Big Bang, Big Data, Big Iron:
High Performance Computing and the Cosmic Microwave Background

Julian Borrill

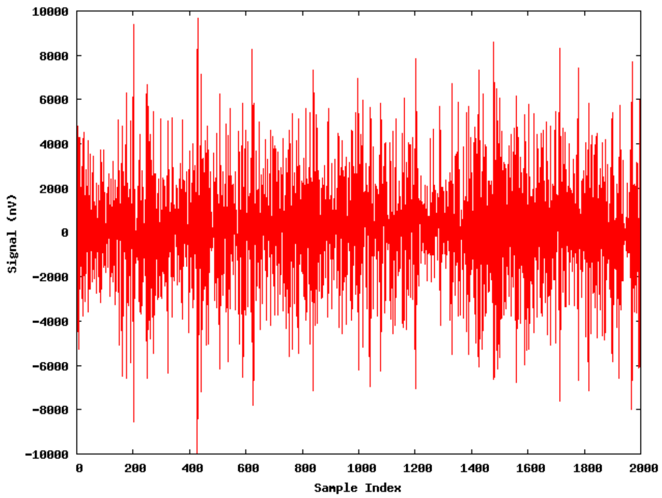
Computational Cosmology Center, Berkeley Lab

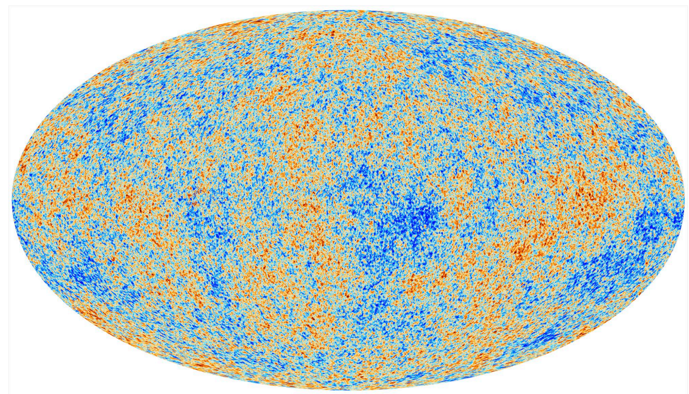
& Space Sciences Laboratory, UC Berkeley

CMB Science

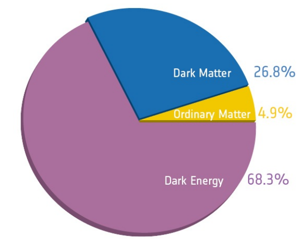
- Primordial photons trace the entire history of the Universe.
- Existence (monopole) distinguishes Big Bang from Steady State cosmology.
- Angular power spectra of tiny temperature & polarization fluctuations constrain fundamental parameters of cosmology & high energy physics.
- Distortions trace intervening matter
  - dark matter via weak lensing
  - galaxy clusters via Sunyaev-Zel'dovich effect
  - neutral hydrogen at recombination via Rayleigh scattering.
- All dismissed as undetectable curiosities when first predicted!
- The challenges are (i) detection and (ii) decoding.

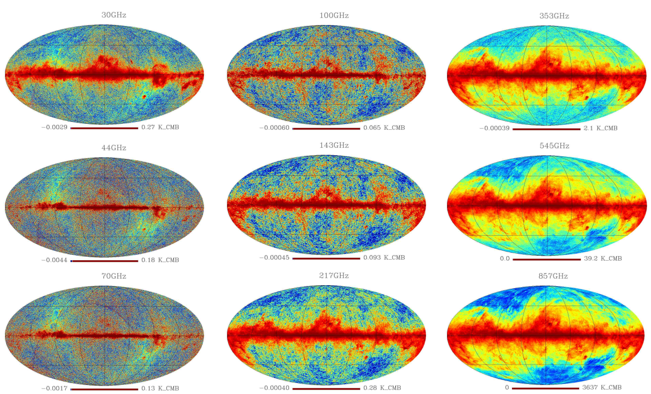
The CMB Data Sequence

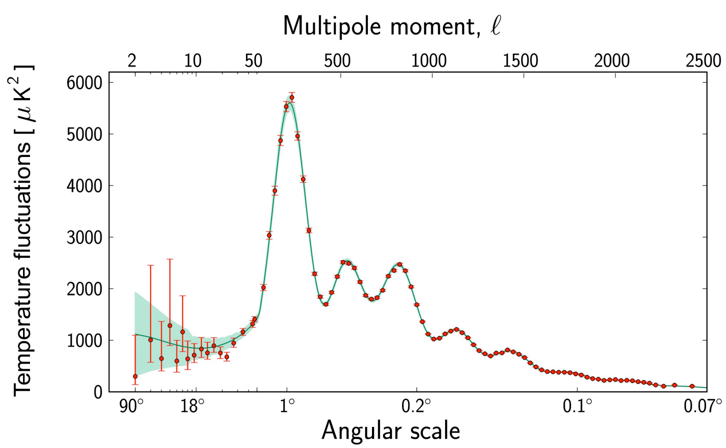




Parameter	*Planck* ("CMB+Lens")
$\Omega_b h^2$	0.02217 ± 0.00033
$\Omega_c h^2$	0.1186 ± 0.0031
Ω_Λ	0.693 ± 0.019
τ	0.089 ± 0.032
t_0 (Gyr)	13.796 ± 0.058
H_0 (km s^{-1}Mpc^{-1})	67.9 ± 1.5
σ_8	0.823 ± 0.018
Ω_b	0.0481^b
Ω_c	0.257^b







Ideal CMB Data Analysis

- Scan the sky measuring its temperature/polarization at many frequencies.
- Reduce the time-stream data to a total sky map at each frequency
  – Maximize a Gaussian likelihood for the map given the data and its noise correlations
- Combine the maps to extract a single CMB map
  – Use the different spectral dependencies to discriminate between the CMB & foregrounds
- Derive the angular power spectrum of the CMB map
  – Maximize a Gaussian likelihood for the power spectra given the map and its noise correlations.
- Derive the parameter likelihoods for any given cosmology
  – Use Monte Carlo Markov Chains over the parameter values via their theoretical spectra.

The CMB Data Challenge

- Extracting fainter signals (polarization, high resolution) requires:
  - larger data volumes to provide higher signal-to-noise.
  - more complex analyses to control fainter systematic effects.

Experiment	Start Date	Observations	Pixels
COBE	1989	10^9	10^4
eg. BOOMERanG	2000	10^9	10^6
WMAP	2001	10^{10}	10^7
Planck	2009	10^{12}	10^9
eg. PolarBear	2012	10^{13}	10^6
eg. Simons Array	2015	10^{14}	10^7
CMBpol, CORE, PRISM	2020+	10^{15}	10^{10}

- 1000x increase in data volume every 15 years – Moore's Law!
  - Need linear analysis algorithms & cutting-edge HPC systems.

C3 COMPUTATIONAL COSMOLOGY CENTER

BERKELEY LAB

A Stage IV CMB Experiment

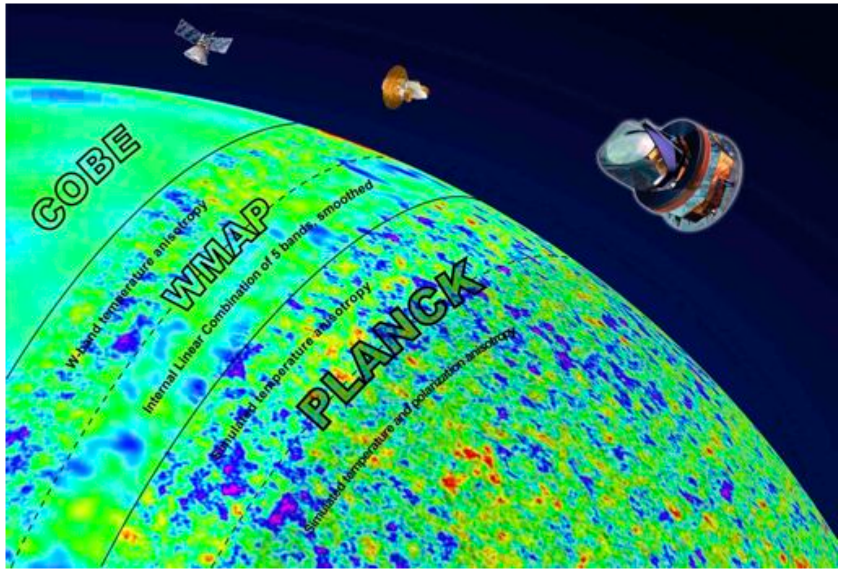
- CMB-S4 is a new proposal within the US DOE/NSF Snowmass process:
  - Search for cosmological B-modes
  - Measure sum of neutrino masses

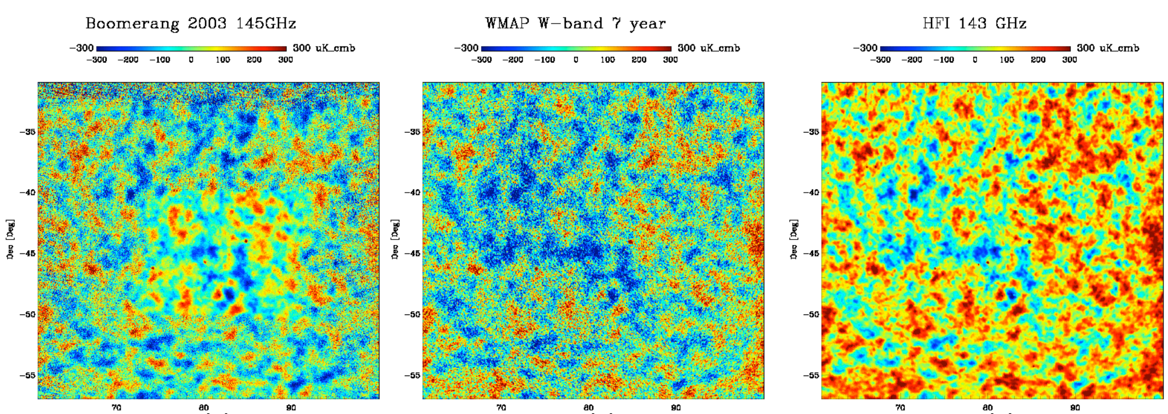
- Field 500,000 background-limited detectors sampling at 100 Hz for 5 years with a 70% duty cycle: $\mathcal{N}_t \sim 10^{16}$

- Survey 50% of the sky (using multiple telescopes/sites) over 40 – 240 GHz at 3 arcminute resolution: $\mathcal{N}_p \sim 10^{10}$

- Science goals require 10^3 times as many samples per pixel as Planck!

Evolving Sensitivity





C3 COMPUTATIONAL COSMOLOGY CENTER

BERKELEY LAB

Practical CMB Data Analysis

- Exact solutions involve both the map and its (dense) correlation matrix.
  - Solutions scale as N_p^2 in memory, N_p^3 in operations
  - Impractical (to date) for $N_p \geq 10^6$ (terabyte, exaflop)
- Instead use approximate solutions:
  - Solve for map only using preconditioned conjugate gradient
    - Scales as $N_i\ N_t$
  - Solve for pseudo-spectra only using spherical harmonic transforms
    - Scales as $N_p^{3/2}$
  - Debias and quantify uncertainty using Monte Carlo methods: simulate and map $10^2 – 10^4$ realizations of the data (10 – 1% UQ)
    - Scales as $N_r\ N_i\ N_t$

COMPUTATIONAL COSMOLOGY CENTER

BERKELEY LAB

CMB Data Analysis Evolution

Data volume & computational capability dictate analysis approach.

Date	Data	System	Map	Power Spectrum
1997 - 2000	B98	Cray T3E x 700	Explicit Maximum Likelihood (Matrix Invert - N_p^3)	Explicit Maximum Likelihood (Matrix Cholesky + Tri-solve - N_p^3)
2000 - 2003	B2K2	IBM SP3 x 3,000	Explicit Maximum Likelihood (Matrix Invert - N_p^3)	Explicit Maximum Likelihood (Matrix Invert + Multiply - N_p^3)
2003 - 2007	Planck SF	IBM SP3 x 6,000	PCG Maximum Likelihood (band-limited FFT – few N_t)	Monte Carlo (Sim + Map - many N_t)
2007 - 2010	Planck AF EBEX	Cray XT4 x 40,000	PCG Maximum Likelihood (band-limited FFT – few N_t)	Monte Carlo (SimMap - many N_t)
2010 - 2013	Planck MC PolarBear	Cray XE6 x 150,000	PCG Maximum Likelihood (band-limited FFT – few N_t)	Monte Carlo (Hybrid SimMap - many N_t)

High Performance Computing

- Supercomputer components:
  - Input/output: moving data between disk and memory
  - Communication: moving data between remote memory locations
  - Calculation: moving data from local memory to processor & acting on it
- Moore's Law: calculation capability doubles every 18 months
  - Clock speed
  - Core count
  - Accelerators
  - What next?
- Computational efficiency is critical
  - Data delivery is the challenge: IO < COMM < CALC

CMB Supercomputing At NERSC

- Almost all CMB experiments have used supercomputers at the DOE's NERSC Center for the last 15 years.
  - Shared allocation for suborbital experiments (~5M CPU-hours/year)
  - Dedicated resources for Planck (~20M CPU-hours/year)

- New top-10 supercomputer every 2-3 years
  - 6 generations of supercomputers
  - 1000x increase in capability

- Open to (almost) anyone in the world
  - https://nim.nersc.gov/nersc_account_request.php & repo mp107
  - provide access to full public data, beyond archive capabilities (MCs)

Simulation & Mapping: Calculations

Given the instrument noise statistics & beams, a scanning strategy, and a sky:

1) SIMULATION: $d_t = n_t + s_t = n_t + P_{tp}\, s_p$
   - A realization of the piecewise stationary noise time-stream:
     - Pseudo-random number generation (caution!) & FFT
   - A signal time-stream scanned & beam-smoothed from the sky map:
     - SHT

2) MAPPING: $(P^T N^{-1} P)\, d_p = P^T N^{-1} d_t$ $(A\,x = b)$
   - Build the RHS
     - FFT & sparse matrix-vector multiply
   - Solve for the map
     - PCG over FFT & sparse matrix-vector multiply

C3 COMPUTATIONAL COSMOLOGY CENTER

BERKELEY LAB

The Planck Challenge

- Analysis *completely* dominated by simulation/map-making (10^4 x 10^2 x 10^{12})
- The first Planck single-frequency simulation & map-making took 6 hours on 6000 CPUs in 2006:
  – 36,000 CPU-hours per realization
- Our goal was 10,000 realizations of all 9 frequencies in 2012
  – With no change => 3 x 10^9 CPU-hours
  – With Moore's Law => 2 x 10^8 CPU-hours
  – NERSC quota => O(10^7) CPU-hours
- Requirements
  – Ability to exploit 4 iterations of Moore's Law, regardless of approach
  – Additional O(20x) algorithmic/implementation speed-up

The Baseline Implementation

Using one MPI task per core:

- For each realization — LOOP: N_r
  - Simulation:
    - Read detector pointing — I/O: $O(N_t)$
    - Simulate detector timestream — CALC: $O(N_t)$
    - Write detector timestream — I/O: $O(N_t)$
  - Map-making
    - Read detector pointing & timestream — I/O: $O(N_t)$
    - At each iteration — LOOP: N_i
      - Make local submap — CALC: $O(N_t)$
      - Allreduce to global map — COMM: $O(N_p \log_2 T)$
    - Write map — I/O: $O(N_p)$

The Current Implementation

Using one MPI task per node + threads

- Read sparse telescope pointing — I/O: $O(N_t)$*
- Reconstruct detector pointing — CALC: $O(N_t)$
- For each realization — LOOP: N_r
  - Simulate detector timestream — CALC: $O(N_t)$
  - For each iteration — LOOP: N_i
    - Make local submap — CALC: $O(N_t)$
    - Scatter/gather global map — COMM: $O(N_p \log_2 T')$**
  - Write map — I/O: $O(N_p)$

* Prefactor reduced by number of detectors & dense/sparse sampling ratio

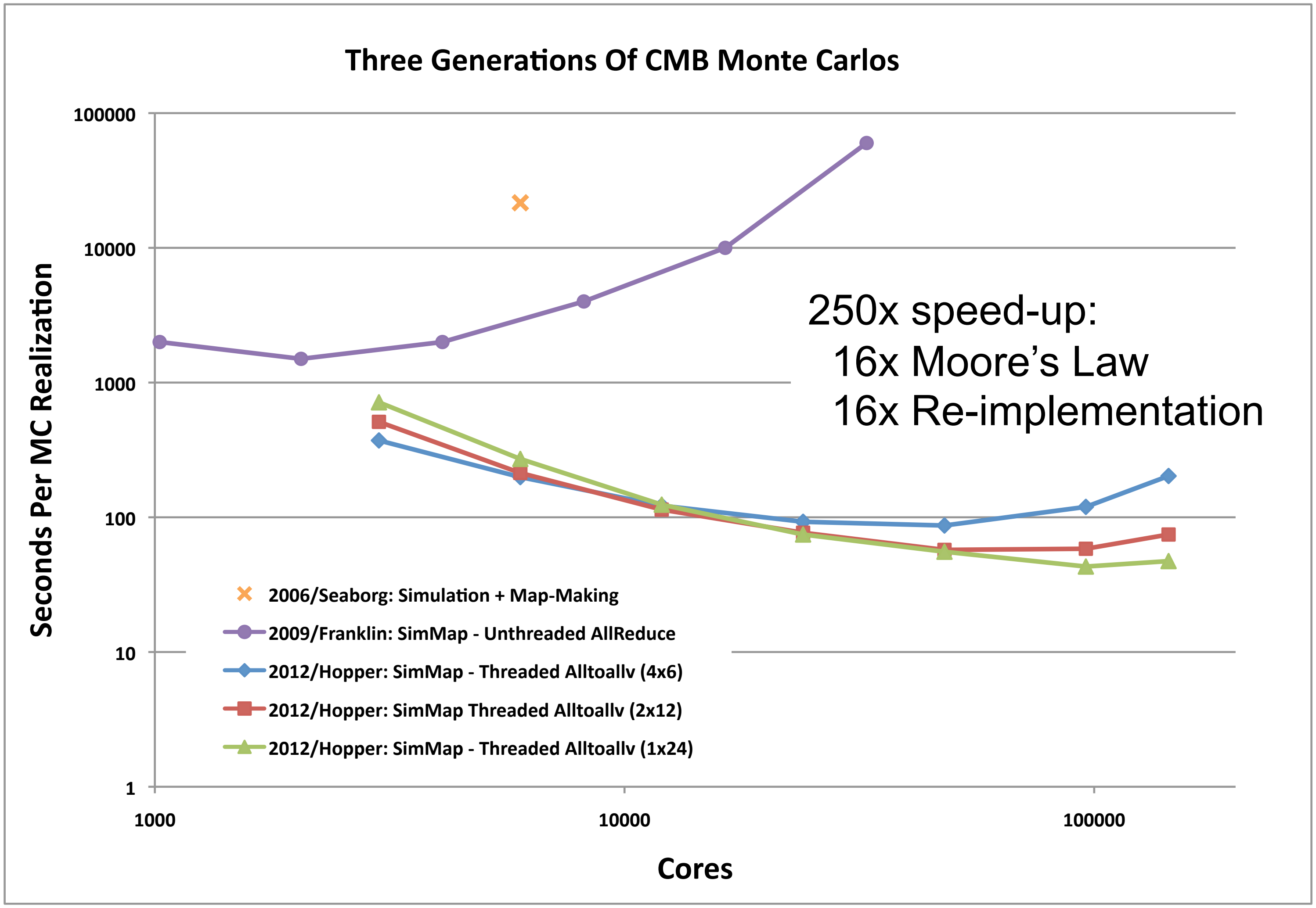
** Prefactor reduced by submap overlap factor

Efficiencies

	I/O	COMM	CALC
BEFORE	$(3 + 1)\ N_r\ N_t$	$N_r\ N_i\ N_p \log_2 T$	$N_r\ (1 + N_i)\ N_t$
AFTER	$10^{-4}\ N_t$	$10^{-2}\ N_r\ N_i\ N_p \log_2 T'$	$1 + N_r\ (1 + N_i)\ N_t$
SPEED-UP	$10^4\ N_r \sim 10^8$	$10^{-2} \log_2 T / \log_2 T' \sim 500$	1

- IO efficiencies:
  - Pull all common data outside of MC loop.
  - Perform pointing reconstruction & simulation on the fly.
- COMM efficiencies
  - Reduce number of MPI tasks by hybridizing the code.
  - Minimize communication volume by calculating pair-wise pixel overlaps.
- IO & CALC scale with N_t, COMM with N_p – sensitivity with N_t/N_p.

Planck Simulations Over Time

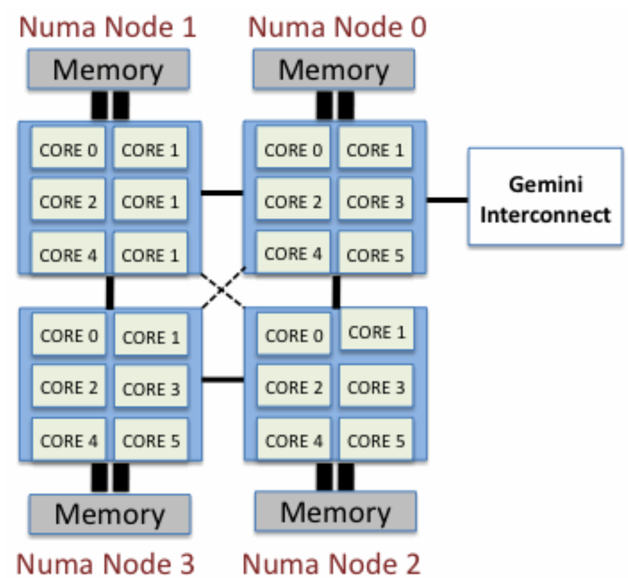


COMPUTATIONAL COSMOLOGY CENTER

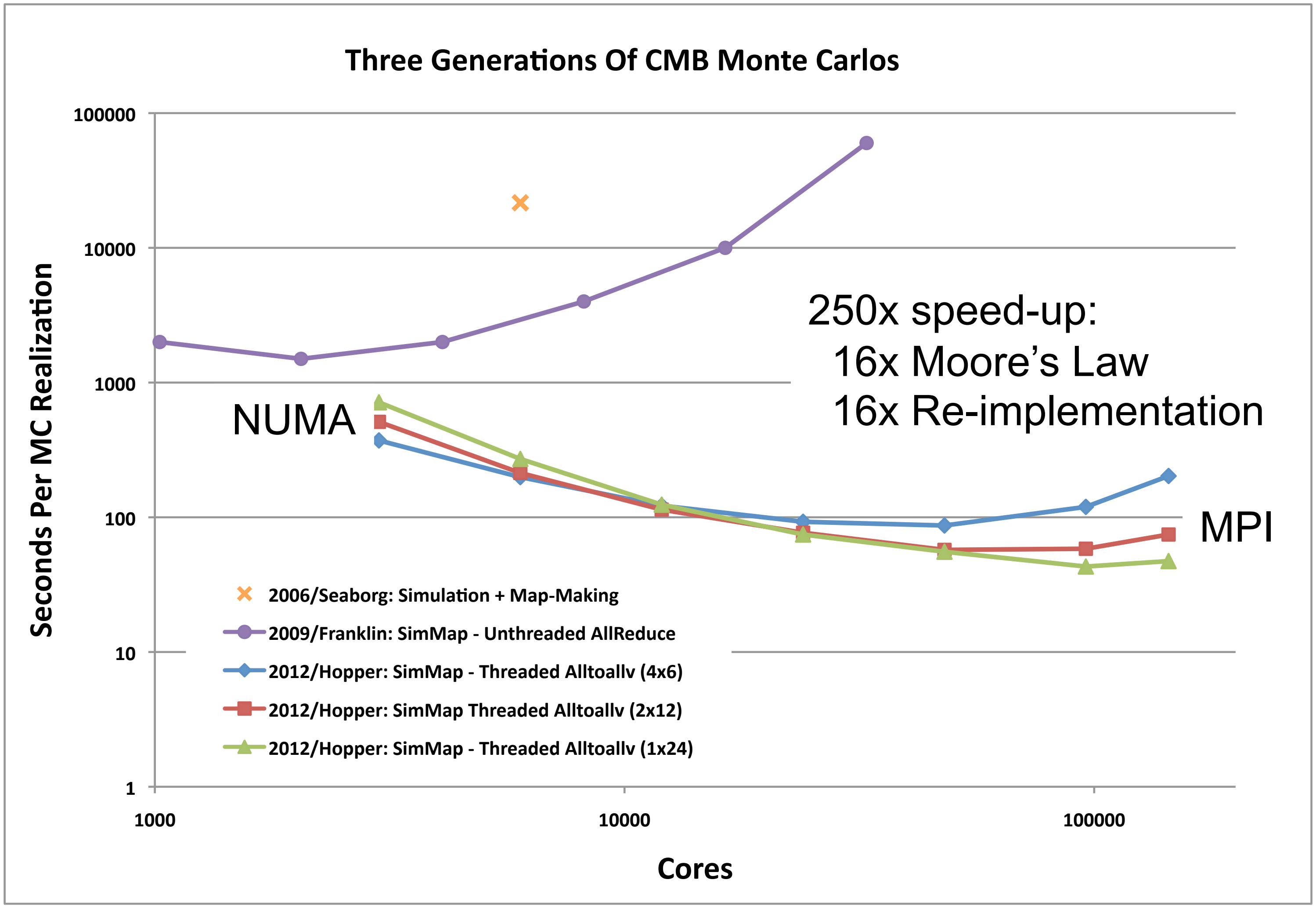
BERKELEY LAB

HPC System Evolution

- Clock speed is no longer able to maintain Moore's Law.
- Multi-core CPU and GPGPU are two major approaches.
- Both of these will require
  - significant code development
  - performance experiments & auto-tuning
- E.g. NERSC's Cray XE6 system *Hopper*
  - 6384 nodes
  - 2 sockets per node
  - 2 NUMA nodes per socket
  - 6 cores per NUMA node
- What is the best way to run hybrid code on such a system?

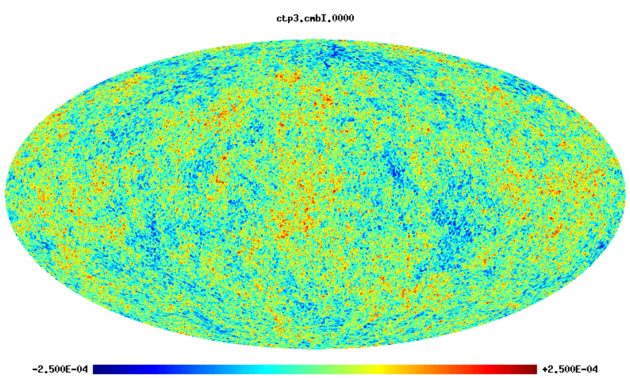


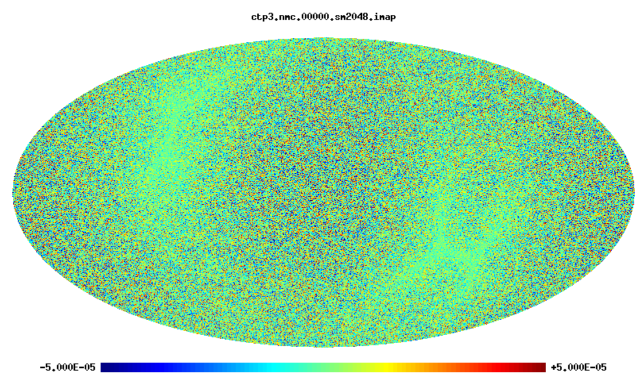
Configuration With Concurrency

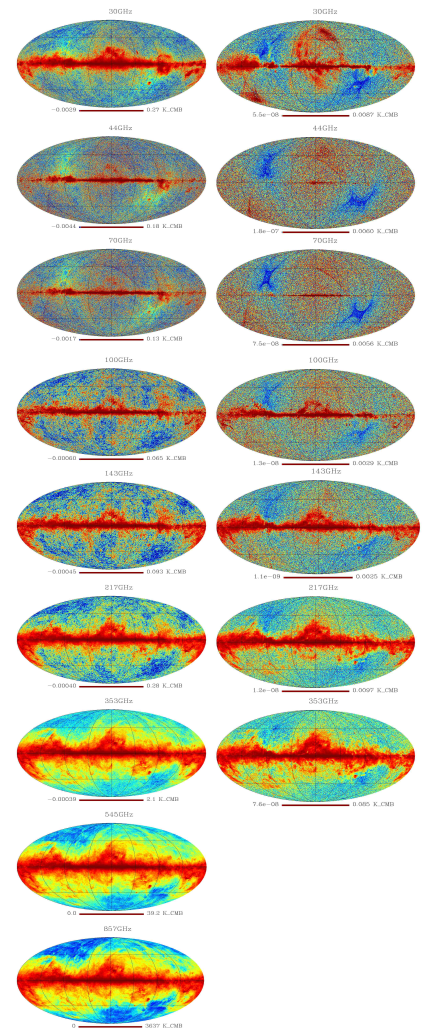


Planck Full Focal Plane 6

- 6th full-mission simulation set used for all 2013 results.
- Single fiducial sky for validation & verification.
- 1,000 CMB & noise realizations for debiasing and uncertainty quantification.
- 250,000 maps in total – largest CMB MC set ever.
- 15M CPU-hours running up to 100,000 way parallel.







COMPUTATIONAL COSMOLOGY CENTER

BERKELEY LAB

Future Prospects

- Next Planck releases (2014 & 2015) will require 10x MC realizations.
- Next-generation B-mode experiments will gather
  - 10x Planck: current suborbital
  - 100x Planck: future suborbital
  - 1000x Planck: future satellite
- Next-generation supercomputers will have
  - Heterogeneous nodes
  - Varied accelerators (GPU, MIC, …)
  - Higher concurrency (?)
  - Limited power

Conclusions

- Planck has been spectacularly successful, and the best is yet to come!
  – As much data again, plus polarization.
  – The definitive CMB data set for the next decade or more.
  – Probing physics and cosmology beyond their standard models.
- Planck (and post-Planck) data analysis is absolutely reliant on HPC capability and capacity
  – Upper bounds on both CPU and wallclock-hours.
  – Guaranteed multi-year NERSC access was critical.
  – Performance goals require Moore's Law *and* improved implementations.

The scientific return of present and future CMB data sets will be constrained by our computational capability and our ability to exploit it.