Princeton University
Department of Physics

Junior Paper

Toward quantification of the presence of the polarized dust foreground in cleaned SPIDER CMB maps

Connor Hainje

Advised by
Prof. William Jones

This paper represents my own work in accordance with University regulations.
/s/ Connor Hainje

Contents

1 Introduction **1**

2 Background **1**
2.1 Stokes maps 2
2.2 Polarized dust foreground 2
2.3 Galactic neutral hydrogen 3

3 Data **3**
3.1 Sky binning scheme: `HEALPix` 4
3.2 *Planck* 353 GHz map 4
3.3 SPIDER experiment 5
3.4 HI map 7

4 Analysis **7**
4.1 Correlation metrics 7
4.2 Correlation validation 9

5 Conclusions **13**

Abstract

Measurements of the polarized cosmic microwave background typically suffer from the presence of galactic signals from dusty regions of the interstellar medium, known as the polarized dust foreground. The polarization maps developed by SPIDER are no exception. The SPIDER maps have been cleaned using data from the *Planck* 353 GHz survey, but it is difficult to know the degree to which this cleaning was successful. As such, this paper presents a method to quantify the degree to which the dusty signals remain in the cleaned data by making use of galactic HI maps. This quantification method, however, does not reproduce important results from a similar study, so we have not applied the pipeline to SPIDER maps.

1 Introduction

The cosmic microwave background (CMB), present throughout all the universe, is extremely old electromagnetic radiation which dates back to the early universe. Measurements of the CMB thus provide crucial evidence for understanding the origin and formation of the universe. One such measurement, of chief importance for this study, is that of the polarization of the CMB. In particular, the existence of a pattern of divergence-free polarization, also known as "B-modes", in the polarized CMB is a generic signature of the presence of primordial gravitational waves. These gravitational waves are predicted by some inflationary cosmological models, and so the detection of the B-modes would serve to validate the inflationary paradigm of the origin of the universe.

The CMB polarization has been mapped by the SPIDER experiment, although interpreting this data is complicated by the presence of the polarized dust foreground. The SPIDER group has attempted to clean their maps of the CMB polarization by comparison with dust signals as mapped by the *Planck* experiment, but it is difficult to quantify the degree to which this was successful. This is the aim of this study: to measure and quantify the presence of dusty signals in the SPIDER CMB maps. We attempt to do so using galactic neutral hydrogen data, which, it has been shown, closely traces the orientation of dust in the interstellar medium.

This paper is organized as follows. In Section 2, we present background information regarding polarization, the polarized dust foreground, and galactic neutral hydrogen. In Section 3, we dicuss the data that are used for our study. Finally, in Section 4, we present our methodology for conducting our analysis, and show our attempts to reproduce important results from [1].

2 Background

The aim of this section is to provide important background information for this study. In subsection 2.1, we begin by describing how one quantifies and characterizes polarization. In subsection 2.2, we introduce the polarized dust foreground and explain its origin. Lastly, in subsection 2.3, we describe galactic neutral hydrogen and how it is linked to the dust.

2.1 Stokes maps

The polarization of electromagnetic radiation may be described uniquely using the *Stokes parameters*. Suppose that we consider an electromagnetic wave propagating in the z-direction. Then, the electric field component of the wave may be considered in terms of its x- and y-components, which we'll write as

$$E_x(t) = E_{0x}\, e^{i\delta_x}, \tag{1}$$
$$E_y(t) = E_{0y}\, e^{i\delta_y}, \tag{2}$$

where E_{0x} and E_{0y} are functions of time, chiefly responsible for the polarization of the wave, and δ_x and δ_y describe the traveling of the wave. They may be written, for example, as $kz - \omega t$, where k is the wave number and ω is the frequency. Note that there may also be a phase difference between the two, such that $\delta_y - \delta_x \neq 0$. With the electric field components described in this way, the Stokes parameters are defined as [4]

$$I = E_x E_x^* + E_y E_y^*, \tag{3}$$
$$Q = E_x E_x^* - E_y E_y^*, \tag{4}$$
$$U = E_x E_y^* + E_y E_x^*, \tag{5}$$
$$V = i(E_x E_y^* + E_y E_x^*), \tag{6}$$

where * represents complex conjugation. The interpretation of the Stokes parameters is as follows. I represents simply the intensity of the wave. Q describes the linear polarization of the wave along the x- and y-axes. U represents the linear polarization of the wave along axes at angles 45° above the positive and negative x-axes. V gives the degree of circular polarization.

The polarization of the CMB is known to be linear, and as such we may ignore V for the rest of this study. I, Q, and U thus comprise a description of the polarization of electromagnetic radiation. A map of the CMB polarization across the sky, then, involves measuring I, Q, and U at each point in the sky, or, at a point in each bin in the sky, assuming some binning scheme.

2.2 Polarized dust foreground

When one makes a map of the CMB polarization across the sky, the data includes a significant amount of signal from galactic sources whose radiative emissions are at similar frequencies. The polarizations of these sources are in no way related to the CMB, and so they must be understood and removed in order to understand the underlying CMB polarization data. It is in this way that these galactic signals are considered a "foreground"; they represent an additional layer in polarization maps that needs to be understood and removed in order to find the CMB polarization signals underneath.

The primary source of these galactic polarization emissions turns out to be dust in the interstellar medium. This dust affects measured polarizations via two phenomena, thermal emission and the polarization of starlight. Both of these come about as a result of the

phenomenon that dust grains tend to align their short axis with the local magnetic field. The first phenomenon, thermal emission, comes about because the thermal radiation of the grains is preferentially polarized in alignment with their long axes, forming a 90° angle with the local magnetic field orientation. The second phenomenon, starlight polarization, is a result of starlight passing through the dusty interstellar medium along our line of sight. As it does so, the component of the starlight which is polarized in alignment with the long axis of the dust grains is preferentially absorbed, so the starlight which reaches us is linearly polarized along the short axis of the dust grains. As such, this polarized starlight measures the orientation of the magnetic field and short axis of the galactic dust. The emissions detected from these two phenomena are known as "dust foreground" and "starlight polarization" data, respectively. We concern ourselves primarily with the former.

2.3 Galactic neutral hydrogen

Throughout the universe, there are pockets of galactic neutral hydrogen (HI). These are simply clouds of neutral hydrogen atoms, and they tend to form linear structures across the diffuse interstellar medium (ISM). The orientations of these structures have been found to be closely correlated with the local magnetic fields, and thus are closely correlated with dust orientations as well [1].

There are three principal facts to support this claim. (In fact, the following argument comes from the work of Clark and Hensley in [1].) The first is that the HI column density in the diffuse ISM traces the dust column density; namely, there are well-mixed in the diffuse ISM. The second fact is that, when limiting maps to narrow spectral channels, HI features are well aligned with plane-of-sky magnetic fields. The third fact is that the "coherence" of HI orientation as a function of velocity, namely the orderedness of HI features in different velocity channels along a line of sight, trace the degree of magnetic-field tangling along the line of sight. In other words, where HI orientations are aligned across velocity channels for a given line of sight, the magnetic field tends to be ordered as well.

When these three facts are considered in concert, one understands that structures in HI are very closely tied to the local magnetic fields therein. Furthermore, given that HI and dust tend to be very well-mixed, and dust similarly aligns itself with the local magnetic field, one can conclude that HI structures trace the orientation of dust in the ISM. It is this fact that we intend to make use of for this study. HI maps give a novel method of probing dusty signals and structures separate from the *Planck* maps, and so comparing the cleaned SPIDER maps to maps of the HI in the ISM is a novel method for testing the presence of dusty signals in the cleaned maps.

3 Data

In this section, we describe the data we use in this study. In subsection 3.1, we discuss `HEALPix`, the binning system used for all maps in this study. In subsection 3.2, we describe the *Planck* polarization maps and, in particular, the importance of the 353 GHz map as a map of polarized dust emissions. In subsection 3.3, we discuss the SPIDER experiment

and SPIDER maps, and the attempts which have been made to clean them of dusty signals. Lastly, in subsection 3.4, we discuss the HI maps from Clark and Hensley that we use.

3.1 Sky binning scheme: `HEALPix`

Throughout this study, all the sky maps considered make use of the `HEALPix` [5] binning scheme. `HEALPix` stands for **h**ierarchical **e**qual **a**rea iso**l**attitude **pix**elization of a sphere. It is a system which, as is implied by the name, subdivides a sphere into pixels of equal surface area. Each pixel is a curvilinear quadrilateral, determined by "hierarchical tesselation"; namely, the cells are subdivided recursively to increase resolution while maintaining the scalability of the pixelization. The number of these subdivisions is called N_{side}, and it directly determines the angular resolution of the binning scheme. A figure which shows the `HEALPix` pixelization for $N_{\text{side}} = 1$, 2, 4, and 8 is shown in Figure 1.

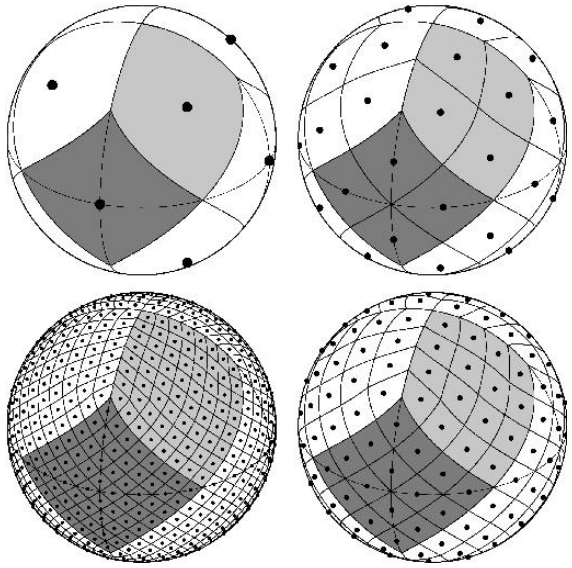

Figure 1: The `HEALPix` pixelization of a sphere using $N_{\text{side}} = 1$ (top left), 2 (top right), 4 (bottom right), and 8 (bottom left). The total number of pixels on the sphere is given by $N = 12 \times N_{\text{side}}^2 = 12$, 48, 192, and 768. Figure reproduced from [5].

3.2 *Planck* 353 GHz map

In 2015, the Planck collaboration measured and produced several polarization maps across the full sky using the *Planck* satellite [3]. These polarization maps were recorded at several different frequencies, the most pertinent of which is the 353 GHz map. This frequency is dominated by galactic dust emissions, and as such, the *Planck* 353 GHz map functions as a map of dust emissions.

The importance of this is that we can use the *Planck* 353 GHz map to better characterize the polarized dust foreground on our measurements of the CMB polarization. To do so, we primarily make use of the 353 GHz R3.01 full-sky map, the 353 GHz map corresponding to release version 3.01. We also make use of one other map for a test in a later section:

the GNILC-identified dust emission component of the 353 GHz R3.00 full-sky map. The Generalized Needlet Internal Linear Combination algorithm (GNILC) is an algorithm which isolates various components in the *Planck* maps corresponding to, for instance, dust emissions, the cosmic infrared background, and more. We use only the dust emission component of the 353 GHz R3.00 map in this study. Both of these maps are `HEALPix` maps, with $N_{\text{side}} = 2048$.

We also make use of a mask on the *Planck* maps. This mask eliminates data which is known to be significantly contaminated from Galactic sources, and as such it eliminates much of the data around the Galactic equator. The R3.01 Q and U maps, masked and unmasked, are shown in Figure 2. The GNILC-identified dust emission component of the R3.00 Q and U maps, masked and unmasked, are shown in Figure 3.

3.3 SPIDER experiment

This study is primarily a study meant to allow us to evaluate the presence of the polarized dust foreground in SPIDER maps. As such, we discuss the SPIDER experiment and its maps. Unfortunately, we were unable to apply the results of our work to these maps, but they are important for contextualizing this study in the larger SPIDER literature.

SPIDER is a balloon-borne polarimeter which was designed to map the CMB polarization with higher sensitivity and fidelity than any maps to date. It collected data during a long-duration balloon flight in 2015, where it mapped approximately 10% of the sky at frequencies of 95 GHz and 150 GHz. These maps come in the form of `HEALPix` maps, with $N_{\text{side}} = 512$ (yielding roughly 6.9 arcminute resolution).

Naturally, these maps initially contained the interstellar dust foreground. As such, efforts have been made by the SPIDER group to clean these maps of the dust foreground to reveal the underlying CMB polarization. The primary way this is done is by comparison with the *Planck* 353 GHz map. This *Planck* map is used a template for dusty signals, and removed from the SPIDER maps, theoretically leaving a dust-free map of the CMB polarization.

As a test of how well this cleaning process has removed the dusty signals, a previous student has compared the cleaned SPIDER maps with starlight polarization data. Because the polarization of starlight traces the orientation of galactic dust, correlations between the CMB polarization in the SPIDER maps and starlight polarization is an indication that the dusty signals have not been entirely removed. The previous student found the existence of "statistically significant" correlation between the two, indicating that the *Planck* -cleaned SPIDER maps have not fully removed dusty signals.

The goal of this paper is to perform a similar study of the presence of dusty signals in the cleaned SPIDER maps. In this study, however, we develop a method to more quantitatively measure the degree to which the dust foreground remains in the cleaned SPIDER maps. This is done using galactic neutral hydrogen (HI) data in a manner similar to the starlight polarization data used before.

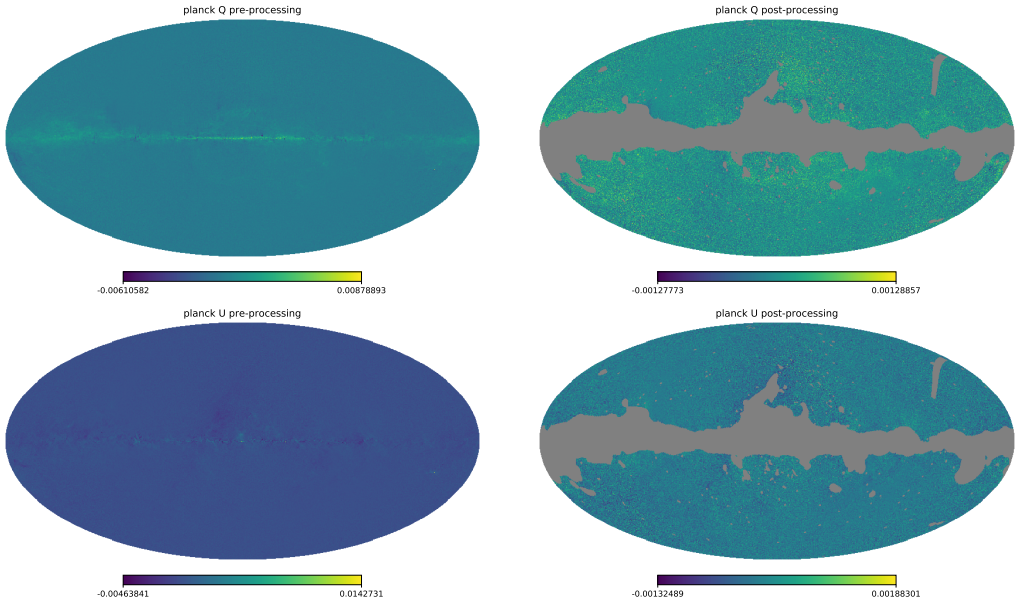

Figure 2: The *Planck* 353 GHz R3.01 full-sky Q and U Stokes maps, shown both with the Galaxy masked and without. The Q maps form the top row and the U maps the bottom. Unmasked maps are in the left column, masked maps in the right. Shown in units of $\mathrm{K_{CMB}}$.

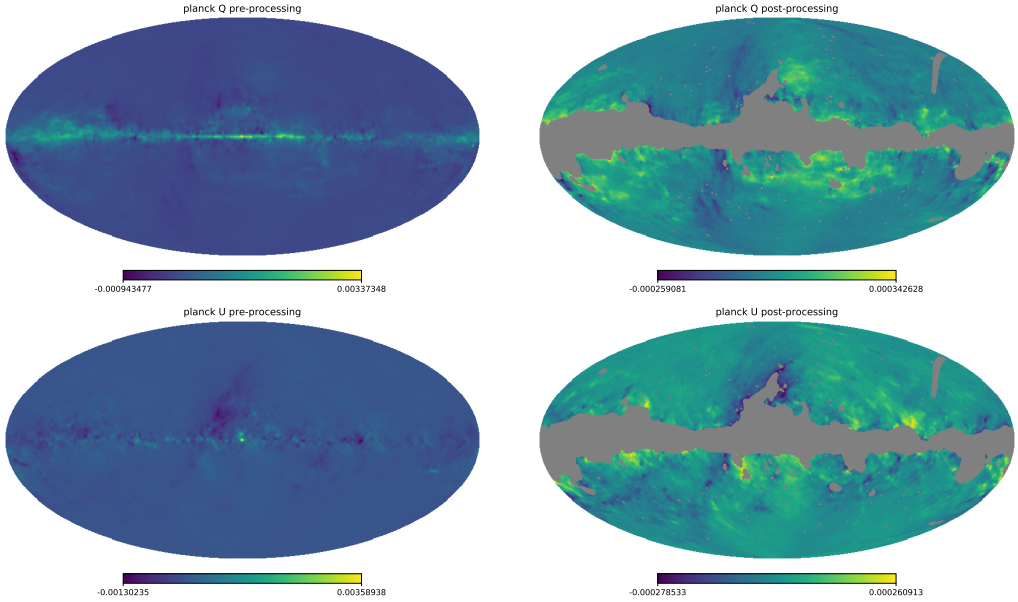

Figure 3: The GNILC-identified dust emission component of the *Planck* 353 GHz R3.00 full-sky Q and U Stokes maps, shown both with the Galaxy masked and without. The Q maps form the top row and the U maps the bottom. Unmasked maps are in the left column, masked maps in the right. Shown in units of $\mathrm{K_{CMB}}$.

3.4 HI map

In this study, we primarily make use of an HI map from Clark and Hensley [1], derived from the HI4PI [2] survey. HI4PI is a full-sky HI survey, with the highest angular resolution of any such full-sky surveys. From this HI survey, Clark and Hensley [1] have developed Stokes I, Q, and U maps describing the orientation of the HI structures in three dimensions: position, position, and velocity. Unless otherwise stated, references to "HI maps" are references to these HI4PI -derived Stokes maps.

For any sky bin, then, there are I, Q, and U values corresponding to a number of velocity channels. For a given velocity range $(v_{\min}, v_{\max})$, one can integrate the Stokes parameters by simple addition:

$$I_{\mathrm{HI}} = \sum_{v=v_{\min}}^{v_{\max}} I(v), \tag{7}$$

$$Q_{\mathrm{HI}} = \sum_{v=v_{\min}}^{v_{\max}} Q(v), \tag{8}$$

$$U_{\mathrm{HI}} = \sum_{v=v_{\min}}^{v_{\max}} U(v). \tag{9}$$

Furthermore, one can find the polarization orientations θ_{HI} for a given velocity range by first computing the above Q_{HI} and U_{HI} for the range, then computing

$$\theta_{\mathrm{HI}} = \frac{1}{2} \arctan \frac{U_{\mathrm{HI}}}{Q_{\mathrm{HI}}}. \tag{10}$$

In this way, we can understand the linear structure of galactic HI and its dependence on velocity, and we can use this to inform our correlation studies between HI data and the SPIDER maps.

In [1], the correlation between these HI Stokes maps and the *Planck* 353 GHz map is studied as a test of the model that galactic HI and galactic dust are tightly correlated. Their correlation analysis involved the use of two primary correlation measures: the Spearman rank correlation coefficient and the mean degree of angular alignment. We intend to use these same measures of correlation for our quantification methods, and as such they are considered in greater detail in the following section.

4 Analysis

To perform our study of the correlation between the SPIDER maps and the Clark and Hensley HI maps, we compute the Spearman rank correlation coefficient and the mean degree of angular alignment, much as is done in [1].

4.1 Correlation metrics

The first correlation coefficient to be used is the Spearman rank correlation coefficient. The Spearman correlation coefficient r between two datasets X and Y is defined as the Pearson

correlation coefficient between the *rank* variables [6]. For the purposes of this study we define the rank variables in the following way: given a list of data points x_i, each point is assigned an integer number corresponding to its rank in the list, which is roughly the same as its index in the sorted list. If any points x_i and x_j are equal, they are assigned the same rank, but the following rank integers are skipped such that the rank of any number minus 1 is exactly the number of data points which are less than it. For example, given the set of measurements $\{1.3, 2.8, 3.1, 2.8\}$, we would assign ranks $\{1, 2, 4, 2\}$. We see that the value 3.1, with three numbers less than it, has rank 4, but both instances of the value 2.8, with only one value less than them, have rank 2.

Thus, for two maps with values X_i and Y_i, one can convert the raw measurements to ranks $\mathrm{rg}\, X_i$ and $\mathrm{rg}\, Y_i$ and compute the Pearson correlation coefficient between them, giving

$$r = \frac{\mathrm{cov}(\mathrm{rg}_X, \mathrm{rg}_Y)}{\sigma_{\mathrm{rg}_X} \sigma_{\mathrm{rg}_Y}}, \tag{11}$$

where rg_X is the set of ranks $\mathrm{rg}\, X_i$ (and similarly for rg_Y), $\mathrm{cov}(\mathrm{rg}_X, \mathrm{rg}_Y)$ is the covariance of rg_X and rg_Y, and $\sigma_{\mathrm{rg}_{X/Y}}$ is the standard deviation of the corresponding rank variable.

We choose the Spearman correlation coefficient because it is sensitive to *monotonic* correlations between two maps. In particular, we seek to understand the degree to which two maps are correlated in structure, so we want to quantify the correlation between the structures in the maps. The particular values could vary greatly between a dust map and an HI map, for instance, but the structures are, in principle, largely the same. If we did not use rank variables, we would be much more sensitive to the specifics of these differences in value magnitudes; using rank variables, however, we uncover more about the actual structures in the maps.

To compute the Spearman correlation coefficient for the HI and SPIDER maps, then, the analysis would look as follows:

1. Identify the sky-bins for which both maps have data, as SPIDER is not a full-sky map.
2. Compute the ranks of the Q values in each of these sky-bins for HI and SPIDER, yielding variables $\mathrm{rg}_{Q_{\mathrm{HI}}}$ and rg_{Q_s} (where the subscript s denotes SPIDER). (This involves choosing a velocity range for the HI maps and computing the integrated Q map for this range using Equation 7.)
3. Compute the correlation coefficient using Equation 11.

For a single point statistic capturing the correlation between the SPIDER and HI maps, this process can be done using the HI maps fully integrated over all velocity channels. To understand the dependence of this correlation on velocity, however, one can repeat the above process using many different velocity ranges.

The other metric we will use to measure correlation is the mean degree of alignment between the orientations of the two datasets. This is computed as follows [1]. For two maps 1 and 2, we compute the polarization orientations θ_1 and θ_2 by Equation 10. The angular difference between θ_1 and θ_2 at each point in the sky may then be computed as

$$\delta\theta = \theta_1 - \theta_2. \tag{12}$$

This information is then collapsed to a point statistic ξ representing the mean alignment between the orientations θ_1 and θ_2 by

$$\xi = \langle \cos(2\delta\theta) \rangle. \tag{13}$$

Employing the definition of θ as $(1/2)\arctan(U/Q)$, we can further simplify the definition of ξ, however. Let $U_1/Q_1 \equiv a$ and $U_2/Q_2 \equiv b$. Then, we find that ξ can be re-written as

$$\xi = \left\langle \cos\left(\arctan(a) - \arctan(b) \right) \right\rangle .$$

Using the cosine addition identity, we can expand this to

$$\xi = \langle \cos(\arctan a)\cos(\arctan b) + \sin(\arctan a)\sin(\arctan b) \rangle .$$

Now we invoke the facts that $\cos(\arctan x) = 1/\sqrt{1+x^2}$ and $\sin(\arctan x) = x/\sqrt{1+x^2}$ to re-write the above as

$$\xi = \left\langle \frac{1}{\sqrt{1+a^2}} \frac{1}{\sqrt{1+b^2}} + \frac{a}{\sqrt{1+a^2}} \frac{b}{\sqrt{1+b^2}} \right\rangle .$$

Fully simplified, we arrive at the simpler definition of ξ as

$$\xi = \left\langle \frac{1+ab}{\sqrt{1+a^2}\sqrt{1+b^2}} \right\rangle , \tag{14}$$

which is defined only in terms of the initial Q and U maps.

Again, we may employ the above equations to compute the mean-alignment between the HI and SPIDER maps in a manner roughly as follows:

1. Identify the sky-bins for which both maps have data.
2. Compute the fractions $a = U_{\rm HI}/Q_{\rm HI}$ and $b = U_s/Q_s$ by using the HI and SPIDER Q and U maps. (As before, this involves choosing a velocity range for the HI data and computing the integrated HI Q and U maps.)
3. Compute the mean-alignment by using Equation 14.

As before, this process can be used to compute a single point statistic of the correlation between the full HI and SPIDER maps by integrating the HI data over the full velocity range. It may also be repeated for many velocity ranges in order to see the dependence of the correlation on the HI velocity.

4.2 Correlation validation

With the Spearman correlation coefficient and mean-alignment computational pipelines implemented as described above, validation for the pipeline is needed. The first and simplest test is to compute r and ξ using a dataset and itself. Because the two datasets are identical, both r and ξ should be 1. Indeed, this is what we find, using each of the *Planck* and HI maps.

As a second, more intensive test, we seek to reproduce Figure 13 of [1]. This figure is reproduced here as Figure 4. This is done by computing the Spearman correlation coefficient r and and the mean-alignment ξ between the HI maps and the Planck 353 GHz map at a range of velocities. In particular, we define velocity-ranges which are always centered roughly at 0 that gradually increase in magnitude over all of the velocity channels.

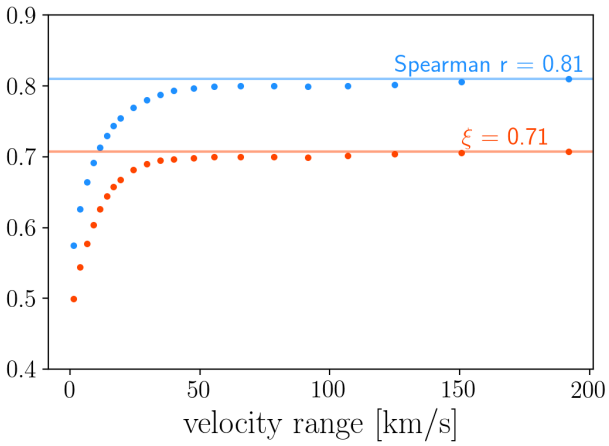

Figure 4: Correlations between HI maps and the *Planck* 353 GHz map as a function of HI velocity range. Figure reproduced from [1], where it is Figure 13.

Unfortunately, we have not as yet been able to completely reproduce the results of Clark and Hensley in their Figure 2. What follows is instead an account of the calculations that we have tried which we believe should accurately reproduce the correlation figure. For the following, unless otherwise stated, the *Planck* 353 GHz R3.01 full-sky map is used. Also for all of the following, both the *Planck* and HI maps are smoothed to an N_{side} value of 256.

First, we simply calculate the Spearman and mean-alignment correlation metrics for all of the cells for which both maps contain data. The result is shown in Figure 5. Note that the significance of the colors red and blue is reversed from that of the Clark and Hensley plot.

Next, we applied a mask to the maps. This mask blocks out the data roughly corresponding to the galactic equator; its purpose is to remove regions of the sky where contamination from other galactic signals may be present. With mask applied, we repeat the calculation of the correlation metrics. The result is shown in Figure 6.

Next, we repeated the above two steps, but first we subtracted the *Planck* 100 GHz R3.01 map from the *Planck* map. Because dusty signals are much more dominant at 353 GHz but other galactic signals dominate at 100 GHz, this has the effect of limiting the *Planck* map to only dusty signals. The results, both with and without masking the Galaxy, are shown in Figure 7. We do not see any significant change from before, indicating that the subtraction of the 100 GHz map has little impact on the correlation.

For our final attempt at reproducing Figure 4, we used a different *Planck* map. This *Planck*

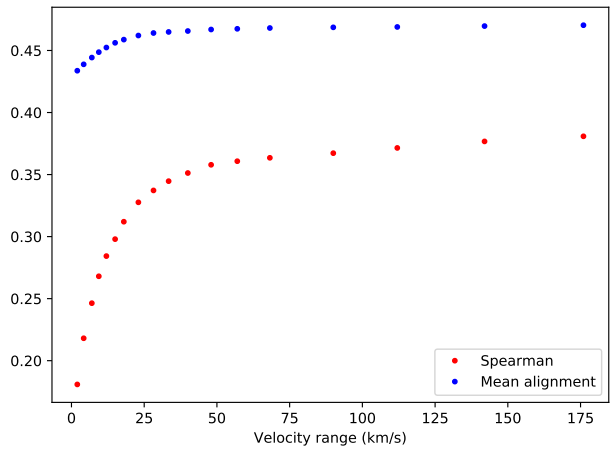

Figure 5: Correlations between *Planck* 353 GHz R3.01 and HI maps across the full sky. N_{side} scaled to 256 for both maps. Note that the significance of blue and red is reversed from that of the Clark and Hensley plot.

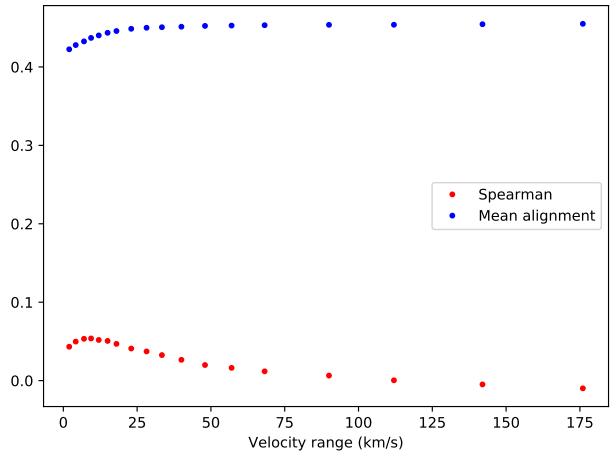

Figure 6: Correlations between *Planck* 353 GHz R3.01 and HI maps across the full sky. N_{side} scaled to 256 for both maps. A mask is applied to remove Galactic sources.

map is the thermal dust emission component resulting from the application of GNILC to the 353 GHz R3.00 map. We expect to see higher correlation using this map, as the non-dusty signals have been filtered out. The results, both with and without masking the Galaxy, are shown in Figure 8. This also appears to be quite similar to the previous results, although the correlation metrics are slightly greater in value. This shows that the dust component does yield higher correlation, as expected, but we are still not reproducing the values reported in [1].

A primary feature of note from the above tests is that our correlations go to zero for large velocity ranges when we mask the Galaxy. This strongly contradicts the results of Clark and Hensley, who claim they find a similar trend to that shown in Figure 4 when restricting the data to $|b| > 30°$ or $|b| > 60°$, which is roughly equivalent to our mask.

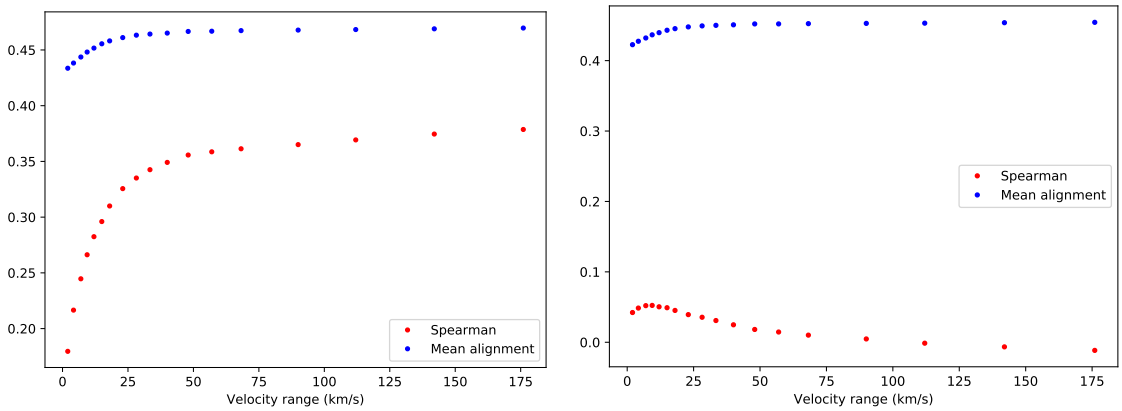

Figure 7: Correlations between *Planck* 353 GHz R3.01 map with *Planck* 100 GHz R3.01 subtracted and the HI map. N_{side} scaled to 256 for all maps. Correlation metrics are shown both without masking (left) and with masking (right).

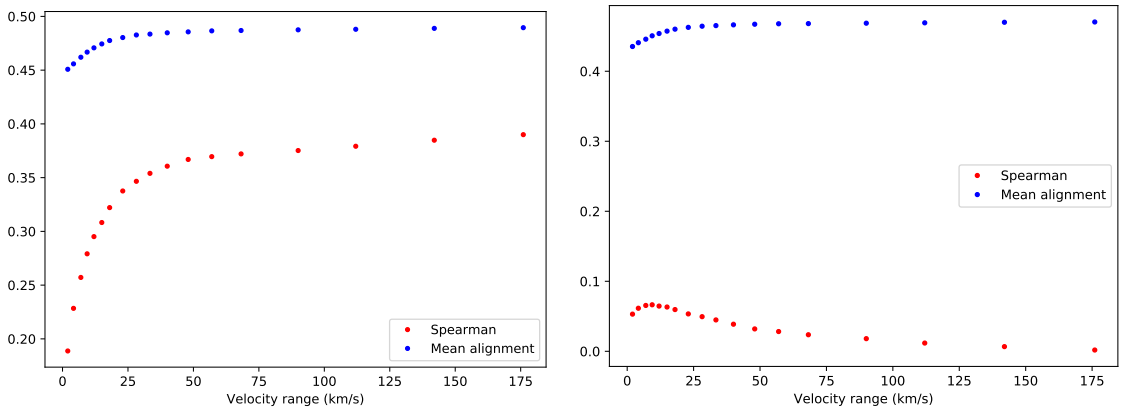

Figure 8: Correlations between the GNILC-identified thermal dust component of the *Planck* 353 GHz R3.00 map and the HI map. N_{side} scaled to 256 for both maps. Correlation metrics are shown both without masking (left) and with masking (right).

The other primary takeaway is that we seem to be able to reproduce the trends found in [1] for unmasked maps. Namely, we see a sharp dependence on the velocity range until around 30 km/s, after which the increase is very slow. Our results show a much less steep dependence for the mean alignment metric; our Spearman coefficient, however, shows a similarly steep increase. Numerically, however, we are rather far from the values we are trying to reproduce. Our Spearman coefficient reaches its maximum at about 0.4 for all trials, where Clark and Hensley report 0.81. Similarly, our mean alignment reaches only 0.45-0.5, where Clark and Hensley report 0.71. It is as yet unknown why there is such a discrepancy. Until this discrepancy is understood, however, we cannot move forward with our analysis of the SPIDER maps, as we cannot conclude that our calculations for correlations with HI maps are correct.

5 Conclusions

In this paper, we have presented first a short review of many important concepts necessary for study of the polarized cosmic microwave background and the effects of galactic dust, such as Stokes parameters, the `HEALPix` binning scheme, and background on the polarized dust foreground. We then presented a pipeline for quantifying and understanding the degree to which two maps are correlated, using the Spearman rank correlation coefficient and the mean degree of angular alignment.

We attempted to verify this pipeline by reproducing a result from Clark and Hensley regarding the correlation between their HI maps and a *Planck* 353 GHz map. Unfortunately, we were unsuccessful in this endeavor. For unmasked maps, we were able to recover the qualitative trends described in [1], but were unable to reproduce these trends for masked maps, in direct conflict with their results.

With more time, this study would have delved more deeply into exactly why our correlation metrics differed from those of Clark and Hensley. With this uncovered, the pipeline would then be able to be used to determine the correlation between HI maps and the SPIDER maps of the polarized CMB. This data would provide new insight about the SPIDER maps, as it would for the first time quantify the degree to which dusty sources are present in the SPIDER maps after being cleaned with *Planck* maps, using an independent dataset.

Acknowledgments

I would like to thank Professor Bill Jones for his advisory and guidance for this study. I would also like to thank Professor Aurelien Fraisse for his invaluable help. It's been a difficult time for us all, and so I am especially thankful that they were able to make time for me and to continue providing an all-around positive research experience!

References

[1] S. E. Clark and Brandon S. Hensley. "Mapping the Magnetic Interstellar Medium in Three Dimensions Over the Full Sky with Neutral Hydrogen". In: *The Astrophysical Journal* 887.2 (Dec. 2019). arXiv: 1909.11673, p. 136. ISSN: 1538-4357. DOI: 10.3847/1538-4357/ab5803. URL: http://arxiv.org/abs/1909.11673 (visited on 04/12/2020).

[2] HI4PI Collaboration et al. "HI4PI: A full-sky HI survey based on EBHIS and GASS". In: *Astronomy & Astrophysics* 594 (Oct. 2016). arXiv: 1610.06175, A116. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201629178. URL: http://arxiv.org/abs/1610.06175 (visited on 04/13/2020).

[3] Planck Collaboration. "Planck 2015 results. I. Overview of products and scientific results". In: *Astronomy & Astrophysics* 594 (Oct. 2016). arXiv: 1502.01582, A1. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201527101. URL: http://arxiv.org/abs/1502.01582 (visited on 04/13/2020).

[4] Edward Collett. *Field Guide to Polarization*. Bellingham, WA: SPIE Press, 2005.

[5] Krzysztof M. Górski et al. *The HEALPix Primer*. 2019. URL: https://healpix.sourceforge.io/pdf/intro.pdf.

[6] J. Myers and A. Well. *Research Design and Statistical Analysis*. 2nd ed. Lawrence Erlbaum Associates, Publishers, 2003.

[7] J. E. G. Peek et al. "The GALFA-H I Survey Data Release 2". In: *The Astrophysical Journal Supplement Series* 234 (Jan. 2018), p. 2. ISSN: 0067-0049. DOI: 10.3847/1538-4365/aa91d3. URL: http://adsabs.harvard.edu/abs/2018ApJS..234....2P (visited on 04/13/2020).