Big Bang, Big Data, Big Iron

Planck Satellite Data Analysis At NERSC

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

The Cosmic Microwave Background

- Hot Big Bang => expanding, cooling Universe.
- After 370,000 years temperature drops to 3000K.
- $p^+ + e^- \Rightarrow H$: Universe become neutral & transparent.
- Photons free-stream to observers today.

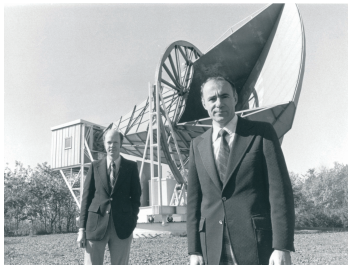
COSMIC: filling the Universe, seen by any observer looking in any direction.

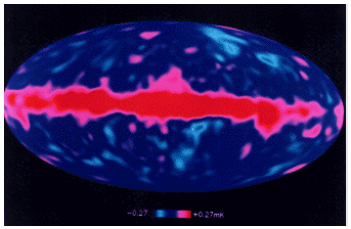
MICROWAVE: red-shifted from 3000K to 3K.

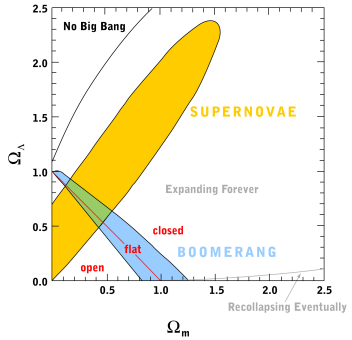
BACKGROUND: primordial photons coming from behind all astrophysical sources.

CMB Science

- Existence confirms Big Bang
  - Penzias & Wilson (1974)
- Tiny fluctuations encode the fundamental parameters of cosmology & high energy physics
  - Mather & Smoot (2006)
- Critical complementarity to astrophysical probes
  - Perlmutter, Riess & Schmidt (2011)



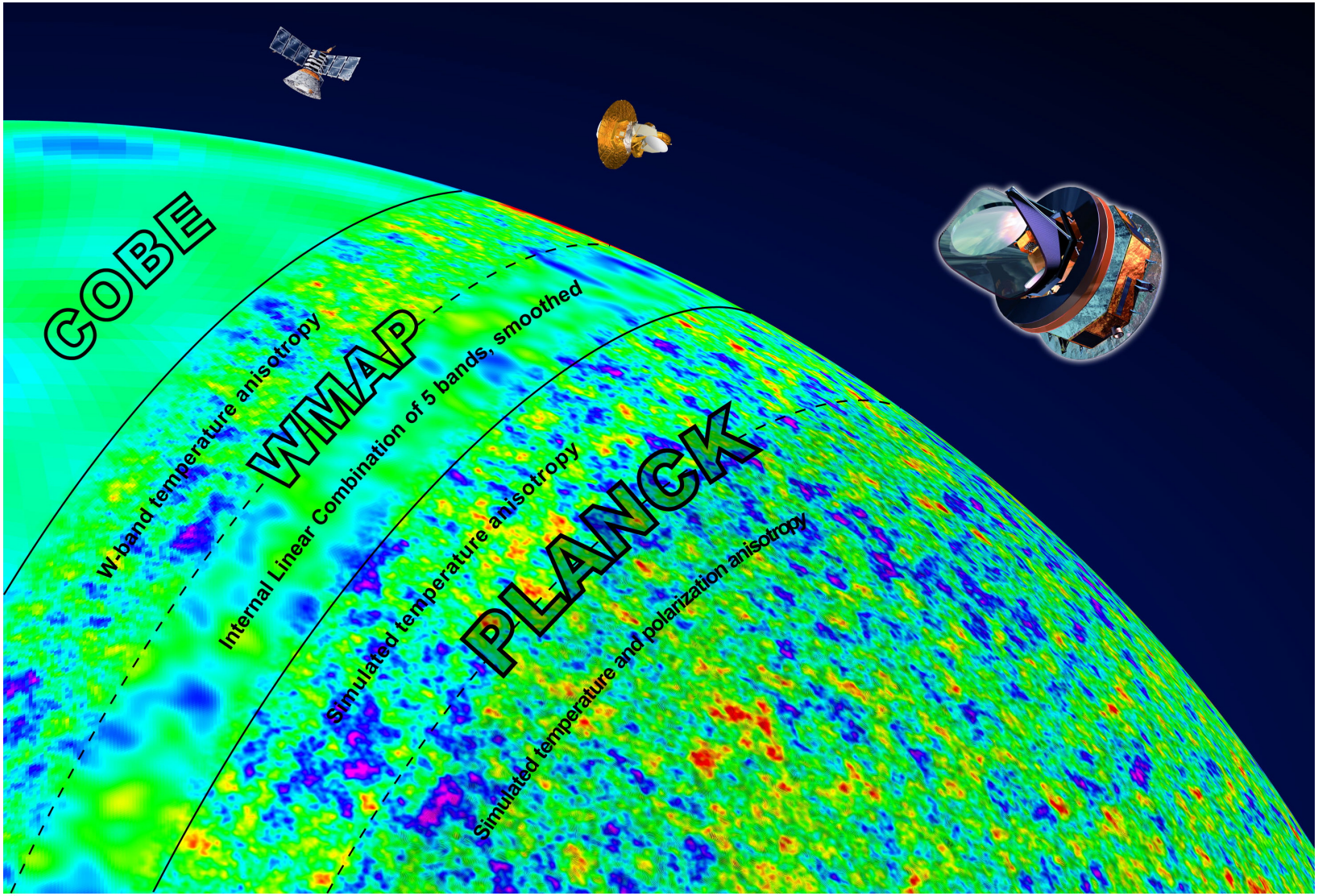




What is Dark Energy? (Planck + DES/BigBOSS/LSST/Euclid)

What is Inflation? (CMBpol + ~~LHC~~)

CMB Satellite Missions



CMB Data Challenge

- Fainter signals require larger datasets
  - COBE: 10^9 samples
  - WMAP: 10^{11} samples
  - Planck: 10^{12} samples
  - CMBpol: 10^{15} samples
- Data volume grows with Moore's Law
  - Weak scaling
- Missions span many years from concept to results
  - Strong scaling

A reliable source of supercomputing is essential.

Planck At NERSC

- Reliability (continuity + growth)
  - 15 years & 6 systems at NERSC
- User support
  - 100 data analysts of varying sophistication
  - International collaboration (24/7 support)
- Mission support
  - Exceptional level of service provision
  - Working meeting & critical deadline boost
  - Early access to new facilities/features
    - NGF, pseudo-user accounts, fractional systems

DOE/NASA MOU

- Signed in 2007, renewed in 2010
- DOE guarantees an annual NERSC/HEP allocation for the mission lifetime:
  - sets minimum level, request/receive much more!
- NASA provides:
  - FTEs to ensure efficient use of these resources
  - funds for exceptional levels of service
    - 32TB => 100TB NGF
    - planck cluster => cabinet of Carver

Provides a model for mission-scale projects.

CMB Data Analysis Evolution

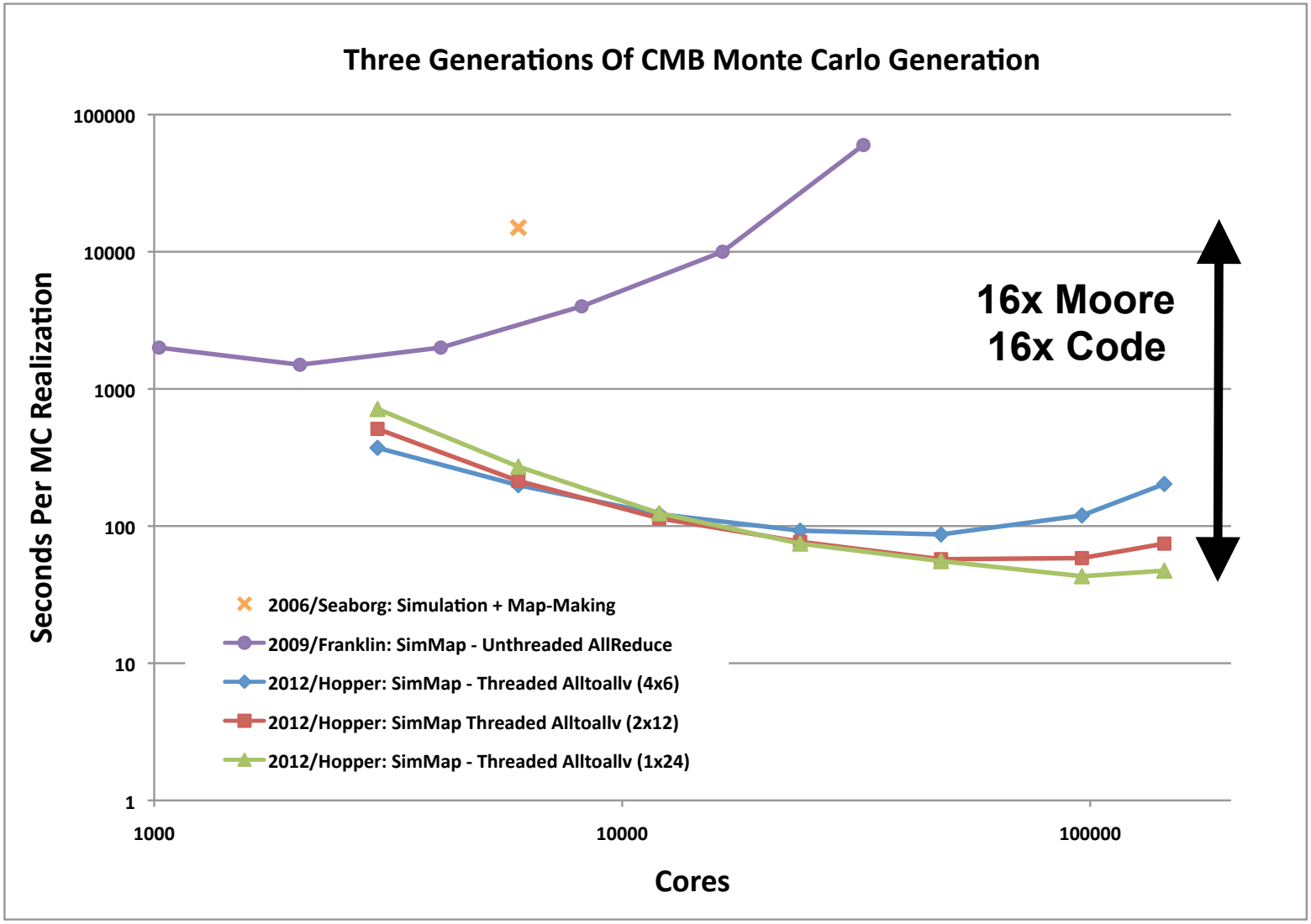
Date	Data	System	Map	Power Spectrum
2000	B98	Cray T3E x 700	Explicit Maximum Likelihood (Matrix Invert - N_p^3)	Explicit Maximum Likelihood (Matrix Cholesky + Tri-solve - N_p^3)
2002	B2K2	IBM SP3 x 3,000	Explicit Maximum Likelihood (Matrix Invert - N_p^3)	Explicit Maximum Likelihood (Matrix Invert + Multiply - N_p^3)
2003-7	Planck subsets	IBM SP3 x 6,000	PCG Maximum Likelihood (FFT – N_t)	Monte Carlo (Sim + Map - many N_t)
2007+	Planck full EBEX	Cray XT4 x 40,000	PCG Maximum Likelihood (FFT – N_t)	Monte Carlo (SimMap – very many N_t)
2010+	Planck MC PolarBear	Cray XE6 x 160,000	Destriping & PCG (FFT – N_t)	Monte Carlo (SimMap – very many N_t)
2015+	Towards CMBpol	Addressing the challenges of 1000x data & next 10 generations of HPC systems.		

Planck Data Analysis

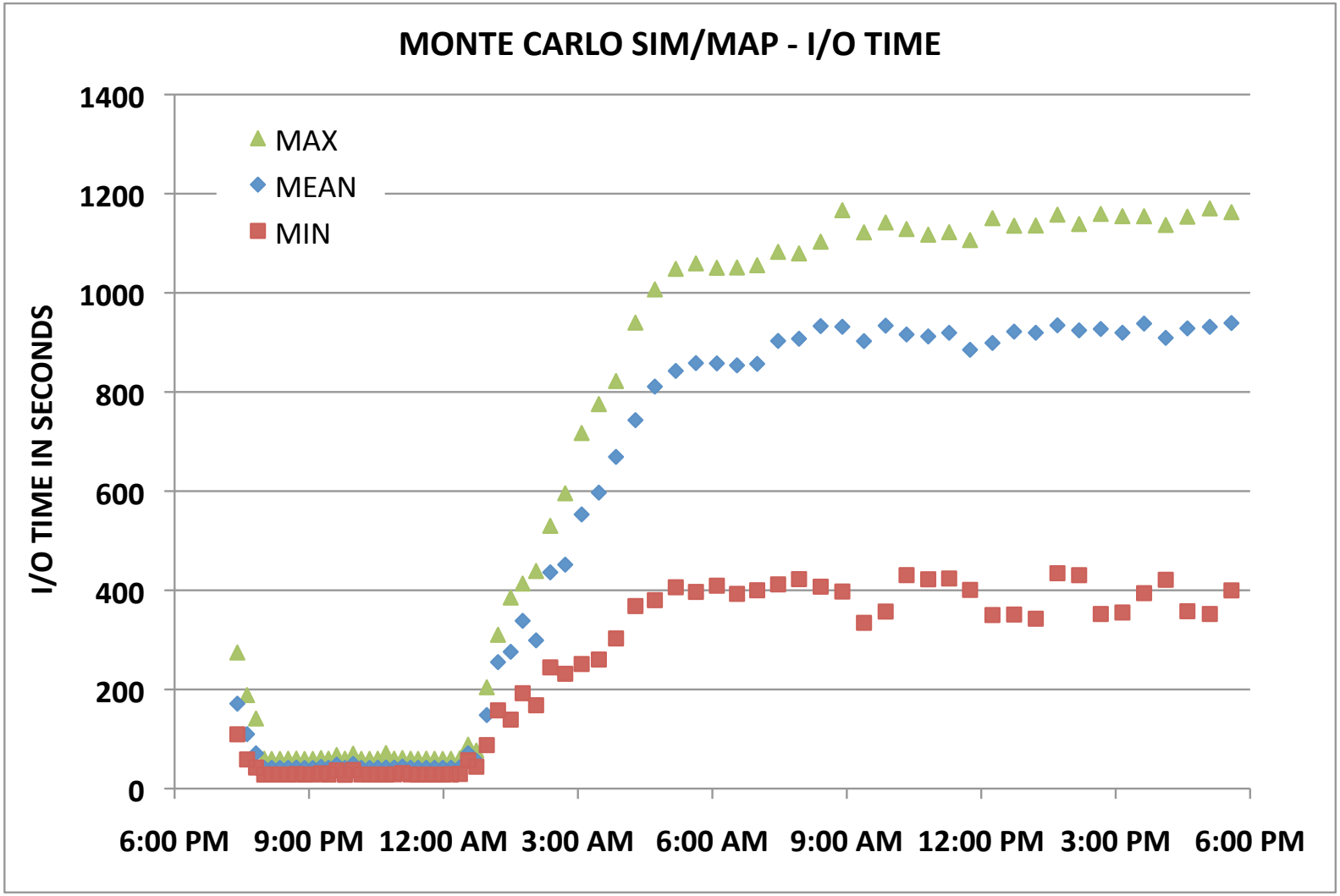
- CMB data analysis is dominated by Monte Carlos
  - Simulate full mission data & reduce to maps
  - Uncertainty quantification, de-biasing, analysis V&V
  - Scales as $N_{mc} \times N_{it} \times N_t \sim 10^4 \times 10^2 \times 10^{12}$ for Planck
- First full-scale Planck simulation & map-making in 2007
  - First science code to run on 6000 cores of Seaborg.
  - 4 wallclock- & 24K CPU-hours for single realization.
- Even allowing for 16x speed-up from Moore's Law*, this projects to 100 days & 240M CPU-hours for in 2013!

*16x => perfect strong scaling to 100K cores

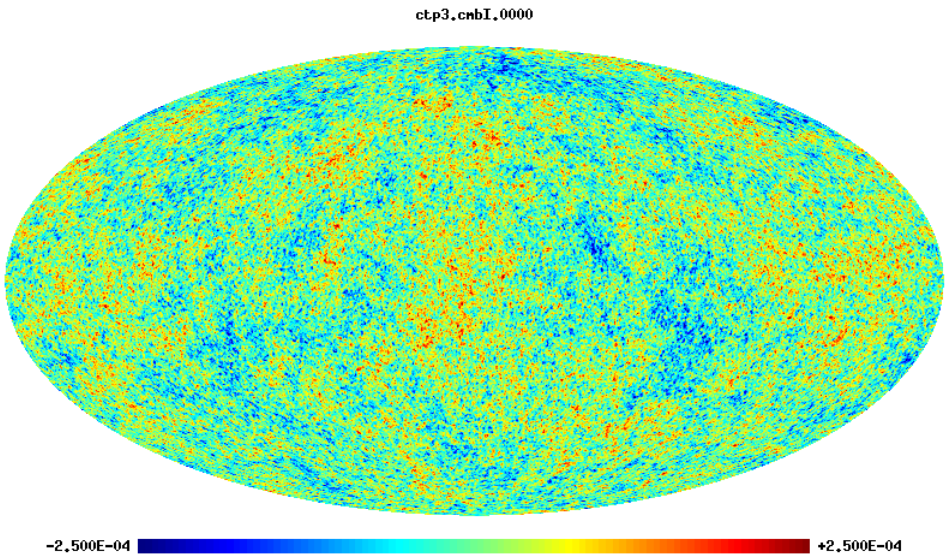
Implementations & Architectures

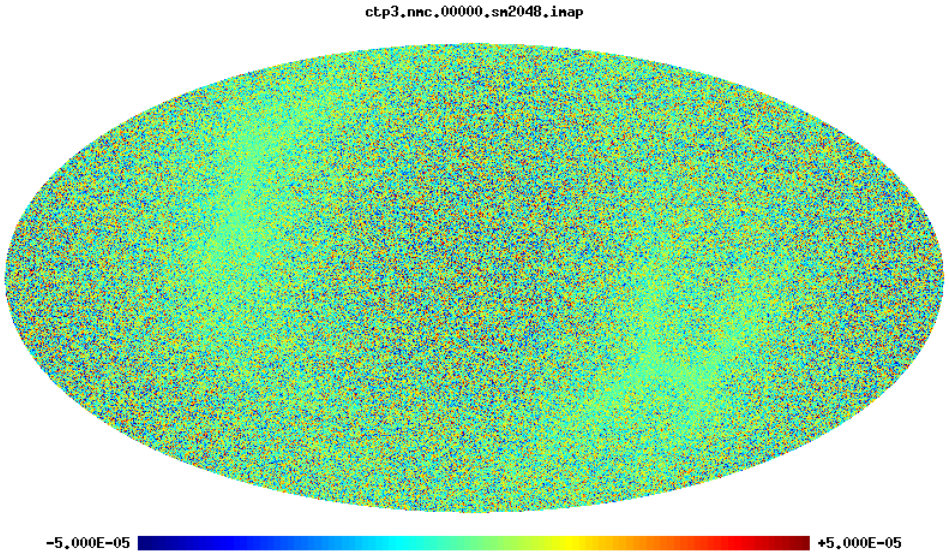


But even then there are bad days …



Results





- 1000 realizations of the Planck mission – CMB & noise
- 250,000 maps & 10M CPU-hours.
- By far the largest CMB Monte Carlo set ever produced.

Planck Results – Cosmology

EMBARGOED
UNTIL MARCH 21ST

Planck Results – Fundamental Physics

EMBARGOED
UNTIL MARCH 21ST

Planck Results – Astrophysics

EMBARGOED
UNTIL MARCH 21ST

Conclusions

- Planck will produce the definitive CMB dataset for the next decade.
  - First results March 21st
  - Full results 2014/15
- Planck results are required/assumed by all Dark Energy experiments.
- These spectacular results would not have been possible without NERSC's exceptional support.
- The Planck/NERSC collaboration can serve as a model for mission-class projects
  - already seeing follow-on from DOE and NASA

Acknowledgements

- NERSC leadership:
  - Horst Simon, Katherine Yelick, Francesca Verdier, Howard Walter
- NERSC staff:
  - Exceptional systems & services teams
- C^3 team
  - Aaron Collier, Reijo Keskitalo, Theodore Kisner
  - Reese Baird, Chris Cantalupo, Rajesh Sudarsan
- And many more …