UNIVERSITY *of the*
WESTERN CAPE

HI intensity mapping: Impact of primary beam effects

Author:
Siyambonga Donald Matshawule

Supervisor:
Prof Mario G. Santos
Dr Marta Spinelli
Prof Roy Maartens

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in Astrophysics

Centre of Radio Cosmology
Department of Physics and Astronomy

January 2023

Declaration of Authorship

I, Siyambonga Donald Matshawule, declare that this thesis titled, 'HI intensity mapping: Impact of primary beam effects' and the work presented in it is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

The work in this thesis has been done by the author under the supervision of Prof Mario G. Santos (University of the Western Cape), Dr Marta Spinelli (Institute for Particle Physics and Astrophysics, ETH Zurich/University of the Western Cape) and Prof Roy Maartens (University of the Western Cape).

Part of this thesis has been published in a peer-reviewed journal:

- Chapters 3 and 4. Siyambonga D. Matshawule, Marta Spinelli, Mario G. Santos, Sibonelo Ngobese. HI intensity mapping with MeerKAT: Primary beam effects on foreground cleaning. *MNRAS, 506(4): 5075–5092*, October 2021.
  doi: 10.1093/mnras/stab1688.

UNIVERSITY OF THE WESTERN CAPE

Abstract

Faculty of Natural Science
Department of Physics and Astronomy

Doctor of Philosophy in Astrophysics

HI intensity mapping: Impact of primary beam effects

by Siyambonga Donald Matshawule

Neutral hydrogen (HI) intensity mapping surveys with upcoming and future radio telescopes such as the MeerKAT, a precursor to the Square Kilometre Array Observatory (SKAO) MID telescope, have great potential for constraining cosmology, particularly in the post-reionization Universe provided that effective cleaning methods are available to separate the strong foregrounds from the cosmological signal. The application of cleaning methods is usually conducted under the assumption of simplistic primary beam models.

In this thesis, I simulate a single-dish wide-area survey with MeerKAT characteristics, and test foreground subtraction with a realistic model for the primary beam that contains a non-trivial frequency dependence. I also probe the impact of strong point sources on the cleaning. To conduct this evaluation, point source maps from a much more realistic full-sky point source catalogue are included as part of the foregrounds present in the sky model.

I find that cleaning can be complicated by the coupling between the foreground structure and the side-lobes of the primary beam, especially when the side-lobes are large. In order to reconstruct the HI signal, strong point sources must be removed and a more aggressive cleaning is required. However, in the case that the beam frequency dependence is smooth, reconstruction of the HI signal is conducted with no difficulty.

A non-trivial frequency dependence of the beam is also considered, in the form of a sinusoidal frequency variation in the width of the MeerKAT beam, expected to be present in all dishes as well as in SKAO-MID. Including this feature in the analysis couples the beam structure with the foreground emissions from both point sources and diffuse synchrotron and biases the reconstruction of the HI signal in frequency space. When not accounted for and removed, this biasing can introduce complexity in cosmological analysis, such as in reconstructing the imprint of the baryon acoustic oscillations on the 21 cm power spectrum. In the power spectrum this effect is contained in a narrow region

of $k_{\parallel}$ space. I find that by carefully re-smoothing the maps to a common lower resolution, the feature in the power spectrum is reduced.

A framework to compute the map-space convolution is constructed to allow for the use of even more accurate beam images, in order for beam asymmetries and non-trivial frequency effects to be included in the pipeline. Computation of the map-space convolution is conducted on the radio sky intensity maps constructed from standard HI IM foregrounds, including point sources from the new catalogue. I find that in the radial power spectrum a more complex feature is present, containing a secondary peak at slightly higher $k_{\parallel}$. Reducing it requires aggressive cleaning ($N_{\rm fg} = 6$) for the HI signal to be recovered, with over-cleaning at lower and intermediate $k_{\parallel}$. The angular power spectrum also reveals the dominance of this frequency effect and in this case more foreground components need to be removed ($N_{\rm fg} = 10$), resulting in the HI signal being recovered within 20 % of the expected signal at intermediate to smaller angular scales (higher ℓ). At higher angular scales there is significant power loss at the same level of cleaning. However, if we assume that point sources can be removed by applying a flux cut to consider only point sources with flux < 100 mJy, we find that this feature is greatly reduced, using a lower number of subtracted components. Furthermore, we observe great improvement in the reconstruction of the HI signal in the angular power spectrum: the signal is recovered within 5% of the expected HI signal.

Acknowledgements

Umbulelo omkhulu kuQamata, ngokundithwala, andigcine kude kube kwelithuba. Ndibulele kumanyange asekhaya, amaNqabe, oBhedla ka Mpulana, oMtshoni, oBhangis'abelungu ngentololwane, izihlahla ezithi zigawulwa zibesihluma. Ndinqule kumaManywabe, oGatyeni oNdondela, onkomo'zibomvu, oMamali.

I would like to express my sincere gratitude to my academic supervisors Prof Mario Santos, Dr Marta Spinelli and Prof Roy Maartens, for their guidance throughout this project. Their enthusiasm, patience, and willingness to lend an ear whenever I had questions or sought clarity made it possible for me to complete this work. "*Inene Umntu ngumntu ngabantu* (Indeed, a person is a person because of other people)". Thank you.

I would like to thank the Department of Physics and Astronomy for their support as I pursued these studies. "*Zenenjenje nakwabanye* (May you do the same for others.)"

Many thanks to the members of the Astronomy Group at UWC past and present, for the advice, the useful discussions, and support. *Enkosi kakhulu.*

A special thank you to my parents, Mkhangeli and Sindiswa Matshawule. Without your love and support, I would not have had the courage to embark on this journey. To my late father, I hope you are smiling and proud. *Enkosi Bhedla.*

To my life partner and best friend, Sisanda Matshawule, *enkosi ngento yonke. Yanga uMdali angasigcina sibumbene ngoluhlobo, ze ndikwazi ukwenza njengokuba wenzile. Enkosi Mamtshatshu, Tubhane, Mawose.*

To my little one, Luhle and my siblings Fezeka and Banele Matshawule, *uMzingisi Akanashwa!* May this be a source of inspiration and motivation.

My friends, Elethu Mvusi and Prof Usisipho Feleni, thank you for the support. Your words of motivation kept me going.

I acknowledge the support from the Centre of High Performance Computing (CHPC), under the project ASTR0945.

Lastly, I would like to acknowledge the Department of Higher Education and Training (DHET) through the New Generation of Academics Programme (nGAP), the National Research Foundation (NRF), and University of the Western Cape for their support.

Contents

Chapter 1

Introduction

Cosmology aims to describe and explain the origins, content, structure, and evolution of the Universe. A simple yet extremely successful model of cosmology, known as the ΛCDM 'concordance' model, has been constructed. With just a few cosmological parameters, this standard model of the Universe has been able to make predictions and provide explanations for all cosmological observations to date. However, certain tensions are being uncovered, in particular between the locally-measured value of the Hubble parameter H_0 and the value inferred from the cosmic microwave background (CMB) [Di Valentino et al., 2021]. Such tensions, and other potential problems in the ΛCDM model, are to be expected in any theoretical model of the Universe. Like any model in physics, ΛCDM needs to be placed under continual questioning, testing, and validating. For the purposes of this thesis, we adopt the standard concordance model.

In this chapter we briefly set out the foundations of the ΛCDM concordance model, explaining only its main features which are relevant for the thesis. We follow broadly the description given in Maartens [2018], together with standard texts, including Liddle [1999], Dodelson [2003] and Ryden [2006].

1.1 Friedmann-Robertson-Walker Metric

In constructing a model to describe the Universe, we need to make some assumption about the geometry of spacetime. This geometry corresponds to the averaged, or smooth, spacetime that approximates the real Universe on large enough scales. We must take into

account that the Universe is expanding, so that it changes in time. The simplest geometry that is compatible with a Universe that changes in time, is given by the ***Cosmological Principle***:

> On average, the Universe is spatially homogeneous and isotropic for observers comoving with the galaxies.

Spatial homogeneity essentially means that for any observer comoving with the galaxies, at any time t_o, physical quantities such as temperature and density are constant, i.e. they are the same at all points in the 3-dimensional space defined by $t_o = \text{const}$. Spatial isotropy means that at any point in $t_o = \text{const}$, all directions are equivalent for a comoving observer.

There are three types of 3-dimensional spaces that are homogeneous: 3-flat (Euclidean 3-space), 3-sphere (positively curved and compact) and 3-hyperboloid (negatively curved and non-compact). These are shown in figure 1.1. Measurements of the temperature anisotropies in the CMB, together with data from supernovae of type Ia and the baryon acoustic oscillation (BAO) scale in the galaxy distribution, show that the amount of curvature is very small [Planck Collaboration, 2020]. The ΛCDM concordance model therefore has zero spatial curvature, i.e. it has flat 3-spaces of constant cosmological time.

Then the spacetime metric in this case is the (spatially flat) Friedmann-Robertson-Walker (FRW) metric:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\Big[\mathrm{d}r^2 + r^2\big(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\big)\Big], \tag{1.1}$$

where we use units with the speed of light $c = 1$, t is the cosmological time (i.e. proper time for comoving observers), $a(t)$ is the cosmological scale factor (dimensionless) and (r, θ, ϕ) are comoving spherical coordinates. The comoving observers have $(r, \theta, \phi) = \text{const}$ (hence the name 'comoving coordinates'). The expansion is given purely by $a(t)$, so that the physical separation between two comoving observers evolves as

$$\Delta r_{\text{phys}}(t) = a(t)\,\Delta r\,, \tag{1.2}$$

where $\Delta r = r_2 - r_1$ and we have chosen coordinates so that $(\theta_1, \phi_1) = (\theta_2, \phi_2)$ – which is possible by isotropy. Equation 1.2 implies that the physical velocity of separation between

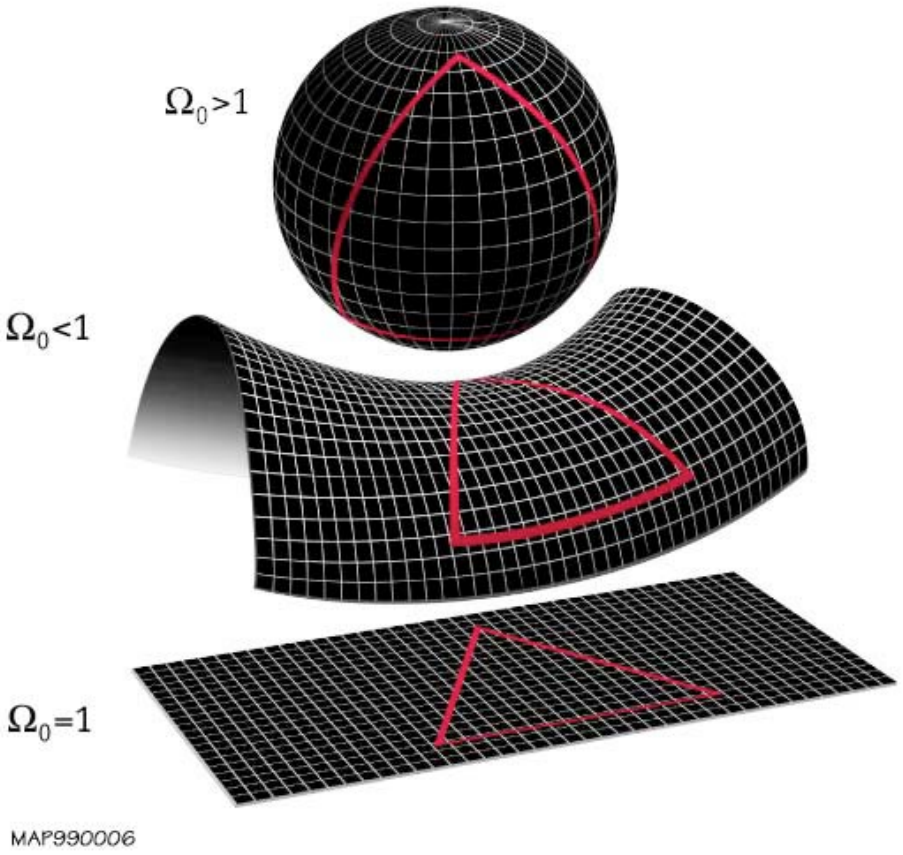

FIGURE 1.1: Homogeneous 3-spaces: 3-sphere ($\Omega_0 > 1$), 3-hyperboloid ($\Omega_0 < 1$); 3-flat ($\Omega_0 = 1$). [From https://wmap.gsfc.nasa.gov/universe/]

the comoving observers is

$$v_{\rm phys} = \frac{\rm d}{{\rm d}t}\Delta r_{\rm phys} = \dot{a}\,\Delta r = H\,\Delta r_{\rm phys}\,, \tag{1.3}$$

where an over-dot indicates a time derivative and we defined the Hubble parameter:

$$H = \frac{\dot{a}}{a}\,. \tag{1.4}$$

An expanding universe has $H > 0$.

1.1.1 Redshift

A particular case of equation 1.2 is the stretching of electromagnetic wavelengths by the expansion of the Universe. Consider a comoving observer at $r = 0$ and a comoving emitter at $r = r_e$. The emitted radiation has wavelength λ_e and the observed wavelength is λ_o.

Taking the ratio of these two physical lengths and using equation 1.2, we find that

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)}\,. \tag{1.5}$$

In an expanding universe, $a(t_o) > a(t_e)$ since $t_o > t_e$. Thus the ratio in equation 1.5 is > 1, and we define the (positive) redshift z by $1 + z = \lambda_o/\lambda_e$. Therefore the cosmological redshift is given by

$$1 + z = \frac{a(t_o)}{a(t_e)}\,. \tag{1.6}$$

It follows that the cosmological redshift increases as the time of emission decreases. This means that higher z indicates earlier emission: the older the emitter, the higher the redshift. Electromagnetic rays travel along the light-cone, so that higher z also indicates greater comoving radial distance between the observer and the emitter. This is easiest to visualise if we use the conformal time coordinate η, defined by

$$a\mathrm{d}\eta = \mathrm{d}t\,. \tag{1.7}$$

Then the metric equation 1.1 becomes

$$\mathrm{d}s^2 = a^2(\eta)\Big[-\mathrm{d}\eta^2 + \mathrm{d}r^2 + r^2\big(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\big)\Big], \tag{1.8}$$

and radial ($\mathrm{d}\theta = 0 = \mathrm{d}\phi$) lightrays ($\mathrm{d}s^2 = 0$) are given by $\eta = \pm r$, as in flat (Minkowski) spacetime (see figure 1.2).

Consider an observer at $r_o = 0, t = t_o, z = 0$, and an emitter at a general comoving coordinate $r > 0$ and time $t < t_o$. The redshift of the emitter satisfies

$$z = \frac{1}{a} - 1\,, \qquad \frac{\mathrm{d}z}{\mathrm{d}t} = -(1+z)\,H\,, \tag{1.9}$$

by equation 1.4 and equation 1.6. Here we have chosen (without loss of generality)

$$a(t_o) = 1\,. \tag{1.10}$$

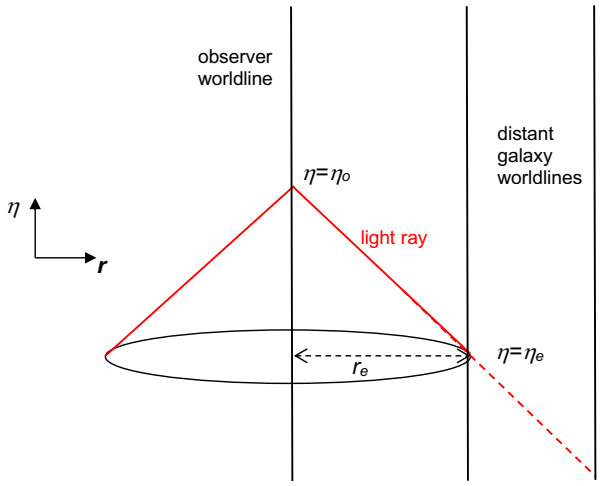

FIGURE 1.2: Light rays reaching the observer from distant sources, using comoving distance and conformal time. (From Maartens [2018].)

1.1.2 Distances

A fundamental distance measure is given by the (comoving) radial line-of-sight distance from the observer ($r_o = 0$) to an emitter at time t:

$$r = \int_t^{t_o} \frac{\mathrm{d}\tilde{t}}{a(\tilde{t}\,)}\,. \tag{1.11}$$

Using equation 1.9, we can rewrite this in terms of redshift:

$$r = \int_0^z \frac{\mathrm{d}\tilde{z}}{H(\tilde{z}\,)}\,. \tag{1.12}$$

The physical distance corresponding to r is the angular diameter distance:

$$d_A = a\,r = \frac{1}{1+z}\int_0^z \frac{\mathrm{d}\tilde{z}}{H(\tilde{z}\,)}\,. \tag{1.13}$$

The observational definition of d_A is the following: if a distant object has physical diameter ΔR in its rest frame, and if it subtends an angle $\Delta\varphi$ at the observer, then the angular diameter distance is

$$d_A = \frac{\Delta R}{\Delta\varphi}\,, \tag{1.14}$$

provided that $\Delta\varphi \ll 1$ (which is the case for cosmological distances). In the FRW spacetime, equation 1.1, we can show that equation 1.14 implies equation 1.13.

The observational definition equation 1.14 leads to the concept of 'standard rulers': if we know the intrinsic physical size of an object, ΔR, then by measuring its angular size $\Delta\varphi$, we can find its distance d_A. Since d_A is related to the Hubble rate by equation 1.13, the standard ruler can give cosmological information. This is the basis for using the BAO feature in the galaxy distribution in order to constrain cosmology.

Another observational distance is the luminosity distance, d_L, defined by

$$d_L^2 = \frac{L_e}{4\pi\, F_o}\,, \tag{1.15}$$

where L_e is the luminosity of the emitter and F_o is the flux measured by the observer. This leads to the concept of 'standard candles': if the intrinsic luminosity of an object is known, then the measured flux gives us the luminosity distance by equation 1.15. This is the basis for using supernova of type Ia in order to constrain cosmology.

In an FRW spacetime, equation 1.15 leads to the relation:

$$d_L = \frac{1}{a}\, r = (1+z) \int_0^z \frac{\mathrm{d}\tilde{z}}{H(\tilde{z}\,)}\,. \tag{1.16}$$

Comparing with equation 1.13 we have

$$d_L = (1+z)^2\, d_A\,. \tag{1.17}$$

This relation is in fact valid in any spacetime. The angular and luminosity distances in the concordance model are shown in figure 1.3. Note that d_A reaches a maximum at $z \sim 1.5$

1.2 The Friedmann Equations

The Cosmological Principle and the assumption of spatial flatness lead to the FRW metric equation 1.1. This is used in Einstein's general relativity field equation:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = 8\pi G\, T_{\mu\nu} - \Lambda\, g_{\mu\nu}\,. \tag{1.18}$$

Here $G_{\mu\nu}$ is the Einstein tensor, made from the Ricci tensor, $R_{\mu\nu}$, the metric tensor, $g_{\mu\nu}$ and the Ricci scalar, $R = R_{\mu\nu}\, g^{\mu\nu}$. On the right-hand side of equation 1.18, G is

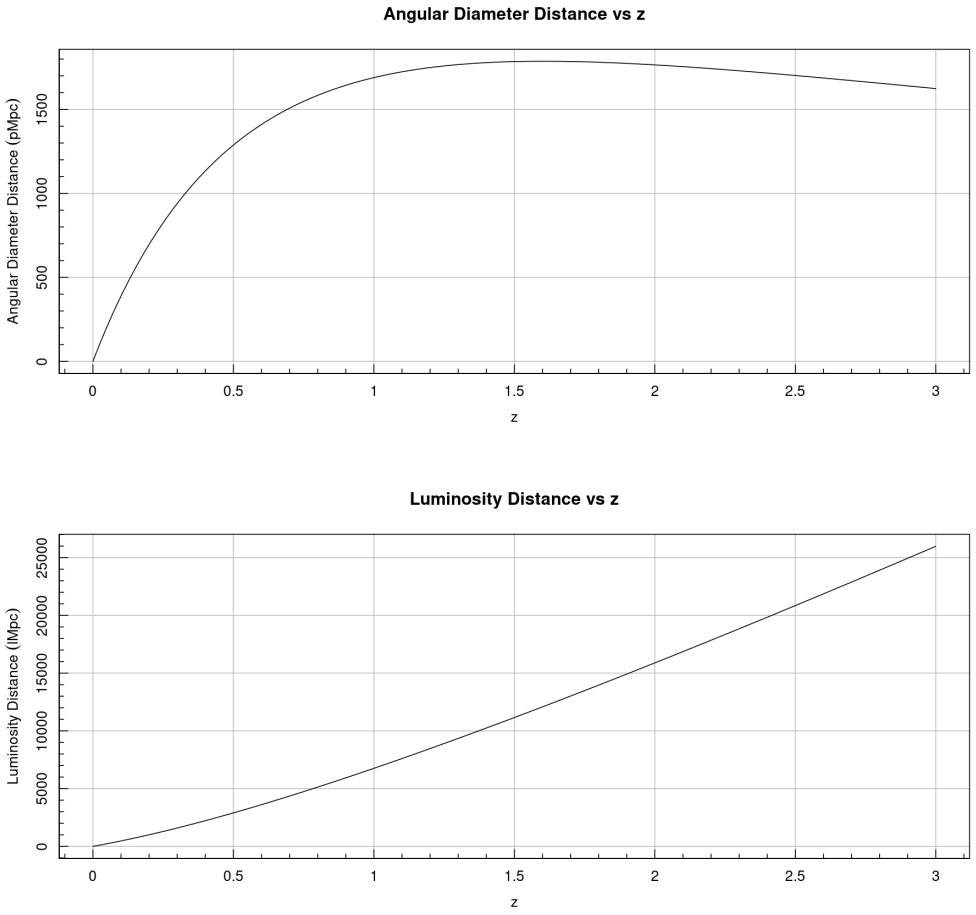

FIGURE 1.3: Angular diameter (top) and luminosity (bottom) distances in the concordance model. (From https://cosmocalc.icrar.org/)

Newton's gravitational constant, $T_{\mu\nu}$ is the energy-momentum tensor of the cosmological fluids (baryons, cold dark matter, photons, neutrinos) and Λ is the cosmological constant, corresponding to 'dark energy'. (We neglect neutrinos in this thesis, since they have negligible effects for our purposes.)

The metric tensor can be deduced from equation 1.1. We change to Cartesian spatial coordinates for simplicity. Then we have the diagonal 4×4 tensor matrix

$$g_{\mu\nu} = \text{diag}\left[-1,\, a^2(t),\, a^2(t),\, a^2(t)\right]. \tag{1.19}$$

This metric is used to find the Einstein tensor. The energy-momentum tensor is

$$T_{\mu\nu} = \text{diag}\left[\rho(t),\, p(t),\, p(t),\, p(t)\right], \tag{1.20}$$

where ρ is the total energy density and p is the total pressure of the fluids. Then the Einstein equations 1.18 are found to reduce to two equations:

$$H^2 \equiv \frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3}\,, \tag{1.21}$$

$$\dot{H} + H^2 \equiv \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p\right) + \frac{\Lambda}{3}\,. \tag{1.22}$$

Equation 1.21 is the Friedmann equation and equation 1.22 is the Raychaudhuri, or acceleration, equation. It is also common to refer to the two equations as 'Friedmann equations'. These equations imply the energy conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0\,. \tag{1.23}$$

Matter has $p_m = 0$ in the late-time Universe, while black-body radiation has $p_r = \rho_r/3$. Although it has positive energy density, Λ has *negative* pressure:

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} = -p_\Lambda\,. \tag{1.24}$$

The first equality follows from equation 1.21 and the second from equation 1.23. By equation 1.23, we have

$$\rho_r = \rho_{r0}\, a^{-4} = \rho_{r0}\,(1+z)^4\,, \tag{1.25}$$

$$\rho_m = \rho_{m0}\, a^{-3} = \rho_{m0}\,(1+z)^3\,, \tag{1.26}$$

$$\rho_\Lambda = \rho_{\Lambda 0}\,. \tag{1.27}$$

In general, the Einstein equations 1.18 imply the energy-momentum conservation equations

$$\nabla^\nu T_{\mu\nu} = 0\,. \tag{1.28}$$

In FRW spacetime, the momentum conservation equations ($\mu = i = 1, 2, 3$) are automatically satisfied, while energy conservation equation ($\mu = 0$) gives equation 1.23.

Equation 1.22 shows that:

- matter (CDM and baryons) and radiation always contribute to deceleration of the universe expansion (i.e., $\ddot{a} < 0$), since $\rho + 3p > 0$;

- the cosmological constant always contributes to acceleration of the universe expansion (i.e., $\ddot{a} > 0$), since $\Lambda > 0$.

Λ is the simplest form of dark energy, i.e. a field that acts to accelerate the expansion of the Universe. All models of dark energy must obey

$$p_{de} < -\frac{1}{3}\rho_{de}\,, \tag{1.29}$$

by (1.22).

Figure 1.4 shows schematically the evolution in the concordance model of the normalised densities $\rho_A(a)/\rho_{A0}$ against scale factor a, where A denotes radiation, matter and Λ. The radiation, matter and dark energy eras are the periods in which radiation, matter and dark energy respectively make the dominant contribution to density. This shows how dark energy is negligible at early times, when radiation dominates. Radiation density reduces faster ($\propto a^{-4}$) than matter density ($\propto a^{-3}$) so that matter begins to dominate after about 60,000 years. In the late matter era, radiation may be neglected, but eventually dark energy becomes important at $z < 1$. Although dark energy is very weak compared to matter for most of the history of the Universe, it does not decrease with expansion, so that eventually it dominates.

We define dimensionless density parameters for the matter, radiation and dark energy:

$$\Omega_A = \frac{8\pi G\,\rho_A}{3H^2} \quad \text{where} \quad A = r, m, \Lambda\,. \tag{1.30}$$

Then the Friedmann equation 1.21 implies that

$$1 = \Omega_r(z) + \Omega_m(z) + \Omega_\Lambda(z)\,, \tag{1.31}$$

$$\frac{H^2(z)}{H_0^2} = \Omega_{r0}(1+z)^4 + \Omega_{m0}(1+z)^3 + \Omega_{\Lambda 0}\,. \tag{1.32}$$

In the radiation era, and the earlier part of the matter era, we can safely neglect dark energy. The radiation era ends at the redshift of matter-radiation equality, which is given by $\rho_r(z_{\rm eq}) = \rho_m(z_{\rm eq})$. By equation 1.25 and equation 1.26, this gives

$$z_{\rm eq} = \frac{\rho_{m0}}{\rho_{r0}} - 1 = \frac{\Omega_{m0}}{\Omega_{r0}} - 1\,. \tag{1.33}$$

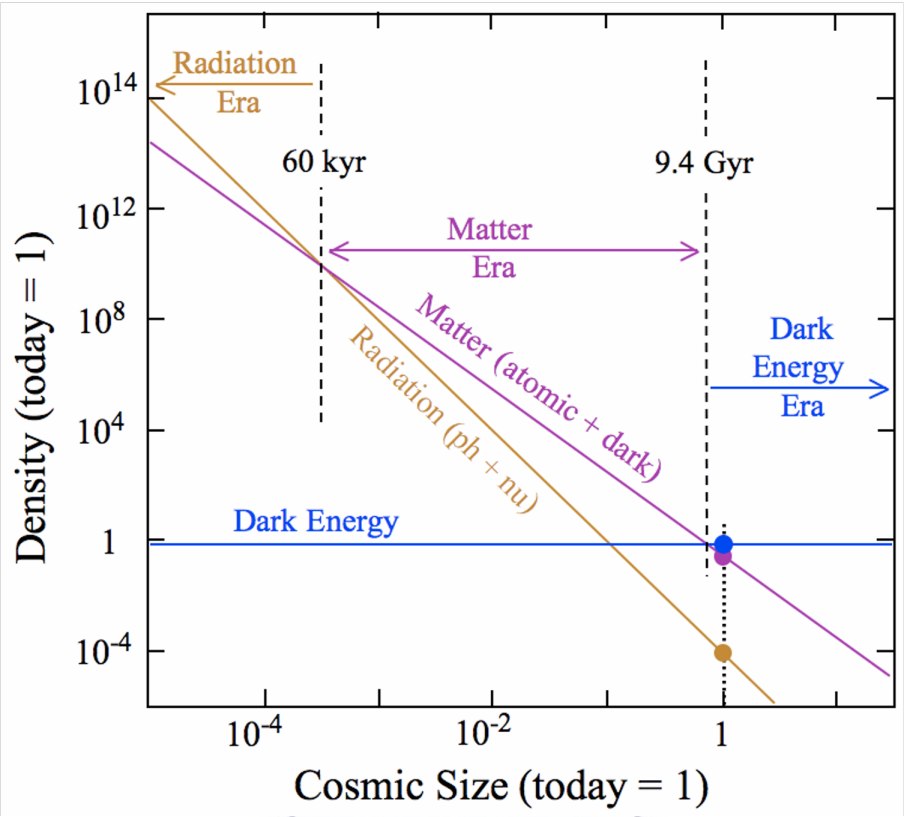

FIGURE 1.4: Evolution of the normalised densities ρ_A/ρ_{A0} against scale factor a in the concordance model. (From Maartens [2018].)

Current observations (see e.g. Planck Collaboration [2020]) give

$$\Omega_{m0} \approx 0.31\,, \quad \Omega_{\Lambda 0} \approx 0.69\,, \quad \Omega_{r0} < 10^{-4}\,, \quad H_0 \approx 67.4 \text{ km/s/Mpc}\,, \tag{1.34}$$

and $z_{\rm eq} \approx 3500$. In this thesis we focus on the late-time Universe, $z \lesssim 3$. For this range of redshift, the radiation contribution to the Friedmann equation is negligible and therefore we use

$$1 = \Omega_m(z) + \Omega_\Lambda(z)\,, \tag{1.35}$$

$$\frac{H^2(z)}{H_0^2} = \Omega_{m0}(1+z)^3 + 1 - \Omega_{m0}\,. \tag{1.36}$$

Note that

$$\Omega_m(z) = \Omega_{m0}(1+z)^3 \frac{H_0^2}{H^2(z)}\,, \tag{1.37}$$

$$\Omega_\Lambda(z) = \left(1-\Omega_{m0}\right)\frac{H^2(z)}{H_0^2}\,. \tag{1.38}$$

1.3 Inflation and the Cosmic Microwave Background

The classical model of the universe originates at $t = 0$ from a singularity, with infinite density, temperature and spacetime curvature. This is known as the 'Big Bang' singularity. Since infinities are unphysical, we need a quantum gravity theory to avoid this singularity, but so far there is no generally accepted quantum cosmology that achieves this. Therefore, for practical purposes, the concordance model can only describe the universe after the quantum gravity era, $t > t_{\rm qg}$, where $t_{\rm qg}$ is thought to be a very small time.

If we assume that the universe is in the radiation era from $t_{\rm qg}$ until matter-radiation equality, then we have difficulty in explaining certain key features of the Universe. In particular, it is difficult to explain how widely separated patches of the sky have the same temperature, apart from very small (1 part in 10^5) fluctuations. The problem is that widely separated patches cannot have been in causal contact since the Big Bang. So the question is: how do they come to have the same temperature?

In the standard model, this problem can be avoided if there is a very brief period of accelerated expansion at or near the beginning of the classical era, which is known as Inflation. Inflation blows up the small spherical region that eventually becomes the observable region of the Universe for any comoving observer. In other words, Inflation ensures that all of the CMB sky that we observe was causally connected in the very early Universe, thus explaining why it has the same temperature (see figure 1.5).

Inflation not only solves the problem of explaining the CMB sky – it also solves the problem of the origin of the tiny fluctuations in the CMB. The inflaton field that drives Inflation also generates the small fluctuations in temperature and density that grow under gravity to produce the temperature fluctuations in the CMB – and later the first stars and galaxies.

When Inflation ends, the Universe reheats to an extremely high temperature and the radiation era begins, when the Universe consists of photons and fundamental particles in an ultra-hot plasma. As the Universe expands, the temperature decreases:

$$T = \frac{T_0}{a} = T_0\,(1+z)\,, \tag{1.39}$$

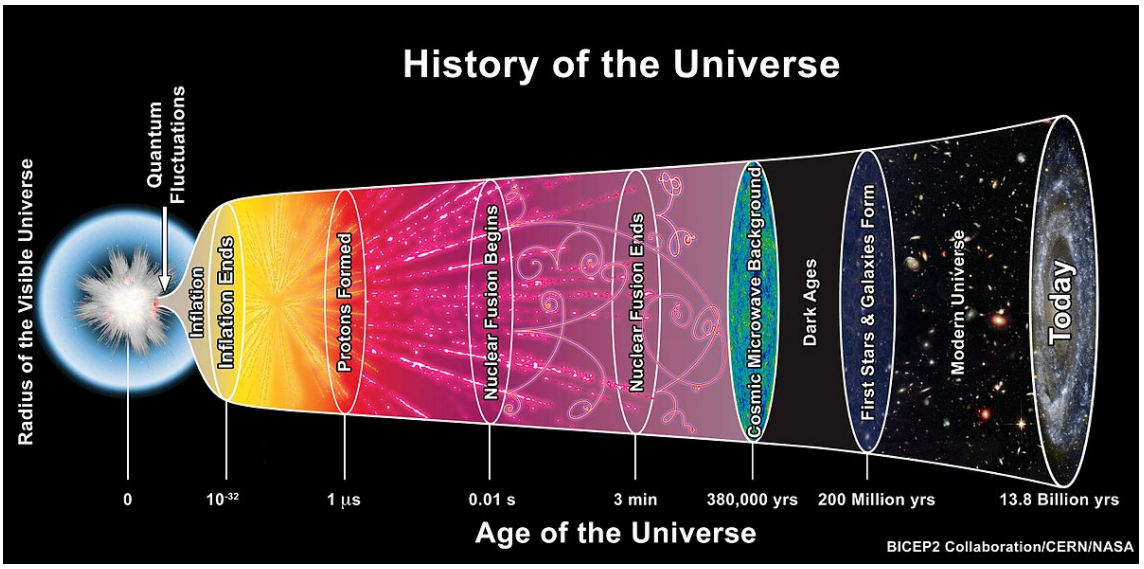

FIGURE 1.5: Schematic of Inflation and the history of the Universe. [From BICEP2 Collaboration/CERN/NASA]

where $T_0 = 2.725\,\mathrm{K}$ is the radiation temperature today. As the temperature drops, fundamental particles can combine to form composite particles like protons and then light nuclei like helium and lithium. This take about 3 minutes. It takes much longer before the first stable atoms (hydrogen) can form without being immediately ionised by super-hot radiation. This happens at

$$T_{\mathrm{dec}} \approx 3000\ \mathrm{K} \quad \text{and} \quad z_{\mathrm{dec}} \approx 1100\,, \tag{1.40}$$

where the 'decoupling' redshift follows from the temperature by equation 1.39. The CMB sky that we observe today corresponds to the time $t_{\mathrm{dec}} \approx 380,000$ years. This is when baryons decoupled from radiation (also known as 'last scattering'), allowing the formation of hydrogen atoms (known as 'recombination').

After decoupling, and still in the early matter era, the Universe enters the period of the 'Dark Ages' – when baryonic matter is mainly in the form of neutral hydrogen (with some helium also). The first generation of stars eventually emerge from the gravitational collapse of hydrogen clouds, lighting up the Universe in the 'Cosmic Dawn' and the subsequent Epoch of Reionization. The first generation of galaxies form and large-scale structure formation is underway.

1.4 Large-Scale Structure

The smooth FRW metric, equation 1.1, describes the Universe on average. The real Universe is not smooth, except on very large scales. In particular, the cosmological radiation, in the form of the CMB, is not smooth, but contains small anisotropies in temperature and polarisation. The distribution of matter in the Universe is even less smooth. The CDM is clustered into halos via gravitational collapse that started in the early Universe. The baryons are drawn into halos by the dominant gravitational attraction of CDM. Large enough halos host galaxies and some baryons are in between galaxies in the intergalactic medium. This process is modelled in N-body and hydrodynamic simulations (see figure 1.6).

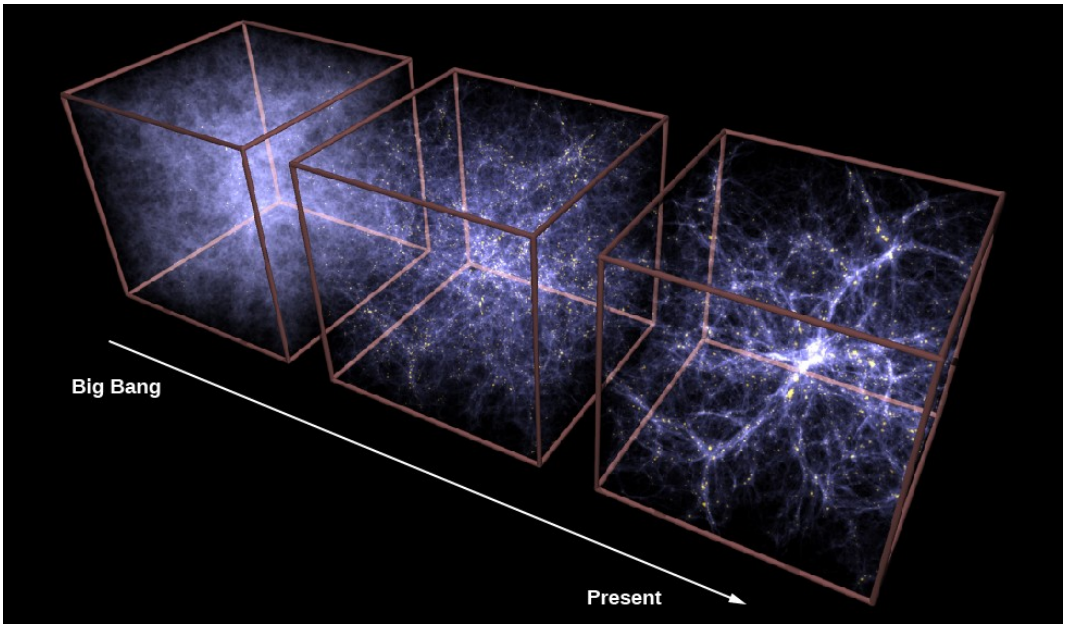

FIGURE 1.6: Schematic of large-scale structure formation in an N-body simulation. [From https://openstax.org/details/books/astronomy]

In order to understand the basics of large-scale structure, we use (linear) perturbation theory. The perturbed metric is

$$\mathrm{d}s^2 = -\big[1 + 2\Phi(t, \boldsymbol{x})\big]\mathrm{d}t^2 + a^2(t)\big[1 - 2\Phi(t, \boldsymbol{x})\big]\,\mathrm{d}\boldsymbol{x}^2\,, \tag{1.41}$$

where Φ is the gravitational potential (with $|\Phi| \ll 1$). The perturbed metric corresponds to perturbed density – since any changes in the matter or the metric are tightly related

to each other in Einstein's general relativity:

$$\rho_m(t,\boldsymbol{x}) = \bar{\rho}_m(t)\big[1+\delta_m(t,\boldsymbol{x})\big] \quad \text{where} \quad \delta_m = \frac{\rho_m - \bar{\rho}_m}{\bar{\rho}_m}\,. \tag{1.42}$$

Here $\bar{\rho}_m$ is the smooth average matter density which obeys the Friedmann equations. The matter contains over-densities (where $\rho_m > \bar{\rho}_m$) and under-densities (where $\rho_m < \bar{\rho}_m$). δ_m is known as the matter density contrast.

Perturbations of the density induce velocity perturbations as matter moves towards neighbouring over-densities and away from neighbouring under-densities. This defines a so-called peculiar velocity of matter $\boldsymbol{v}$, which is the small change to the smooth Hubble-flow velocity. The perturbed version of the radial velocity equation 1.3 is

$$\bar{v}_{\rm phys} + \delta v_{\rm phys} = H\, r_{\rm phys} + \boldsymbol{v}\cdot\hat{\boldsymbol{r}}\,, \tag{1.43}$$

where $\boldsymbol{v}\cdot\hat{\boldsymbol{r}}$ is the radial component of peculiar velocity.

The three perturbed quantities satisfy the following perturbed Einstein equations:

$$\nabla^2\Phi = \frac{3}{2}\,a^2\,\Omega_m H^2\,\delta_m\,, \tag{1.44}$$

$$\dot{\delta}_m = -\boldsymbol{\nabla}\cdot\boldsymbol{v}\,, \tag{1.45}$$

$$\dot{\boldsymbol{v}} + H\,\boldsymbol{v} = -\frac{1}{a}\,\boldsymbol{\nabla}\Phi\,. \tag{1.46}$$

Equation 1.44 is the Poisson equation, equation 1.45 is the perturbed energy conservation equation and equation 1.46 is the perturbed momentum equation. Together, they lead to an evolution equation for the matter density contrast:

$$\ddot{\delta}_m + 2H\,\dot{\delta}_m - \frac{3}{2}\,\Omega_m H^2\,\delta_m = 0\,. \tag{1.47}$$

This evolution equations shows how gravity causes the growth of the density contrast via the source term $\frac{3}{2}\,\Omega_m H^2\,\delta_m$, while the expansion of the universe slows down this growth, via the 'friction' term $2H\dot{\delta}_m$.

The density contrast may be written in terms of a growth factor $D(z)$ as

$$\delta_m(z,\boldsymbol{x}) = D(z)\,\delta_m(0,\boldsymbol{x})\,, \tag{1.48}$$

where we choose to normalise D such that $D(0) = 1$. Then (1.47) implies

$$\ddot{D} + 2H\,\dot{D} - \frac{3}{2}\,\Omega_m H^2\,D = 0\,. \tag{1.49}$$

The time derivative of equation 1.48 implies

$$\dot{\delta}_m = f\,H\,\delta_m \quad \text{where} \quad f = \frac{1}{H}\,\frac{\dot{D}}{D} = -\frac{\mathrm{d}\ln D}{\mathrm{d}\ln(1+z)}\,. \tag{1.50}$$

f is known as the growth rate – a dimensionless measure of how rapidly the density contrast is growing, relative to the expansion of the Universe.

We can see the effect of dark energy on the density contrast by comparing to a model with no dark energy, i.e. a matter-only model, with $\Omega_m(z) = 1$. In this case, the growing-mode solution of equation 1.49 is $D \propto t^{2/3} \propto a$, and then equation 1.50 gives $f = 1$:

$$D = a\,, \quad f = 1 \qquad \text{matter-only model.} \tag{1.51}$$

In the case of a model with dark energy, $\Omega_m < 1$, it is clear that $D < a$ and $f < 1$, since dark energy acts against the clustering of matter. A very good approximation to the growth rate in the standard model is given by

$$f(z) = \left[\Omega_m(z)\right]^{0.55}\,. \tag{1.52}$$

1.5 Power Spectrum

A key statistical quantity derived from the density contrast is the 2-point correlation function ξ_m:

$$\xi_m(z, r) = \left\langle \delta_m(z, \boldsymbol{x})\,\delta_m(z, \boldsymbol{x} + \boldsymbol{r}) \right\rangle, \tag{1.53}$$

where the angled brackets indicate a spatial average. ξ_m quantifies the deviation of the matter distribution from a random distribution – so that it is a measure of gravitational clustering.

In Fourier space, the correlation function defines the matter power spectrum P_m:

$$\left\langle \delta_m(z, \boldsymbol{k})\,\delta_m(z, \boldsymbol{k}') \right\rangle = (2\pi)^3\,\delta_{\mathrm{D}}(\boldsymbol{k} + \boldsymbol{k}')\,P_m(z, k)\,, \tag{1.54}$$

where $\delta_{\rm D}$ is the Dirac delta function. Then we have the relation between P_m and ξ_m:

$$\xi_m(z,r) = \frac{1}{2\pi^2}\int_0^\infty \mathrm{d}k\, k^2\, \frac{\sin(kr)}{kr}\, P_m(z,k)\,. \tag{1.55}$$

Note that P_m has units (length)^3. The power spectrum for the standard model is shown in figure 1.7.

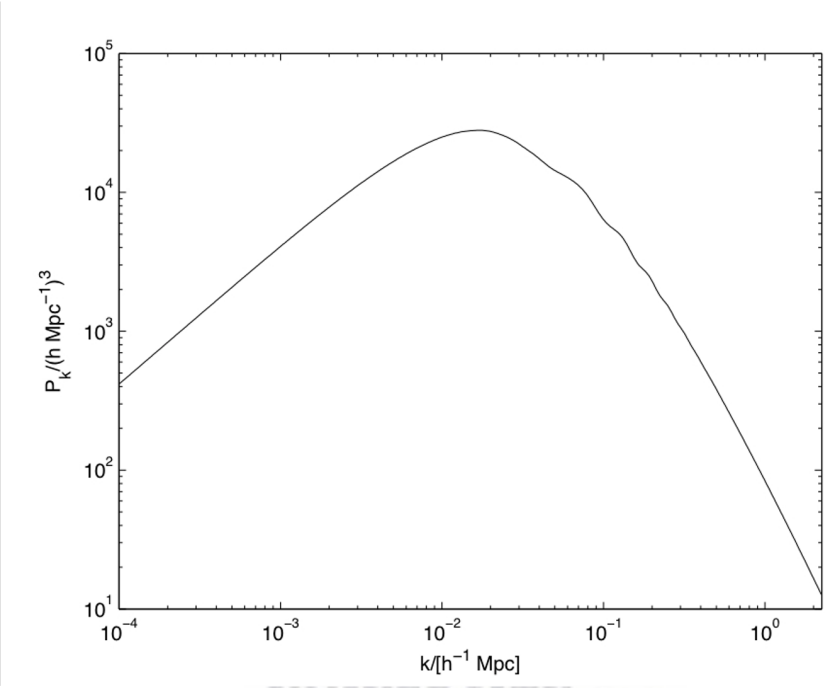

FIGURE 1.7: Linear matter power spectrum in LCDM at $z = 0$.

The key features of P_m are:

- P_m has a maximum (turn-over) at the scale

$$k_{\rm eq} = H(z_{\rm eq})\,, \tag{1.56}$$

  corresponding to the Hubble scale at matter-radiation equality;

- Fourier modes $k < k_{\rm eq}$, to the left of the turn-over, have comoving wavelengths that are greater than the Hubble scale at equality;

- the power spectrum for $k < k_{\rm eq}$ scales approximately as $P_m \sim k$;

- Fourier modes $k > k_{\rm eq}$, to the right of the turn-over, have comoving wavelengths that are less than the Hubble scale at equality, so that they were affected by the radiation era, which suppressed growth relative to the super-Hubble modes;

- the power spectrum for $k > k_{\rm eq}$ scales roughly as $P_m \sim k^{-3}$.

The matter power spectrum is not directly observable. We can find an estimate of P_m via number counts of galaxies or the intensity map of emission lines such as the 21cm line of neutral hydrogen (HI):

$$\delta_g = \frac{n_g - \bar{n}_g}{\bar{n}_g}\,, \qquad \delta_{\rm HI} = \frac{T_{\rm HI} - \bar{T}_{\rm HI}}{\bar{T}_{\rm HI}}\,, \tag{1.57}$$

where n_g is galaxy number density and $T_{\rm HI}$ is the HI temperature. In order to relate these to δ_m, we need to introduce the bias between each tracer and the underlying matter distribution:

$$\delta_g(z, \boldsymbol{x}) = b_g(z)\,\delta_m(z, \boldsymbol{x})\,, \qquad \delta_{\rm HI}(z, \boldsymbol{x}) = b_{\rm HI}(z)\,\delta_m(z, \boldsymbol{x})\,. \tag{1.58}$$

The bias here is assumed to be independent of scale, which is a reasonable assumption on scales where linear perturbation theory is accurate. Then we have a simple relation between galaxy and HI power spectra and the matter power spectrum:

$$P_I(z, k) = b_I^2(z)\,P_m(z, k)\,, \qquad I = g,\,{\rm HI}\,. \tag{1.59}$$

Equation 1.59 models the linear power spectrum (of galaxies or intensity mapping) in real space. However, observations take place in redshift space, which introduces distortions in the clustering signal. The peculiar velocity of a galaxy means that its position in the observer's redshift space is distorted relative to its position in its own rest frame. This redshift-space distortion (RSD) is illustrated in figure 1.8 in the case of galaxies at the same rest frame distance from the centre of a weak over-density.

FIGURE 1.8: Redshift-space distortions in the linear regime. (From Maartens [2018].)

The intrinsically spherical distribution in real space appears as an elliptical distribution, squashed along the line of sight. Since the peculiar velocities transverse to the line of sight do not lead to a distortion, the major axis of the ellipse has the same length as the sphere diameter. This means that the volume of the ellipse is smaller than that of the sphere. The number of galaxies is conserved and therefore the number density in redshift-space is higher than in real space. In other words, the (linear) RSD effect increases the power spectrum. This is known as the Kaiser effect. By contrast, on non-linear scales RSD acts to stretch the spherical distribution along the line of sight (known as the 'finger of god' effect), which damps the power spectrum.

The linear RSD effect on the observed density/ temperature contrast is given by

$$\delta_I^s = \delta_I - \frac{(1+z)}{H}\frac{\partial}{\partial r}\boldsymbol{v}\cdot\boldsymbol{n} \qquad (I = g,\,\mathrm{HI})\,, \tag{1.60}$$

where the superscript s denotes redshift-space and $\boldsymbol{n}$ is the line of sight. Then $\boldsymbol{v}\cdot\boldsymbol{n} \equiv v_\parallel$ is the radial component of peculiar velocity.

In Fourier space, equation 1.60 becomes

$$\delta_I^s = \left(b_I + f\,\mu^2\right)\delta_m \quad \text{where} \quad \mu = \hat{\boldsymbol{k}}\cdot\boldsymbol{n} = \frac{k_\parallel}{k}\,, \tag{1.61}$$

and the redshift-space power spectrum is

$$P_I^s(z,k,\mu) = \left[b_I^2(z) + f(z)\,\mu^2\right]^2 P_m(z,k)\,. \tag{1.62}$$

Later in this thesis we explore the use of neutral hydrogen intensity maps as a probe for large scale structure and study how instruments such as the MeerKAT and SKA will attempt to obtain an estimate of the HI power spectrum described in equation 1.62. In the following sections a brief description of the two radio telescopes is provided.

1.6 Square Kilometre Array Observatory

The Square Kilometre Array Observatory[1] (SKAO) project is a global effort to construct the largest and most sensitive radio telescope in the entire world. On completion of the full SKAO project (i.e. phases 1 & 2) the telescope will have a combined collecting area of 1 km^2 (1 million square metres), it will be 10,000 times faster, and 50 times more sensitive than any of the existing radio telescopes. The telescope will consist of thousands of dishes and close to a million antenna dipoles that will be hosted mainly in two places, the South African Karoo and Australian Murchison desert, with some antennas spread across partner countries within the African continent. These locations were selected for their remoteness, low population densities, dry and arid environments, making them ideal to conduct radio astronomy. In South Africa, the construction of SKAO Phase 1 is expected to begin in 2022 at a radio astronomy site that is about 90 km near the small town of Carnarvon in the Northern Cape. The Astronomy Geographic Advantage Act[2] was passed by the South African government to protect the area and promote astronomy and other science-related activities. Mid- to high-frequency radio astronomy will be conducted by the SKAO-MID array in the South African Karoo, whilst in the Australian desert, low-frequency radio astronomy will be conducted using mainly dishes and dipoles as shown on the right of figure 1.9 forming the SKAO-LOW array. SKAO Phase 1 will consist of a further 133 dishes with an aperture of 15 meters in addition to its precursor telescope the MeerKAT, making a total of 197 dishes.

There are several science goals that the SKAO seeks to achieve which can be put under the following themes:

- Challenging Einstein's Theory of General Relativity,
- Cosmology, Galaxy Evolution, and Dark Energy,

[1] https://www.skatelescope.org/the-ska-project/
[2] https://www.gov.za/documents/astronomy-geographic-advantage-act

FIGURE 1.9: An artistic composite image of the SKAO telescope, comparing the different radio telescopes to be hosted at its two major sites. On the left are the mid-frequency dishes in the upper Karoo desert of South Africa and on the right are the frequency dipole antennas of the SKAO low-frequency array at the Murchison Radio Observatory in Western Australia (From:https://www.skatelescope.org/the-ska-project/).

- Probing the Cosmic Dawn,
- Cradle of Life: the search for intelligent extraterrestrial radio wave emission,
- Cosmic Magnetism,
- Transients,
- Solar and Heliospheric physics.

To demonstrate the capabilities of each host country to perform the type of science the SKAO wishes to embark on, precursor telescopes were constructed. In South Africa, this was the 64 dish MeerKAT radio telescope with an aperture size of 13.5 m, whilst in Australia, the Australian SKA Pathfinder Telescope (ASKAP) was built. ASKAP consists of 36 dishes, with an aperture size of 12 m.

Figure 1.9 is an artistic representation of the two types of radio telescopes that will be constructed for the SKAO. On the left are the dish telescopes that will be mainly used for mid-frequency radio astronomy in South Africa operating in both single dish and interferometric mode. On the right are the small dipole antennas which will mainly perform low frequency radio astronomy and will mainly work as an interferometer.

1.7 MeerKAT

The MeerKAT telescope is a South African radio telescope array and precursor telescope of the SKAO and is hosted by the South African Radio Astronomy Observatory[3](SARAO). The 64 dish MeerKAT radio telescope is a mid-frequency array and will be integrated into SKAO Phase 1 facility, upon completion with construction expected to begin in the year 2022. Of the total number of dishes, 48 are located within a 1 km diameter and the rest distributed outside the core such that a maximum baseline of 8 km is achieved. On the left of figure 1.10 is a satellite view of the MeerKAT foundations prior to installation showing how the dishes are distributed and to the south west of the MeerKAT core is where another pathfinder telescope called the Hydrogen Epoch of Reionization Array[4] (HERA) is located.

FIGURE 1.10: On the left is a Google Earth satellite view of the inner core of the MeerKAT foundations prior to installation. The type of distribution is dependent on the type of science the MeerKAT radio telescope will perform. On the right is a wide view of the MeerKAT dishes shown to illustrate the design of the dishes (courtesy: Asad et al. [2021]).

In terms of construction and design, 75 % of the components that make up the MeerKAT telescope are locally sourced in South Africa. Each antenna is an offset-Gregorian configuration, with no struts in the way to block the radio signal. The main reflector and sub-reflector are shown on the right of figure 1.10, with each telescope containing four receivers and various other electronics. The main reflector has an effective diameter of 13.5 m whilst the sub-reflector is about 3.8 m in diameter. The antenna distribution design is dependent on the science goals which the MeerKAT will seek to achieve [Jonas and MeerKAT Team, 2016].

[3] https://www.sarao.ac.za/
[4] https://reionization.org/

How each radio telescope works is that the cosmic radio signals bounce off the main reflector to the secondary sub-reflector, which focuses the signal onto the feedhorns of one of the four receivers (UHF, L & S) placed on a support structure called an indexer. Currently only the UHF (544–1088 MHz) and L (856–1712 MHz) band receivers are in operation with S-band currently being commissioned. Depending on the observation frequency selected before the observation, the indexer will rotate for the correct receiver to be brought into the focus position. The receiver captures the electromagnetic signal and converts it to a Radio Frequency (RF) voltage signal and applies some amplification. At this stage, the digitizer instruments convert the analog signal into digital for it to be transported via fibre optic cables for further processing.

The MeerKAT telescope will embark on several science projects and which are grouped into different categories such as the Large Survey Projects (LSPs) and Open Time Projects (OTPs). LSPs are surveys that require more than 1000 hrs of observation time with MeerKAT over 5 years. Some of these LSPs are listed below.

- **MIGHTEE**: MeerKAT International GigaHertz Tiered Extragalactic Exploration survey[5] [Jarvis et al., 2016].
- **THUNDERKAT**: The Hunt for Dynamic and Explosive Radio Transients with MeerKAT[6] [Fender et al., 2016].
- **LADUMA**: Looking at the Distant Universe with the MeerKAT Array[7] [Holwerda et al. [2012], Blyth et al. [2016]].
- **MHONGOOSE**: MeerKAT HI Observations of Nearby Galactic Objects[8] [de Blok et al., 2016].
- **TRAPUM**: Search and detection of Pulsars and Transients with the MeerKAT[9] [Abbate et al., 2022]
- **MALS**: The MeerKAT Absorption Line Survey [10] [Gupta et al., 2016]
- **Fornax**: A Fornax Cluster Survey with MeerKAT[11] [Serra et al., 2016]

[5] http://idia.ac.za/mightee/
[6] http://www.thunderkat.uct.ac.za/
[7] http://www.laduma.uct.ac.za/
[8] https://mhongoose.astron.nl/
[9] http://www.trapum.org/
[10] https://mals.iucaa.in/
[11] https://sites.google.com/inaf.it/meerkatfornaxsurvey/home?authuser=0

- **MeerTIME**: Radio Pulsar Timing[12] [Bailes et al., 2016]

About 30% of the MeerKAT array's observation time will be dedicated to OTPs. These are periodic calls to the Astronomy community to submit proposals that are selected on competitive bases in order to use available MeerKAT observation time. Several pilot surveys have already been done in preparation for the MeerKAT Large Area Synoptic Survey (MeerKLASS) using the OTPs. In this dissertation, we focus on illustrating the potential of MeerKLASS in conducting an HI intensity mapping survey with the MeerKAT radio telescope. An overview of the MeerKLASS project is provided in section 2.3.3.

1.8 Summary

In this chapter, we have gone through some of the fundamentals of our current understanding of the ΛCDM model, and highlighted the different epochs in the evolution of the universe and the probes we are using to study them, from the Cosmic Microwave Background to Large Scale Structure. We have then presented the SKA Observatory and its main scientific goals. We have also discussed MeerKAT, the SKAO-MID precursor, already built and taking data in the Karoo semi-desert. In the next chapter, we discuss 21cm intensity mapping and highlight why this technique is important for cosmology.

[12] http://www.meertime.org/

Chapter 2

HI Intensity Mapping

2.1 Neutral Hydrogen and the 21cm Emission Line

Of the total baryonic mass in the Universe, 75% is in the form of Hydrogen, making this element the most abundant one. Although hydrogen has different isotopes, we focus on its simplest form which consists of a single proton with positive charge and an electron with negative charge, making it electrically neutral. With this type of configuration, neutral hydrogen is also referred to as atomic hydrogen or commonly referred to as HI (pronounced as H-one).

The spin of the proton and electron of hydrogen can be in two states either in alignment (parallel) or out of alignment (anti-parallel). When the spins of both proton and electron are aligned, the energy state of the hydrogen atom is a bit higher, about $5.87\text{x}10^{-6}$ eV of energy than its low energy state. The electron in this excited state can undergo a spin flip transition which occurs when the spin change from parallel to anti-parallel. During this process, a photon with a frequency of 1420.4 MHz and wavelength of 21 cm is emitted and the atom returns to a low energy state. The left panel of Figure 2.1 illustrates how this hyperfine spin flip transition occurs. Although this type of transition is rare (the decay time is around 10^6 years), due to the abundance of HI in the Universe, the 21 cm signal can be detected by our radio telescopes and experiences minimal absorption as it travels to the Earth.

Observing at 1420 MHz radio telescopes can detect the presence of local HI. Since the Universe is expanding, observing at lower radio frequencies we can instead observe the

redshifted 21cm signal. The wavelengths of photons get stretched while expansion occurs (see right panel of figure 2.1).

If we write down the definition of redshift we have

$$z = \frac{(\nu_{\text{emitted}} - \nu_{\text{observed}})}{\nu_{\text{observed}}} \tag{2.1}$$

with $\nu_{\text{emitted}} = 1420$ MHz. Observing at a certain frequency ν_{observed}, we are thus looking at the presence of neutral hydrogen at a specific redshift, for example:

- $\nu_{\text{observed}} \sim 900$ MHz corresponds to $z \sim 0.6$,
- $\nu_{\text{observed}} \sim 170$ MHz corresponds to $z \sim 7$,
- $\nu_{\text{observed}} \sim 70$ MHz corresponds to $z \sim 20$.

In the next section we briefly describe to which phases of the evolution of the universe these redshifts corresponds to.

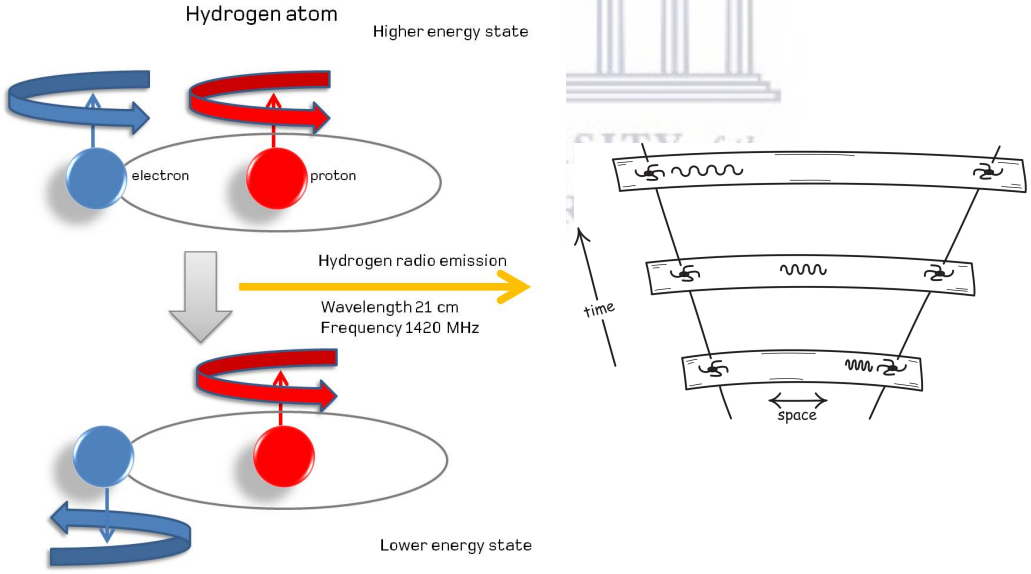

FIGURE 2.1: (*Left*) A schematic showing the HI 21 cm radio radio signal is emitted. Transiting from parallel spin to anti-parallel spin results in the emission of a radio wave with a frequency of 1420.4 MHz and wavelength of 21 cm [From: https://www.skatelescope.org/radio-astronomy/]. (*Right*): A cartoon representation of the effect of the expansion of the Universe on the frequency of the 21 cm line [From: https/www.pitt.edu/ jdnorton/teaching].

2.2 Evolution of the HI 21 cm Signal

In this section we briefly review the evolution of 21 cm signal at different epochs in the cosmic history of the Universe. By studying how the 21 cm signal evolves, information about the formation of the first astrophysical sources such as galaxies, the Intergalactic Medium (IGM) and spatial distribution of neutral HI post the epoch of reionization can be probed. For a thorough discussion, see Mesinger et al. [2016].

The differential brightness temperature is a measure of the HI intensity and can be modelled relative to temperature of the CMB for a chosen line of sight and frequency as a function of redshift ($z = \nu_{21}/\nu - 1$):

$$\delta T_b \equiv T_b - T_{CMB}, \tag{2.2}$$

such that

$$\begin{aligned} \delta T_b(\nu) &= \frac{T_S - T_\gamma}{1+z}(1 - e^{-\tau_{21}}) \\ &\approx 27 x_{\rm HI}(1+\delta_{nl})\left(\frac{H}{H + d\nu_r/dr}\right)\left(1 - \frac{T_\gamma}{T_S}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_{m0}h^2}\right)^{1/2}\left(\frac{\Omega_{b0}h^2}{0.023}\right) \text{ mK}, \end{aligned} \tag{2.3}$$

where T_γ is the CMB temperature, T_s is the spin temperature of the cosmic gas, $x_{\rm HI}$ is the neutral hydrogen fraction, H is the Hubble parameter, $\delta_{nl} = \rho/\bar{\rho} - 1$ is the density contrast, and $d\nu_r/dr$ is the comoving gradient of the line of sight comoving velocity.

As shown above, the evolution of the brightness temperature depends on a number of parameters which arise from the different phases within the evolution of HI in the Universe. The following section briefly describes each of these phases.

Dark Ages – collisional coupling: During this epoch, the Universe is mainly composed of a hot and dense IGM. Stars and galaxies have not yet formed at this stage, with hydrogen being one of the sources of photons via its forbidden 21 cm hyperfine transition. These conditions caused the spin temperature T_S to collisionally couple with the gas kinetic temperature T_k, such that they are in thermal equilibrium $T_K = T_S \lesssim T_\gamma$. Due to the expansion of the Universe, thermal decoupling from the CMB (for $z \lesssim 200 - 300$) occurred, resulting in two different adiabatic cooling rates for the temperatures of the IGM and the gas kinetic temperature. With the IGM cooling at a faster rate as $T_k \propto (1+z)^2$ than the temperature of the CMB, which follows $T_\gamma \propto (1+z)$, this causes the evolution

of the brightness temperature δT_b during the dark ages to be negative. Hence there is a negative evolution in T_b which is shown on the right end of middle panel of figure 2.2 at high redshift. Since hydrogen has not been ionized at these high redshifts, (i.e. $x_{\rm HI} \sim 1$), it is the density fluctuations that contribute to the measured power spectrum since the brightness temperature is directly proportional to the density contrast. The higher redshift range (between $z = 50$ and $z = 200$) in the Dark Ages epoch has the potential to serve as a probe for the matter power spectrum provided it can be measured by future radio telescopes.

Dark Ages – collisional decoupling: Towards the end of the Dark Ages and as the Universe expands, the IGM becomes less dense resulting in less collision between the 21 cm photons and IGM. The spin temperature decouples from the gas kinetic temperature such that $T_k < T_S \lesssim T_\gamma$, and rises towards the CMB temperature T_γ causing δT_b to further increase towards 0 mK as shown graphically in the middle panel of figure 2.2. It is the local gas density that drives the decoupling process of T_s from T_k, beginning from regions of under density of the IGM. Due to the low collisional coupling, there is reduced signal observed on the power spectrum. Graphically we can observe this epoch as the *red to black* phase in the cosmic strip (*top panel*).

Cosmic Dawn: is a period in which the formation of the first stars and galaxies begins. This epoch is driven by two major processes, Lyα coupling at the start, and X-ray heating at the end of the epoch. On figure 2.2, this phase is shown on cosmic strip as the *yellow to blue* transition and corresponds to the dip present in the brightness temperature evolution for $10.5 < z < 29.1$ and the power spectrum exhibiting an overall rise in power in the same redshift range, due to these two processes.

Epoch of Reionization: The EOR arises as a result of the switching on of the first stars and galaxies from the cosmic dawn period, even though beginning of the EOR is not clearly understood, its pre-phase is driven by X-ray Heating from the Cosmic Dawn period. The newly formed astrophysical sources change their surrounding environment by emitting Ultra Violet (UV) radiation which excites the electrons from the neutral hydrogen leaving the atoms ionized. The ionized bubbles grow and overlap interacting with the IGM between the newly formed galaxies causing the majority of hydrogen in the Universe to be further ionized [Cheng, 2019]. The EOR can be probed in a number of ways – here we highlight three techniques.

- **Quasars**: The first approach is to study the spectra of distant Quasi-stellar Radio Sources (Quasars). These are highly energetic and point like AGN, ranking amongst the brightest of astrophysical sources present in the Universe. Observations of their spectra reveal that a majority of them are similar in nature, irrespective of their location in space. Some of these objects are found in the early Universe. An analysis of their spectra reveals an absorption feature referred to as the Gunn-Peterson trough [Gunn and Peterson, 1965] which occurs when photons are absorbed by the intervening neutral hydrogen as the light travels from the quasars to the detectors on Earth. There is a difference observed in the shape of the absorption feature between quasars that are nearby and those that are found in the early Universe. Quasars nearby show sharp absorption lines whilst those found in the early Universe exhibit a Gunn-Peterson trough. The presence of this trough indicates that most of the quasars photons were absorbed by neutral hydrogen, therefore by tracking changes in quasars spectra at different times within this phase, an understanding about the abundance of neutral hydrogen as a function of redshift can be achieved. Studies conducted by Becker et al. [2001] using this method indicate that the epoch of reionization came to an end around $z \sim 6$.

- **CMB:** Observations of the CMB provide a snapshot on the state of the Universe following its formation. CMB photons permeate space, bouncing off of free electrons and causing them to move in different and random directions (via thompson scattering). Within and post reionization, the Universe has undergone some expansion, enough to decrease the amount of free electrons which results in lowering the amount of scattering that occurs. The scattering of free electrons leaves a signature on the CMB maps which results in secondary temperature and polarization fluctuations [Haiman and Knox, 1999]. One of the parameters of the CMB that are sensitive the amount of scattering is the amplitude of the CMB. The scattering of free electrons dampens out fluctuations present in the CMB lowering the amplitude of the CMB [Cheng, 2019]. By tracing the amount of scattering that at different times, one can get a measure of when reionization ended. The cosmological parameters that accounts for the amount of free electron scattering is the optical depth, τ and is given by:

$$\tau = \int n_e \sigma_T dl \tag{2.4}$$

where the number of free electrons are represented by n_e, σ_T is the Thompson cross-section, and dl is the proper length. Various teams have conducted studies to constrain this parameter, with the latest being the Planck Collaboration et al. [2020]. Their investigations constrained the reionization redshift to $z = 7.68 \pm 0.79$.

- **HI 21 cm line:** using the HI intensity mapping technique (to be discussed in detail on section 2.3) to map the distribution of neutral hydrogen in 3-D (angular and as function of redshift), will be a direct way to trace the evolution of hydrogen through the epoch of reionization. HI IM will provide information on the HI abundance and amount of ionized hydrogen gas, giving insight to how the Universe evolved in this era. There has been a tentative measurement of the global HI 21 cm signal by the *Experiment to Detect the Global EoR Signature* (EDGES; Bowman et al. [2018]). Other experiments that aim to probe this epoch include the Precision Array for Probing the Epoch of Reionization (PAPER) [Cheng et al., 2018], Hydrogen Epoch of Reionization Array[1] (HERA).

Post-Reionization: After the reionization epoch, the majority of neutral HI is found within highly dense environments within galaxies such as Damped Lyman Alpha Absorbers (DLAs) [Santos et al., 2015]. Due to the high density of the environment, neutral HI is shielded from being ionized by the UV radiation as it cannot gain access to it [Bagla et al., 2010]. Due to this event of nature, the remaining HI signals within this epoch lie within these galaxies and biasedly traces the underlying large scale structure. If the current understanding is indeed correct, then HI does have the potential to be used as a cosmological probe for the late Universe.

It cannot be safely said that some of the neutral HI signals does not reside in systems outside these dense structures of galaxies, due to limitations in observational 21 cm data from this era. To gain a better understanding of the processes involved during this epoch, cosmologists rely on complicated computer simulated models of the post reionization period such those from Frenk and White [2012], Springel et al. [2018], and Spinelli et al. [2020], with more of these simulations being created. Other computer simulations based on theoretical modeling from studies such as (Popping et al. [2009], Villaescusa-Navarro et al. [2014]) assist in understanding the exact abundance and spatial distribution of HI.

[1] https://reionization.org/

What comes out from these types of studies is that most of the neutral HI is located in dark matter haloes [Villaescusa-Navarro et al., 2018], which are gravitationally bound structures and from which matter can collapse in order for galaxies to form.

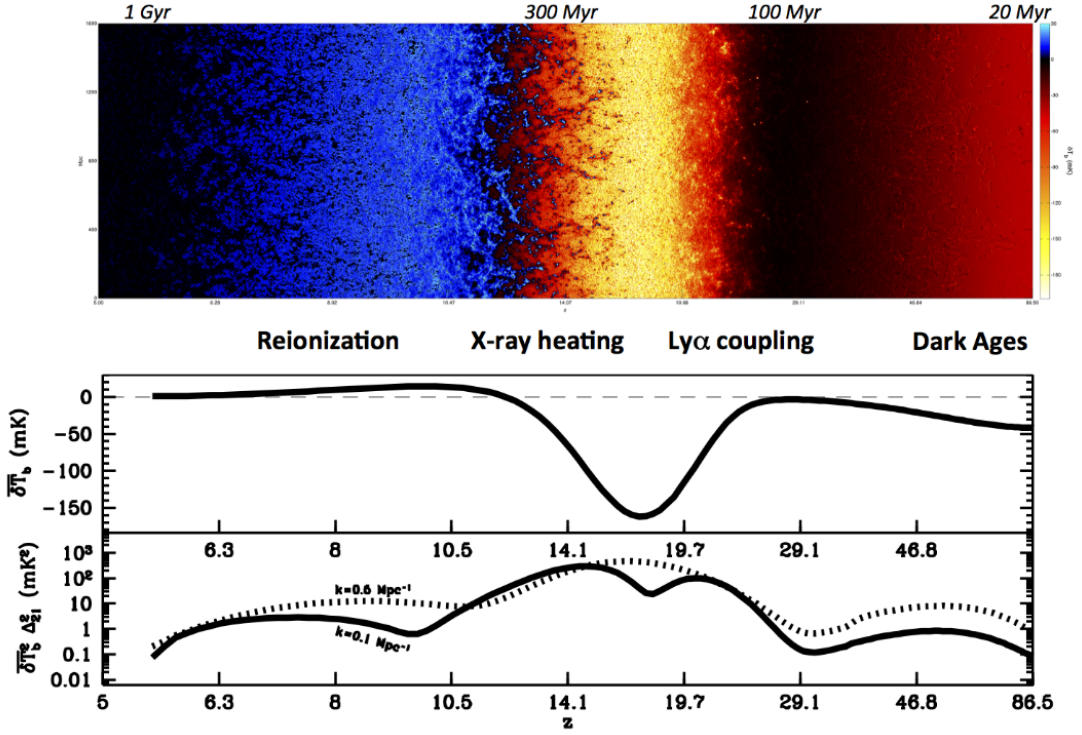

FIGURE 2.2: Shown on the top panel of this figure is a light cone strip depicting graphically what occurs at different times in the evolution of the global 21 cm signal. (*middle*) Model of the evolution of the global 21 cm brightness temperature and the bottom panel shows the amplitude of the power spectrum at $k = 0.1$ Mpc^{-1} (*black solid curve*) and $k = 0.5$ Mpc^{-1}(*black dashed curve*) as a function of redshift shown at different epochs; the dark ages, at cosmic dawn, and during the Reionization period by Mesinger et al. [2016].

2.3 HI Intensity Mapping Technique

We have seen above the role of neutral HI in the evolution of the Universe. In this section we provide a description of an observational technique that makes use of neutral hydrogen to map the large scale structure of the Universe. We conduct a brief overview of the potential instruments such as MeerKAT and SKAO-MID have in order to provide improved constraints on cosmology.

HI Intensity Mapping (HI IM) is an innovative technique whose main method of conducting observations of the sky does not require the resolving of individual galaxies as in traditional galaxy surveys, but instead relies on measuring the integrated intensity of HI

(21 cm) emission on each pixel on the sky (Battye et al. [2004]; McQuinn et al. [2006]; Chang et al. [2008], Mao et al. [2008], Loeb and Wyithe [2008]; Pritchard and Loeb [2008]; Wyithe and Loeb [2008]; Wyithe et al. [2008], Peterson et al. [2009]; Bagla et al. [2010]; Seo et al. [2010]; Lidz et al. [2011]; Ansari et al. [2012]; Battye et al. [2013]). In doing so HI IM provides a way of rapidly scanning across the sky covering large areas in short periods of time producing low resolution maps.

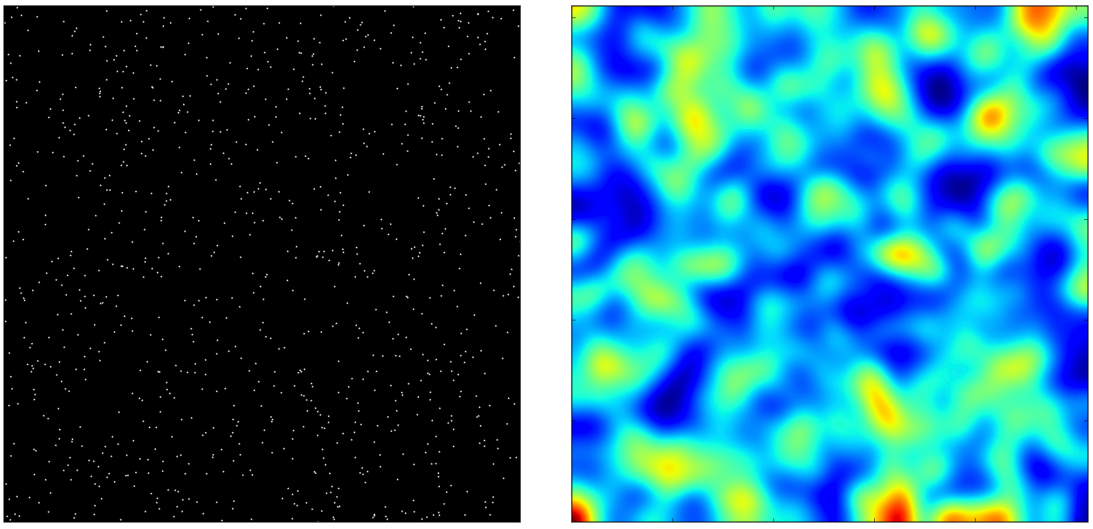

FIGURE 2.3: An example of a single map from a galaxy survey (left) compared to an intensity map obtained using the HI IM technique (right). In the galaxy map, each *white* dot represents a galaxy whereas in the intensity map, the sky is divided into pixels and the combined HI emission from each pixel is measured. Regions of high intensity are represented in *red* whilst those of low intensity are shown in *blue* [From: https://franciscovillaescusa.github.io/im.html].

Figure 2.3 shows the difference between a single galaxy survey map and HI IM map. On the *left*, the *white* dots represent individual galaxies detected in a galaxy survey whilst on the *right*, the observed field is partitioned into pixels in which the combined HI emission is measured, where regions of high intensity are represented in *red* and those of low intensity in *blue*. The HI distribution is mapped in both angular and frequency by combining emission from multiple HI emitting galaxies, thereby increasing the strength of the observed HI intensity. This also means that the observed intensity will trace the underlying density fluctuations which are also mapped by the HI emitting galaxies [Square Kilometre Array Cosmology Science Working Group et al., 2020]. From the HI IM maps, a reconstruction of the HI density can be conducted on sufficiently large scales. Even though the intensity maps obtained are low resolution, they are still adequate for cosmological analysis.

With the HI IM technique, the 21 cm line is being probed, therefore there exists a one to one correspondence between the frequency of the measured line and the redshift, ensuring that a high spectral resolution is obtained and the time required to conduct HI IM observations is reduced when compared to that required by galaxy surveys. In comparison to other spectral lines, the 21 cm line is isolated, therefore experiencing less contamination from other spectral lines [Fonseca et al., 2017]. With respect to spectroscopic galaxy surveys, for a single detection of a galaxy to be made, a high signal to noise ratio must be achieved, which requires more time. By taking the HI observations at different frequencies, HI IM tomography maps can be constructed providing a means to trace the evolution of the HI intensity as a function of redshift as shown in figure 2.4.

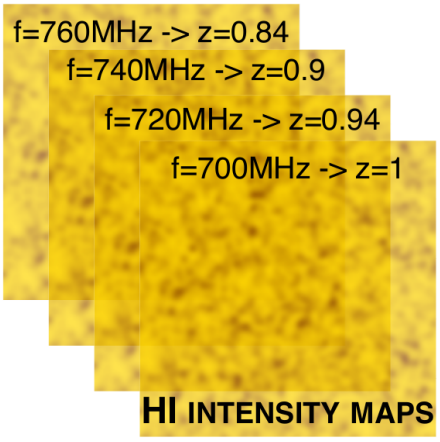

FIGURE 2.4: A cartoon representation of the concept of intensity mapping tomography [Courtesy Laura Wolz].

The HI IM technique does have its own challenges, such as the presence of foregrounds within the same frequency band as the cosmological signal, and instrumental effects which convolve with the measured signal. These challenges have been explored in various studies and are briefly summarized in section 2.4 and efforts to recover the cosmological HI signal are discussed in section 2.5.

Cosmological surveys have been planned with the SKAO and are informed by the instrument design which determines what project can be conducted at a particular phase, during the life time of the project. The Square Kilometre Array Cosmology Science Working Group et al. [2020] (also known as the *Red book*) describes in detail the HI IM surveys to be conducted with SKAO-MID and SKAO-LOW. Here, a brief summary is provided.

HI IM will be possible in both single dish and interferometric mode. In relation to HI IM, SKAO-MID will be used in single dish mode instead of interferometric mode, due to the dish distribution having a low number of short baselines in the array, which are insufficient to achieve the required signal to noise on the necessary scales [Bull et al., 2015]. With the huge number of radio telescope dishes SKAO-MID has, the high survey speeds required for HI IM cosmology are ensured at the required signal to noise levels [Santos et al., 2005].

The first survey with SKAO-MID, called the *Wide Band 1 Survey*, will be a wide but shallow survey, covering a sky area of 20000 square degrees and probing a redshift range between $0.35 < z < 3$ with a total observation time of 10000 hours on sky. Together with this survey, a wide continuum galaxy survey will also be conducted over the same redshift range. The second survey is the *Medium-Deep Band 2 Survey* which will cover 5000 square degrees and using 10000 hours of integration time between redshift range $0 < z < 0.4$. This survey will also be use to conduct other surveys such as an HI galaxy and a weak continuum lensing survey.

With SKAO-LOW, there will be the *Deep SKAO-LOW Survey's* whose scientific goal is to use the HI IM technique to probe the EOR by mapping the distribution of HI within this epoch. This survey will cover a much more smaller sky area of approximately 100 square degrees but probe a deeper in redshift range $3 < z < 6$ (corresponding to a frequency band 200 – 350 MHz) with a total integration time of 5000 hours.

In the following section, we briefly look at some of the forecasts results expected to be obtained when the SKAO HI IM surveys have been conducted.

2.3.1 HI Power Spectrum

As mentioned in section 1.5, the HI power spectrum cannot be directly measured with observations but an estimate can be obtained by measuring the brightness temperature from the 21 cm emission line and computing the 21 cm power spectrum. The brightness temperature is what will be measured with the planned HI IM experiments to be conducted using MeerKAT and SKAO.

The redshift space power spectrum in equation 1.62 can be expressed in terms of the brightness temperature in the following form:

$$P_{\rm HI}(z,k) = \bar{T}_b(z)^2 b_{\rm HI}^2(z)[1+\beta_{\rm HI}(z)\mu^2]^2 P_m(z,k), \tag{2.5}$$

where $\bar{T}_b(z)$ represents the average brightness temperature expressed as a function of the HI density and is given by:

$$\bar{T}_b(z) = 190h\frac{H_0(1+z)^2}{H(z)}\Omega_{HI}(z) \quad \text{mK}. \tag{2.6}$$

The brightness temperature plays a role in determining the amplitude of the HI power spectrum. The HI density describes how the density of HI evolves as a function of redshift and to obtain the HI density we must rely on hydro-dynamical simulations. It is possible to model the HI density via the halo model. Under this assumption the HI density can be modeled by computing the following integral (see for example Villaescusa-Navarro et al. [2017]):

$$\Omega_{HI}(z) = \frac{1}{\rho_c^0}\int_0^\infty n(M,z)M_{HI}(M,z)dM, \tag{2.7}$$

where ρ_c^0 represents the critical density of the Universe today, with the average halo mass function is given by $n(M,z)$ and $M_{HI}(M,z)$ is the mean HI mass of a dark matter halo with mass M at redshift z. There have been a few measurements of the HI density as depicted in figure 2.6.

The second unknown of equation 2.5 is the HI bias that could also play a role in determining the shape of the HI power spectrum. Under the assumption of the halo model, the HI bias can be modeled via the following expression

$$b_{HI} = \frac{1}{\rho_c^0\Omega_{HI}(z)}\int_0^\infty b(M,z)n(M,z)M_{HI}(M,z)dM, \tag{2.8}$$

where $b(M,z)$ is the dark matter halo bias, $n(M,z)$ is the halo mass function.

Prediction for the HI bias can also be obtained with the aid of hydro-dynamical simulations. We can get the evolution of the HI bias by computing the square root of the ratio of the HI power spectrum with the matter power spectrum as shown in Figure 2.5. This figure shows the evolution of the bias as a function of wavelength scale k at different redshifts ($z = 0-5$) [Villaescusa-Navarro et al., 2018].

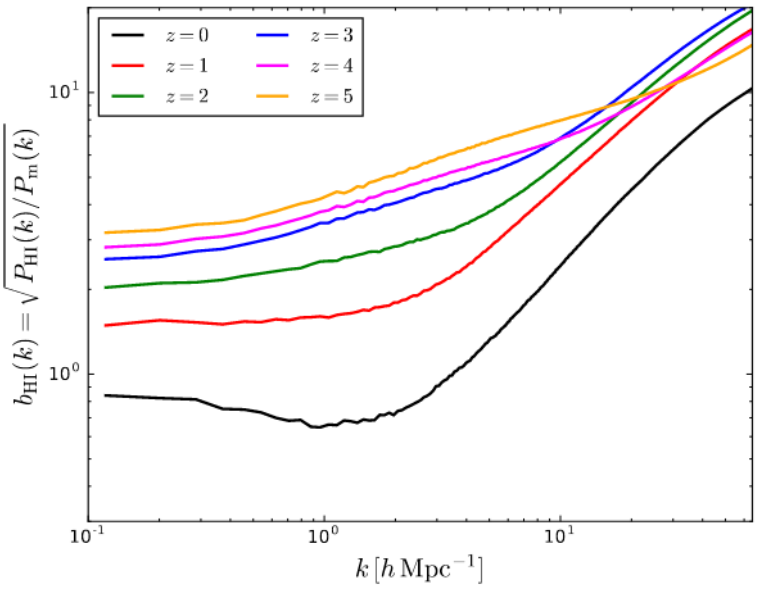

FIGURE 2.5: The evolution of the HI bias as a function of both redshift and comoving wavelength k scale [Villaescusa-Navarro et al., 2018]

2.3.2 Forecasts for SKAO

The solid black circles in figure 2.6 indicate the constraints both SKAO IM experiments can place on the HI density. Notable is how well this parameter can be constraint at low redshift, with the errors bars slightly increasing with redshift. However in these forecast measurements the contamination from systematics and other residual foregrounds were ignored and are subject of future work.

The top panel of Figure 2.7 shows the expected HI power spectrum detection (*black*) by the SKAO-MID radio telescope. This prediction was computed using a redshift bin of $\Delta z = 0.1$ and centered at $z = 0.6$. The error on the measurement given in *cyan* shows the tight constraint on the measurement, with the error shooting up at large k, which corresponds to small spatial scales that HI IM in single dish mode is poor in resolving.

On the *bottom* panel is the power spectrum predicted for a *Deep SKAO-LOW Survey* (*solid black*) and the error on the measurement (*cyan*), computed at $z = 4$ for a redshift bin of $z = 0.3$. However this result ignores contribution from RSDs. The SKAO-LOW measurement depicts the type of constraints that can be achieved by the radio telescope in its attempts to probe a new redshift space for probing the HI abundance and bias.

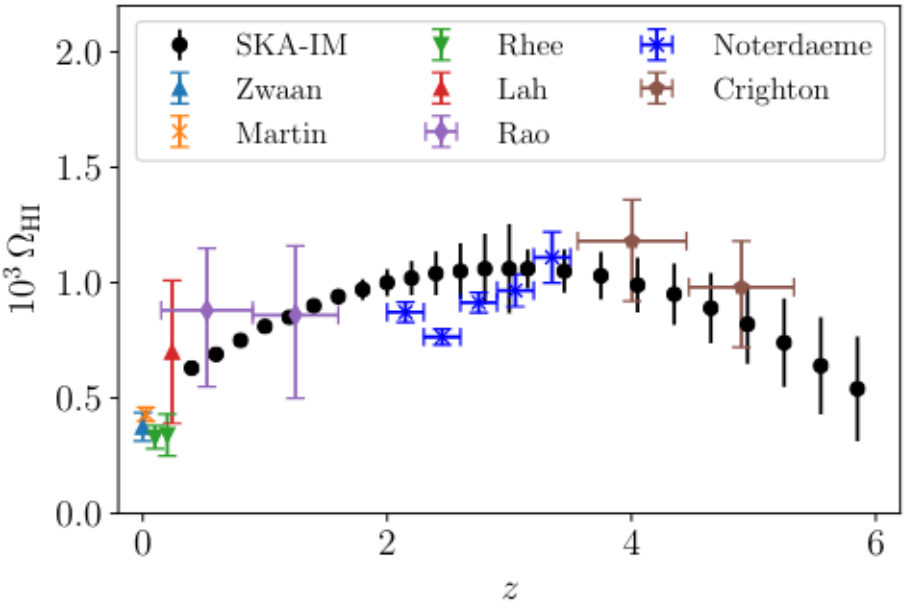

FIGURE 2.6: Forecast measurements (*in solid black points*) of the HI density $\Omega_{\rm HI}$ as a function of redshift were computed using both SKAO Wide Band 1 and *Deep* SKAO-LOW survey. Notable is how well the HI density can be constrained at low redshift. The expected $\Omega_{\rm HI}$ is compared with other data points from work by Crighton et al. [2015] and other teams cited in that study which used a methodology found in Pourtsidou [2017]. This result was obtained by using a very conservative non-linear k_{max} cutoff was used in the Fisher forecast analysis [Square Kilometre Array Cosmology Science Working Group et al., 2020].

The forecast measurements were obtained using a Fisher forecast analysis, which considered information such as survey design and fixed cosmological parameters from Planck Collaboration et al. [2016].

2.3.3 MeerKAT Large Area Synoptic Survey (MeerKLASS)

MeerKLASS is the proposed wide area HI IM survey for the MeerKAT telescope. Its main science goal is to measure the HI auto-correlation power spectrum by mapping the 3-D distribution of neutral HI via the 21 cm signal [Santos et al., 2017]. The survey will cover 4000 square degrees in 4000 hours and conduct observations in the L-Band frequency range (856 - 1712 MHz), overlapping the same sky footprint as the Dark Energy Survey[2] (DES). Listed below are a few other broad aims MeerKLASS wishes to achieve:

[2] www.darkenergysurvey.org

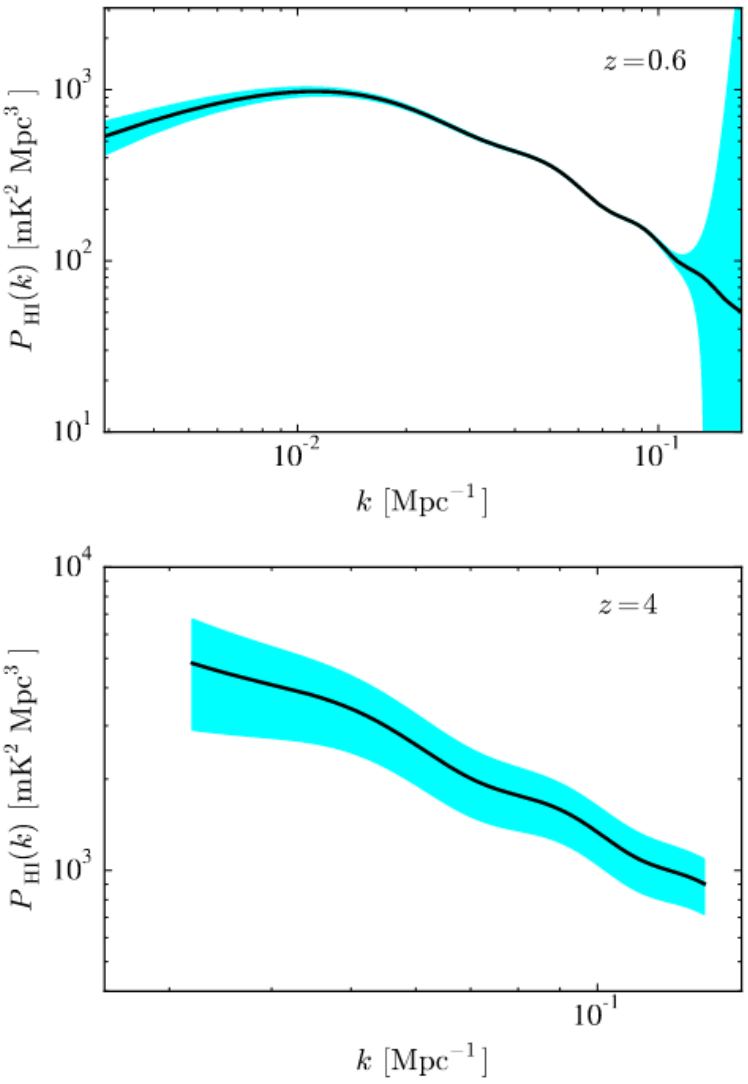

FIGURE 2.7: The *upper panel* shows the expected auto-correlation HI IM power spectrum (solid black line) using survey specifications for the SKAO-MID Wide Band 1 Survey from Square Kilometre Array Cosmology Science Working Group et al. [2020]. Shown in *cyan* is the expected errors in the measurement. The power spectrum measurement is calculated with an assumption of a $\Delta k = 0.01$ Mpc^{-1} binning at $z = 0.6$. On the *lower panel*, survey specifications of the Deep SKAO-LOW survey are used to forecast the expected HI power spectrum at $z = 4$. The solid black curve is the HI detection, whilst in cyan is the measurement errors. A redshift bin of $= 0.3$ and a comoving wavelength k binning of $\Delta k = 0.01$ Mpc^{-1}.

- to use the HI IM technique (in both auto-correlation and cross-correlation) to make the first detection of the BAO and redshift space distortions (RSD's) in order to probe the nature of dark energy and to test other theories of gravity;
- to use the multi-tracer technique to place constraints on non-Gaussianity during the inflationary period of the Universe;
- to perform cross-correlation with CMB data in order to place constraints on the matter power spectrum at high redshifts;
- to use MeerKLASS as an extragalactic survey;
- to use MeerKLASS for the detection of polarized sources;
- to study clusters and galaxy HI emission.

HI IM with MeerKAT: Pilot Study

A feasibility study for a HI IM survey with MeerKAT telescope was conducted by Wang et al. [2021] to show that the MeerKAT radio telescope has the potential to measure the cosmological HI. As proposed with the MeerKLASS survey, key to achieving this goal is the use of the MeerKAT telescope as a collection of single dish telescopes rather than an interferometer. With just six nights of observations and using all MeerKAT dishes, 10.5 hours of data was collected from 200 square degrees of sky area known as the Wiggle Z 11hr field from the DES survey [Drinkwater et al., 2010]. The investigation focused on a frequency range between 856 - 1712 MHz with a 0.2 MHz resolution. A multi-level data reduction pipeline was developed whose foundation relied on the ability to perform RFI flagging at different levels of the calibration, the periodic injection of instrumental noise diodes in order to achieve stability in how the gains drift, and the effective use of multi-component sky models. Applying the multilevel calibration method to the observed data, the full calibration of the auto-correlation data from all MeerKAT dishes, and using both horizontal and vertical polarization was achieved. Maps of diffuse galactic emission were recovered as shown in figure 2.8. In the figure the first panel is the assumed model for synchrotron emission while the second panel shows the final map obtained after calibration and map-making. Point sources in *magenta* with flux greater than 1 Jy at 1.4 GHz are also shown. Note that these were not part of the sky model during the calibration process. The third panel shows the residual after subtracting the

smooth diffuse foreground model. From the residuals plot, one can observe the excess flux introduced by these bright point sources. The total number of time stamps each pixel in each scan and all dishes is shown on the *lower panel*. For comparison with other studies, the HI column density was extracted from the map reconstructed by the pipeline and was shown to be in good agreement with the HI column density map constructed from the Effelsberg-Bonn HI survey [HI4PI Collaboration et al., 2016]. For a detailed analysis on this MeerKAT pilot study, see Wang et al. [2021].

2.3.4 Other HI IM Experiments

In this section, we briefly review some of the radio telescopes that have been used to conducted HI IM surveys thus far, exploring the basic telescope design and science goals. A review by Kovetz et al. [2017] provides details on some of the several planned HI IM experiments using various instruments, including those briefly outlined in this section. The first two instruments to be discussed were not initially built particularly for HI IM studies, but were one of the first instruments in which the technique could be tested and were able to make detections of the HI signal via cross-correlation of HI IM data with optical galaxy surveys.

Parkes: Parkes is a 64 m single dish radio telescope based in New South Wales, Australia and is amongst the largest radio telescopes in the south of the equator. Parkes[3] which is part of the Australian Telescope National Facility (ATNF) became one of the first instrument to provide a detection of the HI signal that was statistically significant [Pen et al., 2009]. See section 2.7 for a summary of the study that was conducted.

Since its construction in 1961, the radio telescope has received a number upgrades in its systems, such as its several receiver systems which include the 13 beam receiver that is capable of observing 20-cm radio wavelengths and has been used to conduct a number of large scale HI surveys such as (Anderson et al. [2018], Li et al. [2021]), the 7-beam receiver used for 'methanol' observations at microwave wavelengths and the ASKAP phased array feed which was tested on the Parkes [Hobbs et al., 2020]. The various receivers can be switched into focus depending on the science objectives. Observations of Pulsars can also be done with the radio telescope using the phased array feeds [Deng et al., 2017]. The Parkes radio telescopes also participates in VLBI research.

[3] https://www.csiro.au/en/about/facilities-collections/atnf/parkes-radio-telescope

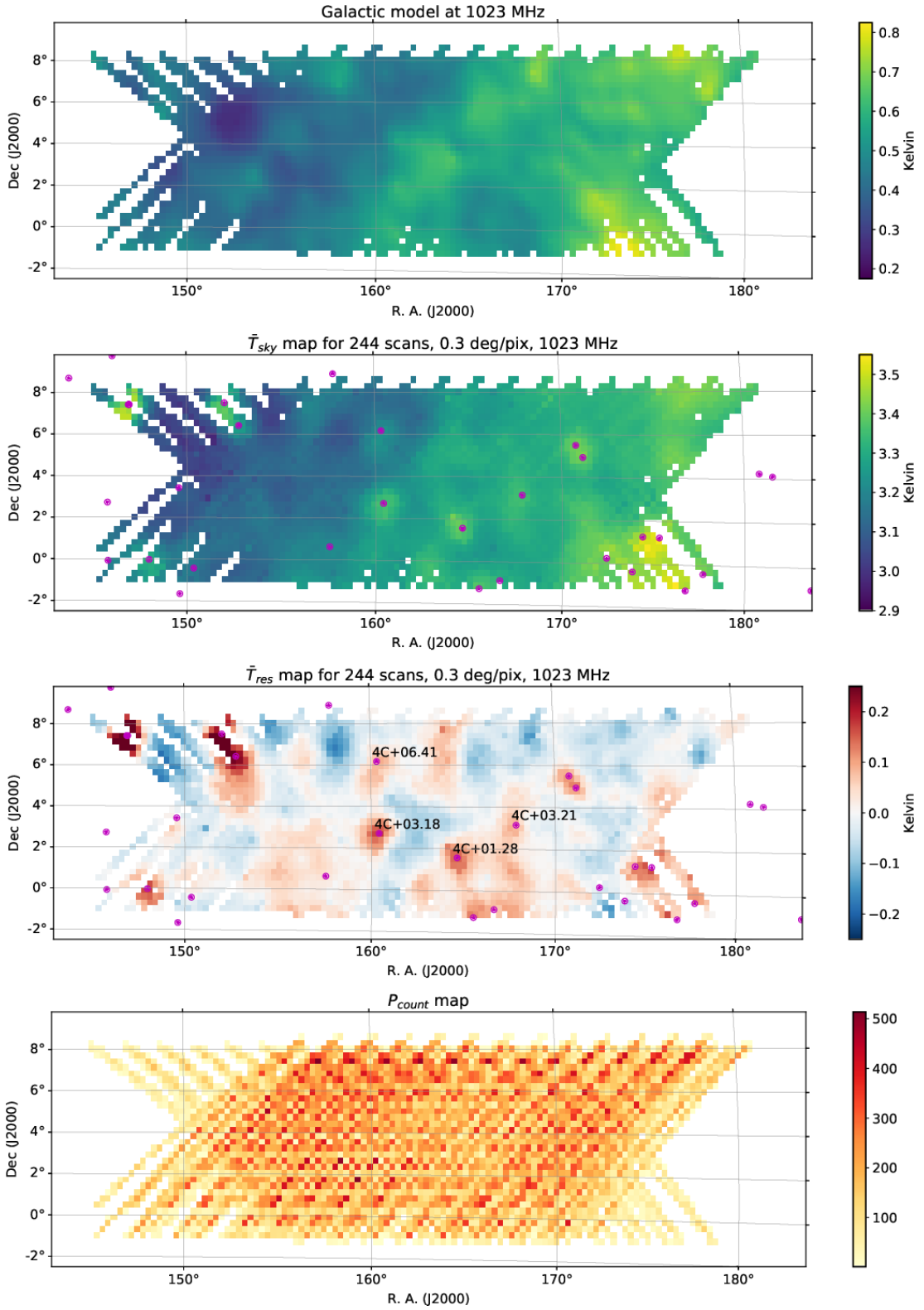

FIGURE 2.8: Following three rounds of RFI flagging are the combined average HI intensity maps of all scans at 1023 MHz. In the *top* panel is a model for the galactic emission obtained from PySM [Thorne et al., 2017] over the same sky area under investigation. Situated in the *middle panel* is the combined sky temperature map (T_{sky}) measured by the dishes and the residual map below. Overplotted on both plots are the point sources with flux > 1 Jy at 1.4 GHz but are not included in the modeling of the sky brightness temperature maps. On the *lower* panel is the total number of time steps across each dish and for all scans [Wang et al., 2021].

FIGURE 2.9: Illustration of the various type of radio telescope designs, located on the *left* is the GBT radio telescope, which is single dish radio telescope located Green Bank, West Virginia, USA [from: https://greenbankobservatory.org/science/telescopes/gbt/]. On the *right* is the Tianlai radio telescope array in Northern China [From: http://tianlai.bao.ac.cn].

GBT: The Green Bank Telescope (GBT) holds the record of being the world's largest fully steerable single dish radio telescope (shown on the *left* of figure 2.9). Hosted by the National Radio Astronomy Observatory (NRAO) in West Virgina, US, the GBT has an aperture of 100m and with its large collecting area and high sensitivity, the radio telescope can detect millimeter wavelengths over a frequency range of 0.1 to 116 GHz. In terms of sky coverage, the GBT is able to observe about 85 % of the celestial sphere.

The GBT telescope is attributed to having produced the first ever HI signal detection using the HI IM technique. This investigation was conducted in a study published by Chang et al. [2010] where HI IM data from the GBT was cross-correlated with an optical galaxy survey data from the Hawaiian based Keck telescope. See section 2.7 for a brief overview of the main results. Following that a few other HI IM studies have been conducted with the GBT such as Masui et al. [2013] and Wolz et al. [2017, 2021].

TIANLAI: is a radio astronomy project whose goal is to measure the BAO imprints

on the large-scale structure in order to detect Dark Energy. Tianlai[4] is a stationary cylindrical array that is 100 m in the North-South and East-West directions. The main telescope consists of an array of dishes and semi-cylindrical dishes that make use of the Earth's rotation in order to observe its intended field of view. The project plan is to conduct HI IM observations in order to map the large scale structure of the Universe at low redshift $0 < z < 2.5$, which corresponds to a frequency band between 0.4 to 1.4 GHz (Chen [2012]; Wu et al. [2021]).

FAST: The Five-hundred-meter Aperture Spherical Telescope[5] (FAST) is a 500 m aperture radio telescope with 300 m fully illuminated. The single dish radio telescope is placed on a karst depression in the Guizhou mountains of China's south western province as seen on the right panel of figure 2.10. FAST holds the title of being one of the largest radio telescopes on Earth, with an aperture size equivalent to 30 football fields. This single dish radio telescope is a multibeam receiver with 19 feedhorns and designed to operate in L-Band with a wide frequency range of 70 MHz to 3 GHz. This high resolution radio telescope has the potential to cover a sky area of 25000 square degrees. By design the telescope will be able to perform drift scanning observations with its field of view passing over as the earth rotates. A HI IM survey with FAST will consider about 6000 square degrees of sky area, and a frequency range between 950-1350 MHz. With such a survey, the 3-D mapping of HI will be conducted in order to probe the BAO's and constrain properties of dark energy. For a thorough discussion on the FAST telescope, proposed HI IM survey and forecasts of cosmological constraints from HI IM with FAST, please see (Nan et al. [2011]; Bigot-Sazy et al. [2015] Smoot and Debono [2017] and Yohana et al. [2019]). Some of the FAST radio telescope's scientific goals are

- to reconstruct images of the Universe at its early formation by mapping neutral hydrogen at very high redshift,
- to investigate hyperfine structures of celestial sources through participating in the Very-Long-Baseline Interferometry (VLBI) network,
- to search and find pulsars, construction of a pulsar timing array, pulsar navigation and in the future, to participate in gravitational wave and detection research,

[4] http://tianlai.bao.ac.cn/
[5] https://fast.bao.ac.cn/

- to participate in the search for intelligent radio signals from within and beyond our galaxy,
- to conduct a spectral survey at high resolution and detect weak signals from space.

FIGURE 2.10: Shown on the left is the semi-cylindrical CHIME radio telescope based in Canada, North America [From https://chime-experiment.ca/en]. On the right is the FAST radio telescope constructed on a karst depression in the Ginzhou mountains of south west China [From https://aasnova.org/2020/04/17/pulsar-discovery-from-an-enormous-telescope/].

CHIME: The Canadian Hydrogen Intensity Mapping Experiment[6] (CHIME) is a radio telescope array at the Dominion Radio Astrophysical Observation, Canada. CHIME consists of fixed position semi-cylindrical dishes as shown the left panel of figure 2.10. Since the telescope does not move, it relies on the earth's rotational axis in order to observe its large field of view of approximately 200 square degrees at an instant and is able to observe a wide frequency band of 400 - 800 MHz.

CHIME has a number of science goals. Primarily the radio telescope was designed to be a cosmology instrument with the goal to observe the neutral HI distribution in 3 dimensions. These efforts are beginning to bear fruit as the first detection of the HI signal between redshifts $0.78 < z <= 1.43$, by cross correlating and stacking of 102 nights of

[6] https://chime-experiment.ca/en

observational data with optical galaxies, has been conducted [CHIME Collaboration et al., 2022]. Secondly, the radio telescope's digital signal processing technology can also be put to use for other science applications such as the everyday surveillance of pulsars and the detection of Fast Radio Bursts (FRB's). In the period between 2018-2019 of CHIME's operation, the telescope has been able to detect around 500 new astrophysical sources of this type. FRB's are quick flashes of light (in the order of milliseconds) that can be picked up at radio frequencies [Chu, 2021].

BINGO: The Baryon Acoustic Oscillations from Integrated Neutral Gas Observations[7] is a multinational radio astronomy project led by the University of Sao Paolo, Brazil, whose aim is to construct the first radio telescope mainly for mapping neutral hydrogen using the 21 cm radio wave emission in order to measure the BAO at low redshift [Abdalla et al., 2021]. Initially presented by Battye et al. [2012], BINGO is a stationary single dish radio telescope that will operate using drift scanning, which means that the telescope makes use of the rotation of the earth to observe a different patch of the sky [Abdalla et al., 2021]. The telescope will consist of a 40 m primary dish, a 36 m secondary dish and 28 feed horns for the first phase. The number of feeds will be increased to 60 upon completion of construction. An HI IM survey of the BINGO project will cover a wide area of 6000 square degrees, observing at a frequency range between 980-1260 MHz [corresponding to a redshift range: $0.13 < z \leq 0.45$] with an angular resolution of 0.67 degrees (40 arcmins). The design of the telescope is informed by its science goals which are:

- to map the distribution of neutral hydrogen in 3 dimensions and at low redshift and from the large scale structure of HI measure the BAO in order to constrain the dark sector and test the standard concordance model (ΛCDM),
- to understand the systematics and develop data analysis tools and machinery to aid in the extraction of the small-scale cosmological HI signal,
- to detect Fast Radio Bursts and Rotating Radio Transients (RRATs).

BINGO also serves as a pathfinder for the SKA telescope in its quest to conduct large scale and ultra deep HI IM Survey. Lessons from BINGO about systematics and data analysis of how to extract the small scale HI cosmological signal will be useful for the SKA.

[7] https://bingotelescope.org/

FIGURE 2.11: An artistic impression of the HIRAX radio interferometer located in Carnarvon, Northern Cape, South Africa (from Crichton et al. [2022]).

HIRAX: The Hydrogen Intensity and Real Time Analysis eXperiment [8] (HIRAX) is an international radio astronomy project that will mainly conduct HI IM and radio continuum surveys, based in South Africa (Newburgh et al. [2016]; Crichton et al. [2022]). This radio telescope array will be deployed in the SARAO Karoo site alongside the MeerKAT and future SKA. It will contain a total of 1024 6m dishes beginning with an initial 256 dishes. With the dishes distributed in a closely packed grid as shown in figure 2.11, the radio telescope will operate as an interferometer and make use of the earth's rotation for observations of its large field of view. HIRAX will operate over a frequency range between 400 to 800 MHz, equivalent to a redshift range of 0.8-2.5, and observe $\sim$ 15000 square degrees of the Southern Hemisphere sky. The survey area will overlap the sky footprints of surveys such as DES, CHIME, KiDS[9], DESI[10], ACTPol[11], the SKA, and its precursor telescopes.

HIRAX has a number of science goals, which are based on the antenna design, dish distribution and instrumental processing capabilities. It will also be a complementary telescope to the MeerKAT and SKA, as there are a number of science goals which will be of benefit to both projects. For example, whilst SKA1 Mid will be performing single dish HI IM, which will be unable to resolve small angular scales, on the other hand HIRAX with its number of small baselines and redundant layout will be able probe those scales, providing greater resolution and probing deeper in redshift. The main goal is to use the

[8] https://hirax.ukzn.ac.za/
[9] https://kids.strw.leidenuniv.nl/
[10] https://www.desi.lbl.gov/
[11] https://act.princeton.edu/overview/camera-specifications/actpol

telescope to map the distribution of neutral HI via the 21 cm intensity mapping technique and use the BAO as a cosmic ruler in order to put constraints on Dark Energy. Forecasts show that HIRAX will be able to place 7 % constraint on the dark energy equation of state when its intensity mapping data is combined with that of the *Planck* satellite [Crichton et al., 2022].

The regular spacing of the radio telescopes provides the array the ability to make detections and monitor radio transients. Studies of FRB's by CHIME which has similar arrangement of its elements have shown the benefits of having such a configuration (CHIME/FRB Collaboration et al. [2019b]; CHIME/FRB Collaboration et al. [2019a]; CHIME/FRB Collaboration et al. [2019c]; Ng et al. [2020]).

2.4 Challenge of Foregrounds in HI IM

The term foregrounds in HI IM refers to radio emission within the same frequency range that HI IM will probe, but is not the desired cosmological HI signal. The term does not only refer a signal that is in front of the HI signal (i.e. in terms of physical source position or amplitude of the radio signal) but also to emission that may be embedded within the radio waves detected by a radio telescope apart from the HI signal. The source of foregrounds can be both terrestrial and extra-terrestrial foreground radio emission. By terrestrial foregrounds, we refer to radio emission that originates from the earth such as Radio Frequency Interference (RFI), thermal and atmospheric noise, etc. Extra-terrestrial foregrounds are radio emission from beyond the earth, i.e. our galaxy and other extra-galactic radio sources. In this section, we will focus on the extra-terrestrial foreground emission, describing their form, and in chapter 3, illustrate how we model the foregrounds for use in the single dish HI IM observation simulations.

For the HI IM technique to be successful it relies on the ability to separate between the weak cosmological HI signal and the strong foreground radio sources emitting in the same frequency range as the signal, $\sim 1420/(1+z)$ MHz. In the next section, we discuss the nature of the dominant foregrounds present at these frequencies.

2.4.1 The Radio Sky Below 1 GHz

Below the frequency range of 1 GHz, some of the major foregrounds that are of concern to HI IM are the following:

Galactic Synchrotron: this is one of the strongest foreground emission from within our galaxy and has an amplitude that is about 5 orders of magnitude larger than the expected HI 21 cm signal. This type of radiation arises as a result of cosmic-ray electrons and positrons that are accelerated within the interstellar magnetic fields of galaxies [Rybicki and Lightman, 1979]. Spiralling at relativistic speeds, the electrons travel along the magnetic field lines emitting photons whose frequency is dependent on the speed of the electron at the time of emission. Figure 2.12 shows how galactic synchrotron radiation is produced. The mechanisms behind the acceleration of these particles within the magnetic fields varies and can originate from physical processes within the AGN's supermassive black hole, the death of stars and supernovae remnants. Galactic synchrotron is non-thermal as the emitted radiation is independent of the temperature of its source. Since most galaxies have magnetic fields, galactic synchrotron radiation is found in all galaxies. Observing the spectra of galactic synchrotron emission, the flux variability exhibits a smooth dependence across frequency which is a sum of the flux from the individual photons. Galactic synchrotron can also be linearly polarized, which means that it can undergo Faraday rotation which changes the smooth spectral variation of polarized galactic synchrotron radiation creating greater complexity for foreground removal techniques in adequately retrieving the HI signal [Spinelli et al., 2018].

Free-free Emission: refers to the thermal Bremsstrahlung emission produced by free electrons scattering off ions (free-free). This type of emission originates from ionized hydrogen clouds (HI regions) which can be found within galaxies or have extragalactic origin [Alonso et al., 2014].

Point Sources: The term point source refers to isotropic radiation from sources that are located beyond our own galaxy. This type of radio wave emission can be either from Active Galactic Nuclei (AGN) or star forming galaxies. While star forming galaxies can be characterized into mainly two categories (i.e star bursts and quiescent galaxies), AGN are galaxies with high luminosity that are not only powered by starlight, but are also driven by an accretion of matter onto a super-massive black hole (SMBH) located in the cores of these types of galaxies. A study by Kormendy and Richstone [1995] showed that in

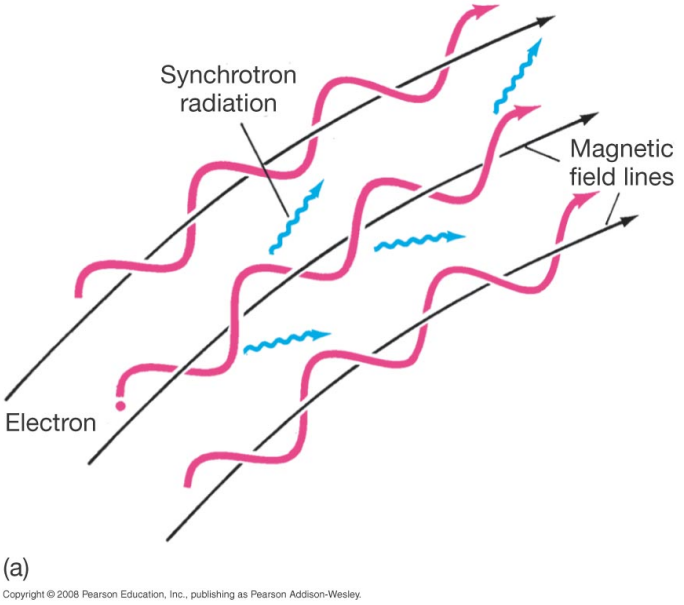

FIGURE 2.12: Galactic synchrotron radiation is produced by electrons spiralling within a magnetic field at relativistic speeds (From https://pages.uoregon.edu/jimbrau/astr123-2010/Notes/Exam1rev.html).

every massive galaxy, there exists a SMBH at its central core with typical masses ranging between $10^6 - 10^9$ $M_\odot$. AGN have higher flux variability in comparison to star forming galaxies such that they are observable across the wide range of the electromagnetic spectrum, with the most of the emission originating from the X-ray and ultraviolet wavelength range [Kubo et al., 1998].

Figure 2.13 shows the typical axial structure of AGN with an accretion disk, a rotating dusty torus containing material surrounding the SMBH, and jets speeding outwards on both sides that span hundreds of kiloparsecs. The narrow and broad line regions within the dusty torus are due to matter being excited by the radiation from the AGN's SMBH.

AGN can be classified into different classes with the main classes being; Quasars, Blazers, Seyfert Galaxies and Radio galaxies. The differences in these classes can be explained using the Unified AGN model (Antonucci [1993]; Urry and Padovani [1995]) which suggests that the different types of AGN have the same central core, and the differences observed arise from the viewing angle of the observer as shown on figure 2.13.

Looking at the frequency evolution of the foregrounds we have briefly summarized, figure 2.14 shows how the brightness temperature of the foregrounds discussed evolve as a function of frequency. In this figure, the amplitude and spectral variation of different types of foregrounds is compared to the HI signal. The complex spectral variation of the polarized

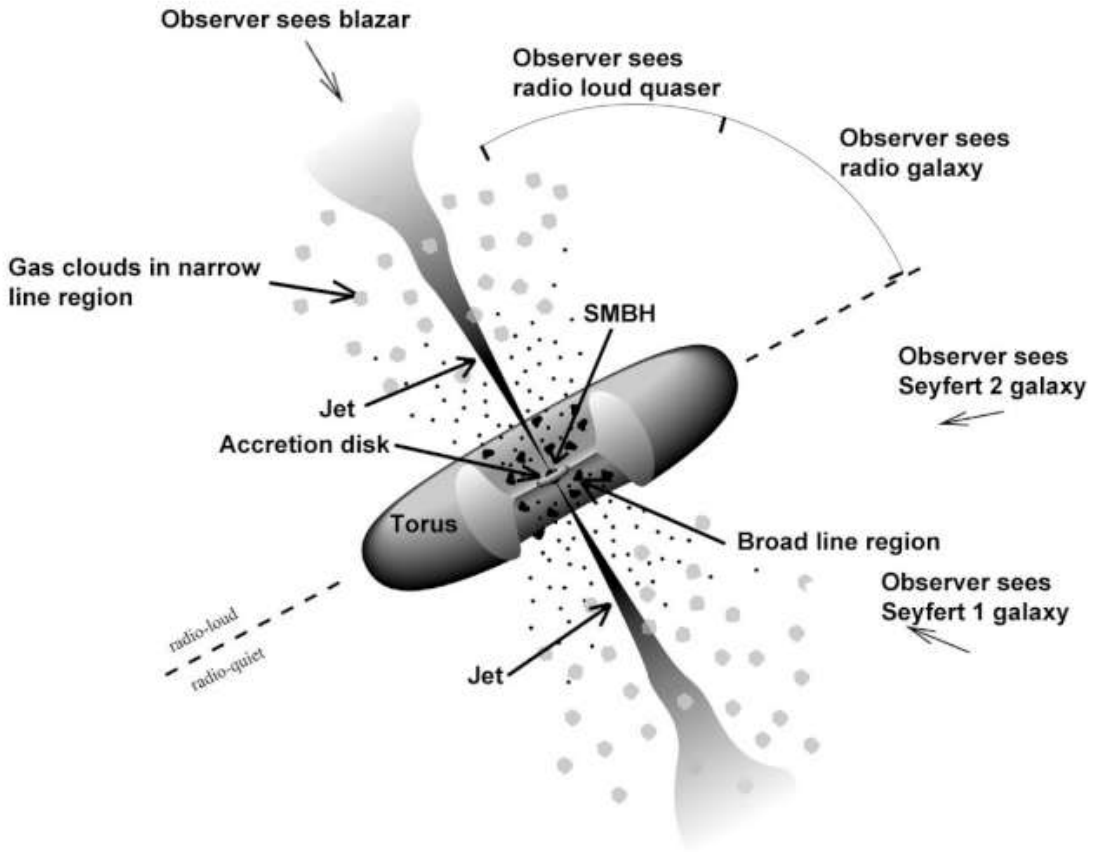

FIGURE 2.13: A simplified schematic showing the structure of a typical AGN with its various components. The figure also shows how the unified model of AGN can be used to differentiate between the various types of AGN. (From: https://fermi.gsfc.nasa.gov/science/eteu/agn/)

synchrotron emission is also shown in the *dashed red* curve, with its complexity increasing as one gets closer to the galactic plane. This complex structure in frequency can cause difficulties for effective foreground removal by the cleaning methods.

A discussion on the modelling of the components that make up the intensity of the radio sky is outlined in section 3.1 of the following chapter.

2.5 Foreground Subtraction Methods

Foreground subtraction (or foreground cleaning) refers to the process of removing foregrounds from the observed data in order to retrieve the cosmological HI signal. Foreground cleaning techniques can be broadly characterized into two categories, blind and non-blind. Blind if they use properties of foregrounds and non-blind when the methods require a detailed modeling of the spectral behaviour of the foregrounds as a function of frequency

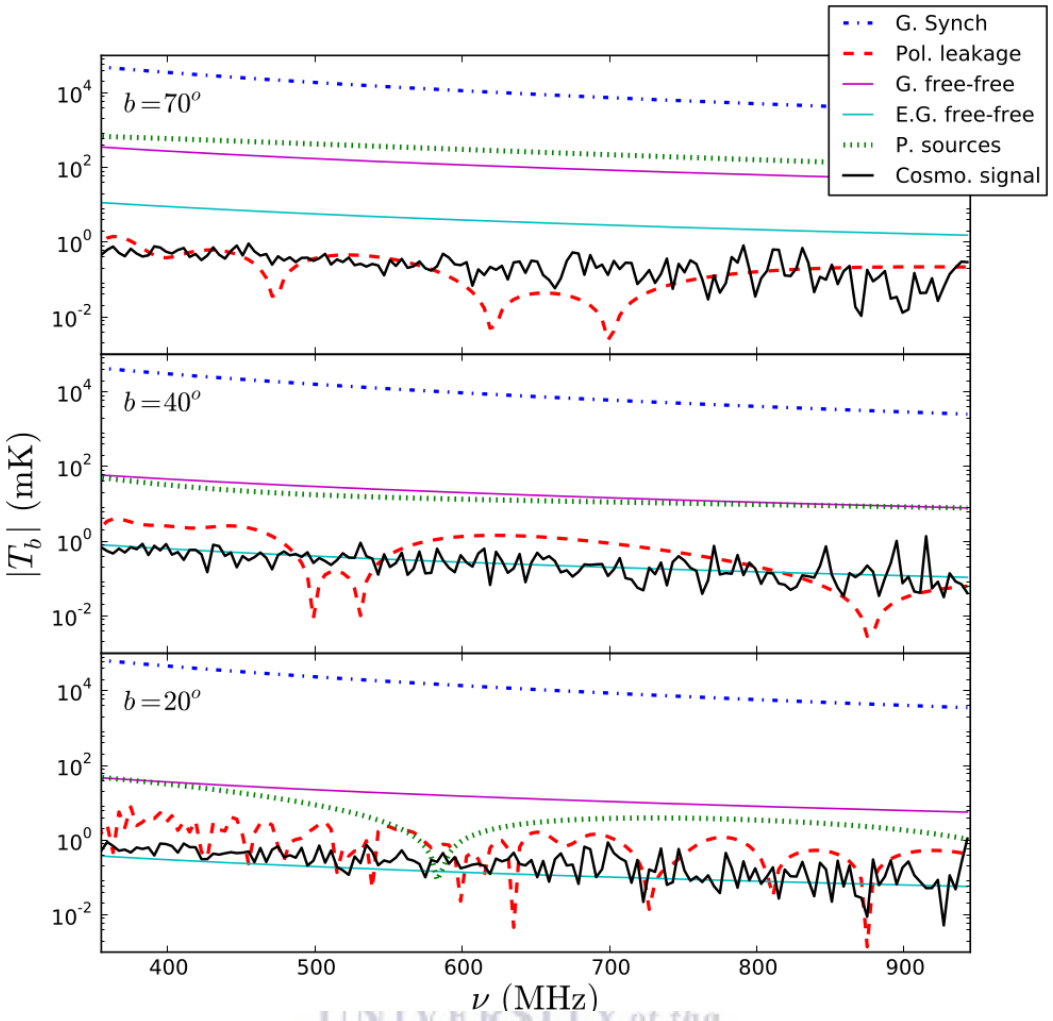

FIGURE 2.14: The frequency evolution of the different foregrounds and the HI signal from Alonso et al. [2014], at different galactic latitudes. The unpolarized foregrounds are smooth in frequency and greater in amplitude compared to the cosmological HI signal. Also polarized galactic synchrotron with 1 % leakage exhibits a much more complex spectral evolution due to being affected by Faraday rotation. This type of complexity in the spectral evolution may be a challenge to foreground removal techniques.

[Alonso et al., 2015]. There are a number of blind foreground techniques available. The focus of this section will be on three blind foreground subtraction techniques referred to in subsections 2.5.1, 2.5.3, and 2.5.2.

Foregrounds are expected to be highly correlated in frequency (smooth), contrary to the fluctuating 21 cm signal (and the instrumental noise). The underlying assumption to all foreground cleaning methods is that the foreground emission is very large in amplitude. For this work, we follow on the three blind foreground cleaning techniques that have been successfully applied in IM studies (Alonso et al. [2015], Carucci et al. [2020], Cunnington et al. [2019], Wolz et al. [2014, 2017]). We will use the Principal Component Analysis

(PCA), Fast Independent Component Analysis (FastICA) [Hyvarinen, 1999] and Logarithmic Polynomial Fitting algorithm (poLOG) (Wang et al. [2006]; Ghosh et al. [2011]; Ansari et al. [2012]; Wang et al. [2013]) to solve for the foreground components, using the code presented in [Alonso et al., 2015]. In chapter 4 (i.e. the results chapter), a single test case will be selected to demonstrate that PCA and ICA yield similar results, and from thereon the results will be presented with the PCA cleaning algorithm having been applied.

For the purposes of illustrating the basis of how the cleaning methods work, the sky temperature in equation 2.6 can be expanded in the following form for a single frequency map

$$T_{obs}(\nu, \hat{\mathbf{n}}) = \sum_{k=1}^{N_{fg}} f_k(\nu) S_k(\hat{\mathbf{n}}) + T_{HI}(\nu, \hat{\mathbf{n}}) + T_{noise}(\nu, \hat{\mathbf{n}}) \tag{2.9}$$

where the parameter N_{fg} represents the number of degrees of freedom to subtract, f_k represents the set of smooth basis functions in frequency to be used to model the foregrounds, $S_k(\hat{\mathbf{n}})$ are the foreground sky maps.

Given a set of N_ν frequencies and individual line of sight ($\hat{\mathbf{n}}$) measured discretely through this set of frequencies, equation 2.9 can be expressed as a linear system

$$\mathbf{x} = \hat{\mathbf{A}} \cdot \mathbf{s} + \mathbf{r}, \tag{2.10}$$

where $x_i = T(\nu_i, \hat{\mathbf{n}})$, $A_{ik} = f_k(\nu_i)$, $s_k = S_k(\hat{\mathbf{n}})$ and $r_i = T_{cosmo}(\nu_i, \hat{\mathbf{n}}) + T_{noise}(\nu_i, \hat{\mathbf{n}})$.

The three foreground cleaning techniques apply different methods of determining the mixing matrix $\mathbf{A}$ and foreground maps $\mathbf{s}$ in order to solve this linear system as accurately as possible such that,

$$\mathbf{r} = \mathbf{x} - \hat{\mathbf{A}}\mathbf{s}. \tag{2.11}$$

2.5.1 Principal Component Analysis (PCA)

The PCA cleaning method relies on taking the strongest $N_{\rm fg}$ eigenvalues of the frequency-frequency covariance matrix of the data, which should be coming from the foreground components. This number of foreground components, $N_{\rm fg}$, depends on the complexity of the problem and is a free parameter in the code. If $N_{\rm fg}$ is too small we will underclean and confuse the signal with residuals from foregrounds and instrumental effects. In

contrast, if $N_{\rm fg}$ is too large, we will start removing the actual 21cm signal, irreparably compromising its interpretation. Establishing the number of components to remove thus requires a careful study with simulations.

The PCA procedure starts from the data $\mathbf{x}$ where each x_i is a 2D map at a different frequency and flatten it in a vector $\mathbf{d}$ at each frequency. Then one computes the frequency-frequency covariance matrix of $\mathbf{d}$:

$$C_{ij} = \frac{1}{N_{\rm pix}} \sum_{p=1}^{N_{\rm pix}} d_i^{(p)} d_j^{(p)} \tag{2.12}$$

where we have made explicit the dependence over pixel p of the vector $\mathbf{d}$. The matrix $\mathbf{C}$ is symmetric and can thus be diagonalized by an orthogonal matrix $\mathbf{O}$:

$$\mathbf{O}^T \mathbf{C} \mathbf{O} = \mathbf{\Lambda} = diag(\lambda_1, \lambda_2, .., \lambda_{\rm N_\nu}), \tag{2.13}$$

where the columns of $\mathbf{O}$ are the eigenvectors of $\mathbf{C}$ and the λ_i are the corresponding eigenvalues, which can be ordered such that $\lambda_i > \lambda_{i+1}$. The fact that the foreground emission is smooth means that it is strongly correlated in frequency; the strongly correlated component of the data concentrates in the first eigenvalues of $\mathbf{C}$. The PCA method retains the first $N_{\rm fg}$ ones assuming they are a good enough description of the data. We call $\mathbf{O}_{\rm fg}$ the matrix composed by the first $N_{\rm fg}$ eigenvectors of $\mathbf{C}$ and we compute the projection of the selected eigenvectors along the data, obtaining the eigen-sources

$$\mathbf{S} = (\mathbf{O}_{\rm fg} {\mathbf{O}_{\rm fg}}^{\rm T})^{-1} {\mathbf{O}_{\rm fg}}^{\rm T} \mathbf{d} \tag{2.14}$$

The matrix $\mathbf{S}$ will have shape (3,$N_{\rm pix}$). Once $\mathbf{S}$ is available, we can obtain the estimation of the foregrounds as

$$\mathbf{d}_{\rm fg} = \mathbf{O}_{\rm fg} \mathbf{S} \tag{2.15}$$

that has now dimension (N_ν,$N_{\rm pix}$). The last step to obtain a clean HI cube is to subtract this description of the data $\mathbf{d}_{\rm fg}$ from the original data $\mathbf{d}$ and reshape it as N_ν 2D maps.

A reasonable first guess for the number of component to describe the foreground could be $N_{\rm fg} = 4$ as there are four different foregrounds (galactic synchrotron, point sources and galactic and extra-galactic free-free), but it is possible that a smaller/higher amount of components would be needed. The mean HI temperature and the effect of the beam are

also smooth functions of the frequency, and they will be inevitably play a role in defying the optimal number. Once $N_{\rm fg}$ is set, the 21cm signal (and the noise) we are looking for, will be contained in the residual maps after PCA is applied. Note that PCA works well in the simplified case of uncorrelated noise. More complicated analysis are required to deal with real data [Masui et al., 2013, Switzer et al., 2013].

2.5.2 *Fast* Independent Component Analysis (*Fast*ICA)

In a nutshell, FastICA is an efficient implementation of the more general Independent component analysis (ICA) statistical method, whose core assumption is that the $N_{\rm fg}$ components s_i are as statistically independent from each other as possible. FastICA formulates the ICA problem as a maximization of the *negentropy* of the transformed variables s_i. Given a random vector $\mathbf{y} = (y_1, y_2, .., y_n)^T$ with density $f(.)$, a differential entropy can be defined as $H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}$. For variables with the same variance, the entropy H is maximal for Gaussian variables. From the concept of entropy, one can further define the *negentropy* $J(y) = H(y_{\rm gauss}) - H(y)$, where $y_{\rm gauss}$ is a Gaussian random variable of the same covariance matrix as $\mathbf{y}$. Note that, for what is stated above on the entropy of Gaussian variables, the negentropy J will be always positive for non-Gaussian variables, and thus a measure of non-gaussianity. In our case our variable of interest are the components $s_i = \mathbf{w}^T \mathbf{d}$, where $\mathbf{w}$ are the row of the matrix $\mathbf{W}$ that relates the sources and the data. Moreover, instead of using the exact definition of negentropy, the FastICA algorithm makes use of approximations based on the maximum entropy principle, leading to the maximization of

$$J_G(\mathbf{w}) \simeq c[E[G(\mathbf{w}^T \mathbf{d})] - E[G(\nu)]]^2, \tag{2.16}$$

where the expectation value $E[.]$ is over the $N_{\rm pix}$ unmasked pixels in the map, $\nu \sim N(0, 1)$, c is a constant and G any non-quadratic function as discussed in Hyvarinen [1999].

This procedure allow the reconstruction of all the rows of the matrix $\mathbf{W}$ and thus to the estimation of $\mathbf{x}$. The result will depend on the user define variable $N_{\rm fg}$ that will not necessary coincide with the actual number of physical foregrounds. For our purposes this is not much of a problem, since we just need FastICA to efficiently remove the smooth components, and not to precisely reconstruct the different foregrounds. Indeed as for PCA, what we are interested in is the residual maps, that will be an estimation of the

21 cm signal we are looking for plus the thermal noise, that, having been simulated as Gaussian, will not be included in the independent component that FastICA is looking for. The noise will therefore need to be further subtracted from the final residual maps. Moreover, since the mean HI temperature is a smooth function of the frequency, it will be inevitably removed by the foreground cleaning method.

2.5.3 Logarithmic Polynomial Fitting (poLOG)

The poLOG method is one of the simplest technique that is used to conduct foreground subtraction. This method assumes that the mixing matrix A, takes a special form; that it contains a set of smooth polynomials in frequency f_k. The number of these smooth polynomials used for the fitting depends on the number of foreground contributions that are present in the maps. Equation 2.17 shows how the fitting, occurs in log-log space for each line of sight

$$\log(T(\nu, \hat{\mathbf{n}})) = \sum_{k=1}^{N_{fg}} \alpha_k(\hat{\mathbf{n}}) f_k(\log(\nu)), \tag{2.17}$$

where $f_k(\log(\nu)) = [\log(\nu)]^{k-1}$ are the set of smooth polynomials of the logarithm of the frequency.

To determine the foreground maps $\mathbf{s}$, a χ^2 minimization is performed on equation 2.11 in the following way:

$$\chi^2 = (\mathbf{x} - \hat{\mathbf{A}} \cdot \mathbf{s})^T \mathbf{N}^{-1} (\mathbf{x} - \hat{\mathbf{A}} \cdot \mathbf{s}), \tag{2.18}$$

where $\hat{\mathbf{N}}$ represents the covariance matrix of the cosmological HI signal and instrumental noise $\mathbf{r}$. The solution to the χ^2 minimization of equation 2.18 is

$$\mathbf{s} = (\hat{\mathbf{A}}^T \hat{\mathbf{N}}^{-1} \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{N}}^{-1} \mathbf{x}. \tag{2.19}$$

Despite the method being simple, there are cavaets to the polynomial fitting method, applied in this analysis. For instance, how the different frequency channels correlate has been ignored and to account for this neglect, a diagonal covariance $N_{ij} = \sigma_i^2 \delta_{ij}$ is assumed and $\sigma = \sqrt{\sigma^2_{noise,i} + \sigma^2_{cosmo,i}}$ is the joint variance from the cosmological signal and the noise. For a thorough discussion of the assumptions and limitations of this method see Alonso et al. [2015].

2.5.4 Other Foreground Cleaning Methods

Beside the three methods described above, some other methods have been explored in the literature for the case of post-reionisation HI intensity mapping. In this section we briefly name some of them. An interesting variation of the standard PCA is Kernel Principal Component Analysis, where some information from the spatial shape of the maps is included. The method has been applied to HI IM in [Irfan and Bull, 2021] who found results similar to the standard PCA analysis. Interestingly, their results suggest that a pre-cleaning done with PCA with a low $N_{\rm fg}$ improves the performances of KPCA. A method previously applied to EoR studies that performs well for our case is Generalised Morphological Component Analysis (GMCA). This foreground subtraction method, assumes sparsity in the wavelet domain and diversity in morphology [Carucci et al., 2020]. Another method inherited from EoR is Gaussian Process Regression, see Soares et al. [2022] for post reionization HI IM application. In Spinelli et al. [2022], the interested reader can find a detailed comparison of the performances of some of these methods.

2.6 Cross-correlation with Galaxy Surveys

Some of the advances that have been made in the quest to make a tentative detection of the HI signal, have occurred through combining two data sets which measure the large scale structure such as galaxy surveys (optical or near-infrared) and HI IM surveys (i.e. cross-correlating different surveys which probe the same region of sky). One of the advantages of cross correlations is that systematics that affect the HI IM technique may not necessarily be a problem for galaxy surveys and vice versa. For example, systematics from optical surveys such as photometric calibration, galactic extinction, and contamination from stellar sources, are expected to be uncorrelated with systematics from the radio such as thermal noise, RFI noise and foregrounds. Secondly by cross-correlating the signal to noise (S/N) ratio increases, which then reduces the error in the measured signal [Cunnington, 2019]. The less similar the two surveys are in terms of instrumental and survey design, the less correlation of the systematics with the required detection will occur. Let us consider the following example to illustrate the benefits of using cross-correlations (following similar work as described in Kitching et al. [2015]).

The observed signal by a telescope can be broadly characterized into two main components, the measured signal and the systematics. The systematics are additive and arise from the instrument and the sky, which then contaminate the true signal. Below we illustrate how best the observed signal can be obtained by going through the two methods in which observations can be conducted, in auto-correlation and cross-correlation. For our purposes we choose large scale structure observations conducted with radio and optical galaxy surveys.

For the radio telescope, the observed signal $\mathbf{X}_{rad}$ is;

$$\mathbf{X}_{rad} = \mathbf{S}_{rad} + \mathbf{N}_{rad}, \tag{2.20}$$

where $\mathbf{S}_{rad}$ is the HI 21 cm signal and $\mathbf{N}_{rad}$ is all systematics relevant in the radio regime (e.g. instrumental noise, foregrounds, RFI, etc). A similar equation can be constructed for an observation $\mathbf{X}_{opt}$ conducted by an optical telescope,

$$\mathbf{X}_{opt} = \mathbf{S}_{opt} + \mathbf{N}_{opt}, \tag{2.21}$$

where $\mathbf{S}_{opt}$ is the photons from galaxies and $\mathbf{N}_{opt}$ is the relevant systematics in optical galaxy surveys such as contamination from stars, photometric calibration and galactic extinction.

To appreciate why cross-correlation of radio and optical galaxy survey works, let us first consider an auto-correlation case. In this observation mode the $\langle \mathbf{S}_{opt}\mathbf{N}_{opt} \rangle$ cross terms do not correlate and therefore fall away,

$$\langle \mathbf{X}_{opt}\mathbf{X}_{opt} \rangle = \langle \mathbf{S}_{opt}\mathbf{S}_{opt} \rangle + 2\underline{\langle \mathbf{S}_{opt}\mathbf{N}_{opt} \rangle} + \langle \mathbf{N}_{opt}\mathbf{N}_{opt} \rangle. \tag{2.22}$$

What remains is the systematic term $\langle \mathbf{N}_{opt}\mathbf{N}_{opt} \rangle$ which correlate, and therefore contaminating the measured signal as shown in equation 2.23. This essentially means that we require systematic/foreground subtraction methods in order to retrieve the signal:

$$\langle \mathbf{X}_{opt}\mathbf{X}_{opt} \rangle = \langle \mathbf{S}_{opt}\mathbf{S}_{opt} \rangle + \langle \mathbf{N}_{opt}\mathbf{N}_{opt} \rangle, \tag{2.23}$$

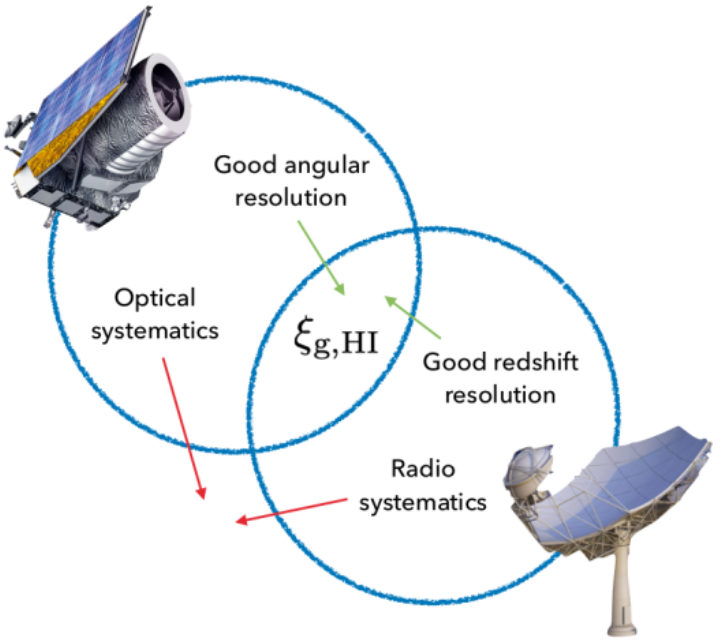

FIGURE 2.15: There is an increased potential of a detection of an HI signal when cross-correlations are conducted between HI 21 cm intensity mapping and optical galaxies [Courtesy: Steve Cunnington].

With cross-correlation, the final measured signal from two different surveys will be the following:

$$\langle \mathbf{X}_{opt}\mathbf{X}_{rad}\rangle = \langle \mathbf{S}_{opt}\mathbf{R}_{rad}\rangle + \underline{\langle \mathbf{S}_{opt}\mathbf{N}_{rad}\rangle} + \underline{\langle \mathbf{S}_{rad}\mathbf{N}_{opt}\rangle} + \underline{\langle \mathbf{N}_{opt}\mathbf{N}_{rad}\rangle}. \tag{2.24}$$

Since the $\langle \mathbf{S}_{opt}\mathbf{N}_{rad}\rangle$, $\langle \mathbf{S}_{rad}\mathbf{N}_{opt}\rangle$ and $\langle \mathbf{N}_{opt}\mathbf{N}_{rad}\rangle$ are uncorrelated, what remains is the cross-correlated signal. Figure 2.15 is an illustration to summarise the advantages of conducting cross correlation measurements where good angular and redshift resolution is achieved and systematics do not correlate and therefore fall away.

Now that the cross-correlation method has been motivated, in the next section we briefly describe some of the detections of the HI signal obtained using this method.

2.7 HI Power Spectrum Detection via Cross-correlation with Galaxy Surveys

The cross-correlation method presented in section 2.6 has been applied in various studies within the HI IM community. In this section, we briefly review some of the studies in

which a HI signal detection was made from cross-correlating data from HI IM and galaxy surveys.

As mentioned in section 2.3.4, the first statistically significant measurement of the HI signal was from cross-correlating data from two galaxy surveys, the HI Parkes All Sky Survey[12] (HIPASS) survey, a HI galaxy survey by the Parkes telescope, and the Six Degree Field Redshift Survey[13] (6dFGS) [Jones et al., 2005], a spectroscopic survey from the Anglo-Australian Telescope[14](AAT). Although in this study by Pen et al. [2009], a direct HI IM survey was not conducted, as it used intensity data from the 21 cm spectral line of HI emitting galaxies instead of the observed area being pixelated and measuring the integrated emission coming from individual pixels. However, by measuring line of sight correlations present in the combined data, a signal could be detected. This result essentially provided early signs that neutral HI emission traces biasedly the large-scale structure as it followed the distribution of the galaxies present on the large-scale structure.

The first proper detection of the HI signal with the HI IM technique came from cross correlating HI IM survey data from the GBT telescope with the DEEP2 optical galaxy redshift survey [Davis et al., 2003] in a study by Chang et al. [2010]. The main scientific goals of this study was to investigate the potential of the HI IM technique to map the cosmic web in 3-D using the redshifted HI emission. The second aspect of this study was to make a high redshift measurement of the 21 cm brightness temperature and density of neutral gas.

To achieve these goals, the GBT measured the redshifted HI emission over a redshift range of $0.53 < z < 1.12$. The DEEP 2 survey is an optical redshift galaxy survey whose aim was to sample galaxies located at high redshifts ($z \sim 1$). Using the Keck II telescope's DEIMOS spectrograph, 50000 optical galaxy redshifts were measured in four survey areas ($\sim 120' \times 30'$). About 10000 of these faint galaxies are located on the same volume of sky probed by the GBT telescope. From the cross-correlation of both the HI IM and optical galaxy data, a 4.0σ detection of the HI 21 cm was made.

Figure 2.16 is the cross-correlation power spectrum computed in this work, where the *crosses* represent measured cross-correlated brightness temperature with 1σ error bars. To verify the measurement, a statistical null test was conducted in which the optical data

[12] https://www.atnf.csiro.au/research/multibeam/release/
[13] http://www.6dfgs.net/
[14] https://aat.anu.edu.au/about-us/AAT

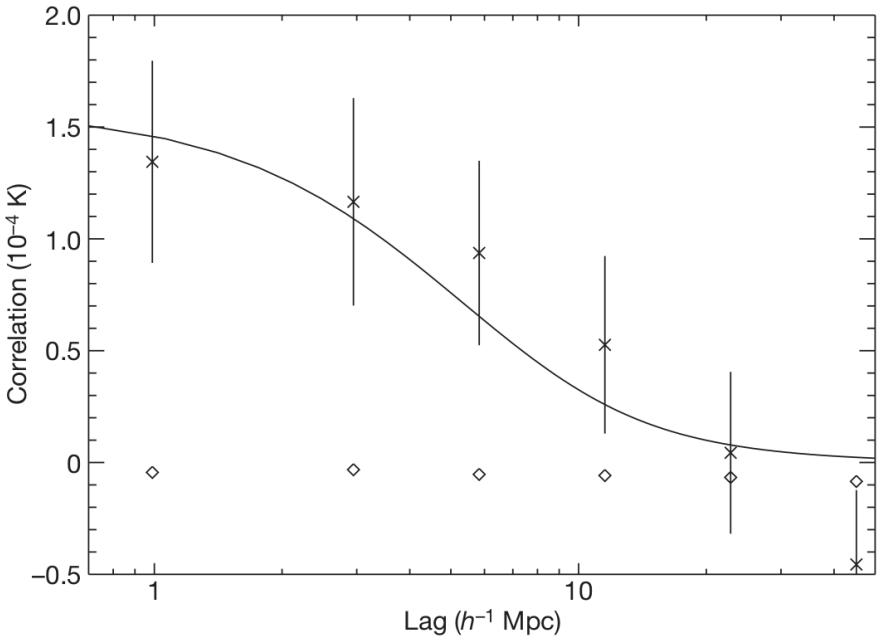

FIGURE 2.16: The measured GBT X DEEP2 cross-correlation power shown by the crosses. The associated error bars where obtained from randomised data from the optical data via the bootstrapping method. The *diamond* represent the null test performed as a verification of the measurement by repeatedly randomizing the optical maps many times and with each repetition re-computing the cross-correlation measurement. The *solid* curve is a predictive model using the auto-correlation function from DEEP2 optical galaxies [Chang et al., 2010].

was randomized repeatedly and with each repeating a re-computation of the correlation calculation was done. The outcome of this test is shown by the *diamond* points and is a significant result as it illustrates that the systematics such as astrophysical continuum sources and residual RFI are least likely to contribute to a false detection of the HI 21 cm signal. The solid curve is the DEEP 2 galaxy auto-correlation function and is used as a comparative model to the measured data points.

Soon after that a number of surveys have been embarked on in the quest to measure the HI signal in cross-correlation such as the GBT-WiggleZ DES result by Masui et al. [2013] which measured a 7.4σ correlation at a redshift of $z = 0.8$ by combining both WiggleZ 1hr and 15hr fields. From this detection, the amplitude of the power spectrum which is a product of HI density fraction $\Omega_{\rm HI}$, the galaxy-HI cross-correlation coefficient $r_{\rm HI,Wig}$ and linear HI bias $b_{\rm HI}$ were constrained to $\Omega_{\rm HI} b_{\rm HI} r_{\rm HI,Wig} = [4.3 \pm 1.1] \times 10^{-4}$.

A low redshift study was conducted by Anderson et al. [2018] in which the Parkes radio telescope was used to conduct an HI IM survey on an area spanning 1300 square degrees. The observed HI IM data was cross-correlated with the 2dF survey [Colless, 1999] in order

to detect a cross correlated HI signal at $z \simeq 1$. However in this study, a low-amplitude level of clustering in the cross-correlated data was found.

One of the most recent examples of the HI 21 cm signal detections with cross-correlation was conducted by Wolz et al. [2021]. In this study, the HI IM observations conducted by the GBT radio telescope were cross-correlated with the following three data samples: Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs) from the extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Prakash et al. [2016]; Dawson et al. [2016]) and WiggleZ sample from the DES survey. In terms of the HI survey, the GBT observed 100 square degrees of sky, on an area overlapping the 1hr WiggleZ field. The telescope's receivers were tuned to detect HI emission within a frequency band of $700 - 900$ MHz which is equivalent to a redshift range $0.6 < z < 1.0$. The observed HI time ordered data was split into four seasons. Across these seasons, the same coverage and integration depth were kept. This ensured that the thermal noise is uncorrelated.

Prior to cross correlating the HI IM maps, foreground subtraction was performed using the FastICA method described in section 2.5.2, in order to remove the unwanted foreground emission present in the data. As a way to verify that FastICA was working effectively, the foreground subtraction method was also tested on mock observations of the HI intensity sky. In the event that there are effects such as signal loss introduced by the foreground subtraction, a transfer function is applied in order to correct for the effects. The HI power spectrum of the foreground subtracted HI data is computed in auto-correlation and also cross-correlation of HI data in between two seasons at a time. From this spectra, the scale dependence and overall amplitude are evaluated.

Figure 2.17 shows GBT HI IM power spectra in both auto-correlation and cross-correlation of HI IM survey data for different seasons, after foreground subtraction. Auto-correlation data points are shown in *crosses* whilst cross-correlations data points with *solid circles* with error bars, both types depicted at different levels of cleaning indicated by IC. From this figure, there is a clear difference that can be observed, all the auto correlation spectra have higher amplitude compared to the cross correlation spectra due to systematics correlating with the HI signal. Despite the contamination from systematics, the auto-correlation signal serves as an upper limit of this type of measurement. By cross-correlating the HI data from different seasons, the contribution from systematics can be reduced resulting in an overall lower amplitude in the power spectrum. Even though this

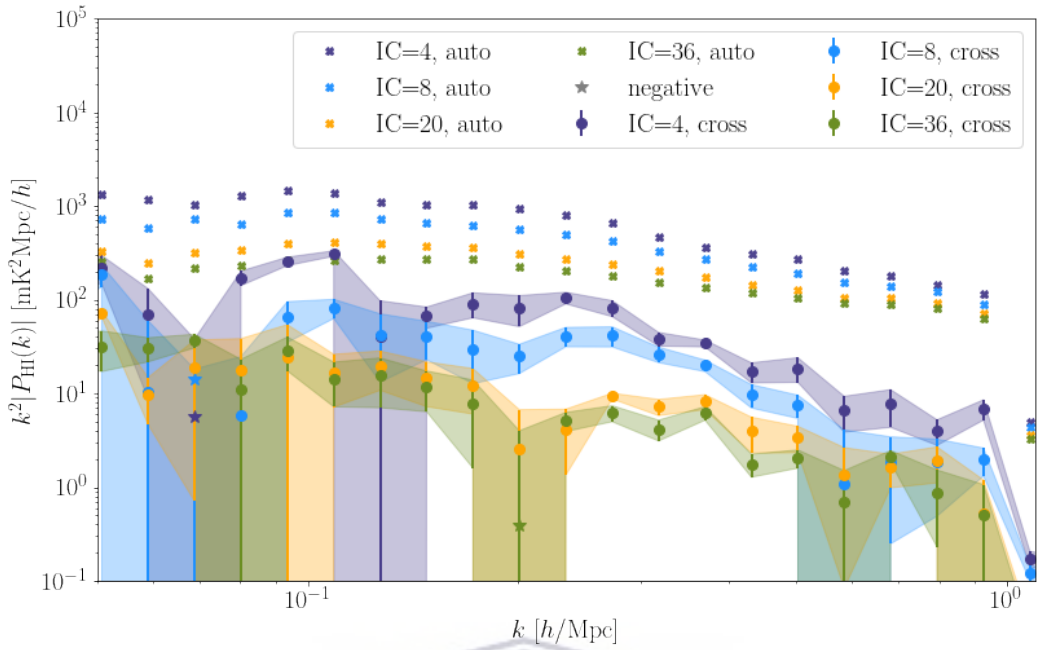

FIGURE 2.17: The foreground cleaned absolute HI power spectrum computed from GBT HI IM observations by Wolz et al. [2021]. Indicated with *crosses* are the auto-correlation spectra and in the *solid circles* are the cross-correlations of the HI data across different seasons shown for different number of components removed (IC = 4, 8, 20, 36). The added advantage provided by cross-correlating the two sets of cleaned HI data from different seasons is the limitation of the contribution from systematics as the thermal noise from each season are uncorrelated. This is shown by the overall lower amplitude of the power compared to the auto-correlation spectra that still has higher levels of systematics within the signal.

is still far from the theoretical HI auto-correlation power spectra expected, it is improvement from the HI auto we seek to observe that is contaminated. In order to observe the auto-correlation power spectrum efforts must be made to effectively deal with systematics affecting its measurement.

Shown on figure 2.18 is one of the GBT cross-correlation power spectra with one of the samples, the Wiggle Z dataset from the study by Wolz et al. [2021]. The HI-galaxy power spectra is shown after foreground cleaning and when the appropriate correction for signal loss has been applied using the transfer function. The data points are presented staggered for illustration purposes in order to highlight the differences introduced when the number of ICs is increased.

The most recent HI power spectrum detection came from cross-correlating data from the MeerKAT Pilot survey from Wang et al. [2021] (see also section 2.3.3) and galaxy data [Drinkwater et al., 2010, Drinkwater et al., 2018] from WiggleZ survey in a study by

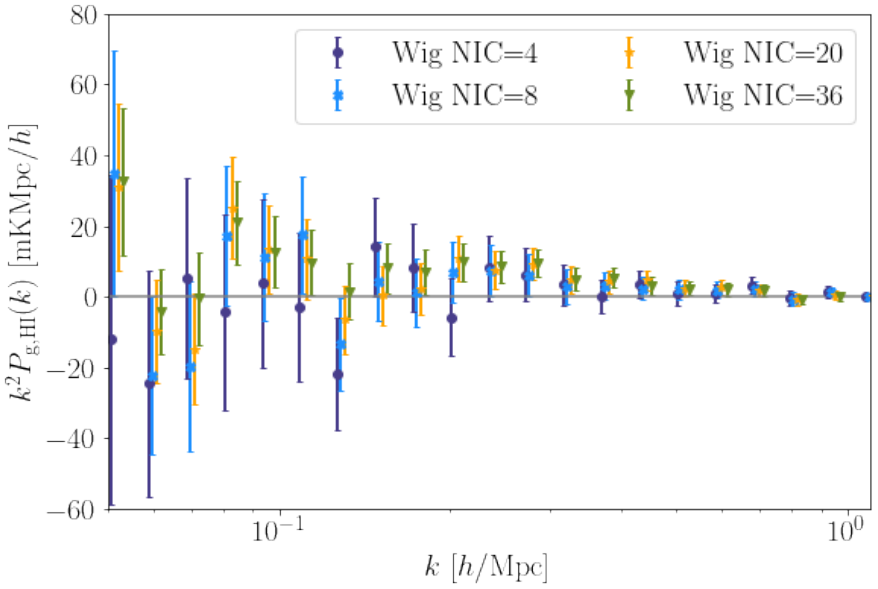

FIGURE 2.18: The cross-correlation HI power spectrum result of the GBT with the WiggleZ galaxy sample. The power spectra are shown after foreground cleaning with varying the number of independent components (ICs) removed and signal loss correction. For illustration purposes each the data points for each spectra appears staggered for each scale value k. This is to show differences in the spectra when IC's are increased. For the other two cross-correlation results using the ELG and LRG samples see Wolz et al. [2021].

Cunnington et al. [2022]. In order to make this detection, the MeerKAT radio telescope was used in single dish mode to scan 200 square degrees of sky with its L-band receiver. The observed sky area overlapped with the 11.5 hr WiggleZ field, resulting in 10.5 hrs of observational data over a frequency band of 937-1015 MHz (equivalent to a redshift range of $0.400 < z < 0.459$. The top panel of figure 2.19 shows the measured power spectrum shown with the blue data points with 1σ error bars computed. A 7.7σ HI signal detection was reported. A series of null tests have been performed in order to validate the result. The lower panel of figure 2.19 shows how nothing is detected if the HI maps or the WiggleZ maps are shuffled in redshift.

This HI power spectrum detection has an amplitude that is directly proportional to the product of the HI density fraction (Ω_{HI}), the HI bias (b_{HI}), and the cross-correlation coefficient (r). Figure 2.20 shows the resulting constrain on $\Omega_{\mathrm{HI}} b_{HI} r$ as a function of the effective scale considered in the power spectrum. The result depends on the number of component subtracted but shows consistent trends.

The main significance of this study is that it showed that for the first time, the MeerKAT radio telescope array can be used in single dish mode to measure the cross-correlated HI

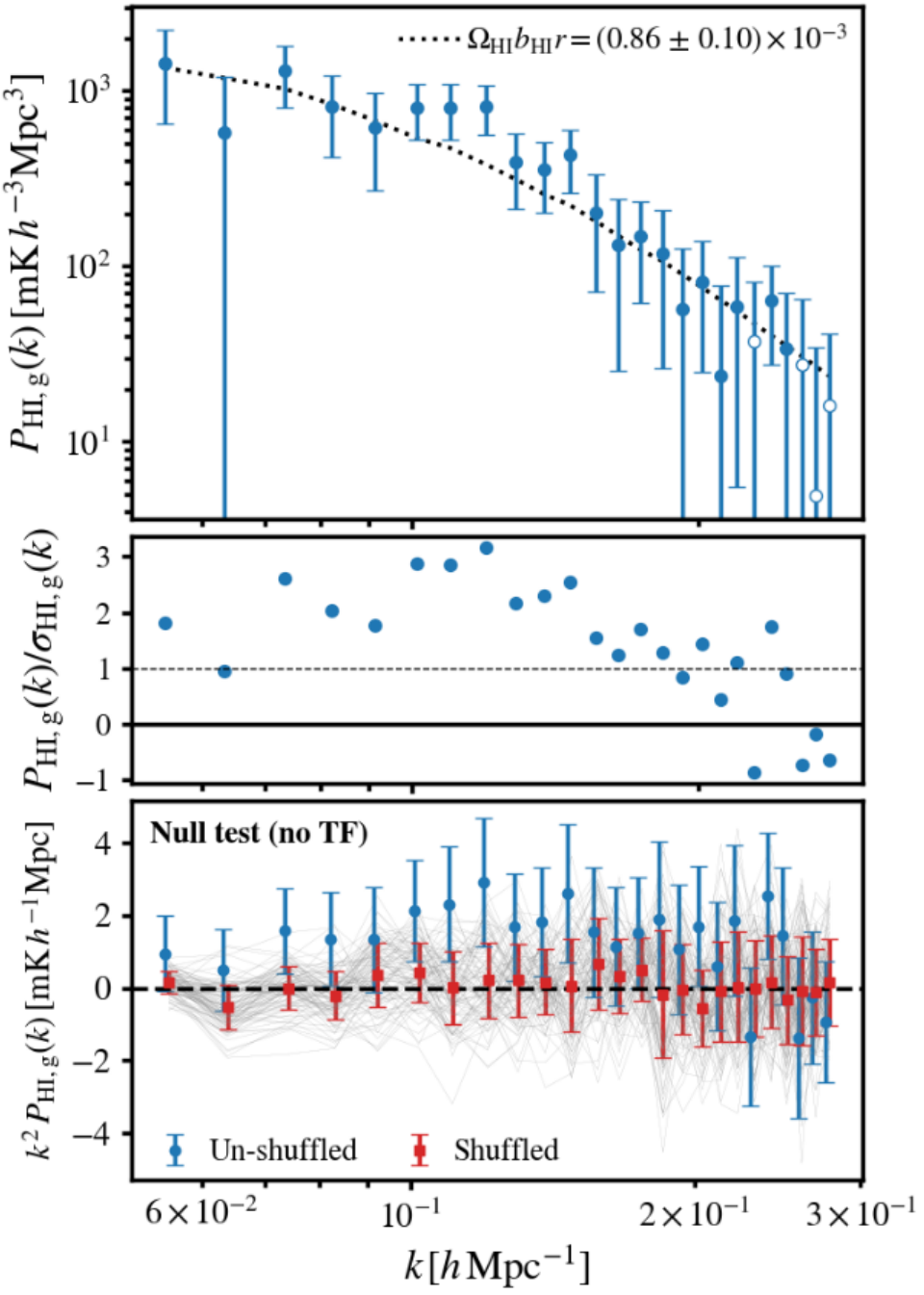

FIGURE 2.19: Shown on the top panel is the power spectrum detected from cross-correlating the HI IM maps from the MeerKAT Pilot Survey data with WiggleZ 11hr field from DES shown by the solid *blue* points with 1σ error bars. To obtain this result $N_{fg} = 30$ PCA modes had to be removed over a redshift range between $0.400 < z < 0.459$ [Cunnington et al., 2022]

signal.

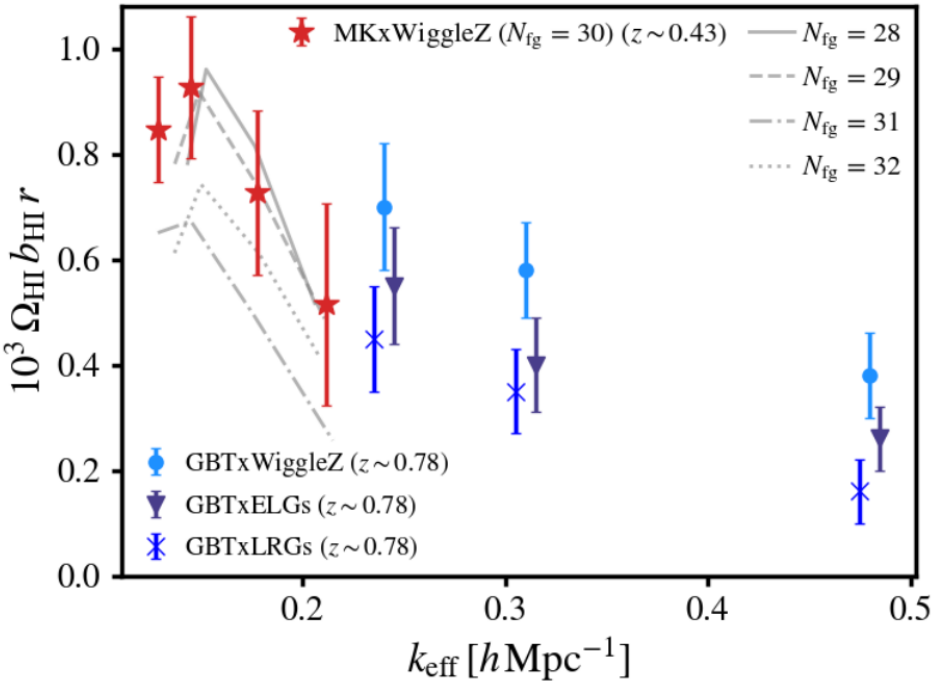

FIGURE 2.20: HI cross-correlation constraints on $\Omega_{HI} b_{HI} r$ [Cunnington et al., 2022].

2.8 Summary

In this chapter we have considered the role of neutral hydrogen as a probe for the large scale structure of the Universe citing its relevance at different epochs during its evolution, particularly in the post reionization era. We have discussed a new observational technique HI IM, how it will assist in the quest to map the 3-D distribution of HI, its limitations and use of the technique by instruments such as the MeerKAT through the MeerKLASS project and SKAO-MID. We have also explored other radio astronomy telescopes involved in HI IM research and briefly described their instrument setup and science goals.

A review of some of the challenges facing HI IM such as the presence of foregrounds that contaminate the cosmological HI signal was conducted and a summary of the some of the widely used blind foreground removal methods was provided.

Apart from the presence of foregrounds, there are also radio telescope systematics that hamper the reconstruction of the HI signal. Cross-correlations of HI IM data with optical galaxy survey data has been found to be one of the methods that are critical in being able to reduce the contribution from systematics and obtain a measurement of the HI signal so far and examples of HI IM results in cross-correlation were provided.

In this chapter we have seen that there is yet to be a detection of the HI 21 cm using single dish radio telescopes due to the number of challenges present, however we have also observed evidence for a detection of the HI signal via cross-correlations. The rest of this thesis is dedicated to investigating some of these systematics further for the single dish case, by using simulations of HI IM observations in single dish mode and that take into account the problematic foregrounds and instrumental effects, all in the quest to measure the HI signal in auto-correlation.

Chapter 3

Simulating Single-dish HI IM Observations

In this chapter we provide details of how we simulate a single dish HI IM observation using the Spherical Harmonic Transforms (SHT) beam convolution approach and briefly describe how the various scientific packages are used to create the required components of the simulation pipeline. This part of the work is an original contribution that resulted in a publication [Matshawule et al., 2021].

3.1 Sky Model

In this section we describe the various components of our sky model: the 21 cm signal (section 3.1.1), the Galactic and extra-Galactic Free-free emission (section 3.1.2), the Galactic synchrotron emission (section 3.1.3), and a new simulated Point Source catalogue (section 3.1.4). The latter is a central ingredient to the analysis presented in this work. Indeed, point source emission can have a non-trivial impact on the cleaning when the standard assumption of a Gaussian beam is relaxed. The other components are obtained using the publicly available code: Cosmological Realizations for Intensity Mapping Experiments (CRIME[1]), presented in Alonso et al. [2014], where the interested reader can find more details. In the following we just give a brief overview. All the components of the sky model are given in full-sky HEALPix maps [Górski et al., 2005] at selected

[1] http://intensitymapping.physics.ox.ac.uk/CRIME.html

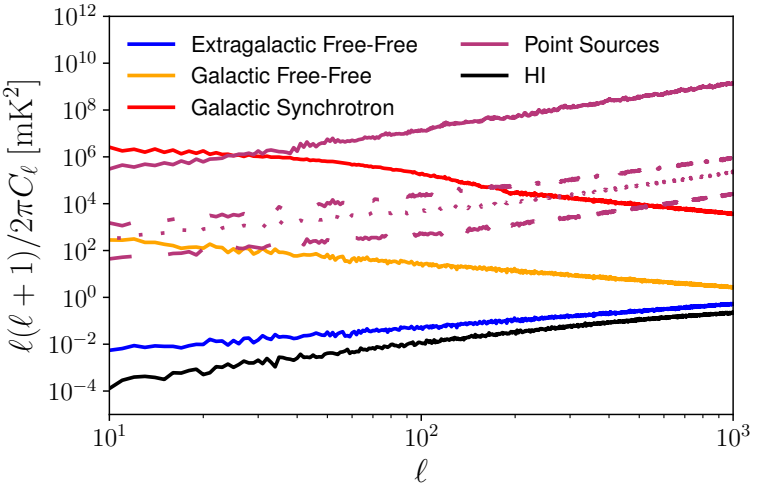

FIGURE 3.1: Angular power spectrum of the HI signal and the foregrounds at 950 MHz. The point source C_ℓ are shown for different flux cuts: full catalogue (solid line), 10 Jy cut (dashed-dotted line), 1 Jy cut (dotted line) and 100 mJy cut (dashed line).

frequencies and then combined and masked to mimic an IM survey with MeerKAT. We will discuss the simulated survey details later in section 3.3.

3.1.1 21 cm HI Signal

To model the 21 cm signal we need to simulate the brightness temperature of the neutral hydrogen emission in every position on the sky and at every frequency of interest. There is a one to one relation between the observed frequency, ν, and the redshift z of the emission: $\nu = \nu_{21}/(1+z)$, where ν_{21} is the frequency of the 21 cm line in its rest frame, ≈ 1420 MHz. The value of the 21 cm brightness temperature can be related to the underlying neutral hydrogen density $\rho_{\rm HI}(\hat{n}, z)$ through [Furlanetto et al., 2006]

$$T_{21\rm cm}(\hat{n}, z) = 0.19 \frac{\Omega_b h (1+z)^2}{\sqrt{\Omega_m (1+z)^3 + \Omega_\Lambda}} x_{\rm HI}(z) \frac{\rho_{\rm HI}(\hat{n}, z)}{\bar{\rho}_{\rm HI}(z)} \ {\rm K}, \tag{3.1}$$

where Ω_b, Ω_m, Ω_Λ are the baryon, total matter and dark energy density fractions respectively, $x_{\rm HI}(z)$ is the neutral hydrogen mass fraction (with respect to baryons) and $\bar{\rho}_{\rm HI}(z)$ is the mean HI density at redshift z.

As already mentioned, the 21 cm signal is simulated using the CRIME code. The code assumes the Planck Collaboration [2014] best fit cosmology and starts from a log-normal dark matter simulation on a cubic grid. The dark matter box is then divided in spherical shells, which are in turn pixelized to yield 21 cm maps. A temperature is associated

TABLE 3.1: Foreground $C_\ell(\nu_1, \nu_2)$ model parameters for the pivot values $\ell_{\rm ref} = 1000$ and $\nu_{\rm ref} = 130$ MHz [Santos et al., 2005]

Foreground	A [mK^2]	β	α	ξ
Galactic free-free	0.088	3	2.15	35
Extragalactic free-free	0.014	1	2.1	35
Galactic synchrotron	700	2.4	2.8	4

to each pixel considering the hydrogen density enclosed within it through equation 3.1. Redshift distortions are also introduced using the velocity field. The evolution of the neutral hydrogen fraction is assumed to be $x_{\rm HI}(z) = 0.008(1+z)$, in agreement with the trend present in existing data [Crighton et al., 2015]. An example of the angular clustering of the signal is presented in figure 3.1.

3.1.2 Galactic and Extragalactic Free-free Emission

In the simulations we use here, from Alonso et al. [2014], statistical isotropy is assumed. Such premise can break down, especially for the Galactic emission, which can be problematic if the fluctuations across the sky are strong, due to the convolution with the primary beam. However, we expect this signal to be a few orders of magnitude below the galactic synchrotron and therefore have a subdominant effect. Full-sky free-free maps are therefore generated, following Santos et al. [2005], as a Gaussian realization of

$$C_\ell(\nu_1, \nu_2) = A \left(\frac{\ell_{\rm ref}}{\ell}\right)^\beta \left(\frac{\nu_{\rm ref}^2}{\nu_1 \nu_2}\right)^\alpha \exp\left(-\frac{\log^2(\nu_1/\nu_2)}{2\xi^2}\right), \tag{3.2}$$

where A denotes the overall amplitude and β parameterize the foreground distribution on angular scales. The parameter α is the foreground spectral index and ξ is the frequency-space correlation length that parameterizes the characteristic frequency scale over which foregrounds are correlated. The parameter values can be found in table 3.1. As before, we make use of the code presented in Alonso et al. [2014] to create the desired foreground maps.

3.1.3 Galactic Synchrotron

Galactic synchrotron radiation is the strongest foreground emission in the frequency range of interest for intensity mapping and is produced by cosmic-ray electrons and positrons

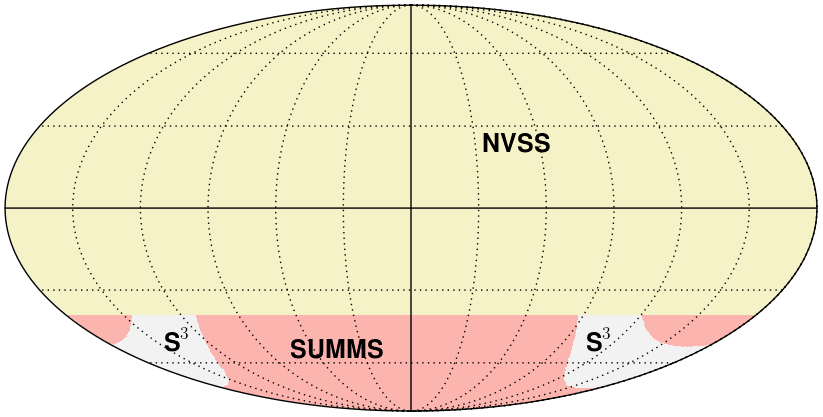

FIGURE 3.2: Survey coverage for NVSS and SUMSS. The yellow area indicates the NVSS coverage while the SUMMS area is in pink. The grey area is the SUMSS masked area *i.e.* galactic latitude $|b|$ <10° which was filled with S^3 sources. Note that below a flux cut of 5 mJy (see text for details) the full sky is filled with S^3 sources.

propagating in interstellar magnetic fields. Since a full physical model would require the knowledge of magnetic fields, cosmic-ray electron distributions and propagation, the standard approach is to resort to data driven modelling that relies on the Haslam full-sky map at 408 MHz [Haslam et al., 1982]. In Alonso et al. [2014], the Haslam map is extrapolated to the frequency range of interest using a spectral index $\alpha(\hat{n})$ taken from the Planck Sky Model [PSM, Delabrouille et al., 2013]. Moreover, since the Haslam map has poor resolution, small angular scales ($\ell > 200$) are filtered out and replaced with the isotropic model of Santos et al. [2005] discussed in the previous section and whose parameters are reported in table 3.1. An example of angular clustering of the galactic synchrotron emission at fixed frequency is reported in figure 3.1, where the transition between the two regimes can be noticed.

3.1.4 Point Sources

Previous studies aimed at modelling of extragalactic point sources (PS) such as (e.g. Santos et al. [2005]; Alonso et al. [2014]) considered this type of emission to be isotropic and that it can be characterized with a frequency space angular power spectrum. In this work we make use of a more realistic model for the PS that utilizes well validated data and simulations to create a full sky catalogue at 1.4 GHz from Ngobese [2018]. Using this new catalogue we produce pixelized maps at a frequency range of interest. The role of such a PS model will be important for the primary beam effects being studied. In the following section, we give a brief overview of how the full sky catalogue was created.

3.1.4.1 Constructing the Full-sky Point Source Catalogue

The point source catalogue is created using two publicly available data sets, the National Radio Astronomy Observatory VLA Sky Survey (NVSS, Condon et al. [1998]) at 1.4 GHz and the Sydney University Molonglo Sky Survey (SUMSS, Mauch et al. [2003]) with frequency 843 MHz and the SKAO Simulated Skies (S^3, Wilman et al. [2008]) simulation.

Our starting point is the SKAO Simulated Skies (S^3, Wilman et al. 2008), a semi-empirical simulation of the extragalactic radio continuum whose simulated sources, that include AGN and star-forming galaxies, are drawn from realistic luminosity functions and follow with appropriate bias the underlying dark matter density field. The minimum flux density, $S_{\rm min}$ obtained from S^3 is 10 μJy. The S^3 simulation covers a patch of 20×20 deg^2. To extend the catalogue to full sky we use the following approach:

- we consider the S^3 area, $\Omega_{\rm sim}$ and group all the flux densities into logarithmic bins. Every flux bin i contains $N^i_{\rm sim}$ sources;
- we create an empty HEALPix map with $N_{\rm side} = 512$, whose pixel area is $\Omega_{\rm pix}$[2]. The resolution of this map will give the resolution of our final catalogue;
- we assume that the number of sources scales linearly with area such that the mean number of sources per pixel corresponding to the i-th flux bin is $\bar{N}^i_{\rm pix} = N^i_{\rm sim}\left(\frac{\Omega_{\rm pix}}{\Omega_{\rm sim}}\right)$;
- we assign to every pixel a certain number of sources for every flux bin, Poisson sampled from $\bar{N}^i_{\rm pix}$;
- we assign to every source a spectral index, α, drawing from a Normal distribution $\mathcal{N}(-0.8, 0.2)$, in broad agreement with Garn et al. [2008]. This spectral index dictates the frequency scaling of the source flux, using the standard power law $S \propto \nu^\alpha$;
- right ascension and declination of every source are obtained from the pixel position in the sphere. Note that multiple sources will have the same (ra, dec). This is not a limitation as long as the map resolution is higher than our target experimental resolution.

With this procedure we construct a full-sky catalogue of S^3-like sources. We note that this catalogue does not include angular clustering seen in real data (e.g. de Oliveira-Costa

[2]Note that $N_{\rm side} = 512$ corresponds to $\Omega_{\rm pix} \sim 0.01\,{\rm deg}^2$.

and Capodilupo 2010, de Oliveira-Costa and Lazio 2010). To make the catalogue more realistic, we therefore added the observational data from two radio surveys, that naturally carry the information on the large-scale distribution of matter. We consider the data from the National Radio Astronomy Observatory VLA Sky Survey (NVSS, Condon et al. [1998]) at frequency 1.4 GHz and the Sydney University Molonglo Sky Survey (SUMSS, Mauch et al. [2003]) with frequency 843MHz. NVSS covers declination $\rho \geqslant -40^\circ$ while SUMSS covers $\rho \leqslant -30^\circ$. To avoid double counting of sources we only consider SUMSS at declination $\leqslant -40^\circ$. Note also that SUMSS has a galactic mask at latitude $|b| < 10^\circ$. Inside this area we keep using the S^3 generated sources.

The resulting sky coverage of the NVSS, SUMMS and S^3 is shown in figure 3.2 in equatorial coordinates. Note that to extrapolate to other frequencies, we assign a spectral index to NVSS and SUMSS sources the same way as for the S^3 sources. We then use these spectral indexes to extrapolate SUMSS flux densities from 843 MHz to the common 1.4 GHz frequency. For both NVSS and SUMSS we only keep the sources higher than a specific flux cut, $S_{\rm cut}$, at 1.4 GHz to avoid spurious detections. We set $S_{\rm cut} = 10\sigma_{\rm NVSS} \simeq 5$ mJy where $\sigma_{\rm NVSS}$ is the NVSS noise flux rms. On the NVSS and SUMSS sky patches, below $S_{\rm cut}$, we include the S^3 sources. One may worry that below $S_{\rm cut}$ the point sources are not correlated across the sky. However, for the objectives of this paper, this lack of correlation should not affect the conclusions from the beam effects. In figure 3.3 we reconstruct the differential source counts diagram at 1.4 GHz. The scatter at high flux density is due to Poisson sampling with a low average number of sources per bin. The catalogue can be downloaded from the CRC repository. A detailed description of its structure can be found in section 3.1.5, together with the description of the polarized part of the catalogue which is not being used in this analysis.

3.1.4.2 Point Source Maps

For our simulation pipeline we need to transform the point source catalogue discussed in the previous section into a pixelized point source temperature brightness HEALPix maps at every frequency of interest. We use $N_{\rm side} = 512$. As we will discuss further later on, this is a safe choice since the expected resolution of a single-dish survey is a factor 10 worse. Given $N_{\rm side}$, the pixel area, $\Omega_{\rm pix}$, is fixed and the coordinates of every source can be associated to a specific pixel. The brightness temperature value for every pixel can

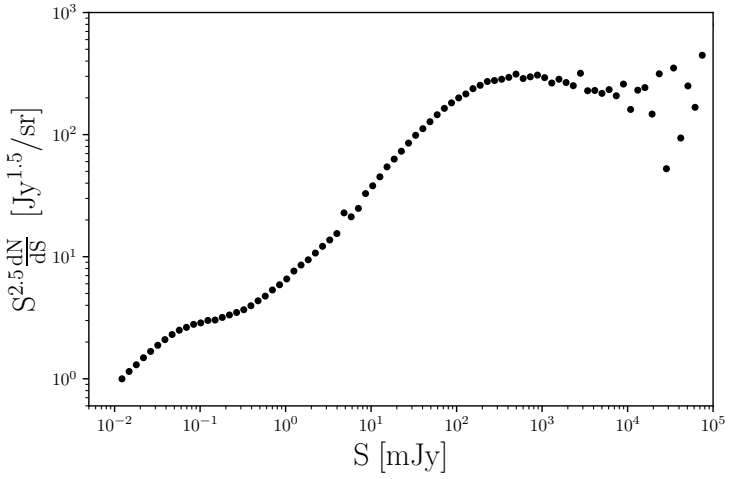

FIGURE 3.3: The normalized intrinsic source counts plot of the catalogue at 1.4 GHz. Below 5 mJy are the sources from the S^3 simulation while above are NVSS sources, extrapolated SUMSS sources or S^3 (see figure 3.2).

then be computed using the Rayleigh-Jeans approximation:

$$T_b^j(\nu) = \left(\frac{c^2}{2k_B\nu^2\Omega_{pix}}\right)\sum_{i=1}^{N_p^j} S_i(\nu), \tag{3.3}$$

where k_B is the Boltzmann constant, c the speed of light and the sum is over all the sources falling in the j-th pixel. $S_i(\nu)$ is the flux density of each these sources rescaled from 1.4 GHz to ν using the source spectral index. An example of such map is given in figure 3.4. In figure 3.1 we show instead an example of the angular power spectrum of the same map compared to the other foreground emissions and the HI signal. Together with the C_ℓ of the map created using the full point source catalogue, we show how the angular power spectrum reduces in amplitude if successive stronger flux cuts are applied to the catalogue.

3.1.5 Point Source Catalogue Extension to Polarization

It is well known that, due to instrumental imperfections, the polarized sky can leak into the intensity mapping signal. Polarization leakage can be particularly challenging for foreground cleaning algorithm since its spectral behaviour is not smooth due to Faraday rotation [e.g. Spinelli et al., 2018]. At radio frequencies, the two main polarized foregrounds are the diffuse Galactic synchrotron and the polarized signal from point sources. Simulations of the former and its impact on intensity mapping has been investigated in

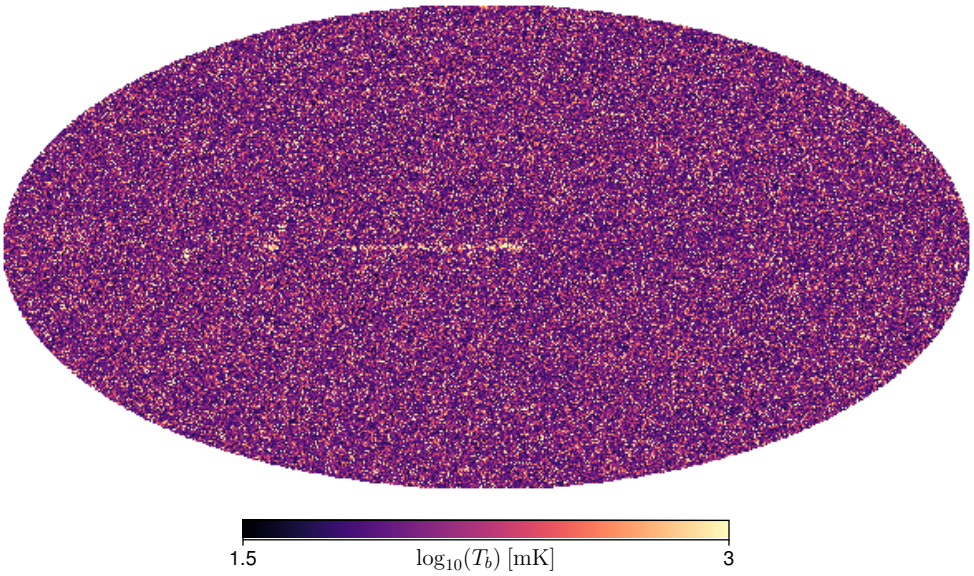

FIGURE 3.4: Our PS model at 950 MHz in galactic coordinates.

Alonso et al. [2014] (see also Shaw et al. 2015 and more recently Carucci et al. 2020, Cunnington et al. 2021), while the latter contaminant remains quite unexplored.

In this section we extend the intensity catalogue presented in section 3.1.4 to include also polarization. An accurate discussion of the impact of our polarized simulation for IM is ongoing and will be presented in a future work. Note that other simulated polarized source catalogues are also available in the literature [e.g Bonaldi et al., 2018]. Our catalogue is the result of the combination of two surveys (NVSS and SUMMS) and the S^3 simulations. Neither SUMMS or S^3 contain information on source polarization so we use the NVSS polarized flux density $|P|$ data and statistically extrapolate them to our simulated sources. We first bin the NVSS flux densities into five bins, to have a reasonable amount of signal to noise ratio. We then obtain the polarization fraction for every NVSS source in a bin using $\Pi = |P|/S$ and we compute the mean polarization fraction $\bar{\Pi}_i$ and its standard deviation σ_{Π_i}) for the i-th bin. We then divide all the sources in our catalogue in the same five bins and assign to each of them a polarization fraction drawn from a Gaussian distribution $\mathcal{N}(\bar{\Pi}_i, \sigma_{\Pi_i})$. Every source in the catalogue now has an assigned polarized flux value $|P| = \Pi S$.

TABLE 3.2: The 1.4 GHz point source catalogue catalogue format for every radio source. The total number of sources is ~ 408.8 million.

ra	right ascension in degrees
dec	declination in degrees
α	source spectral index
$S_{1.4\,\mathrm{GHz}}$	source flux density at 1.4 GHz in mJy
$\lvert P\rvert$	source polarized flux density at 1.4 GHz in mJy
ϕ_0	the intrinsic polarisation angle in degrees
ψ	the rotation measure given in rad/m^2

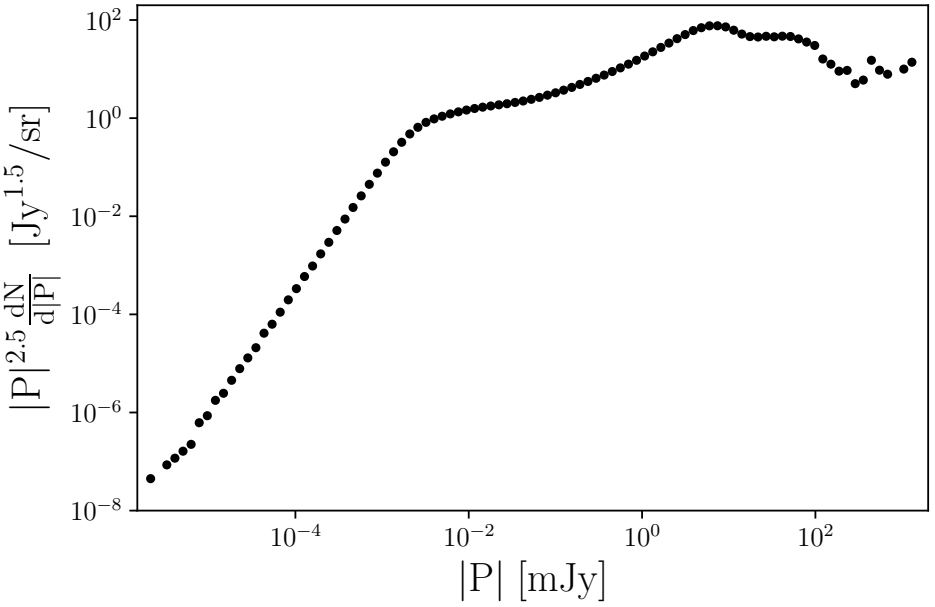

FIGURE 3.5: The normalized intrinsic polarized source counts of our catalogue at 1.4GHz. Polarization fractions are computed from NVSS data and extrapolated to all sources.

To construct Stokes Q and U maps from the catalogue we are still lacking important information: the parameters Q and U are related to $|P|$ via

$$Q + iU = |P|\, e^{2i\phi}, \tag{3.4}$$

where ϕ is the polarisation angle given by $\phi = \phi_0 + \psi\lambda^2$. ϕ_0 is the intrinsic polarisation angle that is rotated by the Faraday depth ψ (also called rotation measure), proportional to λ^2. For our catalogue, we assume that the intrinsic polarisation angle can be drawn from a Uniform distribution $\phi_0 \sim \mathcal{U}(0°, 360°)$. For the Faraday depth ψ instead, we follow Nunhokee et al. [2017], that, motivated by the finding of Taylor et al. [2009], assume ψ values as Gaussian distributed around a mean value of 5.6rad/m^2 and a variance of 20 rad/m^2. We can now assign to every source in the catalogue a value for ϕ_0 and ψ.

The full list of entries of our catalogue era reported in table 3.2. The final differential polarized source counts is reported in figure 3.5. The full catalogue is available in the CRC repository. We recall that the catalogue is based on simplified assumptions. Nevertheless, polarization leakage is still a not well explored topic and thus we believe that this polarized source catalogue could be of general interest for the IM community.

3.2 MeerKAT Beam Model

The simulations that we create focus on MeerKAT single dish observations (i.e. the autocorrelation signal) and assume for simplicity that all dishes have the same primary beam, since our focus are the frequency effects. This is a fair approach since map making is done separately for every dish. We only look at the total intensity (Stokes I) since the 21cm signal is unpolarized to a high degree. Polarization leakage might be an important effect for foreground cleaning (Alonso et al. [2014]; Carucci et al. [2020]; Cunnington et al. [2021]) and we plan to return to it in a follow up study. A final assumption is that our beam model has circular symmetry. Indeed, as seen in figure 3.6, the MeerKAT beam is symmetric to a good degree. There are however small beam asymmetries that become more relevant away from the center (Asad et al. [2021]). Nevertheless, this is more of an issue when we need to "subtract" point sources from the map and is an effect that can be separated from the frequency problem which is our focus here. We can imagine that each pixel on the sky will be observed many times, depending on the scanning and cross-linking pattern. The beam we are modelling will be the final one after averaging over these observations and should therefore be more symmetric. Deviations from this symmetry means that point sources along a given circumference might be multiplied by a different beam value. However, the changes in frequency and its impact on smoothness will still be similar in any given direction and we believe that the models presented here capture the most relevant effects.

Using spherical coordinates, we represent our beam function as $B(\nu, \theta, \phi)$, where ν is the frequency of observation, θ the polar angle and ϕ the azimuthal angle. The beam pattern is related to the dish reflective surface, or more accurately, its effective area set by the aperture illumination function, through a Fourier transform (for details see Wilson and Rohlfs K. [2013]). It is maximal in the direction at which the telescope is pointing ($\theta = 0$) and decreases with θ, away from the pointing direction. As already mentioned, we assume

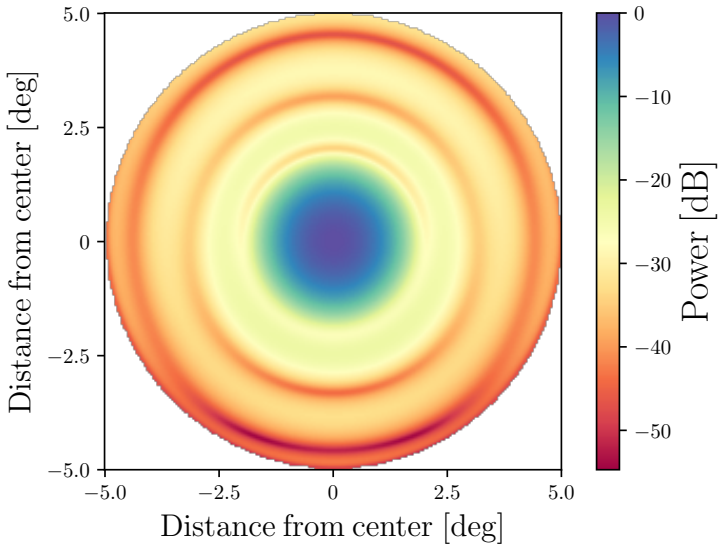

FIGURE 3.6: 2-d image of the MeerKAT primary beam model at 950 MHz obtained using the EIDOS package [Asad et al., 2021].

that gain calibration has been done already and the beams are normalized to 1 at the center.

Figure 3.6 shows the 2-dimensional MeerKAT beam at 950 MHz, using the "EIDOS" package[3] from Asad et al. [2021]. The current version of the package can be used to create MeerKAT L-band beams from Zernike polynomial fits to holographic observations or EM simulations within a maximum diameter of 10 degrees. In this study we use the fit to the MeerKAT holographic measurements. Using this package, we can generate images of the Stokes I beam at any frequency, although only up to 5 degrees from the beam center. We then numerically average this 2d beam over the ϕ direction in order to obtain a beam pattern that is a function of θ only: $B(\nu, \theta)$. We refer to this beam as the EIDOS beam. In figure 3.7 we can see the shape of this beam as a function of θ, at a frequency of $\nu = 950$ MHz (solid cyan line).

Since the beam above only extends to 5 degrees, we explored other models that can be applied to a full sky case. High accuracy is not required but we want a function that captures some of the main trends of the MeerKAT beam: 1) it is an accurate representation within the main lobe; 2) has a full width at half maximum (FWHM) with the same frequency dependence; 3) decays with θ at the same rate and 4) has side-lobes and follows

[3] https://github.com/ratt-ru/eidos

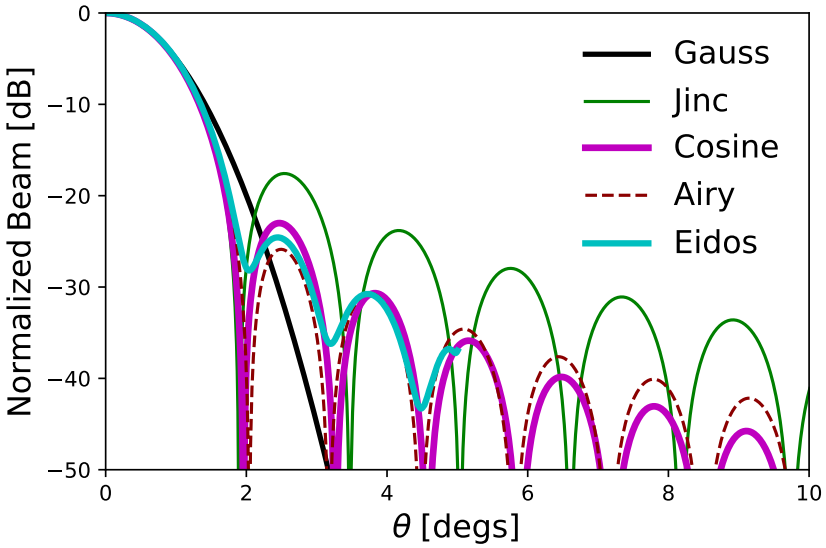

FIGURE 3.7: A comparison of primary beam models at 950 MHz: the standard Gauss model (black), the Jinc beam model corresponding to an unblocked circular aperture [Wilson and Rohlfs K., 2013] (solid green), the Cosine model [Condon and Ransom, 2016] (magenta), the Gaussian tapered airy disk used in Harper and Dickinson [2018] (dashed red) and the one obtained from the EIDOS package presented in Asad et al. [2021] (cyan).

the nulls and peaks with reasonable accuracy. We would also like such function to be easy to calculate in order to quickly deploy it in simulations.

3.2.1 Modeling the Frequency-dependent Ripple

The FWHM, $\Delta\theta$, of a MeerKAT dish is given approximately by

$$\Delta\theta \approx 1.16\frac{\lambda}{D}, \tag{3.5}$$

where λ is the observed wavelength and D the dish diameter (13.5m for MeerKAT). Measurements of the MeerKAT/EIDOS beam shows that the FWHM follows this dependence but also exhibits a low-level frequency-dependent ripple. This effect can be seen clearly in figure 3.8, where the EIDOS FWHM is normalized by λ/D. This ripple is caused by the interaction between the primary and secondary reflector of MeerKAT [de Villiers, 2013] and can be important in the foreground cleaning as it will add extra structure to the frequency spectra.

To address the effect on the extraction of the cosmological signal of such a frequency dependent FWHM, we model its main features. We fit the additional smooth frequency

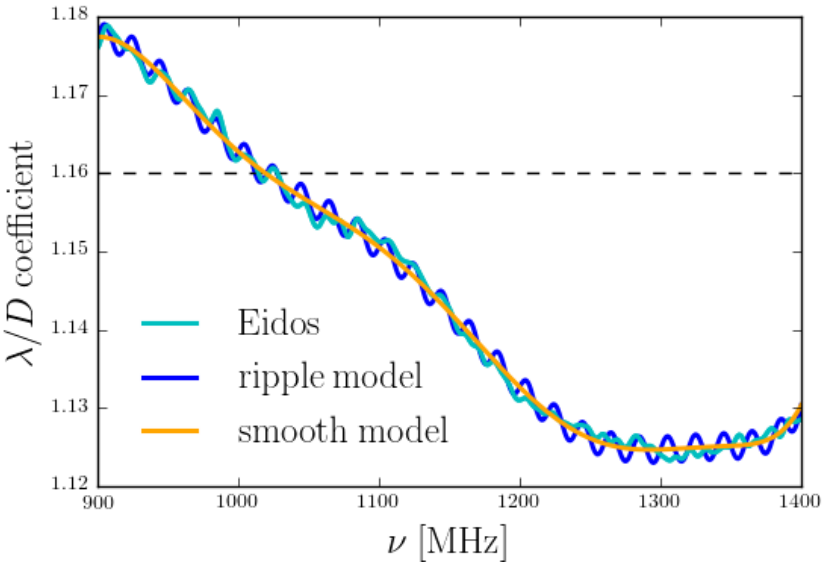

FIGURE 3.8: The FWHM of the azimuthally averaged MeerKAT/EIDOS beam normalized by λ/D as a function of frequency (solid cyan line) compared to our ripple model (solid blue line) which is composed of a sinusoidal oscillation on top of a smooth polynomial frequency dependence (solid orange line) - see equation 3.6.

TABLE 3.3: Numerical values of the coefficient of equation 3.6.

A [arc-min]	T [MHz]	a_n $\{n = 0, ..., 8\}$
0.1	20	{6.7e3, −50.3, 0.16, −3.0e−4, 3.5e−7, −2.6e−10, 1.2e−13, −3.0e−17, 3.4e−21}

dependence with a polynomial. We choose a high order polynomial (8th degree) to accurately describe the beam in the frequency interval of interest. On top of it we superimpose a sinusoidal oscillation with period T and amplitude A arc-minutes,

$$\Delta\theta_r = \frac{\lambda}{D}\left[\sum_{d=0}^{8} a_d \nu^d + A\sin\left(\frac{2\pi\nu}{T}\right)\right]. \tag{3.6}$$

The values of the parameters are summarized in table 3.3 and the ripple model and its smooth component are shown in figure 3.8. The smooth polynomial provides an a-dimensional scaling factor i.e the coefficients a_d have dimension $\mathrm{MHz}^{-\mathrm{d}}$. We note that in Asad et al. [2021] this ripple in the beam width is shown to be asymmetric between the vertical and the horizontal direction (e.g. it is not rotation invariant). Our model can be considered a spherical averaged version of this effect which we believe still captures the main issues of such ripple. The combination of beam asymmetries and sky rotation will probably result in a superposition of sine waves which will leak the ripple across more

scales while reducing its overall amplitude.

So far we have discussed the FWHM of the beam that we are assuming fully describes its frequency behavior. We address now the possible choices to describe how the beam behaves as a function of the polar angle θ. One interesting option for a beam model is the jinc function: $\mathrm{jinc}(x) \equiv J_1(x)/x$, where $J_1(x)$ is the Bessel function of the first kind. This model is quite popular as it corresponds to an unblocked circular aperture with uniform illumination [Wilson and Rohlfs K., 2013]. We use:

$$B_{\mathrm{J}}(\nu, \theta) = 4\ \mathrm{jinc}^2\Big(\pi \frac{\theta}{\Delta\theta}\Big). \tag{3.7}$$

Note that the correct derivation would use $\sin\theta$ instead of θ in the equation above. They give similar results for $\theta < 10°$ but will start to deviate after that. We found that using θ instead provides a better behavior at large angles. We can see from figure 3.7 that this function follows the beam main lobe accurately and captures the nulls of the first side-lobes. However, the amplitude of the side-lobes is higher than the MeerKAT/EIDOS beam.

Another option with smaller side-lobes is the beam pattern resulting from a cosine-tapered field (or cosine-squared power) illumination function (Condon and Ransom 2016),

$$B_{\mathrm{C}}(\nu, \theta) = \left[\frac{\cos\left(1.189\theta\pi/\Delta\theta\right)}{1 - 4(1.189\theta/\Delta\theta)^2}\right]^2. \tag{3.8}$$

Figure 3.7 shows that this model fits the main lobe and the first two side-lobes quite well. Indeed, a simplified MeerKAT beam model based on this function is publicly available[4]. More details of this MeerKAT beam model can be found in Mauch et al. [2020].

Finally, we also consider a Gaussian function:

$$B_{\mathrm{G}}(\nu, \theta) = \exp\Big[-4\ln(2)\Big(\frac{\theta}{\Delta\theta}\Big)^2\Big]. \tag{3.9}$$

This is the simplest case and we include it here for consistency. It is a good approximation to the main lobe (at least down to an order of magnitude) but it neglects completely the effects of the side-lobes. One important point is that all these beams are a function of the ratio $\frac{\theta}{\Delta\theta}$ so that a change in $\Delta\theta$ is equivalent to a rescaling in θ.

[4] https://github.com/SKA-sa/katbeam

A somewhat more accurate option was proposed in Harper and Dickinson [2018] based on the transformation of a Gaussian tapered airy disk. This model is also presented in figure 3.7 after tuning the parameters. We can see from the figure that within 0 to 1 degree (i.e. within the main lobe) all beams do match. Beyond that, the Jinc beam side-lobes drop slower in amplitude compared to the side-lobes of the other beams. The airy beam has a trend with θ similar to the cosine but does not depend directly to the FWHM, making it difficult to control its scaling with frequency since the integration parameters need to be adjusted at each frequency. For these reasons, we focus the rest of our analysis on the Cosine beam as the best description of the MeerKAT true beam and retain the Gaussian and the Jinc respectively as an alternative best and worst case scenario for side-lobes.

3.3 Survey Specifications

Following what was proposed with the MeerKLASS survey, we consider a survey area of almost 10% of the sky, overlapping with the sky area probed by the Dark Energy Survey (DES), in order to allow comparative studies and cross-correlation analysis that will ultimately help in constraining cosmological parameters. We choose the sky patch as in figure 3.9 and consider observations between 900 MHz and 1050 MHz with a frequency resolution of 1 MHz. The 150 channels cover the redshifts range $z \in (0.35, 0.58)$. We store the mock data in HEALPix at $N_{\rm side} = 512$ that corresponds to a pixel resolution of $\theta_{pix} = 0.11$ deg.

Table 3.4 summarizes the survey specifications and instrumental parameters. The instrumental noise can be computed as a function of these parameter choices. In our simulations we consider only thermal noise, that is, Gaussian noise with null mean and a standard deviation σ_N [Wilson and Rohlfs K., 2013]:

$$\sigma_N = \frac{T_{\rm sys}}{\sqrt{2t_{\rm pix}\Delta\nu}}, \tag{3.10}$$

with $T_{\rm sys}$ the system temperature, $\Delta\nu$ the frequency resolution and $t_{\rm pix}$ the total integration time spent on a single pixel,

$$t_{\rm pix} = t_{\rm obs} N_{\rm dish} \frac{\Omega_{\rm pix}}{\Omega_{\rm sur}}. \tag{3.11}$$

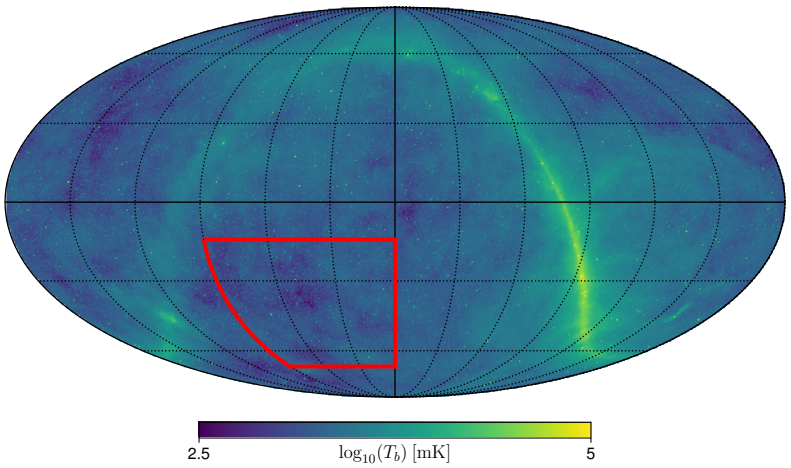

FIGURE 3.9: The simulated survey region in equatorial coordinates (red contours) covering $4k$ deg^2 of the sky ($f_{\rm sky} \sim 0.09$). For illustrative purpose we show in the background the simulated foreground emission at 950 MHz. The simulated survey area avoids the strong emission coming from the Galactic plane.

where $t_{\rm obs}$ is the total integration time, $\Omega_{\rm sur} = 4\pi f_{\rm sky}$ the survey area, $\Omega_{\rm pix}$ the pixel area and $N_{\rm dish}$ the number of telescope dishes[5]. With the assumptions listed in table 3.4 and the choice of $N_{\rm side} = 512$ we obtain a noise rms value of 0.245 mK. The resulting noise map $T^{f_{sky}}_{\rm noise}(\hat{\mathbf{n}})$ takes into account the sky patch considered and every unmasked pixel contains values drawn from a Gaussian distribution. Note that in principle $T_{\rm sys}$ should include a frequency dependent evolution inherited from the sky temperature, but we are neglecting this small variation and assuming the noise to be constant with frequency. We do not expect the specifics of the instrumental parameters assumed here to affect the main conclusions of the paper, which can be easily extrapolated to the SKAO.

3.4 Simulating an HI IM Observation with a Spherical Harmonic Transform Convolution Approach

To simulate how a measured sky brightness temperature map will look like, we need to convolve the sky map with one of the beam models described in section 3.2:

$$\tilde{T}_{\rm sky}(\nu,\theta,\phi) = \int \left[\hat{R}(\theta,\phi)B\right](\nu,\theta',\phi')T_{\rm sky}(\nu,\theta',\phi')d\Omega', \tag{3.12}$$

[5]Note that, although data are auto-correlation of single dish measurements, the final maps will be a combination of the independent measurements obtained with the various dishes, enhancing the signal-to-noise.

TABLE 3.4: Instrumental parameters used to simulate an intensity mapping survey with MeerKAT.

D (dish diameter)	13.5 m
$t_{\rm obs}$	4 000 h
$\Delta\nu$	1 MHz
N_ν	150
$N_{\rm dish}$	64
$T_{\rm sys}$	20 K
$(\nu_{\rm min}, \nu_{\rm max})$	(900, 1050) MHz
$(z_{\rm min}, z_{\rm max})$	(0.35, 0.58)
$f_{\rm sky}$	0.09
survey area ($\Omega_{\rm sur}$)	$\sim$ 3 700 deg^2

where $d\Omega' = \sin(\theta')d\theta' d\phi'$ and $\hat{R}$ is the operator of finite rotations such that $\hat{R}B$ is the rotated beam into the (θ, ϕ) direction [e.g. Wandelt and Górski, 2001]. The direct computation of this convolution integral can be avoided by moving to spherical harmonic space. A sky map can be decomposed into spherical harmonics,

$$T(\nu, \theta, \phi) = \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} a_{\ell m}(\nu) Y_{\ell m}(\theta, \phi), \tag{3.13}$$

where $Y_{\ell m}(\theta, \phi)$ are the spherical harmonic functions and $a_{\ell m}(\nu)$ are the spherical harmonic coefficients,

$$a_{\ell m}(\nu) = \int T(\nu, \theta, \phi) Y^*_{\ell m}(\theta, \phi) d\Omega, \tag{3.14}$$

For a symmetrical beam this simplifies to

$$B(\nu, \theta) = \sum_{\ell} b_\ell(\nu) Y_{\ell 0}(\theta, \phi), \tag{3.15}$$

where the beam harmonic coefficients b_l do not depend on m and can be written as

$$b_\ell(\nu) = \int B(\nu, \theta, \phi) Y^*_{\ell 0}(\theta, \phi) d\Omega. \tag{3.16}$$

The convolution theorem transforms the integral into a simple product in harmonic space. We therefore use it for including the effect of the beam in our simulated sky maps. At each frequency, we compute the spherical harmonic transform of the sky temperature maps and the one of the beam model projected onto a HEALPix map, using healpix routines [Zonca

et al., 2019]. Then, a fast, element by element multiplication is performed,

$$\tilde{a}_{\ell m}(\nu) = \sqrt{\frac{4\pi}{2\ell+1}} a_{\ell m}(\nu) \frac{b_\ell(\nu)}{\sqrt{4\pi} b_0(\nu)}. \tag{3.17}$$

Note that we are assuming that the beam function, $B(\nu, \theta, \phi)$, is defined to be one at the center. We then need to divide by $\sqrt{4\pi} b_0(\nu)$ so that the beam integrated over the sky is normalized to 1. This is the required normalization in order to recover the signal angular power spectrum with the correct amplitudes. To obtain the beam convolved sky maps we simply use $\tilde{a}_{\ell m}(\nu)$ in equation 3.13.

3.4.1 Beam Effects on a Single Point Source

To examine the effect of a frequency dependent beam, we start by looking at the simplified case of a single point source. Since this point source should be represented by a Dirac delta function in terms of the sky temperature, the measured brightness temperature is:

$$T_P(\nu, \theta) = \frac{\lambda^2}{2k_B \int B(\nu, \Omega) d\Omega} S(\nu) B(\nu, \theta), \tag{3.18}$$

where θ is the angle of the point source with respect to the beam pointing and, again, $B(\nu, \theta)$ is the telescope beam normalized to 1 at the center. As discussed in section 3.2, our standard description for the MeerKAT beam is the Cosine model of equation 3.8 and we use the full frequency dependence in equation 3.6. For simplicity, we take the point source flux, $S(\nu)$, to be constant in frequency and equal to 1 Jy. This corresponds to a temperature contribution of about 0.05 K at the peak.

The upper panel of Figure 3.10 shows the effect of the convolution with a frequency dependent beam (ripple model) for a few positions of the point source with respect to the beam center. The angular positions are defined for the standard "λ/D" case at 1 GHz (first peak refers to the first sidelobe peak). The smooth component of the beam model (central panel) creates a slowly varying behavior near the peaks and nulls of the sidelobes but we expect the foreground cleaning algorithms to be able to deal with this type of structure. Of course, the situation can become more complicated when we combine several strong point sources with different spectral indexes. Still, the overall effect is reasonably benign as we will see later.

Assuming that we know the beam well enough to remove the effect from the smooth component, we are left with the residuals seen in figure 3.11 which are much harder to deal with. If no further cleaning can be done, this will give the final contribution from this single point source. Note that we cannot remove this effect through a gain calibration as the amplitude of the effect changes with θ (and is zero at the peak). Interestingly, although the relative effect is stronger near the nulls, a point source at the FHWM gives a stronger overall contribution with a similar shape. However, the signal is around 0.02 mK. Since the 21cm signal rms is about 0.1 mK, this means that a 1 Jy source would have an impact at the 20% level at most. Again, the situation will be complicated once we include more point sources. The strength of the signal goes down as we move away from the center but there is a higher chance of finding more and stronger point sources. A 10 Jy source in the second null would generate contamination at the 10% level. Although we have been considering a single point source, the same analysis could be applied to fluctuations across the sky from diffuse components (a completely smooth component would not suffer from beam effects). A 0.1 K fluctuation over degree scales would show up in the first null and also contribute at the 10% level. Such fluctuations are expected with the galactic synchrotron.

This was just a basic calculation to show the expected contamination level, but a proper simulation is needed to take all the effects into account as we will in section 4.3.2. On the one hand, the combination of more sources across the sky will make the cleaning more complex. On the other hand, the cleaning algorithm might be more robust. Moreover, techniques such as point source subtraction and beam deconvolution can improve the overall outcome.

3.4.2 Mock Final Intensity Maps

To construct our mock sky maps, we start by adding together all the elements described in section 3.1, at every frequency ν and direction $\hat{\mathbf{n}}$,

$$T_{\rm sky}(\nu, \hat{\mathbf{n}}) = T_{\rm HI}(\nu, \hat{\mathbf{n}}) + T_{\rm fg}(\nu, \hat{\mathbf{n}}) \tag{3.19}$$

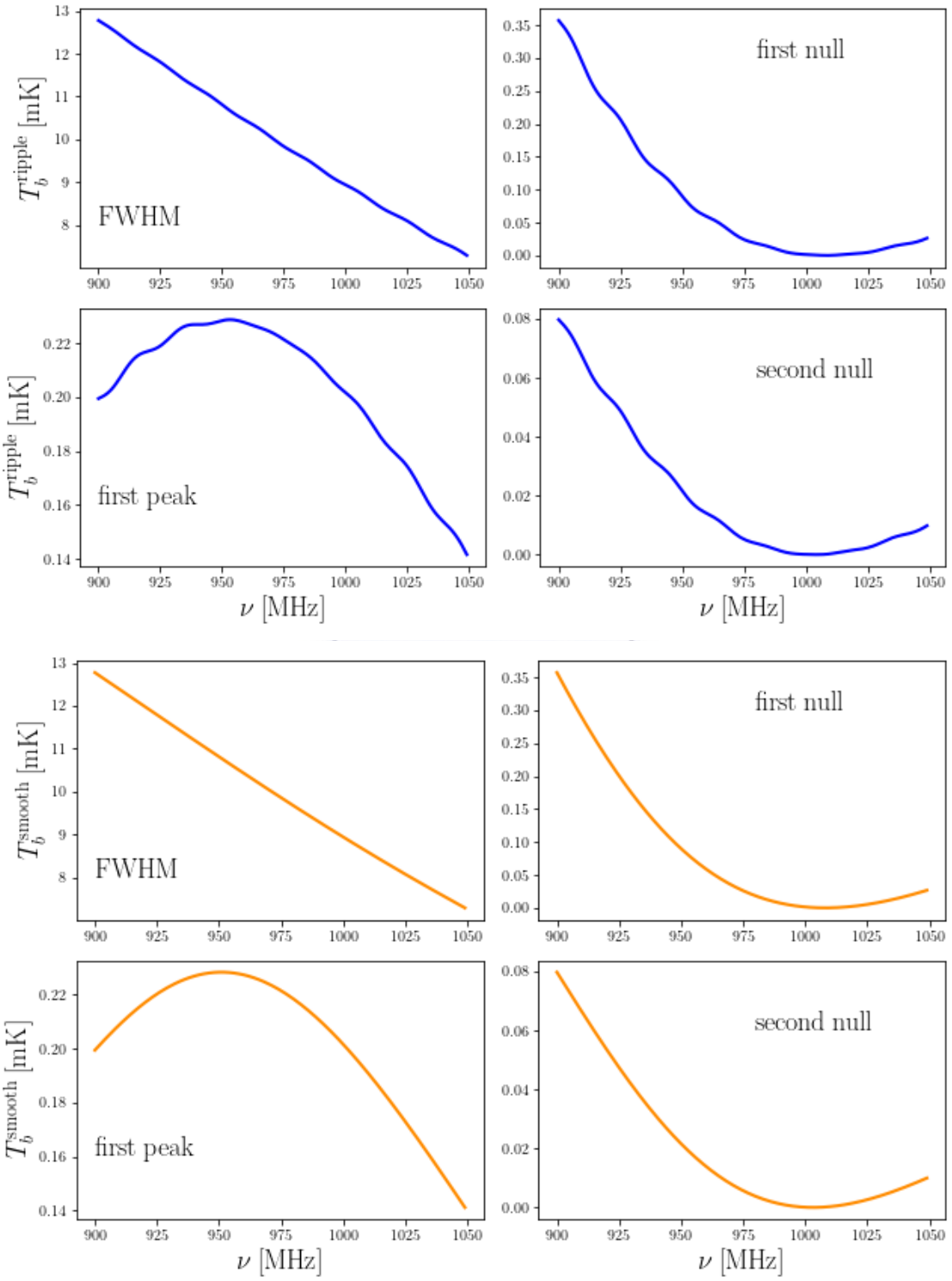

FIGURE 3.10: *Top panel:* Behavior of $T_b(\nu)$ for a single point source in different angular positions with respect to the center of the beam as a function of frequency. The positions are defined at 1 GHz. The beam model considered is the Cosine beam with $\Delta\theta$ of equation 3.6 (ripple model). *Central panel:* same as before but considering the Cosine beam with only the polynomial part of $\Delta\theta$ of equation 3.6 (i.e. with $A = 0$, smooth model).

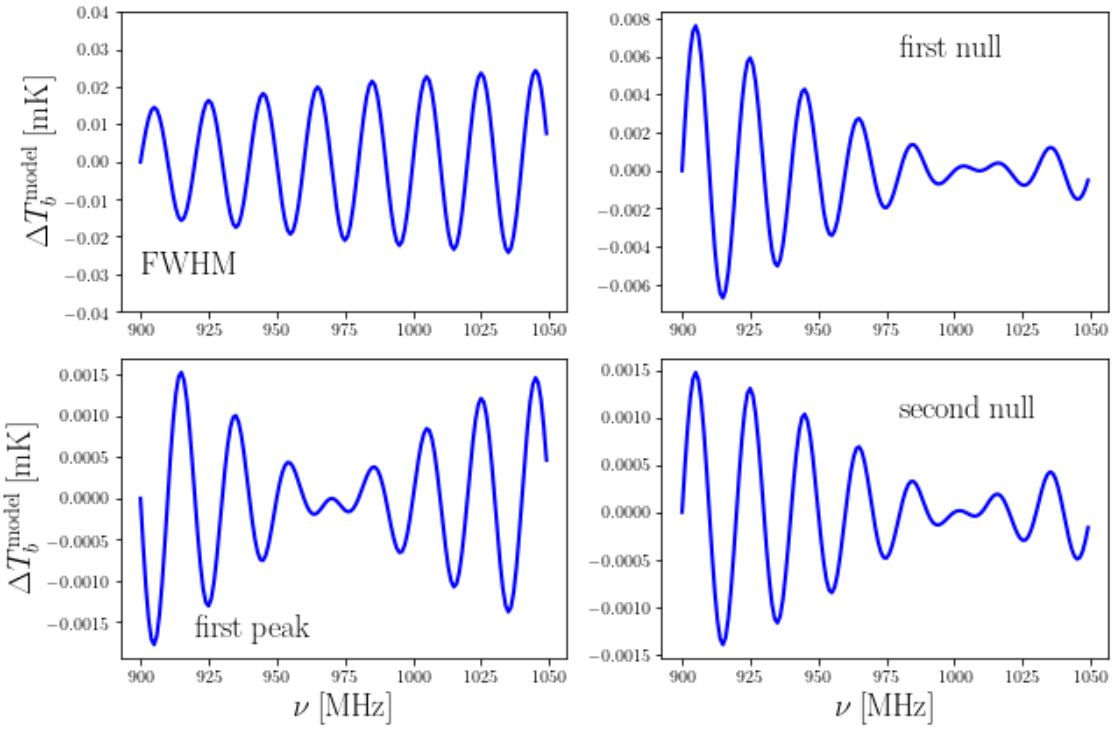

FIGURE 3.11: The residual behavior of $T_b(\nu)$ for the ripple model after the smooth component is subtracted.

where $T_{\rm HI}(\nu, \hat{\mathbf{n}})$ is the brightness temperature of the 21 cm signal, constructed as described in section 3.1.1 using a 1 Gpc^3 box with $N_{\rm grid} = 2048$ and

$$T_{\rm fg}(\nu, \hat{\mathbf{n}}) = T_{\rm PS}(\nu, \hat{\mathbf{n}}) + T_{\rm Gsynch}(\nu, \hat{\mathbf{n}}) + T_{\rm GFF}(\nu, \hat{\mathbf{n}}) + T_{\rm EGFF}(\nu, \hat{\mathbf{n}}), \tag{3.20}$$

where $T_{\rm PS}$ represents the brightness temperature of the point sources, $T_{\rm Gsynch}$ is the Galactic synchrotron emission and $T_{\rm GFF}$, $T_{\rm EGFF}$ are the Galactic and extra-Galactic free-free emission.

The sky signal needs to be convolved with the beam and masked to mock the MeerKAT observations. The beam convolution is performed as described in section 3.4, i.e. at the full sky level to avoid complication at the edges of the mask. Our reference mock maps include the Cosine beam model described in section 3.2, but we produce maps using also the Gaussian and the Jinc beam for comparison. For every beam model we need to specify its frequency dependence: either the simple proportionality to λ/D of equation 3.5, or the smooth/ripple model described in equation 3.6. For each of these beam models we compute the spherical harmonic coefficients b_ℓ defined in equation 3.15 and produce a final convolved map $\tilde{T}_{\rm sky}(\nu, \hat{\mathbf{n}})$ using equation 3.13 and 3.17. The final full sky map is

then masked to the target area and thermal noise is added.

$$T_{\rm obs}^{f_{sky}}(\nu,\hat{\mathbf{n}}) = \tilde{T}_{sky}^{f_{sky}}(\nu,\hat{\mathbf{n}}) + T_{\rm noise}^{f_{sky}}(\hat{\mathbf{n}}). \tag{3.21}$$

This final mock data product will be the input of the foreground cleaning algorithm.

3.4.3 Foreground Cleaning Diagnostics

To assess the degree of success of the foreground cleaning procedure we will use two estimators: the angular power spectrum, which describes the clustering on the angular direction at fixed frequency, and the radial power spectrum which instead describes the clustering along the line of sight. One of the questions we are addressing is what is the effect on foreground cleaning when we have different types of frequency dependence on the beam size (standard, smooth, ripple). To this end, looking separately at the line of sight and the angular direction will help in identifying residual structures in the cleaned maps. The angular power spectrum is well adapted for tomographic measurement as the ones that HI IM would like to perform. The radial power spectrum is instead a powerful tool to check residuals due to the evolution of the primary beam which are dominant in the frequency direction.

3.4.3.1 Angular Power Spectrum

We define the brightness temperature fluctuation contrast ΔT as the difference between the temperature in each pixel and the mean of the sky patch under consideration. At every frequency, the angular power spectrum in the full sky case can be estimated from the spherical harmonic coefficient $a_{\ell m}(\nu)$ of $\Delta T(\nu)$ using

$$\hat{C}_\ell(\nu) \equiv \frac{1}{2\ell+1}\sum_{m=-\ell}^{+\ell} |a_{\ell m}|^2 \tag{3.22}$$

This estimator is no longer valid for sky patches, and finding the correct one is a non-trivial problem. One widely used solution is to apply the Monte Carlo Apodized Spherical Transform Estimator [MASTER, Hivon et al., 2002] to correct the C_ℓ for the effect of the mask. When the telescope scanning strategy probes a small patch of the sky with sharp edges, recovering the signal is also highly non trivial for the MASTER estimator.

Given that for our purposes we are interested in the quality of the cleaning and not in the shape of the signal itself, we will compute the C_ℓ using simply equation 3.22, correcting only for the sky fraction (dividing by $f_{\rm sky}$). The resulting C_ℓ will have approximately the same trend and amplitude as the full-sky ones but will display the well known oscillatory behavior due to the coupling of different scales induced by the presence of the mask. The HI signal with which we compare will suffer the same issue since the signal angular power spectrum is also computed on the selected sky patch, justifying the direct use of equation 3.22.

To show the quality of the PCA cleaning we will also use the quantity:

$$\langle (C_\ell^{\rm rec} - C_\ell^{\rm true})/C_\ell^{\rm true} \rangle \equiv \langle \Delta C_\ell / C_\ell^{\rm true} \rangle, \tag{3.23}$$

where $\langle . \rangle$ indicates averaging over the frequency channels.

3.4.3.2 Radial Power Spectrum

To explore clustering along the line of sight we use the definition of the radial power spectrum as in Alonso et al. [2014]. The frequency band is divided into slices, within which the universe should not evolve significantly and a constant redshift, $z_{\rm eff}$, is assumed. The slices, however, should be big enough for capturing all the relevant scales. We then compute the Fourier transform of the temperature fluctuations per bin along each line of sight, $\tilde{\Delta T}(k_\parallel, \hat{\mathbf{n}})$, where

$$k_\parallel = \frac{\nu_{21} H(z_{\rm eff})}{(1+z_{\rm eff})^2} k_\nu \tag{3.24}$$

and k_ν is the Fourier conjugate of the frequency, e.g. $\delta k_\nu = 2\pi/\Delta\nu$. The radial power spectrum results from an average over all lines of sight in the sky patch,

$$P(k_\parallel) = \frac{\Delta\chi}{2\pi N_{\rm pix}} \sum_{i=1}^{N_{\rm pix}} |\tilde{\Delta T}(k_\parallel, \hat{\mathbf{n}})|^2, \tag{3.25}$$

where $\Delta\chi_s = \chi(z_s^{\rm max}) - \chi(z_s^{\rm min})$ is the slice width.

Foreground removal methods generally struggle at the edges of the input frequency band. To partially avoid this bias we do not include in the analysis the 10 MHz at the beginning and at the end of our frequency range. Given the small redshift range of our simulations, we use a single redshift bin corresponding to $z_{\rm eff} \sim 0.46$. In chapter 4, we will show the

recovered $P(k_\parallel)$ for our cases of study. Moreover, as for the angular power spectrum we will consider the estimator

$$\left[P(k_\parallel) - P^{\text{true}}(k_\parallel)\right]/P^{\text{true}}(k_\parallel) \equiv \Delta P(k_\parallel)/P^{\text{true}}(k_\parallel), \tag{3.26}$$

to better assess the quality of the foreground cleaning. We anticipate that, due to the smooth frequency dependence of the foregrounds, the cleaning is expected to inevitably compromise the largest radial scales.

3.5 Summary

The main aim of this chapter was to describe in detail the modelling process of the single dish HI IM observations. In doing so, two main components of the simulations had to be considered, the radio sky intensity and the primary beam of a single dish radio telescope.

For the sky component, we show how to go about modelling the HI signal and the foregrounds discussed in section 2.4. In particular, we presented a new point source catalog that is more realistic and constructed from NVSS, SUMSS surveys and the S^3 simulation by Ngobese [2018]. Furthermore we described how this intensity catalog is further extended to include polarization.

The second aspect of the simulation that we considered, is how to model the single dish primary beam for the MeerKAT radio telescope with the survey design as discussed in section 3.3. In probing the frequency effects from the primary beam, the first assumption we made in this modelling is that all telescope dishes behave the same way. The second is that the beam has circular symmetry. Defining the primary beam in this way simplifies the modelling in that a direct computation of the convolution integral as defined in equation 5.3 is avoided and instead spherical harmonic transforms are be used by following the method described in section 3.4 to compute the beam convolution.

To address the first aim of this thesis, which is to investigate the role of primary beam sidelobes, we will test our beam convolution pipeline using the primary beam models shown in figure 3.7. The second frequency effect to be investigated is the non-trivial frequency dependent ripple present in the EIDOS/MeerKAT beam which is outlined in section 3.2.1. Within the same section, we described how to model this effect in the simulation pipeline.

As a proof of concept, we also test the effect of the frequency dependent ripple on a single point source in section 3.4.1. In chapter 4, we will present and discuss the results from the various test cases. The foreground cleaned results will be presented using the diagnostics described in section 3.4.3.

Chapter 4

Primary Beam Effects on Foreground Cleaning

In this chapter, we present the results of the single-dish HI IM simulations by conducting various tests to determine the role played by primary beam side-lobes, particularly in the presence of strong point sources. The results presented in this chapter are an original contribution that resulted in a publication [Matshawule et al., 2021].

To determine the role of beam side-lobes we will make use of various primary beams as shown in figure 3.7, run the simulation pipeline and discuss the results obtained. Secondly and the main crux of this work, we will evaluate how the non-trivial frequency dependence of the FWHM discussed in section 3.2.1, impacts our foreground cleaning when using the spherical harmonic approach discussed in section 3.4. With each case of the primary beam the different types of FWHM frequency variations will be tested, starting from the standard case, given by equation 3.5 and moving to those shown in figure 3.8.

In simulating the single dish HI IM observations, the survey specifications as described in table 3.4 are followed and the radio sky intensity maps are constructed as described in equation 3.19. For all cases that will be discussed, the results obtained will be presented in two forms; 1) a visual comparison of the reconstructed HI signal with input HI signal as in figure 4.1 or 2) quantitatively: via the estimators defined in section 3.4.3.

Three blind foreground cleaning methods will be considered (PCA, ICA and poLOG) and using a single test case, one of the cleaning methods will be selected based on how well

the method recovers the HI signal. From thereon, the results will be presented using the best performing foreground subtraction method.

4.1 Simulating Single-dish HI IM with CRIME

Before entering in the details of our pipeline we briefly present in this section the software that is the starting point of our analysis. Indeed, we will show single dish HI IM simulation result obtained with the CRIME software by Alonso et al. [2014] which is one of the publicly available scientific packages for simulating HI IM experiments already mentioned in the previous chapter. The result from this simulation will serve as a benchmark to ascertain that our pipelines are working correctly.

The main purpose of CRIME is to produce fast HI IM simulations. In modeling the radio telescope's primary beam, CRIME makes use of a Gaussian-type beam which is the standard assumption in HI IM studies. In this section, we demonstrate a single dish HI IM simulated with CRIME and perform foreground cleaning on the observed intensity maps in order to recover the cosmological HI signal.

The second aspect of CRIME is that the only frequency modulation considered in the primary beam is the beam width as described in equation 3.5 and does not take into account other frequency modulations that may be present in the primary beam such as the non-trivial frequency-dependent effect as shown in figure 3.8 and which in our simulations we can model as discussed in section 3.2.1 using equation 3.6.

Within the CRIME simulation exists a parameter script called *JoinT* which allows for a single dish HI IM observation to be conducted. The survey specifications, HI, and foreground maps are passed into CRIME, and Gaussian smoothing is applied. The maps are masked and the appropriate instrumental noise is added to obtain the observed sky brightness temperature maps from a MeerKAT-like radio telescope observation. Following this process, these observed brightness temperature maps are used as input for the foreground subtraction process using the *fgrm* script.

The output from the foreground cleaning process is the plots shown in Figure 4.1 which show the foreground cleaned angular and radial power spectrum from the three foreground cleaning methods PCA, ICA, and poLOG. In this visual comparison, the black curve

in both angular and radial spectra represents the 'observed' HI signal whilst the *cyan*, *magenta*, and *red* solid curves represent the reconstructed HI signal after the foreground removal. The N_{fg} parameter shows the degree to which the foregrounds are removed, starting from $N_{fg} = 3$ to $N_{fg} = 5$. From this figure, it can be observed that the foreground methods have no difficulty in retrieving the HI signal, in both angular and radial directions. At low $k_{\parallel}$ however, an overcleaning occurs since everything that is smooth (thus described with power at low $k_{\parallel}$) tends to be subtracted by the cleaning procedure, being it signal or foregrounds. This signal loss is particularly evident in the radial power spectrum but it is problematic also in the case of the 3D power spectrum. In literature [e.g. Cunnington et al., 2022, Switzer et al., 2015, Wolz et al., 2021], this signal loss is often alleviated using a transfer function. In this thesis we are mostly interested in showing possible bias from the cleaning procedure so we do not attempt any signal reconstruction.

4.2 Gaussian Beam Convolution

Having observed how CRIME simulates a single dish HI IM observation, in this section we perform a Gaussian beam convolution to model the same type of observation but now with the use of spherical harmonic transforms as outlined in section 3.4.

The initial test we perform has the primary beam modelled as a Gaussian. Our analysis throughout this study will focus on a frequency sub-band of 900 - 1050 MHz with a frequency resolution of 1.0 MHz of the MeerKAT L-Band. For each frequency channel we follow the method described in section 3.4 and finally analyse the reconstructed HI signal through the angular and 1D radial power spectra as described in section 3.4.3.

Figure 4.2 is a visual comparison of the foreground cleaned angular and line of sight (radial) power spectra of the Gaussian beam convolution with the FWHM$= 1.16\lambda/D$. In all plots the *black* curve is the input HI signal as observed by 1° Gaussian primary beam, whilst the differently colored curves are the reconstructed HI signal. For all three panels we observe that the cleaning methods are able to recover the HI signal successfully on all scales except at high ℓ in the angular power spectra where the beam is unable to resolve the HI structure. The same can be said with line of sight power spectra. At low $k_{\parallel}$, the missing signal is expected as anything that is smooth is at those scales and is subtracted since interpreted as foregrounds.

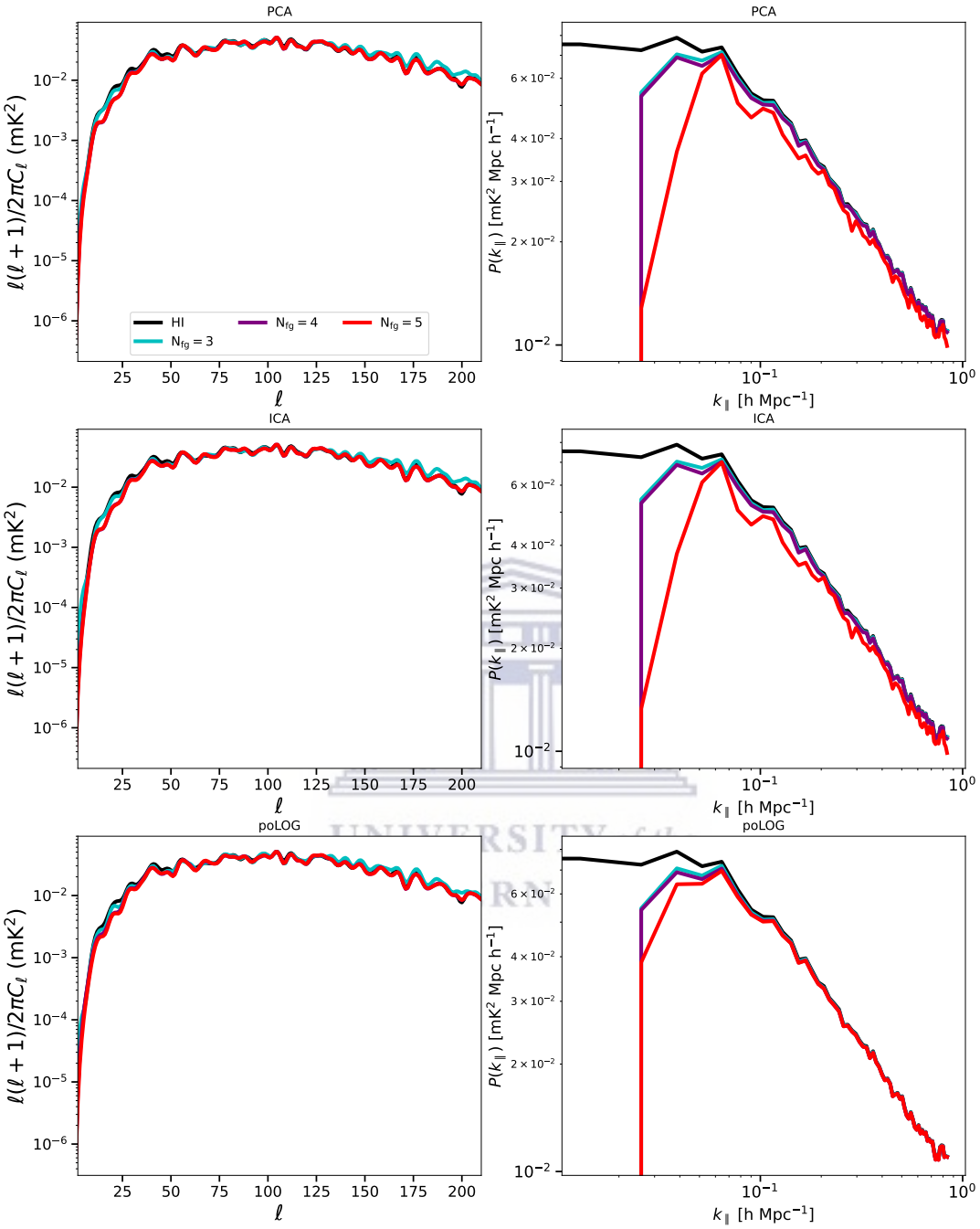

FIGURE 4.1: Shown in this figure are the cleaned angular and line of sight power spectra from the CRIME simulation obtained using the three blind foreground cleaning methods shown in the following order; PCA (*top*), ICA (*middle*) and LOS (*bottom*). The different colored curves represent the degree to which foreground cleaning is performed, with $N_{fg} = 3$ (in cyan) representing the lowest number of foreground modes removed and $N_{fg} = 5$ (red) being the highest number of modes removed in order get to the level of HI signal shown in the black solid curve. Note that here the signal is HI only and does not contain the noise. The noise power spectrum is also subtracted from the cleaned power spectrum.

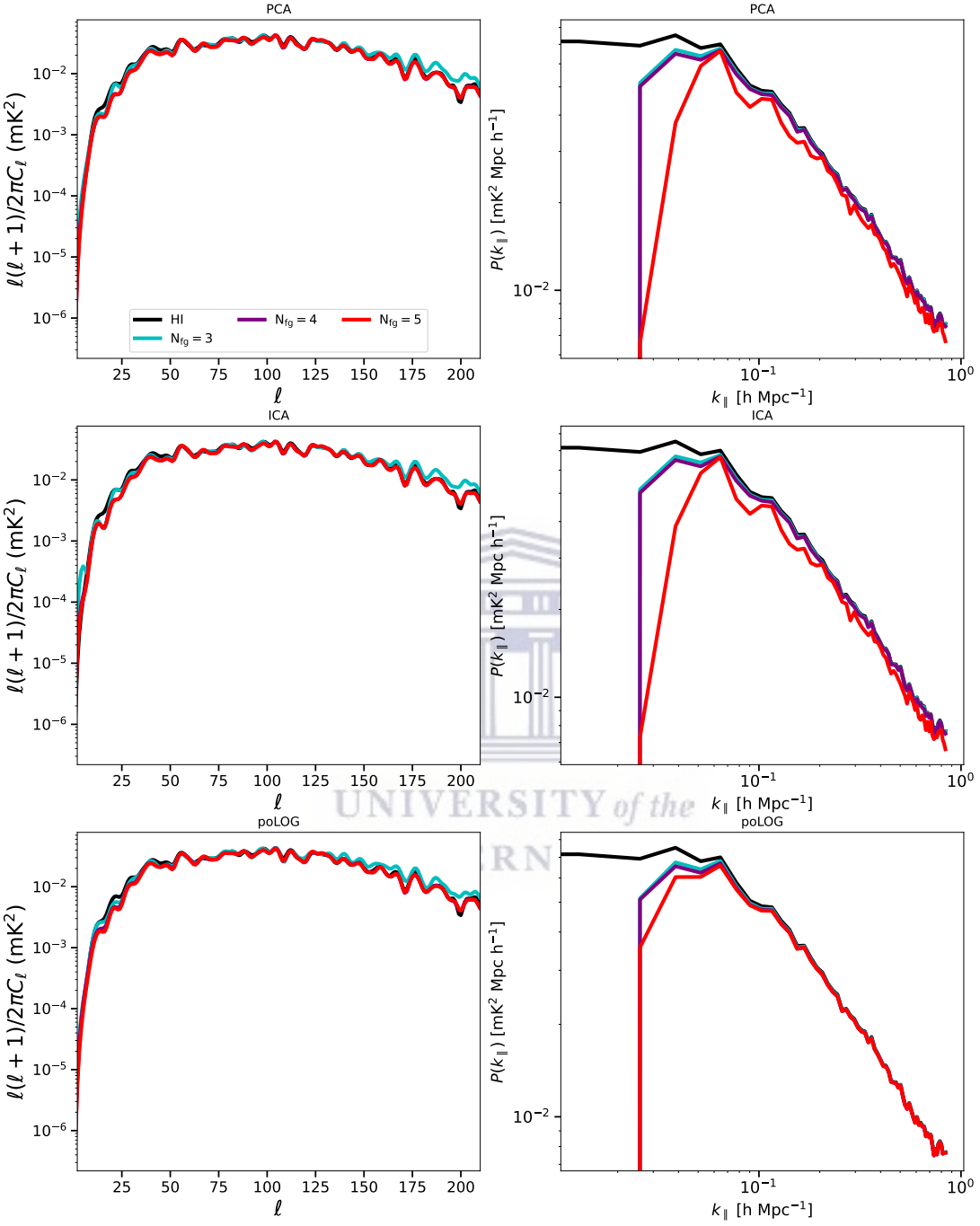

FIGURE 4.2: The foreground cleaned spectra shown is the Gaussian beam convolution with total intensity maps using the standard FWHM given in equation 3.5. The spherical harmonic transform approach is able to reproduce the same result as observed with CRIME. Furthermore when comparing the cleaning methods we see no difference in the manner in which they recover the HI signal. Note that here the signal is HI only and does not contain the noise. The noise power spectrum is also subtracted from the cleaned power spectrum.

We have performed a comparison between the mainstream CRIME package and the spherical harmonic transform approach and saw that the results are the same. We also observed that the foreground subtraction methods are able to reconstruct the HI signal and return similar results. In the sections to follow we will get into the main problem we are trying to address in this thesis, and going forward perform the foreground cleaning using only the PCA method. Note that, when comparing the power spectra of the cleaned maps with the input, we subtract the noise power spectrum from the residual and check it against the HI only one.

4.3 Results

We discuss here the results of our simulation and cleaning pipeline. The mock observations finalized in section 3.4.2 are the input of the foreground cleaning method described in section 2.5. To quantify the performance of the cleaning, we make use of the estimators detailed in section 3.4.3. The main focus of this thesis is to investigate how a non trivial shape and/or frequency dependence for the telescope beam impacts the cleaning. To this aim, we discuss the effect of side-lobes with particular attention to the role of the point source contamination in section 4.3.1, and analyze the increasing difficulties of the cleaning procedure in the presence of a frequency dependent FWHM in section 4.3.2. In section 4.3.3 we investigate how smoothing the maps to a common resolution could help the cleaning process.

4.3.1 Effect of Beam side-lobes

As discussed in section 3.2, a realistic model for the MeerKAT beam should include side-lobes. Since side-lobes should be most problematic in the presence of strong point sources, we show results considering different levels of point source contamination. Our best-case scenario considers that we will be capable of removing point sources with fluxes as low as 100 mJy at 1.4 GHz, but we explore also more pessimistic cases. Note that, at this stage, we are simply considering that the FWHM of the beam scales proportionally to λ/D (see equation 3.5). The more complex models of equation 3.6 will be explored later on.

A good level of approximation for the shape of the side-lobes in the case of MeerKAT, is obtained with the cosine-tapered field illumination function, i.e. the Cosine model. We

present in section 4.3.1.1 the results obtained using mock sky emission convolved with this type of beam. A more pessimistic assumption for the side-lobes is instead presented in section 4.3.1.2, where we consider the Jinc model (see figure 3.7).

4.3.1.1 Realistic side-lobes

We consider here the sky model convolved with a Cosine beam and apply the blind cleaning. An example of the reconstructed angular power spectrum is shown in the left panel of figure 4.3. Despite the side-lobes of the Cosine beam, the foreground cleaning method has no particular difficulties in reconstructing the HI signal and there is no dependence on the level of point source contamination. The results presented in figure 4.3 are for $N_{\rm fg} = 4$, which we found was the optimal number of components to be subtracted. Raising $N_{\rm fg}$ only worsens the over-cleaning at large scales.

When examining the radial power spectrum (right panel of figure 4.3) we find that the overall quality of the cleaning is good for intermediate and small scales. As for the angular power spectrum, we show results for $N_{\rm fg} = 4$. Large scales are inevitably over-cleaned since the strong foregrounds are mostly smooth in frequency thus making power at low $k_\parallel$ the first to be subtracted by blind methods. We note that the cleaning, at fixed $N_{\rm fg}$, is more aggressive for lower point source contamination. This is expected since the complexity created by the interaction between side-lobes and strong point sources could counteract signal loss. Moreover, we report that, for $N_{\rm fg} = 3$, the case with the full point sources still leaves high contamination on large scales. On the other end, as soon as the strongest point sources are not included in the sky model, choosing $N_{\rm fg} = 5$ leads to severe over-cleaning, subtracting power not only at large scales but at intermediate $k_\parallel$.

4.3.1.2 Pessimistic side-lobes

We now consider a more pessimistic model for the beam (worse than what is expected for MeerKAT) by convolving the sky with the Jinc beam and apply again the blind cleaning. We present again only the results with $N_{\rm fg} = 4$. In the left panel of figure 4.4, we can see that in this case the performance of the cleaning algorithm is poorer: the full point source case has residual contamination already at intermediate scales. A higher $N_{\rm fg}$ does not ease this contamination and worsens the signal loss at large scales. We can appreciate

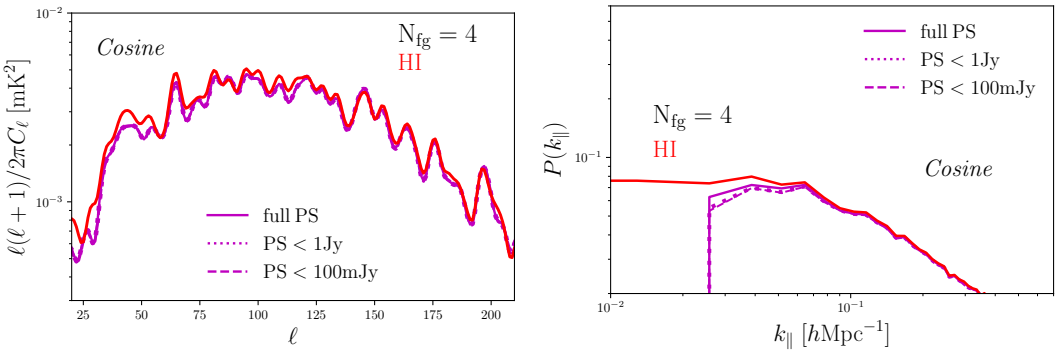

FIGURE 4.3: The results for the foreground cleaning estimators considered in this work: the angular power spectrum defined in equation 3.22 (left panel) for a given frequency (950 MHz in this example) and the radial power spectrum of equation 3.25 (right panel). We assume a Cosine beam model for the simulated observations and that the FWHM scales proportionally to λ/D. Different line-styles correspond to different levels of point source contamination. The solid curves represents the case with no flux cut applied to the catalogue (full PS), PS $<$ 1 Jy is in dotted and PS $<$ 100 mJy in dashed. We present results using $N_{\rm fg} = 4$ for the number of removed components. The retrieved signal is compared to the input HI signal plotted in red.

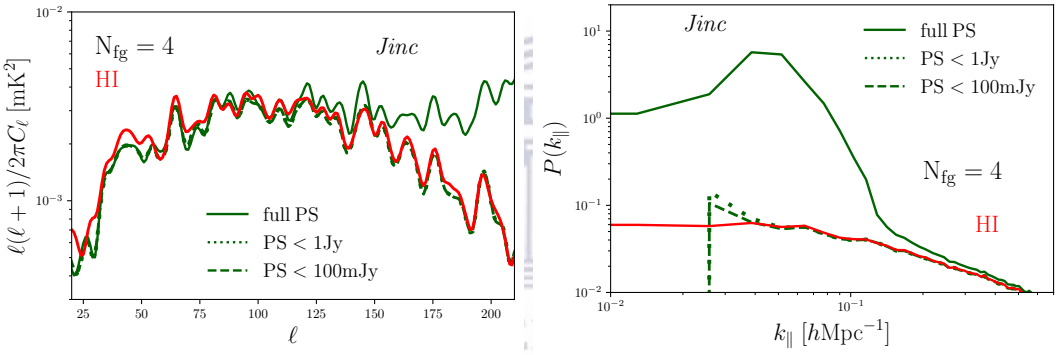

FIGURE 4.4: Same as figure 4.3, but for a Jinc beam model that has stronger side-lobes with respect to the Cosine beam. The input HI is shown in the red solid curve whilst the foreground cleaned signals for different flux density cuts are shown with different line styles.

even better the role of strong side-lobes and point sources in the cleaning procedure when examining the radial power spectrum (right panel of figure 4.4). We observe a "bump" in the low to mid $k_\parallel$ scales for the full point source model, and the HI radial power spectrum is only recovered at large $k_\parallel$. Increasing the number of foreground components $N_{\rm fg}$ has a very small effect in removing this feature. Residual foregrounds are gradually removed when point source flux cuts are applied, and the HI signal is recovered at least at intermediate and small scales.

4.3.1.3 Quantitative Comparison of Beam Models

We compare in figure 4.5 the cleaning results for the Cosine and Jinc beam model with the input signal. We add for reference the case without side-lobes, i.e. the Gaussian model. We show the estimators of equation 3.23 and 3.26. By construction, $\Delta P/P^{\rm true}(k)$ and $\Delta C_\ell/C_\ell^{\rm true}$ will be positive if there are still contaminants in the recovered signal, and negative if the cleaning is too aggressive, resulting in signal loss. We retain the worst and best-case scenario for the point source contamination, considering the full catalogue and point sources with a flux cut of 100 mJy, respectively.

The complexity of the interaction between point sources and side-lobes appears clearly from the figure: for the Gaussian case there is little dependence on the level of point source contamination while, for the two models with side-lobes, the cleaning is easier without strong point sources. Note also that, while $N_{\rm fg} = 3$ is the optimal assumption for the Gaussian case, the Cosine and the Jinc beam requires a higher number of components to be subtracted. When the Jinc beam is convolved with a sky model with no cut on point source flux, the reconstructed signal remains highly contaminated by foregrounds and increasing $N_{\rm fg}$ offers little assistance. Nevertheless, when considering faint point sources (PS $<$ 100 mJy), the angular power spectrum signal is recovered with 10% precision reaching a few % at small scales, and it is possible to recover the radial power spectrum, except for low $k_\parallel$. For the more realistic Cosine beam, side-lobes seems not to be a major limitation in the recovery of the signal even in the presence of strong point sources, and using the optimal value of $N_{\rm fg} = 4$, we can recover the HI signal within 10% precision.

A note of caution is needed: the recovered signal is systematically below the input signal for the optimal value of $N_{\rm fg}$, indicating signal loss. This effect can be, in principle, corrected by constructing a foreground transfer function, generally repeating the cleaning on a set of simulations [e.g. Switzer et al., 2015, Wolz et al., 2021]. Although we do not characterize this function in this thesis, figure 4.5 gives a raw expectation of how the signal loss varies across the number of components to subtract, the level of point source contamination and the primary beam model.

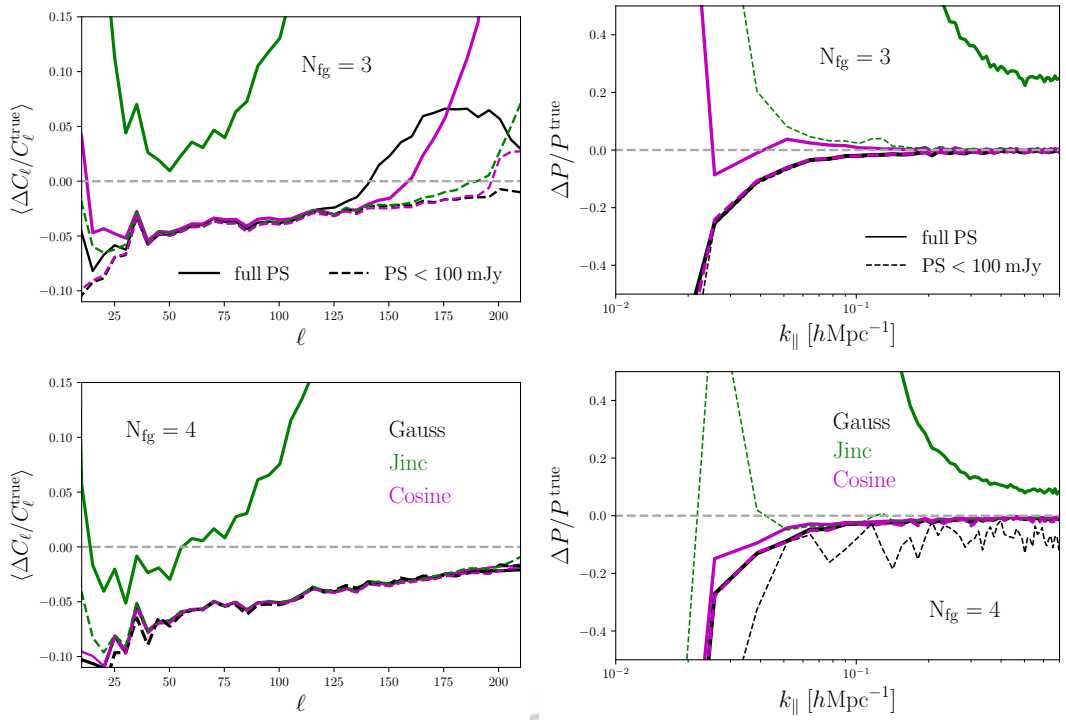

FIGURE 4.5: A quantitative comparison of the foreground cleaned angular power spectrum (left) and radial power spectrum (right) with the true input signal (see equation 3.23 and 3.26). Three types of beam models, *Gauss* (black), Jinc (green) and Cosine (magenta) were used in the beam convolution pipeline. In all cases the FWHM scales simply with λ/D. The performance of the recovery of the HI signal is evaluated for $N_{fg} = 3$ in the top panels and for $N_{fg} = 4$ in the bottom panels. The solid curve represents the case in which the full point source catalogue was considered for the simulated maps, whilst the dotted lines are the cases in which a flux cut of PS < 100 mJy was applied.

4.3.2 Frequency-dependent Beam

Up to now, we have assumed the standard λ/D dependence for the FWHM of the beam. What if we relax this hypothesis? We investigate here how our conclusions change in the presence of either the smooth or the ripple models presented in section 3.2. We note again that this frequency dependence cannot be simply absorbed into the bandpass calibration. Throughout this section, we use our most realistic model for the MeerKAT beam: the Cosine beam model.

Being a function of frequency, the effect of a non trivial FWHM is mostly visible in the reconstructed radial power spectrum. We present in the left panel of figure 4.6 results without subtracting any point sources from the foreground models. We note here, and elaborate on later, that the trends are similar for the cases with lower contamination, although the magnitude of the effect is smaller.

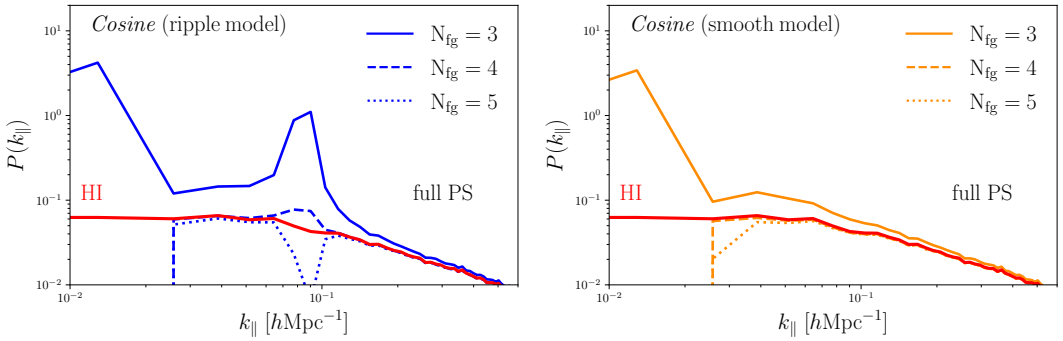

FIGURE 4.6: Comparison of radial power spectra of the recovered signal for the Cosine beam case with a FWHM with non-trivial spectral variation: the full ripple model (left panel) or only the polynomial smooth model (right panel). The input HI is shown in red whilst the reconstructed signals for different number of foreground components removed ($N_{\rm fg}$) are shown with different line styles. These results are obtained without removing the strong point sources.

The radial power spectrum for the ripple model shows a feature around $k_{\parallel} = 0.1$ hMpc^{-1} whose position depends on the period of the oscillation of the FWHM, in this case T =20 MHz. We find that the amplitude of the feature can be reduced by making the cleaning more aggressive, although we quickly start to see a strong depletion of the signal at the same $k_{\parallel}$. The HI signal is best recovered with $N_{\rm fg} = 4$. In the presence of the smooth model (right panel of figure 4.6) the feature caused by the ripple disappears and the performances of the cleaning are similar to what is seen for the Cosine beam with the standard FWHM in figure 4.3.

4.3.2.1 Quantitative Comparison of Spectral Models

In figure 4.7, we evaluate quantitatively the cleaning at $N_{\rm fg} = 4$ and present the results not only for the full point source catalogue, but also for the flux cut at 100 mJy. The HI angular power spectrum is recovered within 10% precision for the smooth model, reaching the same performances of the standard λ/D case, with both strong or low point source emission. In the ripple case instead, if no point source cut is applied, the cleaning methods struggle in recovering the small scales. Moreover, while the radial power spectrum can be successfully recovered for the standard λ/D model and the smooth model, the cleaning method is 60% off at $k_{\parallel} \sim 0.1$ hMpc^{-1} for the ripple case. For this latter case, the removal of the bright point sources transforms the feature of residual contamination into a signal loss at the same scales reaching up to 20%.

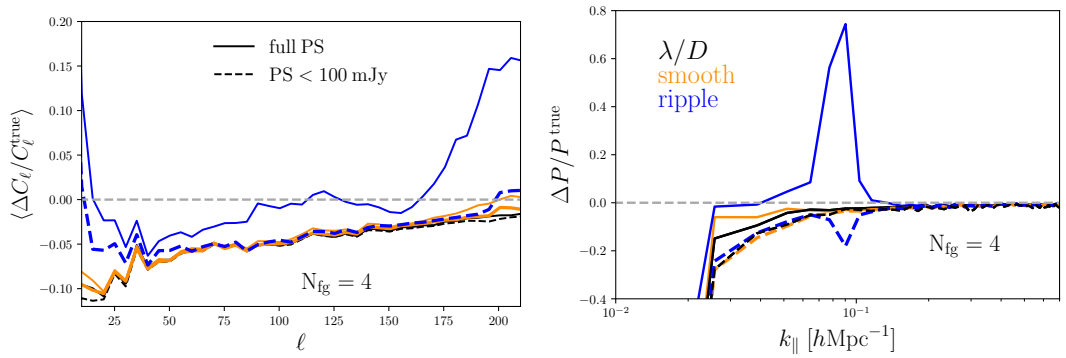

FIGURE 4.7: A quantitative comparison of the foreground cleaned angular power spectrum (left) and radial power spectrum (right) with the true input signal (see equation 3.23 and 3.26). Results for $N_{\rm fg} = 4$ are compared for the different FWHM models: λ/D in black, the smooth model in orange and the ripple model in blue. Results are shown for the worst and best-case scenario for the point source contamination, considering the full catalogue (solid lines) or only point sources with a flux cut of 100 mJy (dotted lines), respectively.

4.3.2.2 Impact of Diffuse Emission

Up to now we have focused our discussion on the interaction between the structure of the beam and the level of point source contamination. We remark that our foreground model includes not only point sources but also free-free and synchrotron emission. A legitimate question is then how much the cleaning performance is impacted by the ripple model in the presence of a diffuse foreground component. We consider a simplified foreground emission comprising only Galactic synchrotron and report in figure 4.8 the results for the HI radial power spectrum. Similarly to figure 4.6 and 4.7, the cleaning struggles around the scale of the FWHM frequency oscillation, although its impact is smaller. Interestingly, the same exercise performed with the Gaussian beam convolution yields similar results, underlying that the effect of the persistent ripple is important also when no side-lobes are present. Note that if the foregrounds were completely smooth across the sky, no effect would be observed.

4.3.3 Applying Re-smoothing

To alleviate the effect of the frequency ripple in the cleaning process, that, as seen in section 4.3.2, compromises the recovery of the radial power spectrum, one could attempt a re-smoothing of the maps to a common resolution. Re-smoothing is a standard technique adopted in Intensity Mapping to prevent systematics and artifacts in the data: all maps are de-convolved and re-smoothed to a common angular resolution, slightly lower than

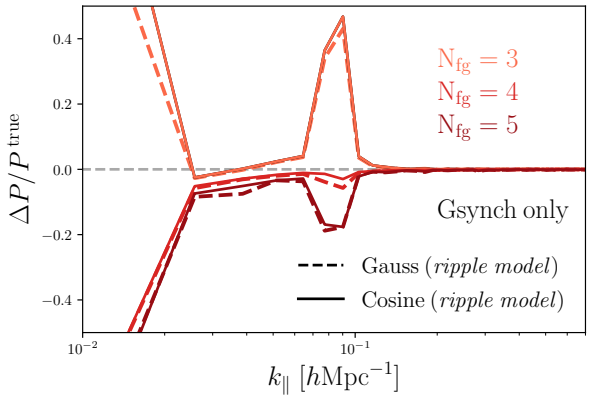

FIGURE 4.8: The estimator of equation 3.26 applied after cleaning an input simulation that includes *only* synchrotron emission and the signal, convolved with a Cosine beam (solid lines) or a Gauss beam (dashed lines), in both cases assuming the frequency ripple model. Different colors show results obtained with different values of N_{fg}. Although the effect is similar, note that for $N_{fg} = 4$ it is only at few % level while it reaches 60% in figure 4.7 in the presence of strong point sources.

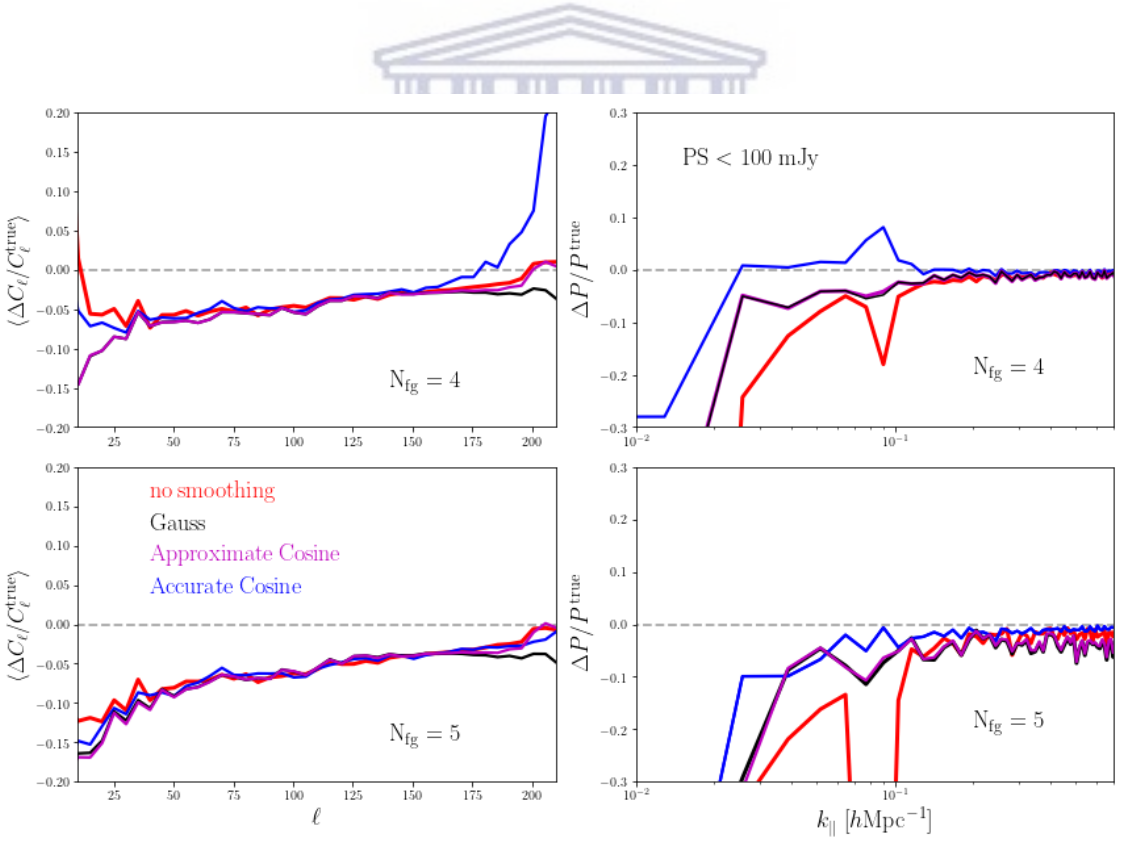

FIGURE 4.9: We compare the signal reconstruction performances for the Cosine beam case with a FWHM evolving with the ripple model, when our standard cleaning procedure is applied and maps are not smoothed (red), with the one obtained after performing a re-smoothing with a Gaussian (black), an approximated cosine (magenta) or a accurate cosine (blue) beam (see text for details), at fixed N_{fg}. We show results for the case with only point sources fainter than 100 mJy, for both the angular power spectrum (*left*) and the radial power spectrum (*right*).

the worst one in the data (see for example Anderson et al. [2018]). In other words, a map measured at frequency ν and thus convolved with a telescope beam with a certain FWHM can be adjusted to a (lower) frequency resolution given by the FWHM of the beam at a lower frequency ν_0. This is achieved in spherical harmonics space via a simple multiplication of the $a_{\ell m}^{\rm mes}(\nu)$ of the measured map with the ratio of the spherical harmonic representations $b_\ell(\nu_0)$ and $b_\ell(\nu)$ of the two beams

$$a_{\ell m}^{\rm rs}(\nu) = a_{\ell m}^{\rm mes}(\nu)\frac{b_\ell(\nu_0)}{b_\ell(\nu)}. \tag{4.1}$$

Although this procedure is exact for a full sky map with known beam and no noise, its application to smaller sky patches requires some precautions due to numerical instabilities, which we briefly introduce here and discuss in more detail in section 4.4.

i) We start by *apodizing* our mask (e.g. smoothing the edges) to have less "ringing" entering the computation of the $a_{\ell m}^{\rm mes}(\nu)$;

ii) assuming a certain beam model, we compute the $b_\ell(\nu)$ of the frequency of interest and the $b_\ell(\nu_0)$ of the common frequency we want to resmooth to;

iii) we regularize the ratio $b_\ell(\nu_0)/b_\ell(\nu)$ imposing a cut-off $\ell_{\rm cut}$ at small scales;

iv) we apply equation 4.1 and obtain a map from $a_{\ell m}^{\rm rs}$;

v) we enforce a conservative mask on the re-smoothed map in order to avoid the inclusion of pixels at the edges of the old mask.

We apply this procedure to both the pessimistic case, where no point source flux cut has been applied (full PS), and the best-case scenario where their emission has been kept lower than 100mJy. All our maps have been initially convolved with the Cosine beam with a peculiar frequency dependence (the ripple of equation 3.6). The deconvolution (dividing by $b_\ell(\nu)$) should be done with a beam that we believe closest to the real one. We then consider three distinct cases for the deconvolution step, assuming increasing knowledge on the beam.

1. A Gaussian beam with a FWHM scaling proportionally to λ/D as in equation 3.5. This is the simplest assumption;

2. A Cosine beam. Since the presence of side-lobes is a known feature of the MeerKAT beam, we can suppose that we can describe them with this model. We instead assume no knowledge of the frequency behavior of the FWHM and rely on the λ/D approximation. We call this case *approximate Cosine*;

3. A Cosine beam with ripple model, assuming exact knowledge of the beam side-lobes and the frequency dependence. We call this case *accurate Cosine*. Note that, even in this last case where we deconvolve using the same b_ℓ used for the original convolution, the cancellation will not be perfect due to the mask and the noise (although the noise effect should be negligible on the foreground map).

The subsequent smoothing (convolution) to a common resolution (using $b_\ell(\nu_0)$), could be done with different beam shapes. A common solution is to use a Gaussian. In this work, we use instead the same beam that was used in the deconvolution step above.

In figure 4.9, we compare the performance of the cleaning algorithm using maps deconvolved/re-smoothed with the three different procedures. We present both the angular power spectrum and the radial power spectrum for $N_{\rm fg} = 4$ and 5. The reference, for comparison, is the case discussed in section 4.3.2. We present, for simplicity, only the case with low point source contamination. This is a conservative choice since we see even more improvement due to the re-smoothing procedure in the case of strong point sources. Note that we apply exactly the same procedure to the HI only signal so that we can make a fair comparison that should be mostly dependent on the cleaning and not on extra effects from the power spectrum calculation itself (such as the window function).

When we subtract $N_{\rm fg} = 4$ components, the C_ℓ are similarly reconstructed for all cases within a $\sim 5\%$ precision for most scales, although some residual small-scale contamination is found for the *accurate Cosine* re-smoothing. This residual contamination disappears for $N_{\rm fg} = 5$, and we find agreement between the different procedures. We note again the angular power spectrum diagnostic is always negative, uncovering a systematic signal loss.

More interesting for our purposes, is the effect on the radial power spectrum. The *accurate Cosine* re-smoothing deals quite efficiently with the effect of the ripple, reaching a precision always better than 10% for a large range of $k_\parallel$, for both $N_{\rm fg} = 4$ and 5. The re-smoothing procedure with the Gaussian or *approximate Cosine* beam erases the presence of the ripple and reconstructs the $P(k_\parallel)$ much better than reference case.

In summary, figure 4.9 suggests that, in the presence of non trivial beam frequency dependent effects, an accurate knowledge of the beam would allow a re-smoothing procedure that should improve the quality of the foreground cleaning in the radial direction and alleviate signal loss. A less accurate re-smoothing would also be enough to retrieve the signal with good accuracy. Probably the best approach will be to include the beam in the mapmaking process (see for example Tegmark [1997]). This is, however, something that will be computationally demanding and we would like to explore in future work. Still, we will never know the beam exactly and a smoothing kernel will always be useful. The choice of smoothing kernel should be validated against simulations in order to quantify any effect on the signal power spectrum.

4.4 Map Deconvolution

In section 4.3.3, we have discussed the performance of the foreground cleaning on maps that have been deconvolved to a common resolution. Here we present in more details some of the steps of the adopted procedure.

Apodization. Since we are interested only in how well we reconstruct the signal and not in the shape of the signal itself, throughout the paper we have computed the angular power spectrum of the maps using simply equation 3.22 and correcting for $f_{\rm sky}$ (see also discussion in section 3.4.3). Nevertheless, to re-smooth a map using equation 4.1, having well behaving $a_{\ell m}$ becomes important. To this purpose we do not compute them directly using the original sharp mask but we use an apodized mask. Apodization is a standard procedure for example in CMB studies, assuring that the mask does not sharply pass from one to zero but that there is a smooth transition at the borders. Apodized masks alleviate the *ringing* in the power spectrum, and thus in the $a_{\ell m}$. To apodize, we smooth the mask with a Gaussian with a FWHM of a few degrees, replacing the zeros with negative numbers before the smoothing to obtain a better behaviour around the edges of the mask. We choose this negative value ensuring that, in the sky patch defined by the initial mask, there are no negative values. The final apodized mask is obtained just by multiplying the smoothed one and the original mask.

Beam b_ℓ ratio. The other important ingredient for equation 4.1 are the spherical harmonics coefficient of the beam decomposition on the sphere, the b_ℓ of equation 3.15. As

discussed in section 3.4, for a symmetric beam these coefficients are only function of ℓ and are real. We report in figure 4.10 the shape of the b_ℓ for two different frequencies and comparing the cosine and the Gaussian beams. Due to the shape of the cosine main lobe (see figure 3.7), the b_ℓ drops faster after $\ell \sim 200$.

In section 4.3.3 we discussed different cases for re-smoothing. In the first case, we have used a Gaussian both to de-convolve and re-convolve, i.e. the ratio of equation 4.1. In harmonic space a Gaussian function remains Gaussian, so we have an easy analytical form for the ratio: it corresponds to the b_ℓ of a Gaussian beam with $\Delta\theta^2 = \Delta\theta(\nu_0)^2 - \Delta\theta(\nu)^2$. Note that, as discussed in Anderson et al. [2018], it is better to choose ν_0 to be slightly lower than the minimum frequency used (we use 890 MHz). For a better comparison, instead of using the analytical form we compute numerically the ratio between the two b_ℓ. The finite machine precision inevitably introduces spurious behaviors at small scales ($\ell \sim 370$ for our frequency range, see top panel of figure 4.11). To avoid the effect of these artifacts, in case of numerical ratio, one can perform a cut in ℓ and artificially put at zero the ratio for smaller scales. We compare the numerical and the analytical case in the top panel of figure 4.11, showing the loss of information imposed by the $\ell_{\rm cut}$.

For the re-smoothing case where we assume that the beam could be described with a Cosine model, the analytical ratio is not straightforward to derive so we always revert to the numerical ratio. In the bottom panel of figure 4.11, we show how the ratio behaves with this beam model. Due to the differences in the b_ℓ seen in figure 4.10, the effect of finite machine precision enters at larger scale so we perform the cut at $\ell \sim 250$.

One could also try to deconvolve the maps from the effect of the Cosine beam and then use a Gaussian re-smoothing to alleviate the effect of the side-lobes. In the bottom panel of figure 4.11 we report also this case, noting however that a very large Gaussian beam is required to avoid spurious numerical behaviour of the ratio. If on the one hand this could be of help in suppressing artifacts in the data, on the other hand we are forced to throw away more information. We have thus decided to use the cosine beam also for re-smoothing.

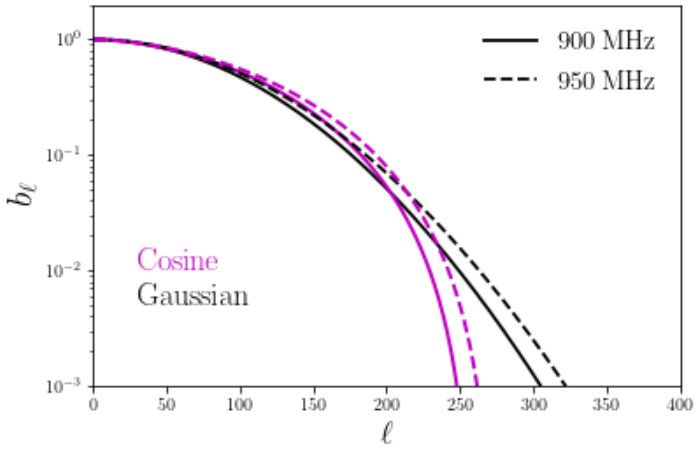

FIGURE 4.10: Spherical harmonic coefficients of the beam decomposition b_ℓ (see equation 3.16) for the cosine beam (magenta) and the Gaussian beam (black). The lower frequency of L-band 900 MHz is plotted with a solid line while 950 MHz is plotted with a dashed line.

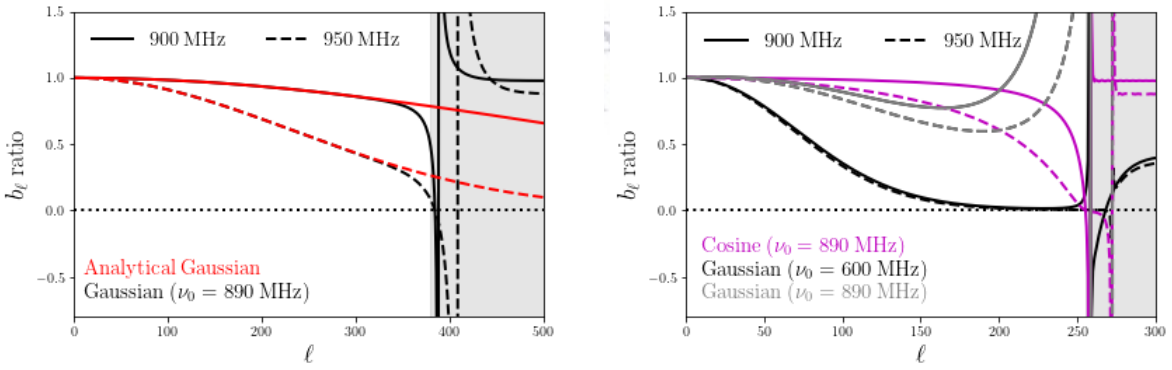

FIGURE 4.11: (*Left panel*) The ratio between the b_ℓ for a Gaussian beam at 900 MHz (solid black) or at 950 MHz (dashed black) with the lower frequency Gaussian beam at $\nu_0 = 890$ MHz. After $\ell \sim 380$ the ratio is not well behaved so we artificially set to zero the smaller scales. This approximated procedure is compared with its the analytical solution for both 900 MHz (solid red) and at 950 MHz (dashed red). (*Right panel*) The ratio between the b_ℓ for a cosine beam at 900 MHz (solid magenta) or at 950 MHz (dashed magenta) with the lower frequency cosine beam at $\nu_0 = 890$ MHz. In this case after $\ell \sim 240$ the ratio is not well behaved so we consider zero all the smaller scales. We report also the behaviour of the ratio when we deconvolve with a cosine beam but re-convolve with a Gaussian. If we use the same ν_0 as before, i.e. 890 MHz the ratio start a fast growth around $\ell \sim 200$ (grey lines), while to obtain a reasonable value for the ratio we are obliged to throw away a lot of information using a very large beam, i.e. a very low ν_0 (black lines).

4.5 Summary

In this chapter we analyzed first the effect of side-lobes in presence of strong point sources. Cleaning is severely impacted by the large side-lobes of the Jinc beam model and residual foreground contamination improved with the subtraction of a larger number of components (up to $N_{\rm fg} = 5$) only if the strongest sources are removed. Fortunately, the Cosine model, which should be a more realistic description for the MeerKAT beam, shows a good performance if 4 modes are subtracted. Note that $N_{\rm fg} = 3$ is enough instead for the Gaussian beam case. We recovered the HI signal within 10% on large angular scales, while reaching few % accuracy at small scales. We find, however, that the foreground subtraction always tends to over-clean the signal. The HI radial power spectrum reconstruction is good at intermediate and small radial scales but fails at low $k_{\parallel}$ as expected, due to the removal of the smooth foreground contamination. This effect at large scales is known, for example, to bias primordial non-Gaussianity studies [e.g. Cunnington et al., 2020]. We found that the performances of the cleaning at low $k_{\parallel}$ depend both on the level of point source contamination and on the amplitude of the beam side-lobes.

We then explored the effect of a realistic frequency dependent beam on the cleaning efficiency. In the literature, the beam FWHM is generally assumed to follow a simple scaling (λ/D). Following the fits of Asad et al. [2021], we constructed two models for the FWHM of increasing complexity using the realistic Cosine beam (see figure 3.8): a polynomial in frequency (i.e the smooth model) and a frequency oscillation on top of this latter (i.e the ripple model). We found that the smooth model does not impact the performance of the cleaning procedure. On the contrary, the coupling of the foreground, and in particular of the strong point sources, with the ripple feature, affected the cleaning more strongly. The effect is more visible for the radial power spectrum, where the $k_{\parallel}$ scales corresponding to the frequency of the ripple are biased. It is interesting to note that this effect is also present when only the Galactic synchrotron is included, showing that the contamination will still be present at some level even after aggressive point source removal. Moreover, this contamination also comes from the frequency fluctuations in the main lobe of the beam and not only the side-lobes. This type of oscillatory behavior of the FWHM of the beam, if not recognized and treated, could contaminate the cosmological analysis, such as the reconstruction of the imprint of the Baryon Acoustic Oscillations on the 21 cm power spectrum.

In the next chapter, we explore in an even more realistic setup the effect of the beam, constructing a pipeline to apply the beam in map space instead of the spherical harmonic approach used in this chapter.

Chapter 5

Effect of the Beam through Map-space Convolution

In the previous chapter we discussed the implications for the foreground cleaning procedure of a realistic primary beam that encodes some of the measured characteristics of the MeerKAT dishes. We analysed the effect of side-lobes and of a non-trivial frequency dependence of the FWHM of the beam.

Nevertheless, there are other potential systematics that simulations should include, for example, the effect of non-homogeneous noise or of the 1/f noise, both due to the scanning strategy. Moreover, we should consider the effect of a heavy RFI flagging that can have consequences on the cleaning procedure, similar to the ones of a mask with fringes and holes, causing spurious features at power spectrum level. When observing for long periods there is also the risk of a modification with time of the dishes and this effect can have consequences for the final co-adding of the single-dish maps.

Another important systematic effect to take into account is polarization leakage, i.e. the contamination of the pristine HI signal coming from the polarized sky. To this end, we created and provided in the work of this thesis and in Ngobese [2018] a point source catalogue (see chapter 3), containing also the polarized information and a code to create Q and U maps from it. In assessing the potential level of polarization leakage in the intensity mapping observations carried out with MeerKAT, the beam is again central. Indeed, the final leakage will depend on the combination of the polarized sky and on the

off-diagonal Mueller matrix term of the beam. In the results presented in the previous chapters, although we have implemented a realistic MeerKAT beam $B(\theta, \phi, \nu)$ with a side-lobe structure and a non-trivial frequency dependence, we have always assumed the beam to be symmetric with respect to its pointing direction, i.e. we have neglected any ϕ dependence.

Under this assumption, the beam $B(\theta, \nu)$ can be easily described in spherical harmonic space and its symmetry provides a way to avoid directly computing the convolution integral described in equation 5.3, instead making use of the convolution theorem as shown in equation 3.17. Computing a product in harmonic space instead of an integral in map-space makes the sky model convolution with the telescope beam much faster. Approximating the beam as symmetric can be justified assuming that the same area on the sky will be scanned more than once for each observation and from different directions, mitigating the effect of asymmetries in the beam. The symmetrization of the beam remains, however, an approximation and its impact has not been explored in the context of intensity mapping and foreground cleaning.

In this chapter, we build a framework to explicitly compute the impact of a realistic beam model with asymmetries on the foreground cleaning results. We make use again of the beam pattern of the EIDOS package [Asad et al., 2021] and instead of imposing an ad hoc frequency ripple as in section 3.2.1, we will include in our results the natural frequency structure of the EIDOS beam. To this end, the convolution integral must be solved directly. In section 5.1, we describe how to project the 2-D grid representation of the EIDOS beam onto the sphere using the HEALPix format. A description of the map-space convolution method is provided in section 5.2.

In section 5.3 we apply the map-space convolution to a sky model constructed in the same manner as for the previous chapter, thus including the various foreground components and the HI signal. We proceed with the foreground cleaning on these maps, with the aim of testing the effect of the EIDOS beam: its frequency dependency and its asymmetries.

5.1 Mapping the EIDOS 2-D Beam into HEALPix Format

The EIDOS package can be used to generate 2-D beam patterns within a frequency band ranging from 900 to 1050 MHz, using the following command:

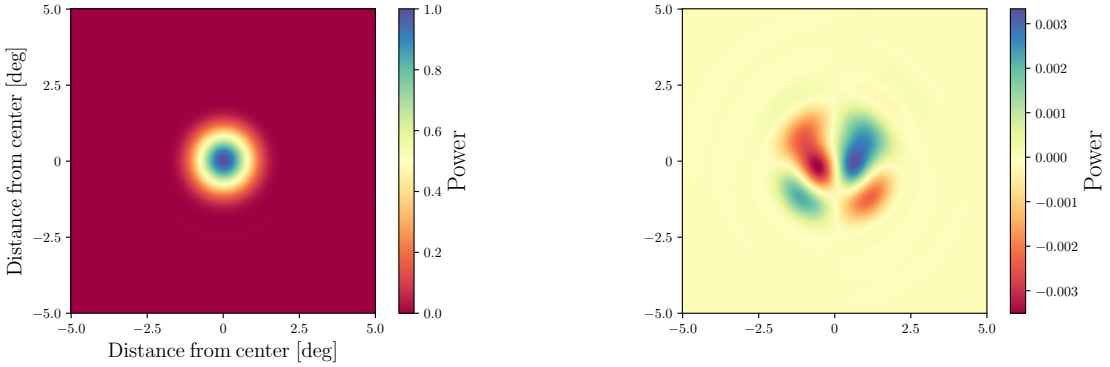

FIGURE 5.1: Two EIDOS 2-D beams at 900 MHz. On the left is a Stokes I beam (a diagonal term of the Mueller matrix) and on the right is a Stokes Q $\rightarrow$ I beam, encoding leakage from Stokes Q to I.

```
eidos -p 512 -d 10 -f 900 1050 -S I
```

The above instruction is interpreted as follows: input parameter -p indicates that each 2-D beam generated must contain 512 pixels per side, -d specifies the diameter of the beam per side which is 10 degrees, -f is the frequency range (with a default frequency resolution of 1 MHz), and -S is the stokes parameter of the beam requested, which can be Stokes I, Q, U or V. This command produces a total of 150 beam images.

The type of primary beam that we use in the main analysis, is a stokes I beam. However for the purposes of testing our pipeline, we use a non-symmetric beam so that it is easier to visually pinpoint the asymmetries of the beam and compare between the 2-D image and the HEALPix representation of the beam. Figure 5.1 shows the EIDOS Stokes I beam (*left*) and Stokes Q $\rightarrow$ I beam (*right*), an off-diagonal term which shows leakage from Stokes Q to I.

To map the 10x10 degree beam we superimpose a 2-D coordinate system centered on the peak of the image at $(\theta, \phi) = (0, 0)$, such that the angular extent in all four quadrants ranges from 0 degrees (centre) to 5 degrees (edge), as indicated by the axis ranges in figure 5.1.

From this we compute the radial distance and azimuth angle using the small-angle approximation as follows:

$$\theta = \sqrt{(\theta_x)^2 + (\theta_y)^2}, \tag{5.1}$$

$$\phi = \arctan\left(\frac{\theta_y}{\theta_x}\right), \tag{5.2}$$

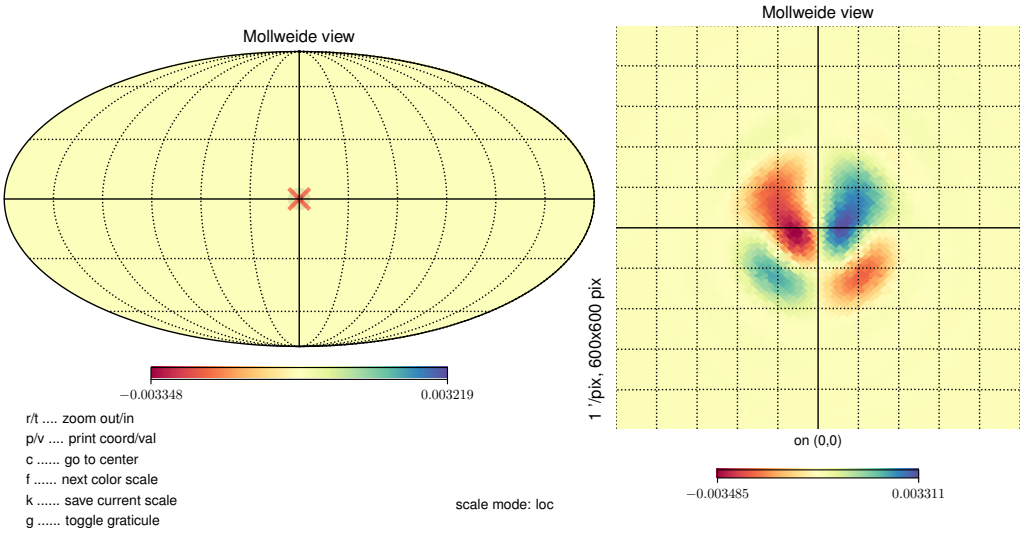

FIGURE 5.2: The final HEALPix format of the original EIDOS beam. We show here an example of the Q → I beam to give emphasis to the asymmetric structure of the beam. This can be compared with the right panel of figure 5.1.

where θ_x and θ_y are the elevation and azimuth angles of the beam in degrees, with the ranges $0 < \theta \leq \pi$, $0 < \phi \leq 2\pi$. While θ is straightforward to compute using equation 5.1, some care is needed to ensure the right sign for ϕ when using the arctan function. In particular, the following conditions must be implemented depending on the quadrant in which (x, y) falls:

- if $x \geq 0$ and $y \geq 0$: $\phi = \arctan\left(\frac{\theta_y}{\theta_x}\right)$ (1st quadrant)
- if $x < 0$ and $y \geq 0$: $\phi = \arctan\left(\frac{\theta_y}{\theta_x}\right) + \pi$ (2nd quadrant)
- if $x < 0$ and $y < 0$: $\phi = \arctan\left(\frac{\theta_y}{\theta_x}\right) + \pi$ (3rd quadrant)
- if $x > 0$ and $y < 0$: $\phi = 2\pi - \arctan\left(\frac{|\theta_y|}{\theta_x}\right)$ (4th quadrant)

Once the (θ_x, θ_y) coordinates are correctly transformed into the polar (θ, ϕ), we can simply use the HEALPix routine ang2pix(θ, ϕ) to obtain the correct pixel position in which to store the associated beam value on a map. We report in figure 5.2 the result.

5.2 Map-space Convolution with Stokes I Beams

We recall that to implement map-space convolution we need to explicitly compute

$$\tilde{T}_{\rm sky}(\nu, \theta, \phi) = \int \left[\hat{R}(\theta, \phi) B\right](\nu, \theta', \phi')\, T_{\rm sky}(\nu, \theta', \phi')\, d\Omega', \tag{5.3}$$

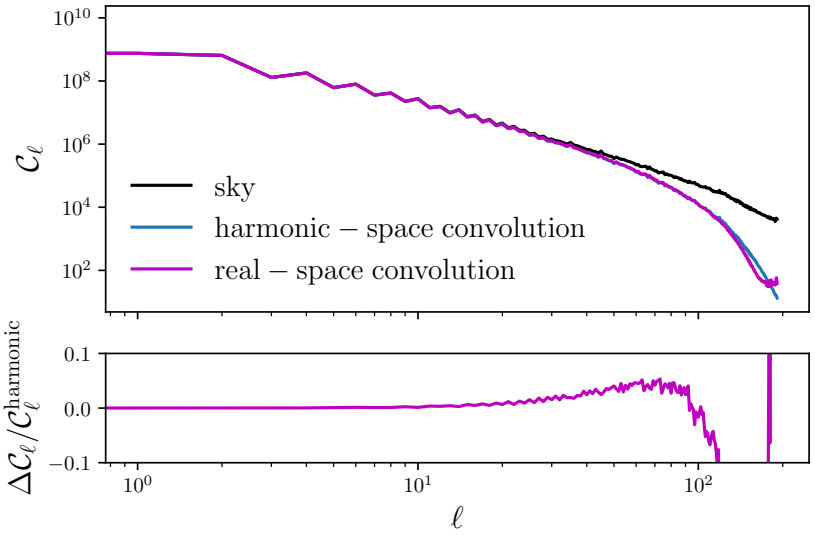

FIGURE 5.3: Full-sky angular power spectrum comparison. The *black* curve is an input sky model including only synchrotron. The *blue* line is the harmonic space result while the *magenta* line is the map-space convolution. The considered beam is the Stokes I beam in the left panel of figure 5.1. In the lower panel we show the relative difference between the real and the harmonic result. The two methods give the same results up to small scales. Note that to speed up the computation we have used in this example a low $N_{\rm side}$.

where $d\Omega' = \sin(\theta')d\theta' d\phi'$ and $\hat{R}$ is the operator of finite rotations such that $\hat{R}B$ is the rotated beam in the (θ, ϕ) direction [e.g. Wandelt and Górski, 2001].

In order to apply equation 5.3 we need to implement a rotation operation on the sky. We make use of the HEALPix routine rotateDirection() and, given a pointing direction defined by (θ_0,ϕ_0), we compose a rotation around the y-axis for θ_0 and a rotation around the z-axis for ϕ_0. With this, we can point the pixelized beam $B(\nu, \theta, \phi)$ to the selected direction. The value of the convolved sky in that point will then be the sum of the pixel-by-pixel multiplication between the sky (non-rotated) and the rotated beam.

To check that our implementation is correct, we compare the standard harmonic space decomposition used in the previous chapters with the one described here. In figure 5.3 we show the angular power spectrum of an original sky map and the $\mathcal{C}_\ell$ of the convolved map with the two different techniques, at a given frequency. We see that the agreement is very good up to small scales. This confirms that our map-space convolution pipeline is correct.

The next section investigates the effect of the map-space convolution on the foreground cleaning performances. We stress that the EIDOS beam can be pixelized on the sphere

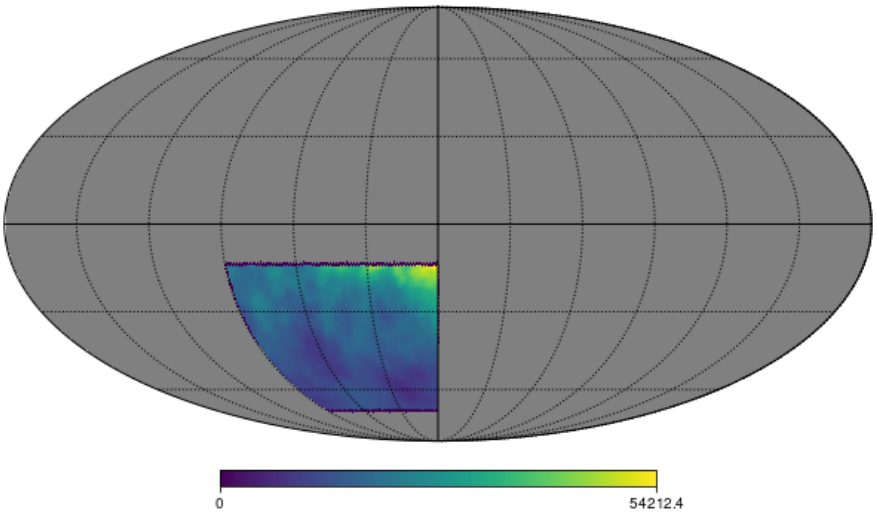

FIGURE 5.4: An example of the output of our map-space convolution procedure. In the selected sky patch we simulate synchrotron emission and we model the effect of the beam, rotating a pixelized version of the EIDOS beam on each pixel. One can see how the structures are smoothed by this effect.

at each frequency in order to naturally include the effect of the frequency ripple. We are, however, making a simplification in our analysis. Even when considering smaller sky patches, we are not implementing a proper scanning strategy but we are simply visiting each pixel in the *ring* ordering in which they are stored by HEALPix.

We consider the same case analyzed in chapter 3, restricting our analysis to a southern hemisphere sky patch. The computation of the map-space convolution will be faster due to the smaller number of pointings compared to the pixels of the entire map (using $f_{\rm sky} \lesssim 0.1$, we reduce the number of pixels by a factor 10). In figure 5.4 we show an example of the beam convolved sky in the chosen patch.

5.3 Map-space Convolution Effect on Foreground Cleaning

We begin by creating Stokes I beams within the specified frequency band and resolution. For the sky model we consider a sky consisting of the cosmological HI signal, galactic synchrotron, galactic and extra-galactic free-free emission and our point sources. Corresponding pixelized sky maps are created per channel. The map-space convolution is computed at each channel and the corresponding instrumental noise is added.

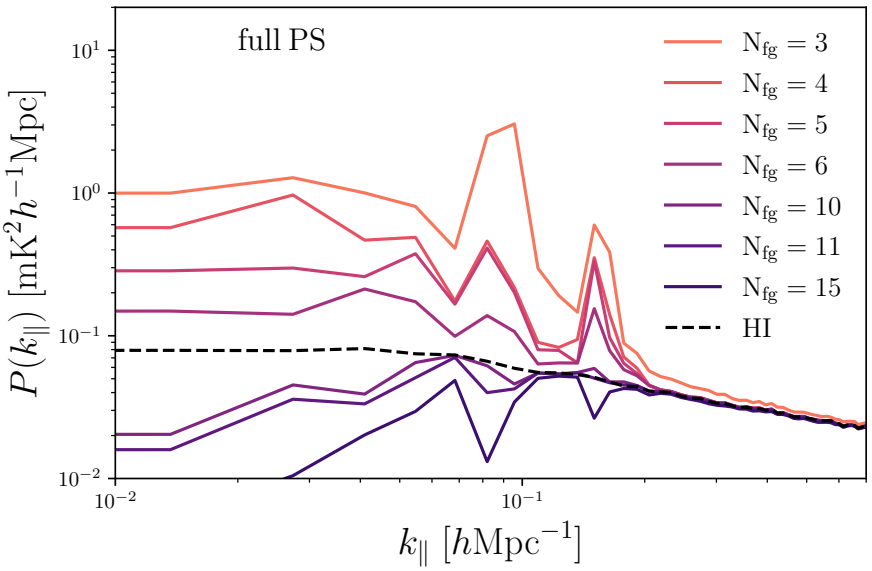

FIGURE 5.5: The results for the foreground cleaned radial power spectrum described in section 3.4.3. We consider the full point source catalogue for the sky model (full PS). We present results using a variety of values for the number of removed components $N_{\rm fg}$. The retrieved signal is compared to the input HI signal (dashed).

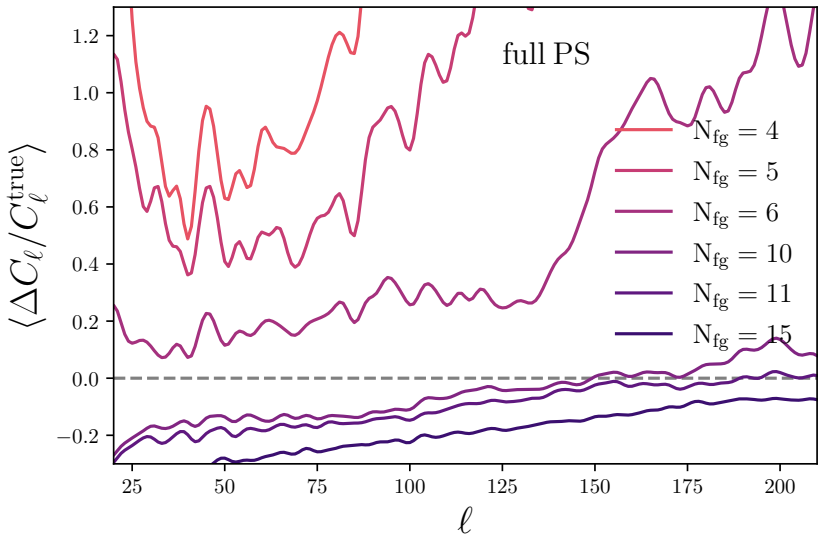

FIGURE 5.6: The results for the foreground cleaned angular power spectrum shown as the mean difference over frequency between the input signal and the recovered one (see section 3.4.3). We consider the full point source catalogue for the sky model (full PS). We present results using a variety of values for the number of removed components $N_{\rm fg}$. The retrieved signal is compared to the input HI signal (dashed).

We make use of the Principal Component Analysis (PCA) cleaning method described in detail in chapter 3 and applied already in chapter 4. Note that we focus on the same sky patch analyzed before.

We show in figure 5.5 the reconstructed radial power spectrum for the conservative assumption of the full point source catalogue. As discussed in the previous chapter, the effect of the beam is less important if we assume that strong point sources can be removed. Here, however, we are interested to see this effect in the stronger case, in order to fully appreciate it. We present results comparing cases where a different number of components have been subtracted, i.e. where we used different values of $N_{\rm fg}$. For all values of $N_{\rm fg}$ we can recognise the peak structure seen in the previous chapter in the presence of the ripple. These results prove that the oscillation of the FWHM is present in the data and not only enforced by our modelling with a sinusoidal. We recall that the frequency of the ripple was tuned to mimic the interaction between the primary and secondary reflector of MeerKAT with a period of about 20 MHz. In figure 5.5 we can see that the structure of the EIDOS beam in frequency is even more complex and there is a secondary peak at a slightly higher $k_{\parallel}$.

As expected, in the presence of this realistic beam the cleaning is more demanding and for $N_{\rm fg} = 6$ we still have strong residual foreground in the final maps. Pushing the number of components to subtract to 10 leads to an excess of cleaning at low $k_{\parallel}$ and to the residual peak transforming into an over-cleaned hollow, similarly to what is shown in figure 4.6.

For completeness we show in figure 5.6 the effect of the map-space convolution on the angular power spectrum. Similarly to figure 5.5, it is noticeable that for $N_{\rm fg} = 6$ we still have strong residual foreground. For $N_{\rm fg} = 10$, the angular power spectrum loses power at large scales but it is within 20% of the expected value. When using a higher value for the number of foreground components the reconstruction is compromised more and more and too much signal is subtracted together with the foregrounds.

These results are obtained by assuming that strong point sources in the field cannot be removed. This is a conservative assumption that, similarly to what is done in chapter 4, can be relaxed by assuming point sources above a certain flux threshold will be peeled from the final map. We show in figures 5.7 and 5.8 the analogue of figures 5.5 and 5.6, considering only point sources below 100 mJy at 1.4 GHz.

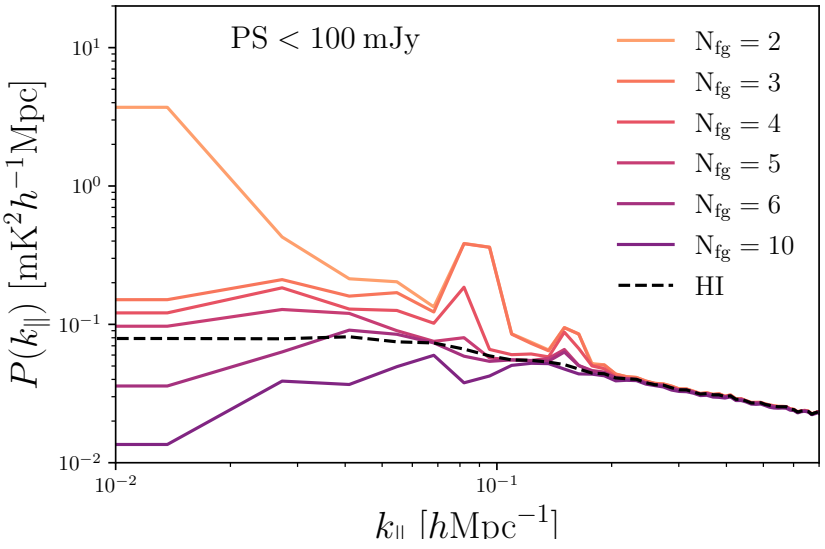

FIGURE 5.7: The results for the foreground cleaned radial power spectrum described in section 3.4.3. We consider only point sources in the catalogue with flux lower than 100 mJy (PS $<$ 100 mJy). We present results using a variety of values for the number of removed components $N_{\rm fg}$. The retrieved signal is compared to the input HI signal (dashed).

The lower level of point sources in the map reduces the residual peaks in the line of sight power spectrum and a decent cleaning is achieved with a lower number of subtracted components, as seen also in the results in chapter 4. The angular power spectrum is also better recovered for a lower $N_{\rm fg}$ and over a larger range of scales. We note however a feature at $\ell > 200$ that could be due to numerical limitations in our pipeline.

5.4 Summary

In this chapter we sought to investigate the full effect on foreground cleaning of non-trivial frequency dependence and asymmetries in the beam, by computing the map-space convolution, using the EIDOS beam images. In order to achieve this goal, a framework was constructed in section 5.1 to map the 2-D beam onto a HEALPix sphere.

Computing the map-space convolution means that the convolution integral described in equation 5.3 must be computed directly, instead of making use of spherical harmonic transforms as implemented in chapter 4. Section 5.2 describes how this computation is conducted. In section 5.3 we present our findings on the impact of the non-trivial frequency dependent effect and ways to minimize this complex feature.

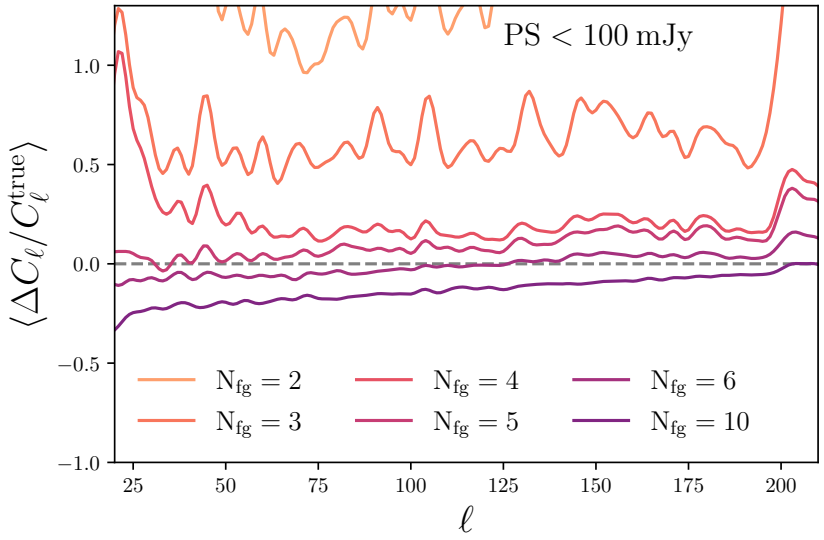

FIGURE 5.8: The results for the foreground cleaned angular power spectrum shown as the mean difference over frequency between the input signal and the recovered one (see section 3.4.3). We consider only point sources in the catalogue with flux lower than 100 mJy (PS $<$ 100 mJy). We present results using a variety of values for the number of removed components $N_{\rm fg}$. The retrieved signal is compared to the input HI signal plotted (dashed).

Chapter 6

Conclusions

The ΛCDM model is the most successful cosmological model to date, in that is able to accurately explain cosmological observations. Using this model the evolution of the Universe can be traced from seconds after its formation right through its present age, 14 billion years later. Although various challenges do exist with the model, such as the Hubble tension highlighted in chapter 1, it is still the best model currently available, and it continues to be tested and validated. The central theme of this thesis, is to explore the issues with using the new HI IM observational technique described in chapter 2, that makes use of neutral hydrogen as a tracer with the ultimate goal of measuring the HI power spectrum.

The SKA precursor MeerKAT telescope in South Africa is a precious test ground for the HI IM techniques and has the potential to provide competitive constraints on cosmological observables using the single dish mode [Santos et al., 2017]. Component separation techniques such as PCA and FastICA are a vital step to extract the HI signal from the bright foreground emission. These techniques, although widely and successfully tested in literature, have been applied mostly under the assumption of symmetric primary beams.

To be successful in detecting the auto-correlation HI signal with the MeerKAT or future SKAO radio telescope, a proper understanding of the various systematics such as primary beam frequency effects will be important. The main focus of the thesis was to: 1) study the impact of non-Gaussian beams (e.g. beams with side-lobes) on extracting the HI signal with single dish MeerKAT observations; 2) investigate the frequency dependent

effects on such extraction, and 3) also test the impact with an even more realistic beam with angular asymmetries and more complex frequency dependence.

In chapter 3 the modelling process of each component is discussed and in chapter 4 a detailed analysis of the impact of a more realistic beam on the recovery of the signal is presented. The sky model used includes a new realistic point source full-sky catalogue constructed from the combination of NVSS and SUMSS data and the S^3 simulations, which allows to test the effect of strong point sources on the signal.

The analysis began with a comparison of the beam convolution method described in section 3.4 with the publicly available CRIME software package, so as to set a benchmark. The results obtained from the exercise showed that indeed the beam convolution method was able to reproduce the same results as those obtained with CRIME.

We explored both the presence of side-lobes and the frequency-dependent effects on the FWHM of the beam. We simulated MeerKLASS-like instrumental noise levels and selected a sky patch in the southern hemisphere covering $\sim 9\%$ of the sky, avoiding strong contamination from the Galactic plane.

The effect of side-lobes in the presence of strong point sources was analyzed and the following findings were made. The large side-lobes present in the Jinc model impact the foreground subtraction severely. We find that by subtracting a large number of modes (up to $N_{\rm fg} = 5$) and removing the strongest point sources, the residual contamination is reduced. A more positive outcome is obtained when using the Cosine model, which is a more realistic description of the MeerKAT beam. We observe that cleaning is greatly improved, requiring 4 modes for the HI signal to be recovered. In the case of the Gaussian beam, $N_{\rm fg} = 3$ is sufficient.

On large angular scales, we find that the HI signal is recovered within 10%, while accuracy reaches a few % at small scales. We also find that cleaning has a propensity of over-cleaning the HI signal. Unsurprisingly, the reconstruction of the HI signal in the radial power spectrum fails at low $k_{\parallel}$, due to the removal of the smooth foreground contamination. This is a known effect, already observed in studies such as those of primordial non-Gaussianity [e.g. Cunnington et al., 2020]. At intermediate to high $k_{\parallel}$ however, the reconstruction of the HI signal is good. We find that the cleaning efficiency at low $k_{\parallel}$ is determined by two factors: the amplitude of the beam side-lobes and the level of point-source contamination.

The second aim of this work was to address the impact of a realistic frequency-dependent beam on the cleaning efficiency. As explained previously, HI intensity mapping studies assume that the FWHM of the beam follows a simple (λ/D) scaling. In this work we followed fits from Asad et al. [2021], where two models of the FWHM with increasing complexity were constructed. As shown in figure 3.8, the first model is a polynomial in frequency (i.e. the smooth model) and the second included a frequency oscillation on top of this polynomial (i.e. the ripple model).

We find that the smooth model poses no problem for the performance of the foreground subtraction. However, for the second model of the FWHM, we find a strong coupling of the foreground with this ripple feature, especially when strong point sources are included. The efficiency of the cleaning is adversely affected. As previously stated, this effect is more visible in the radial power spectrum, where the $k_{||}$ scales corresponding to the frequency of the ripple are biased. Interestingly, when the sky model is constructed with Galactic synchrotron as the only foreground, this effect is also present. This result shows that, even though aggressive cleaning removal of point sources can be applied, some level of contamination will still remain. What is also interesting to note is that frequency fluctuations from the main lobe of the beam and not only the side-lobes, can produce this contamination, as shown by the Galactic synchrotron result in figure 4.8. These non-trivial frequency behaviours in the FWHM of the beam, if left untreated, could negatively impact cosmological analysis of the 21 cm power spectrum.

Nevertheless, we found that the effect of a beam ripple can be greatly reduced if the maps are re-smoothed to a common resolution. Overall, the outcome of considering such a realistic beam is positive: side-lobes should not be an issue even when considering strong point sources and the frequency effects seem well localized in k space and can be improved through deconvolution. This is good news for large cosmology surveys with MeerKAT and the SKA. Further studies with more sophisticated foreground cleaning methods might improve this outcome.

In the previous analysis, the fundamental assumption that paved a way for conducting the simulations described in chapter 3 and 4 was the circular symmetry of the beam. To fully appreciate the frequency effects of the primary beam, this condition must be relaxed, and the convolution integral given in equation 5.3 needs to be solved directly. In chapter 5, a framework for exploring the map-space convolution is built. Instead of using the primary

beam models and non-trivial frequency-dependent models, the actual 2-D EIDOS beams were used so that the asymmetries present are included. However, to use these beams, a method of how to project the 2-D beams onto a HEALPix sphere is developed and validated in section 5.1: note in particular the right panels of figures 5.1 and 5.2, which are before- and after-mapping plots.

In order to compute the map-space convolution, the method outlined in section 5.2 was followed and the map-space convolution effect on foreground cleaning was probed in section 5.3. In studying the primary beam effects using the map-space convolution, the following findings were made.

For the first test, the full point source for the sky models was used and the assumption here was that the point sources could not be removed. We find that the ripple effect is present on the power spectrum as expected and is now much more complex than previously seen. In this case, a secondary peak just a little higher up in $k_{\parallel}$ space is also present (see figure 5.5). This finding shows that the ripple effect does exist and is not constructed as done in section 3.2.1 and included in the primary beam model.

With respect to the foreground subtraction, the EIDOS beams create some complexity in the reconstruction of the HI signal, as one has to aggressively clean by removing higher modes, to $N_{\rm fg} = 6$. At that level of cleaning, the complex feature is still present but greatly reduced, although the HI signal is a slightly over-estimated at lower $k_{\parallel}$. Beyond $N_{\rm fg} = 6$, the HI signal at lower to intermediate $k_{\parallel}$ scales is greatly over-cleaned.

In the case of the angular power spectrum (see figure 5.6), we find that at $N_{\rm fg} = 6$, the residuals from the foregrounds are still dominant, just as with the line of sight power spectrum. We find that increasing the number of modes removed to $N_{\rm fg} = 10$ results in power loss at higher angular scales (lower ℓ). However at intermediate to small angular scales (higher ℓ), the signal is recovered within 20 % of the expected HI signal.

Lastly, we investigated whether a reduction in the level of the ripple is achieved when point sources can be removed. In executing this test, a flux threshold was imposed and sources with flux less than 100 mJy were considered in the construction of the point source maps. Figures 5.7 and 5.8 show the results after this run of the pipeline. Overall, we find that by reducing the flux contribution from point sources, the frequency effect is greatly reduced as observed, with a lower number of subtracted components. In addition,

reconstruction of the HI angular power in this case is greatly improved and it is recovered within 5 % of the expected signal.

The work presented in chapter 5 is an important step toward a better understanding of the impact of a realistic beam on intensity mapping observations. The machinery that we have validated can be improved in two main directions:

- Implement a realistic scanning strategy; this will allow an even better quantification of the impact of the beam asymmetries in the final maps due to beam rotation.
- Perform a proper estimation of the effects of polarization leakage.

These are natural extensions of our pipeline and we plan to explore them in the near future. The goal is to publish a paper based on these investigations and the results of chapter 5.

Bibliography

F. Abbate, A. Ridolfi, E. D. Barr, S. Buchner, M. Burgay, D. J. Champion, W. Chen, P. C. C. Freire, T. Gautam, J. M. Grießmeier, L. Künkel, M. Kramer, P. V. Padmanabh, A. Possenti, S. Ransom, M. Serylak, B. W. Stappers, V. Venkatraman Krishnan, J. Behrend, R. P. Breton, L. Levin, and Y. Men. Four pulsar discoveries in NGC 6624 by TRAPUM using MeerKAT. *Monthly Notices of the Royal Astronomical Society*, 513(2):2292–2301, June 2022. doi: 10.1093/mnras/stac1041.

Elcio Abdalla, Elisa G. M. Ferreira, Ricardo G. Landim, Andre A. Costa, Karin S. F. Fornazier, Filipe B. Abdalla, Luciano Barosi, Francisco A. Brito, Amilcar R. Queiroz, Thyrso Villela, Bin Wang, Carlos A. Wuensche, Alessandro Marins, Camila P. Novaes, Vincenzo Liccardo, Chenxi Shan, Jiajun Zhang, Zhongli Zhang, Zhenghao Zhu, Ian Browne, Jacques Delabrouille, Larissa Santos, Marcelo V. dos Santos, Haiguang Xu, Sonia Anton, Richard Battye, Tianyue Chen, Clive Dickinson, Yin-Zhe Ma, Bruno Maffei, Eduardo J. de Mericia, Pablo Motta, Carlos H. N. Otobone, Michael W. Peel, Sambit Roychowdhury, Mathieu Remazeilles, Rafael M. Ribeiro, Yu Sang, Joao R. L. Santos, Juliana F. R. dos Santos, Gustavo B. Silva, Frederico Vieira, Jordany Vieira, Linfeng Xiao, Xue Zhang, and Yongkai Zhu. The BINGO Project I: Baryon Acoustic Oscillations from Integrated Neutral Gas Observations. *Astronomy & Astrophysics*, art. arXiv:2107.01633, July 2021.

David Alonso, Pedro G. Ferreira, and Mario G. Santos. Fast simulations for intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 444(4):3183–3197, November 2014. doi: 10.1093/mnras/stu1666.

David Alonso, Philip Bull, Pedro G. Ferreira, and Mário G. Santos. Blind foreground subtraction for intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 447(1):400–416, February 2015. doi: 10.1093/mnras/stu2474.

C. J. Anderson, N. J. Luciw, Y. C. Li, C. Y. Kuo, J. Yadav, K. W. Masui, T. C. Chang, X. Chen, N. Oppermann, Y. W. Liao, U. L. Pen, D. C. Price, L. Staveley-Smith, E. R. Switzer, P. T. Timbie, and L. Wolz. Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxy densities. *Monthly Notices of the Royal Astronomical Society*, 476(3):3382–3392, May 2018. doi: 10.1093/mnras/sty346.

R. Ansari, J. E. Campagne, P. Colom, J. M. Le Goff, C. Magneville, J. M. Martin, M. Moniez, J. Rich, and C. Yèche. 21 cm observation of large-scale structures atz 1. *Astronomy & Astrophysics*, 540: A129, Apr 2012. ISSN 1432-0746. doi: 10.1051/0004-6361/201117837. URL http://dx.doi.org/10.1051/0004-6361/201117837.

Robert Antonucci. Unified models for active galactic nuclei and quasars. *Annual Review of Astronomy and Astrophysics*, 31:473–521, January 1993. doi: 10.1146/annurev.aa.31.090193.002353.

K. M. B. Asad, J. N. Girard, M. de Villiers, T. Ansah-Narh, K. Iheanetu, O. Smirnov, M. G. Santos, R. Lehmensiek, J. Jonas, D. I. L. de Villiers, K. Thorat, B. Hugo, S. Makhathini, G. I. G. Jozsa, and S. K. Sirothia. Primary beam effects of radio astronomy antennas - II. Modelling MeerKAT L-band beams. *Monthly Notices of the Royal Astronomical Society*, 502(2):2970–2983, April 2021. doi: 10.1093/mnras/stab104.

J. S. Bagla, Nishikanta Khandai, and Kanan K. Datta. HI as a probe of the large-scale structure in the post-reionization universe. *Monthly Notices of the Royal Astronomical Society*, 407(1):567–580, September 2010. doi: 10.1111/j.1365-2966.2010.16933.x.

Matthew Bailes, Ewan D. Barr, N. D. Ramesh Bhat, J. Brink, Sarah Buchner, Marta Burgay, Fernando Camilo, David J. Champion, Jason W. T. Hessels, G. Janssen, Andrew Jameson, Simon Johnston, Aris

Karastergiou, Ramesh Karuppusamy, Victoria M. Kaspi, Michael J. Keith, Michael Kramer, Maura Mclaughlin, Kavilan Moodley, Stefan Osłowski, Andrea Possenti, Scott M. Ransom, Frederic A. Rasio, Jonathan Sievers, Maciej Serylak, Ben W. Stappers, Ingrid H. Stairs, Gilles Theureau, Willem van Straten, Patrick Weltevrede, and Norbert Wex. Meertime - the meerkat key science program on pulsar timing. *arXiv: Instrumentation and Methods for Astrophysics*, 2016.

R. A. Battye, M. L. Brown, I. W. A. Browne, R. J. Davis, P. Dewdney, C. Dickinson, G. Heron, B. Maffei, A. Pourtsidou, and P. N. Wilkinson. BINGO: a single dish approach to 21cm intensity mapping. *arXiv e-prints*, art. arXiv:1209.1041, September 2012.

R. A. Battye, I. W. A. Browne, C. Dickinson, G. Heron, B. Maffei, and A. Pourtsidou. Hi intensity mapping: a single dish approach. *Monthly Notices of the Royal Astronomical Society*, 434(2):1239–1256, 07 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt1082. URL https://doi.org/10.1093/mnras/stt1082.

Richard A. Battye, Rod D. Davies, and Jochen Weller. Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters. *Monthly Notices of the Royal Astronomical Society*, 355(4):1339–1347, Dec 2004. ISSN 1365-2966. doi: 10.1111/j.1365-2966.2004.08416.x. URL http://dx.doi.org/10.1111/j.1365-2966.2004.08416.x.

Robert H. Becker, Xiaohui Fan, Richard L. White, Michael A. Strauss, Vijay K. Narayanan, Robert H. Lupton, James E. Gunn, James Annis, Neta A. Bahcall, J. Brinkmann, A. J. Connolly, István Csabai, Paul C. Czarapata, Mamoru Doi, Timothy M. Heckman, G. S. Hennessy, Željko Ivezić, G. R. Knapp, Don Q. Lamb, Timothy A. McKay, Jeffrey A. Munn, Thomas Nash, Robert Nichol, Jeffrey R. Pier, Gordon T. Richards, Donald P. Schneider, Chris Stoughton, Alexander S. Szalay, Aniruddha R. Thakar, and D. G. York. Evidence for Reionization at z~6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. *The Astronomical Journal*, 122(6):2850–2857, December 2001. doi: 10.1086/324231.

M.-A. Bigot-Sazy, Clive Dickinson, Richard A. Battye, I. W. A. Browne, Y.-Z. Ma, Bruno Maffei, Fabio Noviello, Mathieu Remazeilles, and Peter N. Wilkinson. Simulations for single-dish intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 454:3240–3253, 2015.

S. Blyth, A. J. Baker, B. Holwerda, A. Bouchard, B. Catinella, L. Chemin, D. Cunnama, R. Davé, A. Faltenbacher, S. February, X. Fernández, E. Gawiser, I. Heywood, D. Kereš, H. R. Klöckner, P. Lah, M. Lochner, N. Maddox, S. Makhathini, K. Moodley, R. Morganti, D. Obreschkow, S. H. Oh, D. J. Pisano, A. Popping, G. Popping, S. Ravindranath, E. Schinnerer, K. Sheth, R. Skelton, M. Smith, R. Srianand, L. Staveley-Smith, M. Vaccari, P. Vaisanen, F. Walter, S. Rawlings, B. A. Bassett, M. A. Bershady, F. H. Briggs, S. M. Crawford, C. M. Cress, J. K. Darling, R. P. Deane, G. de Blok, E. C. Elson, B. S. Frank, P. A. Henning, K. M. Hess, J. P. Hughes, M. J. Jarvis, S. J. Kannappan, N. S. Katz, R. C. Kraan-Korteweg, M. D. Lehnert, A. K. Leroy, G. R. Meurer, M. J. Meyer, D. J. Pisano, A. C. Schröder, O. M. Smirnov, R. S. Somerville, I. M. Stewart, K. J. van der Heyden, M. A. W. Verheijen, E. M. Wilcots, T. B. Williams, P. A. Woudt, J. F. Wu, M. A. Zwaan, J. T. L. Zwart, T. A. Oosterloo, and W. van Drie. LADUMA: Looking at the Distant Universe with the MeerKAT Array. In *MeerKAT Science: On the Pathway to the SKA*, page 4, January 2016.

Anna Bonaldi, Matteo Bonato, Vincenzo Galluzzi, Ian Harrison, Marcella Massardi, Scott Kay, Gianfranco De Zotti, and Michael L Brown. The tiered radio extragalactic continuum simulation (t-recs). *Monthly Notices of the Royal Astronomical Society*, 482(1):2–19, Sep 2018. ISSN 1365-2966. doi: 10.1093/mnras/sty2603. URL http://dx.doi.org/10.1093/mnras/sty2603.

Judd D. Bowman, Alan E. E. Rogers, Raul A. Monsalve, Thomas J. Mozdzen, and Nivedita Mahesh. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*, 555(7694):67–70, March 2018. doi: 10.1038/nature25792.

Philip Bull, Pedro G. Ferreira, Prina Patel, and Mário G. Santos. Late-time Cosmology with 21 cm Intensity Mapping Experiments. *The Astrophysical Journal*, 803(1):21, April 2015. doi: 10.1088/0004-637X/803/1/21.

Isabella P Carucci, Melis O Irfan, and Jérôme Bobin. Recovery of 21-cm intensity maps with sparse component separation. *Monthly Notices of the Royal Astronomical Society*, 499(1):304–319, 09 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa2854. URL https://doi.org/10.1093/mnras/staa2854.

T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald. Baryon Acoustic Oscillation Intensity Mapping of Dark Energy. *Physical Review Letters*, 100(9):091303, March 2008. doi: 10.1103/PhysRevLett.100.091303. URL http://adsabs.harvard.edu/abs/2008PhRvL.100i1303C.

Tzu-Ching Chang, Ue-Li Pen, Kevin Bandura, and Jeffrey B. Peterson. An intensity map of hydrogen 21-cm emission at redshift z~0.8. *Nature*, 466(7305):463–465, July 2010. doi: 10.1038/nature09187.

Xuelei Chen. The Tianlai Project: a 21CM Cosmology Experiment. In *International Journal of Modern Physics Conference Series*, volume 12 of *International Journal of Modern Physics Conference Series*, pages 256–263, March 2012. doi: 10.1142/S2010194512006459.

Carina Cheng. *Methods for the Detection of the Epoch of Reionization.* PhD thesis, UC Berkeley, 2019. URL https://escholarship.org/uc/item/32114694.

Carina Cheng, Aaron R. Parsons, Matthew Kolopanis, Daniel C. Jacobs, Adrian Liu, Saul A. Kohn, James E. Aguirre, Jonathan C. Pober, Zaki S. Ali, Gianni Bernardi, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Matthew R. Dexter, Joshua S. Dillon, Pat Klima, David H. E. MacMahon, David F. Moore, Chuneeta D. Nunhokee, William P. Walbrugh, and Andre Walker. Characterizing Signal Loss in the 21 cm Reionization Power Spectrum: A Revised Study of PAPER-64. *The Astrophysical Journal*, 868(1):26, November 2018. doi: 10.3847/1538-4357/aae833.

CHIME Collaboration, Mandana Amiri, Kevin Bandura, Tianyue Chen, Meiling Deng, Matt Dobbs, Mateus Fandino, Simon Foreman, Mark Halpern, Alex S. Hill, Gary Hinshaw, Carolin Höfer, Joseph Kania, T. L. Landecker, Joshua MacEachern, Kiyoshi Masui, Juan Mena-Parra, Nikola Milutinovic, Arash Mirhosseini, Laura Newburgh, Anna Ordog, Ue-Li Pen, Tristan Pinsonneault-Marotte, Ava Polzin, Alex Reda, Andre Renard, J. Richard Shaw, Seth R. Siegel, Saurabh Singh, Keith Vanderlinde, Haochen Wang, Donald V. Wiebe, and Dallas Wulf. Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment. *arXiv e-prints*, art. arXiv:2202.01242, February 2022.

CHIME/FRB Collaboration, M. Amiri, K. Bandura, M. Bhardwaj, P. Boubel, M. M. Boyce, P. J Boyle, C. . Brar, M. Burhanpurkar, T. Cassanelli, P. Chawla, J. F. Cliche, D. Cubranic, M. Deng, N. Denman, M. Dobbs, M. Fandino, E. Fonseca, B. M. Gaensler, A. J. Gilbert, A. Gill, U. Giri, D. C. Good, M. Halpern, D. S. Hanna, A. S. Hill, G. Hinshaw, C. Höfer, A. Josephy, V. M. Kaspi, T. L. Landecker, D. A. Lang, H. H. Lin, K. W. Masui, R. Mckinven, J. Mena-Parra, M. Merryfield, D. Michilli, N. Milutinovic, C. Moatti, A. Naidu, L. B. Newburgh, C. Ng, C. Patel, U. Pen, T. Pinsonneault-Marotte, Z. Pleunis, M. Rafiei-Ravandi, M. Rahman, S. M. Ransom, A. Renard, P. Scholz, J. R. Shaw, S. R. Siegel, K. M. Smith, I. H. Stairs, S. P. Tendulkar, I. Tretyakov, K. Vanderlinde, and P. Yadav. A second source of repeating fast radio bursts. *Nature*, 566(7743):235–238, January 2019a. doi: 10.1038/s41586-018-0864-x.

CHIME/FRB Collaboration, M. Amiri, K. Bandura, M. Bhardwaj, P. Boubel, M. M. Boyce, P. J. Boyle, C. Brar, M. Burhanpurkar, P. Chawla, J. F. Cliche, D. Cubranic, M. Deng, N. Denman, M. Dobbs, M. Fandino, E. Fonseca, B. M. Gaensler, A. J. Gilbert, U. Giri, D. C. Good, M. Halpern, D Hanna, A. S. Hill, G. Hinshaw, C. Höfer, A. Josephy, V. M. Kaspi, T. L. Landecker, D. A. Lang, K. W. Masui, R. Mckinven, J. Mena-Parra, M. Merryfield, N. Milutinovic, C. Moatti, A. Naidu, L. B. Newburgh, C. Ng, C. Patel, U. Pen, T. Pinsonneault-Marotte, Z. Pleunis, M. Rafiei-Ravandi, S. M. Ransom, A. Renard, P. Scholz, J. R. Shaw, S. R. Siegel, K. M. Smith, I. H. Stairs, S. P. Tendulkar, I. Tretyakov, K. Vanderlinde, and P. Yadav. Observations of fast radio bursts at frequencies down to 400 megahertz. *Nature*, 566(7743):230–234, January 2019b. doi: 10.1038/s41586-018-0867-7.

CHIME/FRB Collaboration, B. C. Andersen, K. Bandura, M. Bhardwaj, P. Boubel, M. M. Boyce, P. J. Boyle, C. Brar, T. Cassanelli, P. Chawla, D. Cubranic, M. Deng, M. Dobbs, M. Fandino, E. Fonseca, B. M. Gaensler, A. J. Gilbert, U. Giri, D. C. Good, M. Halpern, A. S. Hill, G. Hinshaw, C. Höfer, A. Josephy, V. M. Kaspi, R. Kothes, T. L. Landecker, D. A. Lang, D. Z. Li, H. H. Lin, K. W. Masui, J. Mena-Parra, M. Merryfield, R. Mckinven, D. Michilli, N. Milutinovic, A. Naidu, L. B. Newburgh, C. Ng, C. Patel, U. Pen, T. Pinsonneault-Marotte, Z. Pleunis, M. Rafiei-Ravandi, M. Rahman, S. M. Ransom, A. Renard, P. Scholz, S. R. Siegel, S. Singh, K. M. Smith, I. H. Stairs, S. P. Tendulkar, I. Tretyakov, K. Vanderlinde, P. Yadav, and A. V. Zwaniga. CHIME/FRB Discovery of Eight New Repeating Fast Radio Burst Sources. *The Astrophysical Journal*, 885(1):L24, November 2019c. doi: 10.3847/2041-8213/ab4a80.

Jenifer Chu. CHIME telescope detects more than 500 mysterious fast radio bursts in its first year of operation. https://news.mit.edu/2021/chime-telescope-fast-radio-bursts-0609, 2021. Accessed: 2022-02-25.

M. Colless. First results from the 2dF Galaxy Redshift Survey. *Philosophical Transactions of the Royal Society of London Series A*, 357(1750):105, January 1999. doi: 10.1098/rsta.1999.0317.

J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick. The NRAO VLA Sky Survey. *The Astronomical Journal*, 115(5):1693–1716, May 1998. doi: 10.1086/300337.

James J. Condon and Scott M. Ransom. *Essential Radio Astronomy*. Princeton Series in Modern Observational Astronomy. Princeton University Press, 2016. ISBN 9780691137797.

Devin Crichton, Moumita Aich, Adam Amara, Kevin Bandura, Bruce A. Bassett, Carlos Bengaly, Pascale Berner, Shruti Bhatporia, Martin Bucher, Tzu-Ching Chang, H. Cynthia Chiang, Jean-Francois Cliche, Carolyn Crichton, Romeel Dave, Dirk I. L. De Villiers, Matt Dobbs, Aaron M. Ewall-Wice, Scott Eyono, Christopher Finlay, Sindhu Gaddam, Ken Ganga, Kevin G. Gayley, Kit Gerodias, Tim B. Gibbon, Austine Gumba, Neeraj Gupta, Maile Harris, Heiko Heilgendorff, Matt Hilton, Adam D. Hincks, Pascal Hitz, Mona Jalilvand, Roufurd P. M. Julie, Zahra Kader, Joseph Kania, Dionysios Karagiannis, Aris Karastergiou, Kabelo Kesebonye, Piyanat Kittiwisit, Jean-Paul Kneib, Kenda Knowles, Emily R. Kuhn, Martin Kunz, Roy Maartens, Vincent MacKay, Stuart MacPherson, Christian Monstein, Kavilan Moodley, V. Mugundhan, Warren Naidoo, Arun Naidu, Laura B. Newburgh, Viraj Nistane, Amanda Di Nitto, Deniz Ölçek, Xinyu Pan, Sourabh Paul, Jeffrey B. Peterson, Elizabeth Pieters, Carla Pieterse, Aritha Pillay, Anna R. Polish, Liantsoa Randrianjanahary, Alexandre Refregier, Andre Renard, Edwin Retana-Montenegro, Ian H. Rout, Cyndie Russeeawon, Alireza Vafaei Sadr, Benjamin R. B. Saliwanchik, Ajith Sampath, Pranav Sanghavi, Mario G. Santos, Onkabetse Sengate, J. Richard Shaw, Jonathan L. Sievers, Oleg M. Smirnov, Kendrick M. Smith, Ulrich Armel Mbou Sob, Raghunathan Srianand, Pieter Stronkhorst, Dhaneshwar D. Sunder, Simon Tartakovsky, Russ Taylor, Peter Timbie, Emma E. Tolley, Junaid Townsend, Will Tyndall, Cornelius Ungerer, Jacques van Dyk, Gary van Vuuren, Keith Vanderlinde, Thierry Viant, Anthony Walters, Jingying Wang, Amanda Weltman, Patrick Woudt, Dallas Wulf, Anatoly Zavyalov, and Zheng Zhang. Hydrogen Intensity and Real-Time Analysis Experiment: 256-element array status and overview. *Journal of Astronomical Telescopes, Instruments, and Systems*, 8:011019, January 2022. doi: 10.1117/1.JATIS.8.1.011019.

Neil H. M. Crighton, Michael T. Murphy, J. Xavier Prochaska, Gábor Worseck, Marc Rafelski, George D. Becker, Sara L. Ellison, Michele Fumagalli, Sebastian Lopez, Avery Meiksin, and John M. O'Meara. The neutral hydrogen cosmological mass density at z = 5. *Monthly Notices of the Royal Astronomical Society*, 452(1):217–234, September 2015. doi: 10.1093/mnras/stv1182.

Steven Cunnington, Laura Wolz, Alkistis Pourtsidou, and David Bacon. Impact of foregrounds on hi intensity mapping cross-correlations with optical surveys. *Monthly Notices of the Royal Astronomical Society*, 488(4):5452–5472, Jul 2019. ISSN 1365-2966. doi: 10.1093/mnras/stz1916. URL http://dx.doi.org/10.1093/mnras/stz1916.

Steven Cunnington, Stefano Camera, and Alkistis Pourtsidou. The degeneracy between primordial non-Gaussianity and foregrounds in 21cm intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, September 2020. doi: 10.1093/mnras/staa2986.

Steven Cunnington, Melis O. Irfan, Isabella P. Carucci, Alkistis Pourtsidou, and Jérôme Bobin. 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies. *Monthly Notices of the Royal Astronomical Society*, 504(1):208–227, June 2021. doi: 10.1093/mnras/stab856.

Steven Cunnington, Yichao Li, Mario G. Santos, Jingying Wang, Isabella P. Carucci, Melis O. Irfan, Alkistis Pourtsidou, Marta Spinelli, Laura Wolz, Paula S. Soares, Chris Blake, Philip Bull, Brandon Engelbrecht, José Fonseca, Keith Grainge, and Yin-Zhe Ma. HI intensity mapping with MeerKAT: power spectrum detection in cross-correlation with WiggleZ galaxies. *arXiv e-prints*, art. arXiv:2206.01579, June 2022.

Steven D Cunnington. *Synergies Between 21cm Optical Surveys for Probing Large Scale Cosmic Structure*. PhD thesis, University of Portsmouth, 2019.

Marc Davis, Sandra M. Faber, Jeffrey Newman, Andrew C. Phillips, Richard S. Ellis, Charles C. Steidel, C. Conselice, Alison L. Coil, D. P. Finkbeiner, David C. Koo, Puragra Guhathakurta, B. Weiner, Ricardo Schiavon, C. Willmer, Nicholas Kaiser, Gerard A. Luppino, Gregory Wirth, Andrew Connolly, Peter Eisenhardt, M. Cooper, and B. Gerke. Science objectives and early results of the DEEP2 redshift survey. In Puragra Guhathakurta, editor, *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*. SPIE, feb 2003. doi: 10.1117/12.457897. URL https://doi.org/10.1117%2F12.457897.

Kyle S. Dawson, Jean-Paul Kneib, Will J. Percival, Shadab Alam, Franco D. Albareti, Scott F. Anderson, Eric Armengaud, Éric Aubourg, Stephen Bailey, Julian E. Bautista, Andreas A. Berlind, Matthew A. Bershady, Florian Beutler, Dmitry Bizyaev, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton,

Jo Bovy, W. N. Brandt, Jon Brinkmann, Joel R. Brownstein, Etienne Burtin, N. G. Busca, Zheng Cai, Chia-Hsun Chuang, Nicolas Clerc, Johan Comparat, Frances Cope, Rupert A. C. Croft, Irene Cruz-Gonzalez, Luiz N. da Costa, Marie-Claude Cousinou, Jeremy Darling, Axel de la Macorra, Sylvain de la Torre, Timothée Delubac, Hélion du Mas des Bourboux, Tom Dwelly, Anne Ealet, Daniel J. Eisenstein, Michael Eracleous, S. Escoffier, Xiaohui Fan, Alexis Finoguenov, Andreu Font-Ribera, Peter Frinchaboy, Patrick Gaulme, Antonis Georgakakis, Paul Green, Hong Guo, Julien Guy, Shirley Ho, Diana Holder, Joe Huehnerhoff, Timothy Hutchinson, Yipeng Jing, Eric Jullo, Vikrant Kamble, Karen Kinemuchi, David Kirkby, Francisco-Shu Kitaura, Mark A. Klaene, Russ R. Laher, Dustin Lang, Pierre Laurent, Jean-Marc Le Goff, Cheng Li, Yu Liang, Marcos Lima, Qiufan Lin, Weipeng Lin, Yen-Ting Lin, Daniel C. Long, Britt Lundgren, Nicholas MacDonald, Marcio Antonio Geimba Maia, Elena Malanushenko, Viktor Malanushenko, Vivek Mariappan, Cameron K. McBride, Ian D. McGreer, Brice Ménard, Andrea Merloni, Andres Meza, Antonio D. Montero-Dorta, Demitri Muna, Adam D. Myers, Kirpal Nandra, Tracy Naugle, Jeffrey A. Newman, Pasquier Noterdaeme, Peter Nugent, Ricardo Ogando, Matthew D. Olmstead, Audrey Oravetz, Daniel J. Oravetz, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, John A. Peacock, Patrick Petitjean, Matthew M. Pieri, Alice Pisani, Francisco Prada, Abhishek Prakash, Anand Raichoor, Beth Reid, James Rich, Jethro Ridl, Sergio Rodriguez-Torres, Aurelio Carnero Rosell, Ashley J. Ross, Graziano Rossi, John Ruan, Mara Salvato, Conor Sayres, Donald P. Schneider, David J. Schlegel, Uros Seljak, Hee-Jong Seo, Branimir Sesar, Sarah Shandera, Yiping Shu, Anže Slosar, Flavia Sobreira, Alina Streblyanska, Nao Suzuki, Donna Taylor, Charling Tao, Jeremy L. Tinker, Rita Tojeiro, Mariana Vargas-Magaña, Yuting Wang, Benjamin A. Weaver, David H. Weinberg, Martin White, W. M. Wood-Vasey, Christophe Yeche, Zhongxu Zhai, Cheng Zhao, Gong-bo Zhao, Zheng Zheng, Guangtun Ben Zhu, and Hu Zou. The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. *The Astronomical Journal*, 151(2):44, February 2016. doi: 10.3847/0004-6256/151/2/44.

W. J. G. de Blok, E. A. K. Adams, P. Amram, E. Athanassoula, I. Bagetakos, C. Balkowski, M. A. Bershady, R. J. Beswick, F. Bigiel, S. L. Blyth, A. Bosma, R. S. Booth, A. Bouchard, E. Brinks, C. Carignan, L. Chemin, F. Combes, J. Conway, E. C. Elson, J. English, B. Epinat, B. S. Frank, J. Fiege, F. Fraternali, J. S. Gallagher, B. K. Gibson, G. Heald, P. A. Henning, B. W. Holwerda, T. H. Jarrett, H. Jerjen, G. I. Józsa, M. Kapala, H. R. Klöckner, B. S. Koribalski, R. C. Kraan-Korteweg, S. Leon, A. Leroy, S. I. Loubser, D. M. Lucero, S. S. McGaugh, G. R. Meurer, M. Meyer, M. Mogotsi, B. Namumba, S. H. Oh, T. A. Oosterloo, D. J. Pisano, A. Popping, S. Ratcliffe, J. A. Sellwood, E. Schinnerer, A. C. Schröder, K. Sheth, M. W. L. Smith, A. Sorgho, K. Spekkens, S. Stanimirovic, K. J. van der Heyden, L. Verdes-Montenegro, W. van Driel, F. Walter, T. Westmeier, E. Wilcots, T. Williams, O. I. Wong, P. A. Woudt, and A. Zijlstra. An Overview of the MHONGOOSE Survey: Observing Nearby Galaxies with MeerKAT. In *MeerKAT Science: On the Pathway to the SKA*, page 7, January 2016.

Angélica de Oliveira-Costa and John Capodilupo. Clustering at 74MHz. *Monthly Notices of the Royal Astronomical Society*, 404(4):1962–1965, June 2010. doi: 10.1111/j.1365-2966.2010.16412.x.

Angelica de Oliveira-Costa and Joseph Lazio. Clustering of Extragalactic Sources from 151 MHz to 232 MHz: Implications for Cosmological 21-cm Observations. *arXiv e-prints*, art. arXiv:1004.3167, April 2010.

D. I. L. de Villiers. Prediction of aperture efficiency ripple in clear aperture offset gregorian antennas. *IEEE Transactions on Antennas and Propagation*, 61(5):2457–2465, 2013.

J. Delabrouille, M. Betoule, J.-B. Melin, M.-A. Miville-Deschênes, J. Gonzalez-Nuevo, M. Le Jeune, G. Castex, G. de Zotti, S. Basak, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, J.-P. Bernard, F. R. Bouchet, D. L. Clements, A. da Silva, C. Dickinson, F. Dodu, K. Dolag, F. Elsner, L. Fauvet, G. Faÿ, G. Giardino, S. Leach, J. Lesgourgues, M. Liguori, J. F. Macías-Pérez, M. Massardi, S. Matarrese, P. Mazzotta, L. Montier, S. Mottet, R. Paladini, B. Partridge, R. Piffaretti, G. Prezeau, S. Prunet, S. Ricciardi, M. Roman, B. Schaefer, and L. Toffolatti. The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths. *Astronomy & Astrophysics*, 553:A96, May 2013. doi: 10.1051/0004-6361/201220019. URL http://adsabs.harvard.edu/abs/2013A%26A...553A..96D.

X. Deng, A. P. Chippendale, G. Hobbs, S. Johnston, S. Dai, D. George, M. Kramer, R. Karuppusamy, M. Malenta, L. Spitler, T. Tzioumis, and G. Wieching. Observing Pulsars with a Phased Array Feed at the Parkes Telescope. *Publications of the Astronomical Society of Australia*, 34:e026, July 2017. doi: 10.1017/pasa.2017.20.

Eleonora Di Valentino, Olga Mena, Supriya Pan, Luca Visinelli, Weiqiang Yang, Alessandro Melchiorri, David F. Mota, Adam G. Riess, and Joseph Silk. In the realm of the Hubble tension—a review of solutions. *Class. Quant. Grav.*, 38(15):153001, 2021. doi: 10.1088/1361-6382/ac086d.

Scott S. Dodelson. *Modern cosmology*. Princeton Series in Modern Observational Astronomy. Academic Press, 2003. ISBN 9780122191411.

Michael J. Drinkwater, Russell J. Jurek, Chris Blake, David Woods, Kevin A. Pimbblet, Karl Glazebrook, Rob Sharp, Michael B. Pracy, Sarah Brough, Matthew Colless, Warrick J. Couch, Scott M. Croom, Tamara M. Davis, Duncan Forbes, Karl Forster, David G. Gilbank, Michael Gladders, Ben Jelliffe, Nick Jones, I-hui Li, Barry Madore, D. Christopher Martin, Gregory B. Poole, Todd Small, Emily Wisnioski, Ted Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: survey design and first data release. *Monthly Notices of the Royal Astronomical Society*, 401(3):1429–1452, 01 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.15754.x. URL https://doi.org/10.1111/j.1365-2966.2009.15754.x.

Michael J. Drinkwater, Zachary J. Byrne, Chris Blake, Karl Glazebrook, Sarah Brough, Matthew Colless, Warrick Couch, Darren J. Croton, Scott M. Croom, Tamara M. Davis, Karl Forster, David Gilbank, Samuel R. Hinton, Ben Jelliffe, Russell J. Jurek, I. hui Li, D. Christopher Martin, Kevin Pimbblet, Gregory B. Poole, Michael Pracy, Rob Sharp, Jon Smillie, Max Spolaor, Emily Wisnioski, David Woods, Ted K. Wyder, and Howard K. C. Yee. The WiggleZ Dark Energy Survey: final data release and the metallicity of UV-luminous galaxies. *Monthly Notices of the Royal Astronomical Society*, 474 (3):4151–4168, March 2018. doi: 10.1093/mnras/stx2963.

R. Fender, P. A. Woudt, S. Corbel, M. Coriat, F. Daigne, H. Falcke, J. Girard, I. Heywood, A. Horesh, J. Horrell, P. G. Jonker, T. Joseph, A. Kamble, C. Knigge, E. Körding, M. Kotze, C. Kouveliotou, C. Lynch, T. Maccarone, P. Meintjes, S. Migliari, T. Murphy, T. Nagayama, G. Nelemans, G. Nicholson, T. O'Brien, A. Oodendaal, N. Oozeer, J. Osborne, M. Pérez-Torres, S. Ratcliffe, V. A. R. M. Ribeiro, E. Rol, A. Rushton, A. Scaife, M. Schurch, G. Sivakoff, T. Staley, D. Steeghs, I. Stewart, J. D. Swinbank, S. Vergani, B. Warner, K. Wiersema, R. Armstrong, P. Groot, V. McBride, J. C. A. Miller-Jones, K. Mooley, B. Stappers, R. A. M. J. Wijers, M. Bietenholz, S. Blyth, M. Böttcher, D. Buckley, P. Charles, L. Chomiuk, D. Coppejans, W. J. G. de Blok, K. van der Heyden, A. van der Horst, and B. van Soelen. ThunderKAT: The MeerKAT Large Survey Project for Image-Plane Radio Transients. In *MeerKAT Science: On the Pathway to the SKA*, page 13, January 2016.

José Fonseca, Marta B. Silva, Mário G. Santos, and Asantha Cooray. Cosmology with intensity mapping techniques using atomic and molecular lines. *Monthly Notices of the Royal Astronomical Society*, 464 (2):1948–1965, January 2017. doi: 10.1093/mnras/stw2470.

C. S. Frenk and S. D. M. White. Dark matter and cosmic structure. *Annalen der Physik*, 524(9-10): 507–534, October 2012. doi: 10.1002/andp.201200212.

Steven R. Furlanetto, S. Peng Oh, and Frank H. Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. , 433:181–301, October 2006. doi: 10.1016/j.physrep.2006.08.002.

Timothy Garn, David A. Green, Julia M. Riley, and Paul Alexander. A 610-MHz survey of the ELAIS-N1 field with the Giant Metrewave Radio Telescope - observations, data analysis and source catalogue. *Monthly Notices of the Royal Astronomical Society*, 383(1):75–85, January 2008. doi: 10.1111/j.1365-2966.2007.12562.x.

Abhik Ghosh, Somnath Bharadwaj, Sk. Saiyad Ali, and Jayaram N. Chengalur. GMRT observation towards detecting the post-reionization 21-cm signal. *Monthly Notices of the Royal Astronomical Society*, 411(4):2426–2438, March 2011. doi: 10.1111/j.1365-2966.2010.17853.x.

K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622:759–771, April 2005. doi: 10.1086/427976.

James E. Gunn and Bruce A. Peterson. On the Density of Neutral Hydrogen in Intergalactic Space. *The Astrophysical Journal*, 142:1633–1636, November 1965. doi: 10.1086/148444.

N. Gupta, R. Srianand, W. Baan, A. J. Baker, R. J. Beswick, S. Bhatnagar, D. Bhattacharya, A. Bosma, C. Carilli, M. Cluver, F. Combes, C. Cress, R. Dutta, J. Fynbo, G. Heald, M. Hilton, T. Hussain, M. Jarvis, G. Jozsa, P. Kamphuis, A. Kembhavi, J. Kerp, H. R. Kloeckner, J. Krogager, V. P. Kulkarni,

C. Ledoux, A. Mahabal, T. Mauch, K. Moodley, E. Momjian, R. Morganti, P. Noterdaeme, T. Oosterloo, P. Petitjean, A. Schroeder, P. Serra, J. Sievers, K. Spekkens, P. Vaisanen, T. van der Hulst, M. Vivek, J. Wang, O. I. Wong, and A. R. Zungu. The MeerKAT Absorption Line Survey (MALS). In *MeerKAT Science: On the Pathway to the SKA*, page 14, January 2016.

Z. Haiman and L. Knox. Reionization of the Intergalactic Medium and its Effect on the CMB. In A. de Oliveira-Costa and M. Tegmark, editors, *Microwave Foregrounds*, volume 181 of *Astronomical Society of the Pacific Conference Series*, page 227, January 1999.

Stuart E. Harper and Clive Dickinson. Potential impact of global navigation satellite services on total power H I intensity mapping surveys. *Monthly Notices of the Royal Astronomical Society*, 479(2): 2024–2036, September 2018. doi: 10.1093/mnras/sty1495.

C. G. T. Haslam, C. J. Salter, H. Stoffel, and W. E. Wilson. A 408 MHz all-sky continuum survey. II - The atlas of contour maps. *Astronomy & Astrophysics Supplement Series*, 47:1, January 1982. URL http://adsabs.harvard.edu/abs/1982A%26AS...47....1H.

HI4PI Collaboration, N. Ben Bekhti, L. Flöer, R. Keller, J. Kerp, D. Lenz, B. Winkel, J. Bailin, M. R. Calabretta, L. Dedes, H. A. Ford, B. K. Gibson, U. Haud, S. Janowiecki, P. M. W. Kalberla, F. J. Lockman, N. M. McClure-Griffiths, T. Murphy, H. Nakanishi, D. J. Pisano, and L. Staveley-Smith. HI4PI: A full-sky H I survey based on EBHIS and GASS. 594:A116, October 2016. doi: 10.1051/0004-6361/201629178.

Eric Hivon, Krzysztof M. Górski, C. Barth Netterfield, Brendan P. Crill, Simon Prunet, and Frode Hansen. MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *The Astrophysical Journal*, 567(1):2–17, March 2002. doi: 10.1086/338126.

George Hobbs, Richard N. Manchester, Alex Dunning, Andrew Jameson, Paul Roberts, Daniel George, J. A. Green, John Tuthill, Lawrence Toomey, Jane F. Kaczmarek, and et al. An ultra-wide bandwidth (704 to 4 032 mhz) receiver for the parkes radio telescope. *Publications of the Astronomical Society of Australia*, 37:e012, 2020. doi: 10.1017/pasa.2020.2.

B. W. Holwerda, S. L. Blyth, and A. J. Baker. Looking at the distant universe with the MeerKAT Array (LADUMA). In Richard J. Tuffs and Cristina C. Popescu, editors, *The Spectral Energy Distribution of Galaxies - SED 2011*, volume 284, pages 496–499, August 2012. doi: 10.1017/S1743921312009702.

A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

Melis O. Irfan and Philip Bull. Cleaning foregrounds from single-dish 21 cm intensity maps with Kernel principal component analysis. *Monthly Notices of the Royal Astronomical Society*, 508(3):3551–3568, December 2021. doi: 10.1093/mnras/stab2855.

M. Jarvis, R. Taylor, I. Agudo, J. R. Allison, R. P. Deane, B. Frank, N. Gupta, I. Heywood, N. Maddox, K. McAlpine, M. Santos, A. M. M. Scaife, M. Vaccari, J. T. L. Zwart, E. Adams, D. J. Bacon, A. J. Baker, B. A. Bassett, P. N. Best, R. Beswick, S. Blyth, M. L. Brown, M. Bruggen, M. Cluver, S. Colafrancesco, G. Cotter, C. Cress, R. Davé, C. Ferrari, M. J. Hardcastle, C. L. Hale, I. Harrison, P. W. Hatfield, H. R. Klockner, S. Kolwa, E. Malefahlo, T. Marubini, T. Mauch, K. Moodley, R. Morganti, R. P. Norris, J. A. Peters, I. Prandoni, M. Prescott, S. Oliver, N. Oozeer, H. J. A. Rottgering, N. Seymour, C. Simpson, O. Smirnov, and D. J. B. Smith. The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. In *MeerKAT Science: On the Pathway to the SKA*, page 6, January 2016.

J. Jonas and MeerKAT Team. The MeerKAT Radio Telescope. In *MeerKAT Science: On the Pathway to the SKA*, page 1, January 2016.

D. Heath Jones, Will Saunders, Michael Read, and Matthew Colless. Second Data Release of the 6dF Galaxy Survey. *Publications of the Astronomical Society of Australia*, 22(3):277–286, August 2005. doi: 10.1071/AS05018.

T. Kitching, D. Bacon, M. Brown, P. Bull, J. McEwen, M. Oguri, R. Scaramella, K. Takahashi, K. Wu, and D. Yamauchi. Euclid & SKA Synergies. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 146, April 2015.

John Kormendy and Douglas Richstone. Inward Bound—The Search For Supermassive Black Holes In Galactic Nuclei. *Annual Review of Astronomy and Astrophysics*, 33:581, January 1995. doi: 10.1146/annurev.aa.33.090195.003053.

Ely D. Kovetz, Marco P. Viero, Adam Lidz, Laura Newburgh, Mubdi Rahman, Eric Switzer, Marc Kamionkowski, James Aguirre, Marcelo Alvarez, James Bock, J. Richard Bond, Goeffry Bower, C. Matt Bradford, Patrick C. Breysse, Philip Bull, Tzu-Ching Chang, Yun-Ting Cheng, Dongwoo Chung, Kieran Cleary, Asantha Corray, Abigail Crites, Rupert Croft, Olivier Doré, Michael Eastwood, Andrea Ferrara, José Fonseca, Daniel Jacobs, Garrett K. Keating, Guilaine Lagache, Gunjan Lakhlani, Adrian Liu, Kavilan Moodley, Norm Murray, Aurélie Pénin, Gergö Popping, Anthony Pullen, Dominik Reichers, Shun Saito, Ben Saliwanchik, Mario Santos, Rachel Somerville, Gordon Stacey, George Stein, Francesco Villaescusa-Navarro, Eli Visbal, Amand a Weltman, Laura Wolz, and Micheal Zemcov. Line-Intensity Mapping: 2017 Status Report. *arXiv e-prints*, art. arXiv:1709.09066, September 2017.

H. Kubo, T. Takahashi, G. Madejski, M. Tashiro, F. Makino, S. Inoue, and F. Takahara. ASCA Observations of Blazars and Multiband Analysis. *The Astrophysical Journal*, 504(2):693–701, September 1998. doi: 10.1086/306125.

Lin-Cheng Li, Lister Staveley-Smith, and Jonghwan Rhee. An HI intensity mapping survey with a Phased Array Feed. *Research in Astronomy and Astrophysics*, 21(2):030, March 2021. doi: 10.1088/1674-4527/21/2/30.

Andrew A. Liddle. *An Introduction to Modern Cosmology*. John Wiley and Sons Ltd, 1999. ISBN 0471987573.

Adam Lidz, Steven R. Furlanetto, S. Peng Oh, James Aguirre, Tzu-Ching Chang, Olivier Doré, and Jonathan R. Pritchard. Intensity Mapping with Carbon Monoxide Emission Lines and the Redshifted 21 cm Line. *The Astrophysical Journal*, 741(2):70, November 2011. doi: 10.1088/0004-637X/741/2/70.

Abraham Loeb and J. Stuart B. Wyithe. Possibility of Precise Measurement of the Cosmological Power Spectrum with a Dedicated Survey of 21cm Emission after Reionization. *Physics Review Letters*, 100 (16):161301, April 2008. doi: 10.1103/PhysRevLett.100.161301.

Roy Maartens. *Cosmology Lectures (Unpublished)*. 2018.

Yi Mao, Max Tegmark, Matthew McQuinn, Matias Zaldarriaga, and Oliver Zahn. How accurately can 21 cm tomography constrain cosmology? *Phys. Rev. D*, 78:023529, Jul 2008. doi: 10.1103/PhysRevD.78.023529. URL https://link.aps.org/doi/10.1103/PhysRevD.78.023529.

K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek. Measurement of 21 cm Brightness Fluctuations at z ˜ 0.8 in Cross-correlation. *The Astrophysical Journall*, 763:L20, January 2013. doi: 10.1088/2041-8205/763/1/L20. URL http://adsabs.harvard.edu/abs/2013ApJ...763L..20M.

Siyambonga D Matshawule, Marta Spinelli, Mario G Santos, and Sibonelo Ngobese. Hi intensity mapping with MeerKAT: primary beam effects on foreground cleaning. *Monthly Notices of the Royal Astronomical Society*, 506(4):5075–5092, 06 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab1688. URL https://doi.org/10.1093/mnras/stab1688.

T. Mauch, T. Murphy, H. J. Buttery, J. Curran, R. W. Hunstead, B. Piestrzynski, J. G. Robertson, and E. M. Sadler. SUMSS: a wide-field radio imaging survey of the southern sky - II. The source catalogue. *Monthly Notices of the Royal Astronomical Society*, 342:1117–1130, July 2003. doi: 10.1046/j.1365-8711.2003.06605.x. URL http://adsabs.harvard.edu/abs/2003MNRAS.342.1117M.

T. Mauch, W. D. Cotton, J. J. Condon, A. M. Matthews, T. D. Abbott, R. M. Adam, M. A. Aldera, K. M. B. Asad, E. F. Bauermeister, T. G. H. Bennett, H. Bester, D. H. Botha, L. R. S. Brederode, Z. B. Brits, S. J. Buchner, J. P. Burger, F. Camilo, J. M. Chalmers, T. Cheetham, D. de Villiers, M. S. de Villiers, M. A. Dikgale-Mahlakoana, L. J. du Toit, S. W. P. Esterhuyse, G. Fadana, B. L. Fanaroff, S. Fataar, S. February, B. S. Frank, R. R. G. Gamatham, M. Geyer, S. Goedhart, S. Gounden, S. C. Gumede, I. Heywood, M. J. Hlakola, J. M. G. Horrell, B. Hugo, A. R. Isaacson, G. I. G. Józsa, J. L. Jonas, R. P. M. Julie, F. B. Kapp, V. A. Kasper, J. S. Kenyon, P. P. A. Kotzé, N. Kriek, H. Kriel, T. W. Kusel, R. Lehmensiek, A. Loots, R. T. Lord, B. M. Lunsky, K. Madisa, L. G. Magnus, J. P. L. Main, J. A. Malan, J. R. Manley, S. J. Marais, A. Martens, B. Merry, R. Millenaar, N. Mnyandu, I. P. T.

Moeng, O. J. Mokone, T. E. Monama, M. C. Mphego, W. S. New, B. Ngcebetsha, K. J. Ngoasheng, M. T. O. Ockards, N. Oozeer, A. J. Otto, A. A. Patel, A. Peens-Hough, S. J. Perkins, A. J. T. Ramaila, Z. R. Ramudzuli, R. Renil, L. L. Richter, A. Robyntjies, S. Salie, C. T. G. Schollar, L. C. Schwardt, M. Serylak, R. Siebrits, S. K. Sirothia, O. M. Smirnov, L. Sofeya, G. Stone, B. Taljaard, C. Tasse, I. P. Theron, A. J. Tiplady, O. Toruvanda, S. N. Twum, T. J. van Balla, A. van der Byl, C. van der Merwe, V. Van Tonder, B. H. Wallace, M. G. Welz, L. P. Williams, and B. Xaia. The 1.28 GHz MeerKAT DEEP2 image. *The Astrophysical Journal*, 888(2):61, jan 2020. doi: 10.3847/1538-4357/ab5d2d. URL https://doi.org/10.3847/1538-4357/ab5d2d.

Matthew McQuinn, Oliver Zahn, Matias Zaldarriaga, Lars Hernquist, and Steven R. Furlanetto. Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *The Astrophysical Journal*, 653(2):815–834, Dec 2006. ISSN 1538-4357. doi: 10.1086/505167. URL http://dx.doi.org/10.1086/505167.

Andrei Mesinger, Bradley Greig, and Emanuele Sobacchi. The Evolution Of 21 cm Structure (EOS): public, large-scale simulations of Cosmic Dawn and reionization. *Monthly Notices of the Royal Astronomical Society*, 459(3):2342–2353, July 2016. doi: 10.1093/mnras/stw831.

Rendong Nan, Di Li, Chengjin Jin, Qiming Wang, Lichun Zhu, Wenbai Zhu, Haiyan Zhang, Youling Yue, and Lei Qian. The Five-Hundred Aperture Spherical Radio Telescope (fast) Project. *International Journal of Modern Physics D*, 20(6):989–1024, January 2011. doi: 10.1142/S0218271811019335.

L. B. Newburgh, K. Bandura, M. A. Bucher, T. C. Chang, H. C. Chiang, J. F. Cliche, R. Davé, M. Dobbs, C. Clarkson, K. M. Ganga, T. Gogo, A. Gumba, N. Gupta, M. Hilton, B. Johnstone, A. Karastergiou, M. Kunz, D. Lokhorst, R. Maartens, S. Macpherson, M. Mdlalose, K. Moodley, L. Ngwenya, J. M. Parra, J. Peterson, O. Recnik, B. Saliwanchik, M. G. Santos, J. L. Sievers, O. Smirnov, P. Stronkhorst, R. Taylor, K. Vanderlinde, G. Van Vuuren, A. Weltman, and A. Witzemann. HIRAX: a probe of dark energy and radio transients. In , volume 9906 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 99065X, August 2016. doi: 10.1117/12.2234286.

C. Ng, A. Pandhi, A. Naidu, E. Fonseca, V. M. Kaspi, K. W. Masui, R. Mckinven, A. Renard, P. Scholz, I. H. Stairs, S. P. Tendulkar, and K. Vanderlinde. Faraday rotation measures of Northern hemisphere pulsars using CHIME/Pulsar. *Monthly Notices of the Royal Astronomical Society*, 496(3):2836–2848, August 2020. doi: 10.1093/mnras/staa1658.

Sibonelo Ngobese. Point source simulations and Foreground cleaning techniques for HI intensity mapping. Master's thesis, University of the Western Cape, Robert Sobukwe Rd Bellville Western Cape, South Africa, 2018. URL http://hdl.handle.net/11394/6216.

C. D. Nunhokee, G. Bernardi, S. A. Kohn, J. E. Aguirre, N. Thyagarajan, J. S. Dillon, G. Foster, T. L. Grobler, J. Z. E. Martinot, and A. R. Parsons. Constraining polarized foregrounds for EoR experiments. II. polarization leakage simulations in the avoidance scheme. *The Astrophysical Journal*, 848(1):47, oct 2017. doi: 10.3847/1538-4357/aa8b73. URL https://doi.org/10.3847/1538-4357/aa8b73.

Ue-Li Pen, Lister Staveley-Smith, Jeffrey B. Peterson, and Tzu-Ching Chang. First detection of cosmic structure in the 21-cm intensity field. *Monthly Notices of the Royal Astronomical Society*, 394(1): L6–L10, March 2009. doi: 10.1111/j.1745-3933.2008.00581.x.

J. B. Peterson, R. Aleksan, R. Ansari, K. Bandura, D. Bond, J. Bunton, K. Carlson, T.-C. Chang, F. DeJongh, M. Dobbs, S. Dodelson, H. Darhmaoui, N. Gnedin, M. Halpern, C. Hogan, J.-M. Le Goff, T. T. Liu, A. Legrouri, A. Loeb, K. Loudiyi, C. Magneville, J. Marriner, D. P. McGinnis, B. McWilliams, M. Moniez, N. Palanque-Delabruille, R. J. Pasquinelli, U.-L. Pen, J. Rich, V. Scarpine, H.-J. Seo, K. Sigurdson, U. Seljak, A. Stebbins, J. H. Steffen, C. Stoughton, P. T. Timbie, A. Vallinotto, and C. Teche. 21-cm Intensity Mapping. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *ArXiv Astrophysics e-prints*, 2009.

Planck Collaboration. Planck 2013 results. xvi. cosmological parameters. *Astronomy & Astrophysics*, 571: A16, Oct 2014. ISSN 1432-0746. doi: 10.1051/0004-6361/201321591. URL http://dx.doi.org/10.1051/0004-6361/201321591.

Planck Collaboration. Planck 2018 results. *Astronomy & Astrophysics*, 641:A6, Sep 2020. ISSN 1432-0746. doi: 10.1051/0004-6361/201833910. URL http://dx.doi.org/10.1051/0004-6361/201833910.

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, J. Chluba, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, M. Farhang, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, T. Ghosh, M. Giard, Y. Giraud-Héraud, E. Giusarma, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marchini, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, B. Rouillé d'Orfeuil, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, N. Said, V. Salvatelli, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. XIII. Cosmological parameters. *Astronomy & Astrophysics*, 594:A13, September 2016. doi: 10.1051/0004-6361/201525830.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M.

White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

A. Popping, R. Davé, R. Braun, and B. D. Oppenheimer. The simulated H I sky at low redshift. *Astronomy & Astrophysics*, 504(1):15–32, September 2009. doi: 10.1051/0004-6361/200911811.

Alkistis Pourtsidou. HI Intensity Mapping with MeerKAT. *arXiv e-prints*, art. arXiv:1709.07316, September 2017.

Abhishek Prakash, Timothy C. Licquia, Jeffrey A. Newman, Ashley J. Ross, Adam D. Myers, Kyle S. Dawson, Jean-Paul Kneib, Will J. Percival, Julian E. Bautista, Johan Comparat, Jeremy L. Tinker, David J. Schlegel, Rita Tojeiro, Shirley Ho, Dustin Lang, Sandhya M. Rao, Cameron K. McBride, Guangtun Ben Zhu, Joel R. Brownstein, Stephen Bailey, Adam S. Bolton, Timothée Delubac, Vivek Mariappan, Michael R. Blanton, Beth Reid, Donald P. Schneider, Hee-Jong Seo, Aurelio Carnero Rosell, and Francisco Prada. The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Luminous Red Galaxy Target Selection. *The Astrophysical Journal Supplement Series*, 224(2):34, June 2016. doi: 10.3847/0067-0049/224/2/34.

J. R. Pritchard and A. Loeb. Evolution of the 21cm signal throughout cosmic history. *Physics Review D*, 78(10):103511, November 2008. doi: 10.1103/PhysRevD.78.103511. URL http://cdsads.u-strasbg.fr/abs/2008PhRvD..78j3511P.

George B. Rybicki and Alan P. Lightman. *Radiative processes in astrophysics*. John Wiley & Sons, 1979. ISBN 0471827592.

Barbara Ryden. *Introduction to cosmology*. Addison-Wesley, 2006. ISBN 0805389121.

M. Santos, P. Bull, D. Alonso, S. Camera, P. Ferreira, G. Bernardi, R. Maartens, M. Viel, F. Villaescusa-Navarro, F. B. Abdalla, M. Jarvis, R. B. Metcalf, A. Pourtsidou, and L. Wolz. Cosmology from a SKA HI intensity mapping survey. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 19, April 2015.

M. G. Santos, A. Cooray, and L. Knox. Multifrequency Analysis of 21 Centimeter Fluctuations from the Era of Reionization. *The Astrophysical Journal*, 625:575–587, June 2005. doi: 10.1086/429857. URL http://adsabs.harvard.edu/abs/2005ApJ...625..575S.

M. G. Santos, M. Cluver, M. Hilton, M. Jarvis, G. I. G. Jozsa, L. Leeuw, O. Smirnov, R. Taylor, F. Abdalla, J. Afonso, D. Alonso, D. Bacon, B. A. Bassett, G. Bernardi, P. Bull, S. Camera, H. C. Chiang, S. Colafrancesco, P. G. Ferreira, J. Fonseca, K. van der Heyden, I. Heywood, K. Knowles, M. Lochner, Y.-Z. Ma, R. Maartens, S. Makhathini, K. Moodley, A. Pourtsidou, M. Prescott, J. Sievers, K. Spekkens, M. Vaccari, A. Weltman, I. Whittam, A. Witzemann, L. Wolz, and J. T. L. Zwart. MeerKLASS: MeerKAT Large Area Synoptic Survey. *ArXiv e-prints*, September 2017.

Hee-Jong Seo, Scott Dodelson, John Marriner, Dave Mcginnis, Albert Stebbins, Chris Stoughton, and Alberto Vallinotto. A ground-based 21 cm baryon acoustic oscillation survey. *The Astrophysical Journal*, 721(1):164–173, Aug 2010. ISSN 1538-4357. doi: 10.1088/0004-637x/721/1/164. URL http://dx.doi.org/10.1088/0004-637X/721/1/164.

P. Serra, W. J. G. de Blok, G. L. Bryan, S. Colafrancesco, R. J. Dettmar, B. S. Frank, F. Govoni, G. I. G. Jozsa, R. C. Kraan-Korteweg, F. M. Maccagni, S. I. Loubser, M. Murgia, T. A. Oosterloo, R. F. Peletier, R. Pizzo, L. Richter, M. Ramatsoku, M. W. L. Smith, S. C. Trager, J. H. van Gorkom, and M. A. W. Verheijen. The MeerKAT Fornax Survey. In *MeerKAT Science: On the Pathway to the SKA*, page 8, January 2016.

J. Richard Shaw, Kris Sigurdson, Michael Sitwell, Albert Stebbins, and Ue-Li Pen. Coaxing cosmic 21 cm fluctuations from the polarized sky using m -mode analysis. *Physics Review D*, 91(8):083514, April 2015. doi: 10.1103/PhysRevD.91.083514.

George F. Smoot and Ivan Debono. 21 cm intensity mapping with the Five hundred metre Aperture Spherical Telescope. *Astronomy & Astrophysics*, 597:A136, January 2017. doi: 10.1051/0004-6361/201526794.

Paula S. Soares, Catherine A. Watkinson, Steven Cunnington, and Alkistis Pourtsidou. Gaussian Process Regression for foreground removal in H I Intensity Mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 510(4):5872–5890, March 2022. doi: 10.1093/mnras/stab2594.

M Spinelli, G Bernardi, and M G Santos. Simulations of galactic polarized synchrotron emission for epoch of reionization observations. *Monthly Notices of the Royal Astronomical Society*, 479(1):275–283, 2018. doi: 10.1093/mnras/sty1457. URL http://dx.doi.org/10.1093/mnras/sty1457.

Marta Spinelli, Anna Zoldan, Gabriella De Lucia, Lizhi Xie, and Matteo Viel. The atomic hydrogen content of the post-reionization Universe. *Monthly Notices of the Royal Astronomical Society*, 493(4): 5434–5455, April 2020. doi: 10.1093/mnras/staa604.

Marta Spinelli, Isabella P. Carucci, Steven Cunnington, Stuart E. Harper, Melis O. Irfan, José Fonseca, Alkistis Pourtsidou, and Laura Wolz. SKAO H I intensity mapping: blind foreground subtraction challenge. *Monthly Notices of the Royal Astronomical Society*, 509(2):2048–2074, January 2022. doi: 10.1093/mnras/stab3064.

Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, Shy Genel, Paul Torrey, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 475(1):676–698, March 2018. doi: 10.1093/mnras/stx3304.

Square Kilometre Array Cosmology Science Working Group, David J. Bacon, Richard A. Battye, Philip Bull, Stefano Camera, Pedro G. Ferreira, Ian Harrison, David Parkinson, Alkistis Pourtsidou, Mário G. Santos, Laura Wolz, Filipe Abdalla, Yashar Akrami, David Alonso, Sambatra Andrianomena, Mario Ballardini, José Luis Bernal, Daniele Bertacca, Carlos A. P. Bengaly, Anna Bonaldi, Camille Bonvin, Michael L. Brown, Emma Chapman, Song Chen, Xuelei Chen, Steven Cunnington, Tamara M. Davis, Clive Dickinson, José Fonseca, Keith Grainge, Stuart Harper, Matt J. Jarvis, Roy Maartens, Natasha Maddox, Hamsa Padmanabhan, Jonathan R. Pritchard, Alvise Raccanelli, Marzia Rivi, Sambit Roychowdhury, Martin Sahlén, Dominik J. Schwarz, Thilo M. Siewert, Matteo Viel, Francisco Villaescusa-Navarro, Yidong Xu, Daisuke Yamauchi, and Joe Zuntz. Cosmology with Phase 1 of the Square Kilometre Array Red Book 2018: Technical specifications and performance forecasts. *Publications of the Astronomical Society of Australia*, 37:e007, March 2020. doi: 10.1017/pasa.2019.51.

E. R. Switzer, K. W. Masui, K. Bandura, L.-M. Calin, T.-C. Chang, X.-L. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek. Determination of z ˜ 0.8 neutral hydrogen fluctuations using the 21 cm intensity mapping autocorrelation. *Monthly Notices of the Royal Astronomical Society*, 434:L46–L50, July 2013. doi: 10.1093/mnrasl/slt074. URL http://cdsads.u-strasbg.fr/abs/2013MNRAS.434L..46S.

E. R. Switzer, T. C. Chang, K. W. Masui, U. L. Pen, and T. C. Voytek. Interpreting the Unresolved Intensity of Cosmologically Redshifted Line Radiation. *The Astrophysical Journal*, 815(1):51, December 2015. doi: 10.1088/0004-637X/815/1/51.

A. R. Taylor, J. M. Stil, and C. Sunstrum. A Rotation Measure Image of the Sky. *The Astrophysical Journal*, 702:1230–1236, September 2009. doi: 10.1088/0004-637X/702/2/1230. URL http://adsabs.harvard.edu/abs/2009ApJ...702.1230T.

Max Tegmark. How to Make Maps from Cosmic Microwave Background Data without Losing Information. , 480(2):L87–L90, May 1997. doi: 10.1086/310631.

B. Thorne, J. Dunkley, D. Alonso, and S. Næss. The Python Sky Model: software for simulating the Galactic microwave sky. *Monthly Notices of the Royal Astronomical Society*, 469(3):2821–2833, 05 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx949. URL https://doi.org/10.1093/mnras/stx949.

C. Megan Urry and Paolo Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 107:803, September 1995. doi: 10.1086/133630.

Francisco Villaescusa-Navarro, Matteo Viel, Kanan K. Datta, and T. Roy Choudhury. Modeling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping. *Journal of Cosmology and Astroparticle Physics*, 2014(9):050, September 2014. doi: 10.1088/1475-7516/2014/09/050.

Francisco Villaescusa-Navarro, David Alonso, and Matteo Viel. Baryonic acoustic oscillations from 21 cm intensity mapping: the Square Kilometre Array case. *Monthly Notices of the Royal Astronomical Society*, 466(3):2736–2751, April 2017. doi: 10.1093/mnras/stw3224.

Francisco Villaescusa-Navarro, Shy Genel, Emanuele Castorina, Andrej Obuljen, David N. Spergel, Lars Hernquist, Dylan Nelson, Isabella P. Carucci, Annalisa Pillepich, Federico Marinacci, Benedikt Diemer, Mark Vogelsberger, Rainer Weinberger, and Rüdiger Pakmor. Ingredients for 21 cm Intensity Mapping. *The Astrophysical Journal*, 866(2):135, October 2018. doi: 10.3847/1538-4357/aadba0.

Benjamin D. Wandelt and Krzysztof M. Górski. Fast convolution on the sphere. *Physical Review D*, 63 (12), May 2001. ISSN 1089-4918. doi: 10.1103/physrevd.63.123002. URL http://dx.doi.org/10.1103/PhysRevD.63.123002.

Jingying Wang, Haiguang Xu, Tao An, Junhua Gu, Xueying Guo, Weitian Li, Yu Wang, Chengze Liu, Olivier Martineau-Huynh, and Xiang-Ping Wu. Exploring the Cosmic Reionization Epoch in Frequency Space: An Improved Approach to Remove the Foreground in 21 cm Tomography. *The Astrophysical Journal*, 763(2):90, February 2013. doi: 10.1088/0004-637X/763/2/90.

Jingying Wang, Mario G. Santos, Philip Bull, Keith Grainge, Steven Cunnington, José Fonseca, Melis O. Irfan, Yichao Li, Alkistis Pourtsidou, Paula S. Soares, Marta Spinelli, Gianni Bernardi, and Brandon Engelbrecht. H I intensity mapping with MeerKAT: calibration pipeline for multidish autocorrelation observations. *Monthly Notices of the Royal Astronomical Society*, 505(3):3698–3721, August 2021. doi: 10.1093/mnras/stab1365.

Xiaomin Wang, Max Tegmark, Mário G. Santos, and Lloyd Knox. 21 cm Tomography with Foregrounds. *The Astrophysical Journal*, 650(2):529–537, October 2006. doi: 10.1086/506597.

R. J. Wilman, L. Miller, M. J. Jarvis, T. Mauch, F. Levrier, F. B. Abdalla, S. Rawlings, H.-R. Klöckner, D. Obreschkow, D. Olteanu, and S. Young. A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes. *Monthly Notices of the Royal Astronomical Society*, 388:1335–1348, August 2008. doi: 10.1111/j.1365-2966.2008.13486.x. URL http://adsabs.harvard.edu/abs/2008MNRAS.388.1335W.

T. L. Wilson and Huttemeister S. Rohlfs K. *Tools of Radio Astronomy*. Springer Berlin Heidelberg, 2013. ISBN 9783540851226. doi: https://doi.org/10.1007/978-3-540-85122-6.

L. Wolz, F. B. Abdalla, C. Blake, J. R. Shaw, E. Chapman, and S. Rawlings. The effect of foreground subtraction on cosmological measurements from intensity mapping. *Monthly Notices of the Royal Astronomical Society*, 441(4):3271–3283, July 2014. doi: 10.1093/mnras/stu792.

L. Wolz, C. Blake, F. B. Abdalla, C. J. Anderson, T. C. Chang, Y. C. Li, K. W. Masui, E. Switzer, U. L. Pen, T. C. Voytek, and J. Yadav. Erasing the Milky Way: new cleaning technique applied to GBT intensity mapping data. *Monthly Notices of the Royal Astronomical Society*, 464(4):4938–4949, February 2017. doi: 10.1093/mnras/stw2556.

Laura Wolz, Alkistis Pourtsidou, Kiyoshi W. Masui, Tzu-Ching Chang, Julian E. Bautista, Eva-Maria Mueller, Santiago Avila, David Bacon, Will J. Percival, Steven Cunnington, Chris Anderson, Xuelei Chen, Jean-Paul Kneib, Yi-Chao Li, Yu-Wei Liao, Ue-Li Pen, Jeffrey B. Peterson, Graziano Rossi, Donald P. Schneider, Jaswant Yadav, and Gong-Bo Zhao. HI constraints from the cross-correlation of eBOSS galaxies and Green Bank Telescope intensity maps. *arXiv e-prints*, art. arXiv:2102.04946, February 2021.

Fengquan Wu, Jixia Li, Shifan Zuo, Xuelei Chen, Santanu Das, John P. Marriner, Trevor M. Oxholm, Anh Phan, Albert Stebbins, Peter T. Timbie, Reza Ansari, Jean-Eric Campagne, Zhiping Chen, Yanping Cong, Qizhi Huang, Juhun Kwak, Yichao Li, Tao Liu, Yingfeng Liu, Chenhui Niu, Calvin Osinga, Olivier Perdereau, Jeffrey B. Peterson, John Podczerwinski, Huli Shi, Gage Siebert, Shijie Sun, Haijun Tian, Gregory S. Tucker, Qunxiong Wang, Rongli Wang, Yougang Wang, Yanlin Wu, Yidong Xu, Kaifeng Yu, Zijie Yu, Jiao Zhang, Juyong Zhang, and Jialu Zhu. The Tianlai dish pathfinder array: design, operation, and performance of a prototype transit radio interferometer. *Monthly Notices of the Royal Astronomical Society*, 506(3):3455–3482, September 2021. doi: 10.1093/mnras/stab1802.

J. Stuart B. Wyithe and Abraham Loeb. Fluctuations in 21-cm emission after reionization. *Monthly Notices of the Royal Astronomical Society*, 383(2):606–614, January 2008. doi: 10.1111/j.1365-2966.2007.12568.x.

J. Stuart B. Wyithe, Abraham Loeb, and Paul M. Geil. Baryonic acoustic oscillations in 21-cm emission: a probe of dark energy out to high redshifts. *Monthly Notices of the Royal Astronomical Society*, 383 (3):1195–1209, January 2008. doi: 10.1111/j.1365-2966.2007.12631.x.

Elimboto Yohana, Yi-Chao Li, and Yin-Zhe Ma. Forecasts of cosmological constraints from HI intensity mapping with FAST, BINGO and SKA-I. *Research in Astronomy and Astrophysics*, 19(12):186, December 2019. doi: 10.1088/1674-4527/19/12/186.

Andrea Zonca, Leo Singer, Daniel Lenz, Martin Reinecke, Cyrille Rosset, Eric Hivon, and Krzysztof Gorski. healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python. *Journal of Open Source Software*, 4(35):1298, March 2019. doi: 10.21105/joss.01298. URL https://doi.org/10.21105/joss.01298.