Cosmic symphony: orchestrating wavelets, ILC and ML for clearer CMB signals

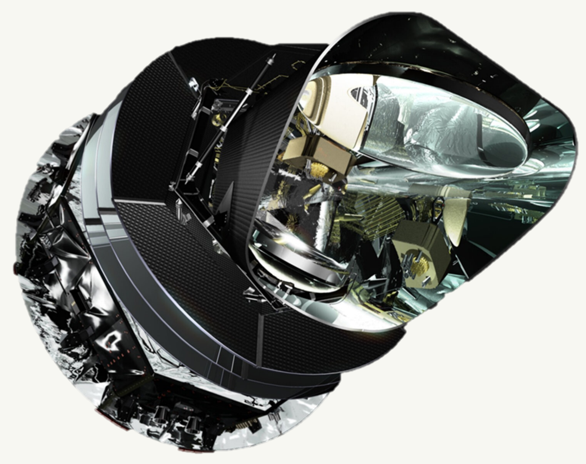
Zhaoqi Wang [1]   Supervised by: Dr. Keir Rogers [2]

[1] Department of Mechanical and Industrial Engineering, University of Toronto
[2] Dunlap Institute for Astronomy and Astrophysics, University of Toronto

Introduction

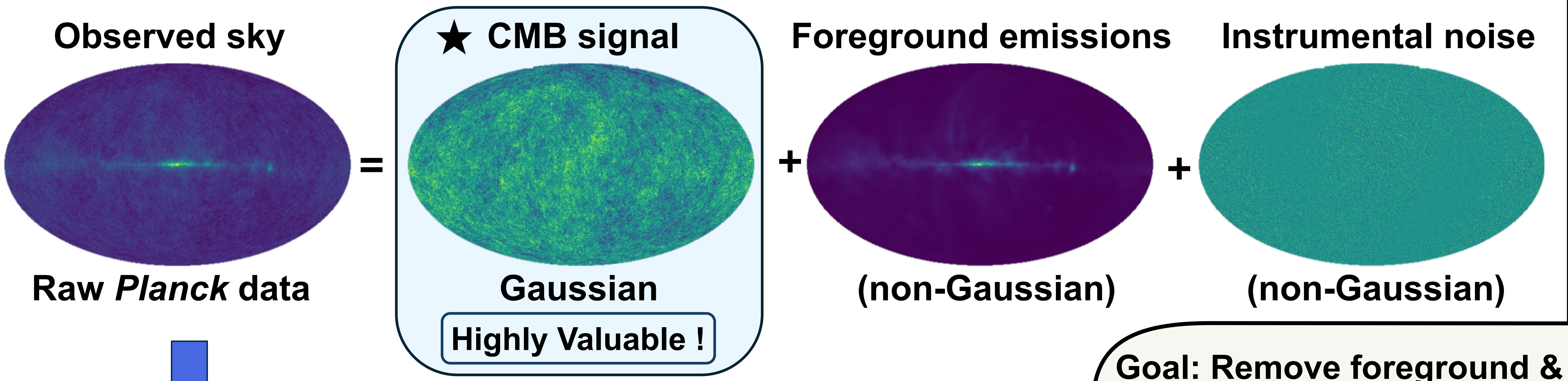
380,000 years after the Big Bang: the first light of the Universe was emitted! As the Universe expanded, this light redshifted into the **Cosmic Microwave Background (CMB)**, now uniformly distributed across the cosmos, with tiny temperature fluctuations ($\sim 10^{-5} K$) revealing the Universe's initial density variations that seeded the formation of galaxies and other large-scale cosmic structures. To capture this snapshot of the early Universe, cosmologists developed instruments to scan the sky. These tools detect faint CMB radiation but also pick up emissions from other astrophysical sources and instrumental noise.

Data source:
***Planck* satellite**

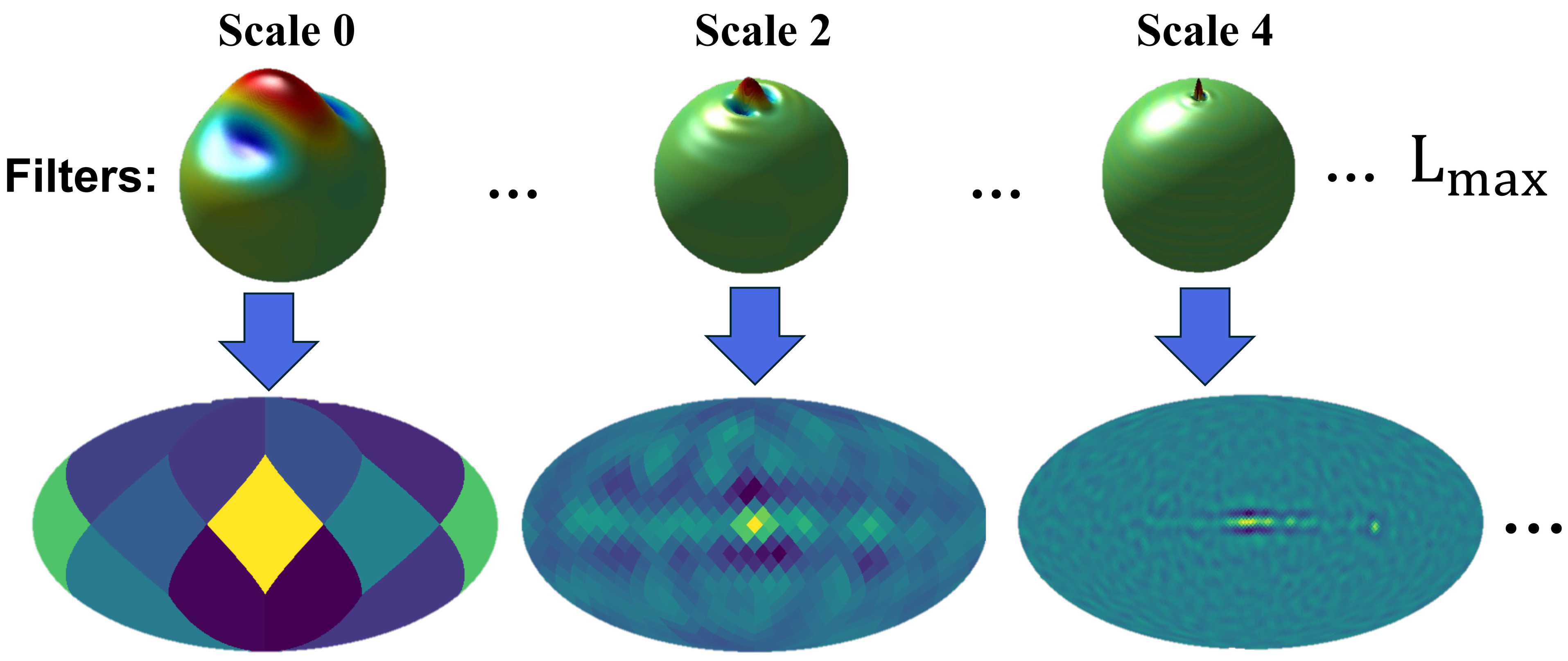


***Planck* Mission [1]** provides the most recent all-sky survey at 9 frequencies:

(30, 44, 70, 100, 143, 217, 353, 545, 857 GHz). Simulated data was used.

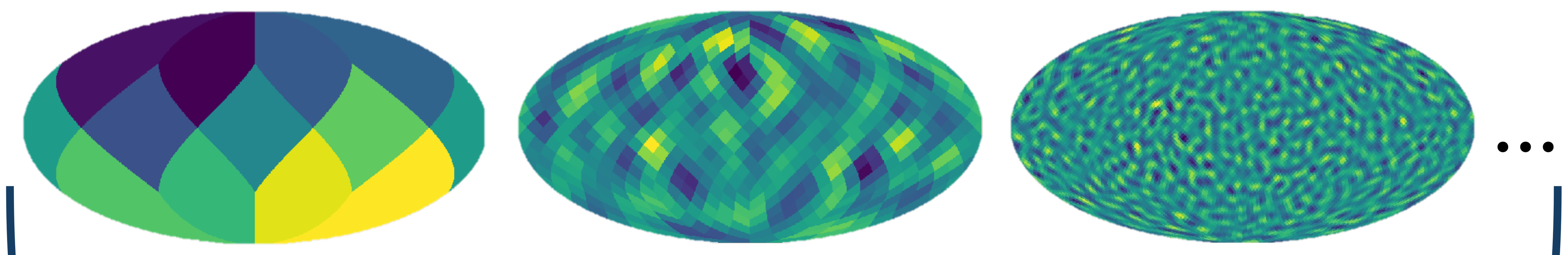


Wavelet transform: Convolve sky with filters



Wavelet coefficient maps

ILC: Combine wavelet maps of nine frequencies at each scale to create single cleaned wavelet map



ILC cleaned wavelet coefficient maps

Goal: Remove foreground & noise while preserving CMB

1. Wavelet Transform [2]

Wavelets decompose sky data into multiple scales for detailed analysis. Wavelet coefficient maps are generated by convolving discrete wavelet filters (from s2wav package) with *Planck* data. The user-defined maximum multipole moment $\mathrm{L_{max}}$, determines the finest angular resolution.

Observation: Larger scale maps capture the overall trend of the observed sky, while the smaller scale maps are sensitive to localized temperature fluctuations.

2. Internal Linear Combination (ILC) [2]

ILC wavelet maps are created by combining all 9 frequency maps with optimized weights w_i to minimize the variance of the sum.

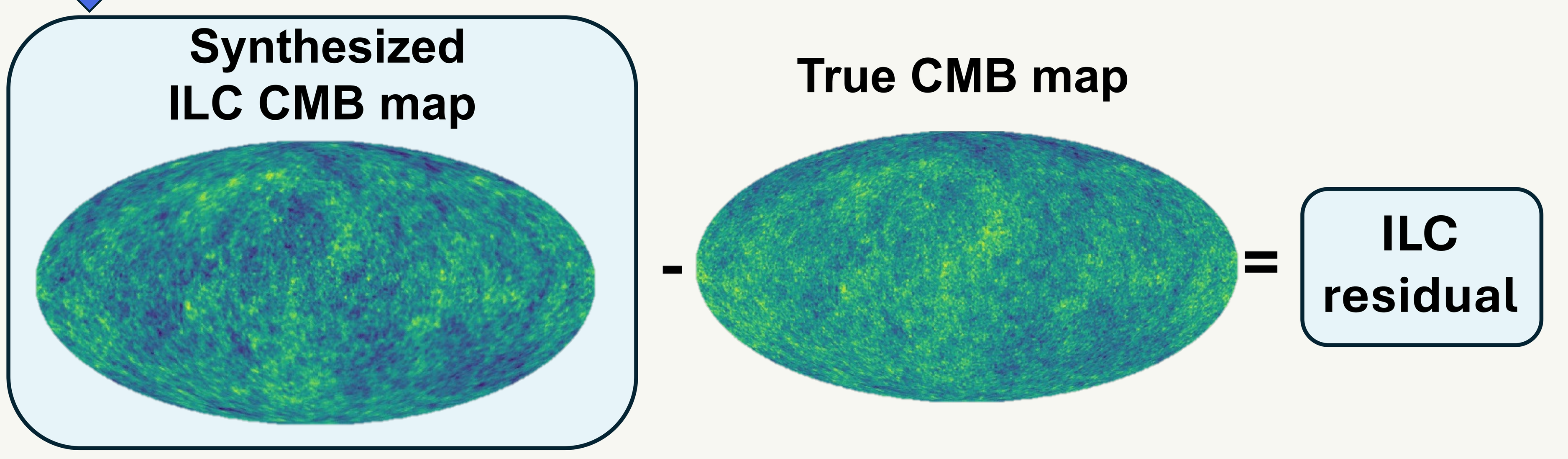
$$i = \text{frequency}; \quad j = \text{scale}$$

$$[\text{ILC cleaned wavelet map}]_j = \sum w_i * [\text{wavelet map}]_i$$

$\sum w_i = 1$: Total contribution from 9 frequencies sums to one to ensure unbiased signal preservation.

Observation: Synthesized ILC CMB map resembles the simulated CMB signal, with frequency-dependent foregrounds and noises minimized.

3. Machine Learning Residual Estimation [3]

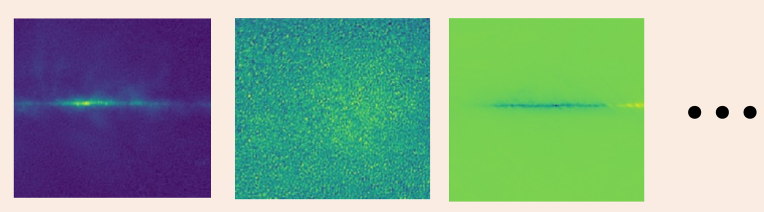


There is still a discrepancy between the ILC CMB and the true CMB map. To close this gap, we designed an encoder-decoder structure inspired by the U-Net architecture to predict ILC residuals from CMB signal-free maps across 9 frequencies.

Training Data Preparation

Inputs: Signal-free maps:
Observed sky – ILC CMB map
This ensures the inputs of the model only contain foreground and noise.

Output: ILC residual map:
ILC CMB map – True CMB map



Note: Maps are trimmed to squares to avoid resizing loss and increase training data.

Inputs: Signal-free maps at 9 frequencies

[Channels: 9, Height: 256, Width: 256]
Double Conv 3×3, Relu
[64, 256, 256]
Max Pool 2×2
[64, 128, 128]   Repeat 4 times

Encoder

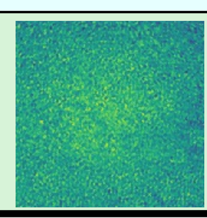
It extracts key features from the input while reducing dimensionality.

[1024, 16, 16]
UpSample 2×2   Repeat 4 times
[512, 32, 32]
Double Conv 3×3, Relu
[512, 32, 32]
Conv 2D 64×1
[1, 256, 256]

Decoder

It reconstructs the encoded, compressed representation to predict the ILC residuals.

Output Residual prediction:



Final Step: Subtract the residual prediction from the ILC CMB map to make a cleaner map.

4. Conclusion & Future: A pipeline for wavelet transform and ILC implementation has been built, and a Python package, **Skyclean**, has been published. The ML prediction model is in the design phase, focusing on data processing, customization, and hyperparameter tuning. The ILC-CMB signal discrepancy, possibly due to ILC bias, will be further explored post-model design.

References: [1] Planck Legacy Archive. (n.d.). https://pla.esac.esa.int/pla/#maps [2] Rogers, K. K., Peiris, H. V., Leistedt, B., McEwen, J. D., & Pontzen, A. (2016). Silc: A new *Planck* internal linear combination CMB temperature map using directional wavelets. *Monthly Notices of the Royal Astronomical Society*, *460*(3), 3014–3028. https://doi.org/10.1093/mnras/stw1121
[3] McCarthy, F., Hill, J. C., Coulton, W. R., & Hogg, D. W. (2024, July 31). *Signal-preserving CMB component separation with machine learning*. arXiv.org. https://arxiv.org/abs/2404.03557

Python Packages: Numpy, Astropy, Healpy, s2fft, s2wav.

```
pip install Skyclean
```