# How Correlated is Dust Polarization between Far-Infrared and CMB Frequencies?


Myra Norton

Advised by Brandon Hensley

A Junior Paper

Submitted to the Department of Astrophysical Sciences
in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Arts

Princeton University

Fall 2021

This paper represents my work in accordance with University regulations.
Myra Norton

# Contents

# 1 Motivation and Background

Physicists have theorized that the Universe underwent exponential expansion very shortly after the Big Bang, during an era referred to as Inflation. Efforts to prove this theory are ongoing, as evidence for Inflation has not yet been discovered. While the gravitational waves from this period are no longer detectable, they may have left their mark on the polarization pattern of the Cosmic Microwave Background (CMB) in the form of B-modes. Since the discovery of these B-modes would carry such tremendous implications, the astrophysics community was extremely interested when Harvard's BICEP team claimed to have found B-modes from primordial gravitational fluctuations (BICEP2 Collaboration et al., 2014). Upon further analysis, fellow cosmologists revealed that BICEP had indeed found B-modes but that these patterns were produced by interstellar dust rather than the aforementioned gravitational waves (Flauger et al., 2014; BICEP2/Keck Collaboration et al., 2015). To facilitate the ongoing search for evidence of Inflation, astrophysicists are working to subtract dust-generated B-modes from the current CMB data so that remaining B-modes are more likely to have been generated by primordial gravitational waves.

Unfortunately, this process is complicated by the fact that current data contains a mix of CMB and dust signals, which means that there is no data that solely represents dust polarization. However, since dust is about ten times hotter than the CMB, it emits electromagnetic radiation at a higher frequency. Therefore, measurements of the sky at increasing frequencies will result in stronger dust signals and weaker CMB signals until the latter disappear altogether, leaving behind data that is, optimistically, pure dust. This higher-frequency dust data must be scaled down in order to subtract it from the current, lower-frequency CMB data, but this scaling is only possible if the dust polarization at CMB frequencies is highly correlated with the dust polarization at far-infrared frequencies. This paper aims to quantify and analyze this correlation.

# 2 Physics Review

## 2.1 What is the CMB?

The Universe is believed to have been born 13.8 billion years ago with the Big Bang. When the Universe started to expand, the fundamental building blocks of matter began to appear. With so much energy and matter confined to such a small space, matter was constantly colliding, and the Universe was extremely hot. It was so hot that protons, electrons, and other particles were not bound together as atoms but rather formed what is formally called the photon-baryon plasma. As the hot Universe expanded, it cooled, and about 380,000 years after its birth, it was cool enough for free-floating electrons and protons to slow down and come together to form atoms. We call this the Epoch of Recombination. At this point, photons that had been trapped in the photon-baryon plasma were now free to soar through the Universe. These photons make up the Cosmic Microwave Background (CMB) signals we detect today.

Since this was the first time that photons were allowed to travel an unobstructed path, this is the oldest light that we can possibly detect and is therefore also the farthest signal we can see. Since it takes time for light to travel, the farther out in space we look, the further back in time we are looking. That's why we say that a star is "light years" away. When you see a star, you are actually seeing the photons that left the star in the past. In the same sense, if we look as far as we possibly can, we see the "wall," another name for the CMB. This is the last light we can see before running into the photon-baryon plasma, which we can't see through. This plasma doesn't exist anymore, but we can only see as far as time has permitted light to travel, and since we are looking into the past, we run into the no-longer-existent plasma.

Interestingly enough, the CMB is detected anywhere and everywhere on the sky. No matter where you look in the sky, you'll find a background signal in the microwave frequency. And the CMB signal is in the microwave frequency because space is expanding. The wavelength of light emitted all those years ago has been stretched as the Universe itself has stretched. In short, the CMB is the cosmic microwave background. It is a microwave signal that appears to be coming from every point in the sky. The CMB is the furthest and oldest thing we can see.

## 2.2 What is Polarized Light?

Light itself is electromagnetic radiation. It is an oscillation of magnetic and electric fields. The magnetic field oscillates perpendicular to the electric field, and the light travels, or propagates, in the direction perpendicular to the plane in which the electric and magnetic fields are oscillating.

To understand light polarization, it is sufficient to know the direction in which the electric field is oscillating since the magnetic field will simply be perpendicular to the electric field. We say that light is polarized if the electric field is preferentially oscillating in one direction. With this definition, it would appear that all light is polarized, which is technically true if just one photon is examined. However, what is more interesting is whether all light emitted by a given source has electric fields oscillating in the same direction. The standard way to measure the direction of light polarization is with the Stokes parameters, $Q$ and $U$, which act similarly to axes when describing polarization.

## 2.3 What are B-modes and E-modes, and why are they important?

The CMB is a signal made up of light. At each point in the sky, we can determine the polarization of the light in the CMB and record it as a map of the sky. This map is the polarization pattern of the CMB and allows us to find two commonly-sought trends: E-modes and B-modes. E-modes are more common patterns that are parallel or perpendicular to the direction of propagation, whereas the rarer B-modes fall 45 degrees from the direction of propagation. B-modes are of particular interest in cosmology because they may be the remnants of primordial gravitational waves and therefore evidence for Inflation.

## 2.4 What is Blackbody Radiation?

A blackbody is an object that absorbs all light, which means that it reflects no light. Any light coming from the object is being emitted by the object itself, and the wavelength of that light is dependent on just one parameter – temperature. Temperature is a measure of how quickly molecules are moving. If an object has a high temperature, then the molecules making up the object are moving quickly and have high kinetic energy. Each time two molecules bump into each other, they transfer energy. An increase in energy will excite electrons, which will put them in a higher orbital, whereas a decrease in energy will allow the electron to fall into a lower orbital or to its ground state. When an electron falls to a

lower orbital, it emits photons. A greater number of photons will be emitted when an electron falls from a higher orbital and when electrons fall to lower orbitals more often. Therefore, a blackbody with a higher temperature will emit more photons, and since more photons correlates to a higher frequency, a hotter blackbody will emit light at relatively higher frequencies.

## 2.5 What is the Planck Function?

$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{\frac{h\nu}{ek_B T} - 1}$$

The Planck function accurately describes the radiation of a blackbody at thermal equilibrium with its surroundings. This function takes a frequency $\nu$ and a temperature $T$ as input and outputs the spectral radiance $B(\nu, T)$ of the light emitted by the object in units of power per area per angle. So, if we know the temperature of an object, we can use this equation to determine its blackbody spectrum. With a fixed temperature, the Planck function is now only dependent on frequency. A graph of the spectral radiance versus frequency shows the distribution of how much light is emitted at different frequencies. Where this distribution peaks is the wavelength at which the object emits the brightest light.

## 2.6 What is a Modified Blackbody?

While the emission spectrum of a blackbody is modeled by the Planck function, a modified blackbody is represented by the Planck function scaled by a power law, specifically frequency to the power of $\beta$.

$$\nu^\beta \cdot B(\nu, T)$$

## 2.7 Frequencies

It is easy to think of light traveling as a wave. Waves travel with a speed, which we call frequency and measure in Hertz (Hz). The frequency tells you how many wavelengths pass a fixed point in space during the span of one second. So, waves with larger wavelengths will have lower frequencies, and conversely, waves with shorter wavelengths will have higher frequencies. For the purposes of this paper, it is important to note that because dust is about 10 times hotter than the CMB, it emits electromagnetic radiation at a higher frequency. The typical frequencies around which the CMB and dust are measured are 150 GHz and 2 THz, respectively.

# 3 Decorrelation of PySM Models at Higher Frequencies

## 3.1 Dust Model Implementation

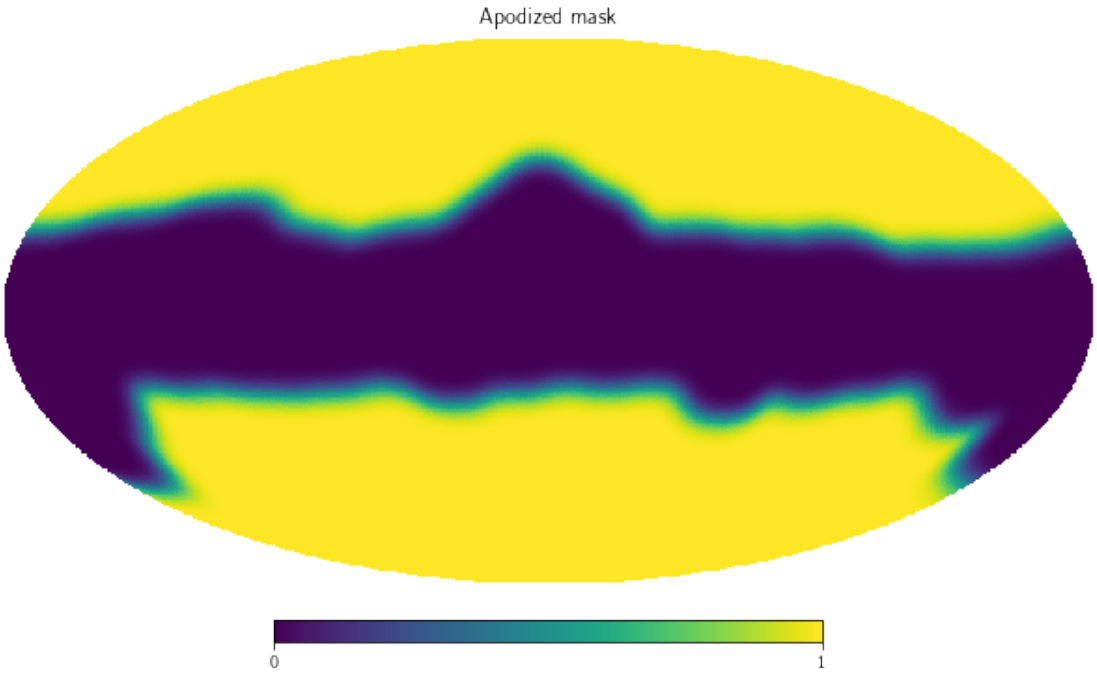
Using the various dust models within the Python Sky Model software library (PySM), we can simulate thermal dust intensity and polarization for a given frequency (Thorne et al., 2017; Zonca et al., 2021). Most models used in this analysis simulate dust with a modified blackbody spectrum that depends on just two parameters - temperature and $\beta$, also known as the spectral index.

- The d0 model uses a modified blackbody spectrum with a fixed temperature of 20K and a fixed spectral index $\beta$ of 1.54. Although implemented differently, the d8 model is extremely similar to the d0 model. Because both models fix parameters across the sky, changes in frequency do not result in any decorrelation. These models, therefore, do not appear on any graphs.
- The d1 model scales Planck data (Planck Collaboration et al., 2020) using a modified blackbody spectrum with data of spatially-varying temperature and $\beta$ values. The d4 model is structured similarly but contains a second dust population. The d2 and d3 models also work similarly to the d1 model, except the spectral indices are chosen at random from a Gaussian distribution with a mean of 1.59 and a standard deviation of 0.2 and 0.3, respectively.
- The d5 model is a physical model of dust that considers dust grain size but not variation in composition. It therefore simulates dust with fluctuating temperatures but a fixed spectral index $\beta$. This model may be closer to reality than d1 since it better appreciates the complexity of grain temperature physics at high frequencies. The d7 model is similar to the d5 model but uses a different fixed spectral index $\beta$ that comes from including iron in the dust composition. This difference only shows up at low frequencies and does not result in any notable difference in our analysis.
- The d12 model, also known as the MKD model, is a 3D model with 6 layers, each with their own spatially-varying temperature and $\beta$ maps.

## 3.2 Masking

In this portion of our analysis, we remove most noise from our Galaxy by applying a galactic plane mask to all calculations, maps, and skies. We downloaded the mask from the

Planck Legacy Archive that leaves 70% of the sky uncovered, lowered its resolution to match those of our maps at an $N_{side}$ of 128, and apodized it on a scale of about three degrees.




### 3.3 Spectral Correlation Ratio

$$R^{BB}_{\ell=80} = \frac{C^{BB}_{\ell=80}(217 \times \nu)}{\sqrt{C^{BB}_{\ell=80}(217 \times 217)\, C^{BB}_{\ell=80}(\nu \times \nu)}}$$

Using this equation, the Planck team analyzed the spectral correlation ratio for B-modes at 217 GHz and at 353 GHz to find a correlation of 99.1% (Planck Collaboration et al., 2020). We attempt to replicate this analysis for increasingly higher frequencies when $\ell = 80$ since this $\ell$-value corresponds to scales of about a degree, which is where B-modes from Recombination appear.

The $R_\ell$ value is essentially the average correlation between any two points that are a degree away, where one point is on one map and the second is on the other map. In the numerator of this ratio is the power of the cross spectrum for the 217 GHz map and the map at a frequency $\nu$ at an $\ell$ of 80. The denominator is the square root of the product of the $C_\ell$ value of the first frequency crossed with itself and the $C_\ell$ value of the second frequency crossed with itself, once again at an $\ell$ of 80.

Note that we perform this statistic for E-modes $R^{EE}_{\ell=80}$ and intensity $R^{TT}_{\ell=80}$.

## 3.4 Procedures

After uploading and apodizing the aforementioned mask from the Planck Legacy Archive, we use PySM to generate a sky and a 217 GHz map according to the d0 model. We then use pymaster, a python wrapper for the NaMaster software package (Alonso et al., 2019), to compute the $C_\ell(217 \times 217)$ value and initialize the spin-2 fields required to calculate $R_\ell^{BB}$ and $R_\ell^{EE}$ and the spin-0 field needed to compute $R_\ell^{TT}$.

With the skies, maps, and spin fields corresponding to the d0 model and the first term under the square root of the spectral correlation ratio available, we use pymaster to calculate the cross power spectra for various frequencies against the 217 GHz map as well as the second term under the square root in order to complete the calculation of the ratio. In other words, we perform the spectral correlation ratio for values of $\nu$ from 300 GHz to 3000 GHz.

Once the correlation statistic is calculated using an $\ell$ of 80 for each of the specified frequencies, the process is repeated for all PySM models.

## 3.5 Results

Plotted on the x-axis of Figure 1 are the frequencies of each map correlated with the 217 GHz map. Because we observe small differences in correlation, we subtract the correlation statistic from one and plot that value on the y-axis. Since a correlation ratio of zero means there is no correlation between the two frequencies, no correlation among all the frequencies would appear on this figure as a horizontal line equal to one. Conversely, a horizontal line at zero would denote perfect correlation among all frequencies. It should be noted that perfect correlation would not appear on this figure since both axes are plotted logarithmically. This is what happens with both the d0 and d8 models because they have perfect correlation due set temperature and $\beta$ parameters.

As mentioned before, the authors of the Planck paper did an analysis of their data to calculate the actual spectral correlation ratio between data at their two highest frequencies, 217 GHz and 353 GHz, and found a correlation limit of 99.1% (Planck Collaboration et al., 2020) This implies that, at 353 GHz, any viable model in PySM would have at least this level of correlation.

Unfortunately this is not the case for the $\beta$-varying models, d2 and d3, which are less correlated than the limit and are therefore not viable models. The rest of the models have sufficient correlation at 353 GHz, with the d4 model having the highest level of correlation.

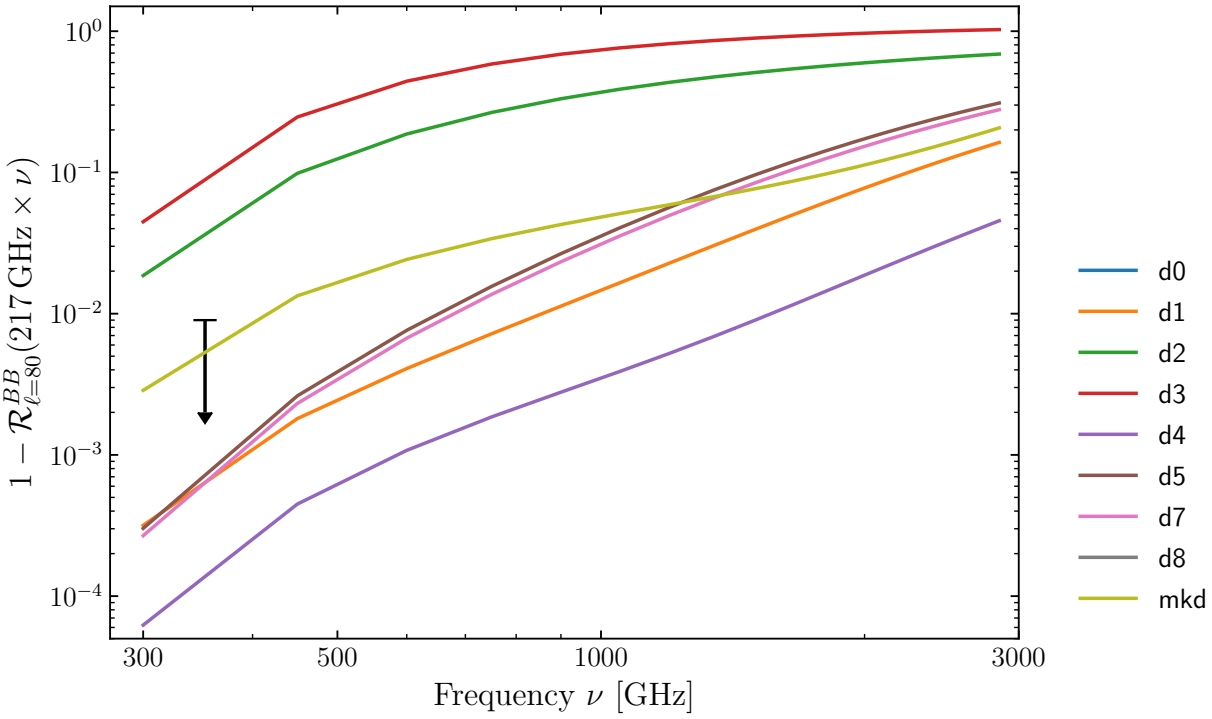



Figure 1: Curves corresponding to different models start at higher correlation and decorrelate as they are extrapolated to higher frequencies. The Planck restriction is plotted as a black dash, and all viable models fall below that dash. The d0 and d8 models have perfect correlation and do not appear on the graph.

The d12, or MKD, model has the lowest correlation of the acceptable models, which we believe is due to the complexity of having multiple layers, each with their own spatially-varying temperature and $\beta$ maps.

We also observe that each model predicts a different level of decorrelation, although we can appreciate that all sufficiently-correlated models remain more than 90% correlated even at 1000 GHz.

For completeness, we plotted the correlation of E-modes as well as intensity for each model. There are slight differences among the models for the three correlation statistics ($R_\ell^{BB}$, $R_\ell^{EE}$, and $R_\ell^{TT}$) computed. We note that the models all seem to be most correlated for the $R_\ell^{EE}$ calculation and least correlated for the $R_\ell^{TT}$ calculation. The relative ordering of the decorrelation appears to remain the same among all three calculations, with the exception that the d12 model is more correlated than the d1 model at the highest frequencies in the $R_\ell^{TT}$ calculation.

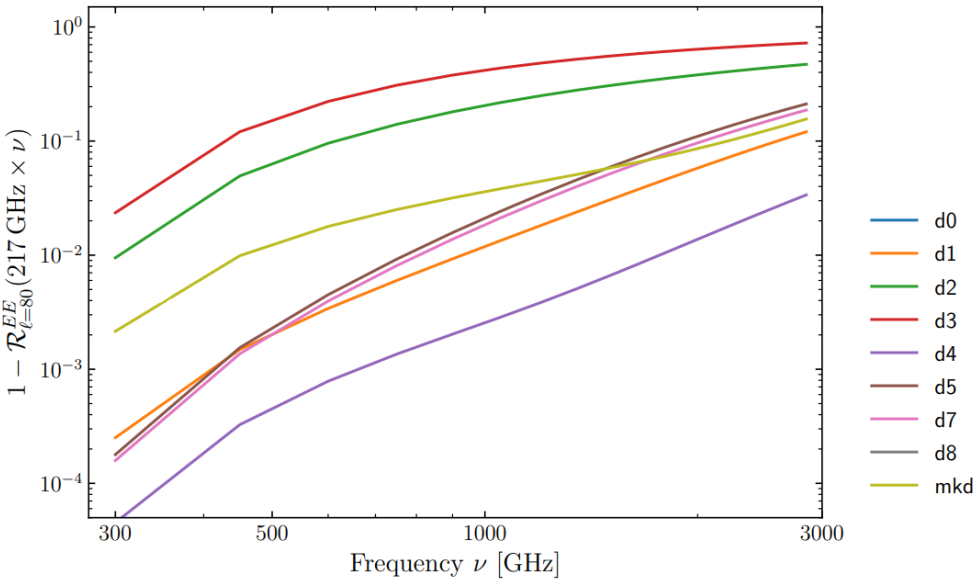



Figure 2: Curves corresponding to different models start at higher correlation and decorrelate as they are extrapolated to higher frequencies. The d0 and d8 models have perfect correlation and do not appear on the graph.

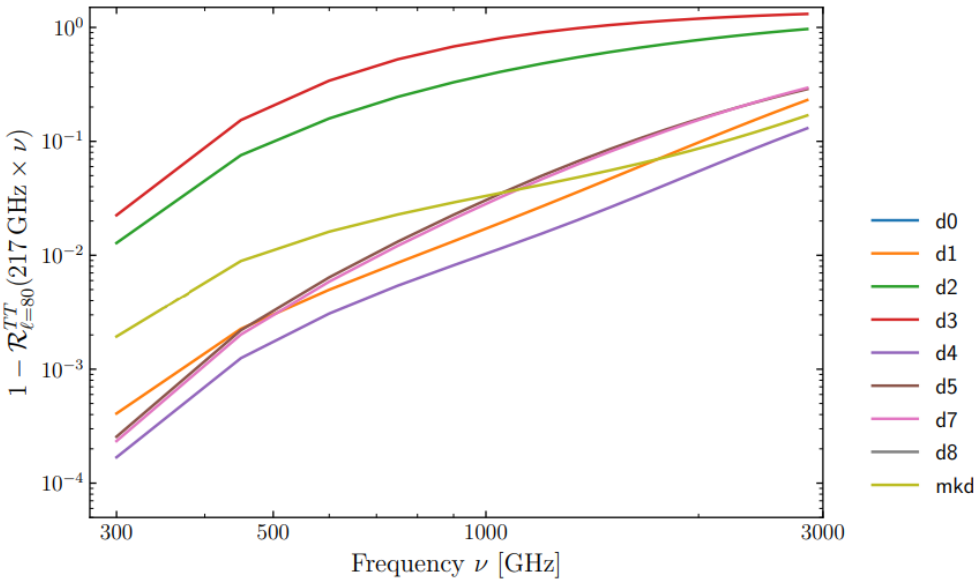



Figure 3: Curves corresponding to different models start at higher correlation and decorrelate as they are extrapolated to higher frequencies. The d0 and d8 models have perfect correlation and do not appear on the graph.

# 4 Determining the Maximum Range for $\beta$ in the d2 and d3 Models

Based on the previous analysis, the d2 and d3 models appear to inaccurately simulate thermal dust since they are less correlated than the actual data. To briefly review, these models scale Planck data using a modified blackbody spectrum with spatially-varying temperatures and $\beta$ values chosen at random from a Gaussian distribution with a mean of 1.59 and a standard deviation of 0.2 and 0.3, respectively. The d2 and d3 models are too decorrelated because their ranges for the $\beta$ value are too large.

## 4.1 Procedures

In order to find the largest range for $\beta$ that satisfies the Planck restriction, we alter the d2 model to have a standard deviation of 0.01 then perform the same correlation statistic from the previous analysis and plot the results. We increase the standard deviation by 0.01 until it reaches a value of 0.2, which is the original $\beta$ value for the d2 model that we know to be too large. To alter the standard deviation of the d2 model, we replace the assigned "map_mbb_index" with a new map with the specified standard deviation $\beta$. To create a new map in a way that is similar to how the original d2 map was created, we create an array of random numbers with length 3145728 (corresponding to an $N_{side}$ of 512), smooth the map to one degree resolution, and manipulate the random values to have a mean of 1.59 and the desired standard deviation.

## 4.2 Results

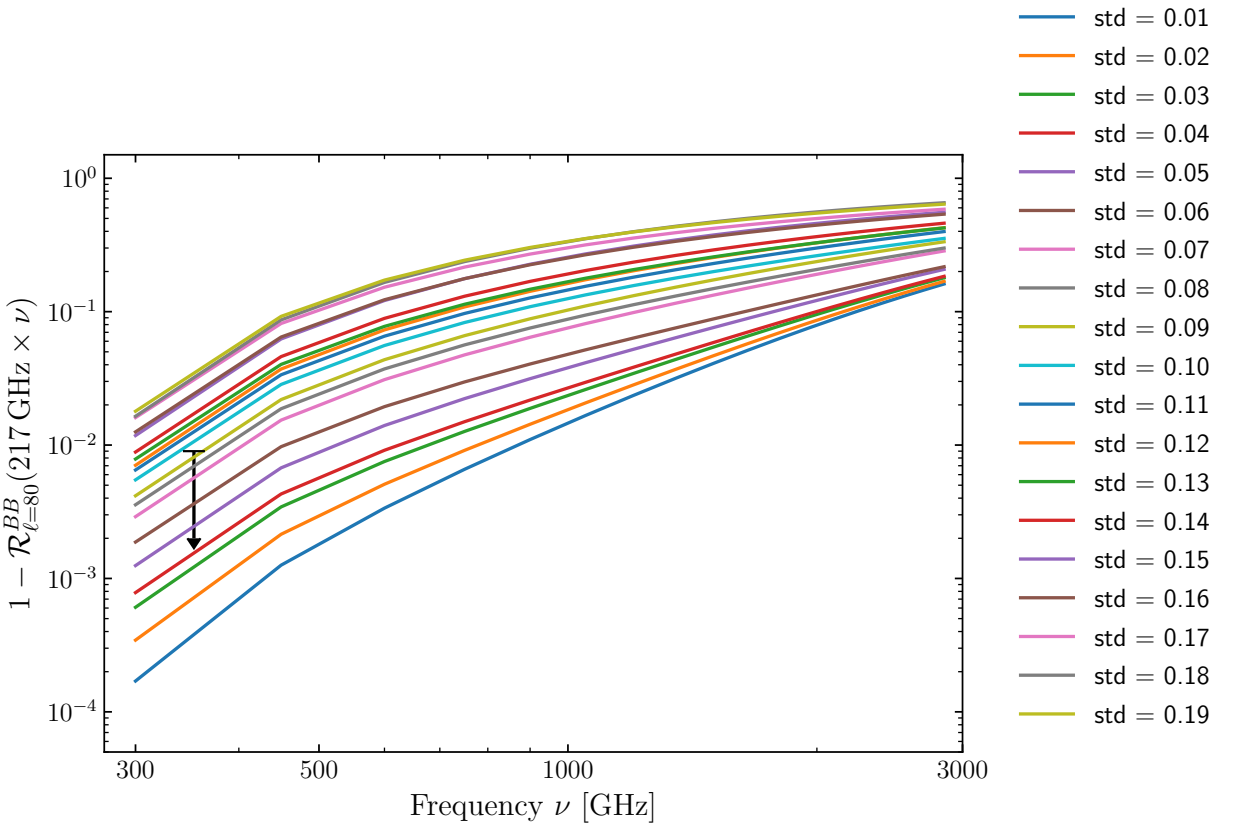



Figure 4: The decorrelation of the d2 model with 19 different standard deviation values from 0.01 to 0.19. The highest possible standard deviation that is sufficiently correlated according to the Planck restriction is 0.09.

According to Figure 4, the maximum standard deviation with sufficient correlation is approximately 0.09.

# 5 Relative Dependence of the d1 Model on $\beta$ and Temperature

After focusing on $\beta$ in the last analysis, we turn our attention to the impact it has on the d1 model. As previously mentioned, the modified blackbody used to simulate thermal dust only depends on two parameters - temperature and $\beta$.

## 5.1 Procedures

To differentiate the influence each parameter has on the d1 model, we compute the correlation statistic for all frequencies using three models: 1) the regular d1 model given by PySM, 2) the d1 model with a fixed temperature of 20K, and 3) the d1 model with a fixed $\beta$ value of 1.54. To simulate the latter two models, we simply create a map of fixed temperature and a map of fixed $\beta$ and use the component class within PySM to alter the original d1 presets.

## 5.2 Results

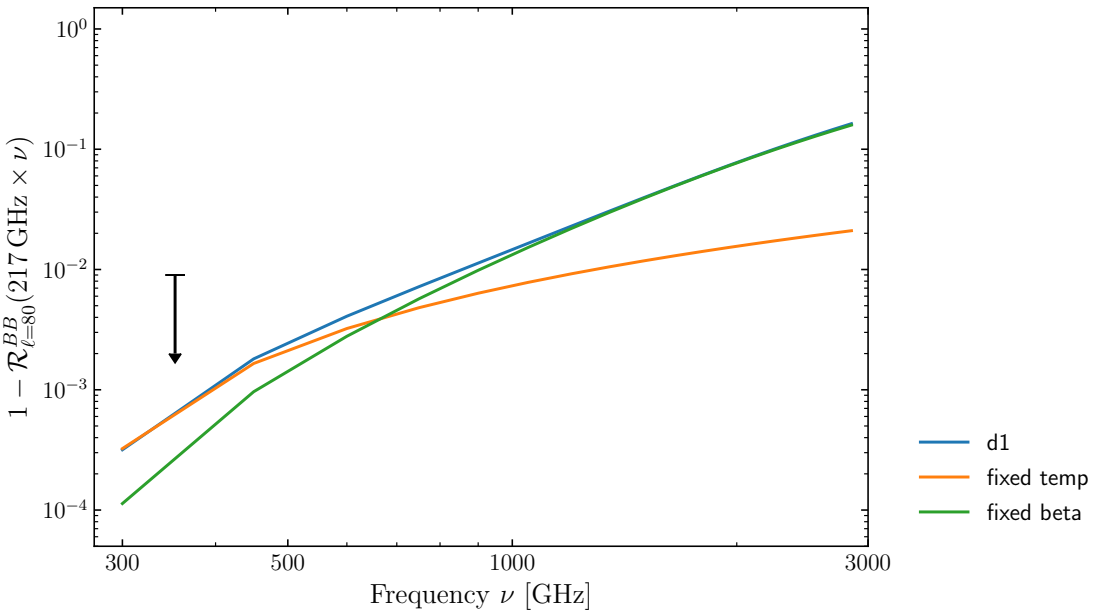



Figure 5: Blue curve represents the regular d1 model, the orange curve represents the d1 model with a fixed temperature and fluctuating $\beta$, and the green curve represents the d1 model with a fixed $\beta$ and fluctuating temperature. Since the d1 model is more similar to the orange curve at low frequencies, we conclude that the d1 model is more sensitive to $\beta$ at low frequencies and temperature at high frequencies.

In Figure 5, the blue curve represents the unaltered d1 model, the green curve shows the d1 model when $\beta$ is fixed at 1.54, and the orange curve represents the d1 model when temperature is fixed at 20K. At low frequencies, the orange curve is most similar to the unaltered d1 model, implying that the model is more sensitive to $\beta$ at low frequencies. After the orange and green curves intersect around 600 GHz, however, the green curve is more similar to the unaltered d1 model. At high frequencies, the model is more dependent on temperature fluctuations than those of $\beta$.

For completeness, we repeated the procedures and plotted the results for the $EE$ and $TT$ calculations.

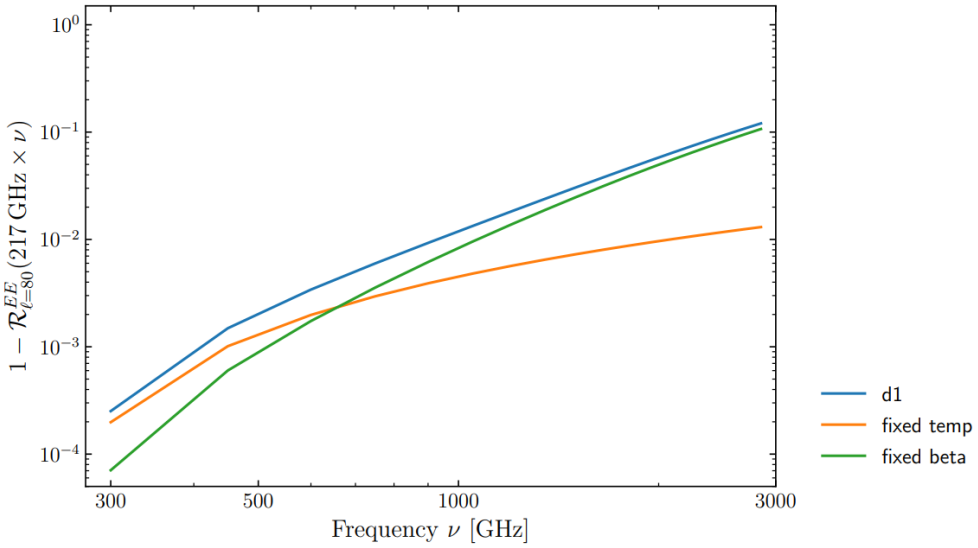



Figure 6: Blue curve represents the regular d1 model, the orange curve represents the d1 model with a fixed temperature and fluctuating $\beta$, and the green curve represents the d1 model with a fixed $\beta$ and fluctuating temperature. Since the d1 model is more similar to the orange curve at low frequencies, we conclude that the d1 model is more sensitive to $\beta$ at low frequencies and temperature at high frequencies.

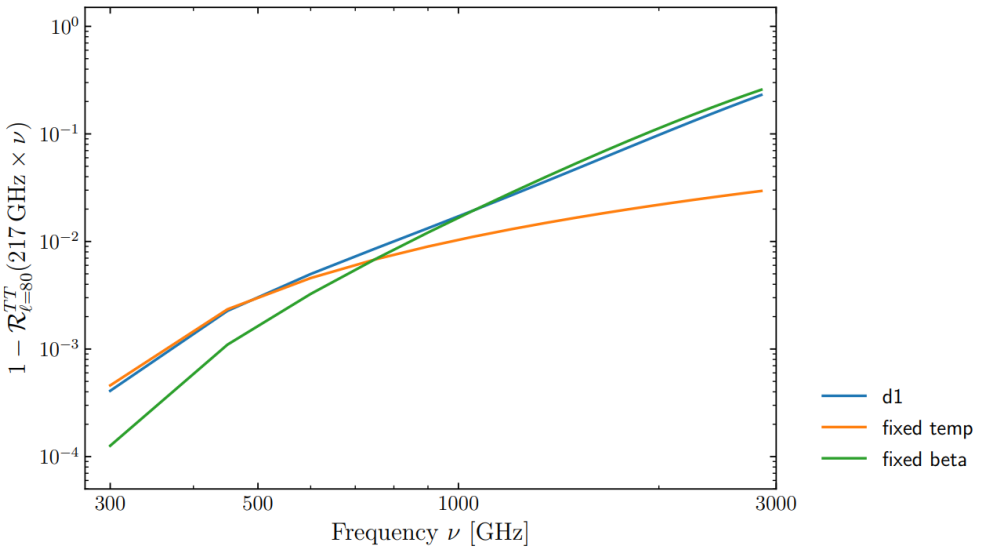



Figure 7: Blue curve represents the regular d1 model, the orange curve represents the d1 model with a fixed temperature and fluctuating $\beta$, and the green curve represents the d1 model with a fixed $\beta$ and fluctuating temperature. Since the d1 model is more similar to the orange curve at low frequencies, we conclude that the d1 model is more sensitive to $\beta$ at low frequencies and temperature at high frequencies.

# 6 Conclusion

From this research, we have come to four conclusions. First, we have quantified the decorrelation of each PySM model, which is useful for determining whether dust at a higher frequency will be sufficiently correlated for CMB calculations. Second, PySM should retire the d2 and d3 models, as they do not agree with actual data and therefore inaccurately simulate thermal dust. If, however, the range for $\beta$ is decreased to about 0.09, the model implementation can be rendered useful. Third, determining the approximate frequency at which temperature becomes more impactful on the d1 model may have implications for data-collection. Fourth, each model depicts a different decorrelation at high frequencies. This makes a compelling argument for collecting high-frequency data, as those measurements would allow us to figure out which model is closest to reality.

# References


Alonso, D., Sanchez, J., Slosar, A., & LSST Dark Energy Science Collaboration. 2019, MNRAS, 484, 4127, doi: 10.1093/mnras/stz093

BICEP2 Collaboration, Ade, P. A. R., Aikin, R. W., et al. 2014, Phys. Rev. Lett., 112, 241101, doi: 10.1103/PhysRevLett.112.241101

BICEP2/Keck Collaboration, Planck Collaboration, Ade, P. A. R., et al. 2015, Phys. Rev. Lett., 114, 101301, doi: 10.1103/PhysRevLett.114.101301

Flauger, R., Hill, J. C., & Spergel, D. N. 2014, J. Cosmology Astropart. Phys., 2014, 039, doi: 10.1088/1475-7516/2014/08/039

Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020, A&A, 641, A11, doi: 10.1051/0004-6361/201832618

Thorne, B., Dunkley, J., Alonso, D., & Næss, S. 2017, MNRAS, 469, 2821, doi: 10.1093/mnras/stx949

Zonca, A., Thorne, B., Krachmalnicoff, N., & Borrill, J. 2021, The Journal of Open Source Software, 6, 3783, doi: 10.21105/joss.03783