Signal-Preserving Cosmic Microwave Background Cleaning with Multi-Scale Learning

Helen Shao [1] **Fiona McCarthy** [2] **Miles Cranmer** [2] **Blake Sherwin** [2]

Abstract

Large-scale Cosmic Microwave Background (CMB) B-mode polarization is a key probe of inflation, but is obscured by complex Galactic dust foregrounds. Reliable foreground removal must suppress this contamination without distorting the underlying primordial signal. This is typically done with Internal Linear Combination (ILC), which requires multi-frequency data and is limited to second-order statistics. We present *single*-frequency cleaning with CNNs that exploit inter-scale correlations in Galactic dust structure to estimate large-scale foregrounds from small-scale modes, while preserving the primordial signal by construction. Using only small-scale B-modes, our model reduces the MSE from $1.4 \times 10^{-3}\,\mu\mathrm{K}^2$ for single-channel ILC to $3.4 \times 10^{-4}\,\mu\mathrm{K}^2$; adding same-frequency temperature and E-polarization further reduces it to $2.1 \times 10^{-4}\,\mu\mathrm{K}^2$. Hybrid models combining multi-frequency information with inter-scale learning improve over ILC residuals by $\mathcal{O}(10^3)$. Our results show that inter-scale and cross-component structure offers substantial foreground cleaning information beyond multi-frequency correlations, enabling ML models that are both physically reliable and complementary to frequency-based CMB foreground removal. However, incorporating realistic instrumental effects and improving generalization across simulations remain key challenges.

[1]Department of Astronomy, Harvard University, Cambridge, MA, USA [2]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. Correspondence to: Helen Shao <helen.shao@cfa.harvard.edu>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

1. Introduction

1.1. Galactic Foregrounds in CMB Observations

The Cosmic Microwave Background (CMB) offers a direct probe of cosmic inflation (Zaldarriaga & Seljak, 1997; Kamionkowski et al., 1997), as it leaves a unique signature on CMB B-mode polarization ("primordial B-modes") at large scales (angular multipole $\ell < 200$) (Seljak & Zaldarriaga, 1997). However, this faint imprint is obscured by complex foregrounds, such as Galactic dust emission (Planck Collaboration, 2016a; Bennett et al., 2013), whose power is over two orders of magnitude larger than the primordial signal on all observed scales (Planck Collaboration, 2016b; 2020a;c; Keck Collaboration, 2026)[1]. Foreground estimation and removal is therefore critical for primordial B-mode detection, a key goal of existing and upcoming CMB experiments such as Bicep/KECK (Bicep/KECK Collaboration, 2022), the Simons Observatory (SO) (Simons Observatory Collaboration, 2019; The Simons Observatory Collaboration, 2025), and LiteBIRD (LiteBIRD Collaboration, 2022).

Internal Linear Combination (ILC) is a widely used foreground-removal method that exploits correlations in foreground behavior across frequencies (Bennett et al., 1992; WMAP Collaboration, 2003; Eriksen et al., 2004; Delabrouille et al., 2008). It computes a weighted sum of multi-frequency CMB observations that minimizes foreground variance to reconstruct the cleaned primordial CMB signal (Eriksen et al., 2004). Crucially, ILC is *signal-preserving*—it fully recovers the primordial signal independent of uncertain foreground models, which makes it reliable for cosmological analysis. However, ILC has critical drawbacks. First, since it relies *only* on frequency correlations of the foregrounds, it requires multi-frequency data. Second, ILC only optimizes for second-order statistics (e.g., the frequency-space covariance matrix), neglecting higher-order correlations that are abundant in Galactic foregrounds (Planck Collaboration, 2016c; Kandel et al., 2017; Vansyngel et al., 2018). Third, it is suboptimal for foregrounds with highly anisotropic features, as in the case of dust emission near the Galactic plane (Planck Collaboration,

[1]We refer to the primordial signal in the absence of gravitational lensing.

2016b; Kritsuk et al., 2018; Caldwell et al., 2017).

1.2. Signal-Preserving ML for Foreground Removal

To capture non-Gaussian and anisotropic structures that ILC neglects, a growing literature applies machine learning (ML) to CMB foreground removal (Petroff et al., 2020; McCarthy et al., 2024; Yan et al., 2024; M et al., 2025; Bachar et al., 2026). Despite achieving high fidelity cleaning, ML-based removal is not automatically reliable as it is sensitive to model misspecifications in training simulations (e.g., Grünwald & van Ommen, 2018). In other words, physical reliability requires more than low foreground residuals: the estimator must fully preserve the primordial B-mode information. This is a nontrivial constraint for ML foreground removal, since a flexible model trained end-to-end may suppress or synthesize components of the primordial signal if doing so improves the training loss, consequently biasing the CMB reconstruction and propagating hard-to-trace systematics into downstream inference(McCarthy et al., 2024). This issue is acute for Galactic dust, whose multi-scale structure remains imperfectly modeled (Kandel et al., 2017; Vansyngel et al., 2018; Kritsuk et al., 2018; Caldwell et al., 2017). We therefore target *signal-preserving* ML foreground removal that provides physically interpretable primordial CMB reconstructions for cosmological analysis.

McCarthy et al. (2024) introduced a ML framework that respects the signal-preserving property of ILC, while learning nonlinear foreground corrections with CNNs. The key is to construct network inputs and target that are *both* uncorrelated with (e.g., blind to) the signal one wishes to preserve — namely large-scale B-modes. Notably, they achieve over five times smaller residuals than the ILC while demonstrating unbiased reconstructions. However, their method still requires multi-frequency inputs. McCarthy et al. (2024) apply the same signal-blind construction with multi-frequency ILC residuals as inputs. Our single-frequency extension replaces frequency information with inter-scale dust correlations and our hybrid models combine both. A direct quantitative comparison on identical simulations is left to future work.

We extend this framework to *single-frequency* foreground removal by leveraging inter-scale correlations within Galactic dust. The key insight for preserving large-scale B-modes is that primordial B-modes at different angular scales are statistically independent (Zaldarriaga & Seljak, 1997), but Galactic dust interactions couple foreground structures across scales (Lazarian et al., 2023; Cho & Lazarian, 2002; Kamionkowski & Kovetz, 2014). This enables the construction of a dust estimator from structural and geometric patterns in both real and harmonic space, rather than from cross-channel frequency information. We therefore formulate *signal-preserving* large-scale dust foreground removal as scale-separated supervised regression from signal-blind, small-scale inputs. In a related work, Philcox et al. (2018) built a single-frequency *analytic* estimator that uses small-scale dust patterns (e.g. $\ell > 90$) to infer large-scale ($\ell < 80$) foregrounds that achieves $\sim 50\%$ cross-correlation with the truth. Their estimator is complementary to multi-frequency methods, but it is limited to low-order summary statistics (the hexadecapole of the 2D power spectrum), which does not capture the full range of foreground structures.

Here, we train convolutional neural networks (CNNs) at the field level to predict and clean large-scale foregrounds ($\ell < 200$), where the inflationary signal resides, using only small-scale modes ($\ell > 200$). Because large- and small-scale modes of the *primordial* CMB are statistically independent, this construction is signal-preserving by design. Relative to ILC variants, multi-frequency CNN cleaning (McCarthy et al., 2024), and the analytic estimator of (Philcox et al., 2018), our contributions are twofold: ***(i)*** signal-preserving, single-frequency CNN regressors for large-scale foregrounds using small-scale B-modes, with optional augmented inputs from temperature (T) and/or E-mode polarization (E) at the same frequency; and ***(ii)*** a hybrid CNN that leverages both inter-scale and multi-frequency correlations. Both variants reconstruct large-scale primordial B-modes in an unbiased manner. We demonstrate these methods on DUSTFILAMENTS simulations (Hervías-Caimapo & Huffenberger, 2022), leaving instrumental noise, beam effects, and real-sky validation to future work.

2. Data: DUSTFILAMENTS Simulations

We train and evaluate with the DUSTFILAMENTS (Hervías-Caimapo & Huffenberger, 2022) simulations, a 3D filament-based model of Galactic dust polarization aligned with the Galactic magnetic field. The simulations are independent full-sky realizations and calibrated against *Planck* 353 GHz polarization maps (Planck Collaboration, 2016d).[2]. We generate 150 realizations at HEALPix resolution $N_{\rm side} = 1024$ (pixel size $\simeq 3.4$ arcmin) for the frequency channels 95, 145, 220, and 270 GHz relevant for SO (Simons Observatory Collaboration, 2019). Training data are 128×128 pixel patches ($\sim 7.3^\circ \times 7.3^\circ$ at this resolution) cut from full-sky maps using a cylindrical equal-area projection and cosine-apodization. We extract $\sim$1,280 patches per realization ($\sim$192,000 patches total), split into training (80%), validation (10%), and test (10%) with stratification by simulation index so patches from the same realization never appear in multiple splits. More DUSTFILAMENTS simulations details are found in App. A.2 and Hervías-Caimapo & Huffenberger (2022).

[2] https://github.com/chervias/DustFilaments

To each patch, we add an independent realization of primordial CMB B-mode polarization generated with CAMB (Lewis & Challinor, 2011), using best-fit *Planck* 2018 cosmological parameters (Planck Collaboration, 2020b) and tensor-to-scalar ratio $r = 0.008$ (without lensing). We do not include instrumental noise, beam smoothing, bandpass mismatch, or gravitational lensing. However, we note that while lensing would couple large- and small-scale B-modes in the late-time universe, this is considered as a foreground coupling and does not violate the statistical independence assumed for *primary CMB* signal-blind inputs. We defer quantifying systematic effects to future work.

3. Methods

3.1. Baseline: Internal Linear Combination (ILC)

We briefly review the Internal Linear Combination (ILC) method, which we use as a baseline for comparison with our U-Nets.[3] At a given frequency i, the CMB observation at line of sight $\hat{\mathbf{n}}$ is

$$B_i(\hat{\mathbf{n}}) = a_i S(\hat{\mathbf{n}}) + F_i(\hat{\mathbf{n}}), \tag{1}$$

where S is the frequency-independent primordial signal, F_i foreground contamination, and a_i are mixing coefficients[4]. The ILC reconstruction is a weighted sum of $B_i(\hat{\mathbf{n}})$ that minimizes foreground contribution,

$$\hat{B}^{\mathrm{ILC}}(\hat{\mathbf{n}}) = \sum_i w_i B_i(\hat{\mathbf{n}}), \quad \sum_i a_i w_i = 1. \tag{2}$$

where w_i is derived in App. A.1. The unit sum constraint implies that

$$\hat{B}^{\mathrm{ILC}} = \sum_i w_i (a_i S + F_i) = S + \sum_i w_i F_i, \tag{3}$$

exactly recovers the primordial signal $S(\hat{\mathbf{n}})$ with residual $\sum_i w_i F_i$ foregrounds uncorrelated with $S(\hat{\mathbf{n}})$. Hence, in its worst-case performance, ILC retains the full cosmological information needed for analysis. However, $\hat{B}^{\mathrm{ILC}}$ requires multi-frequency data since with only a single channel, the unit-response constraint fixes the weight to unity, so the ILC reconstruction is simply the observed map: $\hat{B}^{\mathrm{ILC}}(\hat{\mathbf{n}}) = B(\hat{\mathbf{n}})$. App. A.1 also explicitly shows that ILC is limited to second-order statistics.

3.2. Single-Frequency, Inter-scale Models

First, we investigate single-frequency foreground removal with different input combinations. To construct our inputs and target for this class of experiments, we exploit that primary CMB modes are independent across angular scale while polarized dust correlates small and large scales (Lazarian et al., 2023; Cho & Lazarian, 2002). At a given frequency, the observed CMB, $B(\hat{\mathbf{n}})$, can be separated in scale:

$$\begin{aligned} B_L(\hat{\mathbf{n}}) &= S_L(\hat{\mathbf{n}}) + F_L(\hat{\mathbf{n}}) \\ B_S(\hat{\mathbf{n}}) &= S_S(\hat{\mathbf{n}}) + F_S(\hat{\mathbf{n}}), \end{aligned} \tag{4}$$

where subscript L indicates large-scales $\ell < \ell_{\mathrm{cut}}$ and S indicates small-scales $\ell > \ell_{\mathrm{cut}}$. Since the inflationary imprint is expected to reside at $\ell < 200$, we take the scale cut at $\ell_{\mathrm{cut}} = 200$. This scale separation exploits the statistical independence of CMB modes across different multipoles while leveraging the coherent structure of Galactic dust foregrounds across scales. We have tested the robustness of this scale separation by varying the multipole threshold and find that our results are not significantly affected by changes in the scale cut. Nevertheless, this choice has not been optimized to maximize the information available to the neural network, and a systematic exploration of optimal angular-scale selection for B-mode information extraction is left for future work. Hence, $S_L(\hat{\mathbf{n}})$ is the primordial signal we wish to preserve. We use $B_S(\hat{\mathbf{n}})$ as signal-blind inputs since $B_S(\hat{\mathbf{n}})$ is uncorrelated with $S_L(\hat{\mathbf{n}})$. The target is $F_L(\hat{\mathbf{n}})$, the large-scale foregrounds that we estimate and subtract from observed $B_L(\hat{\mathbf{n}})$ to reconstruct the primordial signal:

$$S_L^{\mathrm{pred}}(\hat{\mathbf{n}}) = B_L(\hat{\mathbf{n}}) - \widehat{F_L(\hat{\mathbf{n}})}. \tag{5}$$

Because the input small-scale B-modes $B_S(\hat{\mathbf{n}})$ are uncorrelated with the target primordial signal $S_L(\hat{\mathbf{n}})$, this estimator is signal-preserving. In other words, if network inputs and training targets are both statistically independent of the large-scale primordial B-modes we wish to preserve, the learned correction cannot systematically absorb that signal, even if foreground simulations are imperfect. We perform all single-frequency experiments at 220 GHz, where dust foregrounds dominate (Planck Collaboration, 2020c). This pipeline is illustrated in Fig. 1.

We also concatenate $B_S(\hat{\mathbf{n}})$ inputs with complementary CMB measurements such as temperature, $T(\hat{\mathbf{n}})$, and E-mode polarization, $E(\hat{\mathbf{n}})$, still at single frequency (Fig. 2). Physical processes in the interstellar medium can generate T,E,B correlations, so these additional inputs can be used to estimate $F_L(\hat{\mathbf{n}})$ (Cukierman et al., 2023; Aghanim et al., 2016; Vacher et al., 2023). Under Gaussian inflationary statistics, primordial T and E are statistically independent of the large-scale primordial S_L, so all-scale T and E remain signal-blind inputs for that component.

3.3. Hybrid Multi-Frequency and Inter-scale Models

Second, we train hybrid U-Nets that leverage both inter-scale and multi-frequency correlations. As in (McCarthy

[3] Here, we use only spatial ILC but there exist ILC variants in other domains, see e.g. (Delabrouille et al., 2008).

[4] for CMB thermodynamic units, $a_i = 1$

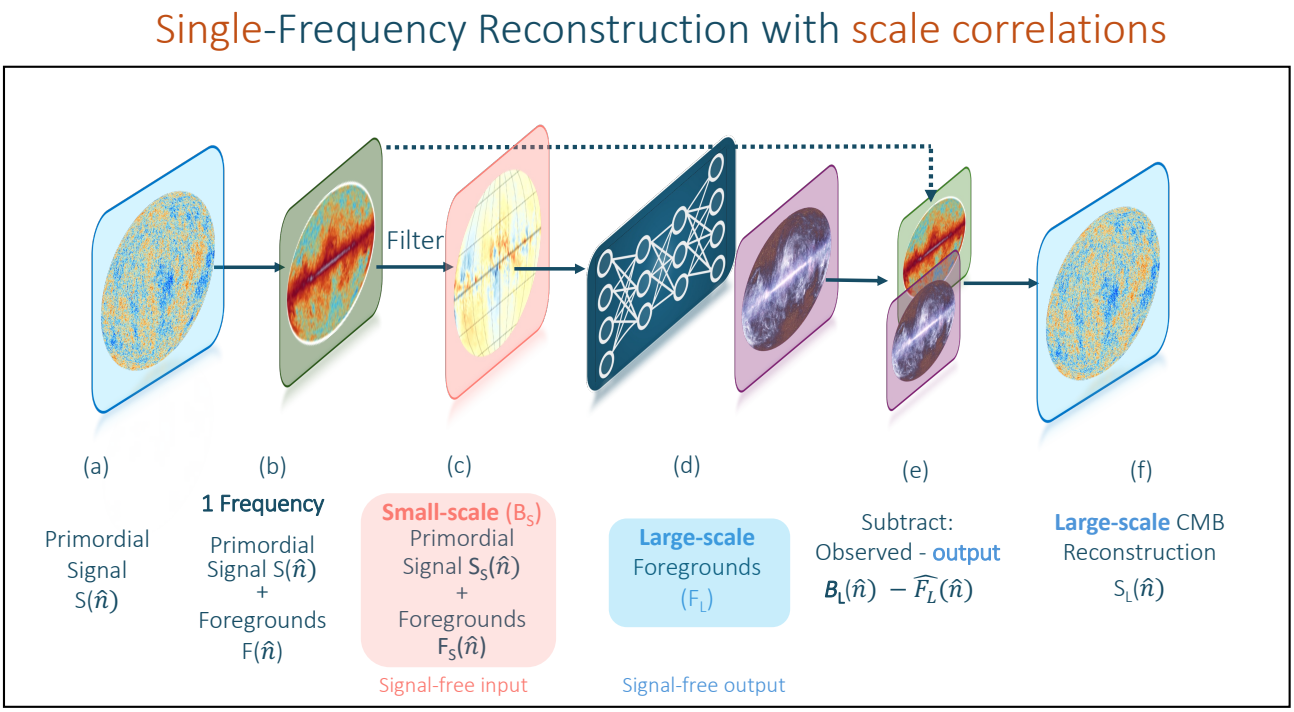


Figure 1. The primordial signal (a) is obscured by foregrounds $F(\hat{\mathbf{n}})$ in full CMB observation, $B(\hat{\mathbf{n}})$ (b). For a signal-preserving pipeline, we never pass full $B(\hat{\mathbf{n}})$ through the network. We split into $B_L(\hat{\mathbf{n}})$ and $B_S(\hat{\mathbf{n}})$ (c), train a U-Net to predict the large-scale foreground $\hat{F}_L(\hat{\mathbf{n}})$ using $B_S(\hat{\mathbf{n}})$ (d), and subtract $\hat{F}_L(\hat{\mathbf{n}})$ from $B(\hat{\mathbf{n}})$ (e) to reconstruct $S_L(\hat{\mathbf{n}})$ (f). Since $B_S(\hat{\mathbf{n}})$ is independent of S_L, the final reconstruction is unbiased. Inputs in (c) are optionally concatenated with $T(\hat{\mathbf{n}})$ and $E(\hat{\mathbf{n}})$ at the same frequency.

et al., 2024), we construct signal-blind multi-frequency inputs:

$$\begin{aligned}\hat{F}_i^{\mathrm{ILC}}(\hat{\mathbf{n}}) &= B_i(\hat{\mathbf{n}}) - \hat{B}^{\mathrm{ILC}}(\hat{\mathbf{n}}) \\ &= \big[S(\hat{\mathbf{n}}) + F_i(\hat{\mathbf{n}})\big] - \big[S(\hat{\mathbf{n}}) + \sum_j w_j F_j(\hat{\mathbf{n}})\big] \\ &= F_i(\hat{\mathbf{n}}) - \sum_j w_j F_j(\hat{\mathbf{n}}),\end{aligned} \tag{6}$$

where $i, j \in \{95, 145, 220, 270\}$ GHz and w_i are the ILC weights introduced in Sec. 3.1. These $\hat{F}_i^{\mathrm{ILC}}$ are linear foreground templates and are *signal-blind* since the primordial signal, $S(\hat{\mathbf{n}})$, cancels upon subtraction.

Unlike single-frequency models, the target is the *ILC residual* at 220 GHz, which contains the remaining non-Gaussian and anisotropic foregrounds that ILC failed to remove:

$$\begin{aligned}\Delta F^{\mathrm{ILC}}(\hat{\mathbf{n}}) &= \hat{B}^{\mathrm{ILC}}(\hat{\mathbf{n}}) - S(\hat{\mathbf{n}}) \\ &= \big[S(\hat{\mathbf{n}}) + \sum_j w_j F_j(\hat{\mathbf{n}})\big] - S(\hat{\mathbf{n}}) \\ &= \sum_j w_j F_j(\hat{\mathbf{n}}).\end{aligned} \tag{7}$$

Here ΔF^{ILC} is the multifrequency foreground component left in the ILC map (linear combination of F_i with weights w_i). In simulations we form regression targets using the known primordial realization $S(\hat{\mathbf{n}})$ as in Eq. (8); *at inference*, only observed maps $B_i(\hat{\mathbf{n}})$ enter $\hat{B}^{\mathrm{ILC}}$ and the network inputs. The trained network outputs $\widehat{\Delta F^{\mathrm{ILC}}}$, and we improve the large-scale ILC reconstruction via

$$\begin{aligned}S_L^{\mathrm{pred}}(\hat{\mathbf{n}}) &= B_L^{\mathrm{ILC}}(\hat{\mathbf{n}}) - \widehat{\Delta F^{\mathrm{ILC}}}_L(\hat{\mathbf{n}}) \\ &= \sum_i w_i B_i(\hat{\mathbf{n}}) - \widehat{\Delta F^{\mathrm{ILC}}}_L(\hat{\mathbf{n}}) \\ &= S_L(\hat{\mathbf{n}}) + \sum_i w_i F_i(\hat{\mathbf{n}}) - \widehat{\Delta F^{\mathrm{ILC}}}_L(\hat{\mathbf{n}}).\end{aligned} \tag{8}$$

which uses only observables (weighted sums of observed B_i and the learned correction). As shown, $S_L^{\mathrm{pred}}(\hat{\mathbf{n}})$ preserves $S_L(\hat{\mathbf{n}})$; training minimizes error on ΔF^{ILC} so that $\widehat{\Delta F^{\mathrm{ILC}}}$ approaches $\sum_i w_i F_i$. Similar to Sec. 3.2, we train U-Nets with different input concatenations: (i) $\hat{F}_i^{\mathrm{ILC}}$ (4 channels), (ii) $\hat{F}_i^{\mathrm{ILC}} + B_{S,i}$ (8 channels), and (iii) $\hat{F}_i^{\mathrm{ILC}} + B_{S,i} + T_i + E_i$ (16 channels).

3.3.1. Model Architecture

For both single-frequency and hybrid models we use a U-Net (Ronneberger et al., 2015). Its encoder–decoder architecture, with skip connections, couples small-scale inputs

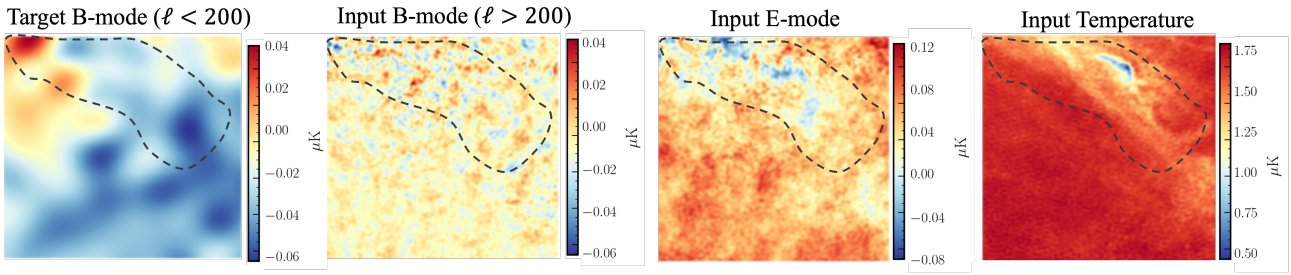


Figure 2. Multi-channel input for single-frequency foreground prediction. (a) Target large-scale B-mode foregrounds (F_L^B, $\ell < 200$). (b) Input small-scale B-modes (B_S, $\ell > 200$). (c) Input E-modes (E, all ℓ). (d) Input temperature (T, all ℓ). The dashed region shows shared morphology between inputs and target that the U-Net can leverage for foreground prediction.

to large-scale targets while preserving spatial resolution. Layer widths, pooling/upsampling, normalization, training hyperparameters, and data augmentation are specified in App. A.3. We did not optimize over network architecture for foreground estimation as it is beyond the scope of this work, but leave this for future investigation.

3.4. Evaluation Metrics

All networks are trained to minimize pixel-wise mean squared error (MSE),

$$L_{\rm MSE} = \mathbb{E}\,\|\hat{Y}(\hat{\mathbf{n}}_p) - Y(\hat{\mathbf{n}}_p)\|^2\,, \tag{9}$$

where $\mathbb{E}$ is the average over training patches, $\hat{Y}(\hat{\mathbf{n}}_p)$ is the network output, and $Y(\hat{\mathbf{n}}_p)$ is the target— large-scale foreground F_L for single-frequency models or ILC residual $\Delta F^{\rm ILC}$ for hybrid models.

For evaluation, we report two metrics. First, the MSE between the predicted and truth averaged across test patches N_p,

$$\mathrm{MSE}_Y = \frac{1}{N_p}\sum_{p=1}^{N_p}\big(\hat{Y}(\hat{\mathbf{n}})_p - Y(\hat{\mathbf{n}})_p\big)^2\,, \tag{10}$$

with $(\hat{Y},\,Y)$ set according to the task: foreground prediction $(\hat{Y},\,Y) = (\hat{F}_L,\,F_L)$ for single-frequency models; residual prediction $(\hat{Y},\,Y) = (\widehat{\Delta F}^{\rm ILC},\,\Delta F^{\rm ILC})$ for hybrid models; and S_L reconstruction $(\hat{Y},\,Y) = (\hat{S}_L,\,S_L)$ when comparing cleaned maps to the true primordial realization. We compare the F_L estimation MSE to the uncleaned observed map in the single-channel limit ILC ($B_L^{ILC} = B_L$), for fair comparison of methods that access only single-frequency information. For hybrid models that have access to multi-frequency information, we compare S_L reconstruction errors to standard ILC. For unbiased estimators of Gaussian primordial B-modes S_L, MSE_{S_L} is a proxy for residual foreground *variance* and is analogous to the variance minimized by standard ILC (Eriksen et al., 2004).

Second, we report harmonic correlations using the normalized cross-power spectrum, $\rho(\ell)$, between predicted and true foregrounds (or reconstructed and true CMB). We work in the flat-sky limit and compute binned 2D Fourier cross-spectra C_ℓ^{XY} for test patches, X and Y, with bin widths $\Delta\ell = 50$. We divide the cross-power spectrum by the geometric mean of the auto-power spectra to yield the normalized harmonic correlation coefficient $\rho(\ell)$,

$$\rho_{X,Y}(\ell) = \frac{C_\ell^{XY}}{\sqrt{C_\ell^{XX}C_\ell^{YY}}}. \tag{11}$$

Unless otherwise stated, we report the mean $\rho(\ell)$ over large-scale multipoles ($\ell < 200$) averaged across the test-set. We also compute the 16th and 84th percentiles of the distribution across patches at each multipole bin, providing uncertainty bands that reflect the patch-to-patch variation in reconstruction quality. Unlike map-level metrics, $\rho(\ell)$ quantifies scale-dependent agreement in harmonic space and is commonly used in CMB analysis.

4. Results: Single Frequency Models

Table 1 shows the performance metrics for single-frequency U-Net configurations. We compare them to the single-channel ILC reconstruction, which reduces to the uncleaned observed map. This provides a fair reference for testing whether inter-scale and cross-component information is useful for foreground estimation at single frequency. Evidently, all U-Nets improve upon this baseline by factors of $\sim$ 4–7 in MSE. Using only small-scale B-modes (B_S) reaches $\rho_\ell = 0.51$, implying that inter-scale foreground structure is predictive even without multi-frequency data. Adding temperature and E-mode channels further improves fidelity, with the best model ($T + E + B_S$) achieving 2.09×10^{-4} MSE and $\rho_\ell = 0.79$. Figure 3a compares absolute ρ_ℓ between the B_S-only and the maximal $T + E + B_S$ models, showing the test-set means with two-tail uncertainties per ℓ-bin. Figure 3b includes the ratio of ρ_ℓ for all single-frequency U-Nets, with respect to B_S-only at the same multipole. Here, ratios >1 quantify improved correlation relative to B_S-only. Together, these panels show that the improvements summarized in Table 1 persist across all large-scales $\ell < 200$.

Figure 4 highlights qualitatively improved dust morphology prediction across representative test patches. Compared with Philcox et al. (2018), our harmonic correlations are higher, but note that their estimator is constructed from the hexadecapole in the 2D dust power spectrum, and their simulations and sky coverage differ from ours. We include this comparison because it is the analogous analytic single-frequency estimator for inter-scale Galactic dust estimation. The ℓ-averaged ρ_ℓ for Philcox+18 in Table 1 is estimated from their Fig. 9.

Although these results are preliminary, they suggest that single-frequency observations with multi-component $T{+}E{+}B_S$ inputs can substantially improve large-scale foreground prediction in the idealized case of noiseless CMB measurements. However, future tests are needed to assess whether this can improve CMB reconstructions under more realistic settings and achieve tighter cosmological constraints. Moreover, as we know from the strong performance of ILC methods, multi-frequency correlations provide substantial information for foreground removal. We therefore explore a hybrid approach that combines multi-frequency and inter-scale correlations in the next section.

5. Results: Hybrid Models

Hybrid models test whether inter-scale and cross-component correlations are complementary or redundant with the multi-frequency information used by standard ILC. Relative to ILC, a U-Net trained on $\hat{F}_i^{ILC}$ alone already reduces MSE_{S_L} from 1.42×10^{-3} to 1.51×10^{-6}, demonstrating significant anisotropic and non-Gaussian information learned by the U-Net that the ILC cannot access. Augmenting with small-scale B-modes yields a further $\sim 2.7\times$ improvement (5.61×10^{-7}), and adding T/E channels reaches the best value, 3.66×10^{-7}. Interestingly, the addition of $B_{S,i}$ improves the harmonic correlation at smaller scales $\ell \sim 60 - 200$, while adding $T_i + E_i$ only increases low-ℓ correlations without providing significant improvement at smaller scales (see Fig. 6). These results indicate that multi-frequency, inter-scale, and T/E information are complementary rather than redundant for residual-foreground prediction. Additional examples of the U-Nets' ability to learn and remove anisotropic foreground features are discussed in App. A.4.

Overall, the hybrid model yields an improvement of about three orders of magnitude in MSE_{S_L} relative to ILC while maintaining its signal-preserving property.

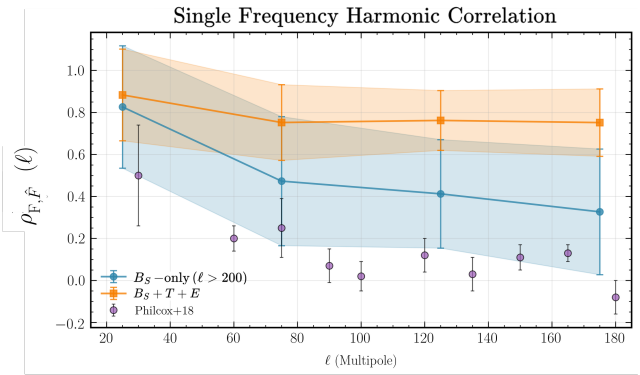
6. Conclusions

Unlike traditional CMB foreground removal with internal linear combination (ILC) methods, machine learning (ML) is more capable of capturing complex inter-scale correlations and learning the non-trivial Galactic physics that couple different scales in dust polarization. However, ML models that have direct access to the primordial signal during training can mistake it for foregrounds and accidentally remove it. This can introduce bias into downstream cosmological inference, making ML-based CMB reconstructions unreliable and physically uninterpretable. We therefore construct signal-preserving CNNs that only have access to signal-blind inputs and are trained to predict signal-blind outputs. With both being independent of the target primordial B-modes enables, the model can safely learn foreground correlations without leakage (and hence potential distortion) of the primordial signal into its non-linear pipeline.

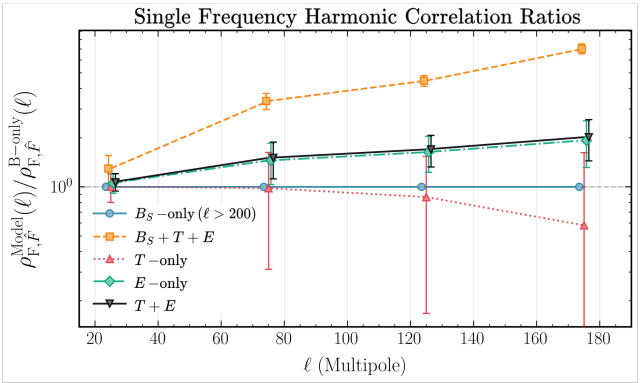
Within this framework, we present new single-frequency dust foreground removal that leverages inter-scale correlations and T/E information. Our key result is that, using data at just a single frequency, CNNs can estimate dust foreground morphology of high fidelity on large angular scales, achieving a significant harmonic correlation with the truth compared to the corresponding analytic estimator of (Philcox et al., 2018). In particular, our best single-frequency model ($T{+}E{+}B_S$) reaches $\rho_\ell \simeq 0.79$, compared with $\rho_\ell \simeq 0.51$ for B_S-only and $\rho_\ell \simeq 0.17$ for the analytic Philcox+18 estimator. Our single-frequency models also surpass the baseline single-channel ILC, indicating that there is significant information at only one frequency for cleaning foregrounds.

We also explore hybrid models that leverage both multi-frequency and inter-scale correlations, giving cleaned CMB patches with near-unity mean harmonic correlations of $1 - \rho(\ell < 200) = 2.71 \times 10^{-5}$ with the truth. This is compared with $1 - \rho(\ell) < 200 \simeq 7.3 \times 10^{-3}$ for standard ILC. This demonstrates the abundance of useful non-Gaussian structure and inter-scale correlations learned by the network that the ILC misses. These results support the use of data-driven, scale-constrained foreground estimation as a promising complement to multi-frequency and analytic estimators for reliable large-scale CMB B-mode cleaning.

In future work, tests with various foreground models (e.g Thorne & et al., 2021; Vansyngel et al., 2018), instrumental effects, and CMB experiment systematics are crucial for implementing our methods on real-sky observations. However, with our signal-preserving framework, mis-specified dust simulations do not bias the preserved primordial B-modes. Namely, in the worst-case scenario, if the network learns no useful correction, one is left with large uncorrelated foreground contamination while the primordial signal remains fully in tact. In such case, one defaults to ILC-level cleaning without losing any cosmological information. Hence, imperfect training physics limits gains over ILC, not signal integrity.



(a) Harmonic correlation ρ_ℓ for single-frequency F_L prediction with different inputs. Curves show test-set means; shaded bands denote 16th–84th percentiles across patches. We also show the analytic estimator from Philcox et al. (2018) for reference, estimated from their Fig. 9. Our B_S-only U-Net outperforms the analytic estimator at all scales. Adding $T + E$ improves correlation from $\rho_\ell \simeq 0.51$ to 0.79.



(b) Relative gain in harmonic correlation versus the B_S-only model for single-frequency inputs. Ratios above unity indicate improvement over B_S-only at given multipole. Inputs containing E provide the largest gains across most of $\ell < 200$.

Table 1. Single-frequency large-scale F_L prediction. $T{+}E{+}B_S$ performs best, and all U-Net variants outperform single-channel ILC. In the single-frequency limit, ILC reduces to the uncleaned observed map ($w_1 = 1$). Uncertainties report mean ± standard deviation across test patches where available.

Single-Frequency Method	**MSE** $[\mu\mathrm{K}^2]$	**Harmonic Correlation** ρ_ℓ
No Foreground Cleaning	1.40×10^{-3}	—
Philcox+18 Analytic ($\ell > 90$)	—	$0.17^\dagger$
B_S-only	$3.44 \times 10^{-4} \pm 2.51 \times 10^{-4}$	0.51
T-only	$3.47 \times 10^{-4} \pm 2.36 \times 10^{-4}$	0.46
E-only	$2.53 \times 10^{-4} \pm 1.89 \times 10^{-4}$	0.71
$T{+}E$-only	$2.35 \times 10^{-4} \pm 1.70 \times 10^{-4}$	0.74
$T{+}E{+}B_S$	$2.09 \times 10^{-4} \pm 1.59 \times 10^{-4}$	0.79

†Estimated from Fig. 9 of Philcox et al. (2018).

7. Software and Data

Simulations use DUSTFILAMENTS (Hervías-Caimapo & Huffenberger, 2022), publicly available at `https://github.com/chervias/DustFilaments`. Primordial CMB realizations are generated with CAMB (Lewis & Challinor, 2011). Models are implemented and trained in PyTorch. We provide implementation details in Sec. 2 and App. A.3 sufficient to reproduce the main experiments. Code to reproduce these results is available at `https://github.com/HelenShao/Interscale-CMB-Reconstruction-Code`.

Acknowledgements

We thank the anonymous reviewers for their helpful feedback that greatly improved the quality of the text. We thank Oliver Philcox, Jacques Delabrouille, Clara Vergès, and Alexandre Adler for insightful discussions that helped clarify the paper draft. We also thank Irene Abril-Cabezas, Carmen Embil Villagra, Louis Legrand, and many others for their helpful comments during early discussions of this work.

Impact Statement

This paper presents work whose goal is to advance the field of Cosmology and Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Aghanim, N. et al. Planck intermediate results - XLIV. Structure of the Galactic magnetic field from dust polarization maps of the southern Galactic cap. *Astron. Astrophys.*, 596:A105, 2016. doi: 10.1051/0004-6361/201628636.

Bachar, L. G., Dvorkin, C., and Supanitsky, A. D. Robustness of neural networks for cmb polarization foreground

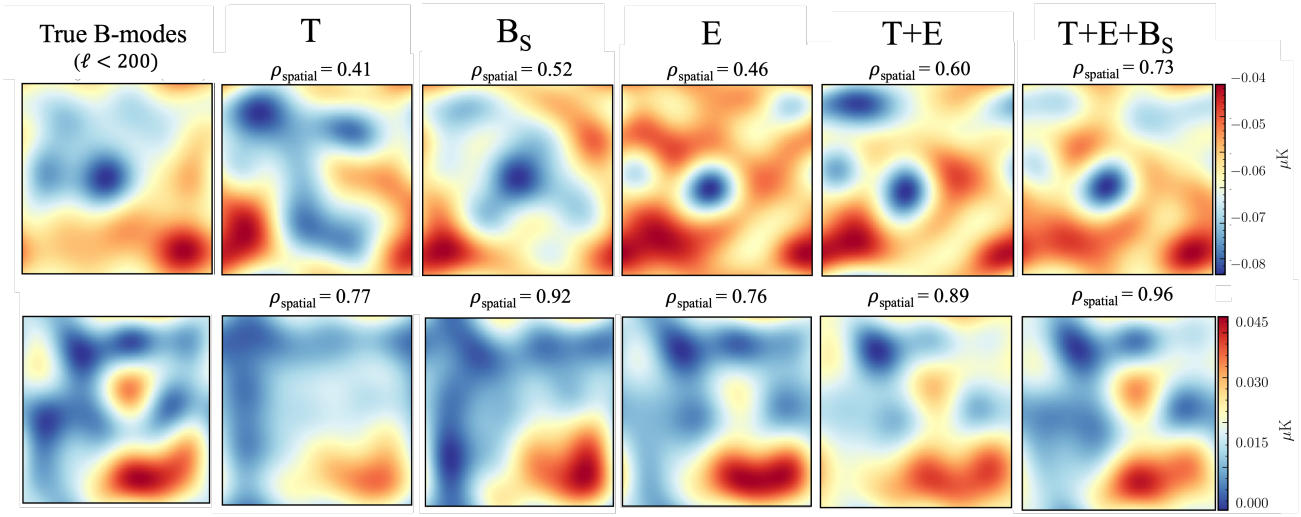


Figure 4. Single-frequency reconstructions with different inputs. Each row corresponds to one target F_L patch; columns show predictions from models using T, B_S, E, $T + E$, and $T + E + B_S$. Harmonic correlation coefficients with the target are shown above each reconstruction. The examples illustrate that E and multi-component inputs generally recover morphology more reliably than T-only or B_S-only inputs.

Table 2. Hybrid multi-frequency/inter-scale S_L reconstruction performance. Relative to standard spatial ILC (4 frequency channels), all hybrid models reduce MSE_{S_L} by orders of magnitude, with the strongest reduction in the 16-channel configuration. MSE_{S_L} improvements are measured relative to standard multi-frequency ILC in noiseless simulations.

Multi-Frequency or Hybrid Method	$\mathrm{MSE}_{\mathbf{S_L}}$ $[\mu K^2]$	ρ_ℓ
Spatial ILC (4-channel)	$1.42 \times 10^{-3} \pm 3.82 \times 10^{-3}$	0.9927
$\hat{F}_i^{\mathrm{ILC}}$-only (4-channel)	$1.51 \times 10^{-6} \pm 1.38 \times 10^{-6}$	0.99989
$\hat{F}_i^{\mathrm{ILC}} + B_{S,i}$ (8-channel)	$5.61 \times 10^{-7} \pm 3.40 \times 10^{-7}$	0.99996
$\hat{F}_i^{\mathrm{ILC}} + B_{S,i} + T_i + E_i$ (16-channel)	$3.66 \times 10^{-7} \pm 2.86 \times 10^{-7}$	0.99997

removal, 2026. URL https://arxiv.org/abs/2603.12364.

Bennett, C. L., Smoot, G. F., Hinshaw, G., Wright, E. L., Kogut, A., de Amici, G., Meyer, S. S., Weiss, R., Wilkinson, D. T., Gulkis, S., Janssen, M., Boggess, N. W., Cheng, E. S., Hauser, M. G., Kelsall, T., Mather, J. C., Moseley, Jr., S. H., Murdock, T. L., and Silverberg, R. F. Preliminary Separation of Galactic and Cosmic Microwave Emission for the COBE Differential Microwave Radiometers. *Astrophys. J. Lett.*, 396:L7, September 1992. doi: 10.1086/186505.

Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K. M., Hill, R. S., Gold, B., Halpern, M., Komatsu, E., Nolta, M. R., Page, L., Spergel, D. N., Wollack, E., Dunkley, J., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., and Wright, E. L. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *Astrophys. J. Suppl.*, 208 (2):20, October 2013. doi: 10.1088/0067-0049/208/2/20.

Bicep/KECK Collaboration. Bicep/keck xv: The bicep3 cosmic microwave background polarimeter and the first three-year data set. *The Astrophysical Journal*, 927 (1):77, March 2022. ISSN 1538-4357. doi: 10.3847/1538-4357/ac4886. URL http://dx.doi.org/10.3847/1538-4357/ac4886.

Caldwell, R. R., Hirata, C., and Kamionkowski, M. Dust-polarization Maps and Interstellar Turbulence. *Astrophys. J.*, 839(2):91, April 2017. doi: 10.3847/1538-4357/aa679c.

Cho, J. and Lazarian, A. Compressible magnetohydrodynamic turbulence: Mode coupling, anisotropy, and scaling laws. *Physical Review Letters*, 88:245001, 2002.

Cukierman, A. J., Clark, S. E., and Halal, G. Magnetic Misalignment of Interstellar Dust Filaments. *Astrophys. J.*, 946(2):106, April 2023. doi: 10.3847/1538-4357/acb0c4.

Delabrouille, J., Cardoso, J.-F., Le Jeune, M., Betoule, M., Fay, G., and Guilloux, F. A full sky, low foreground, high resolution cmb map from wmap. *Astronomy and Astrophysics*, 493(3):835–857, November 2008. ISSN 1432-0746. doi: 10.1051/

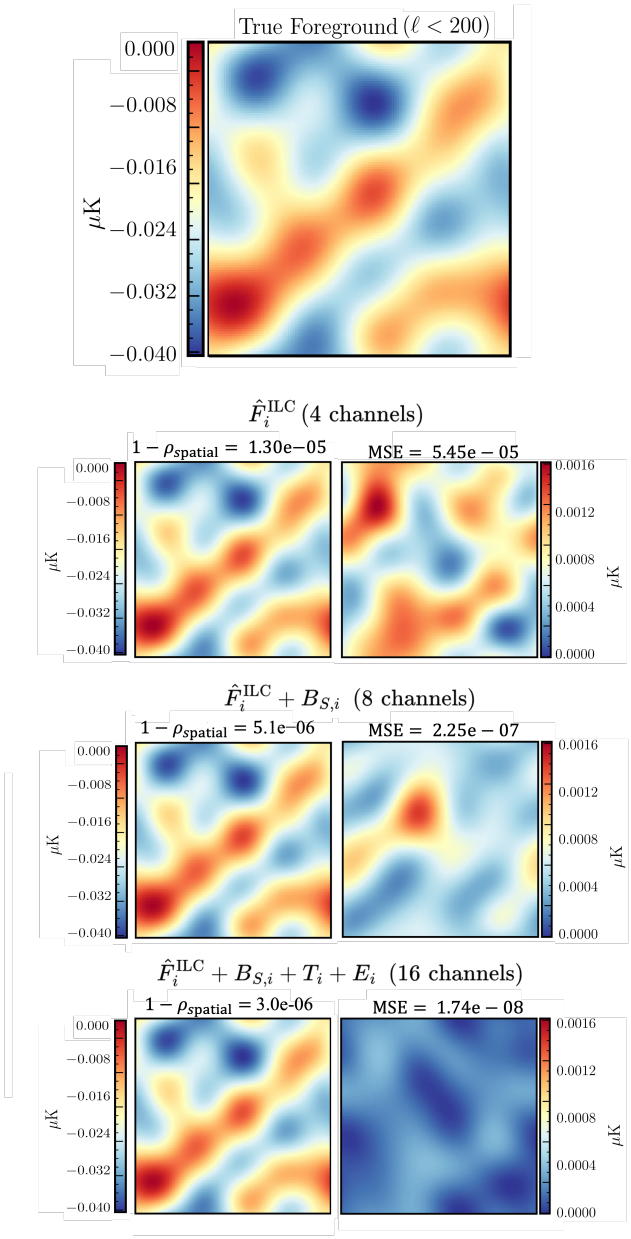


Figure 5. Hybrid-model test patch across model variants. Rows show augmented inputs ($\hat{F}_i^{\rm ILC}$-only, $+\,B_{S,i}$, and $+\,B_{S,i} + T_i + E_i$); columns show (left) predicted F_L at 220 GHz and (right) residuals relative to truth. Residual structure substantially decreases with more inputs, showing abundance of inter-scale and cross-component information that complements multi-frequency correlations.

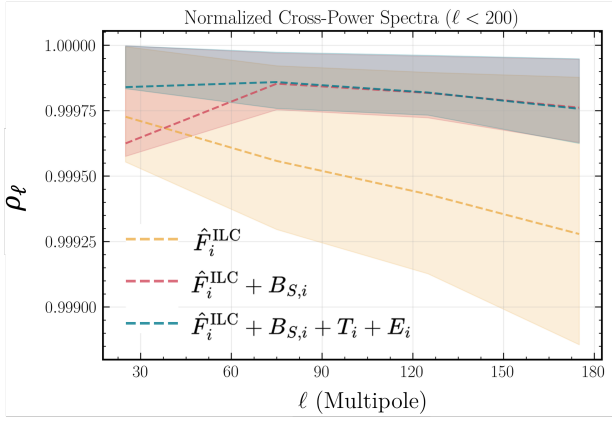


Figure 6. Harmonic correlation ρ_ℓ for hybrid-model S_L reconstructions. Dashed lines denote test-set means and shaded regions show 16th–84th percentiles across patches. Adding $B_{S,i}$ to $\hat{F}_i^{\rm ILC}$ reduces $1 - \rho_\ell$ from 1.10×10^{-4} to 4.27×10^{-5}, and adding T_i, E_i further reduces it to 2.71×10^{-5}.

0004-6361:200810514. URL http://dx.doi.org/10.1051/0004-6361:200810514.

Eriksen, H. K., Banday, A. J., Gorski, K. M., and Lilje, P. B. On foreground removal from the wilkinson microwave anisotropy probe data by an internal linear combination method: Limitations and implications. *The Astrophysical Journal*, 612(2):633–646, 2004. ISSN 1538-4357. doi: 10.1086/422807. URL http://dx.doi.org/10.1086/422807.

Grünwald, P. and van Ommen, T. Inconsistency of bayesian inference for misspecified linear models, and a proposal for repairing it, 2018. URL https://arxiv.org/abs/1412.3730.

Hervías-Caimapo, C. and Huffenberger, K. M. Full-sky, arcminute-scale, 3d models of galactic microwave foreground dust emission based on filaments. *The Astrophysical Journal*, 928(1):65, March 2022. ISSN 1538-4357. doi: 10.3847/1538-4357/ac54b2. URL http://dx.doi.org/10.3847/1538-4357/ac54b2.

Kamionkowski, M. and Kovetz, E. D. Statistical diagnostics to identify galactic foregrounds in b-mode maps. *Physical Review Letters*, 113(19), November 2014. ISSN 1079-7114. doi: 10.1103/physrevlett.113.191303. URL http://dx.doi.org/10.1103/PhysRevLett.113.191303.

Kamionkowski, M., Kosowsky, A., and Stebbins, A. A probe of primordial gravity waves and vorticity. *Physical Review Letters*, 78:2058–2061, 1997.

Kandel, D., Lazarian, A., and Pogosyan, D. Can the observed e/b ratio for dust galactic foreground be explained by sub-alfvénic turbulence? *Monthly No-*

tices of the Royal Astronomical Society: Letters, 472 (1):L10–L14, August 2017. ISSN 1745-3933. doi: 10.1093/mnrasl/slx128. URL http://dx.doi.org/10.1093/mnrasl/slx128.

Keck Collaboration. Bicep/keck xx: Component-separated maps of polarized cmb and thermal dust emission using planck and bicep/keck observations through the 2018 observing season, 2026. URL https://arxiv.org/abs/2509.21648.

Kritsuk, A. G., Flauger, R., and Ustyugov, S. D. Dust-polarization maps for local interstellar turbulence. *Physical Review Letters*, 121(2), July 2018. ISSN 1079-7114. doi: 10.1103/physrevlett.121.021104. URL http://dx.doi.org/10.1103/PhysRevLett.121.021104.

Lazarian, A., Xu, S., and Hu, Y. Cosmic ray propagation in turbulent magnetic fields, 2023. URL https://arxiv.org/abs/2304.02684.

Lewis, A. and Challinor, A. CAMB: Code for Anisotropies in the Microwave Background. Astrophysics Source Code Library, record ascl:1102.026, February 2011.

LiteBIRD Collaboration. Probing cosmic inflation with thelitebirdcosmic microwave background polarization survey. *Progress of Theoretical and Experimental Physics*, 2023(4), November 2022. ISSN 2050-3911. doi: 10.1093/ptep/ptac150. URL http://dx.doi.org/10.1093/ptep/ptac150.

M, O., Jaiswal, S., Das, S., and Parattu, K. M. Deep learning for cmb foreground removal and beam deconvolution: A u-net gan approach, 2025. URL https://arxiv.org/abs/2509.00139.

McCarthy, F., Hill, J. C., Coulton, W. R., and Hogg, D. W. Signal-preserving cmb component separation, 2024.

Petroff, M. A., Addison, G. E., Bennett, C. L., and Weiland, J. L. Full-sky cosmic microwave background foreground cleaning using machine learning. *The Astrophysical Journal*, 903(2):104, November 2020. ISSN 1538-4357. doi: 10.3847/1538-4357/abb9a7. URL http://dx.doi.org/10.3847/1538-4357/abb9a7.

Philcox, O. H. E., Sherwin, B. D., and van Engelen, A. Detection and removal of b-mode dust foregrounds with signatures of statistical anisotropy. *Monthly Notices of the Royal Astronomical Society*, 479(4):5577–5595, July 2018. ISSN 1365-2966. doi: 10.1093/mnras/sty1769. URL http://dx.doi.org/10.1093/mnras/sty1769.

Planck Collaboration. Planck2015 results: X. diffuse component separation: Foreground maps. *Astronomy and Astrophysics*, 594:A10, September 2016a. ISSN 1432-0746. doi: 10.1051/0004-6361/201525967. URL http://dx.doi.org/10.1051/0004-6361/201525967.

Planck Collaboration. Planck intermediate results. XXX. The. *AAS*, 596:A103, 2016b.

Planck Collaboration. Planck intermediate results. XXXVIII. E-. *AAS*, 596:A104, 2016c.

Planck Collaboration. Planck intermediate results: Xlviii. disentangling galactic dust emission and cosmic infrared background anisotropies. *Astronomy & Astrophysics*, 596:A109, December 2016d. ISSN 1432-0746. doi: 10.1051/0004-6361/201629022. URL http://dx.doi.org/10.1051/0004-6361/201629022.

Planck Collaboration. Planck 2018 results. IV. Diffuse. *AAS*, 641:A4, 2020a.

Planck Collaboration. Planck 2018 results. VI. Cosmological. *AAS*, 641:A6, 2020b.

Planck Collaboration. Planck 2018 results. XI. Polarized dust foregrounds. *Astron. Astrophys.*, 641:A11, September 2020c. doi: 10.1051/0004-6361/201832618.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation, 2015.

Seljak, U. and Zaldarriaga, M. Signature of gravity waves in the polarization of the microwave background. *Physical Review Letters*, 78(11):2054–2057, March 1997. ISSN 1079-7114. doi: 10.1103/physrevlett.78.2054. URL http://dx.doi.org/10.1103/PhysRevLett.78.2054.

Simons Observatory Collaboration. The Simons Observatory: Science goals and forecasts. *JCAP*, 2019(02):056, 2019. doi: 10.1088/1475-7516/2019/02/056.

The Simons Observatory Collaboration. The Simons Observatory: forecasted constraints on primordial gravitational waves with the expanded array of Small Aperture Telescopes. *arXiv e-prints*, art. arXiv:2512.15833, December 2025. doi: 10.48550/arXiv.2512.15833.

Thorne, B. and et al. The Python Sky Model 3:. *Monthly Notices of the Royal Astronomical Society*, 504(4):5646–5662, 2021.

Vacher, L., Aumont, J., Boulanger, F., Montier, L., Guillet, V., Ritacco, A., and Chluba, J. Frequency dependence of the thermal dust e/b ratio and eb correlation: Insights from the spin-moment expansion. *Astronomy & Astrophysics*,

672:A146, April 2023. ISSN 1432-0746. doi: 10.1051/0004-6361/202245292. URL http://dx.doi.org/10.1051/0004-6361/202245292.

Vansyngel, F., Boulanger, F., Ghosh, T., Wandelt, B., Aumont, J., Bracco, A., Levrier, F., Martin, P. G., and Montier, L. Statistical simulations of the dust foreground to cosmic microwave background polarization (corrigendum). *Astronomy & Astrophysics*, 618:C4, October 2018. ISSN 1432-0746. doi: 10.1051/0004-6361/201629992e. URL http://dx.doi.org/10.1051/0004-6361/201629992e.

WMAP Collaboration. First-year wilkinson microwave anisotropy. *APJ*, 148(1):97, 2003.

Yan, Y.-P., Li, S.-Y., Wang, G.-J., Zhang, Z., and Xia, J.-Q. Cmbfscnn: Cosmic microwave background polarization foreground subtraction with convolutional neural network, 2024. URL https://arxiv.org/abs/2406.17685.

Zaldarriaga, M. and Seljak, U. All-sky analysis of polarization in the microwave background. *Physical Review D*, 55(4):1830–1840, February 1997. ISSN 1089-4918. doi: 10.1103/physrevd.55.1830. URL http://dx.doi.org/10.1103/PhysRevD.55.1830.

A. Supplementary Material

A.1. ILC derivation and signal-preserving property

We follow the derivation presented in (Eriksen et al., 2004) for the spatial ILC weights. Restating the observation model from Eqn. 1

$$B_i(\hat{\mathbf{n}}) = a_i\, S(\hat{\mathbf{n}}) + F_i(\hat{\mathbf{n}}), \tag{12}$$

where S is the frequency-independent primordial signal, a_i are frequency-dependent mixing coefficients, and F_i denotes foreground contributions. The ILC estimator with unbiasedness constraint is

$$\hat{B}^{\mathrm{ILC}}(\hat{\mathbf{n}}) = \sum_i w_i B_i(\hat{\mathbf{n}}), \quad \sum_i a_i w_i = 1. \tag{13}$$

Hence, the variance-minimization problem is

$$\min_{\mathbf{w}}\ \mathbf{w}^{\mathsf{T}}\mathbf{C}\,\mathbf{w} \quad \text{subject to} \quad \mathbf{a}^{\mathsf{T}}\mathbf{w} = 1, \tag{14}$$

where $\mathbf{C} = \langle \mathbf{B}\,\mathbf{B}^{\mathsf{T}} \rangle$ is the observed covariance matrix in frequency space. Using a Lagrange multiplier gives the standard ILC solution. Define

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w}^{\mathsf{T}}\mathbf{C}\,\mathbf{w} - \lambda(\mathbf{a}^{\mathsf{T}}\mathbf{w} - 1). \tag{15}$$

Taking the first derivative respect to $\mathbf{w}$ gives the minimizing solution

$$\nabla_{\mathbf{w}}\mathcal{L} = 2\mathbf{C}\mathbf{w} - \lambda\mathbf{a} = 0 \ \Rightarrow\ \mathbf{w} = \frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{a}. \tag{16}$$

Imposing $\mathbf{a}^{\mathsf{T}}\mathbf{w} = 1$ yields $\lambda/2 = 1/(\mathbf{a}^{\mathsf{T}}\mathbf{C}^{-1}\mathbf{a})$, hence

$$w_i^{\mathrm{ILC}} = \frac{\sum_{j=1}^{k}(C^{-1})_{ij}a_j}{\sum_{m,n=1}^{k} a_m (C^{-1})_{mn} a_n}. \tag{17}$$

Evidently, the ILC weights are determined entirely by the observed frequency covariance C, so the method uses only second-order information. It is therefore optimal for Gaussian observations (where covariance fully characterizes the fields) but generally suboptimal for non-Gaussian, anisotropic Galactic foregrounds (Caldwell et al., 2017).

A.2. DUSTFILAMENTS Simulation Details

The DUSTFILAMENTS simulations model Galactic dust distribution by populating the foreground sky with dust filaments aligned to the Galactic magnetic field. The filament population naturally produces correlated structures on all scales. At scales ($\ell \lesssim 50$), where the model uses a Planck-based template filled in from observations, the frequency extrapolation follows a standard modified blackbody law with spatially-varying β_{dust} and T_{dust} parameters from Planck GNILC. The fiducial simulations are normalized to reproduce key statistical properties of *Planck* 353 GHz polarization maps (Planck Collaboration, 2016d), including angular power spectra and non-Gaussian structure, following the calibration procedure described in (Hervías-Caimapo & Huffenberger, 2022).

There are two provided options for simulating the Galactic plane. The first inserts the *Planck* GNILC dust template, constructed from the *Planck* DR2 maps (Planck Collaboration, 2016d), instead of populating that region with simulated filaments. The second isotropically samples filament centers in the Galactic-plane region so that those sightlines are populated in the same way as the rest of the sky.

Under the second option, the simulation does not impose an external Galactic-plane template, so filaments in this region still evolve under the magnetic-field model. We adopt this second option in all experiments: it preserves small-scale structure in the plane whereas the GNILC-template option does not simulate small-scale Galactic-plane emission (Hervías-Caimapo & Huffenberger, 2022). Plane regions are typically masked in real analyses due to strong contamination; this choice is nonetheless better matched to learning inter-scale dust morphology than inserting the GNILC template. Filament population, magnetic-field parameters, and spectral energy distribution settings follow Hervías-Caimapo & Huffenberger (2022).

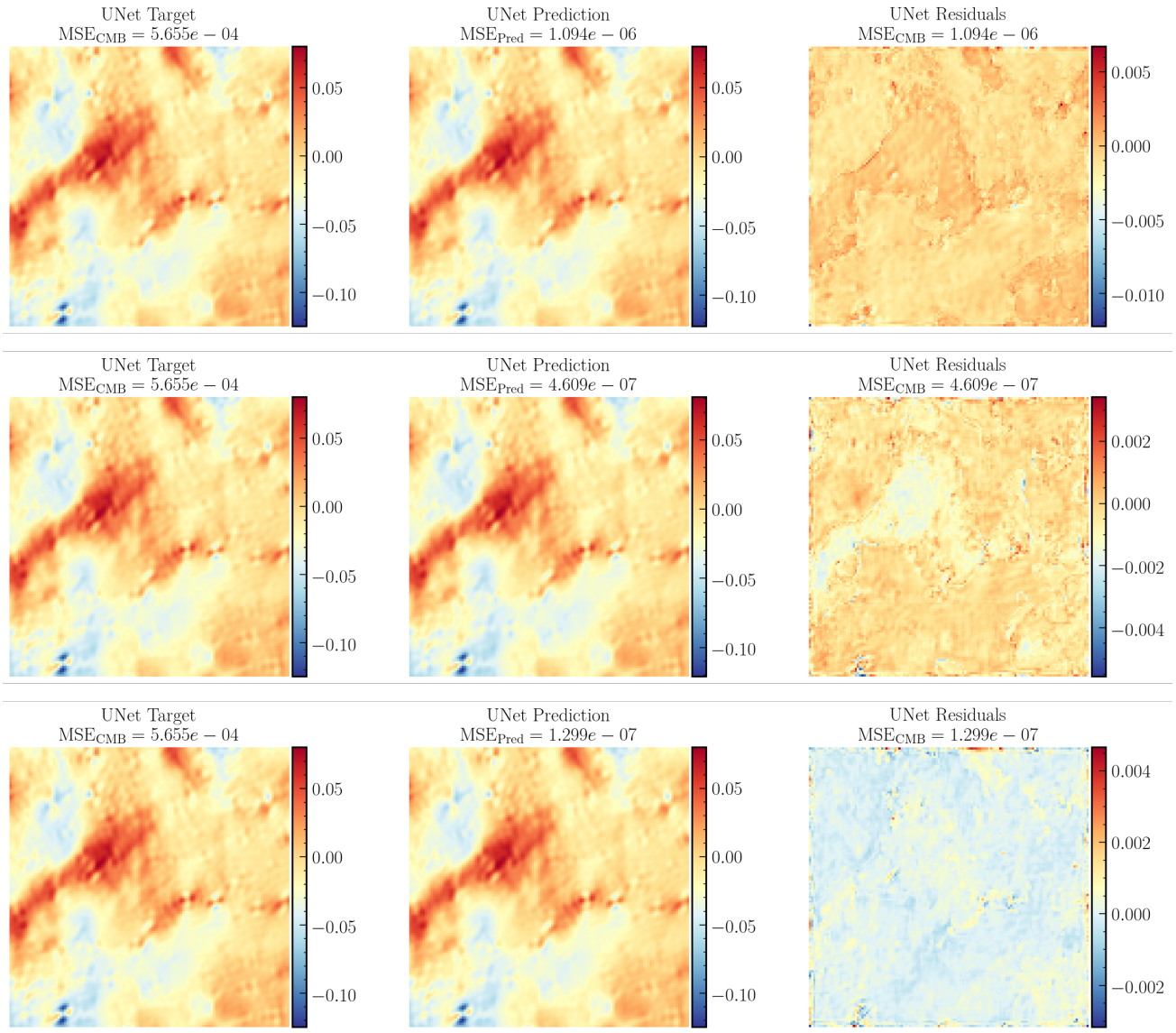


Figure 7. Similar to Fig. 12 in Ref. (McCarthy et al., 2024), showing the target ILC residuals ΔF^{ILC} (first column), the U-Net prediction (second column), and the final residual of the U-Net-corrected reconstruction, which is obtained after removing the U-Net prediction from the ILC solution (third column). As seen in the first column, the ILC residuals are heavily anisotropic. These remnant structures are ones that the ILC cannot remove and reflect the need for non-linear reconstruction. **Top row:** U-Net trained on just $\hat{F}_i$. While producing the highest residuals of the three models, it successfully mitigates anisotropic ILC structures. **Middle row:** U-Net trained on $\hat{F}_i + B_S$, achieving over three orders of magnitude lower in MSE compared to that of the ILC residual. **Bottom row:** U-Net trained on $\hat{F}_i + B_S + T + E$. This network is able to remove almost all of the foreground variance remaining in the ILC solution, reducing the residual by over 99%, from $MSE = 5.65 \times 10^{-4}$ to $MSE_{S_L} = 1.29 \times 10^{-7}$. Overall, increasing network inputs progressively eliminates residual foreground structures from the ILC solution, consistently suggesting that the additional inputs are useful for predicting anisotropic foregrounds.

A.3. U-Net architecture and training

We train the same U-Net architecture (Ronneberger et al., 2015) for both single-frequency and hybrid variants, varying the input channels for these configurations. The network is an encoder–decoder with skip connections and channel widths $[32, 64, 128, 256, 512, 1024]$. Each encoder block stacks two 3×3 stride-one convolutions with padding. We apply batch normalization and LeakyReLU activations with negative slope 0.01. A 2×2 average pool then halves both spatial dimensions. The bottleneck increases the depth from 1024 to 2048 channels. Next, the decoder stages upsample with 2×2 transposed

convolutions at stride two, concatenate matching skips along the channel axis, apply paired convolution–batch-normalization–LeakyReLU blocks that symmetrically narrow channels toward the input widths, and repeat until full patch resolution is restored. When boundary rounding leaves upsampled tensors misaligned with skips, upsampled maps are resized to match skips prior to concatenation. The output head is a pointwise (1×1) convolution producing either $\hat{F}_L$ or $\widehat{\Delta F}^{\mathrm{ILC}}$. Patches of size 128×128 traverse six pooling stages before reaching a 2×2 latent grid.

We train to minimize the mean squared error, using an AdamW optimizer and decoupled weight decay. We warm the learning rates linearly before plateau-driven reductions (multiply by 0.5 whenever validation loss fails to improve for five epochs). Optimization stops after twenty stagnant epochs or after two hundred epochs, whichever occurs first.

Single-frequency models use batch size 32, peak learning rate 10^{-3}, weight decay 5×10^{-4}, five warmup epochs starting from 10^{-5}, and gradient norms clipped at unity. Hybrid multi-frequency models use batch size 64, peak learning rate 10^{-4}, weight decay 10^{-5}, ten warmup epochs starting from 10^{-6}, gradient norms clipped at five, and an optional rescaling of the peak learning rate by $\sqrt{N_{\mathrm{train}}/N_{\mathrm{ref}}}$ when warranted by relative dataset sizes, with reference count $N_{\mathrm{ref}} = 100$.

For training stability, every channel of inputs and targets is z-score standardized using training-split mean and variance estimates computed separately per channel.

A.4. Hybrid Models: Anisotropic Foreground Residuals

Figure 7 exemplifies test predictions and residuals for different input configurations of the hybrid models discussed in Sec. 3.3. As seen in the first column, the ILC residuals, ΔF_L^{ILC}, are heavily anisotropic. These remnant structures are ones that the ILC cannot remove and reflect the need for beyond-Gaussian reconstruction. The third column visualizes the amount of residual structure that is learned and progressively removed by U-Nets with additional input information.