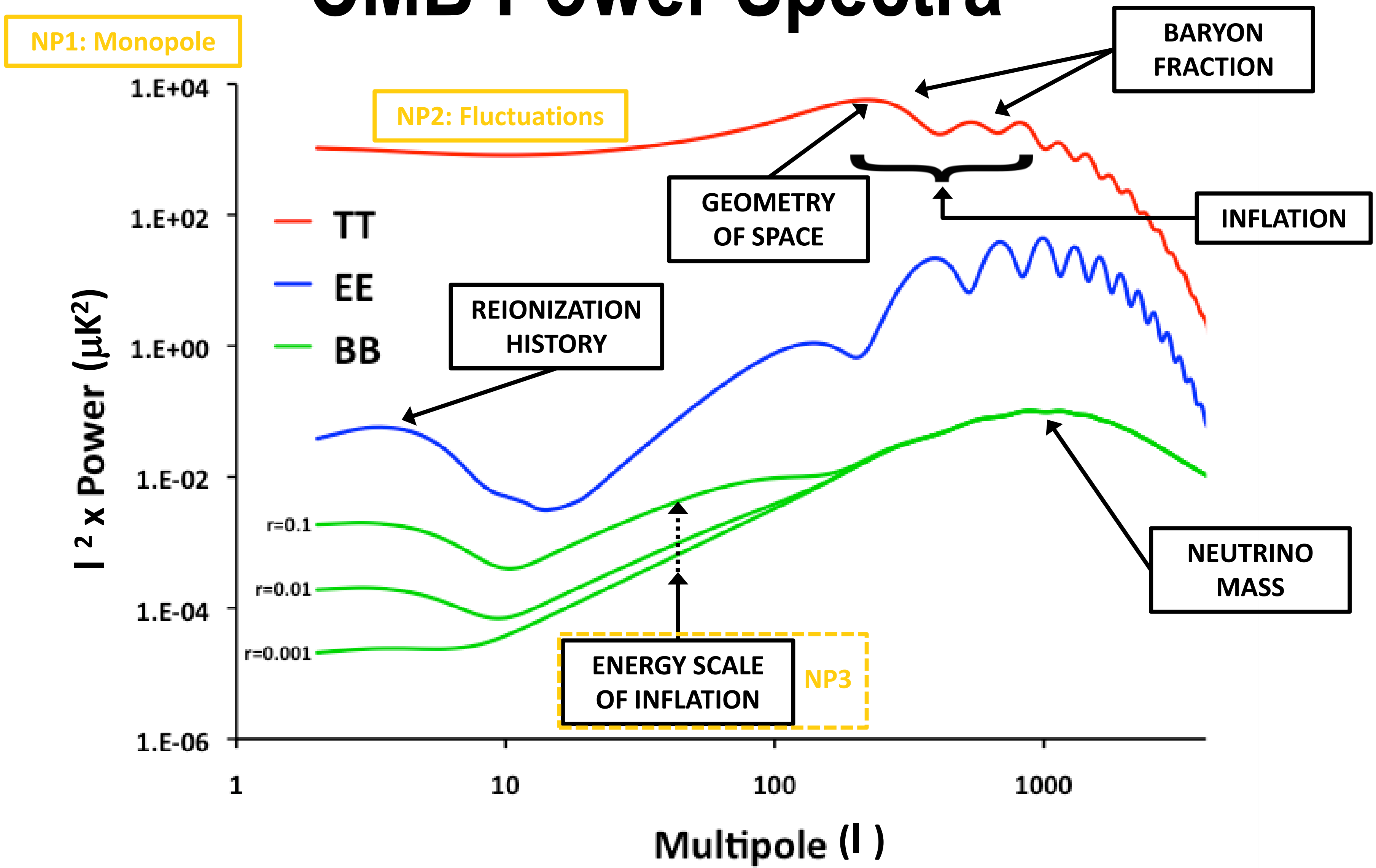
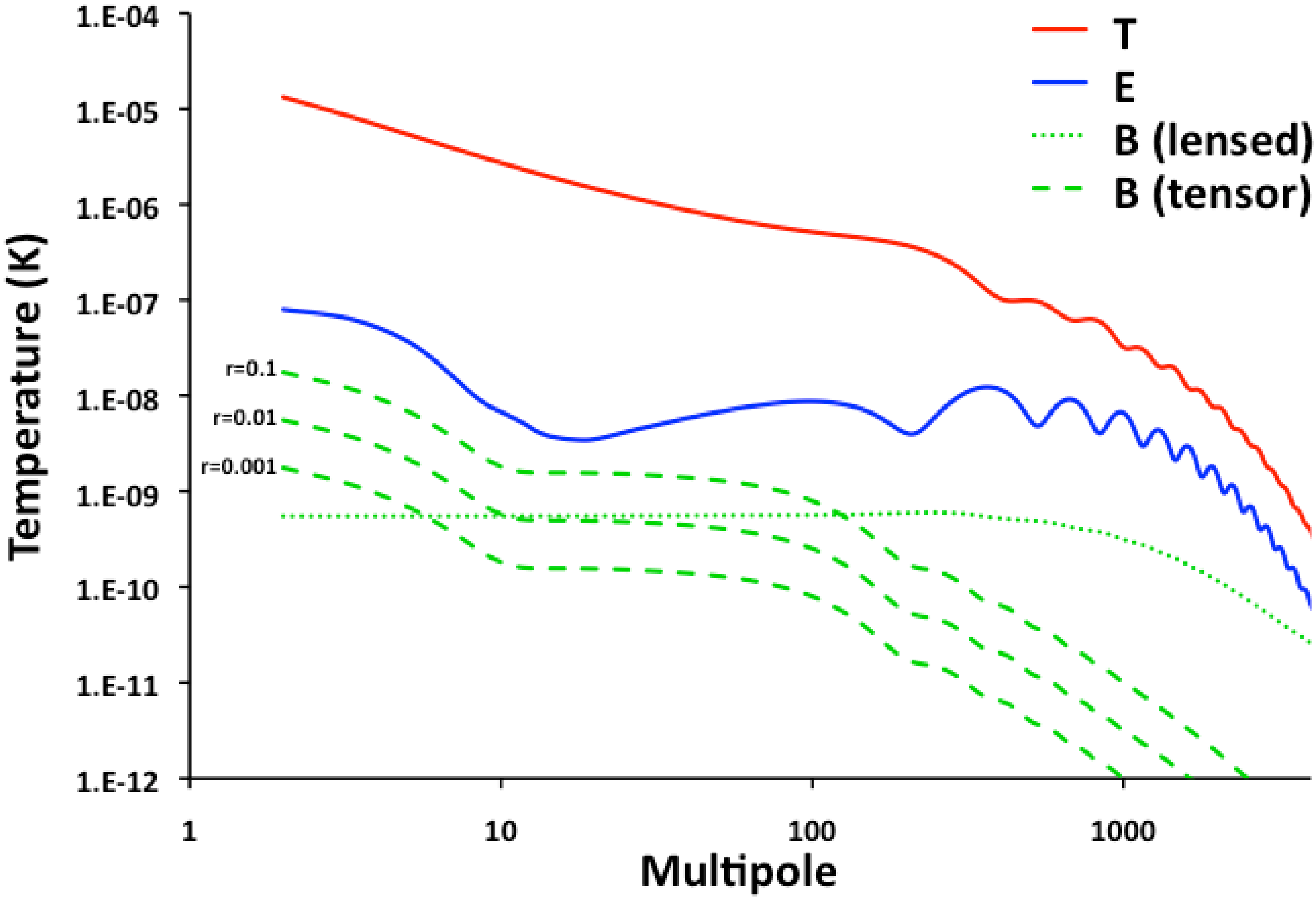
# Data Issues

Julian Borrill
Computational Cosmology Center, Berkeley Lab
& Space Sciences Laboratory, UC Berkeley

# CMB Power Spectra



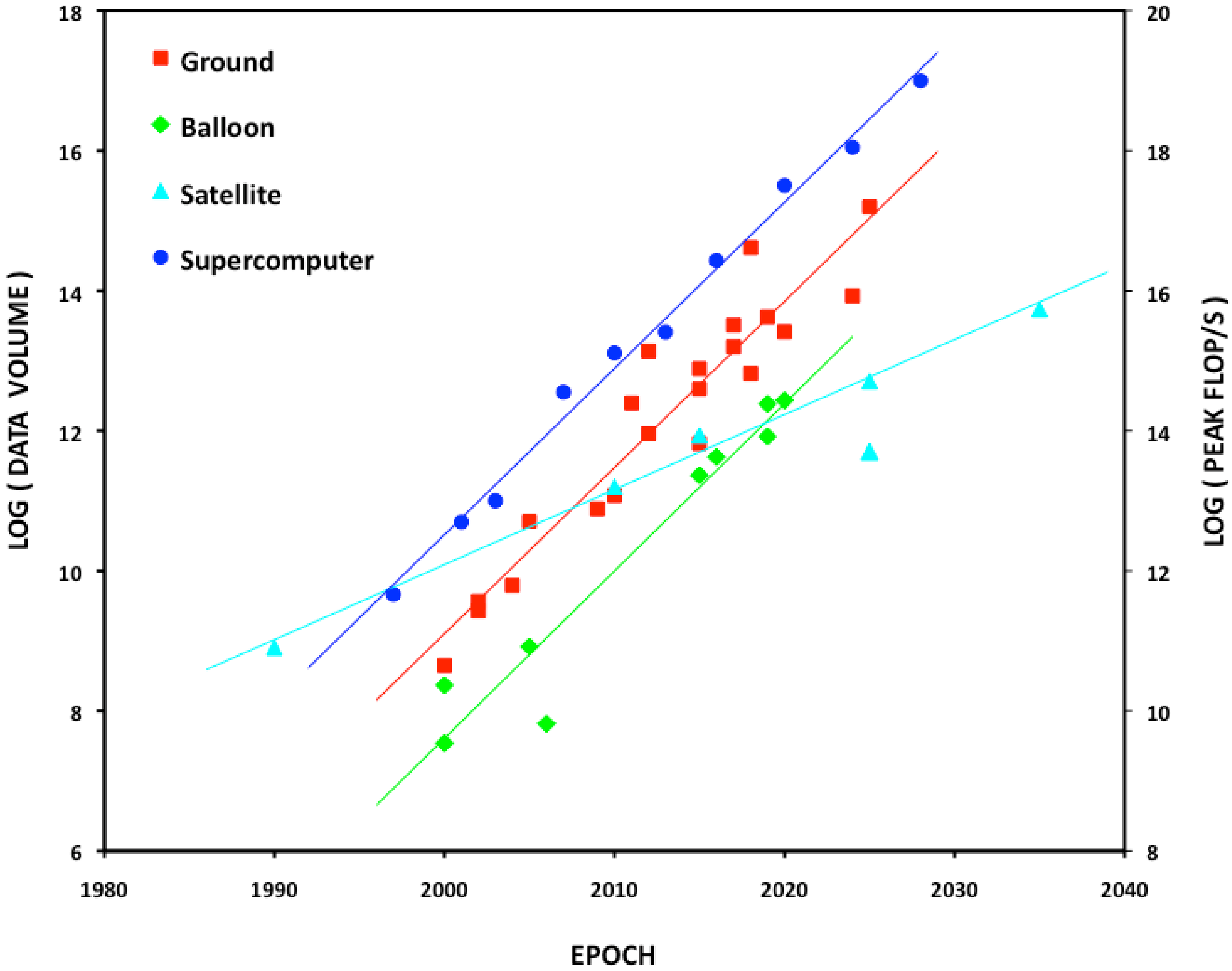
# Another View …



# CMB Signals

| COMPONENT | AMPLITUDE (K) | ERA |
|---|---|---|
| TT : Monopole | 1 | 1968 (Penzias & Wilson) |
| TT : Dipole | $10^{-3}$ | 1971 (Henry/Conklin) |
| TT : Anisotropy | $10^{-5}$ | 1990 (COBE) |
| TT : 1st Peak | $10^{-6}$ | 2000 (BOOMERanG, MAXIMA) |
| EE : Reionization | $10^{-7}$ | 2005 (DASI) |
| BB : Lensing | $10^{-9}$ | 2015 (SPT, POLARBEAR) |
| BB : Gravity Waves | $< 10^{-9}$ | ? |

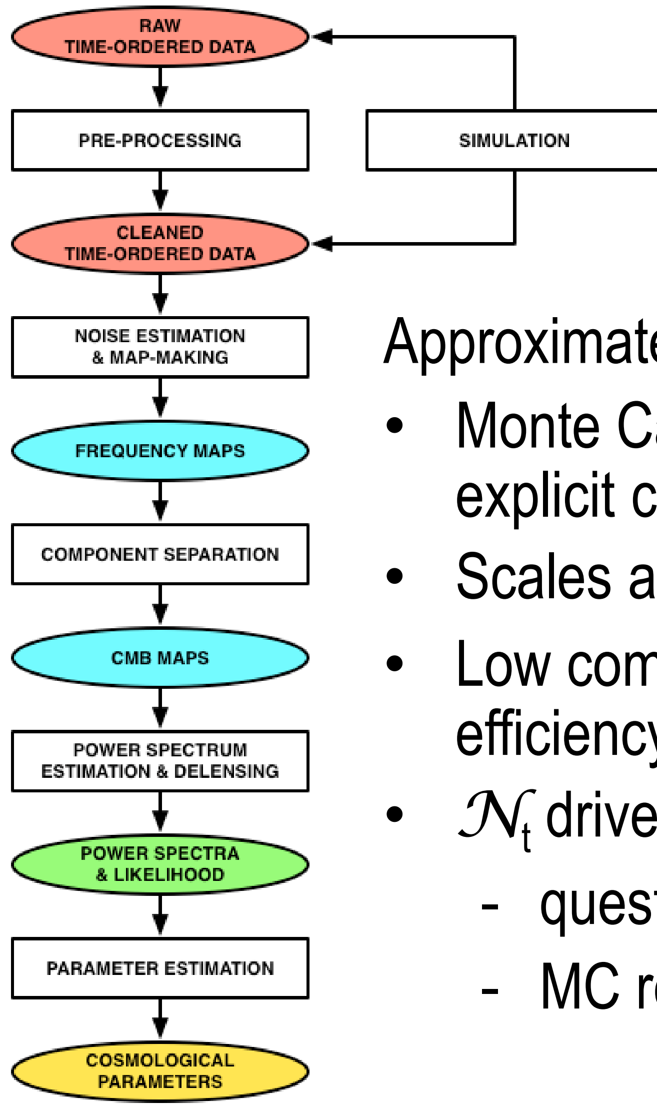
# Fainter Signals Mean Bigger Datasets



# CMB Data Analysis

Exact analysis:

- Maximize Gaussian likelihoods
- Scales as $O(\mathcal{N}_p^3)$
- High computational efficiency
- $\mathcal{N}_p$ driven up by
  - high resolution
  - large sky coverage
  - many frequencies
  - polarization




Approximate analysis:

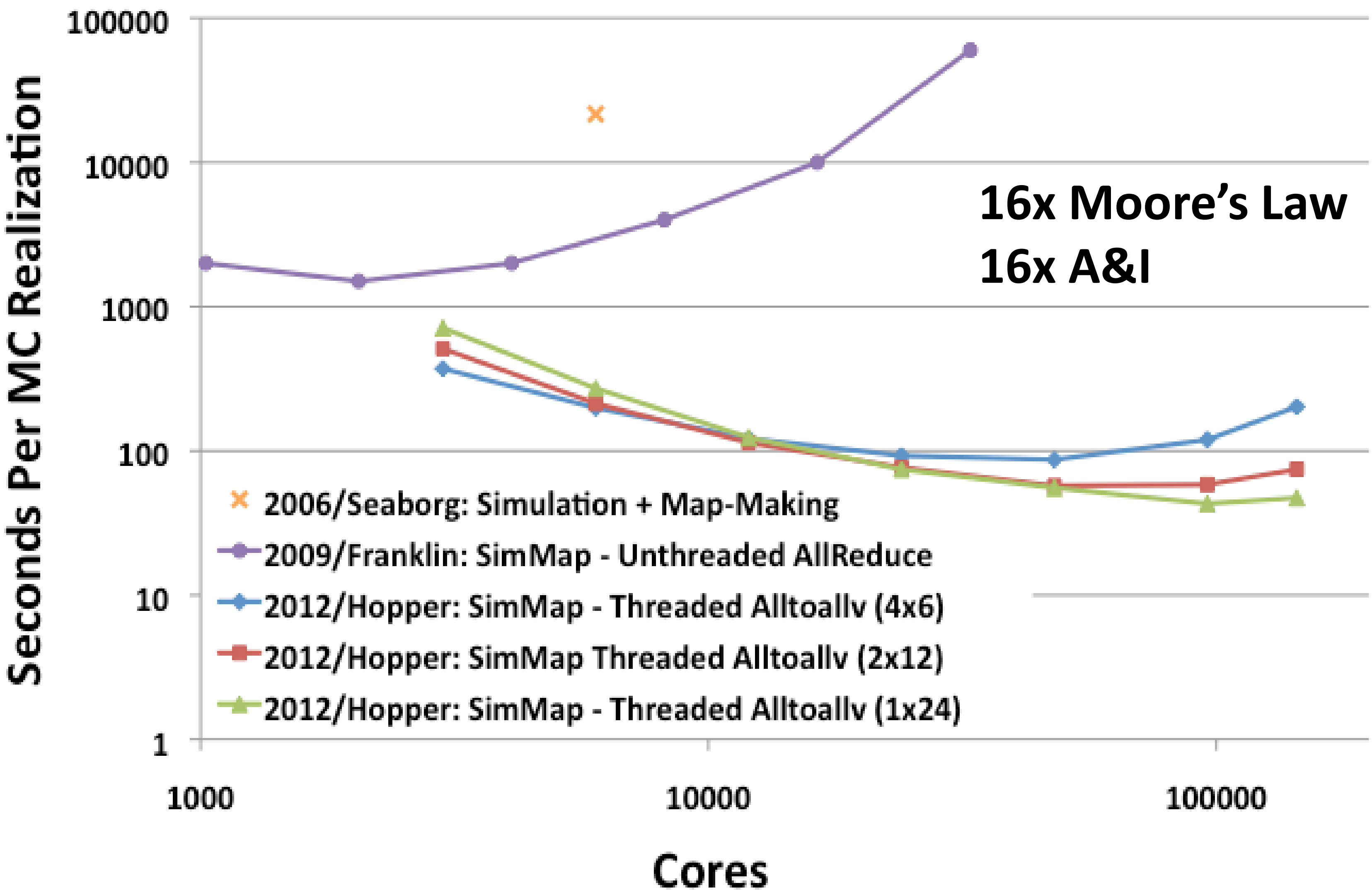
- Monte Carlo replaces explicit covariance
- Scales as $O(\mathcal{N}_t)$
- Low computational efficiency
- $\mathcal{N}_t$ driven up by
  - quest for S/N
  - MC realizations

# Prefactors & Efficiencies

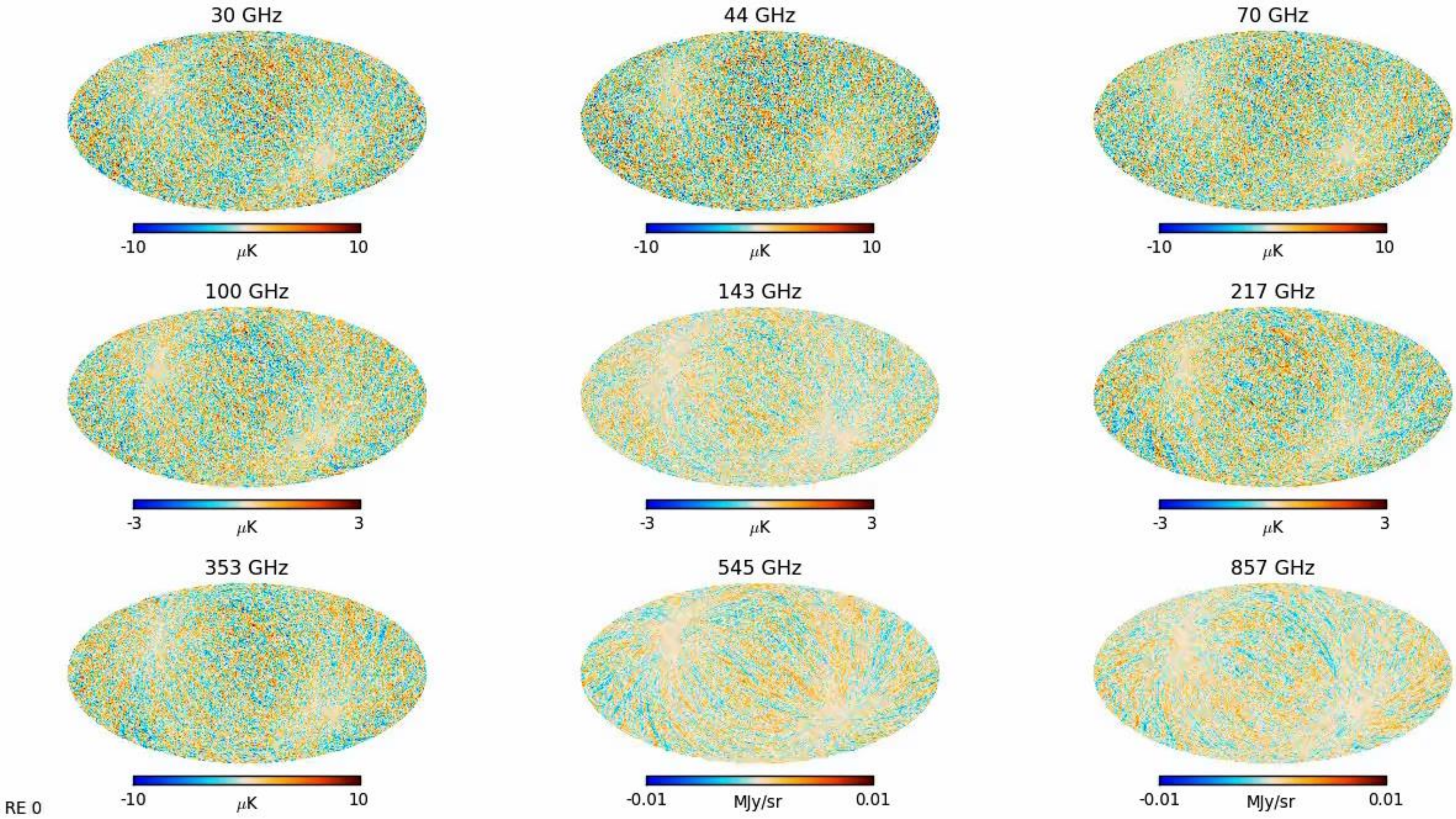
- Exact analysis dominated by covariance matrix operations
  - Floating point operations ~ $10^{2-3}\ \mathcal{N}_p^3$
  - Efficiency > 80%
  - Intractable* for $\mathcal{N}_p > 10^5$ (2000), $10^6$ (2015), $10^7$ (2030)
  - Only viable for small patch/low-resolution data
- Approximate analysis dominated by Monte Carlo map-making
  - Floating point operations ~ $10^4 \times 10^2 \times 10^1\ \mathcal{N}_t$
  - Efficiency < 1%
  - Intractable* for $\mathcal{N}_t > 10^9$ (2000), $10^{12}$ (2015), $10^{15}$ (2030)
  - Viable for current/future missions *if we can maintain efficiency*

* More than 1% NERSC total annual cycles

# Planck Monte Carlo Capability



# $10^4$ MC: Planck Noise




**~1 million CPU-days**

# High Performance Computing

- The most computationally challenging CMB analyses require:
  - Very many cycles
  - High bandwidth/low latency communication
  - Fast parallel I/O
- These features *define* high performance computing (HPC)
  - supercomputers, not grid/cloud/@home/etc.
- How these features are delivered evolves with architecture
  - clock-speed => node-count => accelerators => many-core => ?
  - communication and (especially) I/O grow slower than Moore's Law
- Analysis implementations have to evolve accordingly.

# Coming HPC Challenges

- HPC systems are becoming progressively
  - less homogeneous
  - less balanced
  - more power-constrained
- … likely requiring our implementations to incorporate
  - run-time self-tuning
  - energy-awareness
  - on-the-fly map analysis
  - unknown unknowns
- What do we do when Moore's Law stops?

# Mission-Class Computing

- Long-term CMB missions require:
  - Reliable allocation of persistently state-of-the-art HPC
    - Moore's Law growth during the mission lifetime
    - Over-scoped (shared) systems for bursty usage
  - Access for large numbers of users around the world
    - Cycles and storage both to produce and exploit reduced data
- eg. Planck at NERSC
  - 1% of NERSC cycles per year for 15 years
  - 5 generations of top-10 supercomputer
  - 200+ users from 10+ countries (& 100+ suborbital)

# HPC Centers

| | | Resource | | Access | |
|---|---|---|---|---|---|
| | | Capability (Top 10) | Roadmap (3+ Generations) | Users (100+ Worldwide) | Allocation (10+ Years) |
| USA | DOE/NERSC | #34/(top10) | | | |
| | DOE/LCFs | #2/#5 | | | |
| | NSF | - | | | |
| | NASA | #11* | | | |
| Europe | CINECA/BSC | #32/#77 | | | |
| | CSC | #51 | | | |
| Japan | Riken | #4 | | | |
| China | NSC | #1 (20MW) | | | |

# Conclusions

*The scientific return of next-generation CMB missions will depend on our ability both to access and to exploit the capabilities of the next several generations of challenging HPC architectures.*

*Planck has been the key vehicle for these (and many other) data analysis efforts for the last decade or more; with its ending we risk losing this critical expertise.*

*Analysis-induced systematics are a key – but widely neglected – issue for precision CMB measurements.*