Massive Monte Carlos For CMB Data Analysis: Planck Full Focal Plane Simulations

Julian Borrill

Berkeley Lab & UC Berkeley

On behalf of the Planck collaboration

Types & Roles Of Simulations

- Data Processing Centres (DPC)
  - Single instrument; gory detail
  - Validate & verify characterization & pre-processing
  - Quantify residuals
- Full Focal Plane (FFP)
  - Both instruments; only key effects
    - assumes perfect pre-processing
    - includes flags, bandpasses, beams, noise
  - Validate & verify analysis pipelines
  - Quantify uncertainties & correct biases
    - Monte Carlos on supercomputers

Full Focal Plane Simulations

- Single fiducial realization
  - Foregrounds + CMB + noise together
  - Used for validation & verification
- Monte Carlo simulation sets
  - CMB & noise separately
  - 100 – 10,000 realizations (10 – 1% statistics)
  - Used for uncertainty quantification & debiasing
  - Dominates computational challenge
- Distinct pipelines for fiducial, CMB MC & noise MC.

The Computational Challenge

- ~~"… it is prohibitively expensive to run a large set of end-to-end simulations that would capture all aspects of the map-making pipeline, and the noise characteristics and correlations in the actual data set." Das et al, *ACT 3 Season Power Spectrum*.~~
- Not prohibitive – challenging but critical.
- Operations ~ Realizations x Iterations x Samples
- For Planck 2013: $10^3 \times 10^2 \times 10^{12} \Rightarrow 10^{17}$ flops
  - Numerical prefactors: O(10 – 100)
  - Computational efficiency: O(10 – 1%)
- Requires $10^5 - 10^7$ GHz-CPU-hours

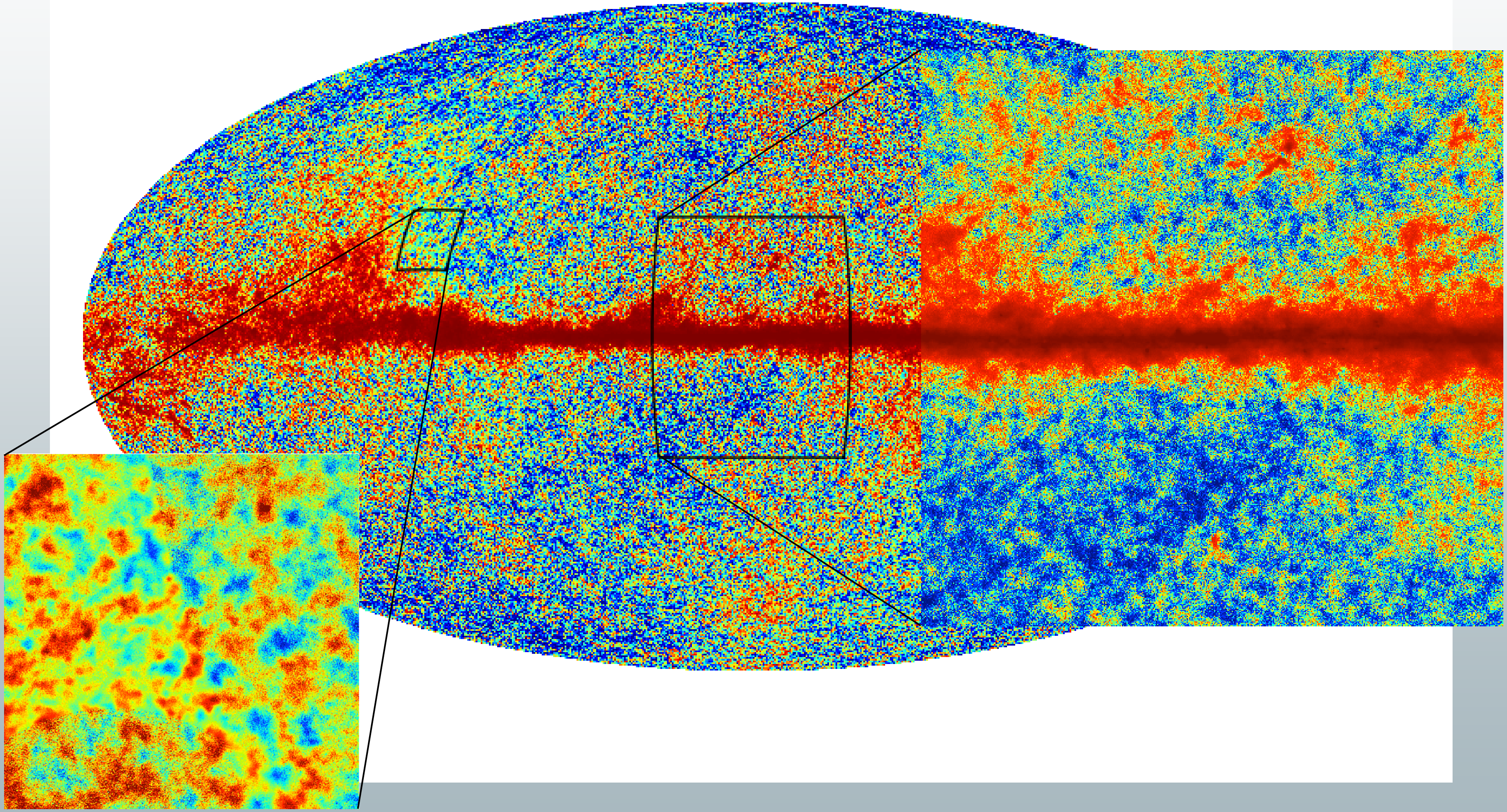
Full Focal Plane 6

- FFP1-5 focused on simulation & analysis validation.
- FFP6 focused on uncertainties & biases.
- Generated as part of Planck 2013 data analysis:
  - 10 component fiducial foreground
  - 1,000 MC realizations
  - 40,000 timestreams
  - 250,000 maps
  - 10,000,000+ CPU hours
- The most massive CMB simulation-set ever fielded.
- 10% of what will be required for final releases.

Fiducial Realization

- Planck Sky Model (Delabrouille et al)
  - Input: foreground component maps & scaling, detector bandpasses
  - Output: per-detector bandpassed component maps, strong point source catalog, CMB map
- LevelS (Reinecke et al)
  - Input: pointing, beams, detector maps & catalog
  - Output: per detector beam-convolved timestream
- MADAM/TOAST (Keihanen et al; Kisner et al)
  - Input: pointing, flags, detector timestreams
  - Output: 117 maps per component & total

Example: Fiducial Sky



CMB Monte Carlo

- Absent bandpass & noise issues, CMB realizations can be generated primarily in the map domain.

- FEBeCoP/TOAST (Mitra et al; Kisner et al)
  - Input: pointing, flags, beams, CMB MC maps
  - Intermediate: effective beam matrix
    - weights of nearby in-pixels in each out-pixel
  - Output: beam-convolved CMB MC maps
  - Repeat for each detector/data subset

Noise Monte Carlo

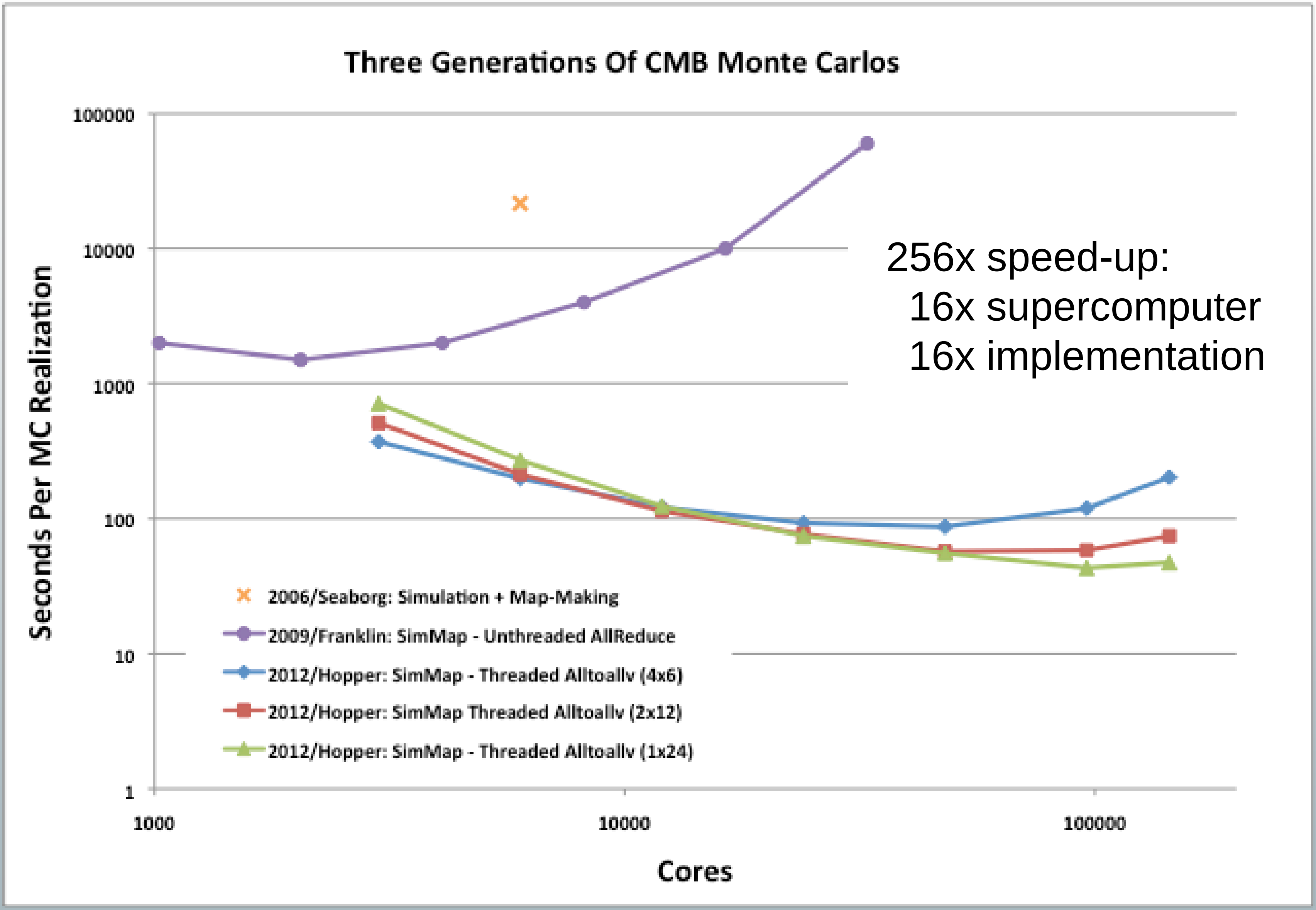
- Capturing full details of noise correlations require simulation in time-domain & mapping.

- MADAM/TOAST (Keihanen et al; Kisner et al)
  - Input: pointing, flags, noise (PSD, parameters)
  - Intermediate: on-the-fly noise timestreams
  - Output: full & half-ring noise maps
  - Repeat for each detector/data subset

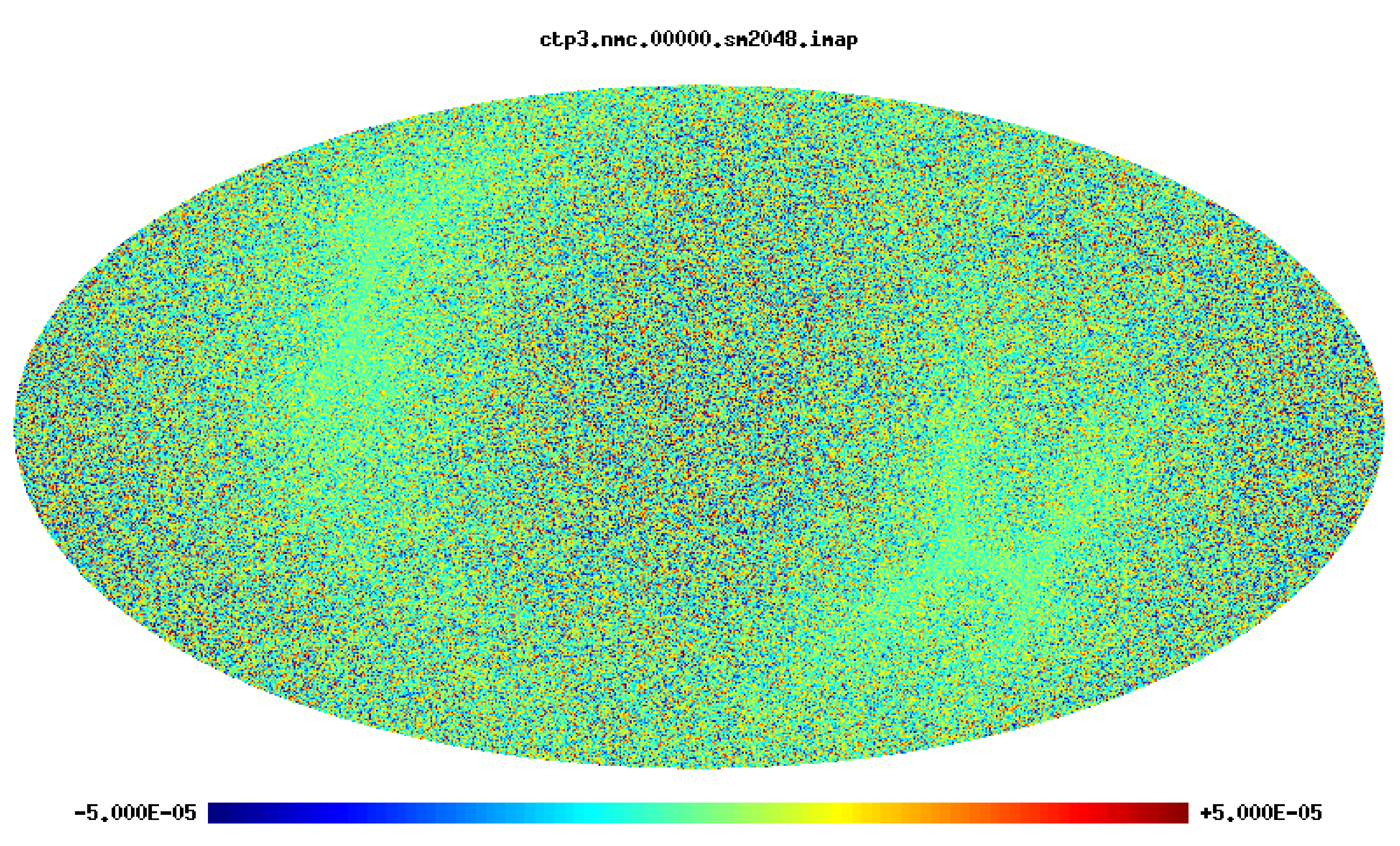
Optimizations

- I/O optimization
  - Replace I/O with (re)calculation
    - On-the-fly pointing reconstruction
      - Sparse boresight => dense detector
    - On-the-fly timestream generation
      - PRNG => independent white noise streams
      - PSD/parameters => coloured noise
- Communication optimization
  - Reduce number and volume of messages
    - Point-to-point over global reduction

Supercomputers & Implementations



Example



Future Prospects

- Final Planck releases 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
  - 1,000,000+ cores
  - Heterogeneous nodes
  - Varied accelerators (GPGPU, MIC, …)
  - Limited power