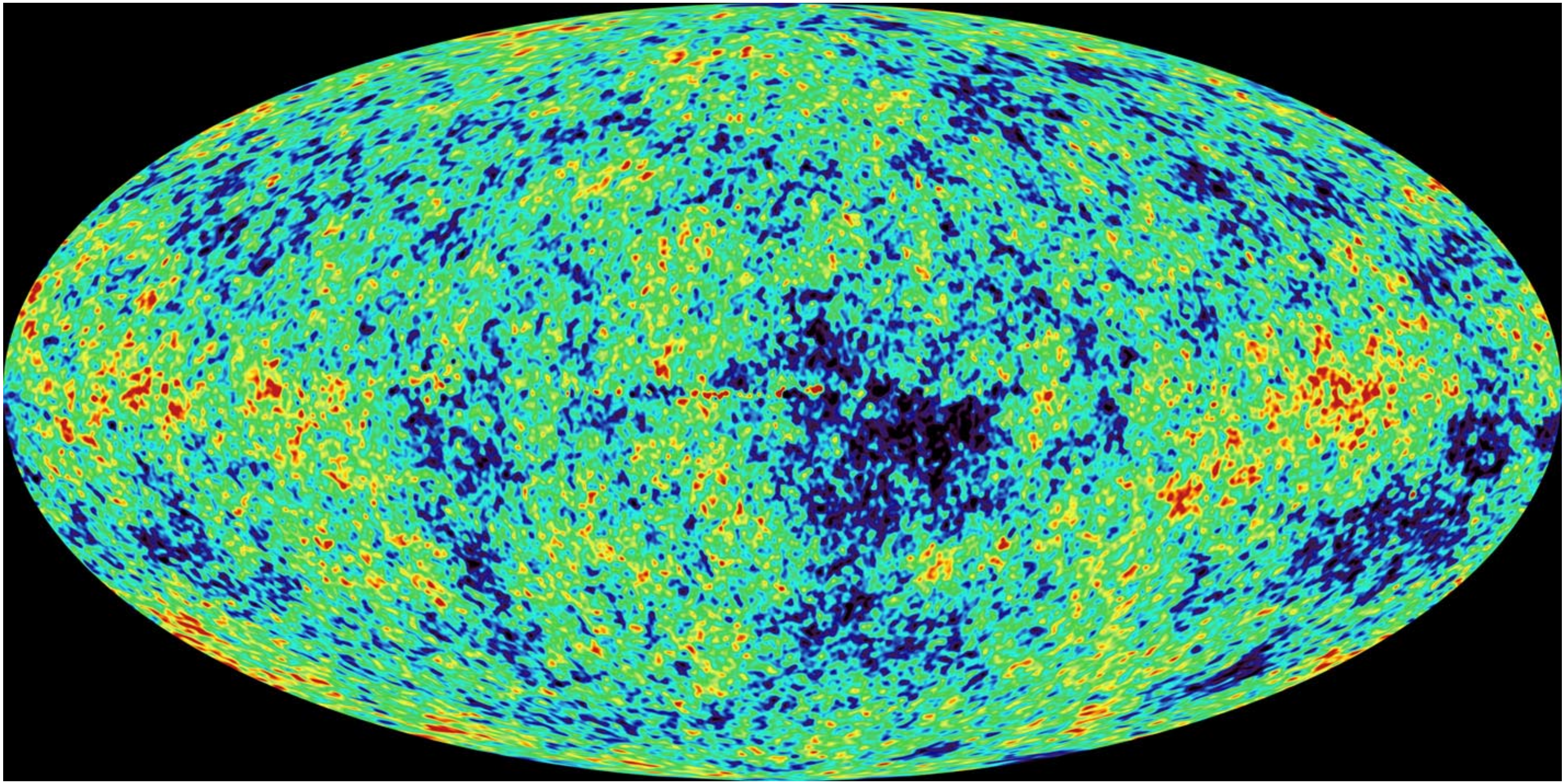
An Approach to Detecting Non-Gaussianity in the Cosmic Microwave Background

Kristofer Jennings

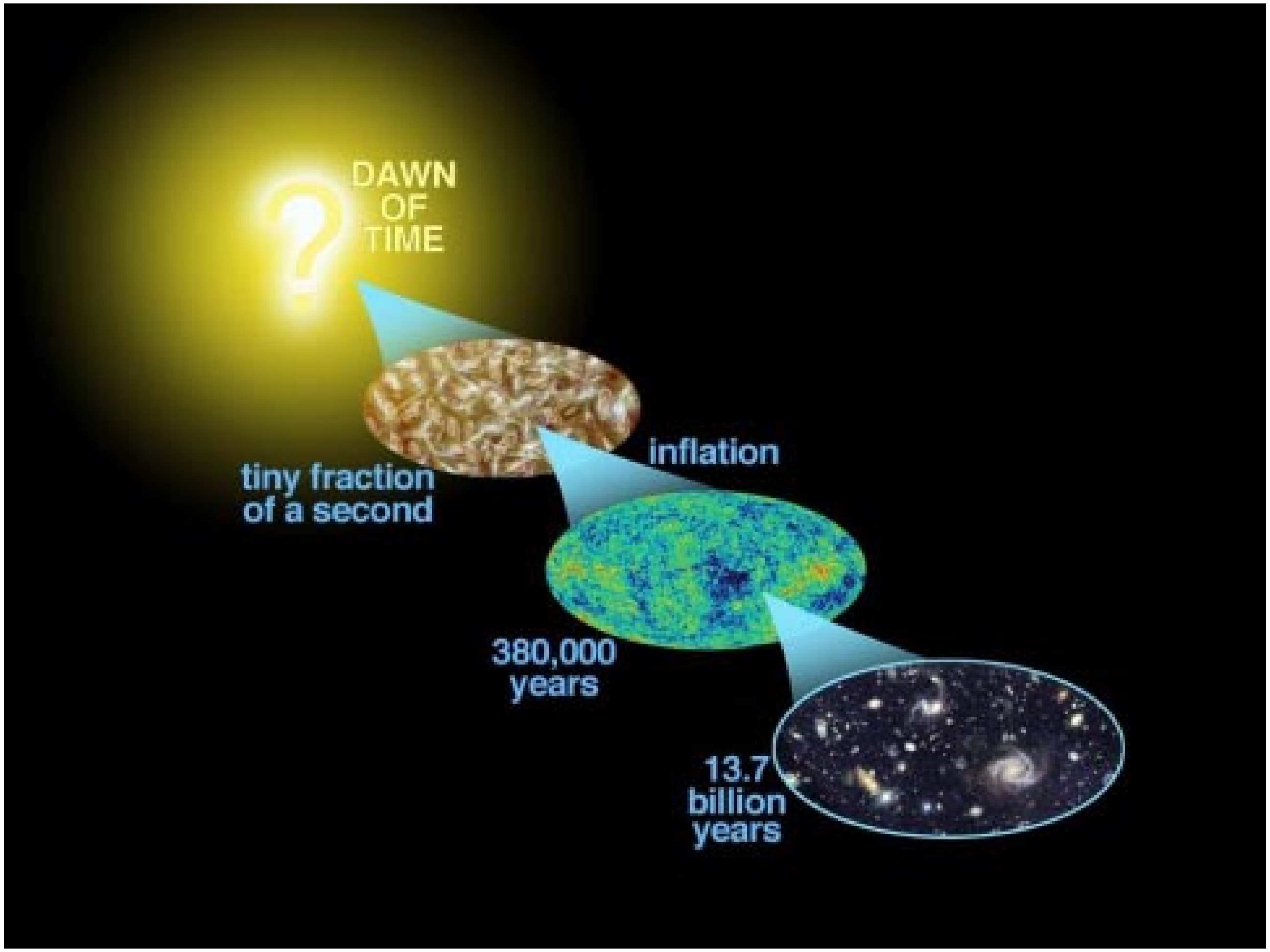
Purdue University

Joint work with Laura Cayón

Cosmic Microwave Background Radiation (CMB)



The Theory of Inflation

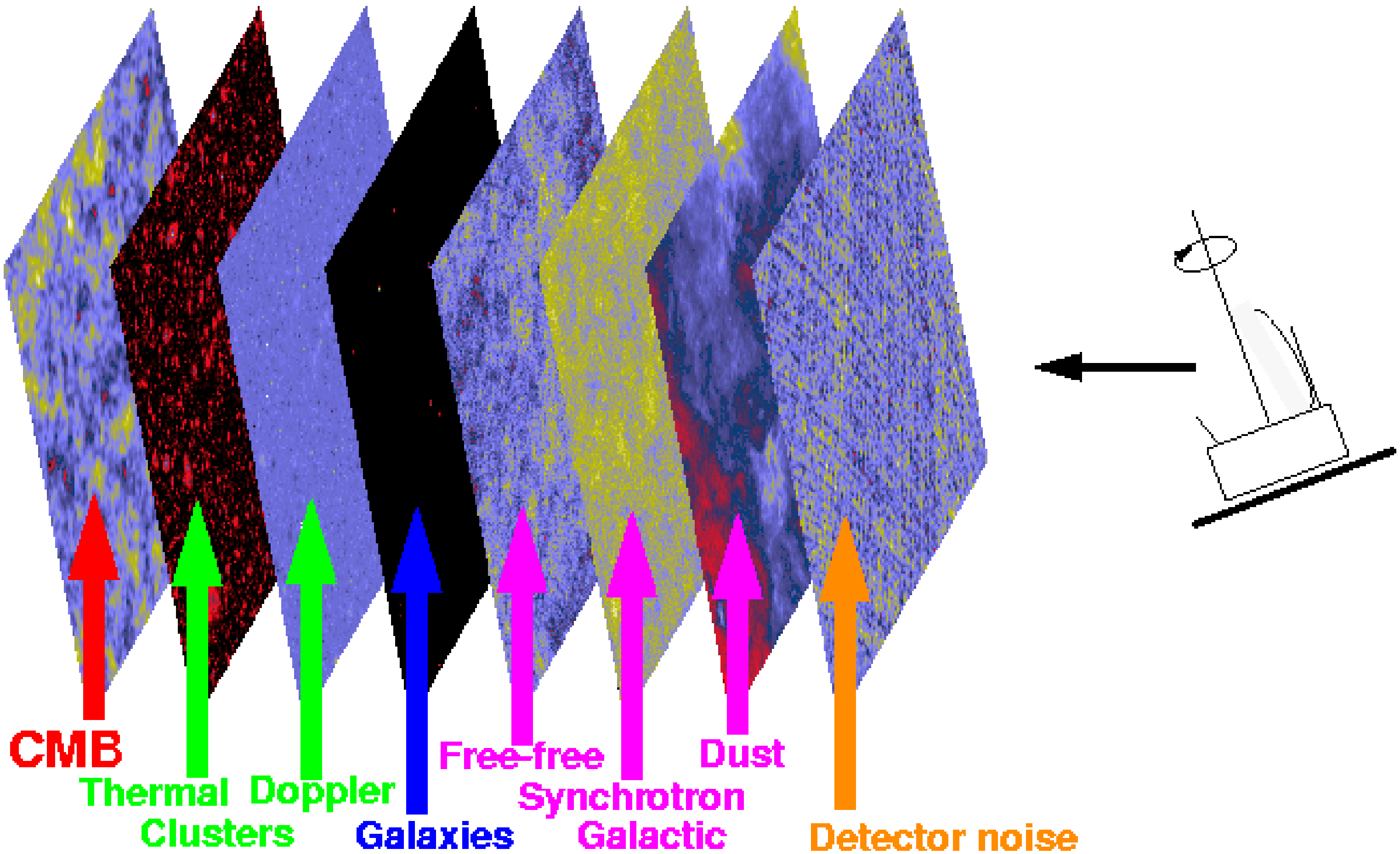


Expansion of the Universe

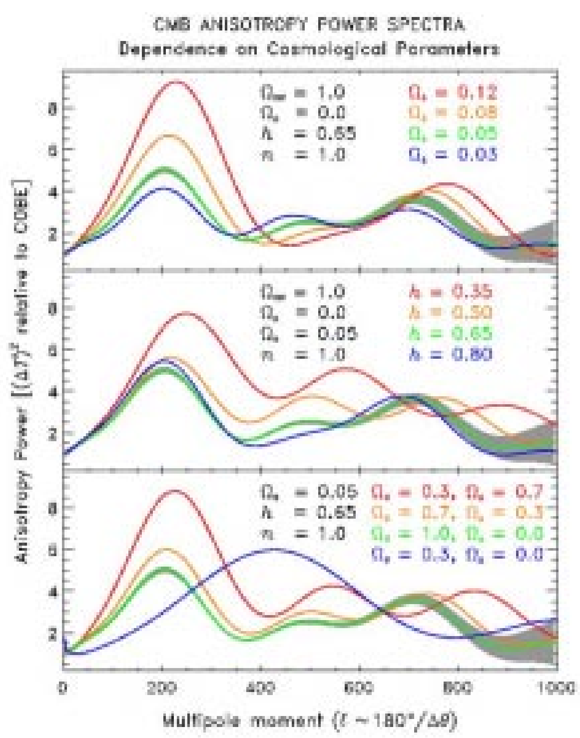
$\longrightarrow$ photons cool down to $T \sim 3K$ at present

$\longrightarrow$ wavelengths in the microwave range

Challenging Data



Why We Care: Clues to Structure/Origin of the Universe



Geometry/energy density of universe

Baryon/Dark Matter Ratio

Baryon/Photon Ratio

Primordial Spectrum

Type of Cosmological Fluctuation

Star Formation History

Gravitational Waves

In so many words, EVERYTHING

Inflationary Model

For celestial angles, $0 < \theta \leq \pi$, $0 < \varphi \leq 2\pi$, Φ is the temperature of the background radiation for that pixel.

Model:

$$\Phi(\theta, \varphi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m}(\theta, \varphi),$$

where $Y_{\ell,m}(\theta, \varphi)$ are spherical harmonics.

- $a_{\ell,m}$ are zero mean, uncorrelated (except $a_{\ell,-m} = (-1)^m a^*_{\ell,m}$) with variances equal to the angular power spectrum of the random field
- **$\Phi(\theta, \varphi)$ is Gaussian if the $\{a_{\ell,m}\}$ come from a Gaussian distribution.**
- ***But what if they don't...?***

Different Methods for Framing Question

- “Real Space” Methods
  - Density of saddle points
  - Looking for “hot” regions
  - Merging Inflation with Cosmic String representation
- Harmonic Analysis
- Wavelet Analysis

“Constrained Non-Gaussianity”

$$\Phi(\theta,\varphi) = \Phi_L(\theta,\varphi) + f_{N,L}\left[\Phi_L^2(x) - \langle\Phi_L^2(\theta,\varphi)\rangle\right]$$

Question: What is $f_{N,L}$?

- The most efficient algorithms require approximately N^3 operations
- $N \sim 3 \times 10^6$ for WMAP, 5×10^7 for Planck (2008)

General Statistical Question

If the likelihood is difficult to form analytically, but simulations are “easy” to produce (for different values of $f_{N,L}$), can this information be used to decide which values of $f_{N,L}$ are most likely?

An Approach

- Each simulation is some “distance” away from the actual data.
- These distances can be simulated to produce a density estimate.
- The density at distance $= 0$ is the “likelihood” of a particular $f_{N,L}$.

(A few of the) Statistical Issues

- Need accurate density estimation technique (particularly since 0 is a boundary and extremely far away from most of the data).
- Input space is very high-dimensional.
- Inference is also necessary.
- Knowing how much simulation will eventually be necessary would be helpful. (Trading analytic problem for computing problem.)

Methodologies

- *Sampling*
  - Use symmetry for antithetic sampling.
  - Compute metric for each value $f_{N,L}$ using same inputs.
  - Adaptive importance sampling – sampling near points of interest

- Using analytic information – use side information (like derivatives) to increase efficiency of estimation and sampling
- Incorporate space-filling sampling schemes such as quasi-Monte Carlo.

- *Density Estimation*
  - Since **relationship** between likelihoods is important, don't need to carry around importance weights in density calculation.
  - **Real question:** What density (kernel) should be used to sample distribution of distances?
  - **Partial answer:** Want to zero out all but first derivative (at distance 0) of sampling density.

Progress

- **Developmental**
  - Incorporating analytic information greatly improves sampling efficiency.
  - Some preliminary results on improved density estimation.
  - Inference follows (more or less) naturally from estimation procedure.

- **Simulation**
  - Preliminary results are crude but promising.
  - *Upshot:* Can only improve with more computation