc., 429, 165

Chapman E., Zaroubi S., Abdalla F., Dulwich F., Jelić V., Mort B., 2014, arXiv e-prints, p. arXiv:1408.4695

Chapman E., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 5

Chapman E., Zaroubi S., Abdalla F. B., Dulwich F., Jelić V., Mort B., 2016, Mon. Not. Roy. Astron. Soc., 458, 2928

Chen X., Miralda-Escudé J., 2004, ApJ, 602, 1

Choi Y.-K., Wang W., Liu Y., Kim M.-S., 2006, Robotics, IEEE Transactions on, 22, 213

Ciardi B., Scannapieco E., Stoehr F., Ferrara A., Iliev I. T., Shapiro P. R., 2006, Mon. Not. Roy. Astron. Soc., 366, 689

Cohen A., Fialkov A., Barkana R., 2016, Mon. Not. Roy. Astron. Soc., 459, L90

Cohen A., Fialkov A., Barkana R., Monsalve R., 2019, arXiv e-prints, p. arXiv:1910.06274

Cunnington S., Irfan M. O., Carucci I. P., Pourtsidou A., Bobin J., 2020, arXiv e-prints, p. arXiv:2010.02907

Das A., Mesinger A., Pallottini A., Ferrara A., Wise J. H., 2017, Mon. Not. Roy. Astron. Soc., 469, 1166

Datta K. K., Mellema G., Mao Y., Iliev I. T., Shapiro P. R., Ahn K., 2012, Mon. Not. Roy. Astron. Soc., 424, 1877

Datta K. K., Jensen H., Majumdar S., Mellema G., Iliev I. T., Mao Y., Shapiro P. R., Ahn K., 2014, Mon. Not. Roy. Astron. Soc., 442, 1491

Daubechies I., Bates B. J., 1993, J. Acoust. Soc. Am., 93, 1671

Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, Mon. Not. Roy. Astron. Soc., 445, 2545

DeBoer D. R., et al., 2017, Publ. Astron. Soc. Pac., 129, 045001

Deep Kaur H., Gillet N., Mesinger A., 2020, arXiv e-prints, p. arXiv:2004.06709

Dickey J. M., 1971, The Annals of Mathematical Statistics, 42, 204

Dillon J. S., Parsons A. R., 2016, ApJ, 826, 181

Dillon J. S., Liu A., Tegmark M., 2013, Phys. Rev. D, 87, 043005

Dillon J. S., et al., 2015, Phys. Rev. D, 91, 123011

Dillon J. S., Orosz N., Parsons A., Ewall-Wice A., Thyagarajan N., 2019, in AAS Meeting Abstracts. p. 413.01

Dulwich F., Mort B. J., Salvini S., Zarb Adami K., Jones M. E., 2009, in Wide Field Astronomy & Technology for the Square Kilometre Array. p. 31

Eide M. B., Graziani L., Ciardi B., Feng Y., Kakiichi K., Di Matteo T., 2018, Mon. Not. Roy. Astron. Soc., 476, 1174

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 831–839

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605

Eisenstein D. J., Hu W., 1999, ApJ, 511, 5

Ellis R. S., 2014, arXiv e-prints, p. arXiv:1411.3330

Ewall-Wice A., Hewitt J., Mesinger A., Dillon J. S., Liu A., Pober J., 2016, Mon. Not. Roy. Astron. Soc., 458, 2710

Ewen H. I., Purcell E. M., 1951a, National Radio Astronomy Observatory NRAO/AUI/NSF. https://www.nrao.edu/whatisra/hist_ewenpurcell.shtml

Ewen H. I., Purcell E. M., 1951b, Nature, 168, 356

Fan X., et al., 2006, Astron. J., 132, 117

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, Mon. Not. Roy. Astron. Soc., 455, 2778

Feroz F., Hobson M. P., 2008, Mon. Not. Roy. Astron. Soc., 384, 449

Feroz F., Skilling J., 2013, in von Toussaint U., ed., Vol. 1553, AIP Conference Series. pp 106–113

Feroz F., Hobson M. P., Bridges M., 2009, Mon. Not. Roy. Astron. Soc., 398, 1601

Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, OJ Ap, 2, 10

Ferrara A., Loeb A., 2013, Mon. Not. Roy. Astron. Soc., 431, 2826

Fialkov A., Barkana R., 2019, Mon. Not. Roy. Astron. Soc., 486, 1763

Fialkov A., Barkana R., Pinhas A., Visbal E., 2014a, Mon. Not. Roy. Astron. Soc., 437, L36

Fialkov A., Barkana R., Visbal E., 2014b, Nature, 506, 197

Field G. B., 1958, Proceedings of the IRE, 46, 240

Field G. B., 1959, ApJ, 129, 536

Finlator K., Keating L., Oppenheimer B. D., Davé R., Zackrisson E., 2018, Mon. Not. Roy. Astron. Soc., 480, 2628

Flechl M., 2015, arXiv e-prints, p. arXiv:1510.01924

Foot C. J., 2005, Atomic Physics. OUP Oxford

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publ. Astron. Soc. Pac., 125, 306

Fragos T., Lehmer B. D., Naoz S., Zezas A., Basu-Zych A., 2013, ApJl, 776, L31

Furlanetto S. R., 2006, Mon. Not. Roy. Astron. Soc., 371, 867

Furlanetto S. R., Furlanetto M. R., 2007, Mon. Not. Roy. Astron. Soc., 379, 130

Furlanetto S. R., Oh S. P., 2005, Mon. Not. Roy. Astron. Soc., 363, 1031

Furlanetto S. R., Oh S. P., 2008, ApJ, 681, 1

Furlanetto S. R., Piran T., 2006, Mon. Not. Roy. Astron. Soc., 366, 467

Furlanetto S. R., Pritchard J. R., 2006, Mon. Not. Roy. Astron. Soc., 372, 1093

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

Furlanetto S. R., McQuinn M., Hernquist L., 2006a, Mon. Not. Roy. Astron. Soc., 365, 115

Furlanetto S. R., Oh S. P., Briggs F. H., 2006b, Phys. Rep., 433, 181

Garaldi E., Compostella M., Porciani C., 2019, Mon. Not. Roy. Astron. Soc., 483, 5301

Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485

Garofalo M., Botta A., Ventre G., 2017, in Brescia M., Djorgovski S. G., Feigelson E. D., Longo G., Cavuoti S., eds, IAU Symposium Vol. 325, Astroinformatics. pp 345–348

Gauss C. F., 1809, Gauss's Derivation of the Normal Distribution and the Method of Least Squares. Springer New York

Gelman A., Rubin D. B., 1992, Statistical Science, 7, 457

Ghara R., et al., 2020, Mon. Not. Roy. Astron. Soc., 493, 4728

Gillet N. J. F., Mesinger A., Park J., 2020, Mon. Not. Roy. Astron. Soc., 491, 1980

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

Goodman J., Weare J., 2010, Comm. App. Math. Com. Sc., 5, 65

Gorce A., Douspis M., Aghanim N., Langer M., 2018, A&A, 616, A113

Gordon C., Pritchard J. R., 2009, Phys. Rev. D, 80, 063535

Gott J. Richard I., Colley W. N., 2017, arXiv e-prints, p. arXiv:1707.06755

Goupillaud P., Grossmann A., Morlet J., 1984, Geoexploration, 23, 85

Graff P., Feroz F., Hobson M. P., Lasenby A., 2012, Mon. Not. Roy. Astron. Soc., 421, 169

Greig B., Mesinger A., 2015, Mon. Not. Roy. Astron. Soc., 449, 4246

Greig B., Mesinger A., 2017a, Mon. Not. Roy. Astron. Soc., 465, 4838

Greig B., Mesinger A., 2017b, Mon. Not. Roy. Astron. Soc., 472, 2651

Greig B., Mesinger A., 2018, Mon. Not. Roy. Astron. Soc., 477, 3217

Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, Mon. Not. Roy. Astron. Soc., 466, 4239

Greig B., et al., 2020a, arXiv e-prints, p. arXiv:2006.03203

Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2020b, arXiv e-prints, p. arXiv:2008.02639

Greig B., Mesinger A., Koopmans L. V. E., 2020c, Mon. Not. Roy. Astron. Soc., 491, 1398

Gull S. F., 1989, Bayesian Data Analysis: Straight-line fitting. Springer Netherlands

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Guth A., 1997, The inflationary universe. The quest for a new theory of cosmic origins

Hahnemann S. Dudgeon R., 1810, Organon der Heilkunst. Forgotten Books

Hamerly G., Elkan C., 2004, Advances in Neural Information Processing Systems 16, p 281-288. MIT Press

Handley W. J., Hobson M. P., Lasenby A. N., 2015, Mon. Not. Roy. Astron. Soc., 453, 4384

Harnois-Déraps J., Pen U.-L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013, Mon. Not. Roy. Astron. Soc., 436, 540

Hartigan J. A., Wong M. A., 1979, JSTOR: Applied Statistics, 28, 100

Hassan S., Liu A., Kohn S., La Plante P., 2019, Mon. Not. Roy. Astron. Soc., 483, 2524

Hastings W. K., 1970, Biometrika, 57, 97

Heavens A., Fantaye Y., Mootoovaloo A., Eggers H., Hosenie Z., Kroon S., Sellentin E., 2017a, arXiv e-prints, p. arXiv:1704.03472

Heavens A., Fantaye Y., Sellentin E., Eggers H., Hosenie Z., Kroon S., Mootoovaloo A., 2017b, Phys. Rev. Lett., 119, 101301

Heger A., Woosley S. E., 2002, ApJ, 567, 532

Higson E., Handley W., Hobson M., Lasenby A., 2019, Mon. Not. Roy. Astron. Soc., 483, 2044

Hills R., Kulkarni G., Meerburg P. D., Puchwein E., 2018, Nature, 564, E32

Hirata C. M., Sigurdson K., 2007, Mon. Not. Roy. Astron. Soc., 375, 1241

Hobson H., Jaffe A., Liddle A., Mukherjee P., D. P., 2009, Bayesian Methods in Cosmology. Cambridge University Press

Hotinli S. C., Kamionkowski M., Bikash D., Binnie T., Muñoz J. B., In prep., OJAp

Hubble E., 1929, P. Natl. Proc. Sci., 15, 168

Hutter A., 2018, Mon. Not. Roy. Astron. Soc., 477, 1549

Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, Mon. Not. Roy. Astron. Soc., 369, 1625

Jacobs D. C., et al., 2013, ApJ, 776, 108

Jaynes E. T., 2003, Probability theory: The logic of science. Cambridge University Press

Kaiser N., 1987, Mon. Not. Roy. Astron. Soc., 227, 1

Kern N., Liu A., 2020, arXiv e-prints, p. arXiv:2010.15892

Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, ApJ, 848, 23

Koopmans L., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 1

Kullback S., Leibler R. A., 1951, Ann. Math. Statist., 22, 79

Kunz M., Trotta R., Parkinson D. R., 2006, Phys. Rev. D, 74, 023503

La Plante P., Battaglia N., Natarajan A., Peterson J. B., Trac H., Cen R., Loeb A., 2014, ApJ, 789, 31

La Plante P., Lidz A., Aguirre J., Kohn S., 2020, arXiv e-prints, p. arXiv:2005.07206

Lahav O., 1999, in Morganti R., Couch W. J., eds, Looking Deep in the Southern Sky. p. 42

Lahav O., Liddle A. R., 2019, arXiv e-prints, p. arXiv:1912.03687

Lazio J., Carilli C., Hewitt J., Furlanetto S., Burns J., 2009, in UV/Optical/IR Space Telescopes: Innovative Technologies and Concepts IV. p. 74360I

Lehmer B. D., et al., 2016, ApJ, 825, 7

Lemaitre G. H., 1927, PhD thesis, MIT

Liddle A. R., 2007, Mon. Not. Roy. Astron. Soc., 377, L74

Liu A., Parsons A. R., Trott C. M., 2014a, Phys. Rev. D, 90, 023018

Liu A., Parsons A. R., Trott C. M., 2014b, Phys. Rev. D, 90, 023019

Loeb A., Furlanetto S., 2013, The First Galaxies in the Universe. Princeton University Press

Loeb A., Zaldarriaga M., 2004, Phys. Rev. Lett., 92, 211301

Madau P., Fragos T., 2017, ApJ, 840, 39

Magueijo J., Sorkin R. D., 2007, Mon. Not. Roy. Astron. Soc., 377, L39

Martin J., Ringeval C., Trotta R., Vennin V., 2014, J. Cosmol. Astropart. P., 2014, 039

Martinez G. D., McKay J., Farmer B., Scott P., Roebber E., Putze A., Conrad J., 2017, Eur. Phys. Jour. C, 77, 761

McGreer I. D., Mesinger A., Fan X., 2011, Mon. Not. Roy. Astron. Soc., 415, 3237

McGreer I. D., Mesinger A., D'Odorico V., 2015, Mon. Not. Roy. Astron. Soc., 447, 499

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815

McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, ApJ, 743, 82

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

Mellema G., et al., 2013, Exp. Astron., 36, 235

Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, Mon. Not. Roy. Astron. Soc., 478, 3640

Mertens F. G., et al., 2020, Mon. Not. Roy. Astron. Soc., 493, 1662

Merz H., Pen U.-L., Trac H., 2011, Software, p. ascl:1102.008

Mesinger A., Furlanetto S., 2007, ApJ, 669, 663

Mesinger A., Furlanetto S., Cen R., 2011, Mon. Not. Roy. Astron. Soc., 411, 955

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem. Phys., 21, 1087

Miralda-Escudé J., 1998, ApJ, 501, 15

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

Mondal R., Bharadwaj S., Majumdar S., Bera A., Acharyya A., 2015, Mon. Not. Roy. Astron. Soc., 449, L41

Mondal R., Bharadwaj S., Majumdar S., 2016, Mon. Not. Roy. Astron. Soc., 456, 1936

Mondal R., Bharadwaj S., Majumdar S., 2017, Mon. Not. Roy. Astron. Soc., 464, 2992

Mondal R., Bharadwaj S., Datta K. K., 2018, Mon. Not. Roy. Astron. Soc., 474, 1390

Mootoovaloo A., Bassett B. A., Kunz M., 2016, arXiv e-prints, p. arXiv:1609.02186

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

Muñoz J. B., 2019, Phys. Rev. D, 100, 063538

Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJl, 638, L51

Muller C. A., Oort J. H., 1951, Nature, 168, 357

NASA/WMAP 2012. https://map.gsfc.nasa.gov/site/citations.html

Nasirudin A., Murray S., Trott C., Greig B., Joseph R., Power C., 2020, arXiv e-prints, p. arXiv:2003.08552

Neal R. M., 2000, arXiv e-prints, p. physics/0009028

Norris R. P., et al., 2020, arXiv e-prints, p. arXiv:2006.14805

Ockham W., 1318, The Four Books of Sentences. Franciscan Institute Publications

Ocvirk P., et al., 2016, Mon. Not. Roy. Astron. Soc., 463, 1462

Ocvirk P., et al., 2018, arXiv e-prints, p. arXiv:1811.11192

Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105

Offringa A. R., Mertens F., Koopmans L. V. E., 2019, Mon. Not. Roy. Astron. Soc., 484, 2866

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Osterbrock D., Ferland G., 2006, Astrophysics Of Gas Nebulae and Active Galactic Nuclei. University Science Books

Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, Mon. Not. Roy. Astron. Soc., 451, 2544

Paciga G., et al., 2013, Mon. Not. Roy. Astron. Soc., 433, 639

Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, Mon. Not. Roy. Astron. Soc., 443, 678

Pagano M., Liu A., 2020, Mon. Not. Roy. Astron. Soc., 498, 373

Park H., Shapiro P. R., Komatsu E., Iliev I. T., Ahn K., Mellema G., 2013, ApJ, 769, 93

Park J., Mesinger A., Greig B., Gillet N., 2019, Mon. Not. Roy. Astron. Soc., 484, 933

Park J., Gillet N., Mesinger A., Greig B., 2020, Mon. Not. Roy. Astron. Soc., 491, 3891

Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012, ApJ, 753, 81

Parsons A. R., et al., 2014, ApJ, 788, 106

Patil A. H., et al., 2017, ApJ, 838, 65

Peacock J. A., 1999, Cosmological Physics. Cambridge University Press

Peebles P. J. E., 1968, ApJ, 153, 1

Pelleg D., Moore A., 2000, in Proc. of the 17th International Conf. on Machine Learning. Morgan Kaufmann, pp 727–734

Perkins D., 2005, Particle Astrophysics. OUP Oxford

Pham D., Dimov S., Nguyen C., 2005, P. I. Mech. Eng. C-J Mec., 219, 103

Planck Collaboration et al., 2016a, A&A, 594, A13

Planck Collaboration et al., 2016b, A&A, 596, A108

Pober J., 2016, Software, p. ascl:1609.013

Pober J. C., et al., 2013, Astron. J., 145, 65

Pober J. C., et al., 2014, ApJ, 782, 66

Press W. H., Schechter P., 1974, ApJ, 187, 425

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2007, Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3 edn. Cambridge University Press, USA

Pritchard J. R., Furlanetto S. R., 2006, Mon. Not. Roy. Astron. Soc., 367, 1057

Pritchard J. R., Furlanetto S. R., 2007, Mon. Not. Roy. Astron. Soc., 376, 1680

Pritchard J. R., Loeb A., 2008, Phys. Rev. D, 78, 103511

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Qin Y., Mesinger A., Park J., Greig B., Muñoz J. B., 2020, Mon. Not. Roy. Astron. Soc., 495, 123

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, Mon. Not. Roy. Astron. Soc., 430, 2427

Reichardt C. L., et al., 2020, arXiv e-prints, p. arXiv:2002.06197

Rojas-Ruiz S., Finkelstein S. L., Bagley M. B., Stevans M., Finkelstein K. D., Larson R., Mechtley M., Diekmann J., 2020, ApJ, 891, 146

Saitou K., Tsujimoto M., Ebisawa K., Ishida M., Mukai K., Nagayama T., Nishiyama S., Gandhi P., 2011, Publ. Astron. Soc. Jpn., 63, S759

Salpeter E. E., 1955, ApJ, 121, 161

Santos M., Ferramacho L., Silva M., Amblard A., Cooray A., 2010, Software, p. ascl:1010.025

Schellart P., et al., 2013, A&A, 560, A98

Schmit C. J., Pritchard J. R., 2018, Mon. Not. Roy. Astron. Soc., 475, 1213

Schmit C. J., Heavens A. F., Pritchard J. R., 2019, Mon. Not. Roy. Astron. Soc., 483, 4259

Schwarz G., 1978, Annals of Statistics, 6, 461

Scoccimarro R., 2004, Phys. Rev. D, 70, 083007

Seager S., Sasselov D. D., Scott D., 1999, ApJl, 523, L1

Shaw J. R., Bridges M., Hobson M. P., 2007, Mon. Not. Roy. Astron. Soc., 378, 1365

Sheth R. K., Tormen G., 1999, Mon. Not. Roy. Astron. Soc., 308, 119

Singh S., et al., 2018, ApJ, 858, 54

Sivia D., Skilling J., 2006, Data Analysis: A Bayesian Tutorial. OUP Oxford

Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, Vol. 735, AIP Conference Series. pp 395–405

Skilling J., 2012, in Goyal P., Giffin A., Knuth K. H., Vrscay E., eds, Vol. 1443, AIP Conference Series. pp 145–156

Sobacchi E., Mesinger A., 2014, Mon. Not. Roy. Astron. Soc., 440, 1662

Sobolev V. V., 1957, Astron. ZH+, 34, 694

Speagle J. S., 2020, Mon. Not. Roy. Astron. Soc., 493, 3132

Speake C. C., Will C. M., 2012, Classical Quant. Grav., 29, 180301

Sun G., Furlanetto S. R., 2016, Mon. Not. Roy. Astron. Soc., 460, 417

Sunyaev R. A., Zeldovich I. B., 1980, ARA&A, 18, 537

Szydłowski M., Krawiec A., Kurek A., Kamionka M., 2015, Eur. Phys. Jour. C, 75, 5

Tegmark M., 1997, Phys. Rev. D, 55, 5895

Thomas R. M., Zaroubi S., 2011, Mon. Not. Roy. Astron. Soc., 410, 1377

Trott C. M., 2016, Mon. Not. Roy. Astron. Soc., 461, 126

Trott C. M., Pober J. C., 2019, arXiv e-prints, p. arXiv:1909.12491

Trott C. M., et al., 2020, Mon. Not. Roy. Astron. Soc., 493, 4711

Trotta R., 2008, Contemporary Physics, 49, 71

Tseliakhovich D., Hirata C., 2010, Phys. Rev. D, 82, 083520

Wallace C., 2005, Statistical and Inductive Inference by Minimum Message Length. Springer

Watkinson C. A., Pritchard J. R., 2014, Mon. Not. Roy. Astron. Soc., 443, 3090

Watkinson C. A., Majumdar S., Pritchard J. R., Mondal R., 2017, Mon. Not. Roy. Astron. Soc., 472, 2436

Watkinson C. A., Giri S. K., Ross H. E., Dixon K. L., Iliev I. T., Mellema G., Pritchard J. R., 2019, Mon. Not. Roy. Astron. Soc., 482, 2653

Watkinson C. A., Trott C. M., Hothi I., 2020, arXiv e-prints, p. arXiv:2002.05992

Wilson T. L., Rohlfs K., Hüttemeister S., 2013, Tools of Radio Astronomy

Wise J. H., 2019, arXiv e-prints, p. arXiv:1907.06653

Wise J. H., Cen R., 2009, ApJ, 693, 984

Wise J. H., Abel T., Turk M. J., Norman M. L., Smith B. D., 2012a, Mon. Not. Roy. Astron. Soc., 427, 311

Wise J. H., Turk M. J., Norman M. L., Abel T., 2012b, ApJ, 745, 50

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, Mon. Not. Roy. Astron. Soc., 442, 2560

Wittke J. P., Dicke R. H., 1956, Phys. Rev., 103, 620

Wouthuysen S. A., 1952, Astron. J., 57, 31

Wu X., Kannan R., Marinacci F., Vogelsberger M., Hernquist L., 2019, Mon. Not. Roy. Astron. Soc., 488, 419

Yamamoto K., Nishioka H., Suto Y., 1999, ApJ, 527, 488

Yue B., et al., 2018, ApJ, 868, 115

Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, Mon. Not. Roy. Astron. Soc., 414, 727

Zaroubi S., 2013, The Epoch of Reionization. p. 45

Zel'Dovich Y. B., 1970, A&A, 500, 13

Zeldovich Y. B., Sunyaev R. A., 1969, Astrophys. Space Sci., 4, 301

Zhang L., Zhang M., Liu X., 2016, Astrophys. Space Sci., 361, 153

Zygelman B., 2005, ApJ, 622, 1356

de Bernardis P., et al., 2000, Nature, 404, 955

van de Hulst H. C., 1951, Astron. J., 56, 144

van der Velden E., Duffy A. R., Croton D., Mutch S. J., Sinha M., 2019, ApJs, 242, 22

Appendix A

Introductory Astrophysics

This appendix provides the astrophysical and cosmological context necessary for interpreting probes of reionisation. The intention here is to bridge the gap to this work from a typical undergraduate course, which might overlap with Sections A.1 and A.2 in a typical cosmology module. We introduce the theory behind calculating quantities in the early universe as well as detailing the current gap in observations and how the cosmological 21cm signal can fill this gap. Then we introduce the hydrogen 21cm line and its theoretical intricacies, leaving the reader up-to-date with the cosmological material required in Section 1.2.

We begin with the foundations of cosmology in Section A.1, including an overview of the $\Lambda-$CDM cosmological model used throughout this work. Section A.2 discusses well established cosmological observations focusing on how they confirm the theoretical work in the previous section and highlighting the areas in-which they need improving. Observations of the red Ly-α damping wing, the Gunn-Peterson trough and the Planck reionisation optical depth are particularly detailed as they are implemented in a Bayesian framework in every chapter (introduced in Section 1.3.2).

Section A.3 details the derivation of the 21cm signal as the hype-fine splitting of hydrogen's atomic line spectra. This is then applied to the radiative transfer that occurs within clouds of primordial hydrogen gas - resulting in the 21cm brightness temperature. Finally, to track the evolution of the 21cm brightness temperature in a cosmological context, we derive the 21cm spin temperature in terms of a UV coupling, a kinetic coupling, and coupling to the CMB. These results lead into the material used in Section 1.2.

A.1 Cosmological theory

In 1915 Albert Einstein, with the help of the mathematician David Hilbert, developed the general theory of relativity (Einstein, 1915), the most precise theory of gravity to date. Due to the equivalence principle, where gravitational mass is assumed equivalent to inertial mass, the trajectories of matter are identified as following geodesics along the surface of what is known as space-time. This holds true to any currently obtainable precision (Speake & Will, 2012). In brief, matter itself describes how the 4-dimensional space-time is shaped. The motion along this surface is described in Einstein's field equations (EFE) as,

$$\mathcal{G}_{\mu\nu} = \frac{8\pi G}{c^4}\mathcal{T}_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{A.1}$$

where $\mathcal{G}_{\mu\nu}$, the Einstein tensor, describes the curvature of space-time; $\mathcal{T}_{\mu\nu}$ is the energy-momentum tensor, i.e. the distribution of energy density (multiplied by the constants $8\pi G/c^4$); Λ is the cosmological constant which represents the vacuum energy density of space and is often referred to in relation to dark energy. The origin of the cosmological constant is unknown but at late times it is responsible for accelerating the expansion of the universe. At earlier times, matter and radiation dominate the energy density of the universe however these dilute as the scale of the universe increases. The expansion of the universe gives rise to the concept of redshift, which can be thought of as the effective Doppler shift caused by the receding source in the context of an expanding universe, $z = \nu_{\rm lab}/\nu_{\rm obs} - 1$ (Figure A.1). Redshift is commonly used to catalogue the age of sources. As John Peacock phrases it: 'cosmology is the task of finding solutions to Einstein's field equations that are consistent with the large-scale matter distribution in the universe' (Peacock, 1999). The EFE are a set of 10 simultaneous partial differential equations, which are not easy to solve. Since Einstein developed this before computers even existed, some simplifying assumptions were necessary, i.e. that the universe is homogeneous and isotropic when viewed on and above a large enough scale ($\sim 250 \times 10^6$ly). This is known as the 'cosmological principle', which is still a valid assumption today (Lahav, 1999). The solutions to the EFE take the form of a metric tensor, $g^{\mu\nu}$ which represents the geometry of space-time. Metric tensors can be used to express a displacement interval in the context of space-time as, $ds^2 = X^\mu g_{\mu\nu} X^\nu$, where X are position 4-vectors. Robertson, Walker, Friedmann and Lemaître (Peacock, 1999; Loeb & Furlanetto, 2013) implemented the cosmological principle with the EFE to produce what is known as the FLRW metric. The displacement

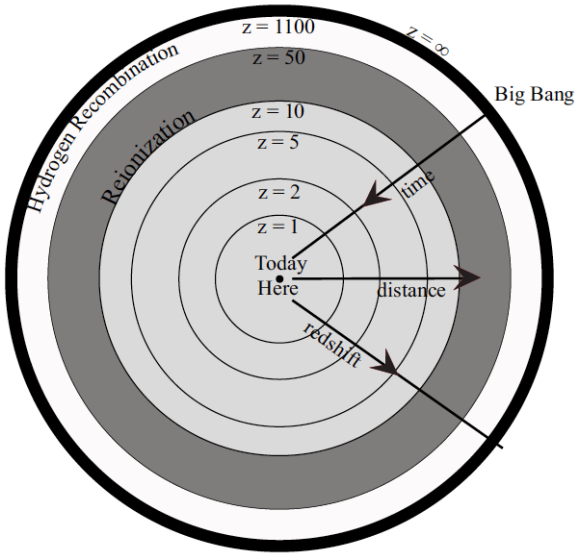


Figure A.1: The geocentric view of the universe in comoving coordinates (coordinates which compensate for cosmic expansion). Due to the finite speed of light and the sheer scale of the universe, signals (and therefore images) detected are up to $\sim$ 13.8 billion years old. The first detectable signal is the CMB, a 2-dimensional snapshot of the universe as hydrogen atoms are able to form for the first time since the big bang in Recombination. Beyond this the universe is opaque. Here the logarithmic nature of redshift and its relevance to the age of the universe becomes apparent (Loeb & Furlanetto, 2013).

interval of this metric is expressed as,

$$ds^2 = -c^2dt^2 + a(t)\left\{\frac{dr^2}{1-kr} + r^2d\theta^2 + r^2sin^2\theta d\phi^2\right\}, \tag{A.2}$$

and this is the simplest solution to the EFE that is consistent with today's matter distribution in the universe. Note that this is based on a Minkowski (flat) metric other than for the addition of the scale factor, $a(t)$, which represents the rate of spatial expansion with respect to time and the possibility for spatial curvature, k.

To gain some physical insight into Equation A.2, we introduce what are known as density parameters, $\Omega_i = \rho_i/\rho_{\text{crit}}$, which represent the content of the universe occupied by species i. These are defined as the ratio between the energy density of the species present, ρ_i to the critical energy density required to halt the universe's expansion, ρ_{crit}. By manipulating some algebra we can represent how the relevant quantities evolve with the scale factor as,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left(\frac{\Omega_{\text{M}}}{a^3} + \frac{\Omega_{\text{R}}}{a^4} + \Omega_\Lambda + \frac{\Omega_k}{a^2}\right). \tag{A.3}$$

Equation A.3 contains the definition of the Hubble parameter H, as a function of the scale

parameter as well as the density parameters for matter (M, non-relativistic species including dark matter), radiation (relativistic energy density), the cosmological constant and the energy density in spatial curvature respectively. Each density parameter is in units of a critical density, i.e. the energy density required to balance the local universal expansion and therefore satisfies,

$$\Omega_{\rm M} + \Omega_{\rm R} + \Omega_{\Lambda} = 1 + \Omega_k. \tag{A.4}$$

The spatial curvature has been measured by Planck to be: $|\Omega_k| < 0.005$ (Planck Collaboration et al., 2016a). The necessity of precise cosmology becomes apparent when an astronomer wishes to work with two equidistant objects objects between two separated by, $\mathbf{r}_{\perp}$, where this vector is perpendicular to the line of sight. In this example, this distance subtends an angle θ on the sky. To maintain some geometrical sense, we know that $|\mathbf{r}_{\perp}| = D_{\rm c}\theta$ must hold for a distance scale, $D_{\rm c}$, which is consistent despite the evolution of the universe due to cosmology. This is known as a comoving distance,

$$D_{\rm c} \equiv \frac{c}{H_0} \int \frac{dz'}{E(z')}, \tag{A.5}$$

where $E(z) = \sqrt{\Omega_{\rm M}(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{\rm R}(1+z)^4 + \Omega_{\Lambda})}$. From Equation A.3 it should be clear that the universe evolves with the expansion. Different epochs are defined depending on which energy density component is dominant as this defines the nature of the expansion that occurs at that time. For example, at early times ($z > 3300$), radiation pressure is thought to dominate but in contrast today $\Omega_{\rm R} \sim 0.0003$ which is negligible (see Figure A.2).

Since the cosmological principle is not precisely true at all scales, we must add the small scale structure with linear perturbation theory. The departure from homogeneity requires studying the evolution of energy density, pressure and gravitational potential (ρ, p and Φ respectively). By perturbing the FLRW metric with a weak gravitational potential (consistent with Newtonian limits) we gain a schematic picture of the EFE, as shown in,

$$ds^2 = -c^2[1 + 2\Phi]dt^2 + a(t)[1 - 2\Phi]d\mathbf{r}^2, \tag{A.6}$$

and,

$$\delta\mathcal{G}_{\mu\nu} \propto \delta\mathcal{T}_{\mu\nu}. \tag{A.7}$$

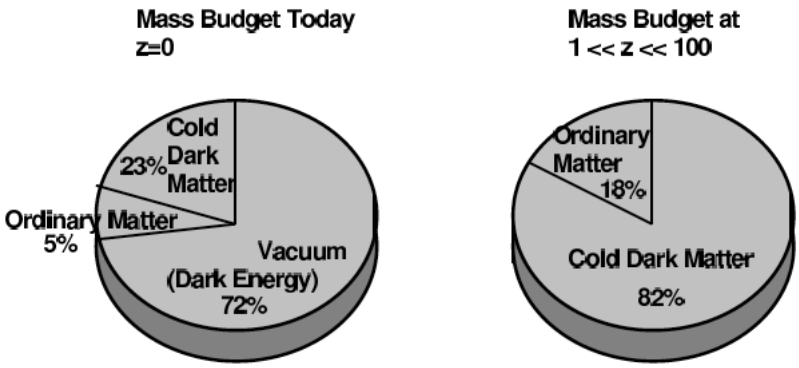


Figure A.2: The density ratios (mass-energy budgets) of the present day ($z = 0$) and during the formation of the first galaxies ($z = [10, 50]$). As the universe expands the densities of matter and radiation decrease, meaning the vacuum density of the universe will eventually dominate causing the expansion to accelerate. $z = 0 \rightarrow z = 1$ corresponds roughly to the most recent 6 billion years in the universe (Loeb & Furlanetto, 2013).

Applying the conservation of mass and momentum respectively to the perturbed EFE gives,

$$\delta' = -(1 + w)(\boldsymbol{\nabla} \cdot \mathbf{v} - 3\Phi') - 3aH(c_s - w)\delta, \tag{A.8}$$

and,

$$(\boldsymbol{\nabla} \cdot \mathbf{v})' = -aH(1 - 3w)\boldsymbol{\nabla} \cdot \mathbf{v} + \frac{c_s^2 k^2}{1 + w} + k^2\Phi, \tag{A.9}$$

where $c_s = \delta P/\delta\rho$ is the sound speed and $w = p/(\rho c^2)$ is the dark energy equation of state. These provides some key insights into structure formation. The Planck team have measured the expansion of the universe with type 1a supernovae to constrain $w = -1.006 \pm 0.045$ consistent with the properties of a cosmological constant, Λ (Planck Collaboration et al., 2016a). $\mathbf{v}$ is the spacial component of the velocity four-vector for the perturbing fluid, and the prime indicates a comoving rate of change. Most importantly, on all scales these equations can be used to study the evolution of perturbations at any epoch - due to their dependence on H.

With an overview of the behaviour of structure, we now turn to look at how structure forms, which we approach chronologically. After the big bang, the universe is radiation dominated, a soup of baryons, photons, and dark matter that steadily cools as the universe expands. As the universe expands, the energy of radiation decreases and baryons of increasing mass condense out of this primordial mix. The first detectable signal occurs at $z \sim 1100$, when the universe cools enough for the first atoms to form. Prior to this, the universe is opaque due to the large number

of Compton scattering events that occur. The rate of these scattering events is proportional to the free electron density, which decreases as the hydrogen atoms start to combine. As the universe expands, the average energy density of the photons decreases too, until they cannot ionise the hydrogen atoms as they try to form. This is known as recombination, which gives rise to the cosmic microwave background (CMB). The CMB is snapshot of the universe as it cooled past this threshold, and is detectable today (we will return to this in Section A.2). Interestingly $\sim 1\%$ of the static picked up by old televisions and radios comes from the CMB. After the CMB the universe continues to expand, the newly formed hydrogen continues to cool, and the epoch known as the dark ages begins. This period earns its name due to a theorised lack of detectable signals - since no stars have formed yet, the universe is literally dark. Although there is little to be detected here, the over and under densities of energy in the CMB seed the distribution of matter later on.

We describe the distribution of large scale structure with the Press-Schechter (P-S) halo mass function (Press & Schechter, 1974),

$$n(m)dm = -\sqrt{\frac{2}{\pi}}\frac{\rho_{\rm M}}{m}\frac{\delta_c}{\sigma(m)}e^{-\frac{\delta_c^2}{2\sigma(m)^2}}d\ln(\sigma(m)), \tag{A.10}$$

where $n(m)dm$ is the number of objects that have a mass $[m, m+dm]$; $\rho_{\rm M}$, is the mean matter density of the universe (baryonic and dark), and $\sigma(M)$ is the standard deviation of the mass distribution. The abundance of halos at a given redshift are obtained by integrating the mass over-densities that can collapse into virial objects, given the density field is Gaussian and therefore described fully by the linear matter power spectrum. Considering purely spherical perturbations, or a 'top-hat' model, allows us to calculate what over-density (density fraction above the mean) is required for a density field to collapse into a halo - within which star forming galaxies will be able to form. During a matter dominated era, where z_c is the redshift at which the structure is collapsing, the critical over-density is,

$$\delta_c \approx 1.686(1+z_c). \tag{A.11}$$

This mathematical approach determines the statistical distribution of structure in a cold (non-relativistic) dark matter dominated universe, but it double counts over-densities that already exist within larger virialised structures. To remedy this, the integral of over-densities needs

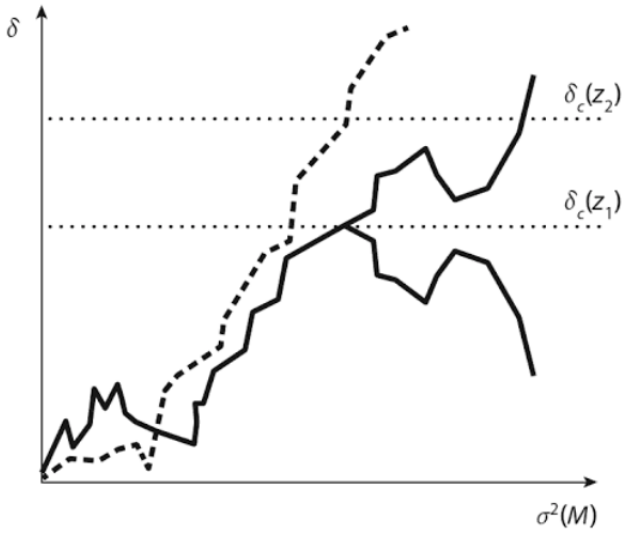


Figure A.3: Two example random walks in density perturbation as a function of mass variance, as part of the excursion-set formalism. The critical density for the structure's collapse threshold is represented by the dotted line (which depends on redshift, Equation A.11). The PS model can be solved with this approach by analysing random walks which cross the black line. Due to the random nature of the walk, just as many trajectories will touch the threshold as pass it, see text for discussion (Loeb & Furlanetto, 2013).

to be halved in the P-S to obtain the result in Equation A.10. It is therefore beneficial to explain the mechanism using the 'excursion set formalism' (Bond et al., 1991), where we view the density field at a given point with a random walk. This implies the smoothing of different mass scales as the random walk in over-density steadily reaches the critical over-density. The mass scale at which the critical over-density is reached counts as the formation of a halo of that mass (see Figure A.3). Running lots of these random-walk excursions leads to the distribution of dark matter halos, i.e. the set of these random walk excursions. At the point of crossing the critical over-density threshold the random walk has an equal chance of crossing the threshold or returning under it - a natural explanation for the factor of two. Another benefit of this approach is that it provides information about the correlations between halos and their different mass scales. Despite the logical inconsistency in the P-S model, it provides an important insight into the physics - the distribution of galaxies closely follows the same distribution as allowing a pressure-less fluid to spherically collapse around density perturbations which are initially generated from a Gaussian random field. A well known improvement to this formalism is the Sheth-Tormen model (Sheth & Tormen, 1999), which uses the collapse of ellipsoidal over-density perturbations instead of spherical ones.

The amalgamation of the FLRW-cosmology and linear perturbation theory is known as $\Lambda-$CDM

	$\Lambda - \mathrm{CDM}$ Parameters
H_0	67.7 ± 0.06 kms^{-1} Mpc^{-1}
Ω_M	0.307 ± 0.006
Ω_Λ	0.689 ± 0.006
n_s	0.966 ± 0.004
σ_8	0.82 ± 0.006
Ω_b	0.0486 ± 0.00014

Table A.1: The 6 $\Lambda - \mathrm{CDM}$ cosmological parameters as recently measured by Planck. H_0, the Hubble parameter is the inverse of the age of the universe; Ω_M, the universe's fractional energy density in matter; Ω_Λ, is the fractional energy density of the vacuum; n_s, the scalar spectral index, which describes how density fluctuations in the CMB vary with scale; σ_8, the matter power spectrum normalisation, a mass variance calculated across 8 Mpc from the observations of galaxies; and Ω_b, the fractional density in baryons (Lahav & Liddle, 2019; Planck Collaboration et al., 2016a).

(a cosmological constant and cold dark matter dominated universe, referring to Equation A.3). This is the concordance cosmology, and is described by six parameters as well as a choice of matter content. The excursion set formalism is a useful tool for quickly describing the matter distribution within $\Lambda - \mathrm{CDM}$. Throughout this work we implement $\Lambda - \mathrm{CDM}$ using the measurements from Planck (Table A.1). The only parameter we haven't mentioned is σ_8, which is the normalisation of the matter power spectrum on a scale of 8 cMpc. Throughout this work redshift is represented by z, distances are co-moving unless specified and we assume no curvature, $k = 1$.

Λ-CDM predicts our universe's cosmology with overwhelming accuracy, but it is not without its pitfalls. We have mentioned the CMB, however its observation is too smooth for a dark energy based expansion. The horizon for particles to interact at this time horizon is expressed by Equation A.5 with $E \propto t^{1/2} \times dz'/dt$, representing the radiation dominated early universe. This leads to ~ 100 Mpc in proper distance, and hence there must be another mechanism to causally relate the equilibrium thermodynamics suggested by the CMB - this is known as the horizon problem. Others issues include the flatness problem, where the sum of the density parameters (Equation A.4) requires $k = 1$ to correctly predict the evolution of our universe; the antimatter problem, which draws attention to the asymmetry in baryonic matter compared to negligible quantities of anti-baryonic matter; the expansion problem where the extreme self-gravity during the big bang must be overcome; and finally an explanation is required to deviate from precisely homogeneous structure formation.

These seemingly daunting problems can be solved with a common solution called inflation

(Guth, 1997). It is theorised that a short inflationary period of the universe's expansion must have occurred shortly after the big bang where a small patch of a larger quantity exponentially grows to become the known universe. If this is the case, the universe can be in thermal equilibrium with itself across the required scales, overcome its own gravitational pull, and baryon number does not have to be conserved within the inflated patch. If a scalar field is responsible, prospects to detect it involve looking for extra heating signatures or a surplus in baryons that the field could have decayed into.

Detecting evidence of the early universe is a tricky business. Other than direct observations of the CMB, the distribution of the galaxies themselves is another useful probe. Dynamics in the early universe lead to a clustering of galaxy distributions known as the baryon acoustic oscillations (BAO). Long after inflation has ended, sound waves in the primordial plasma form due to the opposing baryon-photon pressure and gravitational forces. These waves propagate until recombination occurs ($z \sim 1100$), which causes the sound speed to drop dramatically (Tseliakhovich & Hirata, 2010). The correlation between the CMB power spectrum and the galaxy power spectrum produces a bump around the $\sim 100\ h^{-1}$ Mpc scale with width $\sim 10\ h^{-1}$ Mpc. These represent a characteristic (statistical) scale describing the nature of where structures cluster together and the Jeans scale respectively (Eisenstein & Hu, 1998). Because the difference between the baryonic sound wave and dark matter distribution is preserved through the universe's expansion, the CMB power spectrum traces the primordial baryonic sound wave and the galaxy power spectrum traces the dark matter distribution. Large clouds of neutral hydrogen will contain an imprint of the BAO similarly to the galaxy power spectrum. Star formation in the cosmic dawn is therefore suppressed by the relative velocity of primordial hydrogen and dark matter. This is called the velocity induced acoustic oscillation (VAO) and will be imprinted on the 21cm signal during the cosmic dawn, providing another probe of the early universe (Fialkov et al., 2014a; Cohen et al., 2016; Muñoz, 2019). Precision measurement of the primordial hydrogen will not only help probe the BAO at higher redshift (through the VAO), but it can also help decipher inflation. Whether inflation is driven by one or many scalar fields can be analysed partly by the precise relationships of the photons and baryons that the remaining potential decays into. Subtle detail in this relationship can leave detectable imprints on the 21cm power spectrum during the cosmic dawn (Gordon & Pritchard, 2009; Hotinli et al., prep). There are lots of interesting possibilities surrounding inflation however a deeper analysis into this is beyond the scope of this work, for more detail please see Peacock (1999) or Perkins

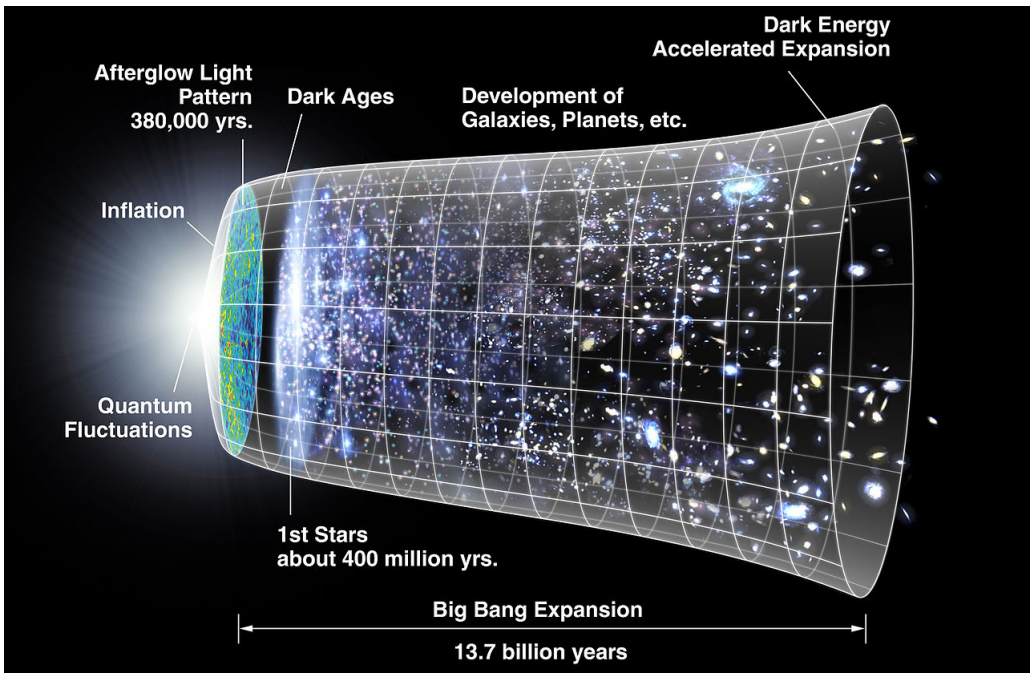


Figure A.4: The currently known epochs of cosmology. The big bang begins on the left, with the Cosmic Microwave Background represented by the 'Afterglow Light Pattern'. The cosmic dawn (the formation of the '1st stars') leads to reionisation, as the radiation form the first stars ionises intergalactic hydrogen (represented by the fading of the purple shade from left to right) (NASA/WMAP (2012), https://map.gsfc.nasa.gov/media/060915/index.html).

(2005) for example.

An overview of the cosmological epochs is shown in Figure A.4. Inflation is thought to have occurred when the universe was 10^{-35} seconds years old, lasting for at most 10^{-31} s. By $z \sim 1090$ the expanding universe has cooled enabling 99.9999% of the universe's baryonic matter to form into hydrogen atoms, (Peebles, 1968; Zeldovich & Sunyaev, 1969). This is thought to remain the case until stars start to form $\sim$ 150Myr after the big bang in the cosmic dawn. As generations of stars grow amidst an increasing background of UV radiation, the neutral hydrogen (HI) is ionised. By $z \sim 6$ galaxies have ionised all $(99.96 \pm 0.03\%)$ the HI in the surrounding inter-galactic medium (IGM) (Fan et al., 2006; Gunn & Peterson, 1965). This is the epoch of reionisation and is the most recent phase change of our universe. Finally we have come to the current, predominantly ionised, state of the universe. The majority of this thesis is aimed at the cosmic dawn and reionisation, whose theory we discuss in Sections 1.2.1 and 1.2.2 at the beginning of the main text. Here, we turn to understanding how such constraints can be made. For this we need observation, the astronomer's equivalent to experiment.

A.2 The current observational picture

The current cosmological picture is bound between two sources of observational data: The CMB, at $z \sim 1100$ and high-redshift quasars (QSOs) around[1] $z \sim 7$ and below. In between these two milestones the universe's first stars form (known as the cosmic dawn), and (re)-ionise the hydrogen left over from recombination (known as reionisation). The cosmic dawn and epoch of reionisation (EoR) are largely untouched in terms of direct observation.

One of the furthest observed QSOs is at $z \sim 7.1$ (Mortlock et al., 2011), meaning the signal has travelled through the IGM for ~ 10 billion years and hence contains information about the universe's history. Their emission also contains large quantities of Lyman-α[2] which is useful for probing reionisation. The majority of cosmological observations are performed on quasars due to their abundance. QSOs also consist of various constituent parts, such as AGN, supernova, stellar remnants and the stars in the galaxy themselves. This means they can have a variety of uses - a classic example is the use of the standardisable properties of type 1a Supernova emission that Hubble (1929) used to measure the expansion of the universe[3]. Observations of the CMB, by definition, pass through all of the observable universe to reach the instrument and therefore contain a vast amount of information. The Boomerang experiment famously measured the statistical clustering of hot and cold patches in the signal across enough of the sky to show the universe as having a flat cosmology (de Bernardis et al., 2000), in favour of an early inflationary period. The CMB reveals a lot about the universe's cosmology - the Planck team have since pinned down the the $\Lambda - \mathrm{CDM}$ parameters with impressive precision through mapping CMB anisotropies across the spherical sky (Planck Collaboration et al., 2016a,b).

This CMB anisotropy power spectrum obtained (Figure A.5) even provides information on the BAO scales mentioned in the previous section, the oscillatory peaks seen help illustrate the scales at which structures cluster in the early universe. Although temperature anisotropies are of order $\sim 2.7 \pm 10^{-5}$K, they map to regions of varying density within the early universe. Over time these over-densities grow until there is enough mass to form virialised structures for the first time. In this section, we take a brief look at what cosmology is constrained, as well as the

[1] Some redshift 10 quasars are beginning to be uncovered with deep surveys (Oesch et al., 2018)

[2] The first excited state to ground emission of a hydrogen atom, $E \sim 10.2$eV (UV).

[3] Interestingly, Lemaitre (1927) discovered this first during his PhD at MIT using pulsating stars known as Cepheid variables but the data was deemed insignificant for a discovery. The physics of Cepheid variable stars was not established until the 1940s, when this result was accepted as an agreement of Hubble's law.

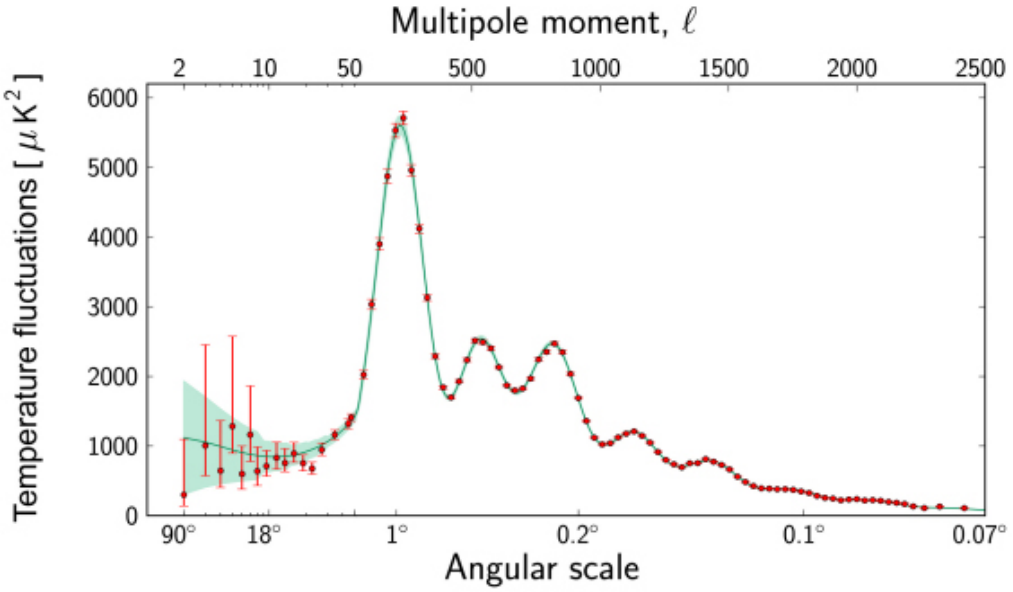


Figure A.5: The power spectrum of the CMB temperature anisotropy as measured by the Planck Collaboration et al. (2016a). Distances between the amplitude peaks in this statistical space are directly related to the clustering that provides the BAO scale (https://sci.esa.int/web/planck/).

limits on the EoR obtained in multiple ways using the CMB and data from QSOs. Please see Peacock (1999) or Loeb & Furlanetto (2013) for more detail.

A.2.1 The CMB

CMB experiments are split into two different categories: first, the temperature anisotropy in the microwave signal relative to the black-body spectrum as measured by experiments like Planck (Planck Collaboration et al., 2016b); secondly, experiments that measure polarisation like the South Pole Telescope (Benson, 2018). To measure both CMB power spectra correctly in either experiment requires the CMB radiation to be seen through reionisation. Hence there is a CMB optical depth that corresponds directly to the timing of reionisation which must be calculated a nuisance parameter along the line-of-sight when observing the CMB. The CMB optical depth, τ, is estimated as,

$$\tau = \int n\sigma dl, \tag{A.12}$$

where σ is the Thomson cross-section, n is the IGM electron density and dl is the comoving distance calculated from today to the reionisation redshift. Thompson scattering in the IGM is caused by the free electrons left after reionisation. These electrons lead to an opacity that

reduces the observed number of CMB photons (when compared to the number emitted at recombination). By measuring the resulting optical depth and assuming an instant reionisation, Planck bounds the midpoint of the EoR as $z = 8.8^{+1.7}_{-1.4}$.

The different polarisation modes of the CMB aim to probe inflation and have the potential to analyse gravitational wave interactions prior to the surface of last scattering. The largest effect however is the Thompson scattering from IGM electrons, which causes about 10% of the CMB radiation to be linear polarised. Recent CMB polarisation observations from the South Pole Telescope suggested a short reionisation, $\Delta z = 1.1^{+1.6}_{-0.7}$ with $\Delta z < 4.1$ at 95% confidence (Reichardt et al., 2020).

Anisotropies and inhomogeneities in density, ionised fraction, and the velocity field ($\mathbf{v}$) of clouds of the electrons in the IGM add artefacts to both observations of the CMB. These can cause secondary effects, such as the kinetic Sunyaev-Zel'dovich effect (Zeldovich & Sunyaev, 1969; Sunyaev & Zeldovich, 1980) where the bulk motion of free electrons will effect the CMB observation. The CMB's temperature fluctuations are expressed as,

$$\frac{\Delta T_{\mathrm{CMB}}(\hat{\mathbf{n}})}{T_{\mathrm{CMB}}} = \int e^{-\tau} \hat{\mathbf{n}} \cdot \mathbf{v} \; dl, \tag{A.13}$$

when observed along the direction $\hat{\mathbf{n}}$ (Park et al. (2013) for example). As reionisation progresses its 'patchiness' (or spatial variability between neutral and ionised hydrogen regions) will increase the second order anisotropy effects. The longer this patchiness exists, the larger the measurable effect is on the CMB since $\mathbf{v}$ will have more directional dependence. Because this also depends on the central reionisation redshift (through τ), it can add a constraint on the EoR's duration.

A.2.2 QSOs

The redshifts of QSOs are identified by least squares fitting various spectra to the observation. When the redshift is known, lots of information about the IGM region local to the QSO can be extracted from the QSO spectrum. The examples of QSO probes for constraining the EoR looked at here are the Gunn-Peterson trough, and the red Ly-α damping wing e.g. Gunn & Peterson (1965) and Miralda-Escudé (1998) respectively.

The IGM is predominately hydrogen (or ionised hydrogen post reionisation), which causes

absorption of the QSO emission as it is redshifted on its journey towards us. At some point the emission that is bluer than Ly-α line will pass through this hydrogen resonance, leading to a streaky pattern in the QSO emission known as the Ly-α forest. If there is enough IGM hydrogen along the line-on-sight the emission will be completely absorbed, leading to what is known as the Gunn-Peterson trough (Gunn & Peterson, 1965). The redshift stamp attached to the observed QSO emission can provide an estimate on the reionisation end point via the dark (no signal) pixels in the Ly-α forest observation. Figure A.6 clearly illustrates this by lining up QSO spectra in order of redshift. Current research (which also include the dark fraction from Ly-β observations) suggests that by $z \sim 6$, reionisation is $99.96 \pm 0.03\%$ complete for hydrogen in the IGM (Becker et al., 2001; Fan et al., 2006; McGreer et al., 2011; Greig & Mesinger, 2017a).

Observations near to the Ly-α line can give information about the size of an HII region, which depends on the ionising luminosity of the embedded QSOs. This requires information on the longer (red) wavelength side of the resonance, known as the red-damping wing. Pedagogically it can be thought of as the QSO emission having to climb through the opacity of its ionised bubble before reaching the end of reionisation. By ignoring recombinations as well as assuming the IGM is iso-thermal and the QSO emission is isotropic we can estimate the radius $R_{\rm HII}$ of the ionised sphere as,

$$R_{\rm HII} \propto \left(\frac{\dot{N}_{\rm QSO} t_{\rm QSO}}{\bar{x}_{\rm HI}} \right)^{\frac{1}{3}}, \tag{A.14}$$

motivated by the ionisation, with a possible maximum of,

$$R_{\rm max} \propto \left(\dot{N}_{\rm QSO} \right)^{\frac{1}{2}}, \tag{A.15}$$

where $t_{\rm QSO}$ is the age of the QSO, $\dot{N}_{\rm QSO}$ is the rate the QSO emits ionising photons. The latter is motivated by the ionisation rate from the UV background becoming dominant over that of the QSO, defining the QSO's sphere of influence. However, when using just Ly-α observation it can be tricky to decipher whether a QSO has reached its luminosity dependant maximum. Other lines are necessary to decipher this mild degeneracy, for example in the Lyman series one can use $2.5 R_{\rm max}[{\rm Ly}\alpha] \approx R_{\rm max}[{\rm Ly}\beta]$ to help resolve this. However this does not overcome sensitivity to scattering throughout the neutral IGM. We have also used a simplistic model to get here, for example the patch of IGM could be over or under-dense, or there could be Lyman

limit systems involved too - more observations at multiple different spectral lines are necessary to do this precisely. Assuming QSO emission is not broadened by the IGM and is far from the Ly-α resonance, we can write the red damping wing optical depth $\tau_{\rm RDW}$ as,

$$\tau_{\rm RDW} \approx \tau_\alpha \bar{x}_{\rm HI} \frac{\Lambda_\alpha}{4\pi^2 \nu_\alpha} \frac{c(1+z)}{H(z)} \left(\frac{1}{R_b} - \frac{1}{R_{\rm e}} \right) \tag{A.16}$$

where Λ_α and ν_α are the Ly-α frequency and decay rate (from Quantum mechanics) and τ_α is the resonant scattering optical depth for HI. $R_{\rm e}$ and $R_{\rm b}$ are the comoving distances from the observer to the end of reionisation and the nearest edge of the QSO's ionised bubble respectively.

The red damping wing is observed as the gradient on the red side of Gun-Peterson troughs in Figure A.6 and can be used to probe the redshift evolution of the sizes of ionised (Strömgren) spheres around the quasars (assuming a step HI gradient at the edge). Equation A.16 is used to match the observed and expected emission, and in the process an estimate of the line of sight neutral fraction is obtained. The variety of different red damping wings in Figure A.6 indicates a patchy end to reionisation. For more information on what information can be achieved from analysing Ly-α radiation and QSOs see Loeb & Furlanetto (2013).

A particularly useful way of analysing the night sky is done by application of the 2nd order statistical moment known as the power spectrum. To measure the statistical clustering (mentioned at the start of this section), one must take a Fourier transform of the correlation function of the signal. We will discuss the statistical origins and some properties of the power spectrum-correlation function relationship in Chapter C but for now, it measures a scale dependent clustering of the quantity at hand. For any quantity I, we can define the fractional perturbation,

$$\delta(x) = \frac{[I(x) - < I >]}{< I >}, \tag{A.17}$$

where x is the spatial position vector, the pointy brackets indicates the mean value (or ensemble average if used upon multiple arguments), bold means vector and the tilde indicates a Fourier transform where the wave vector is k. The power spectrum, $P(\mathbf{k}_1)$, is then defined by the correlation between the two Fourier transformed fractional perturbations at different wave mode scales as,

$$P(\mathbf{k}_1)\delta_D(\mathbf{k}_1 - \mathbf{k}_2)(2\pi)^3 = < \tilde{\delta}(\mathbf{k}_1)\tilde{\delta}(\mathbf{k}_2) >, \tag{A.18}$$

where δ_D is the Dirac delta function. This is typically done using spectral intensity per frequency

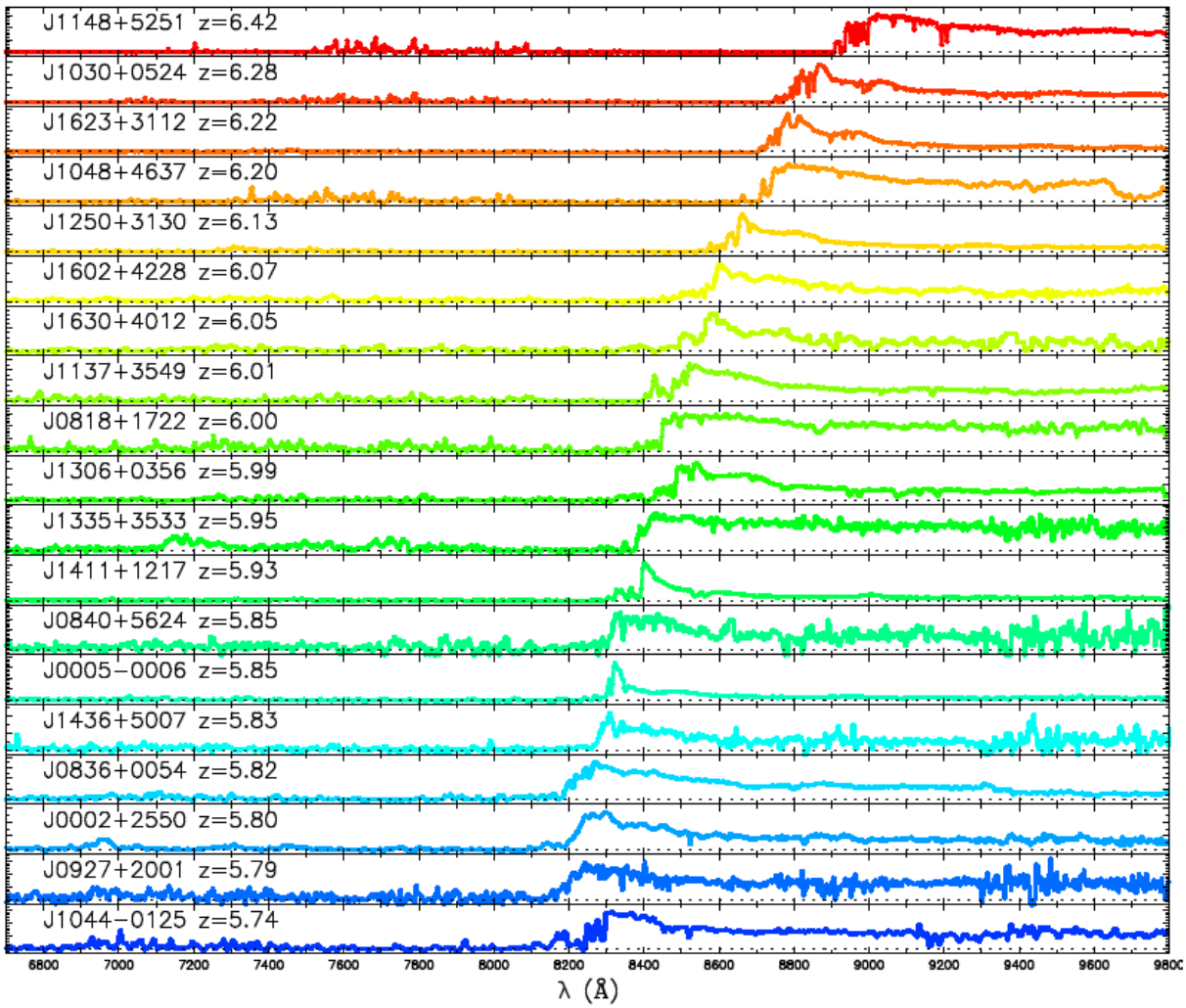


Figure A.6: SDSS (Sloan Digital Sky Survey) data on 19 quasars $z \sim [5.74, 6.42]$. The Gunn-Peterson trough is evident in all cases as the lack of signal to the blue (left) end of each spectrum's maximum. These maxima occurs at the observed Ly-α emission for each quasar hence they scale with redshift. Notice that as redshift decreases, the roughness of the trough decreases - implying more neutral hydrogen in the IGM existed at earlier times. The red damping wing is the gradient of the red (right) side of each trough which contains information about each individual quasar's Strömgren sphere (Loeb & Furlanetto, 2013). (Fan et al. (2006), Figure 1 - © AAS. Reproduced with permission).

inferring a dark matter power spectrum (e.g. when using a constant mass to light ratio, and assuming the observed emission traces the structure accurately). In the case of the CMB, this is done in spherical harmonics as the signal is a series of temperature fluctuations distributed accross the whole 2D spherical sky.

There is a large gap of unobserved universe in between the measurements of the CMB and the observations of QSO spectra - this is where radio astronomy dominates the picture. To tie these observational pieces together, we need the detection of the 21cm brightness temperature. Observing these hydrogen atoms would provide data from when the IGM re-ionises, through the cosmic dawn, up to the release of the CMB as recombination occurs. Current pathfinder experiments such as HERA (Hydrogen Epoch of Reionization Array) and LOFAR (LOw Frequency ARray)[4] are looking for the signal whilst testing technologies and learning about noise removal optimisation. We look at these instruments in more detail in Section 1.3.7 and in Appendix B. These do not provide enough signal to noise to produce a redshift dependent image of reionisation. In contrast the SKA, after its Phase-2 upgrade[5], will hope to do this due to its revolutionary scale. With near to 1000m^2 of receiving area on the ground in the form of $\sim$ 1000000 dipole antennae[6], and with baselines of $\sim$ 40km across the Western Australian desert - it will essentially be the world's most sensitive FM radio. The SKA's impressive detail means that storage alone is estimated to be orders of magnitude larger than the current annual internet stream. This will have to pave the way for big data architectures and machine learning algorithms in astrophysics (Garofalo et al., 2017). We now move onto detailing the 21cm line itself.

A.3 21cm Theory

In 1951 Edward Purcell and Harold Ewen discovered a signal from the galactic spectrum at $1,420.405$ MHz (Ewen & Purcell, 1951b), or 21.106 cm in wavelength. This corresponds to the hyperfine transition of atomic hydrogen. Since its discovery the 21cm line has been used as a tracer of galactic structure, most famously used to map the spiral arms of the Milky way. Using Purcell and Ewen's original data, the rotational velocity of the Milky Way was first measured

[4] http://reionization.org/, http://www.lofar.org/.
[5] http://www.ska.gov.au
[6] A typical house's television antenna is a good example of a type of dipole antenna.

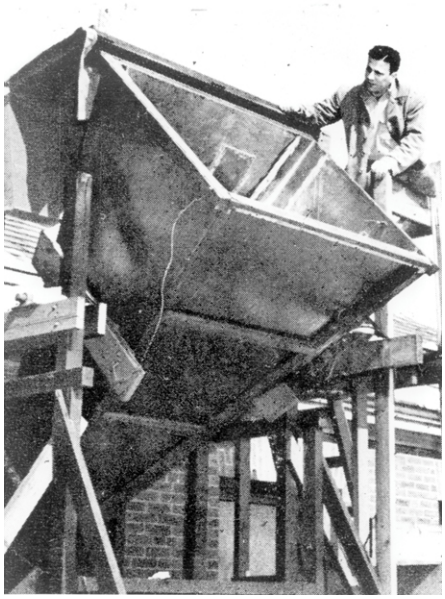


Figure A.7: The radio telescope used by Ewen & Purcell (1951a) to detect the neutral hydrogen signal from within the Milky Way. Due to the unorthodox design of the 'horn antenna' their office at Harvard University became particularly susceptible to flooding during rainstorms (NRAO/AUI/NSF, https://public.nrao.edu/gallery/horn-oplenty-discoveries/).

by Muller & Oort (1951).

The first successful lab detection of the 21cm line was by J. Wittke and R. Dicke in 1954 using a microwave absorption technique (Wittke & Dicke, 1956). Today the hydrogen MASER (Microwave Amplification by the Stimulated Emission of Radiation) is the chosen method for precision measurements of the hydrogen hyper-fine splitting, giving a value of $\nu = 1420405751.7667 \pm 0.0009$ Hz. For more information on the atomic-beam and the hydrogen maser measurement techniques see Foot (2005).

A.3.1 The hyper-fine hydrogen emission line

In brief, this emission line originates from hydrogen's electron flipping spin state with respect to the spin of the proton. The transition is quantified as $\mathbf{F} = 1 \rightarrow \mathbf{F} = 0$, where $\mathbf{F}$ is the total sum of quantum mechanical spin in the atomic system (the sum of the electron and nuclear spin angular momentum, $\mathbf{F} = \mathbf{J} + \mathbf{I}$, respectively).

The Hamiltonian for the hyper-fine structure (HFS) correction is,

$$H_{\rm HFS} = -\mu_{\mathbf{I}} \cdot \mathbf{B}_e, \tag{A.19}$$

where $\mu_{\rm I} = g_{\rm I}\mu_{\rm N}\mathbf{I}$ is the magnetic moment of the nucleus and B_e, is the magnetic flux density caused by the motion of the atomic electron,

$$B_e = \frac{2}{3}\mu_0\{-g_{\rm s}\mu_{\rm B}\mathbf{s}|\psi(r)|^2\}, \tag{A.20}$$

where $\mathbf{s}$ is the spin angular momentum of the electron. The 'Landé g factor' for the proton is labelled, $g_{\rm N}$, and quantifies its gyro-magnetic ratio, the relationship between the proton's spin angular momentum and resulting magnetic field. This magnetic moment is taken as a first approximation to be the interior of a uniformly magnetised sphere in classical electromagnetism, with the magnetisation $\mathbf{M}$ for the atom contained within the curly brackets. Although $\mathbf{s}$-orbitals are spherically symmetric, we must be careful in treating the radial dependency of $\psi(r)$, the electron's wave function. To avoid problems we note that there is a constant maximum, B_e, which we obtain via spatially integrating $|\psi(r)|^2 d^3r$. Since r can vary from 0 to ∞, there must be a cut-off at $r \equiv r_b$ where the distribution outside this sphere no longer contributes to the magnetic field from the nucleus. Luckily for us this happens to be significantly smaller than the Bohr radius and we can therefore make the following approximation,

$$|\psi(r < r_b)|^2 \approx |\psi(0)|^2. \tag{A.21}$$

As we are only concerned with hydrogen here, we only need worry about $\mathbf{s}$-orbitals. Therefore taking the electron's orbital angular momentum to be $\mathbf{l} = 0$ we can use $\mathbf{J} = \mathbf{l} + \mathbf{s}$, to equate the total electron angular momentum to the electron spin angular momentum (since the electron's orbit is spherically symmetrical). Now we are ready to apply Schrödinger's equation to equation A.19, giving,

$$E_{\rm HFS} = \langle\psi|H_{\rm HFS}|\psi\rangle = A\langle\psi|\mathbf{I}\cdot\mathbf{J}|\psi\rangle \tag{A.22}$$

$$= A\{F(F+1) - I(I+1) - J(J+1)\},$$

where the constant $A = \frac{2}{3}g_{\rm I}\mu_{\rm N}\mu_0 g_{\rm s}\mu_{\rm B}|\psi(0)|^2$ has been defined for convenience. Now to evaluate the emission of this transition from $F_{\rm before} = 1$ to the $F_{\rm after} = 0$ we simply evaluate $E_{\rm HFS}$ for both $\mathbf{F}$ values and subtract,

$$\Delta E_{\rm HFS} = \langle\psi(F=1)|H_{\rm HFS}|\psi(F=1)\rangle - \langle\psi(F=0)|H_{\rm HFS}|\psi(F=0)\rangle \tag{A.23}$$

$$= \frac{A}{2}\Delta\{F(F+1)\} = A \equiv E_{10}.$$

Inputting the appropriate constants for a hydrogen atom gives a value of 1.42Ghz, extremely close to the measured value. This theoretical prediction was first done in 1944 by H. van de Hulst (van de Hulst, 1951). For more details, and a generalisation of $\mathbf{l} > 0$, see (Foot, 2005). E_{10} will be referred to in the context of radiative transfer.

A.3.2 Radiative transfer and the brightness temperature

We describe the radiative transfer across an atom with the change in specific intensity at a given observed frequency I_ν as,

$$\frac{dI_\nu}{dl} = \frac{\phi(\nu)h\nu}{4\pi}[n_1 A_{10} + (I_\nu n_1 B_{10} - I_\nu n_0 B_{01})], \tag{A.24}$$

where h is Plank's constant, $\phi(\nu)$ is a line profile (or broadening function per frequency normalised to 1), n_i is the population density of the atomic energy levels (i=0 being the ground state); A_{ij} and B_{ij} are the Einstein spontaneous and stimulated transitional coefficients, typically related by $B_{10}g_1 = g_0 B_{01}$ and $A_{10}c^2 = 2h\nu^3 B_{10}$ (Foot, 2005). These relate respectively to the change in radiation intensity (at a transition $i \to j$) over a proper path length dl. Hydrogen's 21cm spontaneous emission coefficient is $A_{10} = 2.85 \times 10^{-15} s^{-1}$, and B is easily calculable knowing that the transition is from a triplet to a singlet state, giving $g_1/g_0 = 3$ for the quantum degeneracy ratio. For the relative population of states, we use the Boltzmann equation to define the spin temperature T_S as,

$$\frac{n_1}{n_0} = 3e^{\frac{-E_{10}}{k_B T_\mathrm{S}}}, \tag{A.25}$$

and since the 21cm emission temperature is significantly smaller than the other temperature scales of interest ($T_\mathrm{S} \gg E_{10}/k_B$ and $T_\mathrm{CMB} \gg E_{10}/k_B$) this equation can be Taylor expanded to first order as a good approximation. We are also dealing in radio (low) frequencies and therefore are able to apply the Rayleigh-Jeans approximation to the resulting black body radiation curve. This enables us to define our specific intensity in a more intuitive way, using its corresponding brightness temperature as,

$$T_b \approx \frac{I_\nu c^2}{2k_B \nu^2}. \tag{A.26}$$

To apply this to cosmology we need to consider the CMB (introduced in Section 1.2). In the IGM, the hydrogen hyper-fine transition is excited by the CMB. This is the back-light that drives the 21cm spin-flip transition, but first we define the optical depth as in Equation A.12, where the opacity τ of a medium of length l, number density n of particles with cross-section σ. With respect to observed frequency, we can input the radiative transfer information here as,

$$\tau_\nu = \int \frac{\phi(\nu)h\nu}{4\pi}(n_1 B_{10} - n_0 B_{01})dl. \tag{A.27}$$

Since our signal is dependent on the intensity of the CMB it makes sense for us to define δT_b, the brightness temperature in difference to the CMB radiation at that redshift. By using the optical depth (with respect to observed frequency τ_ν) we can write,

$$\delta T_b(1+z) = T_{\rm S}(1 - e^{-\tau_\nu}) + T_{\rm CMB}e^{-\tau_\nu} \approx (T_{\rm S} - T_{\rm CMB})\tau_\nu, \tag{A.28}$$

where $T_{\rm CMB}$ is the temperature of the CMB during the radiative transfer process.

Now all that is required is solving Equation (A.27) for the opacity, which can be done in two ways. The first method requires relating the differential path length to cosmological expansion ($dl = -cdz/[(1+z)H(z)]$) and then using the relationship between observed and light emitted at a given redshift. Alternatively we can use the Sobolev approximation (Sobolev, 1957), where we assume the velocity gradient of the gas is high enough so that a local velocity profile is negligible, hence $ldv = vdl$ when combined with Doppler shift. We also assume some natural (or lifetime) broadening, $\phi(\nu) \propto 1/\Delta\nu$, exists in the emission. The first calculation was done by Field (1958, 1959) which gives the brightness temperature difference as,

$$\delta T_b \approx 9 x_{\rm HI}\,(1+\delta_b)\left(1 - \frac{T_{\rm CMB}}{T_{\rm S}}\right)\left(\frac{H}{\frac{dv_r}{dr} + H}\right)(1+z)^{\frac{1}{2}}\ \ {\rm mK}, \tag{A.29}$$

where δ_b is the mass fraction of baryons and $x_{\rm HI}$ is the mean fraction of neutral hydrogen in the IGM (see Zaroubi (2013) or Loeb & Furlanetto (2013) for a modern explanation).

The main influences of the 21cm signal are down to the IGM neutral fraction and the baryonic density field. The spin temperature is also clearly important in dictating whether the signal is in emission or absorption. When $T_b < 0$, or equivalently $T_{\rm S} < T_{\rm CMB}$ this signal is interpreted as absorption. We need $T_{\rm S} > T_{\rm CMB}$ for emission to occur, and the signal becomes saturated as

$T_{\rm S} \gg T_{\rm CMB}$. The final two terms show the dependency on the cosmology and peculiar velocities of the gas. We assume a constant cosmology throughout every chapter.

A.3.3 The 21cm spin temperature

The spin temperature is dictated by several processes. Firstly, the CMB photons cause resonant scattering and absorption as well as spontaneous emission (represented by the aforementioned Einstein A and B coefficients). Secondly the scattering of UV photons (similarly represented with P_{ij}) can cause ionisation, and add kinetic energy to each atomic system (heating the hydrogen gas in the IGM) in some cases. But on average, the UV photons cause the Wouthuysen-Field effect - a resonant scatter which couples the kinetic gas temperature to the UV radiation colour temperature (discussed further in Section 1.2.3). Finally there are also collisions (C_{ij}) between hydrogen atoms as well as free electrons or protons. We can equate the transition of the hydrogen hyper-fine spin-states by summing these contributions as,

$$n_1(C_{10} + P_{10} + A_{10} + B_{10}I_{\rm CMB}) = n_0(C_{01} + P_{01} + A_{01} + B_{01}I_{\rm CMB}), \tag{A.30}$$

which is solved for the spin temperature using a few approximations. First we Taylor expand equation A.25 into,

$$\frac{n_1}{n_0} \approx 3\left(1 - \frac{E_{10}}{k_B T_{\rm S}}\right). \tag{A.31}$$

We can also apply the principle of detailed balance to the Boltzmann equation and Taylor expand a relation for the collisional coefficients using the kinetic gas temperature, T_K, as,

$$\frac{C_{01}}{C_{10}} \approx 3\left(1 - \frac{E_{10}}{k_B T_K}\right). \tag{A.32}$$

The coupling coefficients for the colour temperature, T_α, for the UV radiation field are defined following suit as,

$$\frac{P_{01}}{P_{10}} \equiv 3\left(1 - \frac{E_{10}}{k_B T_\alpha}\right). \tag{A.33}$$

Now turning to the absorption and stimulated emission terms we combine the Einstein A and B coefficients with the Raleigh-Jean approximation applied to the CMB spectrum to get,

$$B_{01}I_{\rm CMB} = A_{10}T_{\rm CMB}\frac{k_B}{E_{10}}. \tag{A.34}$$

Substituting Equations A.34, A.31, A.32 and A.33 into A.30 yields the spin temperature in the desired form,

$$T_{\rm S}^{-1} = \frac{T_{\rm CMB}^{-1} + x_k T_k^{-1} + x_\alpha T_\alpha^{-1}}{1 + x_k + x_\alpha}, \tag{A.35}$$

where x_α and x_k are the coefficients for the UV radiation field coupling and the collisional coupling respectively, defined as,

$$x_\alpha \equiv \frac{E_{10} P_{10}}{k_B A_{10}} \ , \qquad \text{and} \qquad x_k \equiv \frac{E_{10} C_{10}}{k_B A_{10}} \ . \tag{A.36}$$

Since we have expressed the brightness temperature as a difference from the CMB (Equation A.29) it is worth mentioning that if the CMB was the only source, then $T_{\rm S}$ and $T_{\rm CMB}$ would reach an equilibrium in the timescale,

$$\frac{E_{10}}{k_B T_{\rm CMB} A_{10}} \ \approx \ \frac{3 \times 10^5}{(1+z)} \text{ years.} \tag{A.37}$$

The spin temperature is in equilibrium with the CMB temperature from recombination (when hydrogen first forms) until the cosmic dawn which we discuss in Section 1.2.1.

We have now covered the detail necessary for looking closely at the role of the 21cm signal in unveiling the cosmic dawn and reionisation. The behaviour of the spin temperature can be described by its coupling to the CMB, collisions and Lyman-α radiation (via Equation A.35). The evolution of the 21cm signal brightness depends heavily on the density field (Equation A.29) but importantly: it is dominated by IGM neutral fraction at late times; and by the spin temperature at early times. A detailed description of the physical processes that guide the 21cm brightness temperature signal is provided in Chapter 1.

Appendix B

Introductory Radio Astronomy

Section B.1 introduces radio receivers and their application in interferometers. We provide the basics of interferometry for use in astrophysical observations, as well as a derivation of the flat-sky approximation used in calculating the 21cm power spectrum. In Section B.2 we discuss the application of radio interferometers to observing the 21cm power spectrum and the formulation of the foreground wedge via the delay transform. A detailed overview of the current endeavours to observe the cosmological 21cm signal is provided in Section B.2.2.

B.1 Radio observations

Here we introduce the basics of radio observations before looking into the specifics of interferometers and the simulations of 21cm observations used in the main text. This section is largely based on the text by Burke & Graham-Smith (2002).

B.1.1 An introduction to receivers

The fundamental principle behind any observing device is that the signal power received, $P_{\rm rec}$ depends on the size of the instrument, $A_{\rm eff}$ (effective collecting area), and the signal flux, S,

$$P_{\rm rec} = A_{\rm eff} S. \tag{B.1}$$

$G(\hat{\mathbf{n}})$ is the voltage gain into the instrument which depends on the direction it is pointed, $\hat{\mathbf{n}}$. It therefore is normalised to comply with energy conservation when integrated over a solid angle as,

$$\frac{1}{4\pi}\int_0^{4\pi} G(\hat{\mathbf{n}})d\Omega = 1, \tag{B.2}$$

giving, $G = \frac{4\pi}{\Omega_{\text{inst}}}$, where Ω_{inst} is defined approximately as the field of view of the instrument. Unfortunately our instrument will not convert the gain to the desired signal with perfect efficiency. We parameterise this by introducing a transmitted power, P_{tr}, defined in terms of the flux as,

$$P_{\text{tr}} = \int_0^{4\pi} S d\Omega, \tag{B.3}$$

and relate the gain to the instrument's signal as,

$$S(\hat{\mathbf{n}}) = \frac{G(\hat{\mathbf{n}})P_{\text{tr}}}{4\pi r^2}, \tag{B.4}$$

where r relates to the instrument via Ω_{inst}. Now we have a general result for the effective instrument area,

$$A_{\text{eff}} = \frac{\lambda^2 G}{4\pi}. \tag{B.5}$$

Specifically for radio astronomy we must look at the properties of antennae, which collect the signal. In particular the temperature, T, of the antennae is influenced by the absorbed radiation flux and must be considered. We relate the specific flux intensity B_ν, and the radiation density, u_ν, as $\frac{cu_\nu}{4\pi} = B_\nu$. From thermodynamics we have,

$$B_\nu d\nu = \frac{8\pi h}{c^3}\frac{\nu^3 d\nu}{\exp(h\nu/k_{\text{B}}T) - 1}, \tag{B.6}$$

and since these are radio waves the typical energies are low enough for the Rayleigh-Jeans approximation,

$$B_\nu d\nu = \frac{2k_{\text{B}}T d\nu}{c}. \tag{B.7}$$

Applying this to what we have learnt for a general instrument (Equation B.5), we write the power per frequency as,

$$P_\nu d\nu = \frac{1}{2}\int B_\nu A_{\text{eff}} d^2\Omega d\nu = 2\pi A_{\text{eff}}\frac{2k_{\text{B}}T\nu^2}{c^2}d\nu. \tag{B.8}$$

To dig deeper, we now define the contributions to thermal noise in terms of an antenna temperature, T_a, the 21cm brightness temperature T_b, (derived in Section 1.2). The T we have been using is the black body temperature for a specific intensity emitted at ν. The attenuation reduction (signal loss), α, is defined,

$$T_a = (1 - \alpha)T_b + \alpha T, \tag{B.9}$$

expressing the antennae temperature as a contribution of signal and its own thermodynamics.

Now we turn to interpreting the received signal in terms of the electromagnetic field,

$$E(x, y) = \hat{\mathbf{a}} E_0 e^{2\pi i(\nu t - kz)}, \tag{B.10}$$

where we have assumed the earth is flat so that our antennae align with the z axis ($\hat{\mathbf{a}} = \hat{\mathbf{z}}$). Equation B.10 is then the received electric signal with $\hat{\mathbf{a}}E_0 = (E_x, E_y, 0)$ defining the plane of polarisation. This is required in order to apply the flat sky approximation later on (Equation B.50). Polarisation is conventionally described via the Stokes parameters, I, Q, U, and V, defined as,

$$I = E_0^2 \tag{B.11}$$

$$Q = E_0^2 \cos 2\beta \cos 2\psi \tag{B.12}$$

$$U = E_0^2 \cos 2\beta \sin 2\psi \tag{B.13}$$

$$V = E_0^2 \sin 2\beta, \tag{B.14}$$

with orientation angle ψ, and axial ration β describing the ellipticity of the polarisation (i.e. $\beta = \pm\frac{\pi}{4}$ for circular polarisation or $\beta = 0$ for a plane wave). An alternate parameterisation for the Stokes parameters is,

$$I = E_x^2 + E_y^2 \tag{B.15}$$

$$Q = E_x^2 - E_y^2 \tag{B.16}$$

$$U = 2E_x E_y \cos(\phi_x - \phi_y) \tag{B.17}$$

$$V = 2E_x E_y \sin(\phi_x - \phi_y), \tag{B.18}$$

which can be obtained using,

$$E = \hat{\mathbf{a}} E_0 e^{i\omega t(\nu t - kz)}. \tag{B.19}$$

Both parameterisations are equivalent and satisfy $I^2 = Q^2 + U^2 + V^2$. Later on, in order to describe the sampling pattern of the telescope, we use the Fourier transform of the sky's U-V plane. As well as taking advantage of what's called the 'cross polar product', defined as,

$$\begin{bmatrix} < RR^* > & < RL^* > \\ < LR^* > & < LL^* > \end{bmatrix} \equiv \begin{bmatrix} I + V & Q + iU \\ Q - iU & I - V \end{bmatrix}, \tag{B.20}$$

where R and L are an orthogonal basis, representing the right and left handed description of circular polarisation (used for aperture synthesis).

Now that we have an antenna and the theory to analyse it, we look at the quality of signal that can be detected. The fundamental uncertainty of any measurement cannot be smaller than the noise within the system. To reduce this uncertainty the observation time, t_{obs}, effects the total noise as, $(Bt_{\text{obs}})^{-\frac{1}{2}}$, where B is the bandwidth (discussed in more detail in Section 1.3.7).

Typically the power in noise, N, is modelled as a Gaussian random probability density function,

$$N = \frac{e^{\frac{-V^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}, \tag{B.21}$$

where, V, is voltage. We are only capable of estimating the power spectrum, with errors dictated by our finite observation. To do this we need to define the auto-correlation as (from $t' = 0$ at time τ),

$$R_t(\tau) = \int_{-t/2}^{t/2} \nu(t') V(t' + \tau) dt', \tag{B.22}$$

i.e. for an instantaneous measurement ($\tau = 0$), we are left with the Gaussian random noise variance averaged across the instrument ($R \to \langle \sigma^2 \rangle$). Assuming this time span is long enough, our Gaussian random noise will not correlate with itself. Using this, we can take the Fourier transform (tilde) to produce the signal power spectrum as,

$$\tilde{R}(\tau) = P_\nu. \tag{B.23}$$

A typical linear receiver will consist only of an ideal amplifier and a passive filter. Here, we

write the auto-correlation as a Dirac δ-function, $R(\tau) \rightarrow \sigma^2\delta(t)$. However we have not yet considered constraints from the bandwidth. We define the impedances with the input load, Z_i, and output load, Z_0. For our simple apparatus the impedances are matched, $Z_i = Z_0^*$. To follow suite, we define the signal amplitude, $V_i(t)$, the output amplitude, $V_0(t)$, and the input filter amplification, $V_1(t)$. Allowing us to write the gain as,

$$G = \frac{V_0^2}{V_i^2}\frac{Z_i}{Z_0}. \tag{B.24}$$

We can relate the signal and output via the definition of a transfer function, $H(\nu)$ which accounts for undesired phase shifts in the signal. $|H| = 1$ for a complex H is desired in order to minimise signal loss and allow an accurate Fourier transform to be performed. Typically the response of the instrument will result in a ringing structure or beat pattern. The resonant frequency, ν_0 of this beat pattern depends on the bandwidth as, $B \sim \nu_0$, hence a larger bandwidth will reduce the length of the ringing and provide a more accurate measurement. A time averaged power, P, is needed, as we can unfortunately never have an infinitely wide bandwidth,

$$\langle P \rangle = \int_{-t/2}^{t/2} [V_0]^2 dt. \tag{B.25}$$

By averaging, we can approximate the power spectrum S_d with, N, the detector response for a white noise signal to give,

$$P_d(\nu) \approx N^2\delta(\nu) + 2P(\nu_0) * H(\nu_0). \tag{B.26}$$

The quality of filter can be associated with the response to unit impulses $H(\nu_i)$, which requires a response time τ (the same lag as used in Equation B.22). We can now write the uncertainty in measurement, ΔT, for measuring the system noise temperature, T_s as,

$$\Delta T = \frac{T_\mathrm{s}}{\sqrt{B\tau}}. \tag{B.27}$$

Maximising the bandwidth and response time is necessary for a good measurement (minimising ΔT). This is obvious for the, B, but for τ the subtlety is that we want to maximise the correlation in the output with τ in order to measure the power spectrum accurately. The longer this lag, the longer the instrument has to make the measurement, and the more precisely it can

be done.

The system noise temperature, T_s, can be broken down for analysis into constituent parts,

$$T_s = T_c + T_r + T_n, \tag{B.28}$$

where $T_{\rm n}$ comes from the receiver itself (noise); $T_{\rm c}$ from the desired (celestial) source; and $T_{\rm r}$ accounts for the rest, i.e. unwanted observations from foreground objects, atmospheric effects, as well as system losses. Typically, pointing the radiometer towards and away from the the desired source is enough to calibrate the system well enough to produce a $> 5\sigma$ detection.

Achieving this in practice is difficult. An ideal radiometer will have an rms (root mean square) measurement following Equation B.27 (assuming units of temperature). To generalise this we write the flux sensitivity as,

$$\Delta S = \frac{2k_B T_{\rm n}}{A_{\rm eff}\sqrt{B\tau}}, \tag{B.29}$$

where we made use of $\frac{SA}{2}$, the received power density per polarisation. A conventional technique, applied in what are called 'heterodyne' receivers, is to switch between a black body reference within the device and the detector. The 'Dicke' radiometer shown in Figure B.1 is a typical example if this in application, where $T_{\rm a}$ is the antenna temperature and $T_{\rm ref}$ the temperature from the reference load, which is typically a familiar resistor. Phase sensitivity of the detector is interpreted through, $T_{\rm ref}$: if $T_{\rm a} = T_{\rm ref}$, there is no signal; $T_{\rm a} > T_{\rm ref}$ is a positive detection; and $T_{\rm a} < T_{\rm ref}$ is a detection with opposite sign.

We now have the basics for a simple radiometer, which can be split largely into two categories:

- Spectrometer - measuring spectra directly (e.g. measuring the intensity per frequency for the radio lobes of galaxy jets).
- Interferometer - or cross correlation radiometry, which uses interference (e.g. inferring an object from its shadow with Fraunhofer diffraction).

Generally the simplest 'digital auto-correlation spectrometer' needs to be satisfied by two concepts. The first of these is the Nyquist sampling criterion. This is required to validate our noise assumptions throughout this section as, relating bandwidth to the frequency of measurements as $\nu_{\rm Nyquist} = 2B$. Secondly we must try to hide the discrete nature of the measurement in

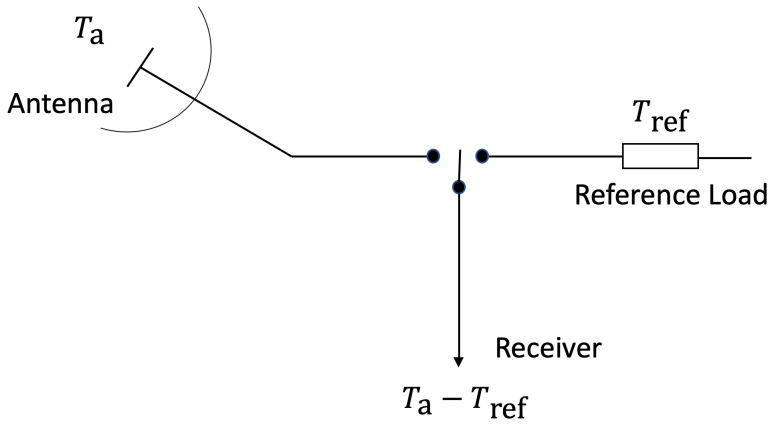


Figure B.1: A simplistic Dicke radiometer with a switch between the antenna, collecting the signal, and a known reference load. T_a is the antenna temperature and T_{ref} the temperature of the reference load. The Dicke radiometer or Dicke switch alternately measures the antenna and reference load in order to reduce systematic from the receiver itself.

order to perform a successful Fourier transform. Therefore we rewrite the correlation function in Equation B.22 as,

$$R_n = \frac{\sum_{i=1}^{M} v(t_i)v(t_i + n\delta t)}{M}, \tag{B.30}$$

where M samples, δt apart, are taken. The Van Vleck approximation is conventionally used to approximate the Gaussian noise per bit, $R_{b,n}$. Using the first term of the summation in Equation B.30 denoted with the superscript (1), we write,

$$R_{b,n} = R(0) \sin\left(\frac{\pi R_n^{(1)}}{\pi}\right). \tag{B.31}$$

If done correctly most of the signal will be contained in the first mode of the sum, meaning this will provide a good approximation. A discrete power spectrum is now obtained, and its quality will depend on the number of samples used. To further the effectiveness of this method, a window function, w can be used to reduced discretisation errors e.g. 22% of a sinc function will remain in terms $i > 1$ of Equation B.30. The result is defined via a convolution as,

$$R_w(\tau) = \int w(\tau')R(\tau' + \tau)d\tau'. \tag{B.32}$$

The Hanning function, $w_H = 1 + \cos(2\pi\tau')$, is the typical window function of choice and is capable of reducing the aforementioned sinc function losses down to 2.6%.

The simplest interferometer is the cross correlation between two radio receivers, which we will

look at in detail in Section B.1.2. Here we generalise the results we have so far for this case. Its output will be a complex number, containing angular as well as spectral information. If we have signal from two antennae, $x(t)$ and $y(t)$, they can be aligned to capture the polarisation information. The distance between these antenna, or baseline, represents a projection into the U-V plane (defined in Equations B.11 or B.15). Their cross correlation function is defined as,

$$R_{x,y} = x(t) \otimes y(t), \tag{B.33}$$

introducing $\otimes$ in place of Equation B.22. As before, this is Fourier transformed for their power spectrum, $S_{x,y}(\tau)$. In practice interferometers will use lots of antennae which are combined to produce a cross-correlation data set, $R_{x,y}(u, v, \tau)$ in U-V space, which is Fourier transformed to $S(\theta, \phi, \nu)$, giving the angular distribution of brightness per frequency for the desired source. What we have learned in Equation B.23 can be transferred to larger instruments as each antenna is a radiometer, requiring its own auto-correlation analysis. With the use of geometric means we can write the noise on interferometer measurements as,

$$\Delta S_{x,y} = 2k_{\mathrm{B}}\sqrt{\frac{T_{n_1}T_{n_2}}{2A_{\mathrm{eff},1}A_{\mathrm{eff},2}B\tau}}. \tag{B.34}$$

Similarly, the discretisation of our data cube (analogous to Equation B.30) is represented as,

$$R_{x,y,n} = \frac{\sum_{j=1}^{M} v_x(t_j)v_y(t_j + n\delta t)}{M}, \tag{B.35}$$

bringing us up-to-date for digital cross-correlation performance.

B.1.2 Radio telescopes

For use in astronomy we need to use the radio receiver as part of a telescope. Radio waves are typically focused with a parabolic receiver, or dish, with focal length f, aperture opening of diameter D, and focal point $\frac{f\lambda}{D}$ away from the dish. When the wavelength is long enough, the ground can be approximated as the parabolic receiver. For a dish, the telescope can be easily aimed at objects. When antennae are placed on the ground, as is often the case for an interferometer, this behaviour can be mimicked as part of the correlation timing analysis. A phase delay is added to one side and subtracted from the other with what is known as the

Butler matrix. This precisely reproduces the tilted dish scenario with ground based antennae.

The beam is analysed from the geometry of the telescope in the small angle approximation. The deviation of current along a given aperture (or grading), $I(x, y)$, for a given wavelength, λ, produces the radiation pattern,

$$F(\theta, \phi) = F_e(\theta, \phi) \int\int_{4\pi} e^{-\frac{2\pi}{\lambda} iz\theta} I(x, y) d\Omega, \tag{B.36}$$

where Ω represents the solid angle with conventional polar components θ and ϕ, or x and y. Polarisation and effects of the instrument are bundled into F_e. The radiation pattern and aperture distribution are related via Fourier transform. So is the power gain and auto-correlation of the of the aperture current density distribution,

$$\tilde{G}(\theta, \phi) = I \otimes I. \tag{B.37}$$

The efficiency of a receiver can be expressed as the ratio of observed area $A_{\rm sky}(\cos\theta, \cos\phi)$, to the geometrical size of the opening. If $S(\cos\theta, \cos\phi)$, represents the sky brightness distribution, the received power in Equation B.5 is generalised to,

$$P_\nu = \int_{4\pi} B \cdot A_{\rm sky}\ d\Omega, \tag{B.38}$$

which in terms of brightness and antenna temperatures is,

$$T_{\rm a} = \lambda^{-2} \int_{4\pi} T_{\rm b} A_{\rm sky}\ d\Omega. \tag{B.39}$$

Single aperture telescopes are useful in producing sky maps (like those used in Section A.2). The loss in detail of scanning the sky be expressed in terms of Fourier coefficients as,

$$\tilde{T}_{\rm a} = \tilde{T}_{\rm b} \cdot \tilde{A}, \tag{B.40}$$

which is conventionally performed in units of λ. $\tilde{A}(U, V)$ is known as the telescope transfer function, e.g. the behaviour of a grating array is obtained by the sum of Dirac δ-functions (one for each small aperture).

We are now equipped to consider the simple two antenna interferometer. With radio frequencies,

single-aperture telescopes suffer badly from diffraction effects. The interferometer is therefore the tool of choice for observing radio waves. Using two detectors for interferometry is akin to Young's double slit experiment for optical wavelengths - fringe amplitude is desired over total power. To start we must define the direction of observation (along the antennae), $\hat{\mathbf{a}}$, and the baseline vector between the antenna, $\mathbf{b}$. The timing lag between the telescopes due to the geometry of their aim, $\tau_{\rm g}$, is,

$$\tau_{\rm g} = \frac{\mathbf{b}\cdot\hat{\mathbf{a}}}{c}. \tag{B.41}$$

For a monochromatic source of frequency, ν, the two antenna receivers will output voltages, $x(t) = V_1\cos(2\pi\nu t)$ and $y(t) = V_2\cos(2\pi\nu t - \tau_{\rm g})$. Using B.33 and time averaging the voltages we obtain,

$$R_{x,y}(\tau_{\rm g}) = A_{\rm sky}(\hat{\mathbf{a}})S\cos\left(2\pi\nu\tau_{\rm g}\right). \tag{B.42}$$

Expressing $\mathbf{b}$ in units of wavelength ($\mathbf{b}_\lambda$) and applying the small angle approximation we define the cross-spectrum power density as,

$$S_{xy}(\nu) \equiv \tilde{x}\tilde{y}^* = A_{\rm sky}(\hat{\mathbf{a}})Se^{(i2\pi\mathbf{b}_\lambda\cdot\mathbf{s})}. \tag{B.43}$$

This is the complex analogue of Equation B.42 and whether this is done in time or frequency units depends on convenience. Now to include the interferometer's ability to point, we use a reference direction or tracking centre, $\mathbf{a}_0$,

$$\mathbf{a} = \mathbf{a}_0 + \sigma, \tag{B.44}$$

where σ must be small (and orthogonal to $\hat{\mathbf{a}_0}$) in order for our geometrical approximations to hold. The telescope time delay (τ_i, which has been set to zero until now) can be simply adjusted in reference to this centre providing an approximate switch between the effective area of our instrument, $A(\mathbf{a})$, and the observed area on the sky, $A_{\rm sky}$. The source will also have an associated bandwidth, $B_{\rm s}$, which will effect the beam pattern. Assuming the flux varies slowly throughout the bandwidth (centred on ν_0); as long as $\theta, \phi \ll \frac{\nu}{B_{\rm s}b_\lambda}$ in the field of view and the single frequency response (Equation B.43) is generalised by integrating over the range of $B_{\rm s}$ to produce,

$$S_{xy}(\nu_0, \mathbf{a}) = A(\nu_0, \mathbf{a})S(\nu_0)B_{\rm s}{\rm sinc}[B_{\rm s}(\tau_{\rm g} - \tau_i)]e^{-2\pi i\nu_0\tau_{\rm g}}. \tag{B.45}$$

Notice the fringe oscillations exponentially reduce the delay beam term (sinc function) if this

delay is at least $1/B_{\rm s}$. The tracking centre is often chosen so its delay is matched to that of the instrument for this reason. Within this limit, the complex visibility is defined as,

$$V_{ij} \equiv S_{xy}(\mathbf{b}_{ij} \cdot \mathbf{a}_0 - \nu\tau_i), \tag{B.46}$$

where i and j represent any pair of antenna in a larger array, allowing the techniques learned here to be applied to interferometry with more than two dipoles. For an array of N antenna, there will be a set of $N(N-1)/2$ visibility measurements. Using Equation B.45 the visibility are expressed as,

$$V_{ij} = \int_{4\pi} B_\nu A_{\rm eff}\left(\frac{\mathbf{a}}{\mathbf{a}_0}\right) e^{2\pi i \mathbf{b}_{ij}\cdot\sigma} d\Omega, \tag{B.47}$$

where deviations from the above methodology are from adjusting the units (i.e. using a relative antenna area normalised to the tracking centre, and $B_{\rm s}$ will integrate to 1.). The flux (B_ν, assumed to be a point source until now) is integrated over solid angle to capture the finite size of the source. The amplitude and phase of the complex visibility are the desirable objects for an interferometric observer. They are recorded directly from antennae as time-stamped voltages which contain all of the instrument's information. Conventionally coordinates (l, m, n) are defined as the cosines of $\mathbf{a}$ so that ω is perpendicular to the U-V plane as,

$$V_{ij}(\mathbf{a}_0, u, v) = \int_{4\pi} A_{\rm eff}\left(\frac{\mathbf{a}}{\mathbf{a}_0}\right) B_\nu(l, m) e^{2\pi i(ul+vm+wn)} d\Omega. \tag{B.48}$$

Since we are already using small angles and small deviations from the tracking centre, it is not extreme to assume that the sky is flat too ($w \approx 0$). We can therefore write

$$d\Omega \approx d\theta d\phi = \frac{dldm}{\sqrt{1-l^2-m^2}}, \tag{B.49}$$

finally giving,

$$V(u, v) \approx \int\int_{\rm Source} B(x, y) e^{2\pi i(ux+vy)} dxdy, \tag{B.50}$$

conventionally known as the flat sky approximation.

To summarise,

$$V(u, v) = \tilde{B}(x, y). \tag{B.51}$$

i.e. the visibility measurements of an interferometer are desirable because they are the Fourier transform of the observed source's flux. Interferometry is a powerful technique and this is

an introduction to its foundation. In order to explore further the astronomical application in detail, a variety of intricacies are required. To briefly touch on a few examples: the curvature of the earth must be taken into account in the time delay calculations if long enough baselines are used; or combining non identical radiometer types requires better mapping of the geometrical means used in Equation B.34; and most modern low redshift interferometers require relaxing the flat sky approximation. The rotation of the earth can even be intricately mapped to better fill the U-V data cube, effectively increasing the number of unique baselines between the antennae. For more information on radio observations please see Burke & Graham-Smith (2002) or e.g. Wilson et al. (2013) and references there-in.

B.2 Detecting the 21cm signal

To map the spin-flip background we ideally must observe the redshifted 21cm brightness temperature between redshifts $z \in [6, 50]$ (Loeb & Furlanetto, 2013). This requires observations in the range $\nu \in [30, 200]$MHz. The lower end of this range is limited by the ionosphere, which becomes opaque above this frequency. To measure the 21cm signal successfully, the radio sources in the astrophysical foreground must be accounted for accurately. At best one hopes the astrophysical foregrounds to be,

$$T_{\rm sky} \approx 180 \left(\frac{\nu}{180{\rm MHz}}\right)^{-2.6} {\rm K}, \tag{B.52}$$

where this observation would be aimed at a 'quiet' patch of high latitude sky[1].

Successful detection requires the system (whose temperature is obtained through a resistor), and the signal power to be matched. Respectively, these are defined as $k_{\rm b} T_a B$ and $B_\nu A_{\rm eff} B/2$. The sensitivity of our detector can therefore be defined as K_a,

$$K_a \equiv \frac{T_a}{S_\nu} = \frac{A_{\rm eff}}{2k_B}, \tag{B.53}$$

and the signal-to-noise estimates are typically $T_a/T_{\rm sys}$. Using Equation B.23 we are able to

[1]The intensity of the cosmological 21cm signal, compared to the radio emission of HI in the milky way has been given optical analogies to help intuitively understand the problem - my favourite being from one of N. Barry's talks, 'Its like detecting a light bulb in the centre of a sports stadium, with that stadium being fully lit, and on Pluto'. The talk also alluded to the latest MWA data measurement being sensitive enough to pick up TV signals reflected from the bottom of aeroplanes.

ensure the Gaussian noise is diminished by having the observation time $t_{\rm obs} \gg \tau$ so that the auto-correlation cycles are negligible. We are therefore left with the error on our source, σ_S, as,

$$\sigma_S = \frac{T_{\rm sys}}{K_a\sqrt{Bt_{\rm obs}}}. \tag{B.54}$$

From geometrical optics we know,

$$\theta_{\rm min} \approx \frac{\lambda}{b_{\rm max}}, \tag{B.55}$$

where the best angular resolution, $\theta_{\rm min}$, is dependent on the maximum baseline magnitude, $b_{\rm max}$. We can now write the noise contribution to our receiver temperature as,

$$\Delta T_n = \frac{b_{\rm max}^2 T_{\rm sys}}{A_{\rm eff,Tot}\sqrt{Bt_{\rm obs}}}, \tag{B.56}$$

where $A_{\rm eff,Tot}$ refers to the effective observing area of the total instrument.

Detecting only the 21cm signal ($\delta T_b \sim 10$mK) requires overcoming foregrounds that are $\sim 10^4$ times brighter than the desired signal, even in a 'quiet' patch of sky. Having $T_{\rm sys}$ matched with $T_{\rm sky}$, and following the above prescription, we can obtain,

$$\Delta T_{n,\rm Global} \approx 0.6\left(\frac{1+z}{10}\right)^{2.6}\left(\frac{Bt_{\rm obs}}{\rm Mhz\ hr}\right)^{-\frac{1}{2}}\ {\rm mK}, \tag{B.57}$$

where the redshift dependency is inputted via the observed wavelength $\lambda = 21{\rm cm}(1+z)$. To generalise to an interferometer we write,

$$\Delta T_{n,\rm Spatial} \approx 2\left(\frac{A_{\rm tot}}{10^5{\rm m}^2}\right)\left(\frac{b_{\rm max}}{0.21{\rm m}}\right)^2\left(\frac{1+z}{10}\right)^{2.6}\left(\frac{Bt_{\rm obs}}{\rm Mhz\ hr}\right)^{-\frac{1}{2}}\ {\rm mK}. \tag{B.58}$$

In a perfect instrument we could calculate the visibilities from this and, via Equation B.51, we could obtain a tomographic image of the EoR and cosmic dawn simply with a Fourier transform. Unfortunately instruments can never be perfect. 'Gain matching' is the real puzzle that needs solving for accurate detection of the cosmological 21cm signal, and it is particularly tricky to ensure the instrument gain is tuned only to detect contributions from the 21cm signal. It is comparatively easy to match these temperature equations theoretically, but the intricate thermodynamics of these detectors is beyond the scope of this work. Not to mention that all of the temperature contributions are oscillatory in nature, however the cosmological 21cm signal's

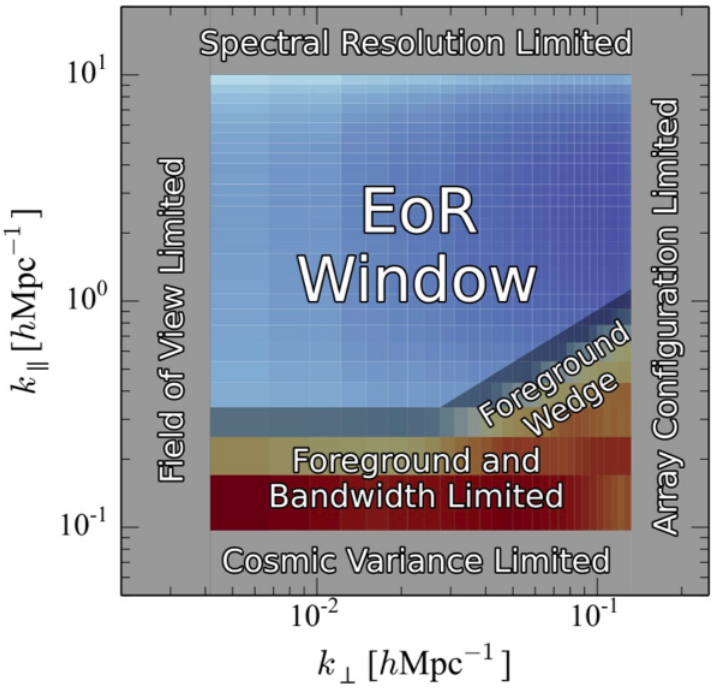


Figure B.2: The typical limitations of instruments like PAPER and MWA. The configuration of the array must be thought through carefully to balance U-V coverage and the signal-to-noise on any pair of voltage measurements. This figure is taken from Liu et al. (2014a) with permission.

slowly varying property may help here. This is very much an active area of research.

B.2.1 Modelling a real interferometer: The foreground wedge

The chromatic effects from the nature of the instrument (Section B.1.1) and the large foregrounds that must be overcome (Section B.2) combine to form a 'wedge' shape in the conventionally used cylindrical Fourier space. Figure B.2 shows the 'window' within which interferometers try to observe the EoR 21cm signal. In this section we guide its derivation as it is essential in understanding how foreground noise is treated in Section 1.3.7. This is where the two dominant techniques of foreground mitigation become involved:

- foreground suppression - aimed at observing within the wedge;
- foreground avoidance - the wedge is ignored and only k-modes outside of it are used.

These are discussed in more detail in Section B.2.3.

To derive the wedge we must mention quadratic estimator formalism, which in essence is a least squares fit of the state of our instrument via the continuity of noise in the multiple processes. This technique is utilised in CMB and Galaxy survey literature (Tegmark, 1997) and has been brought accross to the context of the 21cm power spectrum (Bond et al., 1998; Dillon et al., 2013). The full detailed derivation goes beyond the scope of this text, please see (Liu et al., 2014a) for more detail.

We can choose to analyse the visibilities with conventional trick known as the delay transform. Let us rewrite Equation B.48 emphasising the angular dependence of the observation as,

$$V(\mathbf{b},\nu) = \int I(\boldsymbol{\theta},\nu) A_{\text{eff}}\left(\frac{\boldsymbol{\theta}}{\theta_0}\right) e^{-i2\pi\frac{\nu}{c}\mathbf{b}\cdot\boldsymbol{\theta}} d\Omega, \tag{B.59}$$

where the vector $\mathbf{a}$ has been substituted for its angular counterpart $\boldsymbol{\theta}$, and the flat-sky approximation has not yet been used. We will see that the foreground wedge will emerge from the $\nu\theta$ product in the exponent of Equation B.59. We also must introduce the limitations of the banded nature of the measurement which we can do with a function, f, that depends on relative frequencies and the bandwidth, B. The response for each discrete frequency channel is contained within f, allowing Equation B.59 to be written as,

$$V(\mathbf{b},\nu) = \int\int I(\boldsymbol{\theta},\nu') A_{\text{eff}}\left(\frac{\boldsymbol{\theta}}{\theta_0}\right) f\left(\frac{\nu-\nu'}{B_\nu}\right) e^{-i2\pi\frac{\nu'}{c}\mathbf{b}\cdot\boldsymbol{\theta}} d\Omega \frac{d\nu'}{B_\nu}. \tag{B.60}$$

The previous two equations are a direct result of the experimental machinery. From here we choose to move into delay space to ease the analysis. This is defined as the Fourier transform of the frequency measured by each baseline. Continuing from Equation B.60 the delay-transform is written as,

$$\tilde{V}_{ij}(\mathbf{b},\tau) = \int V(\mathbf{b},\nu)\phi\left(\frac{\nu-\nu_0}{B}\right) e^{2\pi i\nu\tau} d\nu, \tag{B.61}$$

where ϕ is known as the tapering function or band-pass which we assume to be a Gaussian normalised to $\phi(0) = 1$ for simplicity[2]. B is the bandwidth within which we would like to know the power spectrum and it has a central frequency ν_0.

Now we apply whats known as the 'delay-space approximation', i.e. where $w \approx \tau$. Also at

[2]For the interested reader some examples of real choices for ϕ are the Blackman-Harris function or the Blackman-Nuttal function (both are weighted sums of cosines).

$u \approx \nu_0|\mathbf{b}|/c$ we can use $|\tilde{V}(\mathbf{b},\tau)|^2 \sim P$, enabling us to write,

$$P(u,w) \propto A_{b\parallel}\left(\frac{\nu_0 w}{u\theta_0},\right), \tag{B.62}$$

where the power spectrum is proportional to the primary beam (in the delay space approximation). Since the beam has a finite width (for sake of argument this can be the horizon at best) there will always be value of w that allows the power spectrum measurement to rise above the instrumental noise. We therefore have the line,

$$w = \frac{\theta_0}{\nu_0}u, \tag{B.63}$$

that dictates a sharp boundary between the measured power spectrum and the limitations of not only the instrument, but the directional nature of the observing strategy itself. Rewriting Equation B.63 from angular and spectral coordinates into Fourier spatial wavenumbers gives,

$$k_\parallel = \frac{H_0 E(z) D_c \mathbf{a}_0}{c(1+z)}k_\perp, \tag{B.64}$$

where $\mathbf{a}_0$ represents our previous conventions for the characteristic beam width (Equation B.44). Equation B.64 shows that below a threshold defined by the underlying cosmology, a certain portion of the observation will always be dominated by the foregrounds because the frequency and angular nature of the observation is not orthogonal. In other words the the mode mixing from the exponent in Equation B.48 can be interpreted as a linear relationship between the line-of-sight and antenna baseline Fourier modes. Below this line we expect observational trouble due to the limitations of our power spectrum estimator, the design of the instrument, and limitations from the bandwidth.

To show that this is valid is a lengthy process. Briefly, the information obtained per delay mode is able to span accross multiple frequency modes as a function of increasing baseline. We refer the interested reader to the full derivation in Liu et al. (2014a) for the intricate details. Please note that the short derivation we have provided is not strictly true however it provides significantly more intuition than the full derivation whilst obtaining the same analytic result.

By assuming that the 21cm power spectrum varies slowly accross small scales, various PS estimator approximations can increase the size of the EoR window increasing the quality of

measurement (Liu et al., 2014b). Constructing instruments that require narrow window functions can reduce the leakage of foregrounds across the Fourier plane. Since this is was what caused the foreground wedge in the first place, there is hope for measurement inside the foreground wedge with well thought through instrumental design. Both foreground signal and thermal noise must be accounted for if successful detection of the 21cm signal is to be made.

B.2.2 The current state of 21cm radio observations

A number of experiments are trying to detect either the globally averaged 21cm signal, with a radiometer; or the 21cm powers spectrum with interferometers. Both measuring the 21cm radiation as a spectral distortion against the CMB back-light. A first measurement of an absorption signal in the CMB radiation at 78 MHz has been claimed by the EDGES (Experiment to Detect the Global EoR Signature[3]) team (Bowman et al., 2018) in February 2018. This measurement from the sky-averaged radio spectrum could correspond to a 21cm absorption signal at $z \approx 17$ (Bowman et al., 2018), however it has fallen under particular scrutiny. The absorption trough is significantly deeper than previously theorised (Fialkov et al., 2014a). Either the radio background is brighter (Fialkov & Barkana, 2019); the hydrogen is colder e.g. Barkana (2018); or the measurement has incorrectly dealt with the observed foregrounds (Hills et al., 2018). Each argument has a promising case but only in cross-checking the measurement will we know for sure. REACH (The Radio Experiment for the Analysis of Cosmic Hydrogen[4]), SARAS (Singh et al. (2018) Shaped Antenna measurement of the background RAdio Spectrum[5]), and LEDA (Large-aperture Experiment to detect the Dark Ages[6]) all hope to do this in the near future. Interferometer measurements in the 75-100 MHz band are underway, however poor knowledge of extended radio sources and the large fields of view at this redshift do not lean in favour of providing stand-alone measurement with current instruments.

For the 21cm power spectrum at higher frequencies, LOFAR is providing upper limit measurements: at $z \approx 9.1$ there is a $(2\ \sigma)$ upper limits of 5329 mk^2 for $k = 0.075\ h\mathrm{Mpc}^{-1}$ (Mertens et al., 2020); and within the redshift range $z = 9.6 - 10.6$ the upper bound of $(79.6\ \mathrm{mk})^2$ at $k = 0.053\ h\ \mathrm{cMpc}^{-1}$ with a 13 hour observation of the NCP field (Patil et al.,

[3] https://www.haystack.mit.edu/ast/arrays/Edges/
[4] https://www.kicc.cam.ac.uk/projects/reach
[5] http://www.rri.res.in/DISTORTION/saras.html
[6] http://www.tauceti.caltech.edu/leda/

2017). MWA (Murchison Widefield Array[7]) currently support upper limits of 1.8×10^3 mK^2 for $k = 0.14$ hMpc^{-1} at $z = 6.5$ (Trott et al., 2020); and 3.9×10^3 mK^2 for $k = 0.20h$Mpc^{-1} at $z = 7$ using only 21 hrs (Barry et al., 2019). Both show promise in beating down sources of noise, approaching the precision necessary for the data. Although no detection has been made, these can be used to put limits on the fiducial parameters of favoured theoretical models in a parameter estimation framework (Greig et al., 2020a,b). Other interferometers such as GMRT (Giant Metre-wave Radio Telescope [8]) and PAPER (Precicion Array for Probing the Epoch of Reionization[9]) are also working towards this endeavour showing marginally worse precision to those above (Paciga et al., 2013; Ali et al., 2015). The future of radio telescope design is led with HERA and the SKA. In particular the Square Kilometre Array will be capable of tomographically mapping primordial hydrogen from $z = 6$ to 27 (Mellema et al., 2013; Koopmans et al., 2015). Its sensitivity is so impressive that it would be able to detect a mobile phone within 10 light years. Already the prototype mid frequency dishes in ASKAP (Australian SKA Pathfinder[10]), are building the sky-map for SKA with unprecedented precision as well as testing the technology. It has already discovered previously unknown radio sources which need adding to the skymaps (Norris et al., 2020) as well as finding that the polarisation structure of galactic radio lobes of stretch out ($\sim$ kpc) further than previously thought (Anderson et al., 2018). The main challenge with detecting the cosmological 21cm for the EoR is due to foreground sources. As discussed in Section B.2 foreground-to-signal ratios are of at least ~ 1000 (Barry et al., 2019). Although the current limits are still several orders of magnitude larger than the theorised power spectra, they pave the way for the first observations of the EoR. Over time the foregrounds will be better observed, the noise on the observation will reduce and the technology in the new generation of radio telescope will improve. Plans for a space bound global signal detection experiment already exist in the form of The Dark Ages Radio Explorer (DARE[11]). This would remove ionospheric concerns and reduce the human sourced radio interference dramatically. If large scale projects such as the SKA are a success, building a similar radio interferometer on the moon would allow the probing of $z \sim [6, 100]$ and could lead to some major insights in the evolution of cosmological density fields (Burke, 1985; Lazio et al., 2009). The prospect of a first confirmed 21cm detection is on the horizon. At this point

[7] http://www.mwatelescope.org
[8] http://gmrt.ncra.tifr.res.in
[9] https://public.nrao.edu/gallery/paper/
[10] https://www.csiro.au/en/Research/Facilities/ATNF/ASKAP
[11] https://www.colorado.edu/dark-ages-radio-explorer/

in time it is difficult to predict whether the fundamental limitations of the instruments will prevent a detection (Trott & Pober, 2019). More data will reduce systematic error up to a point, and depending on the redshift chosen, lack of foreground sources, and/or the strength of signal may act in our advantage. Nevertheless in order to prepare successful data analysis techniques in time for the completion of the SKA and HERA, simulated data must be used.

B.2.3 Alternate telescope simulations

Throughout the main text we implement 21cmSense when simulating telescope errors on mock observations. However, the methodology behind 21cmSense (discussed in Section 1.3.7) has a few pitfalls. Firstly it only provides the power spectrum while a real radio telescope, which has access to the raw visibility data, can calculate any desired statistic[12]. Secondly, the 21cmSense code iterates per coeval cube and therefore suffers from the light-cone effect. This is discussed further in Section 1.3.4, when we move from observing the 21cm power spectrum with coeval cubes to the light-cone; and in Chapter 5, when we attempt to remedy detail lost through box-car sampling with the Morlet power spectrum.

Foreground avoidance methods (like 21cmSense) are obviously not ideal since the modes in the foreground wedge are lost. Foreground suppression aims to observe within the wedge via a number of methods and is an active area of research. Promise has been shown in developing a sky-map, which is then subtracted from the visibility measurements along with telescope noise. Application of 21cmMC-fg (Nasirudin et al., 2020) within the framework of 21cmMC (discussed in Section 1.3) shows that parameter estimation results can be skewed by up to 5σ if foregrounds are not accurately predicted. An interesting practical example of this foreground removal technique is used by LOFAR. The NCP Sky Model Patil et al. (2017) is a map of the known astrophysical sources constructed around the North Celestial Pole. With every observation the observed sky is better understood and the observational capabilities of LOFAR are iteratively increased. Eventually, LOFAR aims to observe the 21cm power spectrum directly within the foreground wedge.

Most foreground avoidance relies heavily on the density of U-V space measurements which are

[12]Using visibility measurements directly can also be used in astro-particle physics. Radio telescopes can directly measure Čerenkov radiation caused by cosmic rays entering the Earth's atmosphere (Schellart et al., 2013).

to be interpolated. The modes at extreme ends of the measurement tend to be the ones most contaminated, and since they are the anchor points for the interpolation they can have the most effect in skewing the signal reconstruction. Including these modes incorrectly can lead to changes in the smoothness of the interpolated Fourier space leading to precisely inaccurate results. A full analysis of these techniques is beyond the scope of this work. To briefly summarise, avoidance seems to favour small spatial scales ($k_\perp > 0.6$ Mpc^{-1}) while suppression (when using simple frequency independent models) favours observing large line-of-sight scales (small $k_\parallel$) (Chapman et al., 2014). The instrumental effects are often largest at large $k_\perp$ modes. For a full discussion of foreground removal techniques please see e.g. Chapman et al. (2015) and Chapman et al. (2016).

Foreground cleaning is required to separate out the 21cm signal and is also an active area of research. This is particularly tricky because of the mixing proportions of the signal, made up from the noise, foreground sources, and the cosmological signal. The latter of these is unknown and adds variety in the underlying signal structure, spatial distribution and amplitude. A promising technique called global morphological component analysis (GMCA) is currently able to recover the 21cm signal from telescope noise with enough precision to create tomographic imaging (Chapman et al., 2013). GMCA is a Bayesian method that takes advantage of the non-Gaussian nature of the 21cm signal and the inherently Gaussian nature of the telescope noise. Noise structure in the signal is identified by its sparse contributions to the specific Fourier bases. This method is applicable to all but the smallest spacial scales ($<$ 1Mpc) and doesn't require parameterisation, however there is the potential for signal loss if cosmological signal is mistaken to be noise. For a fully specified simulation of SKA interferometry as well as a sky model, the interested reader should turn to OSKAR, The Oxford SKA Radio telescope simulator (Dulwich et al., 2009), however it has a dramatically large simulation time. Deeper analysis of this is beyond the scope of this text.

Appendix C

Statistical Analyses

This appendix contains the excess statistical material that introduces and supplements the Bayesian analyses presented in Chapter 1. Bayesian parameter and model selection are both used heavily in this work. Here we describe the motivation for Bayesian analyses and how to implement them via MCMC algorithms.

We begin discussing Bayesian and frequentist statistical analyses, including the derivation of Bayes theorem with qualitative and quantitative examples in Section C.1. We then look at MCMC methods and how they are implemented via two introductory algorithms - nested sampling and Metropolis-Hastings in Section C.2. These are relevant respectively to the workings of MultiNest and Emcee which are introduced in Chapter 1. One or both of these are used in every subsequent chapter. Section C.3.1 contains some preliminary tests for MultiNest; for Emcee, these were note required as the framework is already established in the form of 21cmMC. We also present a convergence check for each of these algorithms in Section C.3.2. Finally we look at how statistics are tied in with astrophysical measurements in Section C.4, i.e. to make sense of the segue to and from this appendix. Particular attention is applied to the power spectrum as the Fourier transform of a two point correlation function.

C.1 Introduction

Large amounts of this opening section have been based on the Imperial Centre for Inference & Cosmology summer school (ICIC) for which I have both been a student and demonstrator.

More detail can be found online at
https://www.imperial.ac.uk/astrophysics/centre-for-inference-and-cosmology/seminars-and-events/data-analysis-workshops/, or in the texts Jaynes (2003); Hobson et al. (2009); Sivia & Skilling (2006).

C.1.1 Frequentist or Bayesian?

Typically problems in physics are approached with forward modelling (also known as using a generative model). Outcomes are predicted, assuming a combination of physical processes contribute. Applying Bayes theorem can be thought of as the backwards version of this. Given data, we try to infer things about how the data was made. This can be performed in terms of parameter inference, or model selection. If an arbitrary data set is being fit, for example, $y = mx + c$, may be an adequate model but the values of m and c need to be pinned down with parameter estimation. Also, perhaps $y = qx^2 + mx + c$ fits better, or the added parameter, q is unnecessary - either way we can quantify their performance with model selection.

The key difference between frequentist and Bayesian statistics is the interpretation of probability. A frequentist requires infinitely long trials to describe the *relative frequency* of outcomes. Given prior knowledge, a Bayesian expresses probabilities as a *degree of belief*. In both cases detailed information is gained about a desired statistic and its spread (e.g. a mean, μ and variance, σ^2). Our aim in either case is to minimise the distance of our theoretical model to the data.

C.1.2 Bayes theorem

In order to use Bayes' theorem we must derive it, requiring a brief summary of some probability rules. Firstly, $p(x)$ is defined as the probability of an event, x, occurring, and $p(\bar{x})$ represents the same event x *not* occurring. The sum rule is then defined as, $p(x) + p(\bar{x}) = 1$. Taking the sum rule further a probability $p(x)$ must be normalised: $\int p(x)dx = 1$ (also true for probability density functions). For combining probabilistic events we have the product rule: $p(x, y) = p(x|y)p(y)$, where $p(x|y)$ represents the probability that event x has happened *given* event y has occurred in the past. This encodes the possible correlation between the two events,

if x & y are independent for example, $p(x|y) = p(x)$ and $p(x,y) = p(x)p(y)$. Each parameter in a joint probability can be *marginalised* as: $p(x) = \int p(x,y)dy$. By symmetry, we must have $p(x,y) = p(y,x)$ and by rearranging both sides via the product rule we come to Bayes theorem (Bayes & Price, 1763),

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \tag{C.1}$$

C.1.3 The normal distribution and the central limit theorem

In general most (but not all) applications of probability will follow a Gaussian, or normal distribution. To show this, consider a one dimensional random variable. We can estimate a desired location l from a set of $(n+1)$ observations, $[x_0, ..., x_i, ..., x_n]$. Assuming the sampling process is independent between samples it can be factorised: $p(x_i|l) = p(x_0|l)...p(x_i|l)...p(x_n|l)$, is the likelihood distribution. Since we desire the maxima, we can take a logarithm without losing any information,

$$\sum_{i=0}^{n} \frac{\partial}{\partial l} \log p(x_i|l) = 0. \tag{C.2}$$

The most likely (maximum likelihood) estimate for l, labelled $\hat{l}$, will satisfy Equation C.2. It may seem obvious to take the arithmetic mean, $\hat{x} = \sum_{i=0}^{n} x_i/(n+1)$, for the estimate $\hat{l}$ but this does not solve Equation C.2 by substitution. To remedy this we can consider a set in which only x_0 is nonzero. Since these are distances, we can use the change of variables $\log p(x_i|l) = f(l - x_i)$, and also $d = l - x_i$ to give,

$$\sum_{i=0}^{n} \frac{\partial}{\partial l} f(\hat{l} - x_i) = \sum_{i=0}^{n} \frac{\partial}{\partial l} f(d) = 0. \tag{C.3}$$

The set under consideration is then expressed by, $x_0 = (n+1)\hat{x} = (n+1)\hat{d} = (n+1)l$, and $x_{i\neq 0} = 0$ (i.e. the whole length l is represented as the first sample). With the mean $\hat{l} = \hat{d} = l$, and $\hat{l} - x_0 = -n\hat{d}$, this gives,

$$\frac{\partial(-n\hat{d})}{\partial l} + n\frac{\partial(\hat{d})}{\partial l} = 0, \tag{C.4}$$

for $n > 0$. Using $n = 1$, f must be asymmetric, and the only possible solutions to these simultaneous equations is linear: $f(d) = \frac{1}{2}c_1(x-l)^2 + c_2$ with constants c_1 and c_2. Finally,

normalising and returning to our original coordinates gives the desired Gaussian,

$$f(x|l) = \frac{\alpha}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}(x-l)^2}. \tag{C.5}$$

This is the Gauss derivation of the normal distribution (Gauss, 1809). However, since we have used a special data set a more rigorous derivation is required to prove this distribution is true in general. Despite this the principles here remain true: if Equation C.5 satisfies the data, then the peak of the likelihood always has the unique solution of being equal of the sample mean. We have shown that for the maximum likelihood point to be the mean the sampling distribution must be Equation C.5, with freedom only in the scale, α. If this is not the case, it is likely that the samples are not independent (the first assumption) or there are simply not enough samples. We take advantage of this in the particular form of the likelihood statistic used later on (Section 1.4.2) when detailed application of Bayesian theory is looked at.

To explain the central limit theorem we must first define the characteristic function[1],

$$\phi_y(\alpha) = \int_{-\infty}^{\infty} f(y) e^{i\alpha y} dy, \tag{C.6}$$

where $f(y)$ is the pdf of data, y. If we now consider N independent and identically distributed samples drawn as $x_i \sim (\mu, \sigma^2)$. We can transform this set of data by, $y_i = (x_i - \mu)/\sigma$, so that the sample mean becomes standardised with $(\mu = 0, \sigma^2 = 1)$ as,

$$z_{\mathrm{N}} = \frac{N\hat{x} - N\mu}{\sigma\sqrt{N}} = \sum_{i=1}^{N} \frac{y_i}{\sqrt{N}}, \tag{C.7}$$

where z_{N} represents the standardised sample mean. By taking the characteristic functions for z_{N} and N samples of y, we can write,

$$\begin{aligned} \phi_{z_{\mathrm{N}}}(\alpha) &= \left[\phi_y\left(\frac{\alpha}{\sqrt{N}}\right)\right]^N \\ &= \left[1 + \frac{i\alpha\hat{y}}{\sqrt{N}} - \frac{\alpha^2\hat{y^2}}{2N} + ...\right]^N, \end{aligned} \tag{C.8}$$

where we have expanded the exponential in Equation C.6. We also know that $\hat{y} = 0$ and

[1] Characteristic functions are often used to easily generate the statistical moments.

$\hat{y^2} = \sigma^2 + \hat{y}^2 = 1$ due to our transformation from x_i. We therefore have,

$$\lim_{N\to\infty} [\phi_{z_N}(\alpha)] = e^{-\frac{\alpha^2}{2}}, \tag{C.9}$$

where the large N causes moments of order, m, to vanish above $m > 2$ (i.e. terms of order higher than, $\hat{y^2}$, vanish due to the, $N^{\frac{m}{2}}$, on the denominator). Finally we use Equations C.9 and C.6 to infer the form of the pdf as,

$$\lim_{N\to\infty} \left[z_N = \frac{\sqrt{N}(\hat{x} - \mu)}{\sigma} \right] = \mathcal{N}(\mu = 0, \sigma^2 = 1), \tag{C.10}$$

where $\mathcal{N}$ is the normal distribution (similar to the form derived in Equation C.5).

We have found that as $N \to \infty$ our distribution of independent random variables, x_i, has become the normal distribution with mean μ and standard deviation σ. Hence for large data sets, samples that are independent and from identical distributions can be entirely represented by a mean and variance through the normal distribution. This is known as the central limit theorem (CLT) and is regularly used in statistics (Jaynes, 2003).

C.1.4 Qualitative examples

Both statistical methods require a fundamental assumption: in frequentist scenarios this is the hypothesis which should be tested; for the Bayesian it is prior knowledge of the scenario. To approach any problem, one must be equipped with the correct toolbox. Using Bayesian and frequentist methods outside of their respective comfort zones can lead to drastic misinterpretation.

Cosmology is an obvious place to apply a Bayesian approach. Examples include, Heavens et al. (2017b) using Bayesian model selection to show that the $\Lambda - \mathrm{CDM}$ cosmological model is the most likely fit to the Planck CMB data. We look further into this methodology in Chapter 2.

Famously within physics, the BICEP and BICEP-2 experiments[2] lost the Nobel prize by misapplying Bayes theorem (BICEP2 Collaboration et al., 2014). They claimed first evidence of cosmic inflation with the effect of primordial gravitational waves on the CMB, in the form

[2] http://bicepkeck.org

of a large cold patch. Unfortunately this had to be retracted, the dust contribution in the foreground of the observation had not been fully marginalised over (Gott & Colley, 2017).

With only one universe, frequentist methods are not straightforward to apply within cosmology. Particle physics, on the other hand, is naturally frequentist. In order to obtain adequate certainty in a frequentist scenario, a large numbers of data samples must be collected. The idea behind this being from the CLT, where error will reduce as $\frac{1}{\sqrt{N}}$. Frequentist approaches are particularly suited for collider experiments since events are repeated frequently. Experiments are easily repeatable and hypothesis testing is done to check signals are not noise. At CERN, the ATLAS and CMS experiments discovered the Higgs boson to 5 standard deviations (σ) above the noise background of other particles. This means that *if* the specific signature of the Higgs was observed by chance, then it would occur once every 350 million collisions - similar odds to winning the national lottery jackpot for all 6 numbers, on every successful observation (Flechl, 2015).

Application of statistics within medicine is a situation where noise is not easily measurable. Here the conventionally known p-number test is used. This asks whether the sample under question happens to be in the tail of the distributed census. In this case that would be a small sample of clinical trials for a new drug (e.g. a vaccine), compared to the untreated census population known as the control group. Assuming the sampling is plentiful and unbiased (tricky in qualitative science), the p-number is related to a $\chi^2 = \sum_i (x_i - \mu)^2/\sigma^2$ for the trial and control groups. The p-number is defined as $p(\chi^2_{\text{Sample}} \geq \chi^2_{\text{Control}})$ and gives a measure of how effective the sample is reletive to the control. The so called T-test is a simplification of this where the variances of the test distribution and census data are matched. T-tests orient around matching the means of the distributions. When using a T-test, $p < 0.05$ is needed for medicine to be approved for use within the health system, and given our assumptions are met, this will be an adequate test. If enough patients are selected randomly from the population, the sample is an accurate distribution of the population, matching the Bayesian posterior. But sets of samples that are small and only in the tail of the true distribution could have catastrophic reaction when applied to the census population. Another caveat of the p-number test is that the hypothesis is assumed true by default. The doctor knows little about how an exact trial treatment will initially effect a patient (although trial medicine is designed to have similar bio-chemical properties to successful medication). If the hypothesis is objectively false, the method

will always by swayed by the suggestion that more data is necessary to prove this. In the case of trial medicine, the aspect of patient safety outweighs the desire for a rigorous result - trials with non-negative results are therefore abandoned quickly.

Both Bayesian and frequentist applications require responsibility from the user. When there is knowledge of the scenario, the influence of the prior can make Bayesian inference a powerful tool. But both methods can be susceptible to skewed results if misused - the hypothesis must be legitimately plausible, and so must the prior knowledge. A bad example of this occurred in 1810, Samuel Hahnemann assumed arsenic is good for the human body. Today arsenic is known to be poisonous. As any statistician should with either approach, patients were tested in order to gain data. Quickly the conclusion that less arsenic is better for you is reached, and that consuming diluted poisons is therefore good for you[3] (Hahnemann S., 1810).

Particular attention must be given to the prior knowledge applied to any problem. Neither method is fool-proof. In either approach the amount of data **must** outweigh the hypothesis or priors. The aim should always be to minimise the distance of our theoretical model to the observational data.

C.1.5 Quantitative example

Consider, in Cartesian coordinates, a lighthouse on a straight coastline ($y = 0$) at an unknown position $x_{\rm true}, y_{\rm true}$. A series of N flashes are recorded (without direction or intensity) at $x_1, .., x_i, .., x_{\rm N}$. In a Bayesian approach we set up what we want to know, $p(x_{\rm true}, y_{\rm true}|\{x_i\})$, where curly brackets indicate the set of data. We can then apply Bayes theorem as,

$$p(x_{\rm true}, y_{\rm true}|\{x_i\}) \propto p(\{x_i\}|x_{\rm true}, y_{\rm true})p(x_{\rm true}, y_{\rm true}) \tag{C.11}$$

$$\propto \prod_i p(x_i|x_{\rm true}, y_{\rm true}),$$

[3]Samuel Hahnemann is actually the founder of homeopathy. This is its true grounding although poisonous quantities diluted today are statistically negligible (and the book contains the use of many other remedies too). The related hydration, nutrition, and placebo convolute this problem away from decisive conclusions against the methodology.

assuming no prior knowledge (an uninformative prior distribution for x_{true}, and y_{true} see Section 1.4.4). If we now perform a coordinate transform to polar co-ordinates,

$$\frac{x_i - x_{\text{true}}}{y_{\text{true}}} = \tan\theta_{\text{true}}, \tag{C.12}$$

allowing a change of variables as,

$$p(x_i|x_{\text{true}}, y_{\text{true}}) = p(\theta_i|x_0, y_0) \left| \frac{d\theta_i}{dx_i} \right|. \tag{C.13}$$

The signals are therefore uniformly distributed, $\theta_i \in [-\pi/2, +\pi/2]$, independent of the true lighthouse position. The chain rule gives,

$$\sec^2\theta_i \frac{d\theta_i}{dx_i} = \frac{1}{y_0}, \tag{C.14}$$

where from trigonometry we get made use of $\sec^2\theta_i = \left[1 + \frac{(x_i - x_{\text{true}})^2}{y_{\text{true}}^2}\right]$, to produce the likelihood per sample,

$$p(x_i|x_{\text{true}}, y_{\text{true}}) = \frac{1}{\pi y_{\text{true}} \left[1 + \frac{(x_i - x_{\text{true}})^2}{y_{\text{true}}^2}\right]}, \tag{C.15}$$

which are inputted to Equation C.11 to produce,

$$p(x_i|x_{\text{true}}, y_{\text{true}}) = \prod_i \frac{1}{\pi y_{\text{true}} \left[1 + \frac{(x_i - x_{\text{true}})^2}{y_{\text{true}}^2}\right]}, \tag{C.16}$$

the unnormalised posterior. Now that we have this, answering the question becomes a peak finding problem as we simply maximise the likelihood for the true position.

In contrast the frequentist approach requires the definition of an estimator. A sensible choice would be the mean,

$$\mu = \frac{1}{N} \sum_i^N x_i, \tag{C.17}$$

with the hope that as N increases $\mu \to x_{\text{true}}$. However the form of Equation C.15 is known as the Cauchy distribution. This is particularly troublesome as it has no mean and infinite variance. To give the frequentist approach a fair chance, we can approach this with a characteristic

function (Fourier transform of the pdf),

$$\psi = \int_{\infty}^{\infty} \frac{e^{ikx}}{\left[1 + \frac{(x-x_{\rm true})^2}{y_{\rm true}^2}\right]} dx = e^{ikx_{\rm true} - ky_{\rm true}}. \tag{C.18}$$

With convolution theory, the summing the means of the set, $\{x_i\}$, are combined to give the characteristic function ψ^N. We can return to real space to obtain the pdf of the estimators collectively as,

$$p(N\mu) = \frac{1}{\pi N y_{\rm true}\left[1 + \frac{(Nx_i - Nx_{\rm true})^2}{(Ny_{\rm true})^2}\right]}. \tag{C.19}$$

Finally a change of variables $Nx \to x$ and $Ny \to y$ reveals,

$$p(\mu) = \frac{1}{\pi y_{\rm true}\left[1 + \frac{(x_i - x_{\rm true})^2}{y_{\rm true}^2}\right]}. \tag{C.20}$$

In this scenario, we have found that a single data point is as good as any number. Nothing is improved *in this scenario* by using an average due to the infinite variance of the Cauchy distribution. This does not mean the frequentist approach is wrong, we have simply shown a situation where the Bayesian tool box is the right thing to do (Gull, 1989).

C.2 Supplementary detail for Bayesian analysis

C.2.1 Markov Chain Monté Carlo (MCMC) methods

In astrophysics, posterior distributions are rarely analytic functions and evaluating grid searches of the parameter space is too computationally expensive to be a viable option. The solution is to implement a group of algorithms known as MCMC. In essence we are looking for a 'chain' of points that, although they are correlated locally, describe ergodically the distribution of the likelihood as the model at hand fits for the observational data. This is satisfied if consecutive points in the chain satisfy detail balance,

$$\mathcal{L}(s_i)\Pi(s_i)p(s_{i+1}|s_i) = \mathcal{L}(s_{i+1})\Pi(s_{i+1})p(s_i|s_{i+1}), \tag{C.21}$$

where $p(s)$ is the probability of selecting sample s. If this is done successfully it is the idea of physical equilibrium that drives the algorithm to provide an ergodic representation of the parameter space.

The MCMC family consists of a large range of algorithms that can be used for a variety of different purposes. We saw in Sections 1.4.1 and 1.4.3 that parameter estimation is essentially peak finding, and model selection is essentially integrating. Each method implements one of these tasks in a different way, suitable for different intricacies. The following subsections introduce the Metropolis-Hastings algorithm, a simple parameter estimation algorithm designed to find θ_{MAP}; before introducing nested sampling, a model selection algorithm which calculates $\mathcal{Z}$.

The Metropolis-Hastings algorithm

A simple example of an MCMC algorithm is known as the Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970). It is summarised in Algorithm 6.

Algorithm 6: Metropolis-Hastings (M-H)

Initially draw a sample point s_0 from the prior;

$s_i = s_0$;

while *Convergence unsatisfied* **do**

  Sample a 'trial' point s_{trial}, based on a Gaussian pdf

  with a mean of the current point s_i and a user determined covariance;

  Calculate the ratio $r = \Pi(s_{\mathrm{trial}})\mathcal{L}(s_{\mathrm{trial}})/\Pi(s_i)\mathcal{L}(s_i)$;

  Draw a random number $n \in [0, 1]$; **if** $r > n$ **then**

    $s_{\mathrm{trial}} = s_{i+1}$;

  else

    $s_i = s_{i+1}$;

  end

end

Here, p in Equation C.21 would be the Gaussian pdf. The convergence criteria in this case could be n number of consecutive rejections. Typically these algorithms are run for a number of iterations ($\sim 10^4$) and a convergence test (discussed in Section C.3.2) is then implemented afterwards on the resulting chains.

At the start of the algorithm there is a congregation of points which are more sensitive to the initial position rather than the shape of the posterior. Since it is equilibrium that drives the chain, this starting point of the chain might not initially satisfy Equation C.21. This phase is referred to as the 'burn-in' and needs to be removed if the sample is to be an ergodic representative of the posterior. Removing it prevents exaggerating the information content within low likelihood regions of the parameter space.

Nested sampling

Nested sampling (Skilling, 2004) is a Monté Carlo algorithm, aimed primarily at easily producing the Bayesian Evidence. This is done by reducing the dimensionality of the integral (Equation 1.37) to one - that one being a fraction of prior volume, $\mathcal{X}$. The resulting transformation gives,

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX. \tag{C.22}$$

which can be easily updated via the trapezium rule as,

$$\mathcal{Z} \approx \frac{1}{2} \sum_{i=1}^{N_s - 1} (X_{i-1} - X_{i+1}) \mathcal{L}(X_i). \tag{C.23}$$

Its implementation is shown in algorithm 7. Within each iteration, the updated prior volume is the new region defined by the isolikelihood contour.

Algorithm 7: Nested Sampling

Initially scatter N 'live points' throughout the prior distribution;

The first ($i = 0$) prior volume and Evidence values are $X_0 = 1$ and $\mathcal{Z} = 0$ **while**

Convergence unsatisfied **do**

$i \to i + 1$ Select the live point with lowest likelihood s_i;

Draw an isolikelihood contour intersecting s_i;

Remove s_i from the live points and save it;

Update the prior volume,

$X_i = X_{i-1} \frac{N}{N+1}$;

Update the Evidence calculation with $\mathcal{L}(s_i)$ via Equation C.23, $\mathcal{Z} \to \mathcal{Z} + \mathcal{L}(s_i)$;

Check convergence: satisfied if $\mathcal{L}_{\max} X_i = \Delta Z <$ Threshold;

Re-sample s_i from within the isolikelihood contour (which is now the updated prior volume, X_i);

end

Include the remaining live points in the calculation, $Z \to Z + X_i \sum_{j=1}^{N} \mathcal{L}(s_j)/N$

Figure C.1 illustrates this for the isolikelihood contours of a simple two dimensional example. It is worth noting that as the algorithm progresses the prior volume will have shrunk, $X \sim \exp(i/N)$ where, i is the number of iterations.

The simplest form of this is in essence a Metropolis-Hastings algorithm, but with a hard acceptance criteria rather than a probabilistic one. However this implementation becomes horrendously inefficient at high dimension, and will need impractically long run times to achieve ergodicity. Other algorithms must be added to find the iso-likelihood contour (discussed further in Sections 1.5.2 and C.3.3).

A successful application of nested sampling uses Galilean Monté Carlo (Skilling, 2012), where the system is evolved dynamically with Hamilton's Equations,

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = -\frac{\partial H}{\partial \boldsymbol{\theta}}, \quad \text{and} \quad \frac{\partial \boldsymbol{\theta}}{\partial \theta_i} = \frac{\partial H}{\partial \boldsymbol{\theta}}, \tag{C.24}$$

where the canonical momentum, $\boldsymbol{p}$ represents particles instead of walkers. These particles typically have normally distributed velocities, a user defined mass and travel along the parameter space to create the chains according to the Hamiltonian $H = \ln\mathcal{L}$. This differs from Hamiltonian Monté Carlo as when the chain trajectories heading outside of the iso-likelihood contour

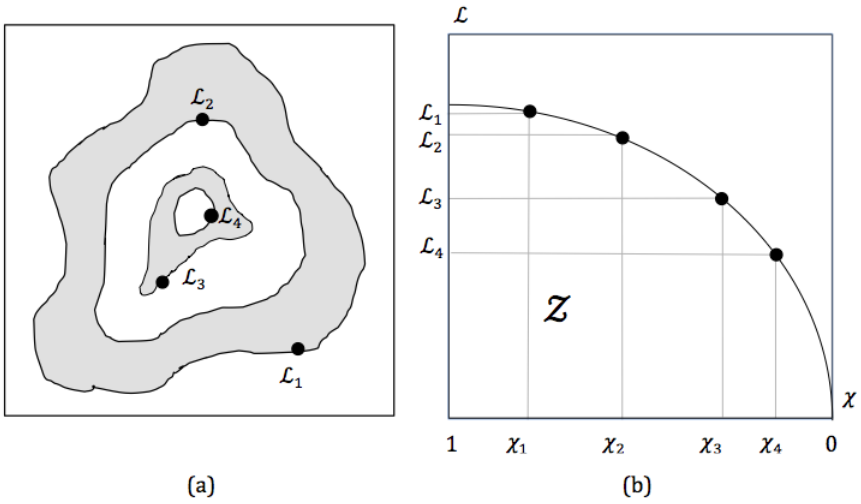


Figure C.1: An illustration of nested sampling (Skilling, 2004). (a) shows 4 iterations of the isolikelihood contours as the algorithm hones in on the posterior peak. In this example the volume of the (uniform) prior distribution is the square. (b) plots the same likelihood points against the parameter volume within the isolikelihood contour as a ratio with the prior volume, $\mathcal{X}$. The integral under the curve is the Bayesian Evidence. In MultiNest, Section 1.5.2) this is evaluated with the trapezium rule as in equation C.23 (this figure is recreated from Feroz et al. (2009) with permission)

have their momentum reflected. It is shown to work on a remarkably varied set of toy models (Feroz & Skilling, 2013), however it requires a differentiable likelihood space.

C.2.2 Alternatives

Before commencing with the implementation of full Bayesian methods, we explored using the maximum likelihood point ($\mathcal{L}_{\max}$) as a model selection test. Typically the maximum likelihood point, $\mathcal{L}_{\max}$, is combined with a penalty for the number of parameters, d, with higher values favouring certain models (in an attempt to implement Occam's razor).

The Bayesian information criterion (BIC) (Schwarz, 1978) is an attempt to approximate the Bayesian Evidence directly and is used later on within within MultiNest (Section 1.5.2). It is designed for use in data sets with orthogonal parameters (dimensions = parameters), and with the same number samples N,

$$\mathrm{BIC} = -2\log(\mathcal{L}_{\max}) + 2d\,\log N. \tag{C.25}$$

There is evidence that this works well on Gaussian likelihoods that are well constrained within

their prior distributions. However we immediately see that that with the same number of samples higher parameters will have their distributions less well sampled. There is also no facility for penalising redundant parameters.

The AIC (Akaike information criterion) (Akaike, 1974) is a similar idea defined as,

$$\mathrm{AIC} = -2\log(\mathcal{L}_{\max}) + 2d. \tag{C.26}$$

The AIC is an approximation to the Kullback-Leiber divergence (Kullback & Leibler, 1951) or relative entropy, defined as,

$$H_{\mathrm{KL}} = \int P(\theta) \left(\log \frac{P(\theta|D)}{P(\theta)} \right) d\theta, \tag{C.27}$$

which measures the difference between competing probability distributions. The AIC tries to capture the information lost from fitting the model to the data by estimating their entropy content.

The MML (minimum message length) (Wallace, 2005) is based on the Shannon information entropy, $H = -\int \mathrm{p}\ \mathrm{lnp}$. Applied here, this would be, $H = -\int \Pi\ \mathcal{L}\ (\log\Pi + \log\mathcal{L})$, however implementing it would be significantly more complicated than just using Equation 1.38, where the prior is easily disentangled. This is not to say it could not be done, but at significantly more effort than the Bayesian framework set out it Section 1.4.3.

Despite these options being simpler than a full calculation of $\mathcal{Z}$, they can be inconsistent due to their differing implementation. They perform badly when the peaks appear near the edge of the prior distributions, or if there are dependencies between the parameters in the model. On top of this, they have no preparation for multi-modal likelihood distributions as only the single maximum likelihood point is considered. Both the AIC and BIC work well in the testing of $\Lambda - \mathrm{CDM}$ (Szydłowski et al., 2015), however both have been shown to disagree on WMAP data (Magueijo & Sorkin, 2007; Liddle, 2007). We will not consider them further.

C.2.3 Future analyses

Some of the techniques that we consider for future work are discussed in this section. Namely the Bayesian complexity as well as the Bayesian model averaging we mentioned at the start of this Section.

Once model selection is used and a selection of models are well established, the most detailed inference that can be obtained in the Bayesian framework is model averaging. Here the posteriors from each model are averaged together. To do this a relative Evidence is constructed from multiple data sets,

$$P(D_i|D, M) = \frac{P(D_i|M)}{\sum_j P(Dj|M)}, \tag{C.28}$$

which is then used as a weight to provide a concordant posterior,

$$P(\boldsymbol{\theta}|D, M) = \sum_i P(\boldsymbol{\theta}|D_i, M) \times P(D_i|D, M), \tag{C.29}$$

providing the most precise constraint on a parameter under question. Bayesian analysis averaging will be looked at in the context of 21cm physics once there is data to average over. In the context of the reionisation this has been done for the high redshift UV luminosity function (Gillet et al., 2020).

The Bayesian complexity, $\mathcal{C}$, is designed to measure how many parameters are necessary to fit the mock data produced by a model. Using the Kullbeck-Leiber divergence defined in Equation C.27 it is defined,

$$\mathcal{C} = -2\left[H_{\rm KL}(\mathcal{P}, \Pi) - \left(H_{\rm KL}|_{\theta_{\rm MAP}}\right)\right], \tag{C.30}$$

where, $H_{\rm KL}|_{\theta_{\rm MAP}}$, is a single point evaluation of $H_{\rm KL}$ at $\theta_{\rm MAP}$. The idea being that the more descriptive the data, the tighter the posterior and $H_{\rm KL}$ distribution becomes closer to its single point equivalent - minimising the complexity. Rearranging (with the use of Bayes theorem) we can write,

$$\mathcal{C} = -2\int p(\boldsymbol{\theta}|D, M)\ln\mathcal{L}(\boldsymbol{\theta})d\boldsymbol{\theta} + 2\ln\mathcal{L}(\theta_{\rm MAP}). \tag{C.31}$$

Since most cases the likelihood is a Gaussian χ^2 distribution we can then write,

$$\mathcal{C} = \hat{\chi}^2(\boldsymbol{\theta}) - \chi^2(\theta_{\rm MAP}), \tag{C.32}$$

providing a measure for the effective number of parameters in the data. This has been shown to work well on $\Lambda - \mathrm{CDM}$ (Kunz et al., 2006), but more work is necessary for its implementation on non-Gaussian data sets.

C.3 Supplementary detail for MCMC algorithms

C.3.1 MultiNest testing

MultiNest only really has 2 tuning parameters - n_live_points and a tolerance criteria on the Evidence used for ending the algorithm (for the full list of possibilities see Section C.3.1). The variation of these has a significant impact on the quality of posterior distributions produced and the calculation of the Evidence.

Should the reader be interested in implementing MultiNest, we recommend leaving all except these two options fixed. The literature recommends infinite number of sample iterations (ensuring the stopping criterion is used) and 2000 live points. Figures C.2 and C.3 show the variety in posterior distributions as these parameters vary respectively for a 3D-Gaussian likelihood ($\mu = 0,\ \sigma = 3$) and uniform priors ($[-10, 10]$). Clearly when the live points are too low, or too few sample iterations are performed the posteriors obtained are patchy and inaccurate.

The Evidence values obtained also have their error bars reduced as the number of live points increases, as can be seen in Figure C.4. In this Figure the results are calculated with respect to the FZH 3 parameter model from 21cmFast (discussed heavily in Chapter 3). The use of 2000 live points produces suitable accurate and precise results. Note that if too few live points are used MultiNest does not disagree with itself, but to ensure this is the case the error bars become uninformatively wide.

Now that we are happy with the MultiNest settings, we test MultiNest with two complicated toy functions. Firstly, the Eggbox function,

$$\ln\mathcal{L} = (2 + \cos(x/2) \times \cos(y/2))^5 \tag{C.33}$$

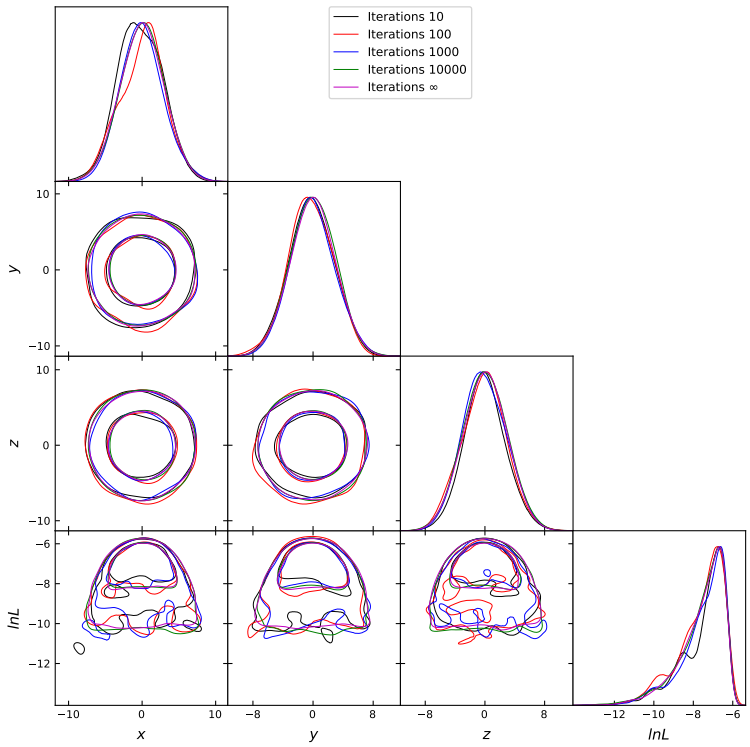


Figure C.2: The posteriors obtained by MULTINEST for a 3D Gaussian likelihood. We vary the number of sample iterations performed whilst using 2000 live points (and otherwise default settings). The colours indicate different numbers of sample iterations.

with uniform priors $(x, y) \in ([0, 30\pi], [0, 30\pi])$. Secondly, a pair of Gaussian shells,

$$\mathcal{L} = \sum_{i=1}^{N=2} \frac{1}{\sqrt{2\pi w^2}} e^{0.5w^{-2}(|\mathbf{r}-\mathbf{c_i}|-r)^2} \tag{C.34}$$

where $w = 0.1$ represents the width of the cylinders, $r = 2$ their radius, and each is centred at $\mathbf{c_1} = [3.5, 0]$ and $\mathbf{c_2} = [-3.5, 0]$. The uniform prior here is $\mathbf{r} = (x, y) \in ([-6, 6], [-6, 6])$ Both of these are included within MULTINEST as test functions.

We plot the posteriors obtained in C.5, and all thought they are not perfect (e.g. the heights of the eggbox peaks or the smoothness of the Gaussian shell centres), the Evidence values agree with the analytic results[4] within error bars (Table C.1).

With enough live points, we are confident the parameter posteriors are as accurate and precise as EMCEE - given our likelihood space does not succumb to jagged edges. The GAMBIT collaboration try a variety of samplers to obtain awkwardly shaped likelihood spaces that arise from particle physics (Martinez et al., 2017). They show MULTINEST can miss small islands in a

[4]The analytic integration here is with the QUADPACK algorithm from SCIPY QUAD.

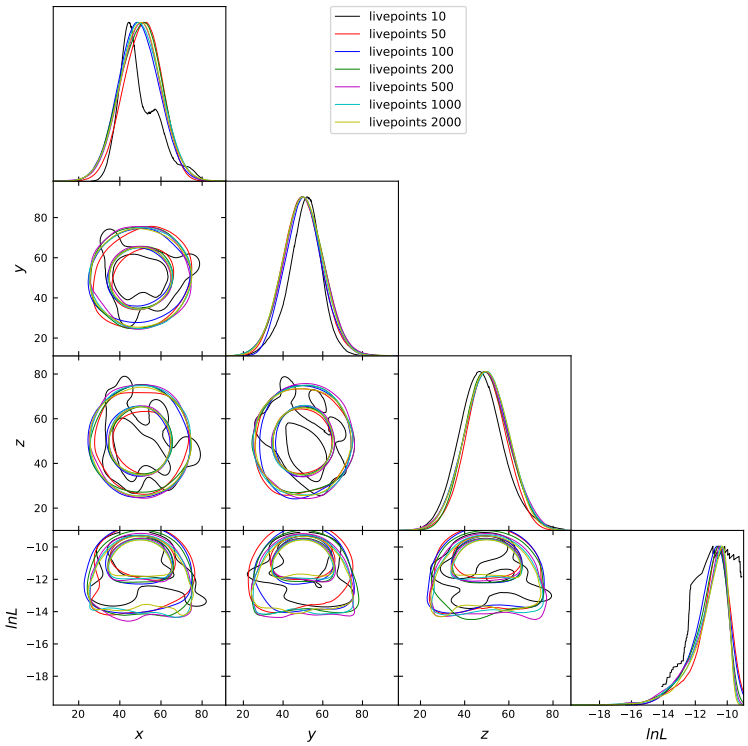


Figure C.3: For the same 3D Gaussian in Figure C.2, but here the colours indicate different numbers of livepoints. We set the sample iterations to infinity to ensure the algorithm reaches the Evidence tolerance stopping criterion. Above 500, the number of livepoints are in undeniable agreement in this toy scenario.

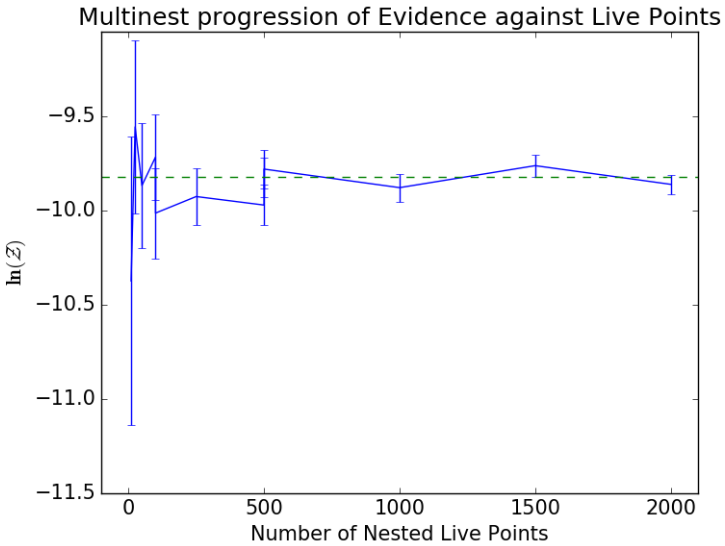
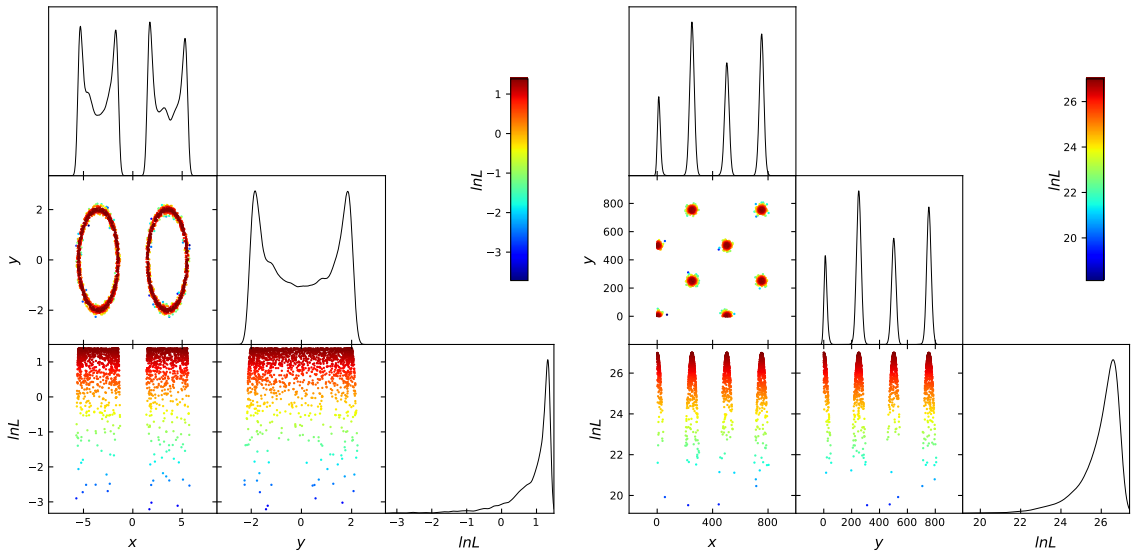


Figure C.4: An example of how the MultiNest Evidence calculation varies with live points. Notice that the statistical error bars are larger for lower number of live points. Although in Figure C.3, using 500 live points was deemed suitable, this is performed on the toy reionisation model (3pFZH Chapter 3), where the parameters are no longer orthogonal, here around 1000 live points would be suitable for reproducible results. For reference the dotted green line represents a value produced with 4000. The authors recommend the value of 2000 live points, above this there is little to be gained.



(a) (b)

Figure C.5: Posteriors of the Gaussian-Shell and Eggbox test likelihoods as found by MultiNest. In order for these histograms to resolve the eggbox, this posterior has been scaled in comparison to Equation C.33.

Function	MultiNest	Samples	Analytic (2 d.p.)
Egg-box	235.88 ± 0.06	51894	235.88
Gaussian Shells	-1.71 ± 0.04	30375	-1.75

Table C.1: We found results within error bars of those produced in Feroz et al. (2009) for the two test functions. See section C.3.1 for function details and Figure C.5 for the parameter posterior distributions.

jagged posterior space due to the ellipsoidal bias brought by the rejection sampling implemented in Section 1.5.2. In this case significantly more expensive algorithms are required to obtain the correct posterior distribution. In subsequent chapters we observe smooth posterior distributions with both Emcee and MultiNest in the parameter spaces relevant to the EoR, so we are confident this is not an issue in this work.

Parallelisation

MultiNest is parallelised with MPI in two places. Firstly when filling the initial set of live points. Secondly, as the acceptance ratio decreases due to the nature of ellipsoidal rejection sampling (discussed in Section 1.5.2) multiple samples are taken to increase the chance that a sample is drawn within the time-span of a single likelihood call. The reciprocal of sampling efficiency is recommended for the number of cores. Particularly as dimensionality increases,

this is the main bottle neck of the algorithm.

Usage

We have used the following settings for MultiNest throughout this work: live_points=2000, max_iter=0, multimodal=True,
evidence_tolerance=0.5, and sampling_efficiency=0.3. All of the other possible variations are left at default[5].

The MultiNest algorithm interface has the following main inputs:

- ndim - the number of dimensions (parameters in our cases).
- sampling_efficiency - recommend 0.3 or 0.8 for parameter estimation or model selection respectively. The reciprocal of this value represents the buffer added to the proposed ellipsoidal fit to the live-points.
- evidence_tolerance - 0.5, once $\Delta\mathcal{Z}$ reaches this value per iteration the algorithm is stopped.
- n_live_points - the number of 'live points' (used to estimate each isolikelihood contour).
- max_iter - sets the maximum number of sample iterations, to be sure the algorithm has hit its stopping criterion we set it to unlimited (max_iter = 0).
- multimodal = $True$ - Runs the clustering algorithm discussed in Section 1.5.2. In brief it performs Equations C.22 and C.23 separately on multiple modes (requires importance_nested_sampling = $False$).

These outputs are useful for book keeping but have no significance statistically:

- n_iter_before_update - how many iterations to perform before writing the output (default 100).
- outputfiles_basename - Define the strings for the output locations. Must be less than 100 characters (hard-coded into the FORTRAN90 memory allocation).
- resume - to continue from the previous run or overwrite (live_points must match).
- write_output - Save output to file (default $True$).
- verbose - how much output to write (bool).
- dump_callback - (default $None$) A callback function can be defined separately and inputted

[5]For the full details please see https://github.com/JohannesBuchner/MultiNest and https://github.com/JohannesBuchner/PyMultiNest

here to analyse information on the current iteration.

• context - (default $None$) Options for the user to pass extra information to MultiNest e.g. should an extra algorithm be included any extra information that is needed one can do it here.

The following options were found to be superfluous for our purposes. and are included for completeness:

• importance_nested_sampling - Include importance nested sampling (discussed in Appendix C.3.1) (default $False$).

• seed - (default -1) The random seed for the sampling, default uses the system clock.

• wrapped_params =None, if parameters have periodic boundary conditions this can applied here.

• const_efficiency_mode=False, the buffer on the ellipsoidal fit the live points varies to ensure a constant sampling efficiency if set to true.

• nparams - should some parameters not be orthogonal and exist in the subspace of other parameters, this can be specified.

• log_zero - likelihood values smaller than this will be ignored (default -1e+100).

• mode_tolerance - the rigour within which points are assigned to modes (if multimodal = $True$).

• n_clustering_params - only used if the clustering of multimodal posteriors occurs in a subset of the initial parameter space (if multimodal = True).

• max_modes - Iteration maximum for the G-means algorithm (requires multimodal = $True$)

• null_log_evidence - minimum Evidence value to define a posterior peak as a mode. (default -1e+90, requires multimodal = $True$)

• init_MPI - initialising MPI. This is only necessary if the user is activate initialising MPI before the call to the sampler - if you are not, MultiNest will do it for you (with init_MPI = $False$).

Importance nested sampling

Importance nested sampling (Feroz et al., 2019) is a method for including the discarded live points into the Evidence calculation. Although we do not use importance nested sampling in this work directly, we have used it as a cross-check for our Evidence calculations. The error in

the nested sampling algorithm is reduced by including more points in the integration. Namely if we consider a sampled estimate of the posterior f, a successful representation of the posterior will yield,

$$1 = \int \frac{\mathcal{P}}{f(\boldsymbol{\theta})} f(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{C.35}$$

We can therefore write,

$$\int f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int \Pi\mathcal{L} d\boldsymbol{\theta} \times \int \frac{f}{\Pi\mathcal{L}} \mathcal{P} d\boldsymbol{\theta}. \tag{C.36}$$

and with some rearranging,

$$\frac{\int f(\boldsymbol{\theta}) \Pi(\boldsymbol{\theta}\mathcal{L}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int f(\boldsymbol{\theta}) \mathcal{P} d\boldsymbol{\theta}} = \int \Pi\mathcal{L} d\boldsymbol{\theta}, \tag{C.37}$$

i.e. the rejected samples (sampled directly from the $\Pi\mathcal{L}$) are summed separately to the $\mathcal{P}$ values (which are calculated with the prior weightings from the live and dead points) and then combined to give an estimate of $\mathcal{Z}$. If both methodologies are successful the value of $\mathcal{Z}$ will be unchanged by including the extra points.

Although including these extra points reduces the error on the calculation it prevents MultiNest's multimodal capabilities from working (Section 1.5.2). It also slows down MultiNest (significantly so in our high, 8 parameter model). Its usage in our work is limited because of this, and only in the toy models testing of MultiNest. When used, the results show agreement with regular nested sampling in Chapter 3, it has not been used in any other chapter.

See Numerical Recipes (Press et al., 2007) for general detail on the algorithm and Feroz et al. (2019) for its implementation within MultiNest.

Algorithms in older versions of MultiNest

The predecessors to the MultiNest of today don't implement the expectation minimisation scheme discussed in Section 1.5.2 (Feroz & Hobson, 2008). Instead either G-means (Hamerly & Elkan (2004); Pham et al. (2005), Algorithm 8) or X-means (Pelleg & Moore (2000), Algorithm 9) clustering is used to decide the how many k's with which the k-means algorithm should proceed.

Algorithm 8: G-means Clustering: deciding the number of Gaussian data subsets within a data set. The resulting distribution will be k Gaussian's will be each data point saved within a Gaussian shaped mode.

Run the k-means (Algorithm 5) with $k = 1$;
Test the resulting distribution with the Anderson Darling test, (Equation 1.34) where f is a Gaussian;
while *Data points unsaved* **do**
  if *Anderson Darling = Passed* **then**
    Assign data to set and save it;
  end
  Split the failure mode in half;
  Repeat k-means on this unsaved modes set with k=k+1;
end

Algorithm 9: X-means Clustering: deciding the K in K-means. Here k_a is estimated as the optimum number of data subsets.

Run two sets of k-means (Algorithm 5)
with $k's$ set as $k_a = 1$ and $k_b = 2$;
Calculate the BIC_a and BIC_b for each set (Equation C.25);
while $\text{BIC}_a < \text{BIC}_b$ **do**
  Increase $k_a = k_a + 1$ and $k_b = k_b + 1$;
  Repeat K-means algorithm k_a times, with new mode started in each $1/k_a$ th;
  Recalculate the BIC_a and BIC_b ;
end

Both produce results similar to a Voronoi cell distribution. X-means was found to be more reliable, but implementations of the k-means algorithm are prepared within MultiNest up to $k = 7$.

C.3.2 Convergence checks

To check the convergence of any set of MCMC chains we perform a Gelman-Rubin test (Gelman & Rubin, 1992). If all of our chains have converged, they will all be in the posterior's global maxima. Therefore the variance on any chain of points selected at random (σ_i), should be the same as that calculated accross all chains ($\sigma_{\rm all}$). Our algorithm passes the Gelman-Rubin convergence test if,

$$\frac{\sigma_{\rm all}}{\sigma_{\rm i}} < 1.03, \tag{C.38}$$

where 1.03 is a threshold defined by the user.

For nested sampling algorithms there is no chain, a different convergence test is required. We implement the the algorithm called NestCheck developed by Higson et al. (2019) for this task specifically. Figure C.6 shows the diagnostic plot produced by NestCheck. This illustrates the uncertainty on the posterior between multiple runs, as well as the changes in posterior mass against change in prior volume (X). For a well converged run these posteriors will agree and the majority of the sampling will have occurred at the posterior peak.

As a final cross check other methods of calculating the Evidence can be used to insure convergence. For example, importance nested sampling and the standard nested sampling method will agree on the Evidence when the algorithm is converged (as long as the posterior is not multimodal).

C.3.3 Alternatives

Here we briefly explore some alternative methods of Evidence calculation.

Bambi (Graff et al., 2012) implements MultiNest however as the likelihood space is explored a neural network is trained. Once the network can give as good likelihood values as the original model, costly computation can be cut saving computation by $\sim 30\%$.

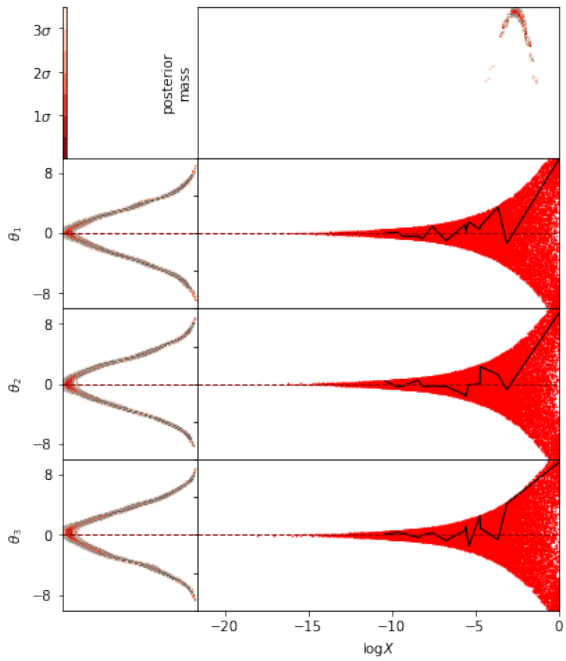


Figure C.6: An example convergence check for nested sampling. Here MULTINEST has been iterated 4 times on a 3D Gaussian likelihood with uniform priors. Since all four runs show similar posteriors (with deviation indicated by the thickness of the red lines on the left) we are confident this has converged. The closest analogy to the MCMC chain are the evaluations along each thread, in this 16 core example the black line indicates the samples produce by one thread selected at random.

As mentioned in Section 1.5.2, MULTINEST suffers from the curse of dimensionality above $n = 10$, becoming unusable by $d = 30$. Above $d = 10$ this the authors recommend the use of POLYCHORD (Handley et al., 2015), where slice sampling (Neal, 2000) is implemented in place of the ellipsoidal rejection sampling. Samples are drawn along the bisectors between live points ensuring the iterations draw points of increasing likelihood. Although this is an effective way of emulating iso-likelihood contours as the prior region reduces, it can become inefficient in the case of correlated parameters.

DYNESTY (Speagle, 2020), implements nested sampling but with a variable number of live points. As the algorithm progresses, functions decide if more live points need sampling or whether the sampler can iterate inwards to the next iso-likelihood contour. The user can also choose one of four methods to approximate the iso-likelihood contours: uniform sampling, random walks, multivariate slice sampling, and Hamiltonian slice sampling. In practice it performs similarly to MULTINEST.

A different approach altogether is PRISM by van der Velden et al. (2019). In brief, the likeli-

hood space is sampled loosely, with the gaps covered with Gaussian regression. The Evidence is them numerically integrated easily from the resulting function, however the likelihood space is biased towards Gaussian shapes.

Heavens et al. (2017a) have developed an approximate method for calculating the Bayesian Evidence directly from MCMC chains called MCEvidence. This is looked at in detail in Chapter 2.

We are considering using BAMBI and POLYCHORD in future work.

C.4 Statistics in Astrophysics

In this section we discuss how the likelihood is created in the context of astrophysics in rest of this work. Namely this is done with a power spectrum, which is detailed in the following sections.

C.4.1 The correlation function

To begin, lets consider the cosmological assumptions made in Chapter A, namely that the universe is homogeneous and isotropic. Considering the distributions of galaxies, if this were exactly the case the number of galaxies, n, would be Poisson distributed around a selected galaxy. This can be written,

$$P_{\text{poisson}}(n|\boldsymbol{x}) = \frac{(r\boldsymbol{x})^n e^{-r\boldsymbol{x}}}{n!}, \tag{C.39}$$

where r is the mean number of galaxies per distance interval $\boldsymbol{x}$ (in any spatial direction). Since this is not exactly the case, we require whats called the correlation function, ξ. This can be defined as the excess probability required to correctly measure the clustering of galaxies when their distribution is compared to Poisson statistics,

$$P_{\text{True}}(n|\boldsymbol{x}) = P_{\text{poisson}}(n|\boldsymbol{x})[1 + \xi(\boldsymbol{x})]. \tag{C.40}$$

As long as our observation is large enough to be ergodic, we can describe precisely the deviation from the cosmological principle.

C.4.2 The power spectrum

A power spectrum is defined as the Fourier transform of the two point correlation function,

$$P(\boldsymbol{k}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\boldsymbol{k}\boldsymbol{x}} \xi(\boldsymbol{x}) \; d\boldsymbol{x}, \tag{C.41}$$

where $k = \frac{2\pi}{\lambda}$ is the wave-number of the distribution and the $\frac{1}{\sqrt{2\pi}}$ factor is a normalisation choice. To Fourier decompose the correlation function, as above, accesses valuable information about clustering. If the signal from galaxies is used (following the example in the previous subsection), we obtain the galaxy clustering power spectrum. This illustrates the number density of galaxies on any particular scale. Another common application is using the power spectrum to measure the temperature oscillation accross the CMB. This is how the BAO scale is measured in Chapter A.

For the 21cm signal in the cosmic dawn and reionisation the brightness temperature is used in the power spectrum. In all subsequent chapters, a spherically averaged analogue is implemented per cube of brightness temperature signal. This is defined per redshift (or line of sight/frequency mode) as,

$$\Delta^2(\boldsymbol{k}) \equiv \frac{k^3}{2\pi^2 V} P(\boldsymbol{k}). \tag{C.42}$$

As we will see in Chapter 5 this runs in to observational intricacies involving observations of the light-cone.

For more information, Peacock (1999) contains a plethora of Cosmological statistics as well as full detail on the power spectrum, correlation function and their applications.

C.4.3 Further astrostatistics

It is an open question as to is whether the power spectrum will capture enough information to satisfy the questions at hand about the EoR and cosmic dawn. Clearly when forward modelling this is not an issue (since it can be measured via an analytic calculation). However the inverse is not obviously true, and so we need be sure that what we put into the likelihood function can determine enough of the system to answer our desired questions. Two 21cm brightness temperature maps with the same power spectrum can appear vastly different (Mondal et al.,

2016, 2015). In the context of 21cm tomography, if reproducing the exact image is desired more information is required than just the power spectrum. To end, we allude the interested reader in three cosmological directions within which the bi-spectrum is showing promise in measuring the non-Guassianity emerging from various signals. Firstly, the clean 21cm bi-spectrum can probe the magnitude and evolution of X-ray heating in the IGM (Watkinson et al., 2017). Secondly, the gravitational and lensing-ISW effect bi-spectrum (Schmit et al., 2019) can improve constraints on cosmological parameters, testing modifications to Einstein's gravity. Finally, the kSZ-kSZ-21cm bi-spectrum shows promise in detailing the reionisation history at scales where the 21cm signal alone would typically suffer from foreground contamination (La Plante et al., 2020).

Appendix D

Supplementary 21cmNest Analysis

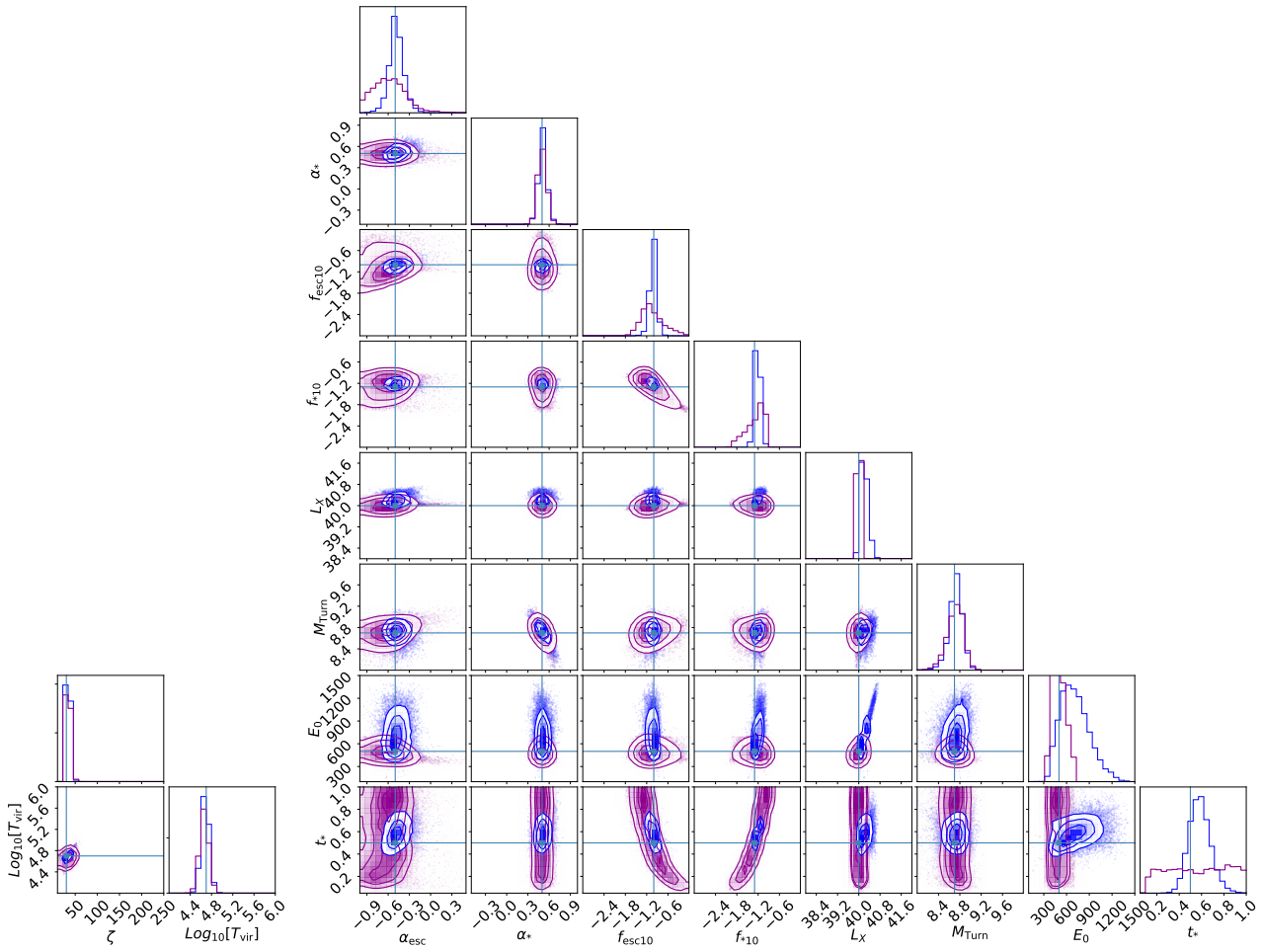
This appendix contains material that supplements Chapter 4.

D.1 *Cross-check*: retrieval of fiducial parameters

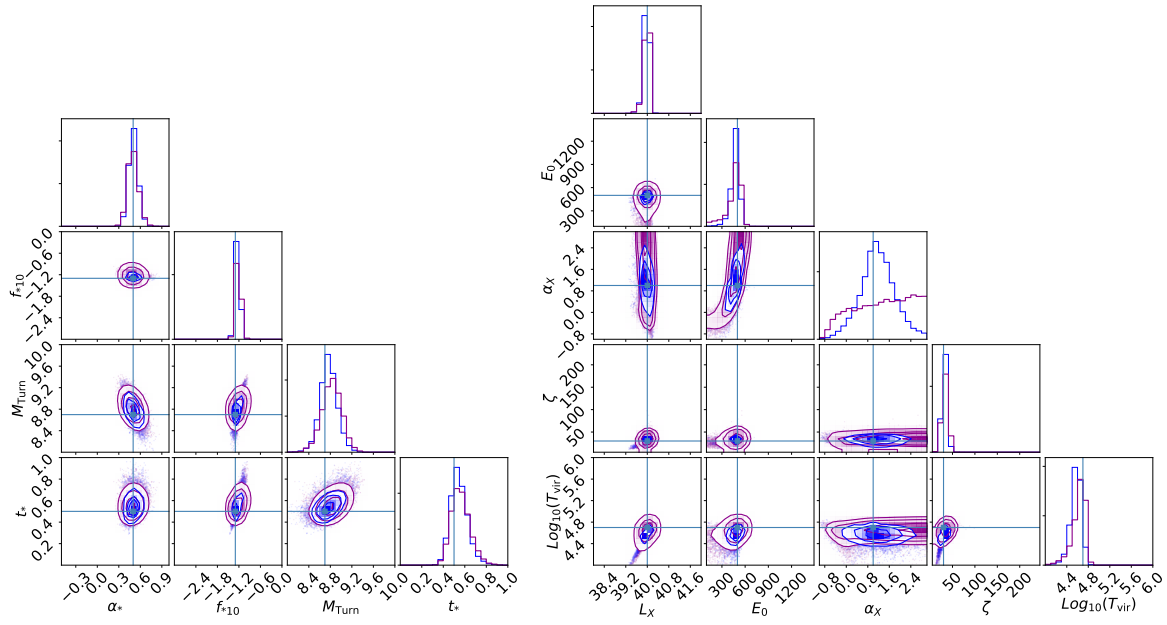
The posteriors obtained by both 21cmNest (magenta) and 21cmMC (blue) as each model fits the mock data set to find the fiducial parameters in Figure D.1. Mock fiducial 21cm PS data is simulated for each model with the parameters in Table 4.1. All Figures D.1(a)-D.1(b) show parameter posteriors and contain blue lines to indicate the fiducial parameter used for the mock observation.

The Bayes factors and MAP parameter sets are shown in Tables 4.4 and 4.5, MAP 21cm PS are plotted for the MAP parameters in Figure 4.5, including every models' best fit against each mock data set.

Note the posterior masses in D.1(d) and D.1(c) for $\alpha_{\rm X}$ and t_* respectively are not peaked, but they contain the fiducial mock data set values within their posterior masses (validating the results). However we are sceptical of their shape and this motivates using them for the SDDR analyses in Section 4.5.4. We will address this again when the posterior distributions calculated using the corrected box size have finished.



(a) A|A (b) D|D



(c) C|C (d) B|B

Figure D.1: Posterior distributions as each model fits for the fiducial parameters from their own mock data set. When the posterior has a clear peak, agreement is obtained in all the cases between the MAP parameters in 21CMMC (blue) and 21CMNEST (magenta) with the fiducial parameter choices indicated by the blue lines.

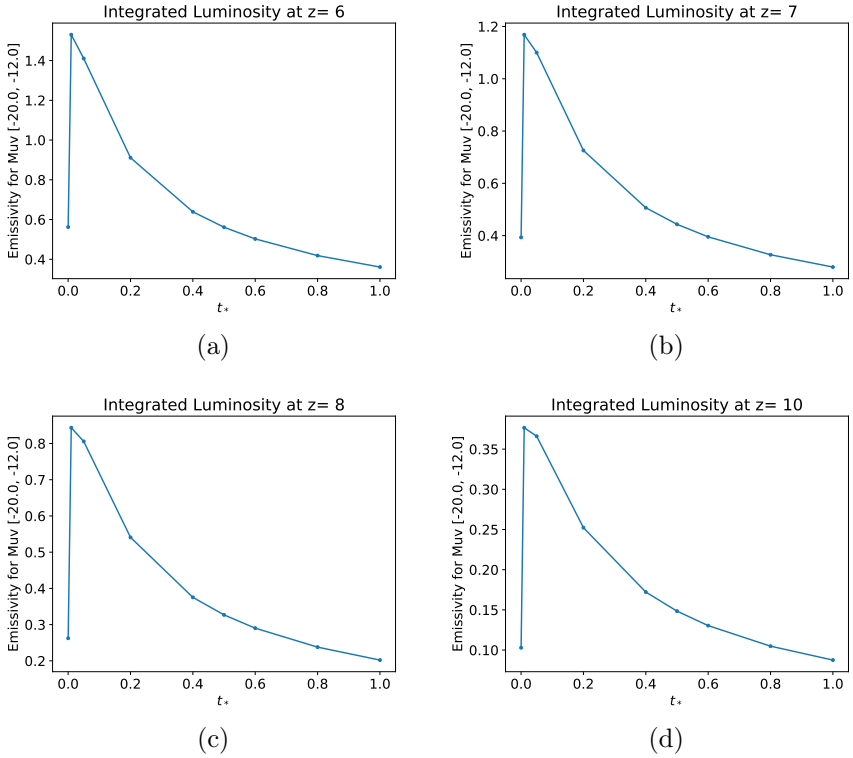


Figure D.2: Integrating the UV Luminosity functions to obtain the specific emissivity of the galactic ensemble. By construction Equation 4.13, our UV luminosity produce emissivity values in agreement with high redshift observation of reionisation galaxies.

D.2 *Cross-check*: integrating the UV luminosity function

The specific emissivity of the galactic ensemble species are obtained by integrating the UV LFs. Since the recipe described in Section 4.3.3 is closely linked to observations the emissivities share a similar shape for each redshift. For a given value of t_* the specific emissivity is decreasing with redshift and therefore agrees with the observational data thanks to Equation 4.13 (Bouwens et al., 2015; Oesch et al., 2018). Figure D.2 shows the integrated luminosity functions as t_* is varied throughout the parameter prior. Each curve is smooth until a sharp drop off as t_* becomes unphysical below 0.05. This confirms the conclusions established with the SDDR in Section 4.5.4.

Appendix E

Supplementary Morlet Power Spectrum Analysis

This appendix contains material that continues the analyses in Chapter 5.

E.1 Interpreting the Morlet wavelet data cube

In an attempt to interpret the Morlet wavelets relationship to physical quantities, we calculate both the FPS and the MPS for five differing light-cones ranging from redshifts $z \in [8, 10]$). These are produced using the simple scenario in Chapter 4 (Model A). Slices of the brightness temperature, density field and ionisation field for each light-cone is plotted in Figure E.1. The choices of fiducial parameter are $[\zeta, \mathrm{Log}_{10}T_{\mathrm{vir}}]$): $[30, 4.7]$, $[5, 4.7]$, $[60, 4.7]$, $[30, 4.0]$, and $[30, 5.3]$. Throughout all the Figures in this section these are colour coded light blue, red, blue, green, and magenta respectively. Although this model is simplistic we have chosen it for its ease of use in these preliminary tests, the parameters are selected to approximately cover the parameter flexibility within the prior distributions used in Chapter 4. The third and fourth parameter selections in particular have a reionisation that finishes unphysically early and are therefore visibly different from the other selections in Figure E.1.

For ease of comparison we have included the power spectra from five chunks of each light-cone in Figure E.2, as well as the averaged quantities themselves directly in Figure E.3. We have included the Fourier power spectra for each quantity, please note that the brightness

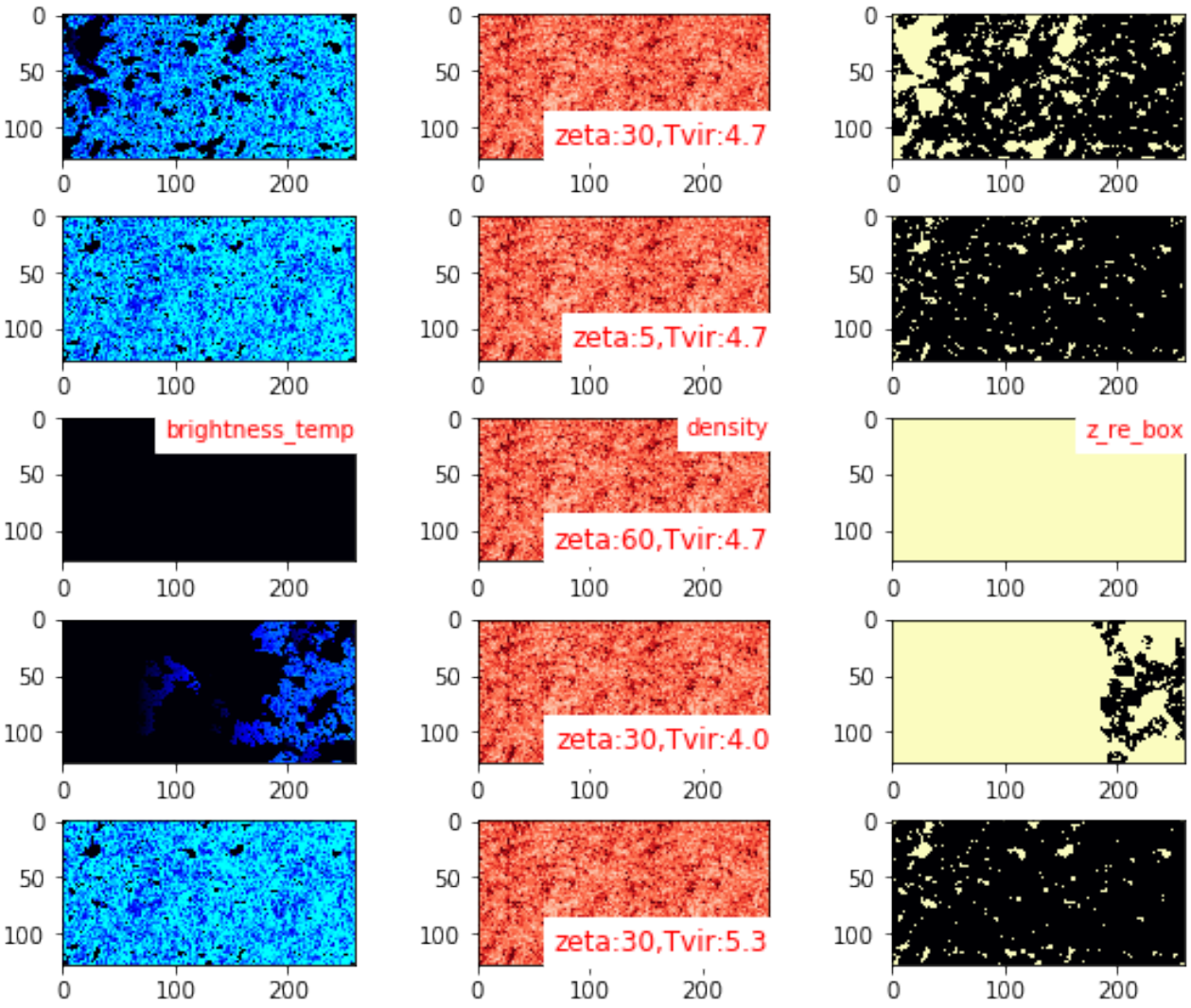


Figure E.1: Slices of brightness temperature, density and ionisation through test light-cones produced by a simple two parameter 21CMFAST model (for more detail see Model A in Chapter 4). Each slice is produced with 128 pixels across a 250 Mpc box with redshifts 8-10.

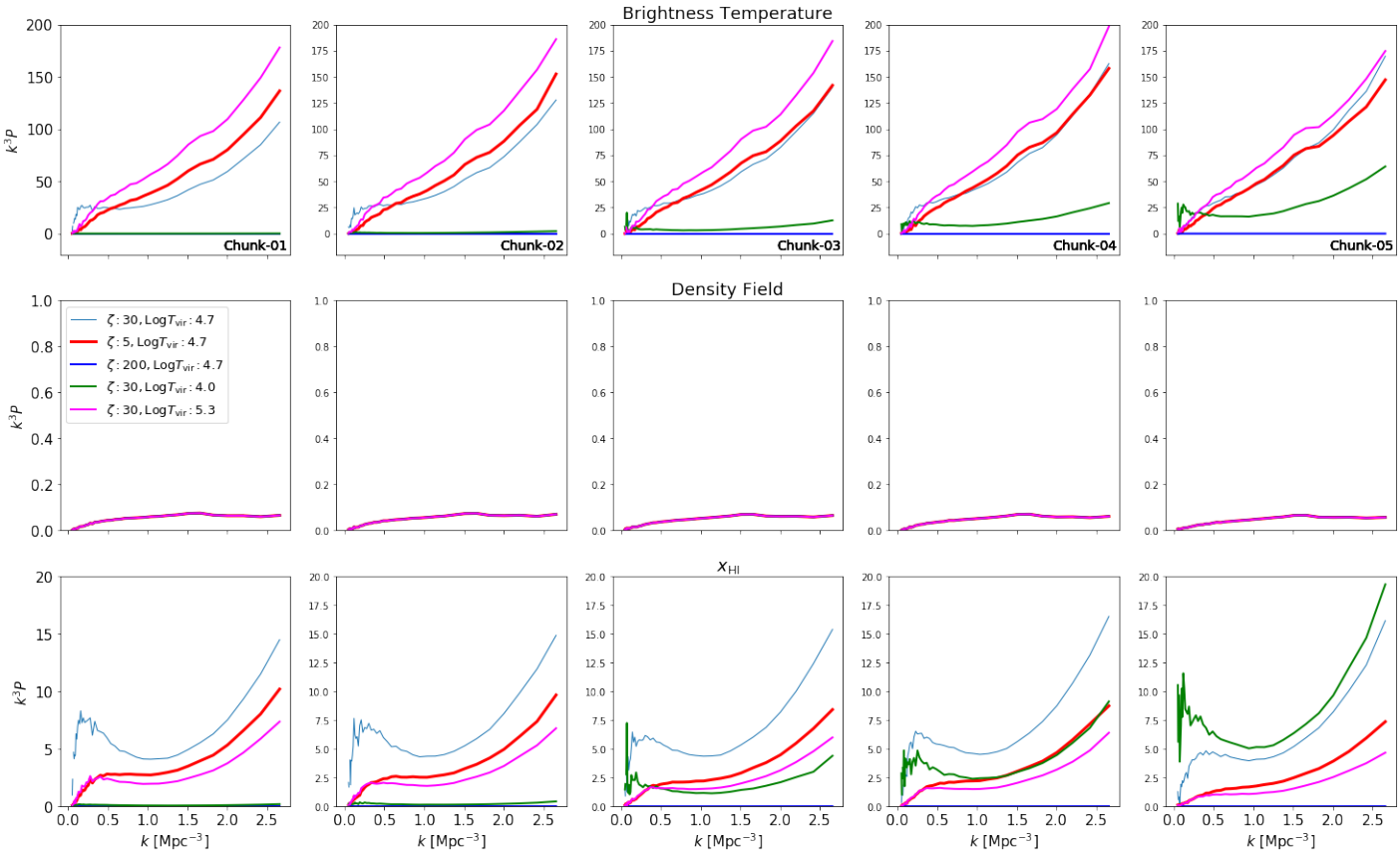


Figure E.2: Fourier power spectrum calculations from each of the light-cone quantities in Figure E.1. As can be seen, the average amplitude of each power spectrum in each light-cone chunk (column) follows closely the behaviour of the light-cone quantities shown in Figure E.3. See text for more discussion.

temperature power spectrum is the one desired for practical use. The others are included to show that the power spectrum is excellent at tracing the behaviour of the quantity it is measured from. Between each chunk the average of the power spectrum for each quantity follows the trajectories of the quantity average directly for both the neutral fraction and brightness temperature. The density field quantity is measured with, $\delta = \frac{\rho}{\bar{\rho}} - 1$, hence it remains constant throughout, the power spectra reflect this until limitations of the box size appear around $k < 0.025\text{Mpc}^{-1}$ $(= 2\pi/250\text{Mpc})$ when power is lost.

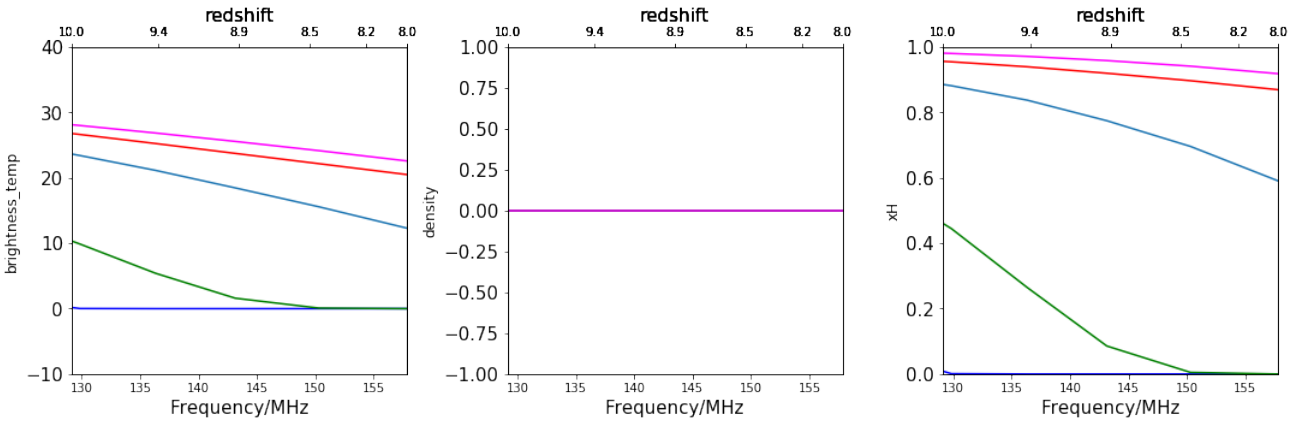


Figure E.3: Averages of the light-cone quantities in Figure E.1. From left to right these are 21cm brightness temperature [mK], density field (in fluctuations from the mean), and neutral fraction.

Every possible slice of the Morlet power spectrum data cube for each of the five light-cones is shown in Figure E.4. This has been done before any normalisation of the k modes, meaning the full data cube consists of MPS[mK^2]($\mathrm{k}_\perp, \mathrm{k}_\perp, \mathrm{k}_\parallel, \eta$). For the extreme cases of the light-cones, when reionisation finishes early, it is clear that signal is lost in slices containing $k_\parallel$. Other than this obvious diagnosis, the figure is more aesthetically pleasing than it is useful within this range of toy light-cones. Figure E.4(a) is normalised across the $k_\perp$ modes before being shown in Figure E.9(a).

Subsequent figures are created after averaging across the $k_\perp$ modes, and selecting the first mode in the array. Figures E.5 and E.6 show the MPS plotted against $k_\parallel$ for various steps of the central wavelet coordinate η as it moves throughout the light-cone. The latter (E.6) has the MPS normalised with a factor of $k_\parallel^3$ to reduce spacial dependency. Evolution across the line of sight of the light-cone is not apparent in E.5 and E.6 but the ordering of the MPS magnitudes resembles that of the FPS brightness temperature plots for chunks 3 and 4. To elaborate at the larger k modes the light-blue, red and magenta lines are grouped at high k, with the light-blue and red lines overlapping. At low k the light-blue and blue lines are significantly lower than the others and becomes negligible when normalised, as is expressed in both the FPS and the averaged brightness temperature quantity. The green light-cone undergoes the fastest change in ionisation field within the selected redshifts, it therefore changes the most between FPS chunks and has the largest change in magnitude across Figure E.5. Consequently a larger magnitude is seen in chunk 5 of the FPS plots than characterised by the averaged quantity, and this becomes zero in chunks 1 and 2 when the average quantity approaches zero. When normalised, Figure

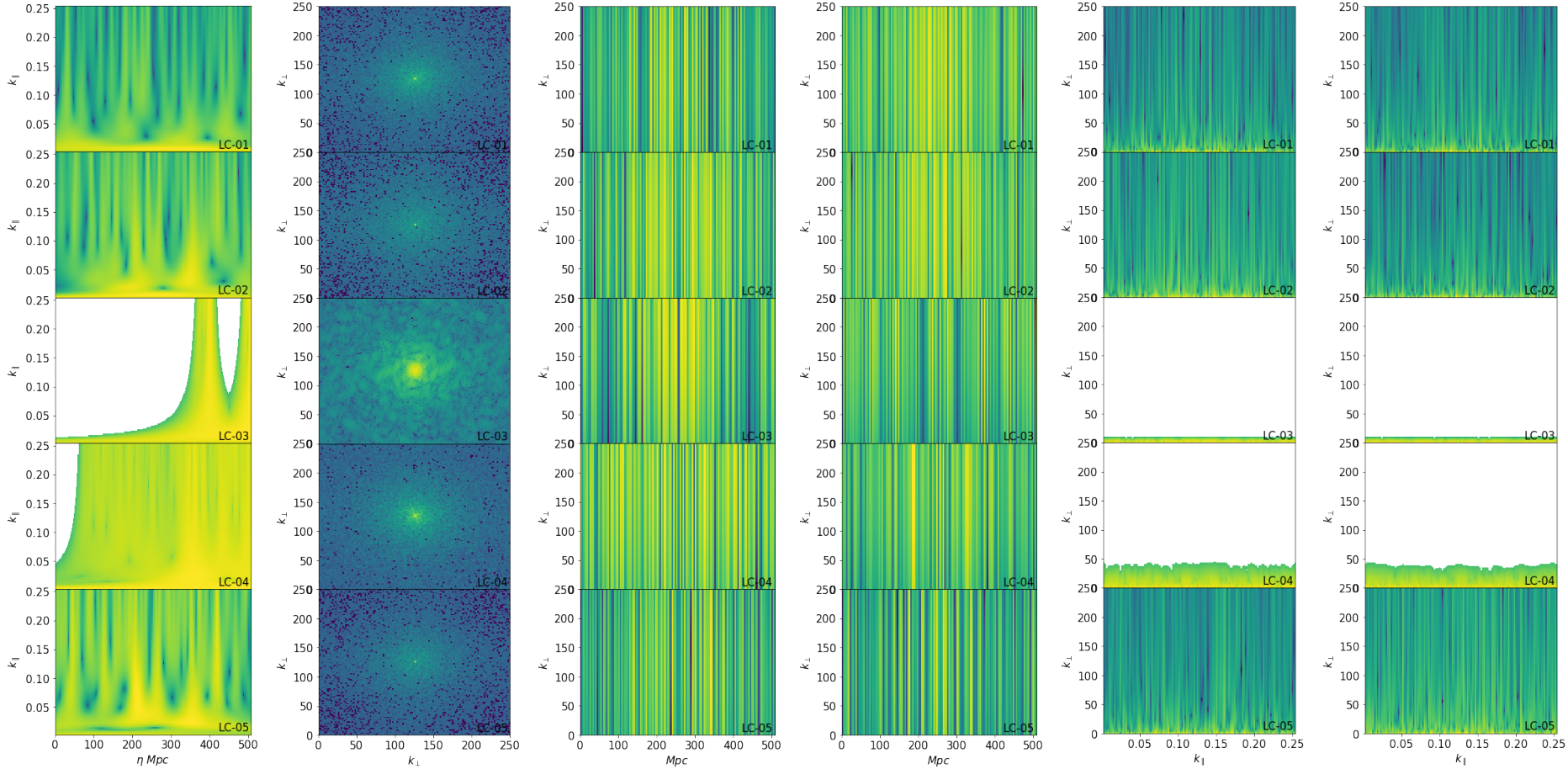


Figure E.4: Slices through the Morlet wavelet produced from the light-cones in Figure E.1. The total data cube consists of $k_\parallel$, η, $k_\perp$, $k_\perp$, each slice here uses the first element of the two unused arrays. From left to right these Figures have axis $(k_\parallel, \eta)$, $(k_\perp, k_\perp)$, $(k_\perp, \eta)$, $(k_\perp, \eta)$, $(k_\perp, k_\parallel)$, and $(k_\perp, k_\parallel)$.

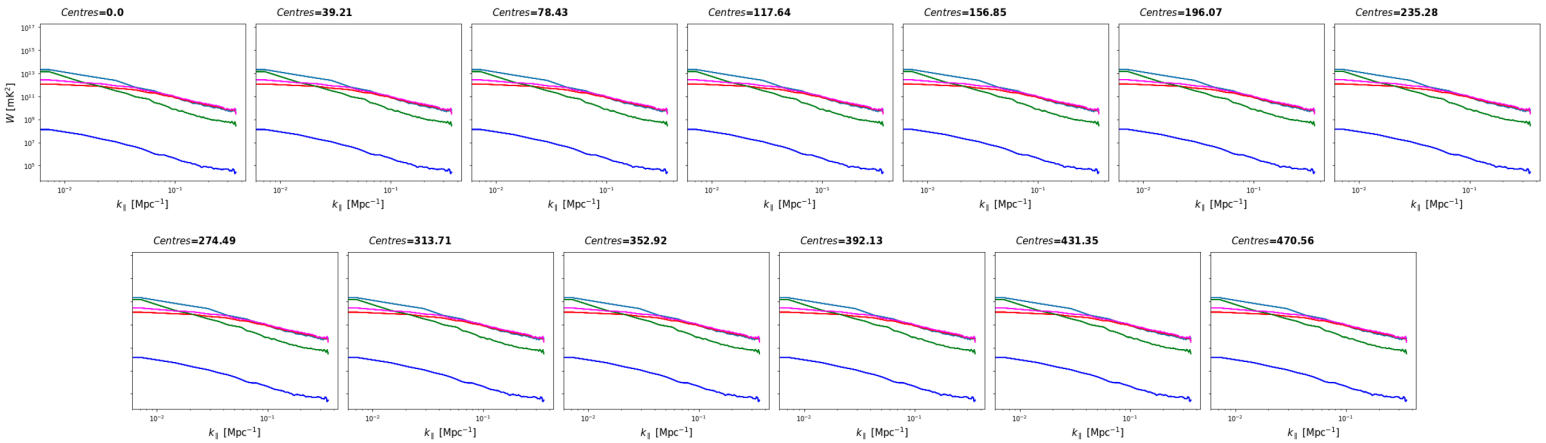


Figure E.5: Here the MPS wavelets are plotted against $k_{\parallel}$. Moving left to right increases the wavelet centre (η) along the light-cone line-of-sight. See text for the discussion (involving Figures E.6, E.7, and E.8).

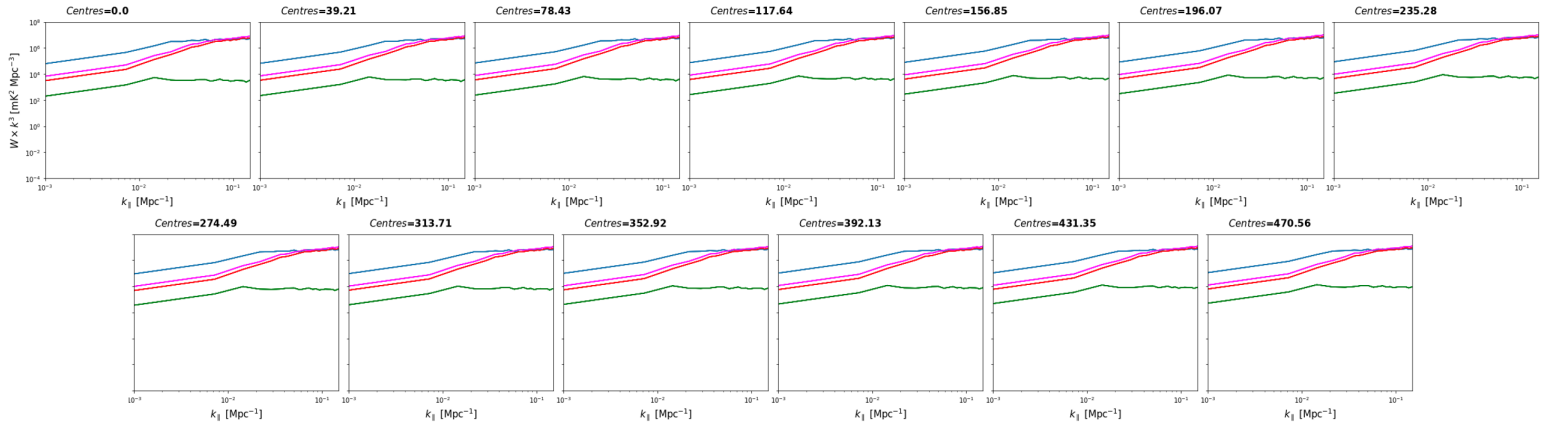


Figure E.6: As in E.5 however the MPS are plotted with a normalising factor of $k_{\parallel}^3$. The dark blue line (the third light-cone with parameters $[\zeta = 200, \mathrm{Log}T_{\mathrm{vir}} = 3.7]$) has the smallest signal by several orders of magnitude and becomes negligible when normalised with $k_{\parallel}^3$. This Figure is the closest of the MPS figures to reproducing the information presented with the FPS.

E.6 shows significantly less signal per mode as is the case with the FPS and the averaged quantity.

Figures E.7 and E.8 are the MPS and MPS$\times$ $k_{\parallel}^3$ plotted against η for various steps through the $k_{\parallel}$ modes. The $k_{\parallel}$ normalisation causes the models to separate a little however the increase in clarity is minor. These Figures show the evolution of the MPS per position of the wavelet centre along the light-cone line-of-sight. More signal comes from the furthest wavelet in all cases. For a constant signal each mode should contain equal amounts of MPS signal because the wavelets are adapting as they move along the light-cone (see Trott (2016) or Equation 5.2). The lack of evolution in the zeroth mode ($k_{\parallel} = 0$) is therefore evidence that the light-cone effect

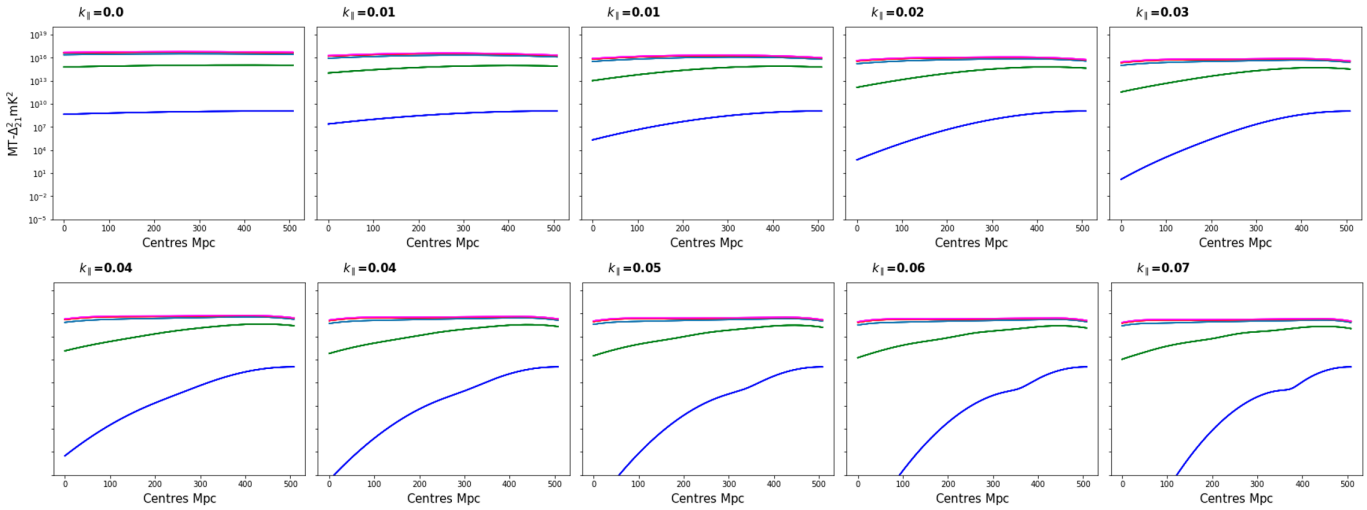


Figure E.7: As in Figure E.5 but here the wavelets are plotted against the centre of the wavelet (η, along the line-of-sight). From left to right the plots show an increase in $k_\parallel$ mode which the data cube has been sliced along.

has been removed from the signal. At larger values of $k_\parallel$, the patchiness of the signal in a late reionisation scenario causes significant decreases to the MPS values. Plotting the MPS$\times$ $k_\parallel^3$ against η is a useful in diagnosing when the end of reionisation is near, or that the wavelets envelope is not correctly adapting to correct for the light-cone effect.

The evolution of Figures E.7 and E.8 follow closely the progression of $x_{\rm HI}$. The red and magenta lines (light-cones 2 and 5) overlap almost exactly throughout since they have the least progression of $x_{\rm HI}$ across Figure E.3. Light-blue (light-cone 1) is initially among this pair but loses MPS power when the wavelet centres approach 0 Mpc, i.e. the near side of the light-cone (or low η). Similarly when comparing the three brightness temperature light-cone slices for these three models, light-cones 2 (red) and 5 (magenta) are visually similar across the redshift range while 1 (light-blue) visually progresses the most of these three. In comparison light-cone 4 (green) has a drastic change in $x_{\rm HI}$ which is more visible in the large scale line of sight modes in Figure E.8 (small $k_\parallel$). When the light-cones are close to the end of reionisation (as in the case of light-cones 3 and 4, the green and blue lines), it is clear that there is a larger power drop from the larger modes. This is the precisely the behaviour of the inside-out morphology prescribed by the FZH toy model since smaller bubbles ionise first. Light-cone 3 (blue line) has a noticeably lower magnitude because reionisation has essentially finished. The signal is not exactly zero because there is a tiny bit of $x_{\rm HI}$ below 130Mhz.

So far, the power spectrum is a better reflection of the light-cone brightness temperature average

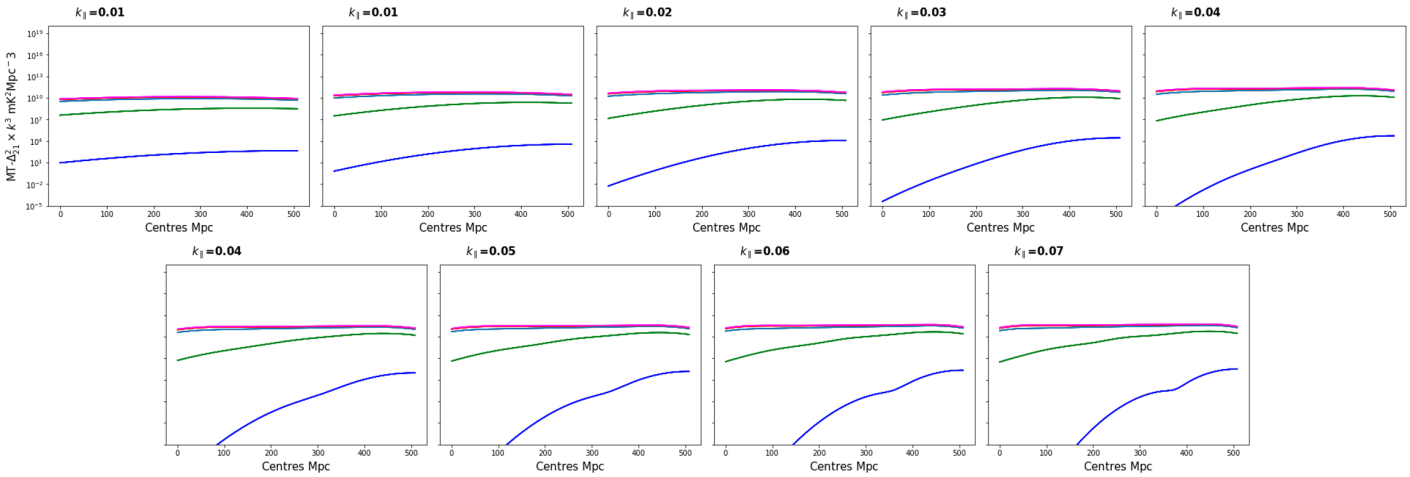


Figure E.8: As in Figure E.7 however the wavelets are plotted with a normalising factor of $k_\parallel^3$. This Figure shows the signal evolution throughout the light-cone across a given mode and resembles the neutral fraction in Figure E.3.

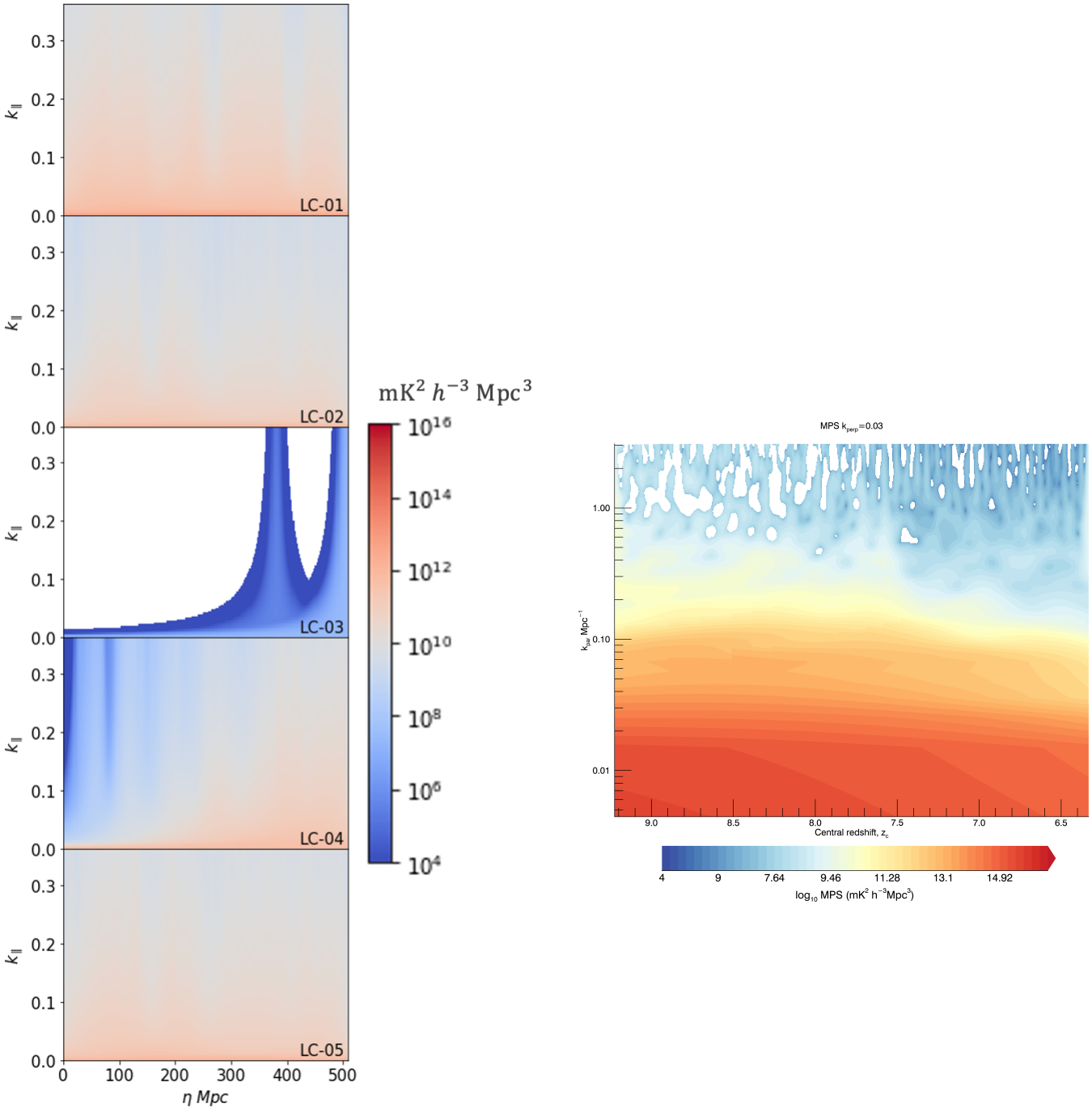
quantity. Although the MPS is ergodic across the light-cone, interpreting the light-cone all at once is difficult. Since the FPS is measured on well defined chunks of the light-cone, the evolution between these chunks is easy to interpret. That's not to say that the MPS can't express similar information to the FPS, but the information content of the MPS is closer to trying to interpreting an average of FPSs made from $\sim$ 100s of chunks (precisely the number of pixels). In future work we would try to interpret smaller chunks of light-cone with the Morlet transform for example. Since we have also only taken the first $|k_\perp|$ mode for Figures E.5, E.6, E.7, and E.8 it is likely there would be a better choice. From Figure E.1, the $(k_\perp, k_\perp)$ direction clearly has some structure however we are currently unable to elaborate on this conclusively. One possibility is that the peaks of each Gaussian wavelet envelope are stacked in the same place causing the increase in power in the centre for each light-cone.

Once the k modes have been cylindrically averaged $(|k_\perp|, k_\parallel)$ we learnt that it is better to include the $k_\parallel^3$ factor when plotting the wavelets in both cases. Figure E.6 is the the most informative single Figure for interpreting the light-cone structure with the MPS. The MPS per $k_\parallel$ expresses the evolution per scale in a similar way to averaging the FPS over all the light-cone chunks used. Figure E.8 expresses information on how much signal content is within each wavelet and is a useful check for the speed of ongoing ionisation and as cross-check for the consistency of power in each wavelet. These lines follow a similar behaviour to the neutral fraction average in Figure E.3. Future analyses could include looking into the relationship of ionisation rate and the slope the MPS$\times k_\parallel$ per η relationship. A combination of these two figures is therefore

the most useful in expressing the information within the five reionisation models. Finally we emphasise that we have experimented with a very limited range of reionisation models. This will need to be expanded in future work if the MPS is to be better understood.

E.2 Comparison with real data from MWA

A comparison of the 5 toy models described in Appendix E.1 are shown next to real MWA data (C. Trott, private communication) in Figure E.9. We include this section somewhat prematurely since the simulation and observation do not undergo the same treatment. The MWA data is an upper limit combining 110 hours of high-band observations with 55 hours of low-band (Trott et al., 2020). Missing bands are filled in with spectral cleaning algorithm CLEAN (Zhang et al., 2016) and the full MWA band is tapered with a Blackman Harris filter before being corrected for flux lost during normalisation. The simulated MPS plots have significantly lower MPS power at all scales, and are therefore in agreement with the MWA upper limits. In future work we will analyse whether tighter constraints on reionisation model parameters can be obtained than with the FPS (Greig et al., 2020b).



(a) (b)

Figure E.9: (a) contains slices through the Morlet wavelet produced from the light-cones in Figure E.1. (b) shows real data processed from MWA (credit: C. Trott) for an upper limit of an MPS measurement. Both figures are sliced at $|k_\perp| = 0.03$, and are plotted on a colour-map scale between 10^4 and 10^{16} mK^2.

Appendix F

Copyright Permissions

F.1 Permissions to reproduce material from publications

Permission for Figures 1.1, 1.2, 1.3, 3.1, A.1, A.2, A.3, A.6, B.2 from Barkana & Loeb (2001); Pritchard & Furlanetto (2006); Fan et al. (2006); Pritchard & Loeb (2012); Loeb & Furlanetto (2013); Zaroubi (2013); Watkinson & Pritchard (2014); Liu et al. (2014a) are listed below; as described in the declaration.

From: **Avi Loeb** aloeb@cfa.harvard.edu
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 7 January 2021 at 15:38
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

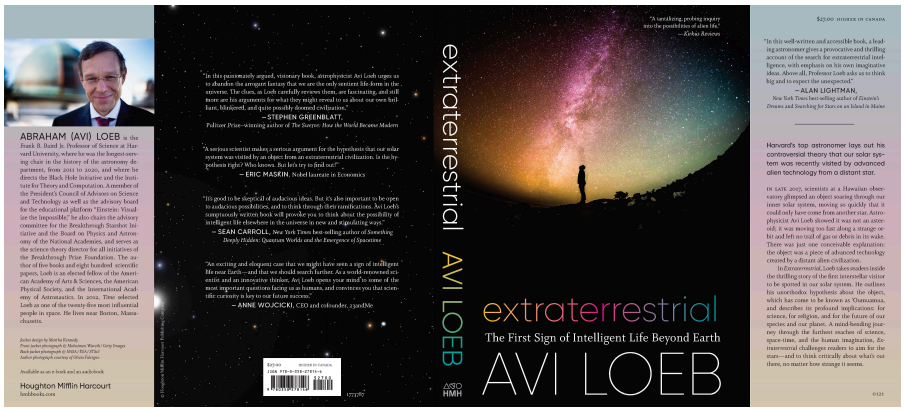
Yes, of course you have my approval, Tom. Good luck with your thesis.

Avi

~~~~~~

Abraham (Avi) Loeb ***(he, him, his)***
Frank B. Baird, Jr., Professor of Science, Harvard University
Director, Institute for Theory and Computation (ITC)
Founding Director, Black Hole Initiative (BHI)
Chair, *Breakthrough Starshot* Advisory Committee
Chair, Board on Physics and Astronomy, National Academies
Member, President's Council of Advisors on Science and Technology
Longest-Serving Chair (2011-2020), Harvard Astronomy Department

~~~~~~



On Thu, Jan 7, 2021 at 9:19 AM Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:

> Dear Prof Loeb,
>
> I wonder if you'd be able to help me.
> I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
> I seek your permission to include figures from your publication that significantly improve the explanations I provide in the introductory sections of my thesis.
>
> The figures in question are Figures 1.3, 1.6 and 3.11 from `The First Galaxies of the Universe' (Loeb & Furlanetto 2013) ; Figure 12 from `In the beginning: the first sources of light and the reionization of the universe' (Barkana & Loeb 2001); and Figure 1 from `21-cm cosmology' (Pritchard & Loeb 2012). Each Figure will contain a reference in the caption.
>
> My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours faithfully,
> Tom Binnie

From: **Steven Furlanetto** sfurlane@astro.ucla.edu
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 7 January 2021 at 17:48
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

SF

Hi Tom,

Congratulations on wrapping up your thesis!

I am happy to grant permission; however, I think the only one of these figures that I personally created was 3.11 from the book. The other two in the book were made by Avi Loeb, and the one from the paper was from Jonathan. So please be sure to check with them too!

Also, from past experience, I believe you need to check with the publishers for permission, as the copyright can transfer to them in some cases.

Good luck finishing up!
Steve Furlanetto

> On Jan 7, 2021, at 6:39 AM, Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:
>
> Dear Prof Furlanetto,
>
> I wonder if you'd be able to help me.
> I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
> I seek your permission to include figures from your publications that significantly improve the explanations I provide in the introductory sections of my thesis.
>
> The figures in question are Figures 1.3, 1.6 and 3.11 from `The First Galaxies of the Universe' (Loeb & Furlanetto 2013); and Figure 1 from `Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line' (Pritchard & Furlanetto 2006). Each Figure will contain a reference in the caption.
>
> My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours faithfully,
> Tom Binnie

Subject: Permission for figure reproduction in PhD thesis
Date: Thursday, 7 January 2021 at 16:43:47 Israel Standard Time
From: Binnie, Thomas J
To: Rennan Barkana

Dear Prof. Barkana,

I wonder if you'd be able to help me.
I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
I seek your permission to include figures from your publication that significantly improve the explanations I provide in the introductory sections of my thesis.

The figure in question Figure 12 from `In the beginning: the first sources of light and the reionization of the universe' (Barkana & Loeb 2001). The figure will contain a reference in the caption.

My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.

If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.

Yours sincerely,
Tom Binnie

I grant you all the permissions requested, and wish you the best of luck.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 07, 2021

This Agreement between Imperial College London -- Thomas Binnie ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4983640433574
License date	Jan 07, 2021
Licensed Content Publisher	Elsevier
Licensed Content Publication	Physics Reports
Licensed Content Title	In the beginning: the first sources of light and the reionization of the universe
Licensed Content Author	Rennan Barkana,Abraham Loeb
Licensed Content Date	Jul 1, 2001
Licensed Content Volume	349
Licensed Content Issue	2
Licensed Content Pages	114
Start Page	125
End Page	238
Type of Use	reuse in a thesis/dissertation

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation
Institution name	Imperial College London
Expected presentation date	Jan 2021
Order reference number	https://doi.org/10.1016/S0370-1573(01)00019-9
Portions	Fig 12, page 152.
Requestor Location	Imperial College London St Marys Manor Road Goring-On-Thames Reading, S. Oxon RG8 9ED United Kingdom Attn: Imperial College London
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

Terms and Conditions

From: **Xiaohui Fan** fan@as.arizona.edu
Subject: Re: [EXT]Permission for figure reproduction in PhD thesis
Date: 7 January 2021 at 15:37
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

Hi Tom,

Please go ahead and use the figure in your thesis.

Cheers,
Xiaohui

> On Jan 7, 2021, at 8:28 AM, Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:
>
> **External Email**
>
> Dear Prof. Fan,
>
> I wonder if you'd be able to help me.
> I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
> I seek your permission to include figures from your work that significantly improve the explanations I provide in the introductory sections of my thesis.
>
> The figure in question is Figure 1 from `Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars' (Fan et al. 2006). Each Figure will contain a reference to this in the caption.
>
> My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours faithfully,
> Tom Binnie

Xiaohui Fan
Regents' Professor of Astronomy
Associate Department Head
Department of Astronomy/Steward Observatory
University of Arizona
Tucson, AZ 85721 - 0065

Email: fan@as.arizona.edu
URL: http://sancerre.as.arizona.edu/~fan
Tel: 520-626-7558
Twitter: @xfan_astro

From: **Permissions** permissions@ioppublishing.org
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 13 January 2021 at 17:54
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

Dear Tom,

Thank you for providing confirmation that you have received the consent of the authors to reproduce content from AAS journals ***(as below)***.

I am happy to confirm that permission is granted and you need take no further action.

Please include the following alongside the material:

- the source of the material, including author, article title, title of journal, volume number, issue number (if relevant), page range (or first page if this is the only information available) and date of first publication. This material can be contained in a footnote or reference;
- for material being published electronically, a link back to the article (via DOI); and
- if practical and IN ALL CASES for works published under any of the Creative Commons licences the words "© AAS. Reproduced with permission".

This permission does not apply to any material/figure which is credited to another source in the AAS publication or has been obtained from a third party. Express permission for such materials/figures must be obtained from the copyright owner.

Kind regards,

Cameron

Copyright & Permissions Team
Sophie Milne - Rights & Permissions Assistant
Sophie Brittain - Rights & Permissions Assistant
Cameron Wood - Legal & Rights Adviser
Contact Details
E-mail: permissions@ioppublishing.org

For further information about copyright and how to request permission: https://publishingsupport.iopscience.iop.org/copyright-journals/
See also: https://publishingsupport.iopscience.iop.org/
Please see our Author Rights Policy https://publishingsupport.iopscience.iop.org/author-rights-policies/
Please note: We do not provide signed permission forms as a separate attachment. Please print this email and provide it to your publisher as proof of permission. **Please note**: Any statements made by IOP Publishing to the effect that authors do not need to get permission to use any content where IOP Publishing is not the publisher is not intended to constitute any sort of legal advice. Authors must make their own decisions as to the suitability of the content they are using and whether they require permission for it to be published within their article.

From: Binnie, Thomas J <t.binnie16@imperial.ac.uk>
Sent: 13 January 2021 14:58
To: Permissions <permissions@ioppublishing.org>
Subject: Re: Permission for figure reproduction in PhD thesis

Dear Cameron,

Thanks for your reply. I have attached the page in my thesis as well as the bibliography entry and a pdf of the permission from the Author.
Please let me know if there is specific wording you require to go with the Figure.

Best wishes,
Tom

On 13 Jan 2021, at 13:50, Permissions <permissions@ioppublishing.org> wrote:

Dear Tom,

Thank you for your enquiry.

AAS grants back to its authors the right to grant permission for reuse of their work. Therefore, please:

- Obtain consent from one of the original authors of the paper(s) from which you wish to reuse content.
- Notify us at this address once consent has been obtained from the original authors
- Give the original citation for the figure(s), table(s), or other material that you plan to reproduce as well as the information about where the material is to appear.

We look forward to hearing from you once you have obtained the author's permission.

Kind regards,

Cameron

Copyright & Permissions Team
Sophie Milne - Rights & Permissions Assistant
Sophie Brittain - Rights & Permissions Assistant
Cameron Wood - Legal & Rights Adviser
Contact Details
E-mail: permissions@ioppublishing.org

For further information about copyright and how to request permission: https://publishingsupport.iopscience.iop.org/copyright-journals/
See also: https://publishingsupport.iopscience.iop.org/
Please see our Author Rights Policy https://publishingsupport.iopscience.iop.org/author-rights-policies/
Please note: We do not provide signed permission forms as a separate attachment. Please print this email and provide it to your publisher as proof of permission. **Please note**: Any statements made by IOP Publishing to the effect that authors do not need to get permission to use any content where IOP Publishing is not the publisher is not intended to constitute any sort of legal advice. Authors must make their own decisions as to the suitability of the content they are using and whether they require permission for it to be published within their article.

From: Binnie, Thomas J <t.binnie16@imperial.ac.uk>
Sent: 12 January 2021 09:40
To: Permissions <permissions@ioppublishing.org>
Subject: Permission for figure reproduction in PhD thesis

Dear Sir/Madame,

I wonder if you'd be able to help me, I tried contacting journals.permissions@oup.com but was encouraged to try here.
I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
I seek your permission to include figures from your publication that significantly improve the explanations I provide in the introductory sections of my thesis.
The figure in question is Figure 1 from `Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars' (Fan et al. 2006) https://iopscience.iop.org/article/10.1086/504836 .
The Figure will contain a reference in the caption.
My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.

Yours faithfully,
Tom Binnie

From: **Saleem Zaroubi** saleem@astro.rug.nl
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 7 January 2021 at 15:20
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

SZ

Dear Tom,

Thanks for asking. Please feel free to use any figure you need from my work.

With kind regards,

Saleem

> On 7 Jan 2021, at 16:53, Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:
>
> Dear Prof. Zaroubi,
>
> I wonder if you'd be able to help me.
> I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
> I seek your permission to include figures from your work that significantly improve the explanations I provide in the introductory sections of my thesis.
>
> The figure in question is Figure 15 from `The Epoch of Reionization' (Zaroubi 2013). Each Figure will contain a reference to this in the caption.
>
> My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours faithfully,
> Tom Binnie

Prof. Saleem Zaroubi
1. Dept. of Natural Sciences, Open Univ., Israel. Tel (office): +972-9-7783376
2. Kapteyn Institute, Univ. of Groningen, Netherlands
Web Page: www.saleemzaroubi.info

RightsLink Printable License 07/01/2021, 15:20

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 07, 2021

This Agreement between Imperial College London -- Thomas Binnie ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4983660761666
License date	Jan 07, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	The Epoch of Reionization
Licensed Content Author	Saleem Zaroubi
Licensed Content Date	Jan 1, 2013
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of	1

figures/tables/illustrations	
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation
Institution name	Imperial College London
Expected presentation date	Jan 2021
Order reference number	https://link.springer.com/chapter/10.1007/978-3-642-32362-1_2
Portions	Figure 15
Requestor Location	Imperial College London St Marys Manor Road Goring-On-Thames Reading, S. Oxon RG8 9ED United Kingdom Attn: Imperial College London
Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

Thursday, January 7, 2021 at 10:59:19 Eastern Standard Time

Subject: Permission for figure reproduction in PhD thesis
Date: Thursday, January 7, 2021 at 10:49:11 AM Eastern Standard Time
From: Thomas Binnie
To: Adrian Liu

Dear Prof. Liu,

I wonder if you'd be able to help me.
I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
I seek your permission to include figures from your work that significantly improve the explanations I provide in the introductory sections of my thesis.

The figure in question is Figure 1 from `Epoch of reionization window. I. Mathematical formalism' (Liu et al. 2014).
The Figure will contain a reference to this in the caption.

My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.

If you are happy to grant me the permission requested, please return a signed copy of this letter.
If you wish to grant only part of the permission requested or you have specific terms you wish for me to abide by, please list these and then sign.

Yours sincerely,
Tom Binnie

American Physical Society Reuse and Permissions License

07-Jan-2021

This license agreement between the American Physical Society ("APS") and Thomas Binnie ("You") consists of your license details and the terms and conditions provided by the American Physical Society and SciPris.

Licensed Content Information

License Number:	**RNP/21/JAN/034928**
License date:	07-Jan-2021
DOI:	10.1103/PhysRevD.90.023018
Title:	Epoch of reionization window. I. Mathematical formalism
Author:	Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott
Publication:	Physical Review D
Publisher:	American Physical Society
Cost:	USD $ 0.00

Request Details

Does your reuse require significant modifications: No

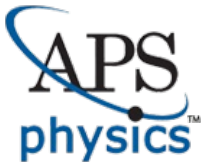
Specify intended distribution locations:	UK & Commonwealth (excluding Canada)
Reuse Category:	Reuse in a thesis/dissertation
Requestor Type:	Student
Items for Reuse:	Figures/Tables
Number of Figure/Tables:	1
Figure/Tables Details:	Figure 1
Format for Reuse:	Print and Electronic
Total number of print copies:	Up to 1000

Information about New Publication:

University/Publisher:	Imperial College London
Title of dissertation/thesis:	Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation
Author(s):	Thomas Binnie
Expected completion date:	Jan. 2021

License Requestor Information

Name:	Thomas Binnie
Affiliation:	Individual
Email Id:	t.binnie16@imperial.ac.uk
Country:	United Kingdom



American Physical Society Reuse and Permissions License

TERMS AND CONDITIONS

The American Physical Society (APS) is pleased to grant the Requestor of this license a non-exclusive, non-transferable permission, limited to Print and Electronic format, provided all criteria outlined below are followed.

1. You must also obtain permission from at least one of the lead authors for each separate work, if you haven't done so already. The author's name and affiliation can be found on the first page of the published Article.
2. For electronic format permissions, Requestor agrees to provide a hyperlink from the reprinted APS material using the source material's DOI on the web page where the work appears. The hyperlink should use the standard DOI resolution URL, http://dx.doi.org/{DOI}. The hyperlink may be embedded in the copyright credit line.
3. For print format permissions, Requestor agrees to print the required copyright credit line on the first page where the material appears: "Reprinted (abstract/excerpt/figure) with permission from [(FULL REFERENCE CITATION) as follows: Author's Names, APS Journal Title, Volume Number, Page Number and Year of Publication.] Copyright (YEAR) by the American Physical Society."
4. Permission granted in this license is for a one-time use and does not include permission for any future editions, updates, databases, formats or other matters. Permission must be sought for any additional use.
5. Use of the material does not and must not imply any endorsement by APS.
6. APS does not imply, purport or intend to grant permission to reuse materials to which it does not hold copyright. It is the requestor's sole responsibility to ensure the licensed material is original to APS and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover or table does not indicate it was reprinted by APS with permission from another source.
7. The permission granted herein is personal to the Requestor for the use specified and is not transferable or assignable without express written permission of APS. This license may not be amended except in writing by APS.
8. You may not alter, edit or modify the material in any manner.
9. You may translate the materials only when translation rights have been granted.
10. APS is not responsible for any errors or omissions due to translation.
11. You may not use the material for promotional, sales, advertising or marketing purposes.
12. The foregoing license shall not take effect unless and until APS or its agent, Aptara, receives payment in full in accordance with Aptara Billing and Payment Terms and Conditions, which are incorporated herein by reference.
13. Should the terms of this license be violated at any time, APS or Aptara may revoke the license with no refund to you and seek relief to the fullest extent of the laws of the USA. Official written notice will be made using the contact information provided with the permission request. Failure to receive such notice will not nullify revocation of the permission.
14. APS reserves all rights not specifically granted herein.
15. This document, including the Aptara Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

From: **Catherine Watkinson** catherine.watkinson@googlemail.com
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 7 January 2021 at 19:19
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

CW

Dear Tom,

I am very happy for you to reproduce this figure in your thesis. Good luck with getting it finished and in all your future research.

All the very best!

Catherine

Catherine Watkinson
Postdoctoral Research Associate
Queen Mary London - Astronomy Unit
https://catherinewatkinson.weebly.com
+44 (0) 791 776 8600

On Thu, 7 Jan 2021 at 18:25, Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:

> Dear Dr Watkinson,
>
> I hope you are well.
> I was wondering if you could give me permission to recreate a figure from one of your publications for use in my thesis.
> The figure in question is Figure 1 from `Distinguishing models of reionization using future radio observations of 21-cm 1-point statistics' (Watkinson & Pritchard 2014).
> It will help massively in explaining the different ionisation thresholds for within toy EoR models.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours sincerely,
> Tom Binnie

From: **Farhan Feroz** farhan.feroz@gmail.com
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 8 January 2021 at 10:05
To: Binnie, Thomas J t.binnie16@imperial.ac.uk
Cc: Mike Hobson mph@mrao.cam.ac.uk

FF

Hi Tom,

I have no objections to you using this figure in your thesis. Are you referring to a letter in a specific format that I need to sign?

Best regards,
Farhan

On Thu, 7 Jan 2021 at 18:18, Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:

> Dear Dr Feroz,
>
> I wonder if you'd be able to help me.
> I am currently completing my PhD under the supervision of Jonathan Pritchard at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
> I seek your permission to recreate a figure from your publication that significantly improves the explanations I provide in the introductory sections of my thesis.
>
> The figures in question is Figure 1 from `MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics' (Feroz et al. 2008). The Figure will contain a reference in the caption.
>
> My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours faithfully,
> Tom Binnie

From: **Pritchard, Jonathan R** j.pritchard@imperial.ac.uk
Subject: Re: Permission for figure reproduction in PhD thesis
Date: 11 January 2021 at 15:11
To: Binnie, Thomas J t.binnie16@imperial.ac.uk

JP

Dear Tom,

I'm very happy for you to make use of these figures for this purpose.

Best regards,
Jonathan

Dr Jonathan Pritchard
Reader in Astrostatistics
Imperial College London
Blackett Laboratory 1018C

j.pritchard@imperial.ac.uk
Office: +44 (0) 207 594 7557
Mobile: +44 (0) 751 841 8007
http://pritchardjr.github.io

> On 11 Jan 2021, at 13:44, Binnie, Thomas J <t.binnie16@imperial.ac.uk> wrote:
>
> Dear Dr Pritchard,
>
> I wonder if you'd be able to help me.
> I am currently completing my PhD under your supervision at Imperial College London entitled 'Towards Improved Statistical Analyses of the Cosmic Dawn and Epoch of Reionisation'.
> I seek your permission to include figures from your publication that significantly improve the explanations I provide in the introductory sections of my thesis.
>
> The figures in question are Figure 1 from `21-cm cosmology' (Pritchard & Loeb 2012); Figure 1 from `Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line' (Pritchard & Furlanetto 2006) and all the Figures from our own work `Bayesian model selection with future 21cm observations of the epoch of reionization' (Binnie & Pritchard 2019).
>
> Each Figure will contain a reference in the caption.
>
> My thesis will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public under a non-commercial Creative Commons license.
>
> If you are happy to grant me all the permissions requested, please return a signed copy of this letter.
> If you wish to grant only some of the permissions requested or you have specific terms you wish for me to abide by, please list these and then sign.
>
> Yours gratefully,
> Tom Binnie

F.2 Figures reproduced where no permission is needed

Figure 1.3 from (Pritchard & Furlanetto, 2006) is under copyright from Oxford University Press, who request that no more than two figures are used and the figure number be stated in supplement to the citation and permission from the author.

Figures 1.7(b), 1.9(d), A.4, A.5, and A.7 are from official websites. These websites contain copyright clauses that request the link and citation (if applicable) be provided, this is done in each figure caption.

Figures A.1, A.2, and A.3 from Loeb & Furlanetto (2013) are under Princeton University Press copyright, permission from the authors as in the previous appendix is all that is required.

Figure A.6 from (Fan et al., 2006) is under similar copyright from the Institute of Physics.

Figure E.9(b) is unpublished, permission has been gained via private communication with C. Trott.

Figures 3.2, 3.3, 3.4, 3.5(a), 3.5(b), 3.7(b), 3.8, 3.9(a), 3.9(e), 3.10, 3.11, 3.12, 3.13, 3.14, 3.15, 3.16, 3.17, 3.19, and 3.21 are from my own publication (Binnie & Pritchard, 2019). I grant myself permission; IOP and OUP request these figures are stated with the publication: these are redistributed from figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16.

All other figures are my own unpublished work.