MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2023/05.16.15.08-TDI

TESTING THE RECOVERY OF THE BINGO 21CM COSMOLOGICAL SIGNAL FROM DIFFERENT MISSION SIMULATIONS

Eduardo Jubini de Merícia

Doctorate Thesis of the Graduate Course in Astrophysics, guided by Drs. Carlos Alexandre Wuensche de Souza, and Vincenzo Liccardo, approved in May 05, 2023.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/4958TLP>

INPE
São José dos Campos
2023

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2023/05.16.15.08-TDI

TESTING THE RECOVERY OF THE BINGO 21CM COSMOLOGICAL SIGNAL FROM DIFFERENT MISSION SIMULATIONS

Eduardo Jubini de Merícia

Doctorate Thesis of the Graduate Course in Astrophysics, guided by Drs. Carlos Alexandre Wuensche de Souza, and Vincenzo Liccardo, approved in May 05, 2023.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/4958TLP>

INPE
São José dos Campos
2023

Cataloging in Publication Data

Merícia, Eduardo Jubini de.

M543t Testing the recovery of the BINGO 21cm cosmological signal from different mission simulations / Eduardo Jubini de Merícia. – São José dos Campos : INPE, 2023.

xxii + 104 p. ; (sid.inpe.br/mtc-m21d/2023/05.16.15.08-TDI)

Thesis (Doctorate in Astrophysics) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2023.

Guiding : Drs. Carlos Alexandre Wuensche de Souza, and Vincenzo Liccardo.

1. 21 cm cosmology. 2. Baryon acoustic oscillations. 3. Radio astronomy. 4. BINGO telescope. 5. Component separation. I.Title.

CDU 524.8

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS
Serviço de Pós-Graduação - SEPGR

DEFESA FINAL DE TESE DE EDUARDO JUBINI DE MERICIA
REG. 138576/2019, BANCA Nº 083/2023

No dia 05 de maio de 2023, por videoconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutor em Astrofísica, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

Novo Título: "Testing the Recovery of the BINGO 21cm Cosmological Signal from Different Mission Simulations"

Observações da banca: A banca examinadora fez algumas considerações para correção do texto e esclarecimentos mais detalhados em alguns parágrafos. As recomendações foram encaminhadas ao candidato a seus orientadores para que avaliem a pertinência ou não de incorporar ao texto final.

Membros da banca:

Dr. Oswaldo Duarte Miranda – Presidente – INPE
Dr. Carlos Alexandre Wuensche de Souza – Orientador INPE
Dr. Vincenzo Liccardo – Orientador – INPE
Dr. Rafael da Costa Nunes – Membro interno – INPE
Dra. Larissa Carlos de Oliveira Santos – Membro externo – YangZhou University/China
Dr. Thyrso Villela Neto – Membro externo – CGEE/MCTI
Dr. Miguel Boavista Quartin – Membro externo – IF/UFRJ

Documento assinado eletronicamente por **Carlos Alexandre Wuensche de Souza**, **Pesquisador**, em 11/05/2023, às 10:45 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Oswaldo Duarte Miranda**, **Coordenador do Gabinete do Instituto Nacional de Pesquisas Espaciais**, em 11/05/2023, às 11:06 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Rafael da Costa nunes (E)**, **Usuário Externo**, em 11/05/2023, às 11:19 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **MIGUEL BOAVISTA QUARTIN (E)**, **Usuário Externo**, em 11/05/2023, às 16:23 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Vincenzo liccardo (E)**, **Usuário Externo**, em 23/05/2023, às 08:17 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Larissa Carlos de Oliveira Santos (E)**, **Usuário Externo**, em 18/06/2023, às 23:30 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Thyrso villela neto (E)**, **Usuário Externo**, em 20/06/2023, às 11:59 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.mcti.gov.br/verifica.html, informando o código verificador **11028036** e o código CRC **85E5CB51**.

Referência: Processo nº 01340.003329/2023-39 SEI nº 11028036

À minha amada mãe ***Sandra*** *(in memoriam).*

ACKNOWLEDGEMENTS

À minha esposa Mayara, por todo amor e dedicação. Sem você nada disso seria possível.

A meu irmão Everton, pela cumplicidade de uma vida toda.

Aos amigos Daniel e Paty, pela conexão que resiste ao tempo e à distância, por me ensinarem o real significado da palavra amizade.

Aos amigos Samantha e Bruno, pela relação leve e profunda, por sempre me acolherem tão bem em São José dos Campos.

À dona Marilene, que cuidou de mim como uma mãe durante a produção desse trabalho.

Aos amigos Williams e Simone, por todo apoio e pelos momentos inesquecíves.

Ao amigo Fernando Machado, pelos ótimos bate-papos e cafés, por todo o apoio dado durante a pandemia.

A meu pai Ricardo, pelas mensagens de apoio e por sempre ter me incentivado a estudar.

A meu orientador Dr. Carlos Alexandre Wuensche, pelos ensinamentos sobre ciência e sobre a vida, pela forma sábia como conduziu as situações mais adversas que enfrentamos nesses últimos anos.

A meu orientador Dr. Vincenzo Liccardo, pelas longas e animadas discussões que iam desde Cosmologia e programação a futebol.

Ao amigo Fred Vieira (*in memoriam*), por todos os momentos incríveis vividos dentro e fora do INPE. Você faz muita falta.

Aos amigos Felipe e Fernando, pela parceria nos momentos mais difíceis e também nos mais divertidos.

À Dra. Larissa Santos e à Dra. Camila Novaes, por todo o suporte, conversas e conselhos, fundamentais para a construção deste trabalho.

Ao companheiro Reitano, por todas as conversas e churrascos maravilhosos.

Ao companheiro Telmo, pelo bom humor contagiante e pelo suporte ao acesso remoto dado durante a pandemia.

Aos professores do curso de pós graduação em Astrofísica e aos demais funcionários do INPE, por contribuírem para minha formação profissional e pessoal.

À CAPES pelo suporte financeiro.

ABSTRACT

The ΛCDM cosmological model predicts that the so-called dark energy (DE) corresponds to about 70% of the energy density of the universe and it is responsible for its observed accelerated expansion. The DE effects on cosmic evolution can be assessed using baryon acoustic oscillations (BAO) measurements. BINGO (BAO from Integrated Neutral Gas Observations) telescope was designed to detect these oscillations using the intensity mapping (IM) technique to observe the 21 cm neutral hydrogen (HI) line, one of the most powerful probes to explore the universe, from recombination to the present times. As a transit telescope, BINGO will observe a fixed declination strip of 15°, centered at $\delta = -15°$, covering about 13% of the sky and probing a redshift interval $0.127 \leq z \leq 0.449$, which corresponds to the frequency range $980 \leq \nu \leq 1260$ MHz, where the BAO signatures should appear with a good signal-to-noise ratio. The BINGO observations at radio frequencies are dominated by strong galactic emissions, which are about 10^4 more intense than the cosmological signal. Therefore, defining an efficient component separation strategy to recover the 21 cm signal from the BINGO data is a crucial step for BAO detection. This work analyses the performance of the Generalized Needlet Internal Linear Combination (`GNILC`), a non-parametric method that combines spatial and frequency information from the components to identify and remove foregrounds, combined with a power spectrum debiasing procedure, in the recovery of the 21 cm power spectrum. This component separation process was tested in BINGO simulated data (HI + foregrounds + thermal noise) produced with different foregrounds models and frequency bin numbers. Furthermore, the effect of using ancillary data from C-Band All-Sky Survey (CBASS) experiment to improve the 21 cm signal estimation was also verified. These tests aim to optimize both the foreground removal and recovery of the 21 cm signal across the full BINGO frequency band, as well as to determine an optimal number of frequency (redshift) bands, which impacts greatly the decisions regarding BINGO hardware configuration before commissioning. The obtained results show that the component separation process adopted is able to recover the 21 cm power spectra from the simulated data, with an average error of 3%, calculated over the BINGO frequency band and in the angular scales interval $30 \leq \ell \leq 270$.

Keywords: 21 cm cosmology. Baryon acoustic oscillations. Radio astronomy. BINGO telescope. Component separation. Foreground removal.

TESTANDO A RECUPERAÇÃO DO SINAL COSMOLÓGICO DE 21 CM DO BINGO A PARTIR DE DIFERENTES SIMULAÇÕES DE MISSÃO

RESUMO

O modelo cosmológico ΛCDM prevê que a chamada energia escura (EE) corresponde a cerca de 70% da densidade de energia do universo e é responsável pela observada expansão acelerada do mesmo. Os efeitos da EE na evolução cósmica podem ser avaliados usando medições de oscilações acústicas bariônicas (do inglês, *baryonic acoustic oscillations* - BAO). O telescópio BINGO (do inglês, *BAO from Integrated Neutral Gas Observations*) foi projetado para detectar essas oscilações usando a técnica de mapeamento de intensidade (do inglês, *intensity mapping* - IM) para observar a linha de 21 cm do hidrogênio neutro (HI), uma das sondas mais poderosas para explorar o universo, desde a recombinação até os tempos atuais. Como um telescópio de trânsito, o BINGO observará uma faixa de declinação fixa de 15°, centrada em $\delta = -15°$, cobrindo cerca de 13% do céu e sondando um intervalo de *redshift* $0.127 \leq z \leq 0.449$, que corresponde à faixa de frequência $980 \leq \nu \leq 1260$ MHz, onde as assinaturas BAO devem aparecer com uma boa relação sinal-ruído. As observações do BINGO em frequências de rádio são dominadas por fortes emissões galácticas, que são cerca de 10^4 mais intensas que o sinal cosmológico. Portanto, definir uma estratégia de separação de componentes eficiente para recuperar o sinal de 21 cm a partir dos dados do BINGO é uma etapa crucial para a detecção de BAO. Este trabalho analisa o desempenho do *Generalized Needlet Internal Linear Combination* (`GNILC`), um método não paramétrico que combina informações espaciais e de frequência dos componentes para identificar e remover foregrounds, combinado com um procedimento de debiasing de espectro de potência, na recuperação do espectro de potência de 21 cm. Este processo de separação de componentes foi testado em dados simulados do BINGO (HI + foregrounds + ruído térmico) produzidos com diferentes modelos de foregrounds e números de bins de frequência. Além disso, o efeito do uso de dados auxiliares do experimento CBASS para melhorar a estimativa do sinal de 21 cm também foi verificado. Esses testes visam otimizar a remoção de *foregrounds* e a recuperação do sinal de 21 cm em toda a banda de frequência do BINGO, bem como determinar um número ideal de bandas de frequência (*redshift*), o que afeta muito as decisões sobre a configuração do hardware do BINGO antes do comissionamento. Os resultados obtidos mostram que o processo de separação de componentes adotado é capaz de recuperar os espectros de potência de 21 cm a partir dos dados simulados, com um erro médio de 3%, calculado sobre a banda de frequência BINGO e no intervalo de escalas angulares $30 \leq \ell \leq 270$.

Palavras-chave: Cosmologia de 21 cm. Oscilações acústicas de bárions. Radioastronomia. Telescópio BINGO. Separação de componentes. Remoção de foregrounds.

LIST OF FIGURES

Page

LIST OF TABLES

Page

CONTENTS

Page

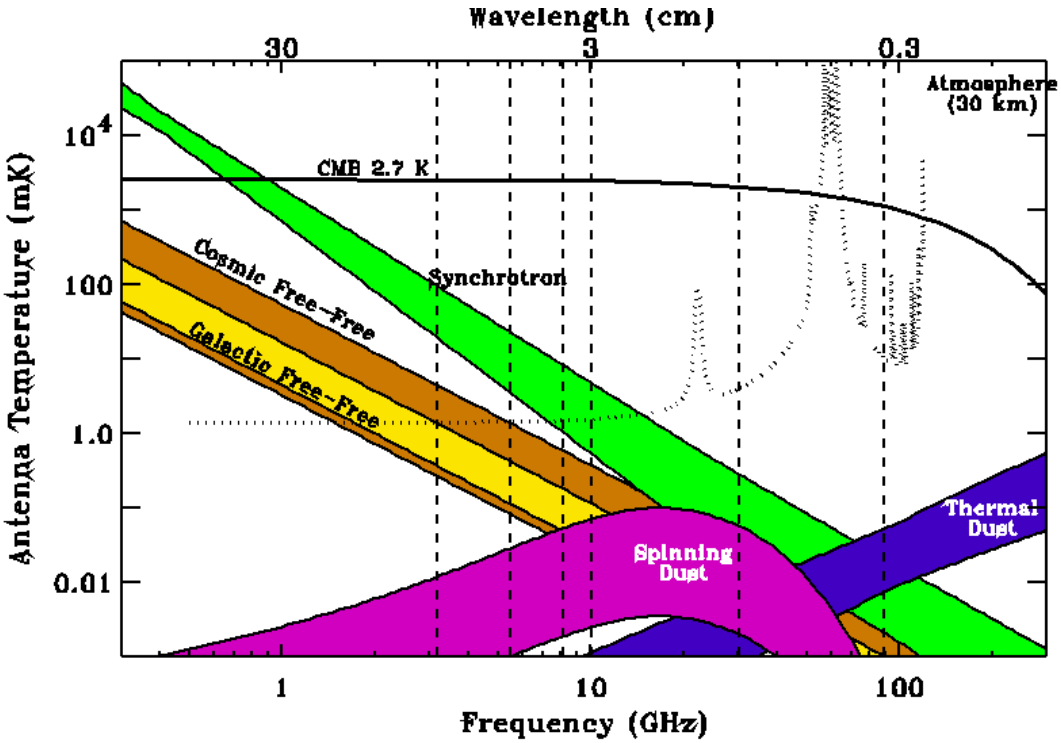
1 INTRODUCTION

There are various cosmological experiments trying to understand different aspects of the origin and evolution of the universe. The current data increasingly strengthen the so-called "concordance model", constraining the cosmological parameters with $\lesssim 1\%$ of precision. However, despite the success of this precision cosmology, some points of the ΛCDM model still need a much deeper investigation. Among them, understanding the nature of dark energy (DE), which comprise about 70% of the present composition of the universe is, arguably, the most important open question both in physics and cosmology. Although its origin is unknown, DE is recognized in the standard model as responsible for the late-time accelerating expansion of the universe.

Baryon acoustic oscillations (BAO) are one of the most efficient observables to measure the history of expansion and growth of structure at low redshifts (WEINBERG et al., 2013). BAO are the imprint left by acoustic waves at recombination and observed in the distribution of matter in the universe. The BAO scale has already been measured in the optical window (EISENSTEIN et al., 2005) and can be detected in radio by integrating the intensity of the 21 cm line of neutral hydrogen (HI) across a large radio frequency band, using a technique known as intensity mapping (IM). It consists in measuring the spatial fluctuations in the emission line from a large sky area at low angular resolution, tracing the evolution of the underlying large scale structures over the redshift interval corresponding the chosen radio frequency band (KOVETZ et al., 2017).

BINGO (BAO from Integrated Neutral Gas Observations) is a radio telescope designed to detect BAO at low redshifts through the intensity mapping technique applied to the HI 21 cm line. However, its observations, made around 1 GHz are contaminated by intense diffuse galactic emissions that dominate the signal collected by any radio telescope. Figure 1.1 shows the most important diffuse cosmic emissions from 300 MHz to 300 GHz (radio and submillimeter bands) measured in antenna temperature. Note that at ~ 1 GHz, the dominant components are, in this order, synchrotron, free-free and CMB emissions.

Figure 1.1 - Radio diffuse cosmic signals (in antenna temperature) in the radio and sub-millimeter bands (300 MHz to 300 GHz).



Source: NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA) (2012).

At BINGO frequencies, the galactic foreground continuum emission is ~ 1 K with spatial fluctuations of ~ 100 mK, while the HI signal is ~ 200 μK (WUENSCHE et al., 2021). This shows that recovering the 21 cm signal from the instrument data is quite a challenging task. Thus, adopting an efficient foreground removal procedure is a crucial step for BAO detection. This work aims to present the results obtained with the combined use of `GNILC` (Generalized Needlet Internal Linear Combination) and a debiasing procedure, to recover the HI power spectrum from different simulated data configurations.

This thesis is structured as follows. Chapter 2 contains a brief description of the physics of the 21 cm line and the intensity mapping technique. Chapter 3 presents a brief description of the BINGO telescope. Chapter 4 describes the models used to produce the simulated BINGO observations, the galactic mask used in the maps and the simulation plan adopted. Chapter 5 presents the foreground removal problem and describes the component separation mehod (`GNILC`) and the power spectrum

debiasing procedure used in this work. Chapter 6 shows the results of the component separation tests carried out with simulated BINGO data. Chapter 7 contains the conclusions and the next steps of this work.

2 21 CM COSMOLOGY

Observations of the redshifted 21 cm line of neutral hydrogen allow us to trace the cosmic evolution, from the recombination period, when the respective atoms were formed, to the post-reionization epoch, when the remaining HI is mostly confined to galaxies, in dense clumps self-shielded from ionizing radiation (PRITCHARD; LOEB, 2012). The 21 cm emission measured after the end of reionization can probe the content of neutral hydrogen in galaxies, serving as a tracer of matter in the universe. However, at these low redshifts ($z \lesssim 1$) only about 1% of the baryons are in the HI form, which considerably reduces the amplitude of the observed signal. Thus, to study the evolution of neutral hydrogen in the post-reionization universe, a technique known as intensity mapping (IM) is used. It maps the distribution of a tracer, which in this work is the 21 cm line, over a volume of the universe, without individually resolving the galaxies that gave rise to it (with low resolution) (PADMANABHAN, 2017). The IM performs a kind of tomography of the distribution of tracer intensities in large-scales, defined over a large area of the sky and in a given range of redshifts.

2.1 The Standard Cosmological Model

The Standard Cosmological Model is supported by the Cosmological Principle, according to which the universe is homogeneous and isotropic on scales greater than 100 Mpc (MUKHANOV, 2005) and whose validity has been proven by several recent observations, which named the model with the term ΛCDM. The greek letter Λ represents a cosmological constant that describes the effect of the still little known dark energy, understood as responsible for the observed acceleration of the expansion of the universe. The acronym CDM for Cold Dark Matter refers to a different type of non-relativistic matter, which is invisible electromagnetically but interacts gravitationally. Three well-established observational pillars support the ΛCDM model: the expansion of the universe, the abundance of light elements and the cosmic microwave background radiation (TURNER, 2022). The ΛCDM model has been constantly tested by a large number of experiments and, so far, has matched the observations with relatively small errors.

The application of General Relativity to cosmic scales allowed the deduction of equations that describe how the content of matter and energy dictates the dynamics and geometry of the universe. The Robertson-Walker metric describes the geometry of an expanding homogeneous and isotropic universe. Thus, the line element $\mathrm{d}s$ that represents the infinitesimal space-time interval between two events in the universe, is given in comoving spherical coordinates by

$$ds^2 = c^2dt^2 - a^2\left(\frac{dr^2}{1-kr^2} + r^2d\theta^2 + r^2\sin^2\theta d\varphi^2\right), \tag{2.1}$$

where dt is the infinitesimal interval of proper time measured by a comoving observer; $a = a(t)$ is the dimensionless scale factor, the ratio of the physical distance between two points at some earlier time t and the comoving distance in present time t_0, such that $a_0 = a(t_0) = 1$ (DODELSON, 2003); r, θ, φ are comoving spatial coordinates; k is the universe curvature parameter, that depends on the matter-energy content of the universe.

The complete dynamic of an expanding homogeneous and isotropic universe is described by the Friedmann equations and the conservation law (continuity equation). The Friedmann equations are obtained combining the Robertson-Walker metric with the Einstein equations written in the cosmological context, which results in

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}G\left(\rho + 3p\right) \tag{2.2}$$

and

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{2.3}$$

where G is the gravitational constant; ρ is the total energy density of the universe; p is the hydrodynamic isotropic pressure; k/a^2 is the Riemannian curvature of space-time. The conservation law is derived from the Einstein field equations and is given by

$$\dot{\rho} = -\frac{3\dot{a}}{a}\left(\rho + p\right). \tag{2.4}$$

The Equation 2.4 can be solved constructing the state equation $p_i = \omega_i\rho_i$ for each component i of the universe. For some cosmological species, ω_i can be considered constant in time. Thus, using the relation $a = (1+z)^{-1}$, one of the solutions for the continuity equation can be written as

$$\rho_i = \rho_{i,0}(1+z)^{3(1+\omega_i)}, \tag{2.5}$$

where $\rho_{i,0}$ is the energy density of the species i in the present day t_0. For non-relativistic matter (baryonic matter and cold dark matter), which will henceforth be called matter, $\omega_{\rm m} = 0$ and $\rho_{\rm m} = \rho_{\rm m,0}(1+z)^3$. For relativistic matter (photons and massless neutrinos), which will henceforth be called radiation, $\omega_{\rm r} = 1/3$ and $\rho_{\rm r} = \rho_{\rm r,0}(1+z)^4$ (RYDEN, 2017).

Unlike the components mentioned earlier, the general solution of the dark energy continuity equation is obtained writing the equation of state with an arbitrary redshift dependence, $\omega_{\rm de}(z)$. Thus, the dimensionless dark energy density is defined as $f(z) = \rho_{\rm de}/\rho_{\rm de,0}$ and is given by

$$f\left(z\right) = \exp\left[3\int_0^z \frac{1+\omega_{\rm de}\left(z'\right)}{1+z'}dz'\right]. \tag{2.6}$$

One of the main goals of modern cosmology is to constrain the possible values of $\omega_{\rm de}(z)$. Recent observations indicates the existence of a positive cosmological constant Λ, such that $\omega_{\rm de}(z) = -1$ and $\rho_{\rm de} = \rho_{\rm de,0} = \rho_\Lambda = \rho_{\Lambda,0} = \Lambda/8\pi G$.

Some parameters need to be presented before going into detail about the dynamics of the universe. The Hubble parameter H measures the rate of expansion and is given by

$$H = \frac{\dot{a}}{a}. \tag{2.7}$$

The Hubble parameter in present time is called Hubble constant, H_0. The density parameter of a component of universe i, $\Omega_{\rm i}$, is is defined as

$$\Omega_{\rm i} = \frac{\rho_i}{\rho_{\rm crit}}, \tag{2.8}$$

where $\rho_{\rm crit}$ is the critical energy density, obtained by making $k = 0$ (flat universe) in the Equation 2.3, which results in

$$\rho_{\rm crit} = \frac{3H^2}{8\pi G}. \tag{2.9}$$

The cosmological parameter Ω, defined as $\Omega \equiv \rho/\rho_{\rm crit}$, where $\rho = \rho_{\rm r}+\rho_{\rm m}+\rho_{\rm de}$, relates

the geometry, the matter and energy content of the universe. universe. Combining this relation with Equation 2.3, it can be rewritten as

$$\Omega = \Omega_{\rm r} + \Omega_{\rm m} + \Omega_{\rm de} = 1 - \Omega_{\rm k}, \tag{2.10}$$

where $\Omega_{\rm r} = \rho_{\rm r}/\rho_{\rm crit}$ is the radiation density parameter; $\Omega_{\rm m} = \rho_{\rm m}/\rho_{\rm crit}$ is matter density parameter; $\Omega_{\rm k} = \rho_{\rm k}/\rho_{\rm crit} = -k/a^2H^2$ is the spatial curvature density parameter; $\Omega_{\rm de} = \rho_{\rm de}/\rho_{\rm crit}$ is the dark energy density parameter, that in the case of the ΛCDM model, $\Omega_{\rm de} = \Omega_\Lambda = \Lambda c^2/3H^2$. If $\Omega > 1$, the universe has closed geometry and $k > 1$; if $\Omega < 1$, the universe has open geometry and $k < 1$; if $\Omega = 1$, the universe has flat geometry and $k = 0$.

The dimensionless deceleration parameter q characterizes the acceleration of cosmic expansion and is defined by

$$q = -\frac{\ddot{a}a}{\dot{a}^2} = \frac{1}{2} + \frac{3}{2}\left(\omega_{\rm r}\Omega_{\rm r} + \omega_{\rm m}\Omega_{\rm m} + \omega_{\rm de}\Omega_{\rm de}\right) + \frac{k}{2a^2H^2}. \tag{2.11}$$

The signal of q determines whether the expansion is slowing down or speeding up (MUKHANOV, 2005). If $\ddot{a} > 0$, the expansion is accelerating and $q < 0$. In the ΛCDM model, it is assumed that $k = 0$ and $\omega_{\rm de} = -1$. Thus, considering that $\Omega_{\rm r,0}$ is negligible, the present value of the deceleration parameter can be written as $q_0 = \Omega_{\rm m,0}/2 - \Omega_{\Lambda,0}$, where is possible to observe the effect of the cosmological constant as a repulsive field that opposes the gravitational attraction.

Rewriting the Equation 2.3, putting the Hubble parameter $H(z)$ in terms of the present values of the density parameters and the Hubble constant, we get

$$H^2(z) = H_0^2\left[\Omega_{\rm r,0}(1+z)^4 + \Omega_{\rm m,0}(1+z)^3 + \Omega_{\rm k,0}(1+z)^2 + \Omega_{\rm de,0}f(z)\right], \tag{2.12}$$

where for the ΛCDM model, $\Omega_{\rm k} = 0$, $f(z) = 1$ and $\Omega_{\rm de,0} = \Omega_{\Lambda,0}$.

One of the main ways to measure $H(z)$ and constrain the cosmological parameters and the nature of dark energy (delimiting $\omega(z)$ values) is through baryon acoustic oscillations (BAO). The details of BAO physics and the use of its characteristic scale as a standard cosmological ruler are presented in Section 2.2.

2.2 Baryon acoustic oscillations

Before the epoch of recombination, the temperature of the universe was so high that protons and electrons formed a hot and dense plasma, tightly coupled with the photons via Thomson scattering. Gravitational potential wells started to emerge as baryons and dark matter were gravitationally attracted to the overdensity regions, arising from quantum fluctuations of the period of cosmic inflation. These wells created pressure in the opposite direction of gravitational attraction by compressing the electron-proton fluid and heating the photons. Then, from these overdensity regions, waves propagating spherically through the plasma were created by the alternation of photon pressure and gravitational force. These spherical density perturbations propagated as acoustic waves with a speed given by (EISENSTEIN et al., 2007)

$$c_{\mathrm{s}}(z) = \frac{c}{\sqrt{3(1+R(z))}}, \tag{2.13}$$

where $R(z) \equiv 3\rho_{\mathrm{b}}/4\rho_{\gamma}$, being that ρ_{b} and ρ_{γ} are the baryons and photons energy densities as a function of redshift, respectively. As $\rho_{\mathrm{b}} \ll \rho_{\gamma}$ at higher redshifts, the acoustic waves started to travel at their maximum speed, $c/\sqrt{3}$. Then, when the universe cooled down enough for the electrons and protons to recombine, which occurred at $z \approx 1100$, the acoustic waves stopped propagating, baryons and photons decoupled, and the latter start to travel freely, forming the observed cosmic microwave background (CMB). The distribution of the baryons after the recombination exhibits a density excess that corresponds to the characteristic radius of the spherical shell that was created when the acoustic waves froze. This radius corresponds to the comoving distance that the sound waves traveled from the Big Bang until the recombination, which occurred at time t_{rec} and redshift z_{rec}, and is given by (WEINBERG et al., 2013)

$$r_{\mathrm{s}} = \int_0^{t_{\mathrm{rec}}} \frac{c_{\mathrm{s}}(t)}{a(t)} dt = \int_{z_{\mathrm{rec}}}^{\infty} \frac{c_{\mathrm{s}}(z)}{H(z)} dz. \tag{2.14}$$

The acoustic scale can be verified as an excess in the distribution of matter (galaxies) in the universe, which is reflected in an increase in the two-point correlation function $\xi(r)$ at a comoving distance $r_{\mathrm{s}} \approx 150$ Mpc. This means that the probability of finding two galaxies separated by the distance r_{s} is higher. Furthermore, harmonic signatures of BAO can be observed in the power spectra of CMB and matter as acoustic peaks.

For the standard model, the comoving line-of-sight distance d_{C} to an object at redshift z can be written as

$$d_{\mathrm{C}}(z) = \frac{c}{H_0} \int_0^z \frac{H_0}{H(z')} dz', \tag{2.15}$$

and the comoving angular diameter distance d_{A}, that relates the comoving size of an object to its angular size, can be written as

$$d_{\mathrm{A}}(z) = \frac{c}{(1+z)H_0\sqrt{-\Omega_{\mathrm{k},0}}} \sin\left[\sqrt{-\Omega_{\mathrm{k},0}} \frac{H_0}{c} d_{\mathrm{C}}(z)\right]. \tag{2.16}$$

As mentioned in the Section 2.1, the BAO scale can be used as a standard ruler to delimit the possible values of $\omega(z)$ and constrain the dark energy properties. A measurement of the Hubble parameter $H(z)$ is given by the acoustic scale along the line-of-sight, $r_{\mathrm{s}_\parallel}$, through

$$H(z) = \frac{c\Delta z}{r_{\mathrm{s}_\parallel}}, \tag{2.17}$$

where Δz is the difference in redshift between the front and back of the BAO spherical shell, and a measurement of the angular diameter distance $d_{\mathrm{A}}(z)$ is given by the tangential characteristic scale $r_{\mathrm{s}_\perp}$ through

$$d_{\mathrm{A}}(z) = \frac{r_{\mathrm{s}_\perp}}{\Delta\theta(1+z)}, \tag{2.18}$$

where $\Delta\theta$ is the angular diameter of the BAO spherical shells. Precise measurements of Δz and $\Delta\theta$, along with the acoustic scale standard value, allow to determine the Hubble parameter and the angular diameter distance of the BAO structure, at different redshifts, through the Equations 2.17 and 2.18. Finally, the Equation 2.16 connects the angular diameter distance, and consequently the BAO scale, to the cosmological parameters.

BAO were first detected as acoustic peaks in the CMB power spectrum, and later, in matter distribution and power spectrum through imaging and spectroscopic redshift surveys. To detect the weak signal of BAO, and obtain precise measurement of r_{s}, it is necessary to probe large volumes of the universe. However, there are some

limitations in the use of galaxy surveys of these cosmic portions. Thus, there are some new experiments, based on IM technique, that intend to apply it to a matter tracer and then, measure the acoustic scale of BAO. As one of those instruments, BINGO intends to make this detection in radio wavelengths, by applying the intensity mapping to the 21 cm line of neutral hydrogen, whose physical foundation is presented in the next section.

2.3 Physics of the 21 cm line

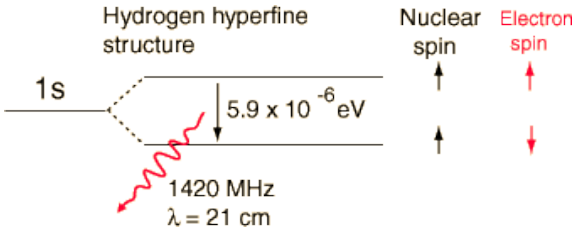
Neutral hydrogen, in its ground energy state, has a proton, with nuclear spin $I = 1/2$, and an electron, with spin $S = 1/2$. As both charged particles have spin, they produce interacting magnetic moments, splitting the ground level $1s$ into two hyperfine states, separated by an amount of energy given by

$$E_{10} = \frac{4}{3}\alpha^4 g_{\rm p}\frac{m_{\rm e}}{m_{\rm p}}m_{\rm e}c^2 = 5.879 \times 10^{-6}\ \text{eV}, \tag{2.19}$$

where α is the dimensionless fine structure constant, $g_{\rm p}$ is the g-factor of a free proton, $m_{\rm e}$ is the mass of the electron, $m_{\rm p}$ is the mass of the proton and c is the speed of light. This splitting is known as the hyperfine ground state structure.

The total angular momentum is given by $F = I + S$, resulting in two possible quantum numbers: $F = 0$ or $F = 1$. The radiation produced by the transition between the two hyperfine levels has a wavelength of 21.106114 cm and, consequently, a frequency of 1420.405752 MHz. Figure 2.1 shows the schematic diagram of the $1s$ ground state hyperfine structure and the transition between the two hiperfine states of HI. On the top line, the hydrogen atom is in the most energetic configuration, $F = 1$, where the proton and electron spins are parallel. On the bottom line, the HI changes to the least energetic configuration (antiparallel spins), $F = 0$, reached after the electron spin flip occurs, and the atom emits a photon with a wavelength $\lambda_{10} \approx 21$ cm or with a frequency $\nu_{10} \approx 1420$ MHz.

Figure 2.1 - Schematic diagram of the hyperfine structure of the ground state $1s$ of the hydrogen atom. A transition between the most energetic state (parallel spins) and the least energetic state (antiparallel spins) emits a photon of 5.9×10^{-6} eV, corresponding to a frequency $\nu_{10} \approx 1420$ MHz and a wavelength $\lambda_{10} \approx 21$ cm.



Source: HyperPhysics (2016).

The 21 cm spectral line can be used as a good tracer of matter in a gas cloud, given a background radio source. Its intensity along the line of sight between the source and the observer, passing through the gas cloud, can be described by the radiative transfer theory.

The variation of the specific intensity $\mathrm{d}I_\nu$, along a distance element $\mathrm{d}s$ over the line of sight is given by the radiative transfer equation

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu + j_\nu, \tag{2.20}$$

where the absorption and emission along the line of sight are given, respectively, by the coefficients α_ν and j_ν.

As in the radio band the specific intensity of a blackbody curve is fully within the Rayleigh-Jeans regime, the relation between the intensity I_ν and the brightness temperature $T_\mathrm{b}(\nu)$ can be written as

$$I_\nu = 2k_\mathrm{B}T_\mathrm{b}(\nu)\frac{\nu^2}{c^2}, \tag{2.21}$$

where c is the speed of light in vacuum and k_B is the Boltzmann constant.

The optical depth τ_ν is defined as

$$\tau_\nu = \int \alpha_\nu (s)\, \mathrm{d}s, \tag{2.22}$$

where s is the position along the line of sight. In a simplified way, the optical depth indicates how difficult it is for a photon to cross a medium without being absorbed. If $\tau_\nu > 1$ the medium is said to be optically thick and photons of frequency ν have a high probability of being absorbed. If $\tau_\nu < 1$, the medium is said to be optically thin and the photons that pass through it have a low probability of being absorbed.

Using the Equations 2.21 and 2.22 to rewrite the radiative transfer equation, one obtains

$$T_\mathrm{b} (\nu) = T_\mathrm{R} (\nu)\, e^{-\tau_\nu} + T_\mathrm{S} \left(1 - e^{-\tau_\nu}\right), \tag{2.23}$$

where $T_\mathrm{b} (\nu)$ is the observed brightness temperature, $T_\mathrm{R} (\nu)$ is the brightness temperature of the background radio source, τ_ν is the optical depth and T_S is the uniform excitation temperature, also known as spin temperature of the gas cloud. It is defined by the ratio between the number densities n_0 and n_1 of the hydrogen atoms at the two hyperfine levels. Subscript 0 refers to the least energetic level $1_0S_{1/2}$ and subscript 1 is associated with the most energetic level $1_1S_{1/2}$.

Considering the system in thermodynamic equilibrium, the relation between the population ratio n_1/n_0 and the spin temperature T_S is given by

$$\frac{n_1}{n_0} = \left(\frac{g_1}{g_0}\right) e^{-T_*/T_\mathrm{S}}, \tag{2.24}$$

where $g_0 = 1$ and $g_1 = 3$ are the numbers of possible states with energy E_0 and E_1, respectively, and $T_* \equiv h\nu_{10}/k_\mathrm{B} = 68$ mK.

The optical depth equation can be rewritten using the explicit form of α_ν, which is obtained from the Equation 2.24 and the Einstein relations. Since in almost all astrophysical situations $T_* \ll T_S$, about 3/4 of the HI atoms are in the excited state. So $n_0 = n_\mathrm{HI}/4$, where n_HI is the number density of neutral hydrogen (FURLANETTO et al., 2006). Therefore, the resulting expression is given by

$$\tau_\nu = \int \sigma_{10} (\nu)\, \frac{n_\mathrm{HI}}{4} \left(1 - e^{-E_{10}/k_\mathrm{B}T_\mathrm{S}}\right) \mathrm{d}s. \tag{2.25}$$

where $E_{10} = 5.879 \times 10^{-6}$ eV is the energy associated with the transition between levels of the hyperfine structure; $n_{\rm HI}$ is the number density of neutral hydrogen; $\sigma_{10}(\nu)$ is the 21 cm cross-section, given by

$$\sigma_{10}(\nu) = \frac{3c^2 A_{10}}{8\pi\nu^2}\phi(\nu), \tag{2.26}$$

where $\phi(\nu)$ is the function that represents the normalized profile of the 21 cm line, whose peak occurs at ~ 1420 MHz and $\int \phi(\nu)\,{\rm d}\nu = 1$; A_{10} is the spontaneous emission coefficient, which gives the probability of occurrence of the hyperfine transition per unit time, equal to $2.85 \times 10^{-15}{\rm s}^{-1}$. It corresponds to a spontaneous decay time of about 10^7 years.

Assuming uniformity throughout the gas cloud, the optical depth can be approximated by

$$\tau_\nu \approx \left(\frac{3c^2 A_{10}}{8\pi\nu^2}\right)\left(\frac{h\nu}{k_{\rm B}T_{\rm S}}\right)\left(\frac{N_{\rm HI}}{4}\right)\phi(\nu). \tag{2.27}$$

In the context of 21 cm cosmology, the line profile $\phi(\nu)$ has to be written taking into account that the gas in the intergalactic medium (IGM) expands uniformly with the Hubble flow. The Sobolev approximation assumes that local variations in the velocity gradient are negligible compared to macroscopic ones, within a region of linear dimension s (SOBOLEV, 1957). According to this approximation, the velocity profile can be written as

$$v = \left(\frac{{\rm d}v}{{\rm d}s}\right)s \approx H(z)\,s, \tag{2.28}$$

where $H(z)$ is the Hubble parameter.

Using Equation 2.28 and Doppler law expression, the line profile equation can be approximated by

$$\phi(\nu) \approx \frac{c}{sH(z)\,\nu}. \tag{2.29}$$

Considering the column density of a segment with length s as a function of the

fraction $x_{\rm HI}$ of neutral hydrogen, so that $N_{\rm HI} = x_{\rm HI} n_{\rm H}(z)\, s$, a final expression for the 21 cm optical depth of the diffuse IGM can be written as (FURLANETTO et al., 2006)

$$\tau_{\nu_{10}} = \frac{3}{32\pi}\left(\frac{A_{10}}{\nu_{10}^2}\right)\left(\frac{hc^3}{k_{\rm B}T_{\rm S}}\right)\frac{1}{(1+z)}\frac{x_{\rm HI} n_{\rm H}}{{\rm d}v_{\parallel}/{\rm d}r_{\parallel}}. \tag{2.30}$$

As the optical depth of this hyperfine transition is small at the observed frequencies (reshifts) ($\tau_\nu \ll 1$), the Equation 2.23 can be approximated by

$$T_{\rm b}(\nu) - T_{\rm R}(\nu) \approx [T_{\rm S} - T_{\rm R}(\nu)]\,\tau_\nu, \tag{2.31}$$

where the absence of a signal or the occurrence of emission or absorption depends on $T_{\rm S}$ and $T_{\rm R}$ values. The differential brightness temperature, measured at a frequency ν, is defined as $\delta T_b(\nu) = T_{\rm b}(\nu) - T_{\rm R}(\nu)$. Equation 2.31 also shows that the brightness temperature of the 21 cm signal, observed at different epochs of the universe, depends directly on the evolution of the spin temperature.

A final expression for the 21 cm signal differential brightness temperature $\delta T_{\rm b}$ is obtained from Equations 2.30 and 2.31, which results in

$$\delta T_{\rm b}(z) \approx 9 x_{\rm HI}(1+\delta_{\rm b})(1+z)^{1/2}\left[1 - \frac{T_{\rm R}(z)}{T_{\rm S}}\right]\left[\frac{H(z)/(1+z)}{{\rm d}v_{\parallel}/{\rm d}r_{\parallel}}\right]{\rm mK}, \tag{2.32}$$

where $x_{\rm HI}$ is the numerical fraction of neutral hydrogen; $\delta_{\rm b}$ is the density contrast of baryons, given by the relative deviation of density from the mean density, which indicates that $\delta T_{\rm b}$ describes indirectly the underlying matter distribution; ${\rm d}v_{\parallel}/{\rm d}r_{\parallel}$ is the gradient of the proper velocity along the line of sight, considering the Hubble expansion and the peculiar velocity.

In the Equation 2.32, one should note that $\delta T_{\rm b}$ saturates if $T_{\rm S} \gg T_{\rm R}$. On the other hand, $\delta T_{\rm b}$ can assume negative values if $T_{\rm S} \ll T_{\rm R}$. Thus, it can be concluded that the observability of the 21 cm signal is directly related to the spin temperature, defined as a function of the ratio between the numerical densities of the populations of the excited state n_1 and the ground state n_0. Rewriting Equation 2.24, one obtains

$$T_{\rm S} = T_* \left[\ln \left(\frac{3n_0}{n_1} \right) \right]^{-1}, \tag{2.33}$$

that is, the greater the number of excited atoms in a gas cloud, the higher the spin temperature. The $T_{\rm S}$ being greater, lesser or equal to the temperature of the radiation field $T_{\rm R}$ is what dictates whether the signal will be observed in emission or absorption, or not.

Three processes define the populations n_0 and n_1, and consequently $T_{\rm S}$, along cosmic evolution: absorption/emission of 21 cm photons from/to the background radiation field (e.g. CMB); collisions of HI atoms with other hydrogen atoms, free protons and free electrons; absorption followed by emission (resonant scattering) of Lyα photons, causing a spin-flip via an intermediate excited state, the so-called Wouthuysen-Field effect.

Let C_{01} and C_{10}, the collisional spin (de-)excitation rates per atom; P_{01} and P_{10}, the spin (de-)excitation rates per atom from Lyα resonant scattering. As the rate of occurrence of these processes is much higher than the spontaneous decay rate of 21 cm transition ($\sim 10^{15}\ s$), the spin temperature is determined by the equilibrium between them as

$$n_1 \left(C_{10} + P_{10} + A_{10} + B_{10} I_{\rm R} \right) = n_0 \left(C_{01} + P_{01} + B_{01} I_{\rm R} \right), \tag{2.34}$$

where B_{10} and B_{01} are the Einstein coefficients and $I_{\rm R}$ is the energy flux of the radiation field photons.

The relation between the rates of excitation C_{01} and de-excitation C_{10} per collision is given by

$$\frac{C_{01}}{C_{10}} = \frac{g_1}{g_0} e^{-T_*/T_{\rm K}} \approx 3 \left(1 - \frac{T_*}{T_{\rm K}} \right), \tag{2.35}$$

where $T_{\rm K}$ is the kinetic temperature of the gas. The effective color temperature $T_{\rm c}$ of the radiation field is defined by

$$\frac{P_{01}}{P_{10}} = 3 \left(1 - \frac{T_*}{T_{\rm c}} \right). \tag{2.36}$$

Substituting the Equation 2.24 and the Einstein relations (equations that relate A_{10}, B_{10} and B_{01}) in the Equation 2.34. Then, considering that T_* is much lower than any astrophysical temperature T, and making the approximation $1 - e^{T_*/T} \approx 1 - T_*/T$, as well as that the photons of the radiation field are in the Rayleigh-Jeans region of the emission spectrum, so that $I_\nu \propto \nu^2 T_\mathrm{b}$, it is obtained

$$T_\mathrm{S}^{-1} = \frac{T_\mathrm{R}^{-1} + x_\alpha T_\alpha^{-1} + x_\mathrm{c} T_\mathrm{K}^{-1}}{1 + x_\alpha + x_\mathrm{c}}, \tag{2.37}$$

where T_R is the temperature of the radiation field, in general $T_\mathrm{R} = T_\mathrm{CMB}$. The temperature of the Lyα photon radiation field at the frequency of the line (2.45×10^{15} Hz) is T_α. It is tightly coupled to the kinetic (thermodynamic) temperature T_K, due to the recoil caused by repeated scattering between photons and atoms. In summary, T_S is expressed as a weighted sum over the temperatures which characterize the efficiency of each of the three processes described above. The strength of these process are given by the coupling coefficients x_α and x_c.

The coupling coefficient x_α is associated with the resonant scattering of Lyα photons, known as the Wouthuysen-Field effect. This process came to dominate the coupling with the formation of the first stars. In this epoch, the hydrogen that is initially in the least energetic hyperfine state $1_0S_{1/2}$, absorbs a Lyα photon and is excited to one of the central 2P hyperfine states. This hydrogen atom can then emit a Lyα photon and return to one of the two hyperfine levels of the ground state. If the atom goes to the more energetic hyperfine level, a spin-flip and consequent 21 cm emission can occur.

The Lyα coupling expression is given by

$$x_\alpha = \frac{16\pi^2 T_* q_\mathrm{e}^2 f_\alpha}{27 A_{10} T_\mathrm{R} m_\mathrm{e} c} S_\alpha J_\alpha, \tag{2.38}$$

where J_α is the Lyα specific flux; S_α is the order-one correction factor for scattering effects in the vicinity of the Lyα line center; m_e and q_e are the mass and charge of the electron; f_α is the oscillator strength of the Lyα transition; $A_{10} = 2.85 \times 10^{-15}$ s^{-1} is the Einstein coefficient of spontaneous emission rate.

The collisional coupling coefficient x_c takes into account the (de-)excitation of the HI hyperfine levels caused by collisions between hydrogen atoms and collisions between a hydrogen atom and a free proton or a free electron. These collisions dominate

coupling in a period of the universe when the density of neutral hydrogen was high (Dark Ages). The coupling coefficient can be written as

$$\begin{aligned} x_{\mathrm{c}} &= x_{\mathrm{c}}^{\mathrm{HH}} + x_{\mathrm{c}}^{\mathrm{pH}} + x_{\mathrm{c}}^{\mathrm{eH}} \\ &= \frac{T_*}{A_{10} T_{\mathrm{R}}} \left[\kappa_{10}^{\mathrm{HH}} \left(T_{\mathrm{K}} \right) n_{\mathrm{H}} + \kappa_{10}^{\mathrm{pH}} \left(T_{\mathrm{K}} \right) n_{\mathrm{p}} + \kappa_{10}^{\mathrm{eH}} \left(T_{\mathrm{K}} \right) n_{\mathrm{e}} \right], \end{aligned} \tag{2.39}$$

where $\kappa_{10}^{\mathrm{HH}}$ is the rate coefficient for spin de-excitation in collisions between hydrogen atoms, $\kappa_{10}^{\mathrm{pH}}$ is the rate coefficient for spin de-excitation in collisions between protons and hydrogen atoms and $\kappa_{10}^{\mathrm{eH}}$ is the rate coefficient for spin de-excitation in collisions between electrons and hydrogen atoms. These coefficients are determined in the scope of Quantum Mechanics.

There are two types of background radio photons sources capable of exciting the neutral hydrogen atoms of the IGM and producing the emission at 21 cm: the CMB and the strong radio point sources (e.g. quasars). After the end of the recombination, the temperature fluctuations (thermodynamics) of the CMB were $\delta T_{\mathrm{CMB}} \sim 10^{-5}$, where $T_{\mathrm{R}} = T_{\mathrm{CMB}}$. In this case, the 21 cm line is observed as a spectral distortion in the blackbody curve of the CMB, in an appropriate frequency range. The interaction between the CMB photons and the HI cloud atoms produces a signal distribution of 21 cm across the entire sky, which allows it to be treated as the anisotropies of the cosmic microwave background. The temperature fluctuations $\delta T_{21\mathrm{cm}}$ can be mapped as a function of celestial coordinates at different redshifts (frequencies). The result is a 3D map of HI distribution.

Radio point sources are objects with intense brightness, such as quasars, whose brightness temperature is much higher than that of the diffuse and weak emission of HI ($T_{\mathrm{R}} \gg T_{\mathrm{S}}$), what causes the IGM gas cloud absorption relative to the background radio source. The occurrence of this 21 cm absorption at different redshifts creates the so-called 21 cm forest. The high intensity of the background radio point sources allows to study the lines with high resolution in frequency, making it possible to solve small-scale structures of the IGM (PRITCHARD; LOEB, 2012).

Summarizing what has been presented so far, the brightness temperature of the 21 cm signal can be written as a function of four variables $T_{\mathrm{b}} = T_{\mathrm{b}} \left(T_{\mathrm{K}}, x_{\mathrm{i}}, J_{\alpha}, n_{\mathrm{H}} \right)$ where T_{K} is the kinetic temperature of the gas; x_{i} is the volume average of the

ionized fraction of hydrogen; J_α is the specific flux at the frequency of Lyα and $n_{\rm H}$ is the number density of hydrogen. The dependence of $T_{\rm b}$ on each of the presented parameters saturates at different times of cosmic evolution. Therefore it is possible to separate the evolution of the 21 cm signal into different regimes, during which each of the variables dominates its correspondent brightness temperature fluctuations.

The following section describes the observational technique that maps the cumulative emission of a spectral line, such as the 21 cm emission presented previously, from galaxies, without resolving them individually, to survey the large-scale structure of the universe.

2.4 21 cm Intensity Mapping

Intensity mapping (IM) has proven to be an efficient technique for probing different epochs of the universe. The basic idea of IM is to measure the integrated emission of a spectral line from large areas of the sky, over a specific range of frequencies, and with low angular resolution. Thereby, the IM is able to detect the global line emission, as a function of redshift, from a large number of individually unresolved galaxies and from the diffuse IGM. Therefore, this three-dimensional mapping of spatial fluctuations in the density of baryonic matter allows revealing the formation and evolution of the large-scale structures (KOVETZ et al., 2017).

The IM method has the potential to explore a wide range of redshifts, covering from the cosmic dark ages, the period following recombination, to after the end of reionization epoch (PADMANABHAN, 2021), and for this to contribute significantly to astrophysics, cosmology and fundamental physics. More specifically, IM surveys can study the cosmic history of star formation; the formation and evolution of galaxies; the composition of the interstellar medium (ISM) and intergalactic medium (IGM); the physics of the epoch of reionization; cosmic inflation models; modified gravity theories; the origin of dark matter and the nature of dark energy (KOVETZ et al., 2019).

The galaxy surveys detect only discrete objects (e.g., individual galaxies) whose emission lies above some flux limit (KOVETZ et al., 2017). This makes them have a low sensitivity, being able to map only a small fraction of the observable universe (PADMANABHAN, 2021). The IM surveys, on the other hand, are sensitive to all sources in a given line-of-sight, which produces large three-dimensional maps of large-scale structures. The low angular resolution of IM experiments makes them less expensive than galaxy surveys, in addition to making it capable of covering

large volumes of the universe in a shorter observation time (KOVETZ et al., 2019).

The IM method can target a variety of atomic and molecular lines. Initially, most efforts were concentrated on probing the epoch of reionization by mapping the 21 cm emission of neutral hydrogen (HI) from the IGM. Then, the IM technique began to be used to observe the universe at low redshifts by measuring the 21 cm emission from galaxies. Finally, IM was also used with other spectral lines, as the rotational carbon-monoxide (CO) transitions, the singly-ionized carbon [CII] fine-structure, the hydrogen Ly-α, H-α and H-β lines; doubly ionized oxygen [OIII] and others, to detail galaxy evolution and star formation (KOVETZ et al., 2017).

IM surveys can contribute to cosmology by detailing the growth of structures through measurements of large-scale redshift-space distortions; constraining the dark energy equation of state and cosmological parameters through BAO measurements; searching for deviations from the ΛCDM model (OXHOLM, 2022).

IM measures spatial fluctuations in the emission intensity of a spectral line at a redshift z, $\delta_{\rm I}(\vec{r}, z)$, which traces the large-scale matter fluctuations as (OXHOLM, 2022)

$$\delta_{\rm I}(\vec{r}, z) = \bar{I}(z) b(z) \delta_{\rm m}(\vec{r}, z) + N(z), \tag{2.40}$$

where $\vec{r}$ is the vector of spatial coordinates; $\bar{I}(z)$ is the average line intensity; $N(z)$ is the map noise, which is assumed to be Gaussian and random at a given redshift (or frequency); $b(z)$ is the redshift-dependent linear clustering bias of the observed galaxies relative to the underlying matter overdensity, $\delta_{\rm m}(\vec{r}, z)$, which is given by

$$\delta_{\rm m}(\vec{r}, z) = \frac{\rho_{\rm m}(\vec{r}, z) - \bar{\rho}_{\rm m}(z)}{\bar{\rho}_{\rm m}(z)}, \tag{2.41}$$

where $\rho_{\rm m}(\vec{r}, z)$ is the matter field density; $\bar{\rho}_{\rm m}(z)$ is the average matter field density over the observed region of the sky.

The intensity field $\delta_{\rm I}$, presented in the Equation 2.40, can be represented by its Fourier transform

$$\tilde{\delta}_{\rm I}(\vec{k}, z) = \frac{1}{(2\pi)^3} \int \delta_{\rm I}(\vec{r}, z) e^{-i\vec{k}\cdot\vec{r}} {\rm d}^3\vec{r}, \tag{2.42}$$

where $\vec{k}$ is the wavenumber (or mode) of the spatial fluctuations described by the coordinates $\vec{r}$.

The intensity power spectrum, $P_{\rm I}(\vec{k}, z)$, describes statistically the Fourier transform of the intensity field, given by the Equation 2.40, $\tilde{\delta}_{\rm I}$, and is written as

$$\left\langle \tilde{\delta}_{\rm I}(\vec{k}, z)\tilde{\delta}_{\rm I}(\vec{k'}, z)\right\rangle = (2\pi)^3 P_{\rm I}(\vec{k}, z)\delta_{\rm D}(\vec{k} - \vec{k'}), \tag{2.43}$$

where $\langle\rangle$ denotes the expected value and $\delta_{\rm D}$ is the three-dimensional Dirac delta function. As the galaxies emitting a spectral line are a discrete and biased tracer of the underlying matter fluctuations, the IM power spectrum consists of two main components, as

$$P_{\rm I}(\vec{k}, z) = \bar{I}^2(z)b^2(z)P_{\rm m}(\vec{k}, z) + P_{\rm shot}(z), \tag{2.44}$$

where $\bar{I}^2(z)b^2(z)P_{\rm m}(\vec{k}, z)$ represents the large-scale clustering and continuous sources contributions, which follows the matter power spectrum $P_{\rm m}(\vec{k}, z)$; $P_{\rm shot}(z)$ is the shot noise power spectrum, which refers to non-clustered line-emitting galaxies, that follow discrete statistics well represented by a Poisson distribution.

The average intensity $\bar{I}(z)$ is proportional to the first moment of the luminosity function (or distribution) of the target emission line $\Phi\left(L, z\right)$, which gives the number density of galaxies with luminosities between L and $L + {\rm d}L$, as

$$\bar{I}(z) \propto \int_0^\infty L\Phi\left(L, z\right){\rm d}L. \tag{2.45}$$

The shot noise power spectrum is scale-independent and has amplitude proportional to the second moment of the luminosity function, as

$$P_{\rm shot}\left(z\right) \propto \int_0^\infty L^2\Phi\left(L, z\right){\rm d}L. \tag{2.46}$$

The 21 cm IM surveys can measure the HI emission from galaxies and the IGM to reveal the three-dimensional picture of large-scale structure in the late universe. This tomographic map of the distribution of matter in large cosmic volumes can be used to detect BAO in radio wavelengths. A 21 cm IM telescope can be used

to look at this standard ruler in the radial and transverse directions across a range of redshifts, to produce estimates of the Hubble parameter and angular diameter distance, respectively. Finally, these measurements can be used to trace the history of universe expansion, and provide very precise constraints on dark energy equation of state (see Section 2.2) (PETERSON et al., 2009).

During the epoch of reionization, the amount of neutral hydrogen in the universe was drastically reduced. It is estimated that the remaining amount of HI from this period corresponds to about 1% of the existing baryons. Furthermore, the preservation of this residual quantity of these atoms at low redshifts was only possible due to the formation of very dense and optically thick regions, which shielded the HI against incident radiation. The observed post-reionization 21 cm signal reveals the density field of matter behind the clustering of collapsed halos (PRITCHARD; LOEB, 2012). However, the 21 cm signal is weak at these redshifts, so that the amplitude of the Galactic synchrotron emission, the dominant foreground at radio frequencies, is a few orders of magnitude greater than the amplitude of the 21 cm emission. This makes the task of disentangling the HI signal from the astrophysical foregrounds the major challenge of the 21 cm IM approach.

BINGO is a radio telescope that will use the 21 cm IM to measure the HI distribution over an area of about 13% of the sky, in a range of redshifts $0.127 \leq z \leq 0.449$, with the aim of detecting the scale from BAO and hence to constrain dark energy parameters. A brief description of this instrument is presented in Chapter 3. The foreground removal problem, as well as the component separation method used in this thesis to recover the HI signal from the BINGO simulated data, are presented in Chapter 5.

3 BINGO TELESCOPE

BINGO is a radio telescope designed to measure the 21 cm emission of HI in the frequency band $980 \leq \nu \leq 1260$ MHz, using the IM technique described in Chapter 2. Additionally, given the properties of the telescope, it will be able to describe phenomena at very short time scales, making it a useful tool for studying pulsars and fast radio bursts (FRBs) (ABDALLA et al., 2022a). The instrument will observe an $\sim 15°$-wide declination strip, centered at $\delta = -15°$, which corresponds to 13% of the sky. The angular resolution of the instrument is ~ 40 arcmin, which is suitable for resolving BAO angular scales in the range of observed redshifts $0.127 \leq z \leq 0.449$, where dark energy becomes dominant (WUENSCHE et al., 2022).

The telescope is being built in Aguiar, Paraíba, Brazil, at a site with very low radio frequency interference (RFI), located at coordinates $7° \, 2' \, 27.6'' \, \mathrm{S}$; $38° \, 16' \, 4.8'' \, \mathrm{W}$. The site selection process and description of the RFI measures are presented in Peel et al. (2019).

In the BINGO frequency band, the mean brightness temperature of the 21 cm emission is ~ 200 μK. Assuming the estimated system temperature of ~ 70 K, a channel bandwidth of $\Delta\nu = 9.33$ MHz (30 frequency bins), a pixel size of ~ 27.5 arcmin (HEALpix $N_{\mathrm{side}} = 128$) and an observation time of 1 year at 60% duty cycle, BINGO should achieve an estimated sensitivity of 102 μK (WUENSCHE et al., 2022).

BINGO has a compact optical configuration with no moving parts, consisting of two reflectors, the secondary one illuminating 28 horn-type antennas (planned for phase 1), which will collect, each one, two hands of circularly polarized signal, each of which will be fed into magic tees connected to stable correlation receivers with dual polarization. More specifically, each magic tee will send the information to two amplifier chains and, after some electronic processing, such as filtering and phase switching, the signal will be recombined in a second hybrid and then sent to a digital backend for frequency decomposition (WUENSCHE et al., 2022). Figure 3.1 shows an artistic view of the telescope, displaying a west-east view (left-right) of the site, with the sheltering hill in the background (ABDALLA et al., 2022a).

Figure 3.1 - The artistic view of the BINGO telescope site. This figure displays the sheltering hill and the use of the natural inclination of the terrain to place the primary and secondary mirrors, in addition to the horn structure.

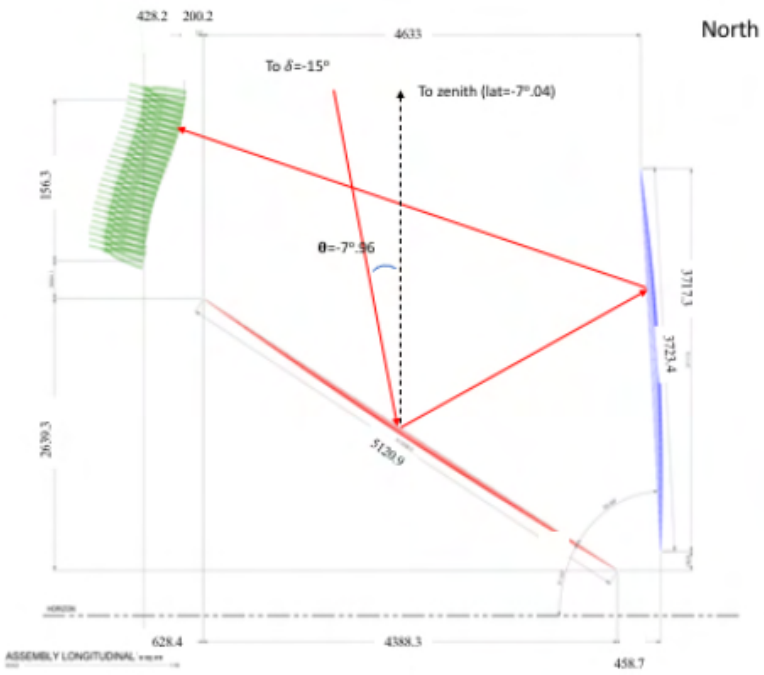
Source: Abdalla et al. (2022a).

The following sections describe the main subsystems of the instrument.

3.1 Optics

The optical system configuration is an off-axis crossed-Dragone (DRAGONE, 1978) type, a compact arrangement that allows a wide field of view. The system consists of two reflectors with no moving parts and has a effective focal length of 63.2 m. The primary is a paraboloid with a semi-major axis of 25.5 m and the secondary is a hyperboloid with a semi-major axis of 18.3 m. The primary mirror has an inclination with respect to the ground and points to a declination $\delta = -15°$. The secondary is supported by a structure that maintains it with a slight inclination in relation to the vertical. The schematic is shown in Figure 3.2.

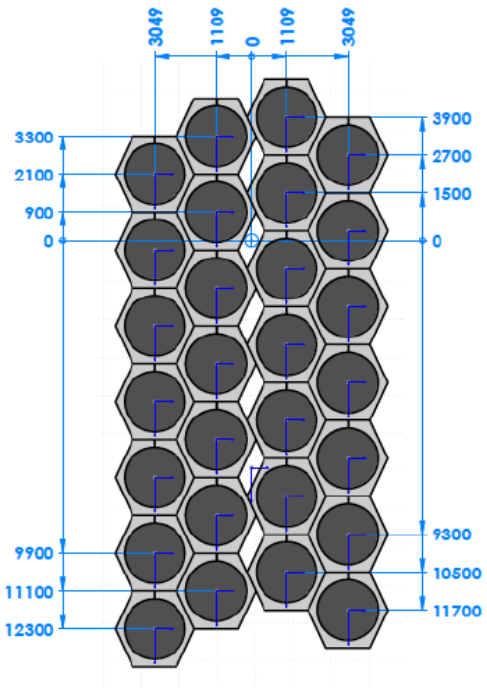
Figure 3.2 - BINGO optics schematics: the primary reflector (red) is facing north pointing to $\delta = -15°$; the secondary reflector (blue) is slightly inclined with respect to the vertical and pointing towards the horn arrangement (green). The dimensions are in millimetres.



Source: Wuensche et al. (2022).

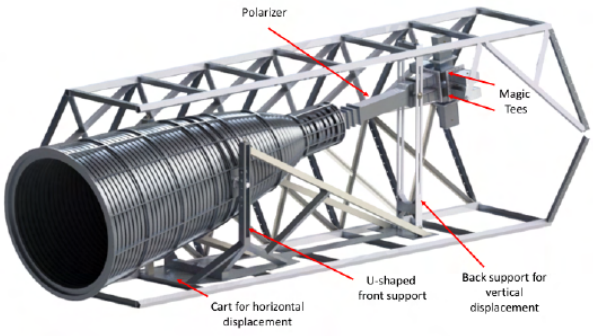
To reduce spillover, spurious signals picked up near the edges of the reflector, as well as far out sidelobes, the secondary is under-illuminated by an array of horn-like antennas. The BINGO Phase 1 predicts a focal plane arrangement with $N_{\text{horns}} = 28$ as shown in Figure 3.3. Each of these horns is supported by the hexagonal structure shown in Figure 3.4. The BINGO optical design ensures that the forward gain of each beam, resulting from the combination between each feed horn and the secondary reflector, does not vary by more than 1 dB for all pixels across the focal plane. In addition, the simulations results show low optical aberration and ellipticity for all beams (WUENSCHE et al., 2021). The full width at half maximum (FWHM) of the telescope's resulting beam should be $\approx 0.67°$ at the central frequency of the BINGO band, the original angular resolution of the project (ABDALLA et al., 2022b).

Figure 3.3 - Horn arrangement in the focal plane for BINGO Phase 1. In this way, each horn covers a different declination strip. Hexagonal structures allow movement and repositioning of the horns during the mission. The dimensions are in millimetres.



Source: Wuensche et al. (2022).

Figure 3.4 - Hexagonal cage horn. The horn is attached to a U-shaped support by the ring crossing its center of mass, for pivoting and vertical displacement, and to another structure connected to the end of the polarizer, on the opposite side of the horn mouth. The U-shaped support and the back structure are mounted on top of a cart that allows for longitudinal positioning of the horn.



Source: Wuensche et al. (2022).

This optical arrangement makes BINGO a transit instrument, covering an instantaneous $\sim 15°$-wide declination strip centered at $\delta = -15°$. This results in a field of view that covers a 15.4° wide declination band. The horns are arranged in the focal plane so that each one points to a different declination strip and there is uniformity in the sky coverage. Furthermore, the design foresees that the longitudinal axis of each horn can be shifted vertically, allowing an increase in declination coverage of $\approx 0.56°$ after 5 years of observation (one horn repositioning per year) and eliminating the gaps in the sky coverage. This repositioning of the feed horns results in a more homogeneous distribution of the noise level in the pixels of the covered area, as discussed in Section 4.1.4.

The component separation techniques used in the 21 cm IM surveys rely on the spectral smoothness of the astrophysical foregrounds, as described in the Chapter 5. However, the frequency dependence of the instrument beam (chromaticity), such as the frequency evolution of the beam size, can complicate the foreground emission. More specifically, the chromatic beam side lobes can couple the large angular fluctuations in the diffuse Galactic emission into spectral structures, distorting the otherwise smooth spectral shapes of the intrinsic foreground, what can confuse global 21 cm signatures. Although the amplitude of the side lobes is typically much lower than the major lobe, they can cover a large solid angle, and so contribute a non-negligible fraction of the total detected emission after integrating over the entire beam profile. Thus, the chromatic structure of the beam must be properly modeled through previous measurements, and corrected by calibration, to enable the separation of the faint cosmological component from bright foreground emission. In short, understanding how damaging these frequency effects to the intrinsic spectral characteristics of the foregrounds is indispensable for the 21 cm signal recovery (ALONSO et al., 2015; ASAD et al., 2021; CUNNINGTON et al., 2021; IRFAN; BULL, 2021; MATSHAWULE et al., 2021; MOZDZEN et al., 2016; TAUSCHER et al., 2020).

Due to very clean beam obtained with optical design simulations (ABDALLA et al., 2022b), and measurements of the beam profile of a prototype horn (WUENSCHE et al., 2020), it was decided to use the conventional circular Gaussian beam approximation in the simulated observations of this work. However, this simplified model does not take into account the side lobe structure and its frequency-dependence.

Current simulations of the optical aberrations suggest the horns most distant from the center of the focal plane are slightly distorted, despite meeting the criteria of low illumination (- 50 dB attenuation) in the outer parts of the beam. These results,

as well as a complete description of the optical design of the BINGO telescope, are presented in Abdalla et al. (2022b).

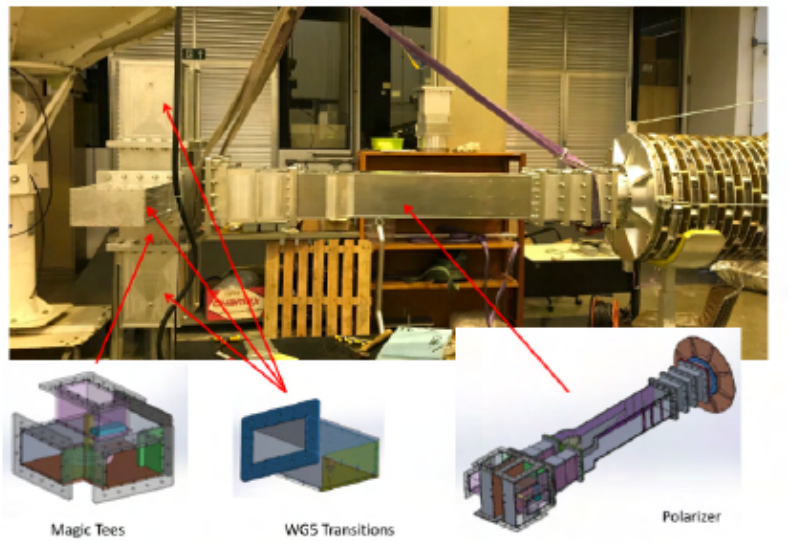
The component separation procedure proposed here must be tested in future works with a more realistic beam model, which include the above simulations, together with the actual beam profile measured for the prototype horn (as described in Wuensche et al. (2020)) coupled to the full telescope optics.

3.2 Feed horns and front end

The BINGO feed horns were designed in a corrugated conical shape, which attenuates spurious signals through successive reflections, and proves low side lobes across a broad frequency range and good polarisation performance. In order to produce the nominal field of view of the experiment, the horns dimensions are ~ 1.7 m in diameter and ~ 4.9 m in length. The horns were projected to optimally illuminate the secondary mirror and have excellent insertion and return loss measurements. Measurements made with a horn prototype show that the instrument beam will have secondary lobes $\lesssim -25$ dB and distortion from the circular shape $\lesssim -30$ dB inside the main beam. Such beam meets the necessary requirements to avoid the leakage of the linearly polarized portion of Galactic foreground emission into the HI signal, and to enable the separation of the bright astrophysical components from the faint 21 cm emission (ABDALLA et al., 2022b; WUENSCHE et al., 2019; WUENSCHE et al., 2021). The polarization leakage effects were not implemented in the simulations of this thesis, but should be tested in future work.

The front end is composed by transitions, polarizer and magic tees, as shown in the Figure 3.5. A circular to square transition was used to connect the horn throat to the polarizer. The polarizer was designed to take the two incoming hands of circular polarisation from the horn and convert them to linear with high isolation between them (WUENSCHE et al., 2020). Its outputs are a rectangular guides that connects to the hybrids (magic tee), each one responsible for feeding the two amplifier chains with the two linearly polarized signal, as described in the next section.

Figure 3.5 - Front end, with parts indicated at the bottom of the figure, connected to the feed horn prototype at INPE.



Source: Wuensche et al. (2022).

For a complete description of the design and construction of the feed horn and front end, see Wuensche et al. (2020).

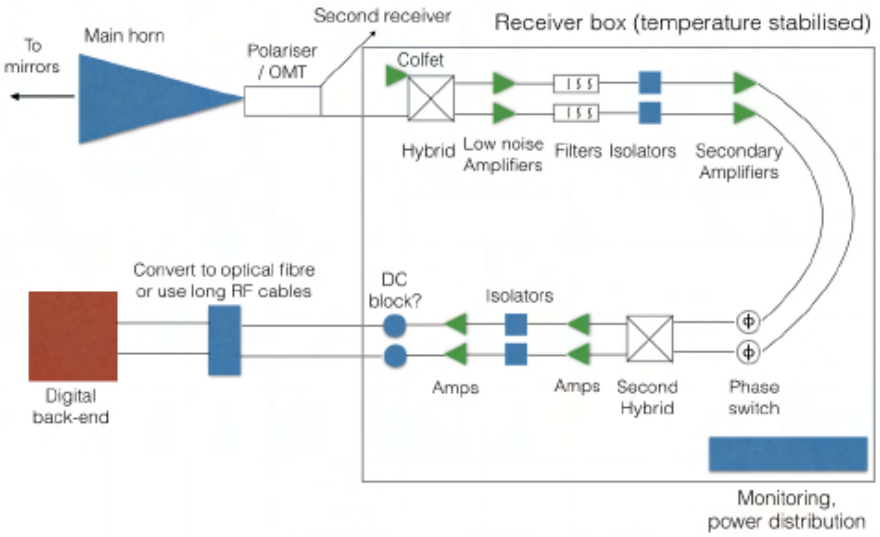
3.3 Receiver

The best choice for BINGO receivers is the full correlation configuration, operating at ambient temperature. With this receiver type, it is expected to minimize the gain variations and, consequently, the 1/f noise content in the TODs. Its estimated knee frequency, at which the thermal and $1/f$ noise power spectral densities (PSD) are equal, is $f_{\text{knee}} = 1$ mHz (WUENSCHE et al., 2022). The block diagram of the BINGO full correlation receiver is shown in the Figure 3.6.

As discussed in the Section 3.2, each magic tee will feeds two amplifier chains. Then, after filtering and phase switching, the signals will be recombined in a second hybrid and sent to a digital backend for frequency decomposition (WUENSCHE et al., 2022).

The receiver chain is based on room temperature low noise amplifiers. First stage filters define BINGO operational frequency band: 980MHz $\leq \nu \leq$ 1260MHz. This frequency window is subject to contamination from ADS-B airplane transponders (1090 MHz) and harmonics from geostationary satellites, which must be removed in

Figure 3.6 - Block diagram of the BINGO correlation receiver.



Source: Wuensche et al. (2021).

the data processing. The effects of mobile phone signal at 956 MHz can be avoided using filters with a very steep cutoff at the lower end of the band. Also, to prevent signal distortions caused by intense RFI emitted out of band, the first stage low noise amplifiers should have a large dynamic range.

The amplifiers gain fluctuations discussed above are correlated over some timescale for all BINGO receivers. This is reflected in the TODs as the $1/f$ noise component. Unlike the contribution of thermal origin, this correlated noise was not included in the simulations presented in Chapter 4, as it was considered that it can be removed using appropriate map making techniques, which is expected to be the scope of the next steps of this work.

The following chapter presents the models used to generate the simulated BINGO observations used in the component separation tests.

4 SIMULATED DATA

The following are the models of the 21 cm signal, astrophysical foregrounds and thermal noise used to produce the simulated observations in this work, as well as a description of how the galactic mask used in the maps was constructed. The simulated maps were produced with the `HEALpix` package (GORSKI et al., 2005) using a $N_{\rm side} = 256$, which corresponds to a pixelization resolution of 13.7 arcmin. At the end of this chapter, the simulation plan containing the data set configurations used in the component separation tests is detailed.

4.1 Simulated sky

4.1.1 21 cm signal

As discussed in the Chapter 2, in the range of low redshifts observed by BINGO ($z \lesssim 0.5$), the 21 cm emission is assumed to be produced by the HI confined to discrete elements such as galaxies, bubbles and filaments. In this case, the HI brightness temperature can be written as the sum of two components, as

$$T_{\rm HI} = \bar{T}_{\rm HI} \left(1 + \delta_{\rm HI}\right), \tag{4.1}$$

where $\delta_{\rm HI} = \delta T_{\rm HI}/\bar{T}_{\rm HI}$ is the density contrast and $\bar{T}_{\rm HI}$ is the mean HI brightness temperature, given by Hall et al. (2013)

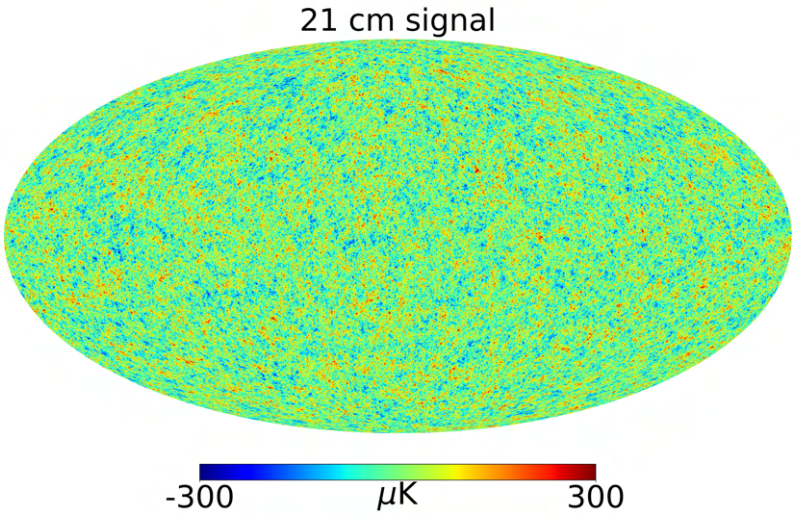
$$\bar{T}_{\rm HI} = 188\ h\ \Omega_{\rm HI} \frac{(1+z)^2}{E\left(z\right)} {\rm mK}\,, \tag{4.2}$$

where h is the dimensionless Hubble constant, $\Omega_{\rm HI}$ is the density parameter for HI and $E(z) = H(z)/H_0$, defined by the Hubble parameter $H(z)$ and by the Hubble constant $H_0 = 100h\ {\rm km\,s^{-1}}$ Mpc. This is the HI mean brightness temperature value assumed for the simulated data used in this work. The fluctuations around $\bar{T}_{\rm b}$, given by the Equation 2.32, are due to differences in densities of structure and they can be calculated assuming that the HI signal is a tracer of the dark matter.

The simulated 21 cm emission maps were generated using the Full-Sky Log-normal Astro-Fields simulation Kit (FLASK) (XAVIER et al., 2016), a package that allows generating full-sky maps of cosmological large-scale structures (LSS) observables, such as matter density tracers (galaxies, quasars, dark matter halos), CMB temper-

ature anisotropies and weak lensing convergence and shear fields (LICCARDO et al., 2022). The set of maps generated with FLASK provides a kind of tomography of the observable, divided into the number of redshift slices (bins) chosen by the user. The statistical properties of the observable are defined by two FLASK inputs: the theoretical angular temperature auto- and cross-power spectra calculated for each redshift bin and the multivariate distribution (log-normal). The power spectra were generated using the Unified Cosmological Library for C_ℓs (UCLCL) (MCLEOD et al., 2017; LOUREIRO et al., 2019), a library for computing two-point angular correlation function of various cosmological fields that are related to large scale structures surveys. The parameters of the log-normal were chosen so that the distribution was approximately a Gaussian. The mean temperature of the HI signal fluctuations in the BINGO redshift range is ≈ 200 μK. Figure 4.1 shows a 21 cm full-sky emission map generated with FLASK, referring to the frequency channel centered at 1115 MHz (near to the BINGO band central frequency) and with a bandwidth of 9.33 MHz.

Figure 4.1 - Full-sky maps of the 21 cm signal for the frequency channel centered at 1115 MHz and with a bandwidth of $\delta\nu \sim 9.33$ MHz. Temperatures are given in μK.



4.1.2 Astrophysical foregrounds

The detailed knowledge of the astrophysical foregrounds that contaminate the cosmological signal measurements is fundamental for the recovery of the 21 cm signal. The sky simulations used in this work include contributions from galactic and extragalactic sources. The galactic foregrounds is due to synchrotron radiation, free-free emission, the anomalous microwave emission (AME) and thermal dust (TD). The extragalactic emissions include unresolved faint radio point sources (FRPS) and the thermal and kinetic Sunyaev Zel'dovich (SZ) effects. Bright radio point sources were not included in the BINGO simulated sky since they can be masked out during the analysis, hence reducing their contribution to the sky.

The cosmic microwave background (CMB) signal is very well characterized in the BINGO frequencies, being responsible for a 2.7 K constant background level, with fluctuations of $\approx 100\mu$K (LICCARDO et al., 2022). The CMB fluctuations can be directly removed from the data using a spatial template from PLANCK COLLABORATION (2020b). For this reason, the CMB component was also not included in the simulated data.

Thus, the brightness temperature of the simulated BINGO sky, defined at each frequency ν and at each pixel (or sky direction) p, is given by

$$\begin{aligned} T_{\rm sky}\left(\nu,p\right) = T_{\rm synch}\left(\nu,p\right) + T_{\rm ff}\left(\nu,p\right) + T_{\rm AME}\left(\nu,p\right) + \\ + T_{\rm TD}\left(\nu,p\right) + T_{\rm FRPS}\left(\nu,p\right) + T_{\rm SZ}\left(\nu,p\right) + T_{\rm HI}\left(\nu,p\right), \end{aligned} \tag{4.3}$$

where the brightness temperatures are: $T_{\rm synch}$, from synchrotron radiation; $T_{\rm ff}$, from the free-free emission; $T_{\rm AME}$, from anomalous microwave emission; $T_{\rm TD}$, from the thermal dust emission; $T_{\rm FRPS}$, from faint radio pont sources; $T_{\rm SZ}$, from thermal and kinetic Sunyaev Zel'dovich (SZ) effects.

The foreground models used to produce the BINGO simulated sky are presented in the following sections. All the astrophysical foreground simulations were generated using the Planck Sky Model (PSM) software (DELABROUILLE et al., 2013).

Galactic synchrotron emission

The synchrotron radiation is the electromagnetic radiation emitted when a relativistic charged particle is accelerated by a magnetic field in a direction perpendicular to its velocity. The Galactic synchrotron emission is predominantly of diffuse origin

and highly polarized, being produced by cosmic rays, mostly high energy electrons ejected from supernova, spiralling around Milky Way magnetic field (ANDERSEN et al., 2020). The intensity of this component is spatially variable, depending on the number density and energy distribution of cosmic ray electrons, and the strength of the Galactic magnetic field, which vary across the sky. Furthermore, this is the strongest foreground for 21 cm experiments, dominating the radio sky at frequencies below ~ 10 GHz (PADOVANI et al., 2021).

For energies $E \gtrsim 10$ GeV, the energy distribution of cosmic ray electrons is well-approximated by a power-law, as (NAVARRO, 2015)

$$N(E)\mathrm{d}E \propto E^{-\alpha}\mathrm{d}E, \tag{4.4}$$

where $N(E)\mathrm{d}E$ is the number of electrons per unit volume with energies between E and $E + \mathrm{d}E$; α is the energy-dependent electron power-law index. This distribution is relatively smooth, both spectrally and spatially over the sky, and the measured values of α are close to 3 at frequencies above 408 MHz (PADOVANI et al., 2021).

Given the distribution of electron energy presented above, the intensity of the synchrotron radiation at a frequency ν can be written as

$$I_{\mathrm{synch}}(\nu) = \varepsilon_{\mathrm{synch}}(\nu) \int_z n_{\mathrm{e}} B_{\perp}^{(1+\alpha)/2} \mathrm{d}z, \tag{4.5}$$

where n_{e} is the electron number density; z is the line-of-sight along which the integral is calculated; $B_{\perp}$ is the component of the magnetic field perpendicular to the line-of-sight; $\varepsilon_{\mathrm{synch}}(\nu)$ is the emissivity term, which can be written as the power law

$$\varepsilon_{\mathrm{synch}}(\nu) = \varepsilon_{\mathrm{synch}}(\nu_0)\nu^{-(\alpha-1)/2}. \tag{4.6}$$

The brightness temperature of the synchrotron emission can be written in terms of its intensity through the Rayleigh-Jeans approximation,

$$T_{\mathrm{synch}}(\nu) = \frac{c^2 I_{\mathrm{synch}}(\nu)}{2k_{\mathrm{B}}\nu^2}. \tag{4.7}$$

Therefore, the synchrotron brightness temperature spectrum, at a particular line-

of-sight, can be written as

$$T_{\rm synch}(\nu) = T_{\rm synch}(\nu_0)\left[\frac{\nu}{\nu_0}\right]^{\beta_{\rm s}}, \tag{4.8}$$

where $T_{\rm synch}(\nu_0)$ is brightness temperature of synchrotron emission, at a particular line-of-sight, defined in the reference frequency ν_0, and β_s is the synchrotron spectral index, given by

$$\beta_{\rm s} = -\frac{\alpha+3}{2}. \tag{4.9}$$

There is evidence that $\beta_{\rm s}$ is non-uniform over the sky and that it varies with frequency, as for example in Lawson et al. (1987), Reich and Reich (1988), and Banday and Wolfendale (1991). This occurs due to the aging of cosmic rays, that is, the older the electron, the more energy it lost due to synchrotron radiation. Since this lost energy is proportional to E^2, the synchrotron spectrum becomes steeper over time. Furthermore, how most cosmic rays are produced in supernova explosions, which occurs mainly close to Galactic plane, and are diffused away to higher Galactic latitudes losing some of their energy, $\beta_{\rm s}$ is expected to be more negative at higher Galactic latitudes (NAVARRO, 2015). The curvature of the synchrotron spectrum is modeled in (KOGUT, 2012). However, detailed models and observations both suggest that the synchrotron spectrum may be closely approximated by a constant power-law at different frequency bands (ANDERSEN et al., 2020).

To simulate the synchrotron component of this work, it was used as template the reprocessed 408 MHz all-sky map of Haslam et al. (1982), presented by Remazeilles et al. (2015). This template map is extrapolated to the BINGO frequencies (980-1260 MHz) through the Equation 4.8. Among the components selected to produce the simulated sky used here, synchrotron radiation is the dominant one in the BINGO frequency band. Thus, to evaluate the response of the component separation procedure used in this work to different sky models, synchrotron maps produced with two models of non-uniform spectral indices were used: the Miville-Deschênes et al. (2008) model, hereafter synchrotron MD, which uses WMAP data at 23 GHz, and Giardino et al. (2002) model, hereafter synchrotron GD, which is the result of the combination of the full-sky map of synchrotron emission at 408 MHz from Haslam et al. (1982), the northern-hemisphere map at 1420 MHz from Reich and Reich (1986) and the southern-hemisphere map at 2326 MHz from Jonas et al. (1998). The sta-

tistical properties of the spectral index maps, referring to the different synchrotron models used here, are presented in the Table 4.1.

Table 4.1 - Synchrotron emission models spectral indices.

Model name	Reference	Mean β_s	RMS β_s
synchrotron MD	Miville-Deschênes et al. (2008)	-3.00	0.06
synchrotron GD	Giardino et al. (2002)	-2.90	0.10

Figure 4.2 shows the synchrotron emission maps generated with the two pixel dependent spectral index models described in the Table 4.1 and Figure 4.3 presents their respective power spectra. The maps in this work are presented in celestial coordinates to allow the sky area observed by BINGO to be shown as a horizontal band (see region delimited by the dashed lines in the Figure 4.6).

Galactic free-free emission

Also known as thermal *bremsstrahlung*, the free-free radiation is emitted when a free electron is accelerated by an ion (generally a proton), without being captured, i.e., the electron is free before and after interacting. In the context of Galactic astrophysics, the free-free emission at radio frequencies is produced in warm ionized gas clouds of the interstellar medium, where the temperature reaches values that correspond to the hydrogen binding energy ($10^3 - 10^4$ K), enough to keep a considerable amount of electrons unbounded (ANDERSEN et al., 2020).

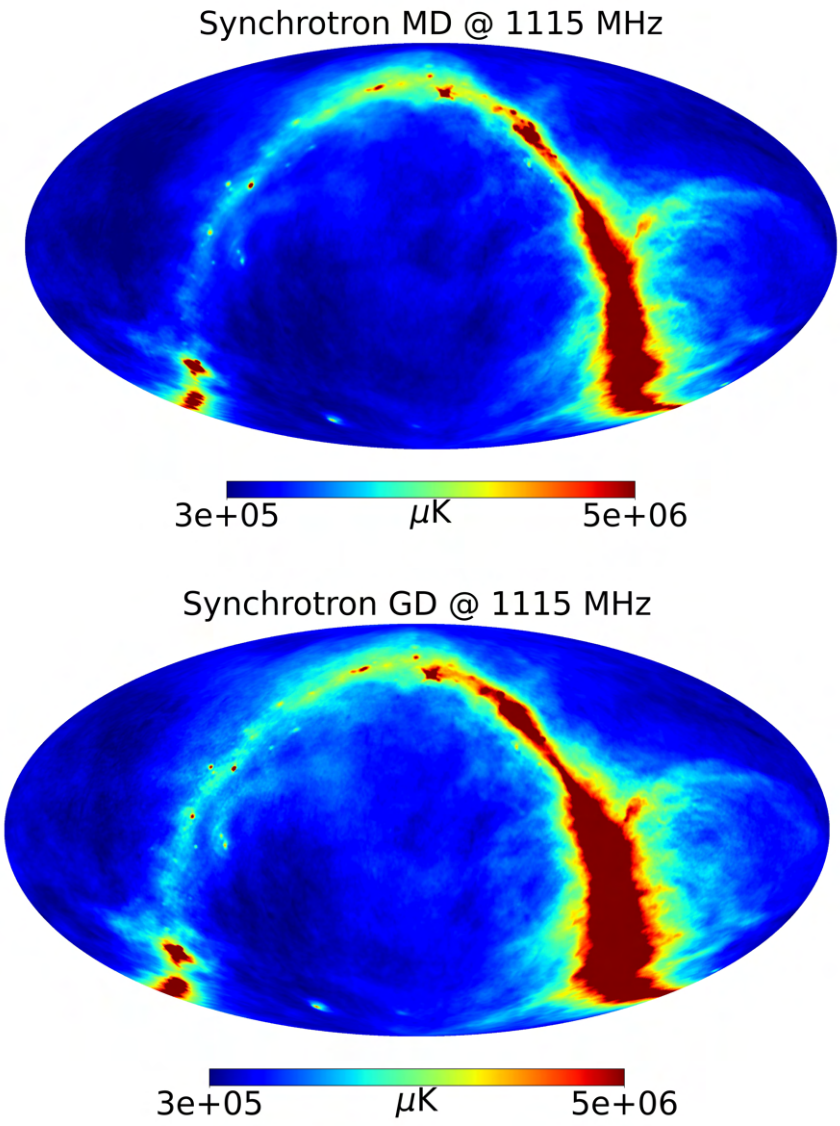
The free-free spectrum depends essentially on the number of free protons and electrons along the line-of-sight, which can be quantified by the emission measure, EM, given by

$$\mathrm{EM} = \int n_e^2 \mathrm{d}l, \tag{4.10}$$

where n_e is the electron number density, which is assumed to be equal to the number density of protons, and $\mathrm{d}l$ is the distance element over the line-of-sight. The EM is conventionally expressed in pc cm^{-6} and in the Milky Way assumes values between 0 and 1000 (PLANCK COLLABORATION, 2016b).

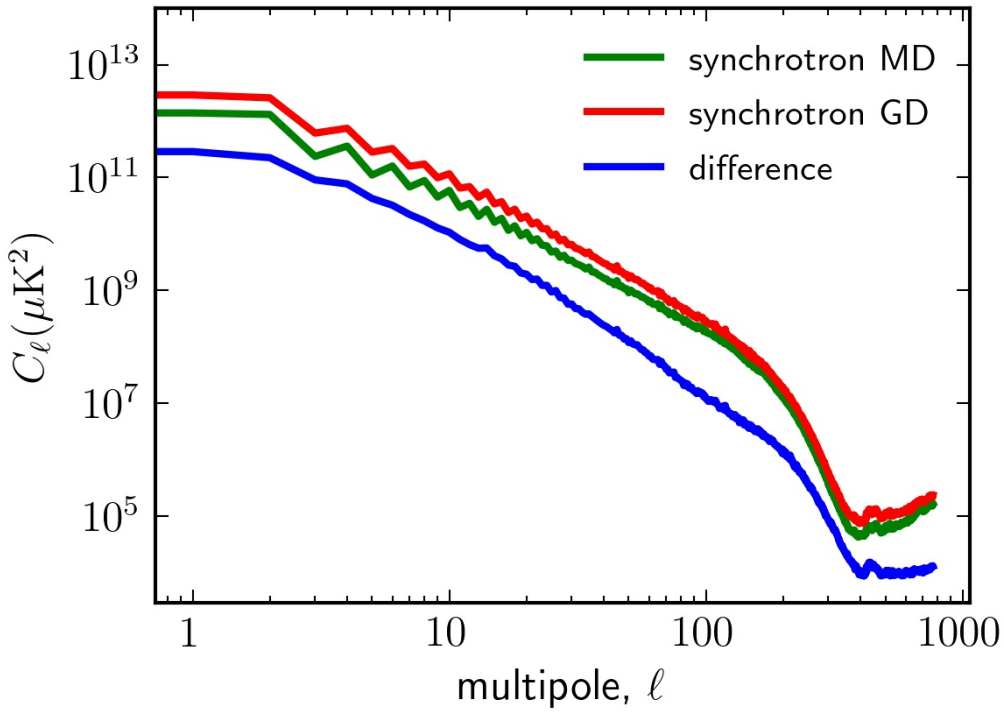
The optical Hα line is a good tracer of free-free emission, because it also depends on electron-ion interactions that occurs in regions where hydrogen is being ionized.

Figure 4.2 - Synchrotron emission full-sky maps for the MD model (*Top*) and the GD model (*Bottom*), defined in the frequency channel centered at 1115 MHz and with a bandwidth of $\delta\nu \sim 9.33$ MHz (30 bins). The maps are presented in celestial coordinates and were convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam.



In addition, free-free and Hα emissions are proportional to the emission measure, defined in Equation 4.10. However, the estimation of free-free emission from optical Hα line observations requires a correction for dust absorption, which is the most important uncertainty-causing factor, mainly in dense regions of the Galaxy. Therefore, an estimate of the EM over the sky is obtained from the map of Hα emission corrected for dust extinction presented in Dickinson et al. (2003).

Figure 4.3 - Power spectra referring to the MD (green) and GD (red) full-sky synchrotron maps shown in Figure 4.2, as well as the power spectrum calculated from the difference between them (blue), defined in the frequency channel centered at 1115 MHz and with a bandwidth of $\delta\nu \sim 9.33$ MHz (30 channels configuration). The maps were convolved with a $\theta_{\rm FWHM} = 40$ arcmin beam.



The brightness temperature of free-free emission can be written as

$$T_{\rm ff}(\nu) = T_{\rm e}(1 - e^{-\tau_\nu}), \tag{4.11}$$

where $T_{\rm e}$ is the electron temperature and τ_ν is the free-free opacity, which is well approximated by (DICKINSON et al., 2003; DRAINE, 2011)

$$\tau_\nu = 0.05468 \cdot T_{\rm e}^{-3/2} \cdot \left(\frac{\nu}{\rm GHz}\right)^{-2} \cdot {\rm EM} \cdot g_{\rm ff}. \tag{4.12}$$

The $g_{\rm ff}$ is known as Gaunt factor, which accounts for quantum and other corrections, and can be written as

$$g_{\rm ff} = \log\left\{\exp\left[5.960 - \frac{\sqrt{3}}{\pi}\log\left(\left(\frac{\nu}{\rm GHz}\right)\cdot\left(\frac{T_{\rm e}}{10^4{\rm K}}\right)^{-3/2}\right)\right] + e\right\}, \tag{4.13}$$

where e is the natural constant.

Since at GHz frequencies the ionized gas is optically thin ($\tau_\nu \ll 1$), the brightness temperature spectrum can be approximated by

$$T_{\rm ff}(\nu) \approx \tau_\nu T_{\rm e}. \tag{4.14}$$

Therefore, using the Equations 4.12, 4.13 and 4.14, the free-free spectrum can be approximated by (DICKINSON et al., 2003)

$$T_{\rm ff}(\nu) \approx 90\ {\rm mK}\left(\frac{T_{\rm e}}{\rm K}\right)^{-0.35}\left(\frac{\nu}{\rm GHz}\right)^{\beta_{\rm ff}}\left(\frac{\rm EM}{\rm cm^{-6}\ pc}\right), \tag{4.15}$$

where $\beta_{\rm ff}$ is the Galactic free-free spectral index, which is described by a slowly varying function of ν and $T_{\rm e}$, as (BENNETT et al., 1992)

$$\beta_{\rm ff} = -\left(2 + \frac{1}{10.48 + 1.5\ln\left(T_{\rm e}/8000\ {\rm K}\right) - \ln\left(\nu/{\rm GHz}\right)}\right). \tag{4.16}$$

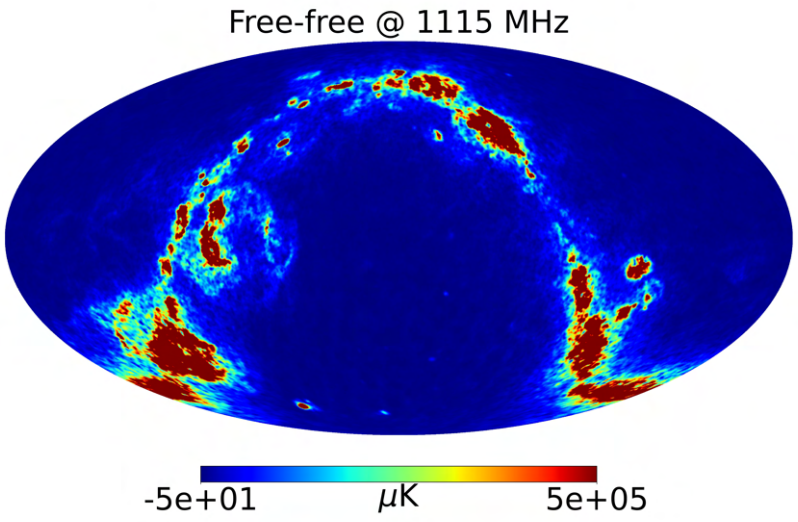
At GHz frequencies, the free-free spectral index takes the average value $\beta_{\rm ff} \approx -2.1$, with very little variation with electron temperature.

The free-free model used by PSM to produce the simulated sky takes two parameters as input: the free-free template map the temperature of free-free electrons. Thus, the free-free emission was simulated using the Hα template map corrected for dust extinction as described in Dickinson et al. (2003) and an uniform warm ionized gas temperature $T_{\rm e} = 7000$ K. The resulting emission map generated with the model described above is presented in the Figure 4.4.

Anomalous Microwave Emission

The most accepted model for anomalous microwave emission (AME) observed in the Milky Way is the spinning dust emission, which is produced in the interstellar medium mostly by sub-nanometer dust grains with electric dipole moment rotating

Figure 4.4 - Free-free emission full-sky map defined in the frequency channel centered at 1115 MHz and with a bandwidth of $\delta\nu \sim 9.33$ MHz (30 bins). The maps are presented in celestial coordinates and were convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam.



at GHz frequencies. The spinning dust emission depends directly on the dust grain size distribution, the type of dust, and environmental conditions, as density, temperature, interstellar radiation field. However, although the physics of this emission is well established, the main limitation in theoretically predicting the AME spectrum is the lack of knowledge about the characteristics of ultra-small dust grains in the interstellar medium (ISM) (DICKINSON et al., 2018).

The spinning dust emission mechanism was first described theoretically by Erickson (1957). This Galactic emission component was identified as a new anomalous component by Leitch et al. (1997) and presented in more details by Draine and Lazarian (1998) since it has been observed by CMB experiments and other instruments at radio and microwave frequencies during the 1990s. AME is one of the dominant foregrounds between 10 GHz and 60 GHz (DICKINSON et al., 2018), reaching its peak at ∼30 GHz. However, although AME is subdominant at BINGO frequencies (∼1 GHz), it was included in the simulations to produce a complete foreground sky.

Spinning dust and thermal dust emissions are typically correlated because both

are produced by the same dust grains. Furthermore, the dust-correlated AME was observed in different astrophysical environments, but the definitive evidence of spinning dust emission was provided by the Planck mission data from molecular clouds. The AME model predict that this emission is dominated by the smallest grains ($\sim< 0.001$m) as they are more sensitive to the processes that affect the rotational dynamics of the grain. The main processes are gas-grain interactions, plasma-grain interactions, infrared emission and the electric dipole radio emission (NAVARRO, 2015).

The AME is modelled here using a version of a 353 GHz thermal dust template, obtained with the application of `GNILC` to separate components of Planck satellite observations (PLANCK COLLABORATION, 2016e), scaled to the spinning dust emission at 22.8 GHz, taking into account a ratio between AME and thermal dust as described in (PLANCK COLLABORATION, 2016d). Then, this emission map is scaled to BINGO frequencies (980-1260 MHz) through the emission law modelled by (DRAINE; LAZARIAN, 1998), using as input to PSM (DELABROUILLE et al., 2013) a combination of 96.2% warm neutral medium (WNM) and 3.8% reflection nebulae (RN) as different environments of the ISM. This parameterization was adopted because it fits well to the average spectrum presented in Miville-Deschênes et al. (2008).

Thermal dust

This process of thermal (vibrational) emission occurs due to the interaction between the dust grains and the interstellar radiation field (IRF), both in thermal equilibrium. A large range of dust grains sizes (from a nanometer to a fraction of micrometer) and compositions are heated by the IRF, and this energy is then re-emitted mainly in the far-infrared (FIR) frequencies.

The specific intensity I_ν from an uniform dust grains population, in thermal equilibrium with IRF at temperature T_d, and evaluated in the optically thin limit, is well described from the data by a modified black-body as

$$I_\nu = B_\nu(T_\mathrm{d})\tau_\nu, \tag{4.17}$$

where $B_\nu(T_\mathrm{d})$ is the Planck function for a temperature T_d and τ_ν is the dust optical depth at frequency ν, which modifies the black-body shape of I_ν. The optical depth τ_ν is given by

$$\tau_\nu = \kappa_\nu M_\mathrm{d}. \tag{4.18}$$

where κ_ν is the dust emission cross section per mass unit, that in this case depends only on the chemical composition and structure of dust grains, and M_d is the dust mass column density, which can be written as

$$M_\mathrm{d} = r\ \mu\ m_\mathrm{H}\ N_\mathrm{H}, \tag{4.19}$$

where r is the dust-to-gas mass ratio, μ is the mean molecular weight, m_H is the mass of a hydrogen atom and N_H is the hydrogen column density. κ_ν is typically described by a power law (COMPIÈGNE et al., 2011)

$$\kappa_\nu = \kappa_{\nu_0} \left(\frac{\nu}{\nu_0}\right)^{\beta_\mathrm{d}}, \tag{4.20}$$

where κ_{ν_0} is the dust emission cross section per mass unity evaluated at a reference frequency ν_0 and β_d is the dust emission spectral index. Substituting the Equations 4.18, 4.19 and 4.20 into Equation 4.17, the following expression is obtained for dust emission in thermal equilibrium and with a given composition and structure

$$I_\nu = \kappa_{\nu_0}\ \left(\frac{\nu}{\nu_0}\right)^{\beta_\mathrm{d}}\ r\ \mu\ m_\mathrm{H}\ N_\mathrm{H}\ B_\nu(T_\mathrm{d}). \tag{4.21}$$

Three main parameters define the shape of the observed thermal dust emission spectrum: the strength of the IRF and the grain size distribution, which affect the equilibrium temperature, and the dust structure and composition, that determine the emission cross section, which is modeled in the simplest form by a power law (see Equation 4.20). In a more complex modeling, the dust emission spectral index β_d can be a function of frequency and/or grain size and/or grain temperature (PLANCK COLLABORATION, 2014a).

The distribution of dust grain sizes in a given volume element along a given line of sight is related with the distribution of equilibrium temperatures. Moreover, the dust grains properties can also vary along a sky direction and the IRF strength can change along some lines of sight. Thus, the observed thermal dust emission spectrum is given by the sum of different modified black-bodies. However, as dust emission

was observed in a few frequency bands, it is common to adjust the respective data to a simplified parameterization of a single black-body function, as (IRFAN et al., 2019)

$$I_\nu = \tau_{\nu_0} \left(\frac{\nu}{\nu_0}\right)^{\beta_{\mathrm{d,obs}}} B_\nu(T_{\mathrm{d,obs}}), \tag{4.22}$$

where τ_{ν_0} is the optical depth estimated at the reference frequency ν_0. This modified black-body model with three parameters was used by PSM to produce the thermal dust maps of this work.

Although the emission of dust grains is subdominant in the BINGO frequencies, this component was included in the simulated data used in this work to obtain a more realistic sky. To produce the thermal dust maps, it was used `GNILC` methodology applied to the Planck 2015 data release (PLANCK COLLABORATION, 2016e). The fitted parameters are Galactic dust optical depth τ_{ν_0}, evaluated at the reference frequency $\nu_0 = 353$ GHz; in addition to the dust temperature $T_{\mathrm{d,obs}} = 19.4 \pm 1.3$K and the dust spectral index $\beta_{\mathrm{d,obs}} = 1.6 \pm 0.1$, both averaged over the entire sky.

Faint radio point sources

There is a large number of compact Galactic and extragalactic sources emitting at radio frequencies. Among the compact Galactic sources are molecular clouds, supernova remnants, HII regions, and other radio sources. The extragalactic sources include a large number of radio galaxies, clusters of galaxies, and quasars (BATTYE et al., 2013; DELABROUILLE et al., 2013).

Point sources are objects whose angular size is much smaller than the angular resolution of an IM experiment, and which therefore appear in the data as point-like objects, after the convolution with the instrument beam (MARTINEZ et al., 2009; OLIVARI, 2018).

The brightest objects can be identified individually by a set of parameters as type, position, flux, etc, as a function of frequency, in a catalogue form. Hereafter called strong radio point sources, this category of objects is represented by populations that can be identified and removed using the respective information (BIGOT-SAZY et al., 2015).

The other objects compose a diffuse background from the integrated emission of a large number of extragalactic objects that are too faint to be detected individ-

ually. Hereafter called faint radio point sources (FRPS), they form a background continuum of unresolved radio sources, represented through frequency-dependent brightness fluctuation maps, produced by summing the contribution of large population of sources (DELABROUILLE et al., 2013; FORNAZIER et al., 2022).

To extrapolate the flux density of each point source to the BINGO frequencies, it is necessary to know the shape of its spectrum, which is usually very complex. However, the model used by PSM considers a power law approximation for the point sources spectra, of the form

$$S_\nu = S_{\nu_0} \left(\frac{\nu}{\nu_0} \right)^{-\alpha}, \tag{4.23}$$

where S_{ν_0} is the flux density at the reference frequency ν_0 and α is the flux density spectral index, which holds for a limited frequency range. The point sources spectrum can be classified as steep of flat, depending on its α value compared to the boundary value $\alpha = 0.5$. The mean spectral index of each population are $\langle \alpha_{\rm steep} \rangle = 1.18$ and $\langle \alpha_{\rm flat} \rangle = 0.16$. The flux density spectral index distributions are then approximated by Gaussians with the previous mean values, and variances $\sigma_{\rm steep} = \sigma_{\rm flat} = 0.3$, and so that each source is associated with a spectral index drawn randomly from the distribution (DELABROUILLE et al., 2013).

The point sources brightness temperature, $T_{\rm ps}$ can be written as a function of the differential source count, $\mathrm{d}N/\mathrm{d}S$, which represents the number of sources per steradian, N, per unit flux density, S. Considering that all sources with $S > S_{\rm max}$ can be subtracted (the strong radio point sources), the mean temperature is given by

$$\bar{T}_{\rm ps} = \left(\frac{\mathrm{d}B}{\mathrm{d}T} \right)^{-1} \int_0^{S_{\rm max}} S \frac{\mathrm{d}N}{\mathrm{d}S} \mathrm{d}S, \tag{4.24}$$

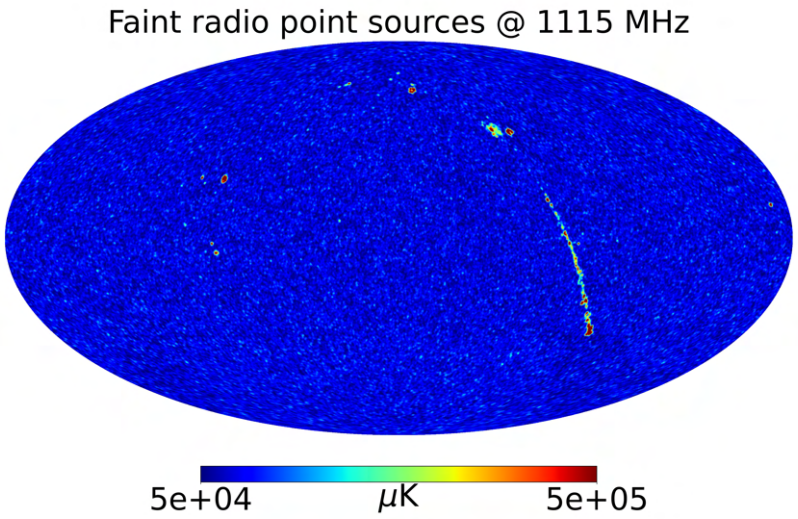
where $\mathrm{d}B/\mathrm{d}T = 2k_{\rm B}\nu^2/c^2$ is the conversion factor between intensity units to brightness temperature units, $k_{\rm B}$ is the Boltzmann constant and c is the speed of light in vacuum.

The FRPS model used here consists of simulated, low-flux, infrared sources, mostly in the Galactic region and from local galaxies, as well as the scaled contributions of extrapolated radio sources fluxes observed in three different frequency bands: 0.843, 1.4 and 4.85 GHz. It was used the set of frequencies 1, 30, 70, 353, 857 GHz and the

flux density limits 0.320, 0.427, 0.501, 0.304, 0.791 Jy to discriminate between strong and faint point sources, with only the latter being included in the simulations.

The generated FRPS map is shown in the Figure 4.5.

Figure 4.5 - Faint radio point sources emission full-sky map defined in the frequency channel centered at 1115 MHz and with a bandwidth of $\delta\nu \sim 9.33$ MHz (30 bins). The maps are presented in celestial coordinates and were convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam.



Thermal and kinetic Sunyaev-Zeldovich (SZ) effects

First described by Sunyaev and Zeldovich (1972), the Sunyaev-Zeldovich (SZ) effect occurs due to inverse Compton scattering of the CMB photons by hot electrons (at temperatures of a few keV) from ionised media, primarily in clusters of galaxies. This interaction shifts the energy of the incoming photons, producing secondary anisotropies in the CMB and causing a small spectral distortion in its spectrum (DELABROUILLE et al., 2013).

The SZ effect can be divided into two different contributions: thermal (tSZ) and kinetic (kSZ). The thermal SZ effect is attributed to the interaction between the

CMB photons and a hot and thermalised electrons in the intracluster gas of a cluster of galaxies. The kinetic SZ effect is associated with the interaction between CMB photons and a population of high energy electrons in a cluster that moves with a line of sight peculiar velocity in the CMB rest frame (MARTINEZ et al., 2009; ERLER et al., 2018).

The thermal SZ effect causes a distortion in the CMB spectrum that can be expressed by a temperature variation, ΔT_{tSZ}, at the dimensionless frequency $x \equiv h\nu/k_{\mathrm{B}}T_{\mathrm{CMB}}$, as

$$\frac{\Delta T_{\mathrm{tSZ}}}{T_{\mathrm{CMB}}} = yf(x, T_{\mathrm{e}}). \tag{4.25}$$

The Compton y-parameter describes the degree of interaction between photons and electrons, and is given by

$$y = \int n_{\mathrm{e}} \frac{k_{\mathrm{B}}T_{\mathrm{e}}}{m_{\mathrm{e}}c^2} \sigma_{\mathrm{T}} \, \mathrm{d}l, \tag{4.26}$$

where n_{e} is the electron number density, k_{B} is the Boltzmann constant, T_{e} is the temperature of the electron population, $m_{\mathrm{e}}c^2$ is the electron rest mass energy, σ_{T} is the Thomson cross-section, and $\mathrm{d}l$ is the distance element along the line of sight. The function $f(x, T_{\mathrm{e}})$ describes the frequency dependence of the thermal SZ effect as

$$f(x, T_{\mathrm{e}}) = \left(x \frac{e^x + 1}{e^x - 1} - 4 \right) (1 + \delta_{\mathrm{tSZ,rel}}(x, T_{\mathrm{e}})), \tag{4.27}$$

where $\delta_{\mathrm{tSZ,rel}}(x, T_{\mathrm{e}})$ is the relativistic correction to the frequency dependence of the thermal SZ effect (CARLSTROM et al., 2002).

The kinetic SZ effect causes a distortion in the CMB spectrum due to the Doppler effect of the cluster velocity on the scattered CMB photons, which can be expressed by a temperature variation, ΔT_{kSZ}, as

$$\frac{\Delta T_{\mathrm{kSZ}}}{T_{\mathrm{CMB}}} = -\tau_{\mathrm{e}} \left(\frac{v_{\mathrm{pec}}}{c} \right), \tag{4.28}$$

where v_{pec} is the line of sight peculiar velocity with which a population of electrons

moves in the CMB rest frame and τ_e is the optical depth of the plasma, given by $\tau_e = \sigma_T \int n_e dl$. The relativistic corrections for the kinetic SZ can be neglected when compared to the sensitivity of current experiments(ERLER et al., 2018).

The thermal and kinetic SZ effects were simulated with the PSM through cluster number counts from cluster mass function $N(M, z)$, as described in Delabrouille et al. (2002). The function $N(M, z)$ provides, for a given set of cosmological parameter values, the number density $(dN/dM)dz$ of clusters of mass M at redshift z. The lower and upper mass limits of clusters included in the simulations, in units of 10^{15} solar masses, are 1 and 10, respectively. The cluster distribution in the mass-redshift plane (M, z) is drawn from a Poisson distribution whose mean is given by mass function presented in Tinker et al. (2008). The clusters are then randomly distributed on the sphere, with an uniform probability distribution function.

The three-dimensional clusters velocities are drawn from a Gaussian distribution with zero mean and standard deviation given by the power spectrum of density fluctuations (PEEBLES, 1993). This velocity field and the electron density of clusters are used to produce the map of the kinetic component of the SZ effect.

The cluster catalogues generated with the method described above is improved by adjusting the simulated data to observed cluster catalogues. Then, a subset of randomly drawn clusters which best matches the observed ones (same bin of mass and redshift, proximity of positions in the sky) were replaced by known real catalogues. In the simulations used in this work were used the clusters catalogues from Planck, South Pole Telescope (SPT), Atacama Cosmology Telescope (ACT), *Röntgensatellit* (ROSAT) and Sloan Digital Sky Survey (SDSS).

The temperature change caused by the SZ effect are below the targeted 21 cm signal and were included to produce a simulated sky maps closer to reality at BINGO frequencies.

4.1.3 Galactic mask

As seen in Section 4.1.2, astrophysical foregrounds are more intense in the galactic plane and there is an intersection of it and the region of the sky observed by BINGO. Thus, to remove this region from the BINGO simulated maps, and consequently facilitate the component separation process, it was used a galactic mask whose production is described in this section.

Figure 4.6 shows the combined emission at 984.7 MHz of all foregrounds presented in

the Section 4.1.2, as well as the region of the sky covered by BINGO. The observed region is a declination strip of $\approx 15^\circ$ centered at $\delta = -15^\circ$. Note that the region where the galactic foregrounds are more intense crosses the area covered by the instrument. The brightness temperature in this intersection region reaches 56 K, while the HI signal amplitude is $\approx 200\ \mu$K.

To create the final BINGO mask, it was used the galactic mask shown in Figure 4.6 which covers 20% of the sky, cutting off the most intense emission region of the galactic plane. The result of applying this galactic mask to the all foregrounds map can be seen in the Figure 4.6. The maximum temperature of the area covered by BINGO after masking the galactic plane is about 6.5 K.

The BINGO observed region is defined by the feed horn arrangement in the focal plane and by the observation strategy (WUENSCHE et al., 2022). The effective masked region is the intersection between the galactic mask discussed above and the observed region. To avoid boundary artifacts in calculating the power spectrum of the masked partial maps of the sky, it was used the `NaMaster` [1] (ALONSO et al., 2019) package to apodize the BINGO galactic mask. The `NaMaster` apodization used is type C2, a function that gives the factor by which each pixel is multiplied, and with a width of 5°, the maximum angular separation between the apodized pixel and the nearest masked (null) pixel. The resulting mask is shown in the Figure 4.6, with a visible area of 12.2 % of the sky.

4.1.4 Thermal noise

The temperature fluctuations produced by the thermal agitation of charged particles in electronic components are called thermal noise. Due to its spectral and statistical properties, this component can be accurately modeled by a Gaussian white noise distribution. The white noise level (or root mean square, RMS, value) per pixel, that defines the BINGO theoretical sensitivity per pixel, is given by

$$\sigma_{\rm t} = \frac{T_{\rm sys}}{\sqrt{t_{\rm pix}\delta\nu}}, \tag{4.29}$$

where $T_{\rm sys}$ is the system temperature, which is assumed to be 70 K for all BINGO receivers; $\delta\nu$ is the frequency channel width; $t_{\rm pix}$ is the integration time per pixel, given by

[1] `https://namaster.readthedocs.io`

$$t_{\rm pix} = N_{\rm horns} \frac{t_{\rm obs}}{\Omega_{\rm obs}/\Omega_{\rm beam}}, \tag{4.30}$$

where $N_{\rm horns}$ is the number of horns; $t_{\rm obs}$ is the total observation time; $\Omega_{\rm obs}$ is the observed area of the sky; $\Omega_{\rm beam}$ is the beam area.

Due to the arrangement of horns in the BINGO focal plane, the observation time $t_{\rm obs}$ is not uniformly distributed over the sky area covered by the instrument. In this case, each horn covers the pixels of a fixed-latitude ring. Therefore, the total observation time of a pixel depends on the latitude.

The focal plane configuration adopted for the BINGO Phase 1 predicts $N_{\rm horns} = 28$, with less horns positioned at the vertical edges than inner region of the arrangement, which makes the white noise level at the extreme declinations of the map more intense (shorter observation time per pixel). To homogenize the white noise level map, at least in the innermost region, a repositioning of the horns is considered every year of mission, within a total of $t_{\rm obs} = 5$ years. In this way it is possible to obtain a better distribution of observation time over the pixels in the covered area of the sky (LICCARDO et al., 2022).

The left panel of Figure 4.7 shows the RMS noise per-pixel map, produced with the feed horn arrangement and observation strategy described above. It can be noticed the highest noise level in the extreme declinations and approximately homogeneous in the innermost region. The center panel shows the white noise map, which is the result of multiplying the RMS white noise map by a Gaussian map with null mean and standard deviation equal to 1. The map on the right shows the result of applying the apodized mask described in the Section 4.1.3 to the white noise map. It can be observed how the apodized mask attenuates the noise in the region of the edges of the map, homogenizing the white noise map and facilitating the process of reconstruction of the 21 cm signal.

In this analysis the most common systematics for a non-ideal instrument such as $1/f$ noise and radio-frequency interference (RFI) are not taken into account. The impact of these systematic effects on the recovery of the cosmological signal will be addressed in the future.

4.2 Simulation plan

The models used to produce the maps of the cosmological, astrophysical and systematic noise components were described in Section 4.1. This section presents the different configurations used to produce the sets of simulated BINGO maps, generated with the components presented above, used to investigate the efficiency of the component separation pipeline used in this work. The sets of simulated maps contains observations with different foreground models, different numbers of frequency bins (number of maps) and made at a frequency outside the BINGO band by another experiment.

In order to test the robustness of the component separation method against different foregrounds models, two sets of simulated maps were created. Each set contains a different synchrotron map, produced from the MD and GD spectral index models, in addition to all other components described in the Section 4.1. These two sets were binned into 30 frequency maps, following the BINGO project baseline configuration, and were called MD30 and GD30.

The sky at low radio frequencies (< 10 GHz) is dominated by astrophysical foregrounds, mainly galactic synchrotron emission. Independent foreground observations in frequencies outside the BINGO band may improve the characterization of the synchrotron contribution and other components, facilitating its removal. The C-Band All-Sky Survey (CBASS) is an all-sky survey at a frequency of 5 GHz and 1 GHz bandwidth, with a sensitivity $\lesssim 0.1$ mK RMS (per beam) and a resolution of 45 arcmin, designed to provide complementary data to the CMB surveys (JONES et al., 2018). Therefore, it was performed a component separation test adding the simulated CBASS sky containing the same components of the BINGO sky, as well as the CBASS noise, to the MD30 base data set. The combination of these simulated data was called MD30+CBASS. The CBASS noise level is ~ 437 μK for a `HEALpix` $N_{\rm side} = 256$. The simulated CBASS all-sky foreground and noise emission maps are shown in Figure 4.8.

To evaluate the pipeline efficiency in recovering the 21 cm signal from the input data arranged in different numbers of frequency channels, were generated sets of simulated maps with 20, 30, 40, 60 and 80 bins. For this analysis, the simulated sky was generated using the synchrotron MD model, in addition to the other components described in the Section 4.1. The map sets created were called MD20, MD30, MD40, MD60 and MD80.

The simulated BINGO maps used in the tests described above are the result of the sum of HI and foregrounds, convolved with a 40 arcmin beam, plus the estimated BINGO noise map. In the case of combining BINGO and CBASS data to improve the 21 cm signal reconstruction, in addition to the BINGO maps, it was added the CBASS foregrounds plus noise map. In this configuration, the maps of the cosmological and astrophysical components, both from BINGO and CBASS, are convolved with a beam of 45 arcmin. The CBASS foregrounds map contains the same components considered in the BINGO data (see Section 4.1). A summary of the simulation plan of this work is presented in Table 4.2.

Table 4.2 - Simulation plan.

Set	**Foregrounds**	**Number of channels**
MD30	Synchrotron MD + freefree + AME + TD + SZ + FRPS	30 (BINGO)
MD30 +CBASS	Synchrotron MD + freefree + AME + TD + SZ + FRPS	30 (BINGO) + 1 (CBASS)
GD30	Synchrotron GD + freefree + AME + TD + SZ + FRPS	30 (BINGO)
MD20	Synchrotron MD + freefree + AME + TD + SZ + FRPS	20 (BINGO)
MD40	Synchrotron MD + freefree + AME + TD + SZ + FRPS	40 (BINGO)
MD60	Synchrotron MD + freefree + AME + TD + SZ + FRPS	60 (BINGO)
MD80	Synchrotron MD + freefree + AME + TD + SZ + FRPS	80 (BINGO)

Figure 4.6 - Description of the masking process. *Top left*: map with the sum of all foregrounds considered in this work, including synchrotron MD model (see Section 4.1.2), in the lowest frequency bin of a total of 30 channels, centered on 984.7 MHz and limited to 10 K. *Top right*: result of applying the galactic mask to the all foregrounds map. *Bottom left*: BINGO apodized (5 deg) galactic mask, preserving a sky fraction of 12.2%. *Bottom right*: result of applying the BINGO mask to the all foregrounds map centered at 984.7 MHz. All maps are in celestial coordinates and the dashed lines delimit the BINGO covered area.

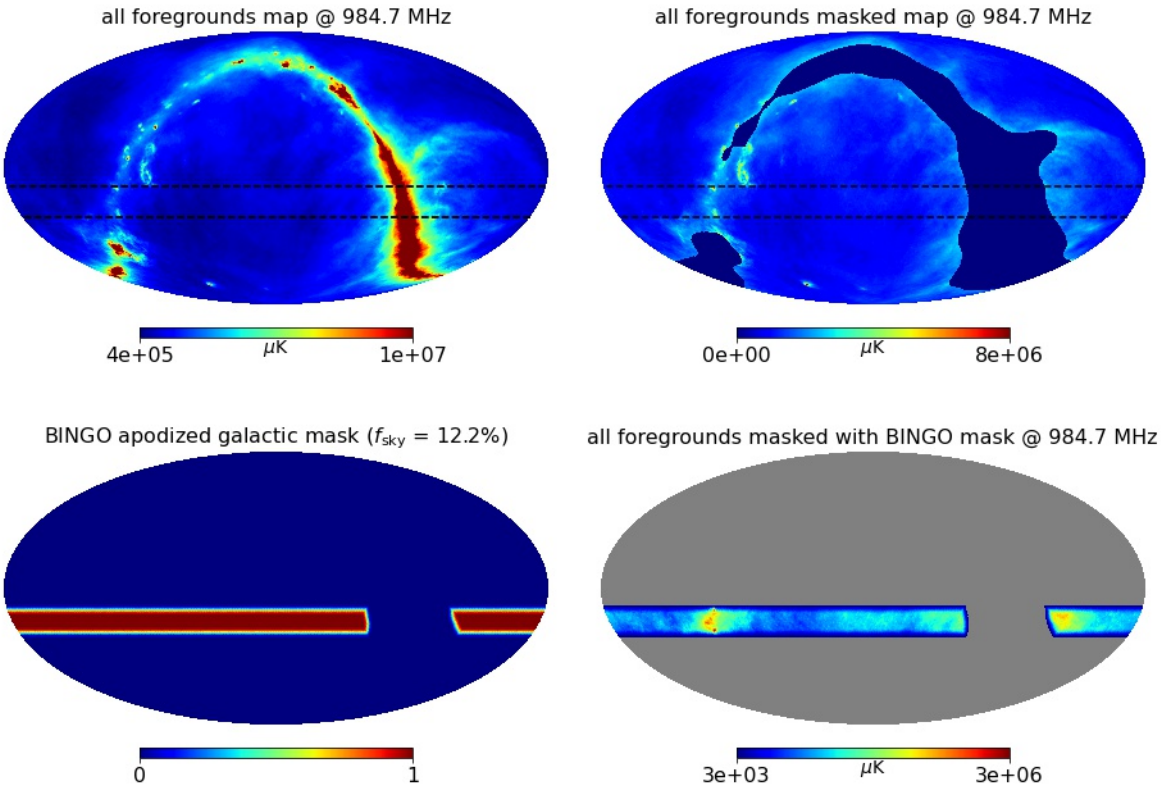


Figure 4.7 - White noise maps for the BINGO Phase 1 horn arrangement ($N_{\text{horns}} = 28$). The maps were generated with a $T_{\text{sys}} = 70$ K, 30 frequency bands, `HEALpix` $N_{\text{side}} = 256$ and a fraction of the sky f_{sky} equal to 13 %. *Left*: white noise level (RMS) map. *Center*: white noise map, a Gaussian realisation of the corresponding RMS map (see text for details). *Right*: White noise realisation map with the apodized mask. The color scale is saturated at 3 times the RMS of a map with homogeneous coverage and same sky fraction. The maps are presented in Gnomonic projection and in a resolution given by the `HEALpix` $N_{\text{side}} = 256$. They are in celestial coordinates centered at $\alpha = 0$ and $\delta = -18°$.

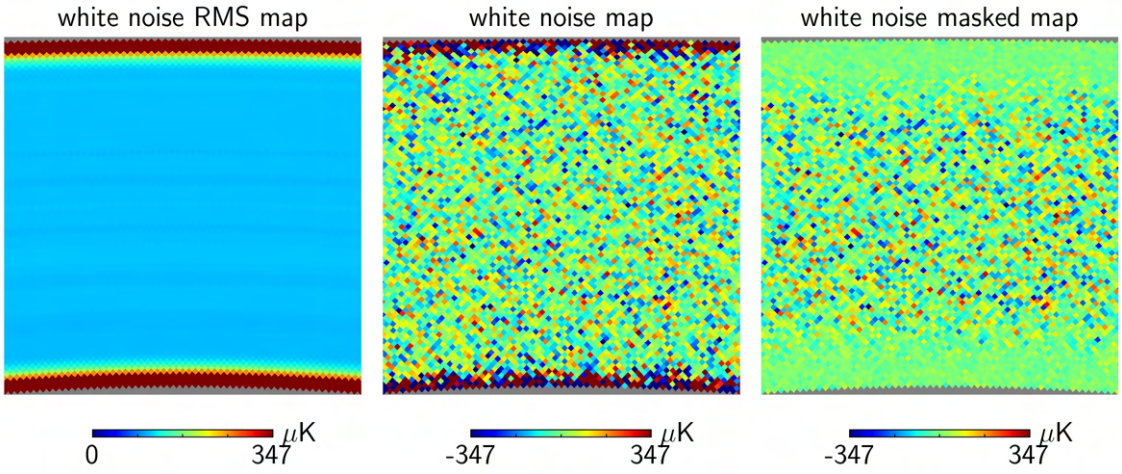
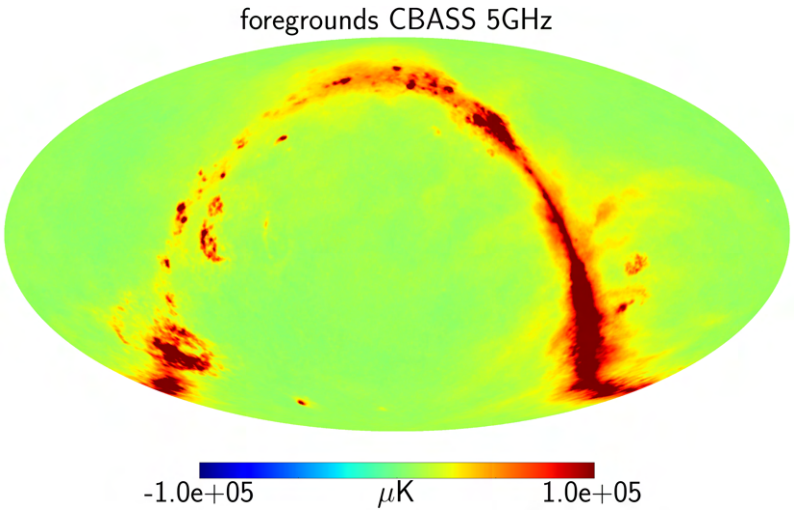
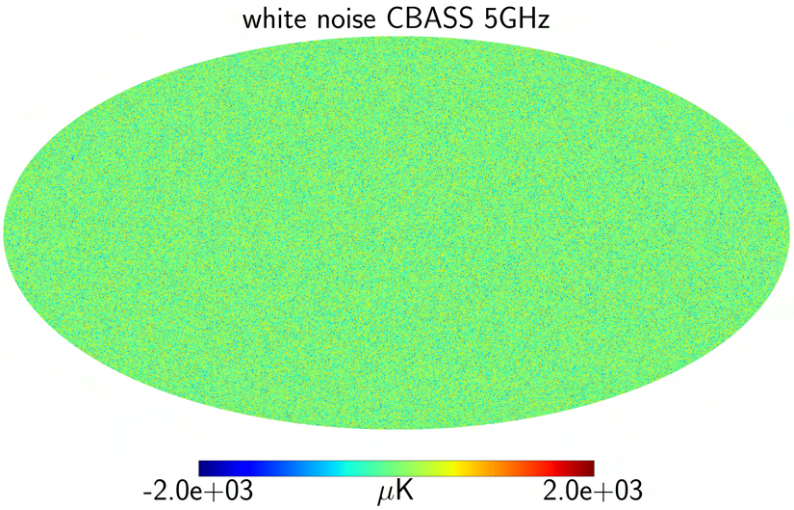


Figure 4.8 - *Left*: CBASS all-sky foregrounds map, result of the sum of the components described in the Table 4.2. *Right*: CBASS white noise map. The temperature scale of the foregrounds map is saturated at $\pm 10^5 \mu K$. The temperature scale of the noise map is saturated at $\pm 2 \times 10^3 \mu K$ and the noise level is $\sim 437 \mu$K.





5 THE PROPOSED COMPONENT SEPARATION PROCESS

As seen in Chapter 4, recovering the 21 cm signal from the data obtained by an IM experiment is a big challenge, which makes component separation a crucial step in measuring BAO with BINGO. Then, this chapter presents the problem of component separation (Section 5.1), as well as a detailed description of the GNILC method (Section 5.2), in addition to a power spectrum debiasing procedure (Section 5.3).

5.1 Component separation problem

Before carrying out the component separation process, the foregrounds that contaminate IM data can be reduced by some observational strategies: selecting the observed sky region, avoiding observing the Galactic plane; masking regions of the sky with high contamination, as the Galactic plane and strong (bright) radio point sources; choosing the instrument frequency band; subtracting an estimate of the foregrounds based on observations, such as the CMB (DELABROUILLE; CARDOSO, 2008).

As discussed in Chapter 4, the observed sky in radio frequencies is a mixture of several physical sources of emission, including cosmological signals from the early universe, Galactic and extragalactic sources emitting in the late universe (astrophysical foregrounds), instrumental noise (thermal and 1/f), atmospheric noise, and radio frequency interference (RFI) (OLIVARI, 2018). However, these components cannot be properly studied in data sets in which they appear combined. Thus, component separation consists of identifying and isolating the emission of a component of interest (here the 21 cm signal) from all other components present in the data, making use of the existing correlations in multi-frequency observations, external constraints and physical modeling. In addition to pixel emission, the component separation process may be able to estimate other parameters that statistically describe the component of interest, such as the power spectrum (DELABROUILLE; CARDOSO, 2008; LEACH et al., 2008).

The component separation problem can be described by instrument observations modeled as a superposition of components at frequency ν and pixel p, as

$$d_\nu(p) = s_\nu^{\text{cosmo}}(p) + f_\nu(p) + n_\nu(p) \tag{5.1}$$

where $d_\nu(p)$ is the observational data map, $s_\nu^{\text{cosmo}}(p)$ is the cosmological signal of

interest map (e.g., 21 cm emission), $f_\nu(p)$ is the sum of the foregrounds emission maps (e.g., galactic foregrounds and extra galactic point sources) and $n_\nu(p)$ is the systematic noise map (e.g., thermal and 1/f noise).

The estimate of the signal to be recovered can be written in vector form, with dimension $N_{\rm ch} \times 1$, where $N_{\rm ch}$ is the number of frequency channels,

$$\hat{\mathbf{s}}^{\rm cosmo}(p) = \mathbf{s}^{\rm cosmo}(p) + \epsilon(p), \tag{5.2}$$

where $N_{\rm ch}$ is the number of frequency channels and $\epsilon(p)$ is the estimation error vector. Thus, the component separation is summarized as an estimation problem for the signal $\mathbf{s}(p)$, minimizing the error $\epsilon(p)$ under some restriction. In general, this is done by exploring the correlations between data obtained at different frequencies, using or not a prior information about foregrounds, noise and the cosmological signal.

In this section, component separation is presented in detail, starting with the physics of the problem, going through the classification of available methods and ending with the motivation for choosing GNILC as the option to be used in this work.

The physics of 21 cm component separation

In the low-frequency radio sky, the weak 21 cm radiation from neutral hydrogen is obscured by a much brighter astrophysical foregrounds emissions, which makes the detection of the cosmological signal a challenging task. These emissions can be of Galactic and extragalactic origin, and the knowledge of their physical properties is indispensable to mitigate the foregrounds in the 21 cm cosmology experiments data. Diffuse Galactic emissions are produced in the local interstellar medium (ISM) of the Milky Way, are mainly located in the Galactic plane and dominate at large angular scales ($> 1°$). The dominant Galactic components in radio are synchrotron and free-free emissions. The extragalactic radio emissions are generated by a background of resolved and unresolved sources, that dominates at small angular scales. The global astrophysical foreground emission in radio is three to four orders of magnitudes brighter than the cosmological 21 cm signal. For more details about the astrophysical foregrounds relevant at BINGO frequencies, see Section 4.1.2.

The component separation methods for IM try to obtain a foregrounds model to estimate and remove them from the observations, leaving the cosmological 21 cm signal, instrumental noise (depending on the adopted strategy) and a residue of

foregrounds. In the case where the component separation strategy isolates the target signal along with the instrumental noise, the power spectrum of the latter can be estimated from the data and then removed to recover the noiseless 21 cm power spectrum (MESINGER, 2019).

As the dominant mechanism in Galactic and extragalactic radio emissions is synchrotron radiation, the foreground spectra are expected to be smooth (or correlated) in frequency (MESINGER, 2019), which means that they can be approximated by power laws in the frequency band of the instrument. This physical property is used by some component separation methods, including the one used in this work, as it allows the foregrounds to be statistically differentiated from the 21 cm signal, since the latter is not spectrally smooth.

Component separation methods

There are a variety of component separation methods available, most of which have already been tested in CMB data analysis before being used in the IM context. In both CMB and IM, observations are dominated by astrophysical foregrounds. However, due to the spectral characteristics of the CMB, the component separation is more complicated in the case of the IM.

A good component separation method greatly contributes to the accurate reconstruction of the cosmological signal. On the one side, it works towards minimizing the foreground contamination in the data, on the other side, it prevents the loss of the cosmological signal. An efficient component separation procedure reduces the uncertainties arising during the foreground removal process and their propagation into the cosmological signal power spectrum, avoiding introducing some bias in the estimation of cosmological parameters.

The component separation methods can be divided into two groups: parametric and non-parametric (or blind) methods. The parametric methods use a model to describe some physical properties of the foregrounds and remove them from the data. The non-parametric methods dispense prior knowledge about the foregrounds and recover the cosmological component using only the data and the cosmological signal information (OLIVARI, 2018).

Among the parametric methods are Parametric Fitting (BRANDT et al., 1994; ANSARI et al., 2012; LIU; TEGMARK, 2012; BIGOT-SAZY et al., 2015), Karhunen-Loeve Decomposition (BOND, 1995; VOGELEY; SZALAY, 1996; TEGMARK et al., 1997; SHAW

et al., 2014), Correlated Component Analysis (CCA)(BEDINI et al., 2005; BONALDI; BROWN, 2015), Wiener Filtering (BUNN et al., 1994; TEGMARK; EFSTATHIOU, 1996), Gibbs sampling (JEWELL et al., 2004; WANDELT et al., 2004; LARSON et al., 2007; ERIKSEN et al., 2008) and Commander is an Optimal Monte-carlo Markov chAiN Driven EstimatoR (Commander) (ERIKSEN et al., 2004; ERIKSEN et al., 2006; ERIKSEN et al., 2008; PLANCK COLLABORATION, 2016c; PLANCK COLLABORATION, 2020a).

Non-parametric methods include Principal Component Analysis (PCA) (MURTAGH; HECK, 1987; SWITZER et al., 2013; ALONSO et al., 2015; BIGOT-SAZY et al., 2015), Independent Component Analysis (ICA)(AMARI; CICHOCKI, 1998; BACCIGALUPI et al., 2000; BACCIGALUPI et al., 2004), Spectral Matching Independent Component Analysis (SMICA) (DELABROUILLE et al., 2003; PATANCHON et al., 2005; CARDOSO et al., 2008; BETOULE et al., 2009), Spectral estimation via expectation maximisation (SEVEM) (MARTÍNEZ-GONZÁLEZ et al., 2003; LEACH et al., 2008; FERNÁNDEZ-COBOS et al., 2012; PLANCK COLLABORATION, 2014b; PLANCK COLLABORATION, 2016a), Quadratic Estimators (TEGMARK, 1997; TEGMARK et al., 1998; LIU; TEGMARK, 2011; SWITZER et al., 2015), HI Expectation-Maximization Independent Component Analysis (HIEMICA)(ZHANG et al., 2016), Generalized Morphological Component Analysis (GMCA)(BOBIN et al., 2007; LEACH et al., 2008; CHAPMAN et al., 2013), Fast Independent Component Analysis (FastICA)(MAINO et al., 2002; MAINO et al., 2003; WOLZ et al., 2015), Internal Linear Combination (ILC) and its variants (TEGMARK; EFSTATHIOU, 1996; TEGMARK et al., 2003; BENNETT et al., 2003; ERIKSEN et al., 2004; SAHA et al., 2006; DELABROUILLE et al., 2009; BASAK; DELABROUILLE, 2012; BASAK; DELABROUILLE, 2013; REMAZEILLES et al., 2011a) and Generalized Needlet Internal Linear Combination (GNILC)(REMAZEILLES et al., 2011b; OLIVARI et al., 2016).

`GNILC` as the main choice

The Generalized Needlet Internal Linear Combination (`GNILC`) is a component separation method that uses both frequency and spatial information to isolate the target signal from the data (OLIVARI et al., 2016). This method does not require previous information about foregrounds (non-parametric method). However, it uses a theoretical estimate of the target signal power spectrum, which here is called prior. In addition, it uses the concept of wavelets to perform component separation by pixel regions and angular scale intervals, adapting to different local contamination conditions. The `GNILC` method was initially successfully used in the context of CMB, separating components of Planck data (PLANCK COLLABORATION, 2016e), and then it was adapted to be used in HI IM experiments (OLIVARI et al., 2016; LICCARDO et

al., 2022; FORNAZIER et al., 2022; MERICIA et al., 2023). A detailed description of the method is presented in Section 5.2.

When trying to recover the HI signal as a single component using GNILC, it was observed that the reconstructed power spectrum was contaminated by a residual content of thermal noise, mainly at small scales (OLIVARI, 2018), which contributes to the 21 cm component estimation error. Since it is possible to obtain the thermal noise characteristics with good precision in an experiment, it was chosen to reconstruct the HI and noise signals as a single component in the GNILC stage. Then, it was used a debiasing procedure, described in the Section 5.3, to estimate and remove the noise content of the GNILC output power spectra, in order to reconstruct the pure HI signal.

5.2 The GNILC method

The GNILC method was developed for use in removing foregrounds from CMB experiment data (REMAZEILLES et al., 2011b) and later adapted to the HI IM (OLIVARI et al., 2016). The GNILC basic idea is to use not only spectral information but also spatial information (angular power spectrum) to discriminate between foregrounds and the targeted signal. GNILC is a blind method, that is, it does not assume prior knowledge about the properties of foregrounds, but only about the power spectrum of the target signal. The following is a description of the method, as applied to this work.

Considering 21 cm plus thermal noise as the target signal $s_\nu(p)$, such that $s_\nu(p) = s_\nu^{21\mathrm{cm}}(p) + n_\nu(p)$, as discussed in Section 5.1, Equation 5.1 can be rewritten as

$$d_\nu(p) = s_\nu(p) + f_\nu(p). \tag{5.3}$$

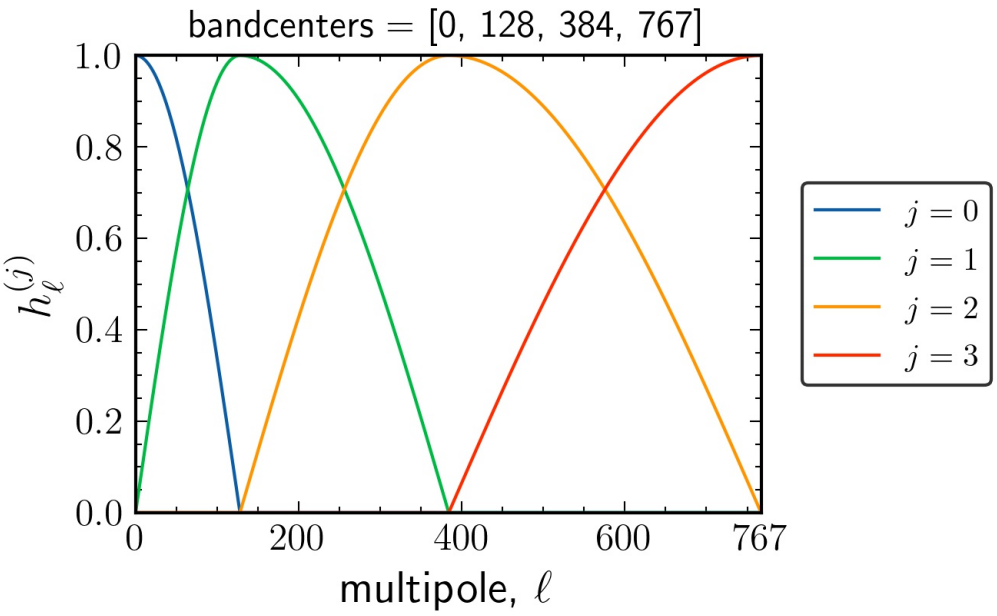
Transforming the data in the pixel domain to the spherical harmonic domain, we obtain the harmonic coefficients $d_\nu(\ell, m)$. Then, each frequency map can be decomposed in harmonic space, using a kind of bandpass filters defined by different multipole intervals ℓ. These filters are called needlets, an application of the concept of wavelets to the sphere.

A needlet window is defined by the function $h_\ell^{(j)}$, where the index (j) indicates the range of angular scales ℓ selected by the bandpass. A set of needlets respects the following relation

$$\sum_j \left[h_\ell^{(j)}\right]^2 = 1. \tag{5.4}$$

To define the set of needlets used as input for `GNILC` in this work, different bandpass combinations were tested. Figure 5.1 presents the best match for the simulations, a set of cosine-shaped needlets, with peaks centered at $[0, 128, 384, 767]$.

Figure 5.1 - Set of cosine-shaped needlets with peaks located at $\ell = [0, 128, 384, 767]$.



Thus, each $d_\nu(p)$ is decomposed into a map $d_\nu^{(j)}(p)$, which is the result of the inverse spherical harmonic transform of $d_\nu(\ell, m) \times h_\ell^{(j)}$, according to

$$d_\nu^{(j)}(p) = \sum_\ell \sum_{m=-\ell}^{\ell} d_\nu(\ell, m)\, h_\ell^{(j)} Y_{\ell m}(p)\,, \tag{5.5}$$

where (j) defines the angular scales isolated by the respective needlet function. For each pixel p and each needlet scale (j), a covariance matrix $\mathbf{R}^{(j)}(p)$ is calculated, whose elements for a frequency pair ν, ν' are given by

$$R^{(j)}_{\nu\nu'}(p) = \frac{1}{N^{(j)}_p} \sum_{p' \in \mathfrak{D}_p} d^{(j)}_{\nu}(p') d^{(j)}_{\nu'}(p'), \tag{5.6}$$

where $N^{(j)}_p$ is the number of pixels that constitute the domain $\mathfrak{D}_p$, centered on the pixel p and whose size is chosen accordingly to avoid creating artificial anti-correlations between the target signal and the foregrounds. These anti-correlations created when working with small regions of the sky (with reduced number of pixels), can result in power loss of the reconstructed signal. This can be measured by the multiplicative ILC bias b, according to Delabrouille et al. (2009). Thus, the number of pixels $N^{(j)}_p$ is adjusted, for each needlelet scale (j), according to the choice of the ILC bias b value, that is defined as one of the `GNILC` input parameters. The relation between the number of pixels in the $\mathfrak{D}_p$ domain and the ILC bias is given by

$$N^{(j)}_p = \frac{n^{(j)}_p \, (n_{\rm ch} - 1)}{|b| \, N^{(j)}_m}, \tag{5.7}$$

where $n^{(j)}_p$ is the total number of pixels in the j-needlet map, $N_{\rm ch}$ is the number of frequency channels, and $N^{(j)}_m$ is the number of modes in the j-needlet window, which can be written as

$$N^{(j)}_m = \sum_{\ell=0}^{\ell_{\rm max}} (2\ell + 1) \, [h^{(j)}_\ell]^2. \tag{5.8}$$

`GNILC` uses an estimate of the target signal covariance matrix, $\hat{\mathbf{R}}^{(j)}_s(p)$, produced with map realisations from prior knowledge of the HI + thermal noise power spectrum, to transform and diagonalize the data covariance matrix, $\mathbf{R}^{(j)}(p)$, disentangling foregrounds and signal subspaces. Omitting needlet scale (j) and pixel p, the diagonalization of the matrix resulting from the transformation $\hat{\mathbf{R}}_s^{-1/2} \mathbf{R} \hat{\mathbf{R}}_s^{-1/2}$, can be approximated by

$$\hat{\mathbf{R}}_s^{-1/2}\mathbf{R}\hat{\mathbf{R}}_s^{-1/2} \simeq$$

$$\simeq \begin{bmatrix}\mathbf{U}_f & \mathbf{U}_s\end{bmatrix} \times \begin{bmatrix} 1+\mu_1 & & & & & & \\ & \ddots & & & & & \\ & & 1+\mu_m & & & & \\ & & & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \end{bmatrix} \times \begin{bmatrix}\mathbf{U}_f^T \\ \\ \mathbf{U}_s^T\end{bmatrix} \quad (5.9)$$

The eigenvalues of $\hat{\mathbf{R}}_s^{-1/2}\mathbf{R}\hat{\mathbf{R}}_s^{-1/2}$ that are $\simeq 1$ do not contain any relevant power of the foregrounds, i.e. the sky emission is dominated by the target signal. The corresponding subset of eigenvectors $\mathbf{U}_s$ spans the target signal subspace. The subset of m eigenvectors $\mathbf{U}_f$, for which eigenvalues of $\hat{\mathbf{R}}_s^{-1/2}\mathbf{R}\hat{\mathbf{R}}_s^{-1/2}$ depart significantly from unity ($1+\mu_i \gg 1$), spans the foregrounds subspace. Then a Principal Component Analysis (PCA) (MURTAGH; HECK, 1987) is performed on the diagonalized covariance matrix to disentangle foregrounds and target signal subspaces. For each needlet scale (j) and each pixel p, is selected the subset of $m^{(j)}(p)$ eigenvectors $\mathbf{U}_f$ of covariance matrix $\hat{\mathbf{R}}_s^{-1/2}\mathbf{R}\hat{\mathbf{R}}_s^{-1/2}$, for which eigenvalues are $\gg 1$ by using a statistical Akaike Information Criterion (AIC) (AKAIKE, 1974). Therefore, the best value for $m^{(j)}(p)$ is chosen finding the minimum of the following function

$$\mathbf{AIC}\left(m^{(j)}(p)\right) = N_m^{(j)}\left(2m^{(j)}(p) + \sum_{k=m+1}^{n_{\rm ch}}\left[\lambda_k - \log\lambda_k - 1\right]\right) \quad (5.10)$$

where $N_m^{(j)}$ is the number of modes in the (j) needlet window, $m^{(j)}(p)$ is the number of degrees of freedom used by `GNILC` to describe the foregrounds and λ_k the eigenvalues of the transformed covariance matrix presented in the Equation 5.9.

After estimating the dimension of the foregrounds subspace, a $m^{(j)}(p)$-dimensional ILC is performed for each needlet scale (j). First, the matrix $\mathbf{U}_s^{(j)}$, whose dimension is $n_{\rm ch} \times (n_{\rm ch} - m)$ is determined. Then, the mixing matrix of the target signal, $\hat{\mathbf{S}}$, is calculated through

$$\hat{\mathbf{S}}^{(j)} = \left[\hat{\mathbf{R}}_s^{(j)}\right]^{1/2}\mathbf{U}_s^{(j)}. \quad (5.11)$$

After determining the matrix $\hat{\mathbf{S}}$, the multidimensional ILC filter is applied to the data vector $\mathbf{d}^{(j)}(p)$ and make a estimate of HI signal $\hat{\mathbf{s}}^{(j)}(p)$. Omitting p and (j) to simplify the notation, the estimate of the target signal is determined by

$$\hat{\mathbf{s}} = \hat{\mathbf{S}} \left(\hat{\mathbf{S}}^T \mathbf{R}^{-1} \hat{\mathbf{S}} \right)^{-1} \hat{\mathbf{S}}^T \mathbf{R}^{-1} \mathbf{d}, \tag{5.12}$$

where the ILC weights matrix is given by

$$\mathbf{W} = \hat{\mathbf{S}} \left(\hat{\mathbf{S}}^T \mathbf{R}^{-1} \hat{\mathbf{S}} \right)^{-1} \hat{\mathbf{S}}^T \mathbf{R}^{-1}. \tag{5.13}$$

Finally, the reconstructed needlet (j) maps, $\hat{\mathbf{s}}^{(j)}(p)$, are synthesized on a single map. They are transformed to spherical harmonic space, their harmonic coefficients are again band-pass filtered by the respective needlet window, $h_\ell^{(j)}$, and the filtered harmonic coefficients are transformed back to maps in pixel space. This results in one reconstructed map per needlet scale (j), according

$$\hat{s}_\nu^{(j)}(p) = \sum_\ell \sum_{m=-\ell}^{\ell} \hat{s}_\nu^{(j)}(\ell, m)\, h_\ell^{(j)} Y_{\ell m}(p). \tag{5.14}$$

These maps are then added to give, for each frequency channel ν, the complete reconstructed map $\hat{s}_\nu(p)$

$$\hat{s}_\nu(p) = \sum_{j=1}^{n_w} \hat{s}_\nu^{(j)}(p), \tag{5.15}$$

where n_w is the number of needlets windows chosen as input. The reconstructed maps, $\hat{s}_\nu(p)$, are the HI plus thermal noise maps recovered with `GNILC`.

As seen before, `GNILC` has two input parameters that control the location both in the pixel and in the harmonic domains, used in the calculation of covariance matrices: the set of needlets and the ILC bias. We tested the efficiency of `GNILC` in reconstructing target signal maps considering different combinations of these input parameters. Thus, the best choice for our BINGO simulated data was the set of cosine-shaped needlets with peaks at $\ell = [0, 128, 384, 767]$ and the ILC bias $b = 0.005$. This was the configuration of `GNILC` input parameters used here. The following section presents the debiasing procedure used to obtain the 21 cm signal power spectrum from the

HI plus thermal noise maps reconstructed with `GNILC`.

5.3 The Debiasing procedure

Section 5.2 describes the `GNILC` method, referring to the signal of interest as the component to be recovered. As discussed in Section 5.1, due to the characteristics of the 21 cm signal and the noise, it was chosen to recover both as a single component. After obtaining the reconstructed 21 cm plus noise maps, it was used a debiasing procedure to reconstruct the pure power spectrum of the 21 cm signal. This method consists of estimating the residual noise content and the loss of the HI signal, after passing the data through the ILC filter (see Section 5.2), and correcting their effects on the power spectra of the `GNILC` output maps. This procedure was also used by Fornazier et al. (2022) and a description of the technique is presented below.

Equation 5.1 can be rewritten as

$$d_\nu(p) = s_\nu^{21\mathrm{cm}}(p) + n_\nu(p) + f_\nu(p), \tag{5.16}$$

where $s_\nu^{21\mathrm{cm}}(p)$ is the 21 cm signal, $n_\nu(p)$, the thermal noise and $f_\nu(p)$, the foreground emission. As seen in Section 5.2, the `GNILC` output maps at all frequencies are the result of the projection of N_{ch}-dimensional BINGO data, which can be written as (omitting the index (j)),

$$\hat{s}_\nu(p) = \sum_{\nu'} W_{\nu\nu'}(p) d_{\nu'}(p), \tag{5.17}$$

into a target signal subspace of lower dimension, defined by the matrix $\mathbf{U}_s$, as can be seen in the Equation 5.9. Although the `GNILC` projection matrix (see Equation 5.12) is defined so as to ensure minimal projection of the foregrounds into the target signal subspace and minimal projection of the target signal into the complementary foreground subspace, still a small part of the 21 cm signal, noise and foregrounds can project into each complementary subspace.

Formally, by inserting Equation 5.16 into Equation 5.17 and using a matrix-vector notation, the `GNILC` output maps read as

$$\hat{\mathbf{s}}(p) = \mathbf{W}(p)\mathbf{s}^{21\mathrm{cm}}(p) + \mathbf{W}(p)\mathbf{n}(p) + \mathbf{W}(p)\mathbf{f}(p), \tag{5.18}$$

which highlights *additive* errors due to projected noise, $\mathbf{W}(p)\mathbf{n}(p)$, and projected foreground residuals, $\mathbf{W}(p)\mathbf{f}(p)$. In addition, a small *multiplicative* error may arise from a possible partial loss of the 21 cm signal into the foreground subspace, i.e.

$$\mathbf{W}(p)\mathbf{s}^{21\mathrm{cm}}(p) \simeq (\mathbf{I} - \epsilon)\mathbf{s}^{21\mathrm{cm}}(p), \tag{5.19}$$

where $\mathbf{I}$ is the identity matrix, while ϵ is a positive-definite diagonal matrix which describes possible small corrections to identity. Taking the spherical harmonic transform of Equation 5.18, the angular auto power spectrum of the `GNILC` output map at a given frequency band ν thus reads as

$$C_\ell^{\mathrm{GNILC},\nu} \simeq b_\ell^\nu C_\ell^{21\mathrm{cm},\nu} + C_\ell^{\mathrm{noise-proj},\nu} + C_\ell^{\mathrm{fgds-proj},\nu}, \tag{5.20}$$

where $b_\ell^\nu \lesssim 1$ is the 21 cm power spectrum multiplicative bias for that frequency channel, while $C_\ell^{\mathrm{noise-proj},\nu}$ and $C_\ell^{\mathrm{fgds-proj},\nu}$ are power spectrum additive biases due to projected noise and foreground residuals, respectively.

Neglecting the residual foregrounds into `GNILC` output maps, the debiasing procedure to reconstruct the 21 cm power spectra can be divided into two steps:

a) Estimate the projected noise power spectra, $\hat{C}_\ell^{\mathrm{noise-proj},\nu}$, and debias the `GNILC` map power spectra from this additive bias;

b) Estimate the multiplicative bias, $\hat{b}_\ell^\nu$, and correct the noise-debiased `GNILC` power spectra from it.

These two steps can be summarized by the equation

$$\hat{C}_\ell^{21\mathrm{cm},\nu} = \frac{C_\ell^{\mathrm{GNILC},\nu} - \hat{C}_\ell^{\mathrm{noise-proj},\nu}}{\hat{b}_\ell^\nu}, \tag{5.21}$$

where $\hat{C}_\ell^{21\mathrm{cm},\nu}$ is the final estimate of the 21 cm power spectrum at the frequency channel ν.

The additive noise bias $C_\ell^{\mathrm{noise-proj},\nu}$ is estimated by generating N_{realis} white noise map realisations for each frequency channel ν and projecting them through the `GNILC` weight matrix Equation 5.13 computed in Section 5.2 for the data. The power spectra of the resulting projected noise realisations are then averaged over all realisations.

The multiplicative bias b_ℓ^ν is estimated by generating N_{realis} realisations of 21 cm signal maps at all frequency channels ν and computing the projected 21 cm signal by applying again the `GNILC` weight matrix given by Equation 5.13 to the pure 21 cm map realisations. For each frequency channel ν, the ratios between the power spectra of the projected 21 cm realisations and the power spectra of the input 21 cm realisations, which we average over all realisations in order to estimate b_ℓ^ν, are then computed. The accuracy of the estimation of additive and multiplicative biases depends directly on the number of realizations N_{realis} used. However, the choice of this number is not free, but limited by the available computational capacity. The greater the number of channels adopted, the longer the debiasing processing time. Thus, considering the available computing resources and the BINGO simulated data configurations, the debiasing procedure was tested with two different N_{realis}, as can be seen in Chapter 6. The preliminary results of the 21 cm signal recovery from the BINGO simulated observations, using the foregrounds removal pipeline presented in this chapter, are presented below.

6 RESULTS

As seen in the previous chapter, the component separation pipeline used in this work can be divided in two steps: the foreground removal stage, where `GNILC` method, described in the Section 5.2, is used to recover the 21 cm plus thermal noise maps from the BINGO simulated data; and the debiasing stage, where the procedure described in the Section 5.3 is used to obtain the 21 cm power spectra from the `GNILC` output maps.

Our BINGO simulated data maps are the sum of astrophysical foregrounds, HI and thermal noise, as described in Section 4.1. The reference case, adopted for comparison of results, is the MD30, the BINGO project baseline configuration. In the simulations that include an independent observation, it was added a channel with the CBASS 5 GHz map, containing foregrounds and thermal noise, to the set of BINGO simulated data.

The angular power spectra were plotted in the range of multipoles $30 \leq \ell \leq 270$, which is equivalent to angular scales $\sim 6^\circ$ to $\sim 0.7^\circ$. Very large angular scales were not considered, $\ell < 30$, due to the area of the sky covered by BINGO, and small angular scales, $\ell > 270$, due to the angular resolution of the instrument, 40 arcmin. Furthermore, as BINGO covers a fraction of the sky, the power spectra of its maps suffer from a loss of angular resolution. Thus, the choice of bin size must respect the restriction $\Delta\ell \sim 180^\circ/\gamma_{\max}$, where $\gamma_{\max}$ is the maximum extent of the observed area (ANSARI; MAGNEVILLE, 2010). In the case of BINGO, considering the repositioning of the horns during the mission, as described in Wuensche et al. (2022) and Liccardo et al. (2022), $\gamma_{\max} = 17.5^\circ$. Thus, in order to respect this limitation and better adapt to the range of multipoles adopted in this work ($30 \leq \ell \leq 270$), we chose bin the angular power spectra with multipole resolution of $\Delta\ell = 12$.

The results obtained in the recovery stages of the 21 cm plus thermal noise maps (`GNILC` step) and the 21 cm power spectra (debiasing step), considering the simulation plan in the Table 4.2, are presented in the following sections.

6.1 21 cm plus thermal noise maps reconstruction

To quantify the maps reconstruction efficiency, the Pearson coefficient was used, defined as

$$\rho_\nu = \frac{\sum_p (\hat{s}_\nu(p) - \mu_{\hat{s}_\nu})(s_\nu(p) - \mu_{s_\nu})}{\sqrt{\sum_p (\hat{s}_\nu(p) - \mu_{\hat{s}_\nu})^2 \sum_p (s_\nu(p) - \mu_{s_\nu})^2}} \ , \tag{6.1}$$

where $\hat{s}_\nu(p)$ and $s_\nu(p)$ are the reconstructed and expected HI plus noise maps, respectively, in a given pixel p and frequency channel ν. The $\mu_{\hat{s}_\nu}$ and μ_{s_ν} terms are the mean value over the pixels of the reconstructed and expected HI plus noise maps, respectively. A perfect reconstruction of the target signal maps occurs when $\rho_\nu = 1$. The results in the pixel domain of the tests proposed in Section 4.2 are presented below.

Testing different synchrotron models

The first test proposed is to check the `GNILC` response against different synchrotron models, the dominant foreground component in the BINGO frequencies. For that, it was performed the component separation in the MD30 and GD30 simulated data (see Table 4.2).

The Pearson coefficients between the input and reconstructed HI plus thermal noise maps are presented in Figure 6.1. It can be seen that the correlation between the expected and reconstructed maps varies very little between the two configurations. Furthermore, the average Pearson coefficient over all channels is the same for both cases, $\bar{\rho} = 0.853$, which shows that `GNILC` performs well against different synchrotron models, in an analysis in the pixel domain.

Figures 6.2 and 6.3 show the input (expected), the reconstructed and the residual 21 cm plus noise maps for MD30 and GD30 configurations, observed near to the central frequency of the BINGO band (1120 MHz). Note the effect of the apodized galactic mask on the edges of the apparent region of the maps.

Figure 6.1 - Pearson coefficients calculated for each pair of expected and reconstructed 21 cm plus thermal noise maps, in each frequency bin of the MD30 and GD30 input data. The average Pearson coefficient over all channels is the same for both cases, $\bar{\rho} = 0.853$.

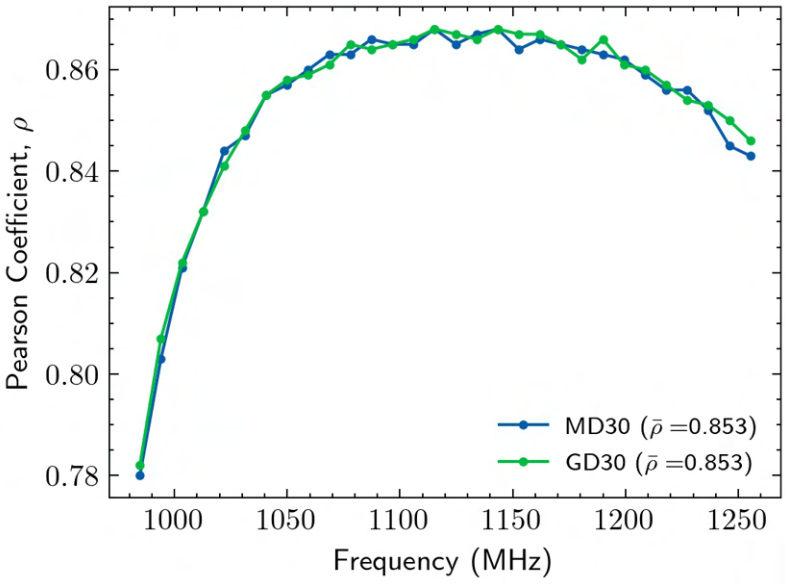


Figure 6.2 - Simulated 21 cm plus thermal noise input map (*Top*), `GNILC` reconstructed map (*Middle*) and respective residuals, the difference between the two previous maps (*Bottom*), corresponding to a channel centered at 1115 MHz in MD30 configuration. The maps are in celestial coordinates and are covered by the apodized galactic mask defined in Section 4.1.3. The HI component was convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam.

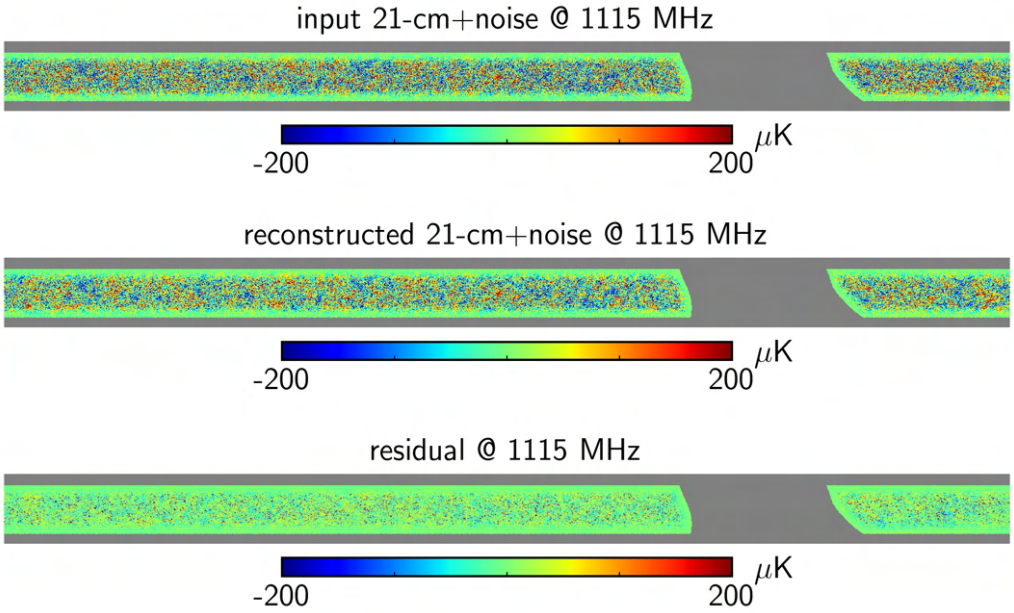
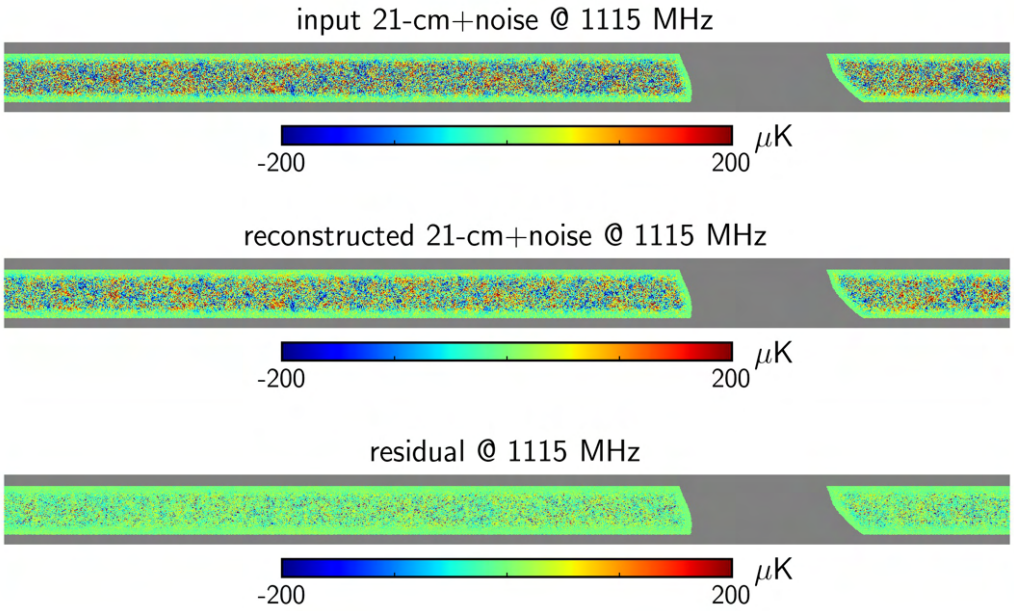


Figure 6.3 - Simulated 21 cm plus thermal noise input map (*Top*), `GNILC` reconstructed map (*Middle*) and respective residuals, the difference between the two previous maps (*Bottom*), corresponding to a channel centered at 1115 MHz in GD30 configuration. The maps are in celestial coordinates and are covered by the apodized galactic mask defined in Section 4.1.3. The HI component was convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam.



Figures 6.4 and 6.5 show the total foreground emission and the respective residual map (projection of foregrounds into the target signal subspace) after component separation with `GNILC`. The maps show the effectiveness of `GNILC` in removing contaminants, since the content of foregrounds in the data is reduced from about 10^5 to tens of μK.

Figure 6.4 - Map containing the sum of all foregrounds described in Section 4.2 (*Top*) and their respective residuals after `GNILC` (*Bottom*) at 1115 MHz in MD30 configuration (see Table 4.2). The maps are in celestial coordinates and are covered with the apodized galactic mask defined in the Section 4.1.3. The foregrounds were convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam. Note the region of extreme declinations where the signal is attenuated by the apodized mask.

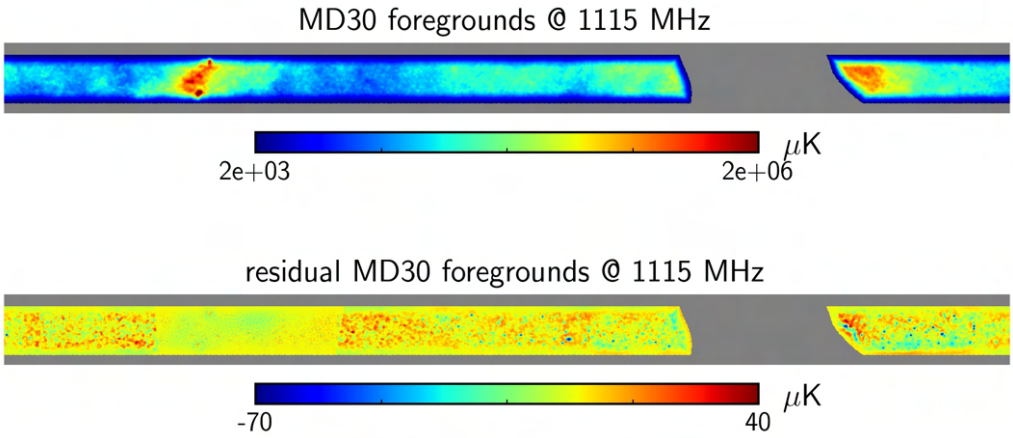
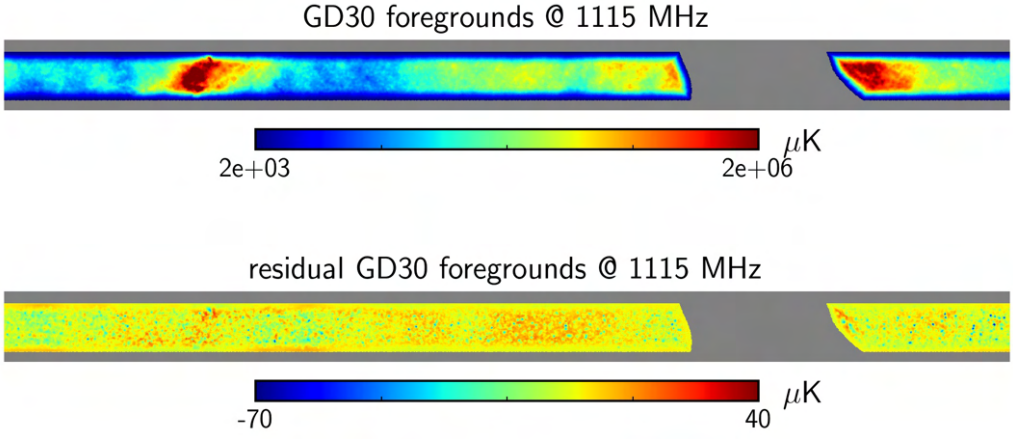


Figure 6.5 - Map containing the sum of all foregrounds described in Section 4.2 (*Top*) and their respective residuals after `GNILC` (*Bottom*) at 1115 MHz in GD30 configuration (see Table 4.2). The maps are in celestial coordinates and are covered with the apodized galactic mask defined in the Section 4.1.3. The foregrounds were convolved with a $\theta_{\mathrm{FWHM}} = 40$ arcmin beam. Note the region of extreme declinations where the signal is attenuated by the apodized mask.



In the MD30 foregrounds projection map it is observed that the `GNILC` removed a

little more foregrounds from the unmasked region of the sky observed by BINGO where the galactic emission is more intense. This can be explained by the maps in Figure 6.6, which contain the number of degrees of freedom $m^{(j)}(p)$ assigned by `GNILC` to describe the foregrounds at each pixel p and each range of angular scales (j), defined by the needlets filters. The map referring to the angular scales delimited by $0 \leq \ell \leq 384$, $m^{(1)}(p)$, shows that `GNILC` dedicates 3 degrees of freedom to describe and remove foregrounds in the unmasked galactic region observed by BINGO, where the synchrotron emission is more intense. Outside this region, the dimension of the foregrounds subspace is equal to 2. Thus, contaminant removal is slightly more effective in the residual map region of Figure 6.4, discussed earlier. However, the choice of the dimension of the foregrounds subspace, $m^{(j)}(p)$, may reflect on the efficiency of the reconstruction of the HI plus noise signal, since the number of degrees of freedom available to reconstruct the target signal is limited by $N_{\rm ch} - m^{(j)}(p)$. As seen in Chapter 5, the best choice for $m^{(j)}(p)$ is made calculating the minimum of the AIC function given by the Equation 5.10, which considers the compromise between estimating the number of dimensions used to best describe the subspace of the signal of interest and the foregrounds subspace.

Figure 6.6 - Maps containing the number of degrees of freedom dedicated to describe the MD30 foregrounds, $m^{(j)}(p)$, per pixel p. Each map corresponds to a range of angular scales (j), isolated by each needlet window from the set defined as input to the GNILC (see Section 5.2). The cosine-shaped needlets are centered at $[0, 128, 384, 767]$ and the angular scale intervals are described in the title of the maps. Note that the map $m^{(0)}(p)$ is associated with larger angular scales and therefore requires a lower resolution ($N_{\text{side}} = 64$). Furthermore, `GNILC` assigns 3 dimensions to the foregrounds subspace at angular scales between $0 \leq \ell \leq 128$ ($j = 0$), as well as to the galactic region of the map corresponding to the range $0 \leq \ell \leq 384$ ($j = 1$). For the maps generated with the other needlet windows ($j = 2$ and $j = 3$), `GNILC` dedicates 2 degrees of freedom to describe the foregrounds.

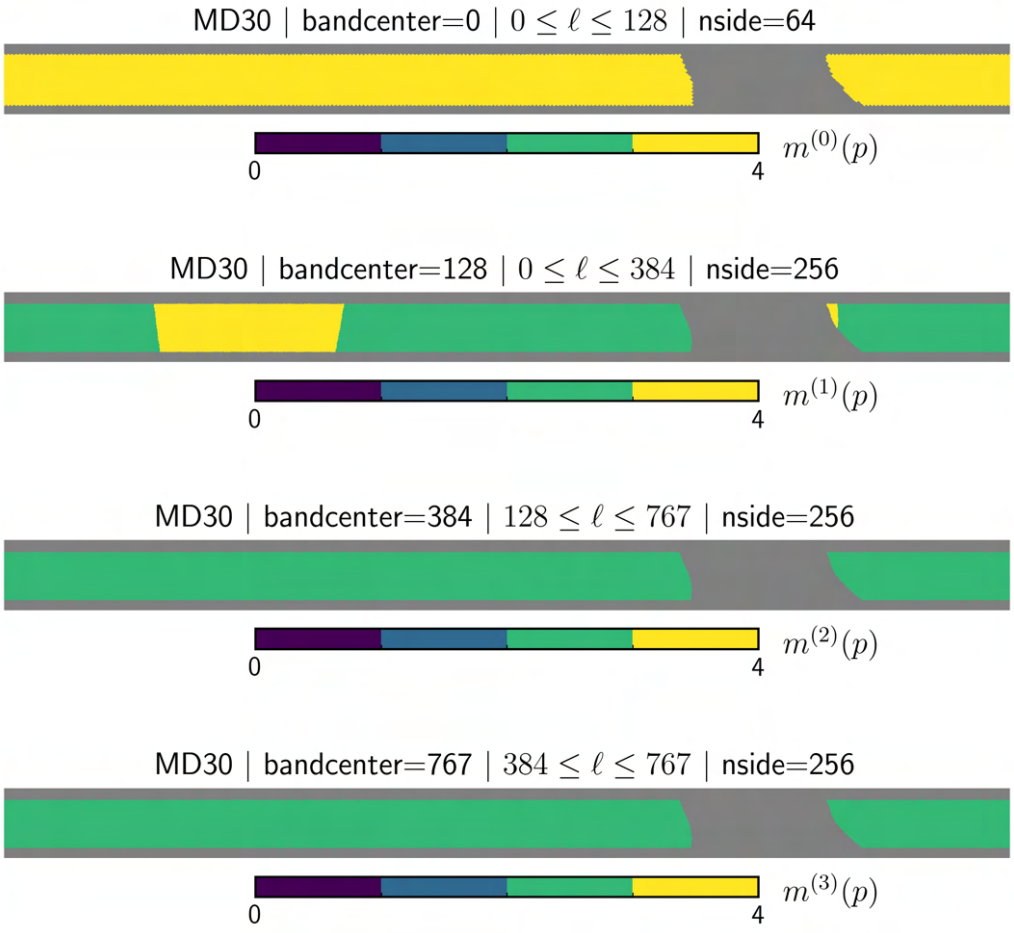


Figure 6.7 - Maps containing the number of degrees of freedom dedicated to describe the GD30 foregrounds, $m^{(j)}(p)$, per pixel p. Each map corresponds to a range of angular scales (j), isolated by each needlet window from the set defined as input to the GNILC (see Section 5.2). The cosine-shaped needlets are centered at $[0, 128, 384, 767]$ and the angular scale intervals are described in the title of the maps. Note that the map $m^{(0)}(p)$ is associated with larger angular scales and therefore requires a lower resolution ($N_{\text{side}} = 64$). Furthermore, `GNILC` assigns 3 dimensions to the foregrounds subspace at angular scales between $0 \leq \ell \leq 128$ ($j = 0$), and 2 dimensions for the maps generated with the other needlet windows ($j = 1$, $j = 2$ and $j = 3$).

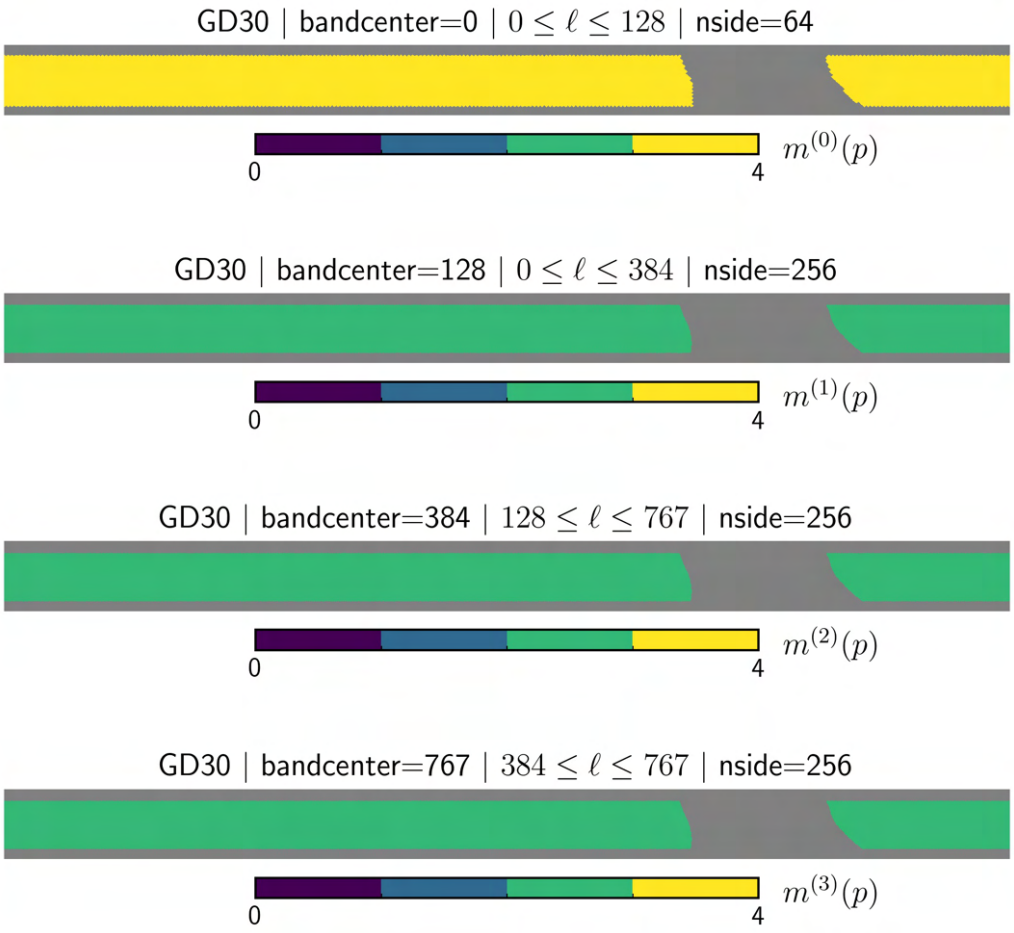


Figure 6.8 shows the comparison between the real and reconstructed power spectra of HI plus noise, as well as the residual foregrounds after `GNILC`, plotted for 3 different frequency channels. As with the Pearson coefficients in Figure 6.1, the power spectra show a worse reconstruction of the HI plus thermal noise signal at lower frequencies. This is because the synchrotron emission increases with the decrease in the observed frequency (see Figure 1.1), making it difficult to remove foregrounds. This reduction

in the 21 cm plus thermal noise signal reconstruction efficiency at low frequencies can also be observed in the pixel domain, as shown in Figure 6.1.

Figure 6.8 - Input 21 cm plus noise (black), reconstructed 21 cm plus noise (dashed green) and residual foregrounds (dashed red) power spectra, obtained from simulated data with MD30 and GD30 configurations, for frequency channels centered at 985 MHz (*left*), 1115 MHz (*center*) and 1255 MHz (*right*). The multipole range considered is $30 \leq \ell \leq 270$, with a multipole bin size of $\Delta\ell = 12$.

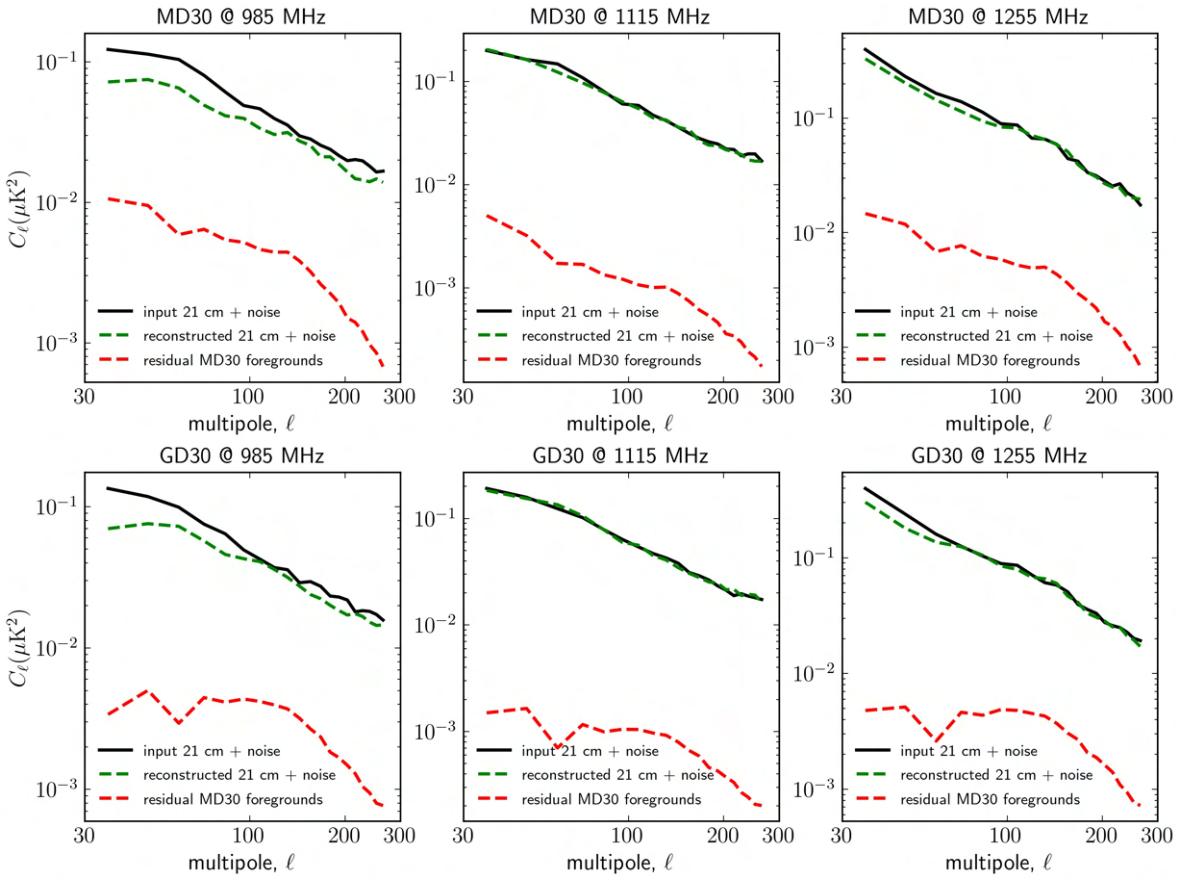
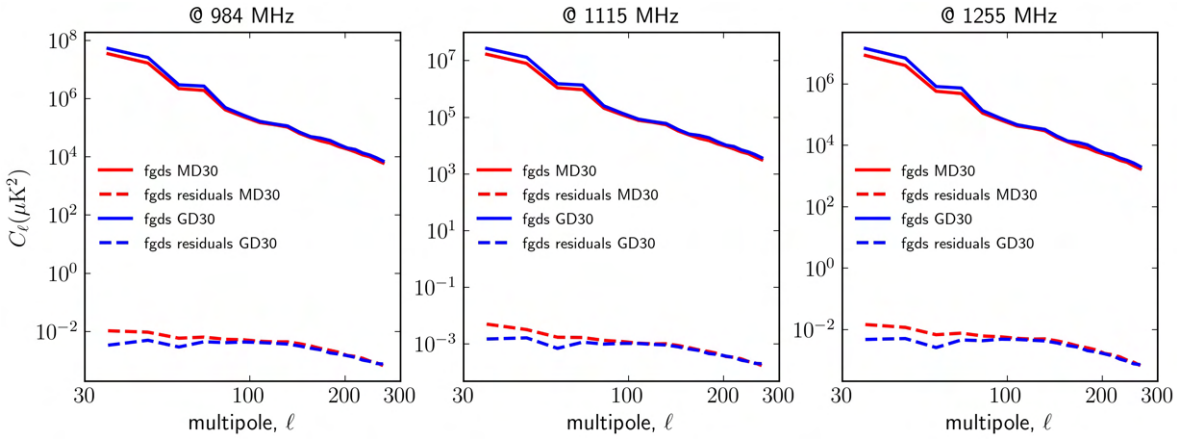


Figure 6.9 shows the MD30 and GD30 foregrounds power spectra and their respective residuals after component separation with `GNILC`. It is possible to notice that at scales $\ell < 100$, `GNILC` removes more foregrounds from the GD30 configuration than from the MD30. Furthermore, this residuals are preserved after the debiasing step, according to the Equations 5.20 and 5.21, contributing to the 21 cm signal power spectrum estimation error.

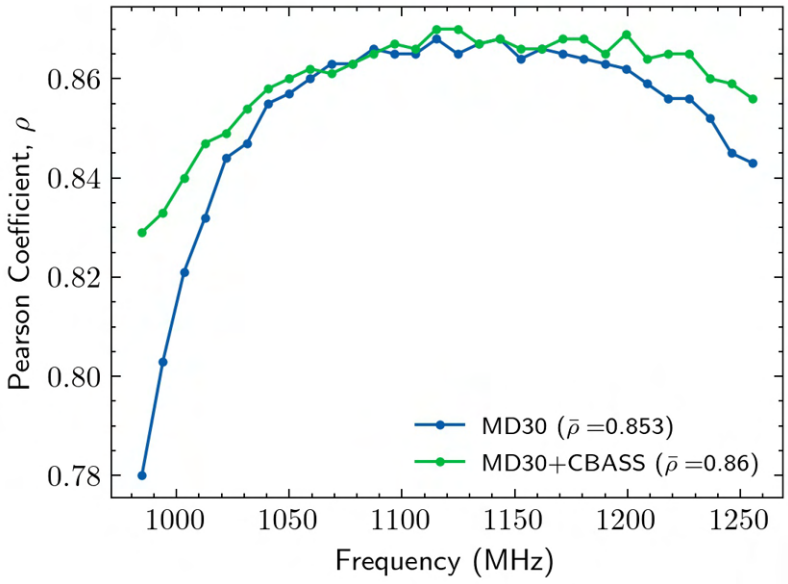
Figure 6.9 - MD30 (red) and GD30 (blue) foregrounds power spectra, as well as their respective residuals present in the signal reconstructed with the GNILC (dashed red and dashed blue). The power spectra are plotted for frequency channels centered at 985 MHz (*left*), 1115 MHz (*center*) and 1255 MHz (*right*). The multipole range considered is $30 \leq \ell \leq 270$, with a multipole bin size of $\Delta\ell = 12$.



Testing the addition of an independent observation

As seen in Section 4.2, to verify the effect of adding an extra channel to the MD30 configuration, with the CBASS simulated data, two cases were compared: MD30 and MD30+CBASS. The Pearson coefficients between the GNILC output maps and the expected HI plus noise maps are plotted in Figure 6.10. As expected, it can be noted an improvement in the reconstruction of HI plus noise maps, when a channel with an independent observation (CBASS 5 GHz) is added to the GNILC input data. The average Pearson coefficient calculated across all channels is $\bar{\rho} = 0.853$ without CBASS map while is $\bar{\rho} = 0.860$ with the addition of the extra channel. Furthermore, the Pearson coefficient shows a more significant improvement in the extreme band frequencies.

Figure 6.10 - Pearson coefficients calculated for each pair of expected and reconstructed 21 cm plus thermal noise maps, in each frequency bin of the sets MD30 and MD30+CBASS (see Table 4.2). The average Pearson coefficient over all channels is $\bar{\rho} = 0.853$, for the MD30 case, and $\bar{\rho} = 0.86$, for the MD30+CBASS case.



Testing different numbers of channels

The accuracy with which `GNILC` reconstructs the HI plus noise maps depends mainly on the number of frequency channels adopted. Theoretically, as the number of frequency bins is increased, more efficient `GNILC` is in removing astrophysical foregrounds from the data, with the counterpart of the increase in the noise level. The reason is that `GNILC` is able to adapt the number of degrees of freedom $m^{(j)}(p)$ (see Section 5.2), dedicated to describing the foregrounds in the angular scales (j) and the pixel p, to the number of channels in which the data is binned, optimizing the target signal reconstruction (OLIVARI et al., 2016). Therefore, the efficiency of the `GNILC` in reconstructing 21 cm plus noise maps was tested considering 20, 30, 40, 60 and 80 channels. For that, the datasets in the MD20, MD30, MD40, MD60 and MD80 configurations (see Table 4.2) were used. The Pearson coefficients for data configurations with different numbers of channels are shown in Figure 6.11. It is possible to observe that, in general, the Pearson coefficients values increase with the number of channels. However, the points seem to tend towards a limit. The difference in reconstruction efficiency improvement seems to decrease with increasing number of bins. Between the MD60 and MD80 configurations, this difference is

already very small, indicating that there is a near optimal choice of the number of channels, which in our case appears to be around $N_{\rm ch} = 80$. The average Pearson coefficient, calculated over all channels, is also directly related to the number of frequency bins. The calculated values range from $\bar{\rho} = 0.835$, in the MD20 configuration, to $\bar{\rho} = 0.872$, in the MD80 case. As expected, an increase in the number of bins gives the method more degrees of freedom to describe and remove the foregrounds, improving the recovery of the HI plus noise signal. However, there seems to be a limit in the number of channels, above which there is negligible improvement in the reconstruction of the signal of interest.

Figure 6.11 - Pearson coefficients calculated for each pair of expected and reconstructed HI plus thermal noise maps, in each frequency bin of the sets MD20, MD30, MD40, MD60 and MD80. It is possible to notice the reconstruction error reduction with the increase in the number of channels, especially at lowest and highest frequencies.

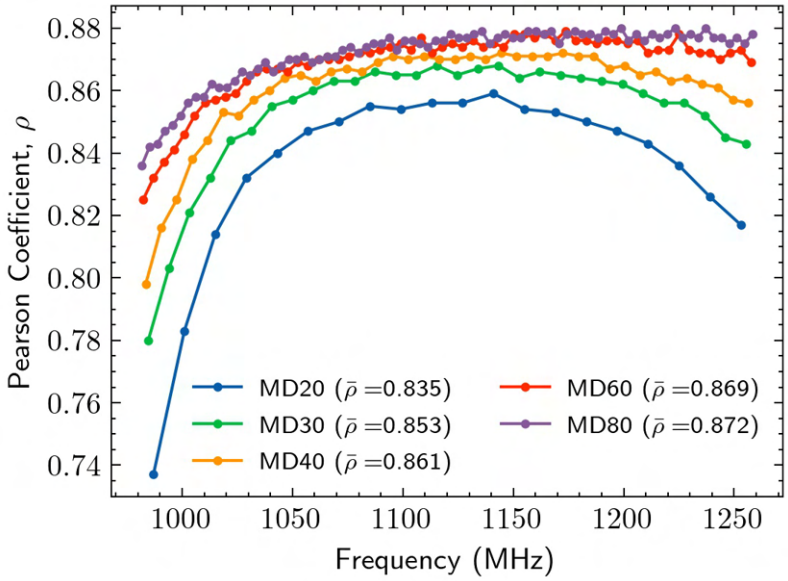
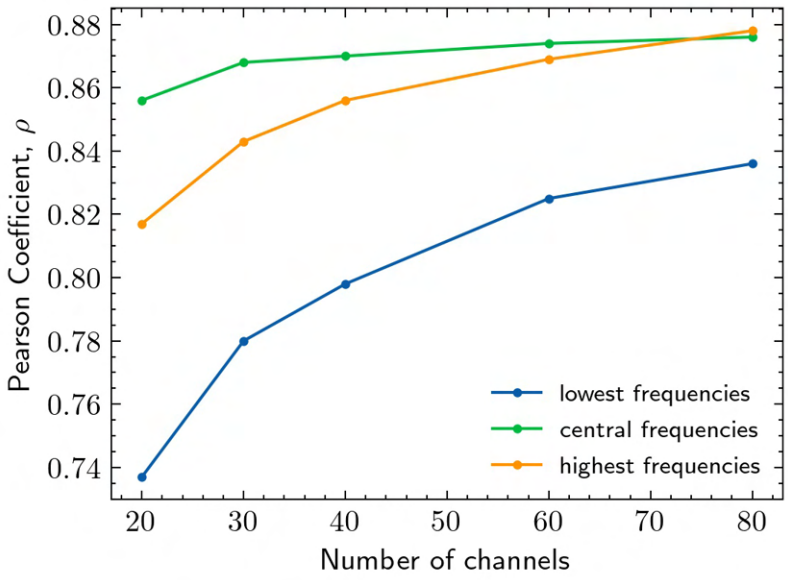


Figure 6.12 shows the Pearson coefficient as a function of the number of channels, calculated in channels centered at lower, central and higher frequencies, within the BINGO band (980-1260 MHz). In this figure it is possible to see better the stabilization trend in the efficiency of the reconstruction of the 21 cm plus noise maps with the increase in the number of channels. This effect is more evident in the central frequency channels, where the Pearson coefficients are already higher than in the other bins. This result reinforces what was said earlier regarding obtaining a near optimal reconstruction in the MD80 configuration.

Figure 6.12 - Pearson coefficients calculated for each pair of expected and reconstructed HI plus thermal noise maps. In this figure, the parameter is calculated for simulated data with different number of channels (MD20, MD30, MD40, MD60 and MD80 configurations), in channels centered at three different frequencies: lowest frequencies (channels centered at 987, 984.7, 983.5, 982.3 and 981.7 MHz), central frequencies (channels centered at 1113, 1115.3, 1116.5, 1117.7 and 1118.2 MHz) and highest frequencies (channels centered at 1253, 1255.3, 1256.5, 1257.7 and 1258.2 MHz).



The performance of `GNILC` in the reconstruction of 21 cm plus thermal noise maps, considering different BINGO simulated data configurations, was presented in this section. The results on recovery the 21 cm power spectra from the `GNILC` output maps, using the debiasing method described in the Section 5.3, are described in the following section.

6.2 Noiseless 21 cm power spectrum reconstruction

The results of using the debiasing procedure to reconstruct the 21 cm power spectra are presented in this section. The power spectra of the `GNILC` output maps, $C_\ell^{\mathrm{GNILC},\nu}$, obtained in the previous section, are used here as one of the inputs for the bias correction method, according to Equation 5.21. To measure the reconstruction error of the 21 cm power spectrum, the $\eta(\nu)$ index was used, defined as the absolute mean difference between the recovered $C_\ell^{21\mathrm{cm,r}}(\nu)$ and real $C_\ell^{21\mathrm{cm,s}}(\nu)$, normalized by the real power spectrum, according to

$$\eta(\nu) = \frac{1}{N_\ell} \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \left| \frac{C_\ell^{21\mathrm{cm,r}}(\nu) - C_\ell^{21\mathrm{cm,s}}(\nu)}{C_\ell^{21\mathrm{cm,s}}(\nu)} \right| \times 100\% \ . \tag{6.2}$$

This index is calculated for each frequency channel ν and for a multipole range $30 \leq \ell \leq 270$, as defined in Section 6.1.

To perform the debiasing for all cases proposed in the Table 4.2 in a reasonable time, it was chosen to work with a low number of realizations. Therefore, a base number $N_{\mathrm{realis}} = 10$ was adopted to correct the bias of the spectra obtained for all data configurations proposed in this work. Then, a test increasing this value to $N_{\mathrm{realis}} = 50$ was performed using the configuration that gave the best result.

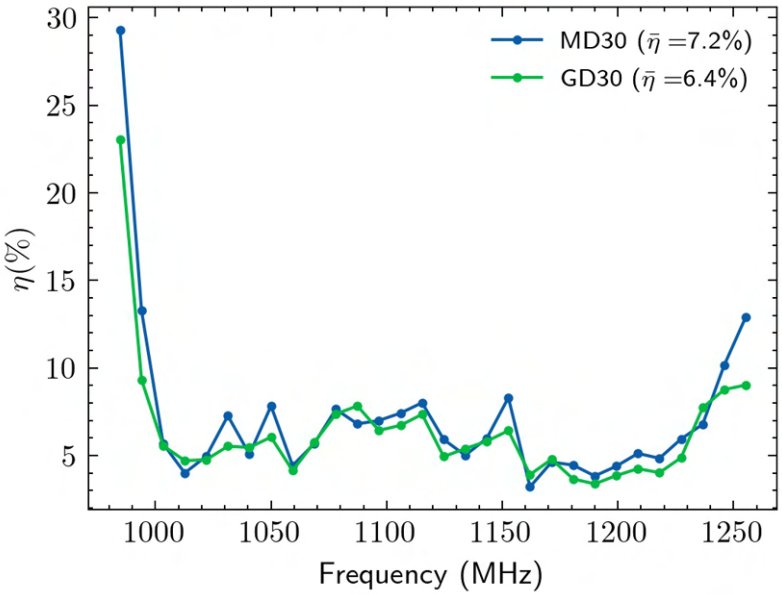
Testing different synchrotron models

The first test on debiasing the `GNILC` output power spectra was performed with input data considering different synchrotron models. Figure 6.13 shows the error η of power spectrum reconstruction of the 21 cm signal in each frequency channel of the MD30 and GD30 input data. The mean reconstruction error for the MD30 case is 7.2% and for the GD30 case is 6.4%. This difference is due to the ability of `GNILC` to adapt to the spatial variation of foregrounds to describe and remove them (see Section 5.2). Thus, maps generated with different models (MD and GD) may require different numbers of degrees of freedom to represent the contaminants in a given region of the sky, as well as to reconstruct the target signal, as seen in the Figures 6.4, 6.5, 6.6 and 6.7.

Testing the addition of an independent observation

The results of the debiasing procedure, considering the addition of a channel with simulated CBASS 5 GHz map to the BINGO data, in its basic project configuration (MD30 case), are presented below. Figure 6.14 shows the error η of power spectrum reconstruction of the 21 cm signal in each frequency channel. The reconstruction error for the MD30 case is 7.2% and for the MD30+CBASS case is 5.7%. As expected, the inclusion of an extra channel with CBASS data increased the number of degrees of freedom available to describe the foregrounds, improving the reconstruction of the 21 cm signal.

Figure 6.13 - Reconstruction error η of the 21 cm signal power spectrum, calculated as an average over the multipole range $30 \leq \ell \leq 270$, on each frequency channel of the data sets with different synchrotron models (MD30 and GD30). The value $\bar{\eta}$ in parentheses is the average error over all channels in each dataset.



Testing different numbers of channels

Figure 6.15 shows the reconstruction error η of the 21 cm signal in the harmonic domain, in each frequency channel, considering simulated data with different numbers of channels: MD20, MD30, MD40, MD60 and MD80. An improvement in the cosmological signal reconstruction can be observed with the increase in the number of channels because the number of dimensions available to describe the components increases, as discussed in Sections 5.2. The smallest average error over all channels is $\bar{\eta} = 4.7\%$, obtained with 80 frequency channels.

Figure 6.14 - Reconstruction error η of the 21 cm signal power spectrum, calculated as an average over the multipole range $30 \leq \ell \leq 270$, on each frequency channel of the data sets MD30 and MD30+CBASS (addition of an independent observation). The value $\bar{\eta}$ in parentheses is the average error over all channels in each data set.

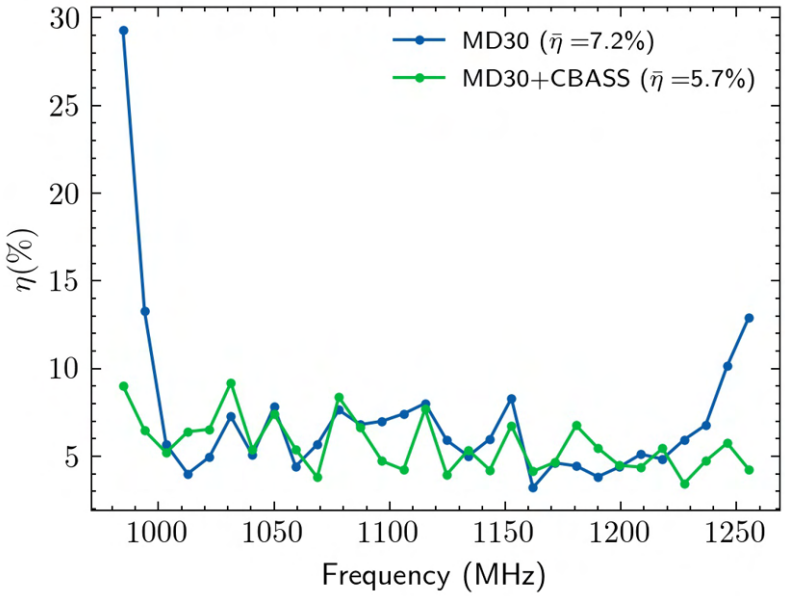


Figure 6.15 - Reconstruction error η of the 21 cm signal power spectrum, calculated as an average over the multipole range $30 \leq \ell \leq 270$, on each frequency channel of the data sets MD20, MD30, MD40, MD60 and MD80 (different numbers of frequency bins). The value $\bar{\eta}$ in parentheses is the average error over all channels in each data set.

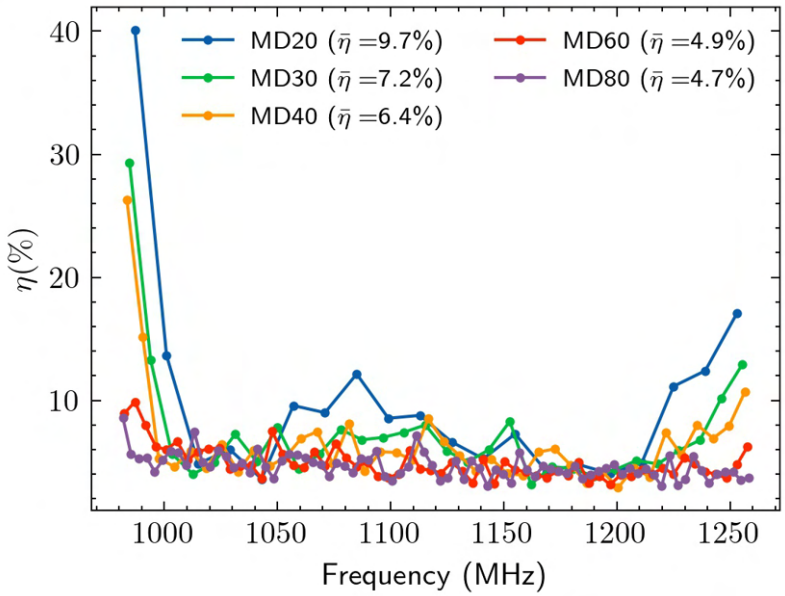
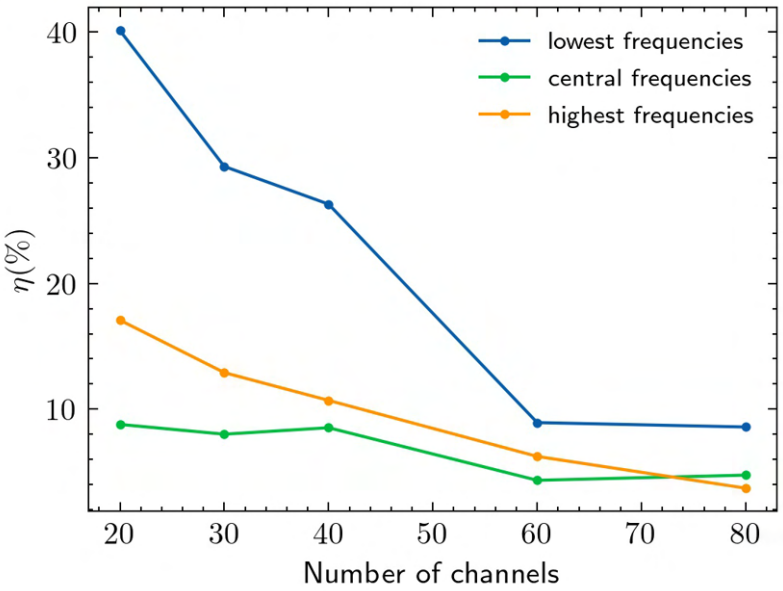


Figure 6.16 shows the power spectrum reconstruction error, calculated in three different channels, corresponding to the lowest, central and highest frequencies, within the BINGO band (980-1260 MHz). In this figure, it is possible to notice a reduction in the estimation error with the increase in the number of channels, in addition to a stabilization trend of this reduction. This effect is more evident in the lowest frequency channels, where the estimation error are already higher for smaller numbers of channels. This result corroborates what was presented in Section 6.1, regarding obtaining a near optimal reconstruction with 80 frequency bins.

Figure 6.16 - 21 cm power spectrum reconstruction error, calculated as an average over multipole range $30 \leq \ell \leq 270$, for simulated data with different number of channels (MD20, MD30, MD40, MD60 and MD80 configurations), in channels centered at three different frequency channels: lowest frequencies (channels centered at 987, 984.7, 983.5, 982.3 and 981.7 MHz), central frequencies (channels centered at 1113, 1115.3, 1116.5, 1117.7 and 1118.2 MHz) and highest frequencies (channels centered at 1253, 1255.3, 1256.5, 1257.7 and 1258.2 MHz).

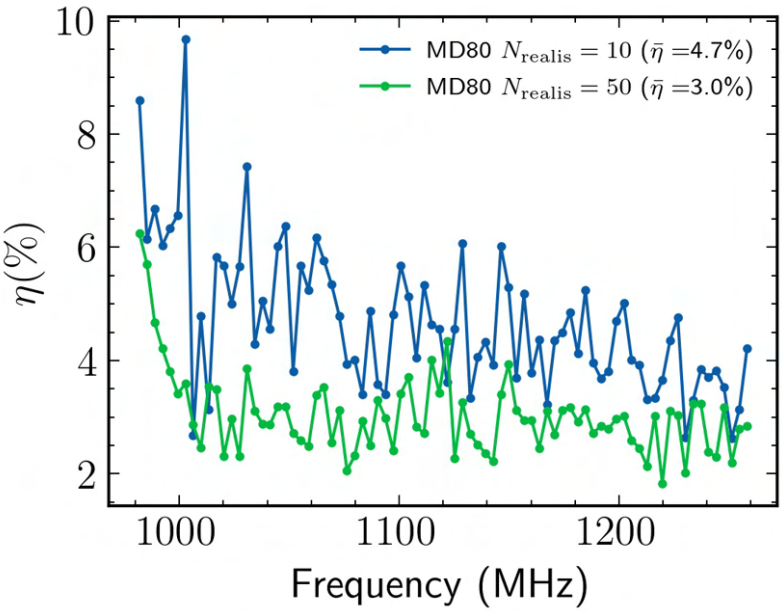


Testing different numbers of realisations N_{realis}

To evaluate the effect of increase the number of realisations on the 21 cm spectrum reconstruction efficiency, the debiasing procedure was reprocessed with $N_{\text{realis}} = 50$. The `GNILC` output power spectra used in this test was the one obtained from

the MD80 data configuration, the case with the best results in the pixel space analysis. Then, $N_{\text{realis}} = 10$ and $N_{\text{realis}} = 50$ results were compared, as presented in the Figure 6.17. The 21 cm signal reconstruction error for the case MD80 with $N_{\text{realis}} = 10$ is 4.7% and with $N_{\text{realis}} = 50$ is 3.0%. As expected, the efficiency of the debiasing procedure increases when using a greater number of HI and thermal noise realisations. The reason is that with more realisations, the method improves the estimation of the reconstructed spectra bias. The objective of this test is not to optimize the number of realizations, which would require much more computational processing time, but to perform a sensitivity analysis of the debiasing method with this parameter.

Figure 6.17 - Reconstruction error η of the 21 cm signal power spectrum, calculated as an average over a multipole range, defined here as $30 \leq \ell \leq 270$, on each frequency channel of a dataset with MD80, considering two different number of realisations ($N_{\text{realis}} = 10$ and $N_{\text{realis}} = 50$). The value $\bar{\eta}$ in parentheses is the average error over all channels in each dataset. The mean error for the $N_{\text{realis}} = 10$ case is $\bar{\eta} = 4.7\%$ and for the $N_{\text{realis}} = 50$ case is $\bar{\eta} = 3.0\%$.



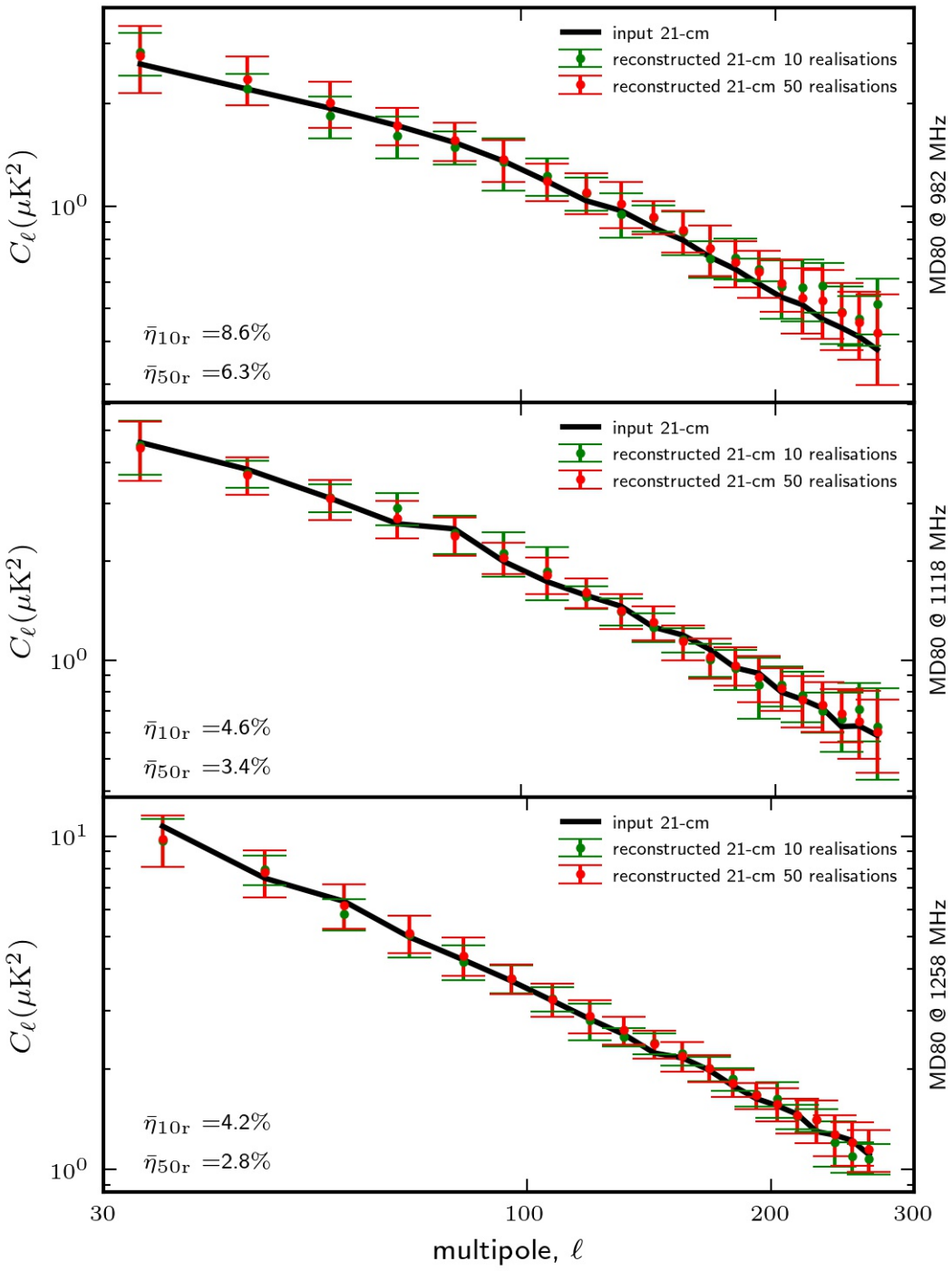
Final estimate of the 21 cm power spectrum

To make the final estimate of the 21 cm power spectrum, the entire component separation procedure presented so far was repeated for N_{realis} times, considering different N_{realis} realizations of the `GNILC` input maps (different realizations of HI and

noise plus foregrounds). Each `GNILC` output power spectra debiasing was performed using N_{realis} independent realizations of HI and noise. Finally, it is calculated the mean and the standard deviation (error bars) of the debiased power spectra. This entire procedure was performed for $N_{\text{realis}} = 10$ and $N_{\text{realis}} = 50$, using the data in the MD80 configuration. Figure 6.18 shows the 21 cm all-sky real and estimated debeamed power spectra in three different frequency channels. It can be observed that the spectrum relative to $N_{\text{realis}} = 50$ is closer to the real 21 cm sky power spectrum in all the frequencies presented. The best result is obtained for $N_{\text{realis}} = 50$ at 1258 MHz, where the average power spectrum reconstruction error is $\bar{\eta}_{50\text{r}} = 2.8\%$.

It is expected that a better estimate of the 21 cm power spectrum can be made with a larger N_{realis}. However, due to the available computational capacity, the number of realizations was limited to $N_{\text{realis}} = 50$. The complete set of simulations for the MD80 data set took two and a half months to be ready. For this it was used 56 cores of the processor Intel Xeon Gold 5120 2.20 GHz and 512 GB of RAM.

Figure 6.18 - Real (black) and reconstructed (green for $N_{\rm realis} = 10$ and red for $N_{\rm realis} = 50$) 21 cm signal all-sky power spectra, with respective error bars for the MD80 case (see Table 4.2). The power spectra plotted refers to the channels centerd at 982 MHz (*Top*), 1118 MHz (*Center*) and 1258 MHz (*Bottom*).



7 CONCLUSIONS AND NEXT STEPS

As discussed in the Chapter 5, the component separation procedure used in this work is divided in two steps: first, the HI + thermal noise signal, recovered by applying `GNILC` to the BINGO simulated data; second, the noiseless 21 cm power spectra, reconstructed passing the first step results through a debiasing procedure. The analysis presented in this work uses as default MD30, the BINGO project baseline configuration, and $N_{\text{realis}} = 10$, which is the number of HI and noise realisations used in bias correction step.

This thesis proposal presents two new results relevant to the BINGO data analysis and hardware configuration. The first result points that `GNILC` has a good performance against different foregrounds models. For this, the efficiency of the cosmological signal recovery process was tested considering simulated observations with two different synchrotron models, the dominant component in the BINGO frequency band. The results show that the reconstructed HI + thermal noise maps in both cases present no significant differences in the pixel domain, except for a small difference in residual foreground in the region where the galactic plane crosses the unmasked area observed by BINGO. In the harmonic domain, the average reconstruction error of the HI power spectra calculated over all frequency channels was 7.2%, for the MD30 case, and 6.4% for the GD30 case, in the interval $30 \leq \ell \leq 270$. The difference between foreground residuals of the two models is evident for $\ell < 100$, which contributes for the observed difference in the harmonic domain results. This is due to the `GNILC` ability to adapt the foregrounds removal not only to different regions of the sky, but also to different angular scales.

The second important result is a near optimal reconstruction of the HI signal in a BINGO 80-bins configuration. In this case, comparing the component separation results with simulated data binned into different numbers of frequency channels produced a mean error in the power spectrum reconstruction of 3% in the multipole interval $30 \leq \ell \leq 270$. The foreground removal results indicated a stabilization trend in the reduction of the 21 cm signal estimation error with the increase in the number of channels. More precisely, the recovery quality of the HI signal seems to start to stabilize at $N_{\text{ch}} = 60$ to $N_{\text{ch}} = 80$, suggesting that it would not be worth testing the component separation process with more redshift bins. The definition of the optimal number of frequency (redshift) bins will have a significant impact in the hardware configuration, helping to define the number of binning channels to preserve information from the raw data, when using a sampling mode < 1 ms.

This work confirms that adding an extra channel, with an independent observation (CBASS 5 GHz simulated sky), to the BINGO map set, improves the 21 cm power spectrum reconstruction, as expected. Considering the MD30 input data configuration, the mean error in the power spectrum reconstruction decreases from 7.2%, without CBASS information, to 5.7% when the CBASS map is added to the 30-bins BINGO simulated data.

The effect of the increase in the number of realizations of HI and noise, used in the debiasing procedure, on the quality of the 21 cm signal estimate was also tested. The mean power spectrum reconstruction error, calculated over all channels, obtained with $N_{\text{realis}} = 10$ is $\bar{\eta} = 4.7\%$ and with $N_{\text{realis}} = 50$ is $\bar{\eta} = 3.0\%$. These results indicate that a larger number of HI and noise realizations allows a better estimation and correction of additive and multiplicative biases of the `GNILC` output power spectra, also as expected. For an optimal choice of N_{realis} a more detailed analysis is necessary, with longer computational processing time.

Finally, an estimate of the reconstructed 21 cm spectra is obtained repeating the entire component separation process for 10 and 50 different realisations of the BINGO simulated input data (HI + thermal noise + foregrounds), produced in MD80 configuration. The results were compared with the power spectrum of an independent 21 cm full-sky realization, considered here as the real map of HI. The estimated spectra are in good agreement with the independent realizations. As expected, it is possible to obtain a more accurate reconstruction of the HI signal with 50 realisations, compared to 10 realisations. The mean reconstruction error for the MD80 50 realisations case, calculated in the range $30 \leq \ell \leq 270$, is equal to 6.3%, 3.4% and 2.8% for channels centered at 982 MHz, 1118 MHz and 1258 MHz.

The results obtained in this work are summarized in the Tables 7.1 and 7.2. In Table 7.1 is presented the mean reconstruction efficiency of the 21 cm plus noise maps and the mean reconstruction error of the 21 cm signal power spectrum obtained with $N_{\text{realis}} = 10$, both obtained for the different data configurations proposed in this work. In Table 7.2 is presented the mean reconstruction error of the 21 cm signal power spectrum obtained with $N_{\text{realis}} = 10$ and $N_{\text{realis}} = 50$, both obtained for the MD80 data configuration.

The results obtained so far indicate that the debiasing procedure described here should work efficiently in BINGO data analysis pipeline. This suggests that, despite recovering the HI signal with good efficiency in the harmonic space, a more detailed analysis of the debiasing procedure should be carried out in future works, particularly

Table 7.1 - Mean 21 cm plus noise map reconstruction efficiency and mean 21 cm power spectrum reconstruction error for $N_{\text{realis}} = 10$.

	MD30	GD30	MD30+CBASS	MD80
$\bar{\rho}$	0.853	0.853	0.860	0.872
$\bar{\eta}$	7.2%	6.4%	5.7%	4.7%

Table 7.2 - Mean 21 cm power spectrum reconstruction error for MD80 case.

	$N_{\text{realis}} = 10$	$N_{\text{realis}} = 50$
$\bar{\eta}$	4.7 %	3.0 %

because the $1/f$ contribution to the instrumental noise was not considered in this analysis.

The main objective to be achieved with the continuity of this work is to test the foregrounds removal process presented in this thesis using a more complete set of simulations. They will be generated using the BINGO IM pipeline and will require some adaptations. The next steps to implement and test this new configuration are:

- Consider a more realistic beam profile, using the BINGO collaboration measurements, as well as the beam size dependence on frequency;
- Enable and test different map-making modules/models in the IM pipeline;
- Include the effect of 1/f noise in the simulated time ordered data and test the efficiency of the map-making methods in removing it from the BINGO maps;
- Include some components not considered in the simulations used in this

proposal, such as radio frequency interference (RFI) and atmospheric contamination.

REFERENCES

ABDALLA, E. et al. The BINGO project. I. Baryon acoustic oscillations from integrated neutral gas observations. **Astronomy and Astrophysics**, v. 664, p. A14, aug. 2022. 23, 24

ABDALLA, F. B. et al. The bingo project-iii. optical design and optimization of the focal plane. **Astronomy & Astrophysics**, v. 664, p. A16, 2022. 25, 27, 28

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974. 62

ALONSO, D.; BULL, P.; FERREIRA, P. G.; SANTOS, M. G. Blind foreground subtraction for intensity mapping experiments. **Monthly Notices of the Royal Astronomical Society**, v. 447, n. 1, p. 400–416, 2015. 27, 58

ALONSO, D.; SANCHEZ, J.; SLOSAR, A.; COLLABORATION, L. D. E. S. A unified pseudo-cℓ framework. **Monthly Notices of the Royal Astronomical Society**, v. 484, n. 3, p. 4127–4151, 2019. 48

AMARI, S.-i.; CICHOCKI, A. Adaptive blind signal processing-neural network approaches. **Proceedings of the IEEE**, v. 86, n. 10, p. 2026–2048, 1998. 58

ANDERSEN, K. et al. Beyondplanck i. global bayesian analysis of the planck low frequency instrument data. **arXiv preprint arXiv:2011.05609**, 2020. 34, 35, 36

ANSARI, R.; CAMPAGNE, J.-E.; COLOM, P.; GOFF, J. L.; MAGNEVILLE, C.; MARTIN, J.; MONIEZ, M.; RICH, J.; YECHE, C. 21 cm observation of large-scale structures at z˜ 1-instrument sensitivity and foreground subtraction. **Astronomy & Astrophysics**, v. 540, p. A129, 2012. 57

ANSARI, R.; MAGNEVILLE, C. Partial cmb maps: bias removal and optimal binning of the angular power spectrum. **Monthly Notices of the Royal Astronomical Society**, v. 405, n. 3, p. 1421–1430, 2010. 67

ASAD, K. et al. Primary beam effects of radio astronomy antennas–ii. modelling meerkat l-band beams. **Monthly Notices of the Royal Astronomical Society**, v. 502, n. 2, p. 2970–2983, 2021. 27

BACCIGALUPI, C. et al. Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps. **Monthly Notices of the Royal Astronomical Society**, v. 318, n. 3, p. 769–780, 2000. 58

BACCIGALUPI, C.; PERROTTA, F.; ZOTTI, G. D.; SMOOT, G. F.; BURIGANA, C.; MAINO, D.; BEDINI, L.; SALERNO, E. Extracting cosmic microwave background polarization from satellite astrophysical maps. **Monthly Notices of the Royal Astronomical Society**, v. 354, n. 1, p. 55–70, 2004. 58

BANDAY, A.; WOLFENDALE, A. Fluctuations in the galactic synchrotron radiation–i. implications for searches for fluctuations of cosmological origin. **Monthly Notices of the Royal Astronomical Society**, v. 248, n. 4, p. 705–714, 1991. 35

BASAK, S.; DELABROUILLE, J. A needlet internal linear combination analysis of wmap 7-year data: estimation of cmb temperature map and power spectrum. **Monthly Notices of the Royal Astronomical Society**, v. 419, n. 2, p. 1163–1175, 2012. 58

_____. A needlet ilc analysis of wmap 9-year polarization data: Cmb polarization power spectra. **Monthly Notices of the Royal Astronomical Society**, v. 435, n. 1, p. 18–29, 2013. 58

BATTYE, R.; BROWNE, I.; DICKINSON, C.; HERON, G.; MAFFEI, B.; POURTSIDOU, A. H i intensity mapping: a single dish approach. **Monthly Notices of the Royal Astronomical Society**, v. 434, n. 2, p. 1239–1256, 2013. 43

BEDINI, L.; HERRANZ, D.; SALERNO, E.; BACCIGALUPI, C.; KURUOĞLU, E.; TONAZZINI, A. Separation of correlated astrophysical sources using multiple-lag data covariance matrices. **EURASIP Journal on Advances in Signal Processing**, v. 2005, p. 1–13, 2005. 58

BENNETT, C. et al. First year wilkinson microwave anisotropy probe (wmap1) observations: Preliminary. **The Astronomical Journal Supplement Series**, v. 148, n. 1, 2003. 58

BENNETT, C. L. et al. Preliminary separation of galactic and cosmic microwave emission for the cobe differential microwave radiometer. **The Astrophysical Journal**, v. 396, p. L7–L12, 1992. 39

BETOULE, M.; PIERPAOLI, E.; DELABROUILLE, J.; JEUNE, M. L.; CARDOSO, J.-F. Measuring the tensor to scalar ratio from cmb b-modes in the presence of foregrounds. **Astronomy & Astrophysics**, v. 503, n. 3, p. 691–706, 2009. 58

BIGOT-SAZY, M.-A.; DICKINSON, C.; BATTYE, R. A.; BROWNE, I.; MA, Y.-Z.; MAFFEI, B.; NOVIELLO, F.; REMAZEILLES, M.; WILKINSON, P. Simulations for single-dish intensity mapping experiments. **Monthly Notices of the Royal Astronomical Society**, v. 454, n. 3, p. 3240–3253, 2015. 43, 57, 58

BOBIN, J.; STARCK, J.-L.; FADILI, J.; MOUDDEN, Y. Sparsity and morphological diversity in blind source separation. **IEEE Transactions on Image Processing**, v. 16, n. 11, p. 2662–2674, 2007. 58

BONALDI, A.; BROWN, M. L. Foreground removal for square kilometre array observations of the epoch of reionization with the correlated component analysis. **Monthly Notices of the Royal Astronomical Society**, v. 447, n. 2, p. 1973–1983, 2015. 58

BOND, J. R. Signal-to-noise eigenmode analysis of the two-year cobe maps. **Physical Review Letters**, v. 74, n. 22, p. 4369, 1995. 57, 58

BRANDT, W.; LAWRENCE, C.; READHEAD, A.; PAKIANATHAN, J.; FIOLA, T. Separation of foreground radiation from cosmic microwave background anisotropy using multifrequency measurements. **The Astrophysical Journal**, v. 424, p. 1–21, 1994. 57

BUNN, E. F.; FISHER, K. B.; HOFFMAN, Y.; LAHAV, O.; SILK, J.; ZAROUBI, S. Wiener filtering of the cobe differential microwave radiometer data. **Astrophysical Journal**, v. 432, p. L75–L78, 1994. 58

CARDOSO, J.-F.; JEUNE, M. L.; DELABROUILLE, J.; BETOULE, M.; PATANCHON, G. Component separation with flexible models—application to multichannel astrophysical observations. **IEEE Journal of Selected Topics in Signal Processing**, v. 2, n. 5, p. 735–746, 2008. 58

CARLSTROM, J. E.; HOLDER, G. P.; REESE, E. D. Cosmology with the sunyaev-zel'dovich effect. **Annual Review of Astronomy and Astrophysics**, v. 40, n. 1, p. 643–680, 2002. 46

CHAPMAN, E. et al. The scale of the problem: recovering images of reionization with generalized morphological component analysis. **Monthly Notices of the Royal Astronomical Society**, v. 429, n. 1, p. 165–176, 2013. 58

COMPIÈGNE, M.; VERSTRAETE, L.; JONES, A.; BERNARD, J.-P.; BOULANGER, F.; FLAGEY, N.; BOURLOT, J. L.; PARADIS, D.; YSARD, N.

The global dust sed: tracing the nature and evolution of dust with dustem. **Astronomy & Astrophysics**, v. 525, p. A103, 2011. 42

CUNNINGTON, S.; IRFAN, M. O.; CARUCCI, I. P.; POURTSIDOU, A.; BOBIN, J. 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies. **Monthly Notices of the Royal Astronomical Society**, v. 504, n. 1, p. 208–227, 2021. 27

DELABROUILLE, J.; CARDOSO, J.-F. Diffuse source separation in cmb observations. In: MARTÍNEZ, V. J.; SAAR, E.; MARTÍNEZ-GONZÁLEZ, E.; PONS-BORDERÍA, M. J. (Ed.). **Data analysis in cosmology**. [S.l.]: Springer, 2008. p. 159–205. 55

DELABROUILLE, J.; CARDOSO, J.-F.; JEUNE, M. L.; BETOULE, M.; FAY, G.; GUILLOUX, F. A full sky, low foreground, high resolution cmb map from wmap. **Astronomy & Astrophysics**, v. 493, n. 3, p. 835–857, 2009. 58, 61

DELABROUILLE, J.; CARDOSO, J.-F.; PATANCHON, G. Multidetector multicomponent spectral matching and applications for cosmic microwave background data analysis. **Monthly Notices of the Royal Astronomical Society**, v. 346, n. 4, p. 1089–1102, 2003. 58

DELABROUILLE, J.; MELIN, J.-B.; BARTLETT, J. Simulations of sunyaev-zel'dovich maps and their applications. In: CHEN L. W.; MA, C. P. N. K. W. P. U. L. E. (Ed.). **AMiBA 2001: high-Z clusters, missing baryons, and CMB polarization**. [S.l.: s.n.], 2002. v. 257, p. 81. 47

DELABROUILLE, J. et al. The pre-launch planck sky model: a model of sky emission at submillimetre to centimetre wavelengths. **Astronomy Astrophysics**, v. 553, p. A96, 2013. 33, 41, 43, 44, 45

DICKINSON, C.; DAVIES, R.; DAVIS, R. Towards a free–free template for cmb foregrounds. **Monthly Notices of the Royal Astronomical Society**, v. 341, n. 2, p. 369–384, 2003. 37, 38, 39

DICKINSON, C. et al. The state-of-play of anomalous microwave emission (AME) research. **New Astronomy Reviews**, v. 80, p. 1–28, 2018. 40

DODELSON, S. **Modern cosmology**. Massachusetts, USA: Academic Press, 2003. 6

DRAGONE, C. Offset multireflector antennas with perfect pattern symmetry and polarization discrimination. **Bell System Technical Journal**, v. 57, n. 7, p. 2663–2684, 1978. 24

DRAINE, B. **Physics of the interstellar and intergalactic medium**. [S.l.]: Princeton: Princeton University Press, 2011. 38

DRAINE, B.; LAZARIAN, A. Diffuse galactic emission from spinning dust grains. **The Astrophysical Journal**, v. 494, n. 1, p. L19, 1998. 40, 41

EISENSTEIN, D. J. et al. Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. **Astrophysial Journal**, v. 633, n. 2, p. 560, 2005. 1

EISENSTEIN, D. J.; SEO, H.-J.; WHITE, M. On the robustness of the acoustic scale in the low-redshift clustering of matter. **The Astrophysical Journal**, v. 664, n. 2, p. 660, 2007. 9

ERICKSON, W. C. A mechanism of non-thermal radio-noise origin. **The Astrophysical Journal**, v. 126, p. 480, 1957. 40

ERIKSEN, H.; JEWELL, J.; DICKINSON, C.; BANDAY, A.; GÓRSKI, K.; LAWRENCE, C. Joint bayesian component separation and cmb power spectrum estimation. **The Astrophysical Journal**, v. 676, n. 1, p. 10, 2008. 58

ERIKSEN, H. K.; BANDAY, A.; GÓRSKI, K.; LILJE, P. On foreground removal from the wilkinson microwave anisotropy probe data by an internal linear combination method: limitations and implications. **The Astrophysical Journal**, v. 612, n. 2, p. 633, 2004. 58

ERIKSEN, H. K. et al. Cosmic microwave background component separation by parameter estimation. **The Astrophysical Journal**, v. 641, n. 2, p. 665, 2006. 58

ERIKSEN, H. K.; O'DWYER, I.; JEWELL, J.; WANDELT, B.; LARSON, D.; GÓRSKI, K.; LEVIN, S.; BANDAY, A.; LILJE, P. Power spectrum estimation from high-resolution maps by gibbs sampling. **The Astrophysical Journal Supplement Series**, v. 155, n. 2, p. 227, 2004. 58

ERLER, J.; BASU, K.; CHLUBA, J.; BERTOLDI, F. Planck's view on the spectrum of the sunyaev–zeldovich effect. **Monthly Notices of the Royal Astronomical Society**, v. 476, n. 3, p. 3360–3381, 2018. 46, 47

FERNÁNDEZ-COBOS, R.; VIELVA, P.; BARREIRO, R. B.; MARTÍNEZ-GONZÁLEZ, E. Multiresolution internal template cleaning: an application to the wilkinson microwave anisotropy probe 7-yr polarization data. **Monthly Notices of the Royal Astronomical Society**, v. 420, n. 3, p. 2162–2169, 2012. 58

FORNAZIER, K. S. F. et al. The BINGO project. V. Further steps in component separation and bispectrum analysis. **Astronomy and Astrophysics**, v. 664, p. A18, aug. 2022. 44, 58, 59, 64

FURLANETTO, S. R.; OH, S. P.; BRIGGS, F. H. Cosmology at low frequencies: the 21 cm transition and the high-redshift universe. **Physics Reports**, v. 433, n. 4-6, p. 181–301, 2006. 13, 15

GIARDINO, G.; BANDAY, A.; GÓRSKI, K.; BENNETT, K.; JONAS, J.; TAUBER, J. Towards a model of full-sky galactic synchrotron intensity and linear polarisation: a re-analysis of the parkes data. **Astronomy and Astrophysics**, v. 387, n. 1, p. 82–97, 2002. 35, 36

GORSKI, K. M.; HIVON, E.; BANDAY, A. J.; WANDELT, B. D.; HANSEN, F. K.; REINECKE, M.; BARTELMANN, M. Healpix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. **Astrophysical Journal**, v. 622, n. 2, p. 759, 2005. 31

HALL, A.; BONVIN, C.; CHALLINOR, A. Testing general relativity with 21-cm intensity mapping. **Physical Review D**, v. 87, n. 6, p. 064026, 2013. 31

HASLAM, C.; SALTER, C.; STOFFEL, H.; WILSON, W. A 408 mhz all-sky continuum survey. ii-the atlas of contour maps. **Astronomy and Astrophysics Supplement Series**, v. 47, p. 1, 1982. 35

HYPERPHYSICS. **The hydrogen 21-cm line**. 2016. Available from: <http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/h21.html>. 12

IRFAN, M. O.; BOBIN, J.; MIVILLE-DESCHÊNES, M.-A.; GRENIER, I. Determining thermal dust emission from planck hfi data using a sparse, parametric technique. **Astronomy & Astrophysics**, v. 623, p. A21, 2019. 43

IRFAN, M. O.; BULL, P. Cleaning foregrounds from single-dish 21 cm intensity maps with kernel principal component analysis. **Monthly Notices of the Royal Astronomical Society**, v. 508, n. 3, p. 3551–3568, 2021. 27

JEWELL, J.; LEVIN, S.; ANDERSON, C. Application of monte carlo algorithms to the bayesian analysis of the cosmic microwave background. **The Astrophysical Journal**, v. 609, n. 1, p. 1, 2004. 58

JONAS, J. L.; BAART, E. E.; NICOLSON, G. D. The rhodes/hartrao 2326-mhz radio continuum survey. **Monthly Notices of the Royal Astronomical Society**, v. 297, n. 4, p. 977–989, 1998. 35

JONES, M. E. et al. The c-band all-sky survey (c-bass): design and capabilities. **Monthly Notices of the Royal Astronomical Society**, v. 480, n. 3, p. 3224–3242, 2018. 50

KOGUT, A. Synchrotron spectral curvature from 22 mhz to 23 ghz. **The Astrophysical Journal**, v. 753, n. 2, p. 110, 2012. 35

KOVETZ, E. D. et al. Line-intensity mapping: 2017 status report. **arXiv preprint arXiv:1709.09066**, 2017. 1, 19, 20

_____. Astrophysics and cosmology with line-intensity mapping. **arXiv preprint arXiv:1903.04496**, 2019. 19, 20

LARSON, D.; ERIKSEN, H. K.; WANDELT, B.; GÓRSKI, K.; HUEY, G.; JEWELL, J.; O'DWYER, I. Estimation of polarized power spectra by gibbs sampling. **The Astrophysical Journal**, v. 656, n. 2, p. 653, 2007. 58

LAWSON, K.; MAYER, C.; OSBORNE, J.; PARKINSON, M. Variations in the spectral index of the galactic radio continuum emission in the northern hemisphere. **Monthly Notices of the Royal Astronomical Society**, v. 225, n. 2, p. 307–327, 1987. 35

LEACH, S. M. et al. Component separation methods for the planck mission. **Astronomy & Astrophysics**, v. 491, n. 2, p. 597–615, 2008. 55, 58

LEITCH, E.; READHEAD, A.; PEARSON, T.; MYERS, S. An anomalous component of galactic emission. **The Astrophysical Journal**, v. 486, n. 1, p. L23, 1997. 40

LICCARDO, V.; MERICIA, E. J. de; WUENSCHE, C. A.; ABDALLA, E.; ABDALLA, F. B.; BAROSI, L.; BRITO, F. A.; QUEIROZ, A.; VILLELA, T.; PEEL, M. W.; WANG, B.; COSTA, A. A.; FERREIRA, E. G. M.; FORNAZIER, K. S. F.; NOVAES, C. P.; SANTOS, L.; SANTOS, M. V. dos; REMAZEILLES, M.; ZHANG, J.; DICKINSON, C.; HARPER, S.; LANDIM, R. G.; MARINS, A.;

VIEIRA, F. The BINGO project. IV. Simulations for mission performance assessment and preliminary component separation steps. **Astronomy and Astrophysics**, v. 664, p. A17, aug. 2022. 32, 33, 49, 58, 59, 67

LIU, A.; TEGMARK, M. A method for 21 cm power spectrum estimation in the presence of foregrounds. **Physical Review D**, v. 83, n. 10, p. 103006, 2011. 58

_____. How well can we measure and understand foregrounds with 21-cm experiments? **Monthly Notices of the Royal Astronomical Society**, v. 419, n. 4, p. 3491–3504, 2012. 57

LOUREIRO, A. et al. Cosmological measurements from angular power spectra analysis of boss dr12 tomography. **Monthly Notices of the Royal Astronomical Society**, v. 485, n. 1, p. 326–355, 2019. 32

MAINO, D.; BANDAY, A.; BACCIGALUPI, C.; PERROTTA, F.; GÓRSKI, K. Astrophysical component separation of cobe-dmr 4-yr data with fastica. **Monthly Notices of the Royal Astronomical Society**, v. 344, n. 2, p. 544–552, 2003. 58

MAINO, D.; FARUSI, A.; BACCIGALUPI, C.; PERROTTA, F.; BANDAY, A.; BEDINI, L.; BURIGANA, C.; ZOTTI, G. D.; GÓRSKI, K.; SALERNO, E. All-sky astrophysical component separation with fast independent component analysis (fastica). **Monthly Notices of the Royal Astronomical Society**, v. 334, n. 1, p. 53–68, 2002. 58

MARTÍNEZ-GONZÁLEZ, E.; DIEGO, J.; VIELVA, P.; SILK, J. Cosmic microwave background power spectrum estimation and map reconstruction with the expectation-maximization algorithm. **Monthly Notices of the Royal Astronomical Society**, v. 345, n. 4, p. 1101–1109, 2003. 58

MARTINEZ, V. J.; SAAR, E.; GONZALES, E. M.; PONS-BORDERIA, M. J. **Data analysis in cosmology**. [S.l.]: Springer, 2009. 43, 46

MATSHAWULE, S. D.; SPINELLI, M.; SANTOS, M. G.; NGOBESE, S. H i intensity mapping with meerkat: primary beam effects on foreground cleaning. **Monthly Notices of the Royal Astronomical Society**, v. 506, n. 4, p. 5075–5092, 2021. 27

MCLEOD, M.; BALAN, S. T.; ABDALLA, F. B. A joint analysis for cosmology and photometric redshift calibration using cross-correlations. **Monthly Notices of the Royal Astronomical Society**, v. 466, n. 3, p. 3558–3568, 2017. 32

MERICIA, E. J. de et al. Testing synchrotron models and frequency resolution in bingo 21 cm simulated maps using gnilc. **Astronomy & Astrophysics**, v. 671, p. A58, 2023. 58, 59

MESINGER, A. **The Cosmic 21-cm revolution; charting the first billion years of our universe**. [S.l.: s.n.], 2019. 57

MIVILLE-DESCHÊNES, M.-A.; YSARD, N.; LAVABRE, A.; PONTHIEU, N.; MACIAS-PEREZ, J.-F.; AUMONT, J.; BERNARD, J.-P. Separation of anomalous and synchrotron emissions using wmap polarization data. **Astronomy and Astrophysics**, v. 490, n. 3, p. 1093–1102, 2008. 35, 36, 41

MOZDZEN, T. J.; BOWMAN, J. D.; MONSALVE, R. A.; ROGERS, A. E. Limits on foreground subtraction from chromatic beam effects in global redshifted 21 cm measurements. **Monthly Notices of the Royal Astronomical Society**, v. 455, n. 4, p. 3890–3900, 2016. 27

MUKHANOV, V. **Physical foundations of cosmology**. [S.l.]: Cambridge University Press, 2005. 5, 8

MURTAGH, F.; HECK, A. **Multivariate data analysis**. Dordrecht: Kluwer Academic xvi, 1987. 58, 62

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). **Galactic emission**. 2012. Available from: <https://asd.gsfc.nasa.gov/archive/arcade/science_galaxy.html>. 2

NAVARRO, M. V. **Diffuse radio foregrounds: all-sky polarisation, and anomalous microwave emission**. [S.l.]: Springer, 2015. 34, 35, 41

OLIVARI, L.; REMAZEILLES, M.; DICKINSON, C. Extracting h i cosmological signal with generalized needlet internal linear combination. **Monthly Notices of the Royal Astronomical Society**, v. 456, n. 3, p. 2749–2765, 2016. 58, 59, 77

OLIVARI, L. C. **Intensity mapping: a new approach to probe the large-scale structure of the universe**. [S.l.]: The University of Manchester, 2018. 43, 55, 57, 59

OXHOLM, T. M. A beginner's guide to line intensity mapping power spectra. **arXiv preprint arXiv:2204.00685**, 2022. 20

PADMANABHAN, H. Neutral hydrogen in the post-reionization universe. **Proceedings of the International Astronomical Union**, v. 12, n. S333, p. 216–221, 2017. 5

_____. Intensity mapping: a new window into the cosmos. In: PARANJAPE, M. B.; MACKENZIE, R.; THOMOVA, Z.; WINTERNITZ, P.; WITCZAK-KREMPA, W. E. (Ed.). **Quantum theory and symmetries**. [S.l.]: Springer, 2021. p. 457–464. 19

PADOVANI, M.; BRACCO, A.; JELIĆ, V.; GALLI, D.; BELLOMI, E. Spectral index of synchrotron emission: insights from the diffuse and magnetised interstellar medium. **Astronomy & Astrophysics**, v. 651, p. A116, 2021. 34

PATANCHON, G.; CARDOSO, J.-F.; DELABROUILLE, J.; VIELVA, P. Cosmic microwave background and foregrounds in wilkinson microwave anisotropy probe first-year data. **Monthly Notices of the Royal Astronomical Society**, v. 364, n. 4, p. 1185–1194, 2005. 58

PEEBLES, P. J. E. **Principles of physical cosmology**. [S.l.]: Princeton University Press, 1993. 47

PEEL, M. et al. Baryon acoustic oscillations from integrated neutral gas observations: radio frequency interference measurements and telescope site selection. **Journal of Astronomical Instrumentation**, v. 8, n. 01, p. 1940005, 2019. 23

PETERSON, J. B. et al. 21 cm intensity mapping. **arXiv preprint arXiv:0902.3091**, 2009. 22

PLANCK COLLABORATION. Planck 2013 results. XI. All-sky model of thermal dust emission. **Astronomy & Astrophysics**, v. 571, p. A11, 2014. 42

_____. Planck 2013 results. XII. Diffuse component separation. **Astronomy & Astrophysics**, v. 571, p. A12, 2014. 58

_____. Planck 2015 results. IX. Diffuse component separation: CMB maps. **Astronomy & Astrophysics**, v. 594, p. A9–A9, 2016. 58

_____. Planck 2015 results-x. diffuse component separation: foreground maps. **Astronomy & Astrophysics**, v. 594, p. A10, 2016. 36

_____. Planck 2015 results. X. Diffuse component separation: foreground maps. **Astronomy & Astrophysics**, v. 594, p. A10, 2016. 58

_____. Planck 2015 results. XXV. Diffuse low-frequency galactic foregrounds. **Astronomy & Astrophysics**, v. 594, p. A25, sep. 2016. 41

_____. Planck intermediate results. XLVIII. Disentangling galactic dust emission and cosmic infrared background anisotropies. **Astronomy & Astrophysics**, v. 596, p. A109, 2016. 41, 43, 58

_____. Planck 2018 results. IV. Diffuse component separation. **Astronomy & Astrophysics**, v. 641, p. A4, 2020. 58

_____. Planck 2018 results. VI. Cosmological parameters. **Astronomy & Astrophysics**, v. 641, p. A6, sep. 2020. 33

PRITCHARD, J. R.; LOEB, A. 21 cm cosmology in the 21st century. **Reports on Progess in Physics**, v. 75, n. 8, p. 086901, 2012. 5, 18, 22

REICH, P.; REICH, W. A radio continuum survey of the northern sky at 1420 mhz. ii. **Astronomy and Astrophysics supplement series**, v. 63, p. 205–288, 1986. 35

_____. A map of spectral indices of the galactic radio continuum emission between 408 mhz and 1420 mhz for the entire northern sky. **Astronomy and Astrophysics Supplement Series**, v. 74, p. 7–23, 1988. 35

REMAZEILLES, M.; DELABROUILLE, J.; CARDOSO, J.-F. Cmb and sz effect separation with constrained internal linear combinations. **Monthly Notices of the Royal Astronomical Society**, v. 410, n. 4, p. 2481–2487, 2011. 58

_____. Foreground component separation with generalized internal linear combination. **Monthly Notices of the Royal Astronomical Society**, v. 418, n. 1, p. 467–476, 2011. 58, 59

REMAZEILLES, M.; DICKINSON, C.; BANDAY, A.; BIGOT-SAZY, M.-A.; GHOSH, T. An improved source-subtracted and destriped 408-mhz all-sky map. **Monthly Notices of the Royal Astronomical Society**, v. 451, n. 4, p. 4311–4327, 2015. 35

RYDEN, B. **Introduction to cosmology**. [S.l.]: Cambridge University Press, 2017. 7

SAHA, R.; JAIN, P.; SOURADEEP, T. A blind estimation of the angular power spectrum of cmb anisotropy from wmap. **The Astrophysical Journal**, v. 645, n. 2, p. L89, 2006. 58

SHAW, J. R.; SIGURDSON, K.; PEN, U.-L.; STEBBINS, A.; SITWELL, M. All-sky interferometry with spherical harmonic transit telescopes. **The Astrophysical Journal**, v. 781, n. 2, p. 57, 2014. 57, 58

SOBOLEV, V. The diffusion of lalpha radiation in nebulae and stellar envelopes. **Soviet Astronomy**, v. 1, p. 678, 1957. 14

SUNYAEV, R.; ZELDOVICH, Y. B. The observations of relic radiation as a test of the nature of x-ray radiation from the clusters of galaxies. **Comments on Astrophysics and Space Physics**, v. 4, p. 173, 1972. 45

SWITZER, E. et al. Determination of $z \sim 0.8$ neutral hydrogen fluctuations using the 21 cm intensity mapping autocorrelation. **Monthly Notices of the Royal Astronomical Society: Letters**, v. 434, n. 1, p. L46–L50, 2013. 58

SWITZER, E. R.; CHANG, T.-C.; MASUI, K. W.; PEN, U.-L.; VOYTEK, T. C. Interpreting the unresolved intensity of cosmologically redshifted line radiation. **The Astrophysical Journal**, v. 815, n. 1, p. 51, 2015. 58

TAUSCHER, K.; RAPETTI, D.; BURNS, J. O. Formulating and critically examining the assumptions of global 21 cm signal analyses: how to avoid the false troughs that can appear in single-spectrum fits. **The Astrophysical Journal**, v. 897, n. 2, p. 132, 2020. 27

TEGMARK, M. How to measure cmb power spectra without losing information. **Physical Review D**, v. 55, n. 10, p. 5895, 1997. 58

TEGMARK, M.; EFSTATHIOU, G. A method for subtracting foregrounds from multifrequency cmb sky maps. **Monthly Notices of the Royal Astronomical Society**, v. 281, n. 4, p. 1297–1314, 1996. 58

TEGMARK, M.; HAMILTON, A. J.; STRAUSS, M. A.; VOGELEY, M. S.; SZALAY, A. S. Measuring the galaxy power spectrum with future redshift surveys. **The Astrophysical Journal**, v. 499, n. 2, p. 555, 1998. 58

TEGMARK, M.; OLIVEIRA-COSTA, A. de; HAMILTON, A. J. High resolution foreground cleaned cmb map from wmap. **Physical Review D**, v. 68, n. 12, p. 123523, 2003. 58

TEGMARK, M.; TAYLOR, A. N.; HEAVENS, A. F. Karhunen-loeve eigenvalue problems in cosmology: how should we tackle large data sets? **The Astrophysical Journal**, v. 480, n. 1, p. 22, 1997. 57, 58

TINKER, J.; KRAVTSOV, A. V.; KLYPIN, A.; ABAZAJIAN, K.; WARREN, M.; YEPES, G.; GOTTLÖBER, S.; HOLZ, D. E. Toward a halo mass function for precision cosmology: the limits of universality. **Astrophysical Journal**, v. 688, n. 2, p. 709, 2008. 47

TURNER, M. S. The road to precision cosmology. **arXiv preprint arXiv:2201.04741**, 2022. 5

VOGELEY, M. S.; SZALAY, A. S. Eigenmode analysis of galaxy redshift surveys i. theory and methods. **arXiv preprint astro-ph/9601185**, 1996. 57, 58

WANDELT, B. D.; LARSON, D. L.; LAKSHMINARAYANAN, A. Global, exact cosmic microwave background data analysis using gibbs sampling. **Physical Review D**, v. 70, n. 8, p. 083511, 2004. 58

WEINBERG, D. H.; MORTONSON, M. J.; EISENSTEIN, D. J.; HIRATA, C.; RIESS, A. G.; ROZO, E. Observational probes of cosmic acceleration. **Physical Report**, v. 530, n. 2, p. 87–255, 2013. 1, 9

WOLZ, L.; ABDALLA, F. B.; ALONSO, D.; BLAKE, C.; BULL, P.; CHANG, T.-C.; FERREIRA, P. G.; KUO, C.-Y.; SANTOS, M. G.; SHAW, R. Foreground subtraction in intensity mapping with the ska. **arXiv preprint arXiv:1501.03823**, 2015. 58

WUENSCHE, C. A. et al. The bingo telescope: a new instrument exploring the new 21 cm cosmology window. **Journal of Physics: Conference Series**, v. 1269, n. 1, p. 012002, 2019. 28

_____. Baryon acoustic oscillations from integrated neutral gas observations: broadband corrugated horn construction and testing. **Experimental Astronomy**, v. 50, n. 1, p. 125–144, 2020. 27, 28, 29

_____. Baryon acoustic oscillations from integrated neutral gas observations: an instrument to observe the 21cm hydrogen line in the redshift range 0.13< z< 0.45–status update. **Anais da Academia Brasileira de Ciências**, v. 93, 2021. 2, 25, 28, 30

_____. The BINGO project. II. Instrument description. **Astronomy and Astrophysics**, v. 664, p. A15, aug. 2022. 23, 25, 26, 29, 48, 67

XAVIER, H. S.; ABDALLA, F. B.; JOACHIM, B. Improving lognormal models for cosmological fields. **Monthly Notices of the Royal Astronomical Society**, v. 459, n. 4, p. 3693–3710, Jul 2016. 31

ZHANG, L.; BUNN, E. F.; KARAKCI, A.; KOROTKOV, A.; SUTTER, P.; TIMBIE, P. T.; TUCKER, G. S.; WANDELT, B. D. Bayesian semi-blind component separation for foreground removal in interferometric 21 cm observations. **The Astrophysical Journal Supplement Series**, v. 222, n. 1, p. 3, 2016. 58