A Novel Approach to Optimize Clustering of Parametric Map-Based Component Separation for Upcoming CMB Polarization Satellites

Wassim Kabalan[1], Wuhyun Sohn[1], Benjamin Beringue[1], Artem Basyrov[1], Pierre Chanial[1], Alexandre Boucaud[1], Josquin Errard[1]

[1]Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France

SciPol — Science from the large scale cosmic microwave background polarization structure

Tensor-to-Scalar Ratio & Foreground Removal

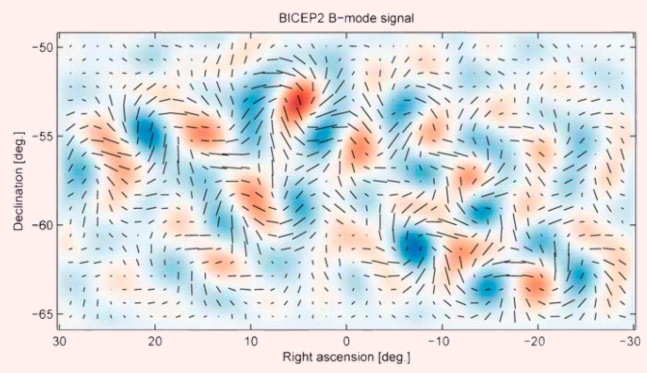

Figure 1: CMB B-mode polarization signal from primordial gravitational waves

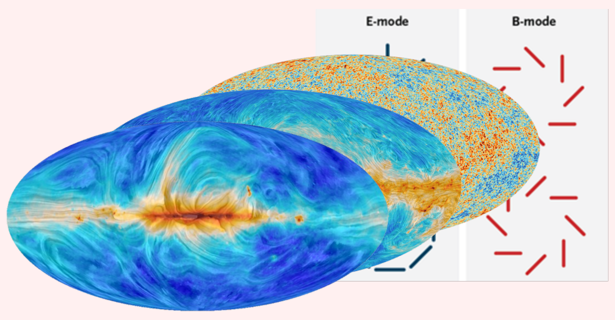

Figure 2: Galactic foreground contamination

The Challenge:

- Measuring tensor-to-scalar ratio r requires detecting faint B-mode polarization
- Galactic foregrounds dominate CMB signal by orders of magnitude
- Accurate foreground removal is critical for $r < 0.001$ detection

Parametric Component Separation

The Standard Model: The observed sky signal ($\mathbf{d}$) in each pixel is modeled as a linear combination of astrophysical components ($\mathbf{s}$) and instrumental noise ($\mathbf{n}$).

$$\mathbf{d} = \mathbf{A}(\boldsymbol{\beta})\,\mathbf{s} + \mathbf{n} \tag{1}$$

The goal is to solve for the components $\mathbf{s}$, particularly the CMB, by estimating the spectral parameters $\boldsymbol{\beta}$ that define the mixing matrix $\mathbf{A}$.

Spectral Likelihood:

$$\ln \mathcal{L}_{\rm spec}(\boldsymbol{\beta}) \propto (\mathbf{A}^\top\mathbf{N}^{-1}\mathbf{d})^\top(\mathbf{A}^\top\mathbf{N}^{-1}\mathbf{A})^{-1}(\mathbf{A}^\top\mathbf{N}^{-1}\mathbf{d}) \tag{2}$$

[1]

Modeling Spatially-Varying Foregrounds

Foreground Spectral Parameters Vary Across Sky:

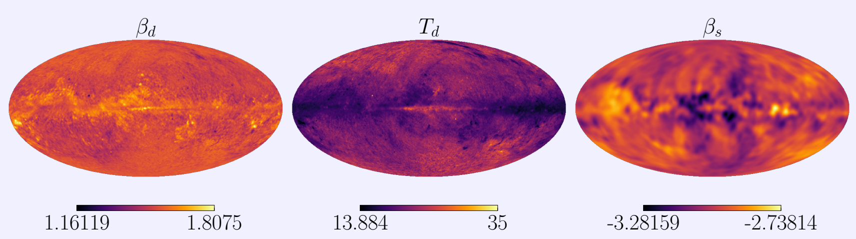

Figure 3: Modifid blackbody spectral index map showing spatial variability in dust emission properties across the sky

Clustering Approach:

- Spherical K-means clustering groups pixels with similar spectral properties
- Balances statistical uncertainty with modeling flexibility

Optimized K-means Parameters:

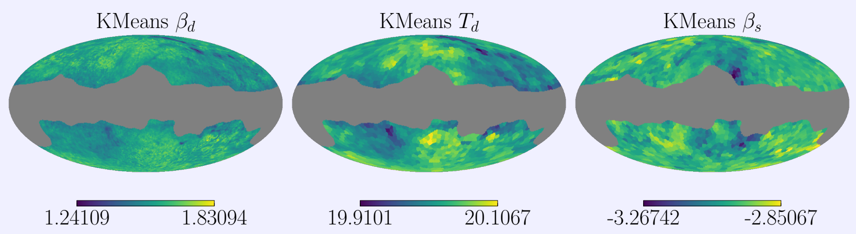

Figure 4: Optimized spectral parameter distributions from K-means clustering showing data-driven adaptation to foreground complexity

The FURAX Framework: A Scalable JAX-Native Engine

Core Features:

- **JAX-Native**: End-to-end differentiable & GPU accelerated. [2]
- **Modular Design**: Composable algebraic operators for complex models.
- **Memory Efficient**: Matrix-free operators enable high-resolution analysis.
- **Scalable**: Distributed parallelism for HPC clusters.

Elegant Implementation:

```
cmb = CMBOperator(nu, ...)
dust = DustOperator(nu, temp=params["T_d"],
    beta=params["beta_d"], ...)
synchrotron = SynchrotronOperator(nu,
    beta=params["beta_pl"], ...)
A = MixingMatrixOperator(cmb, dust, synchrotron)
AND = (A.T @ N.I)(d)
s = (A.T @ N.I @ A).I(AND)
likelihood = AND @ s
```

Unprecedented Speed:

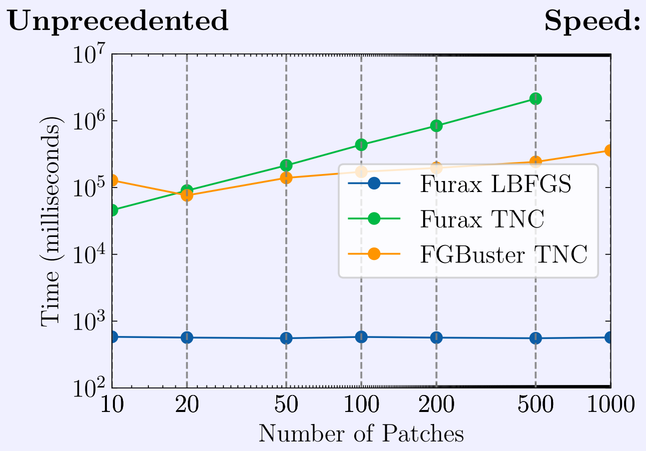

Figure 5: *FURAX (LBFGS) scales robustly with the number of patches, consistently outperforming older frameworks like FGBuster.*

Enabling a Grid Search at Scale: The performance of FURAX allowed us to run ~1.92 million independent component separation fits to find the optimal clustering configuration, a task infeasible with previous tools.

Selection Criterion: Minimizing CMB Variance

Selection Criterion - CMB Variance Minimization:

$$\sigma^2_{\rm CMB} = \left\langle \mathrm{Var}_i\left[\hat{s}^{(i)}_{\rm CMB}\right]\right\rangle_{\rm pixels} \tag{3}$$

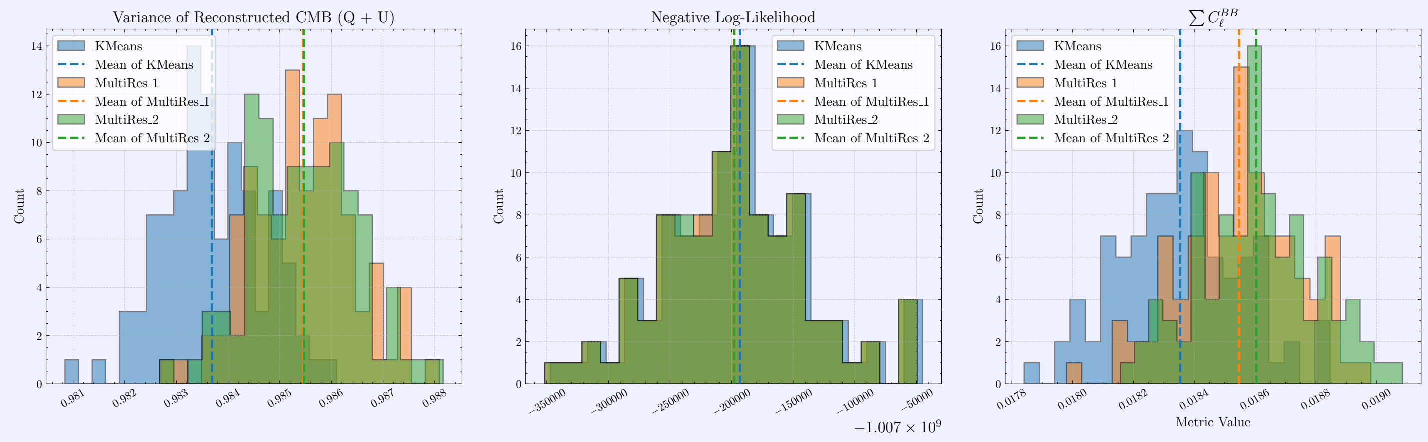

Figure 6: Distribution of CMB variance, spectral likelihood, and B-mode power for different spatial modeling approaches. K-means clustering achieves optimal balance.

Key Insight: Variance minimization acts as proxy for residual foreground contamination, leading to more robust cosmological constraints.

Comparing Clustering Strategies: Data-Driven vs. Fixed-Resolution

K-means Clustering (Our Method)

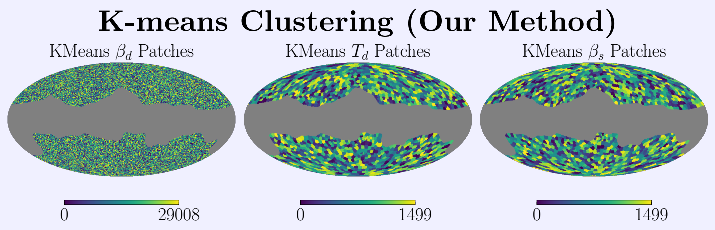

Figure 7: *Flexible, data-driven patches adapt to foreground complexity by minimizing CMB variance. This corresponds to Figure 7 in the paper.* [3]

- Generates irregular, equal-area patches.
- Patch structure is learned from the data itself.

Multi-Resolution (Baseline)

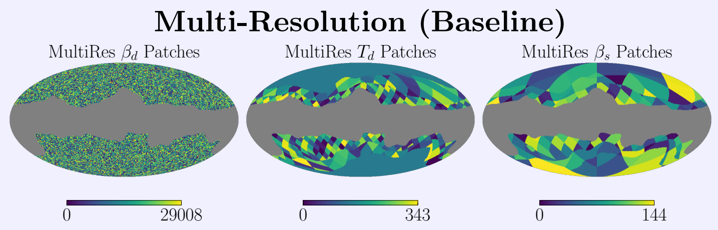

Figure 8: *Fixed, regular patches based on HEALPix downgrading, independent of sky variations. This corresponds to Figure 8 in the paper.* [4]

- Imposes a rigid, uniform grid.
- Lacks flexibility for complex regions.

Results

Tensor-to-Scalar Ratio Constraints:

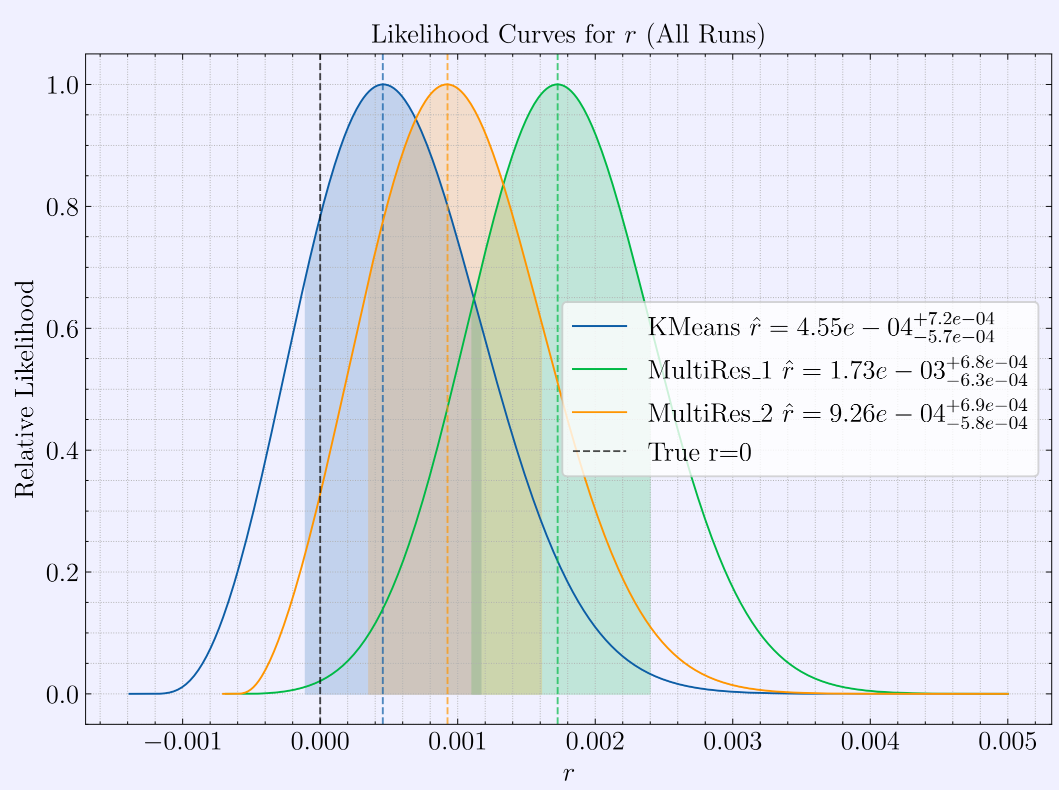

Figure 9: r likelihood distributions: K-means clustering (blue) yields $\hat{r} = 4.55 \times 10^{-4}$ with lowest bias and tightest constraints compared to multi-resolution approaches

Residual B-mode Spectra:

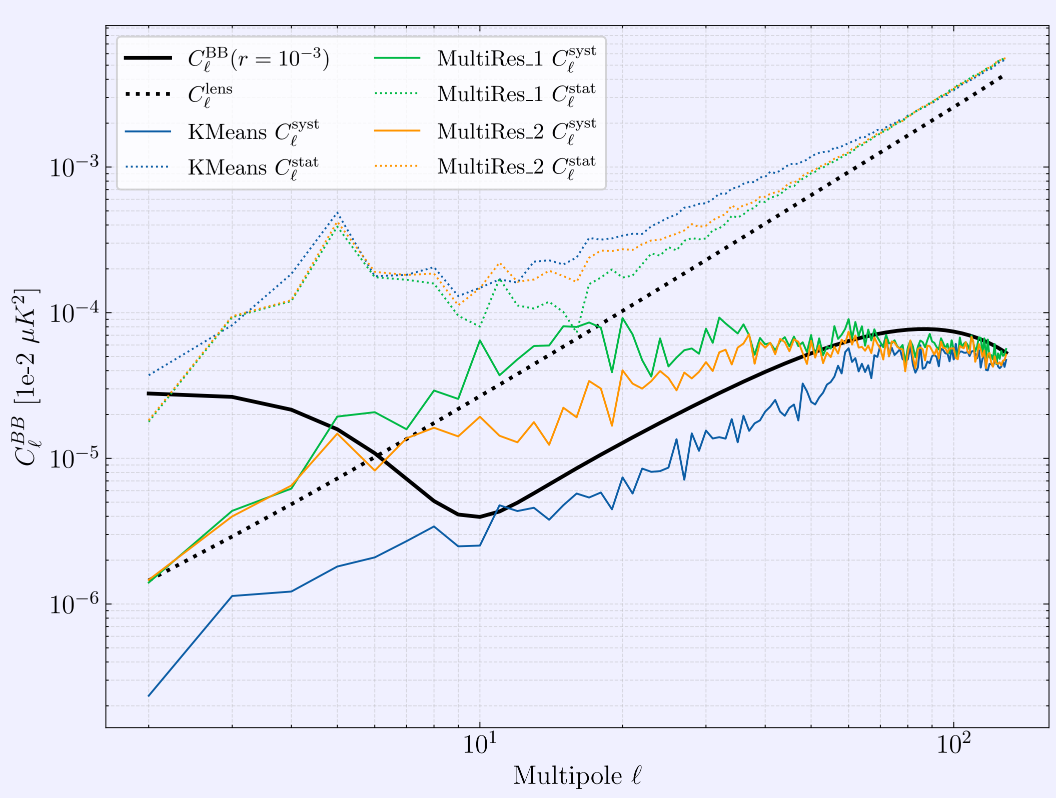

Figure 10: Residual B-mode power spectra: K-means clustering (blue) achieves significantly lower systematic residuals compared to multi-resolution approaches, falling below target sensitivity levels

References & Acknowledgements

References

[1] Radek Stompor. Maximum likelihood algorithm for parametric component separation in cosmic microwave background experiments. *Monthly Notices of the Royal Astronomical Society*, 392(1):216–232, January 2009.
[2] P. Chanial and the CMBSciPol team. Furax: Flexible unified regression analysis for component separation. `https://github.com/CMBSciPol/furax`, 2024.
[3] Pierre Chanial, Simon Biquard, and Wassim Kabalan. jax-healpy: Implementation of HEALPix related functions and extensions in JAX. `https://github.com/CMBSciPol/jax-healpy`, 2024.
[4] et al Allys. Probing cosmic inflation with thelitebirdcosmic microwave background polarization survey. *Progress of Theoretical and Experimental Physics*, 2023(4), November 2022.

Acknowledgements: This work was supported by the SciPol project (ERC Grant No. 101044073, PI: Josquin Errard) and performed using HPC resources at IDRIS Jean Zay supercomputer. Full code available at: `https://github.com/ASKabalan/furax-compsep-paper`