



CMB-S4

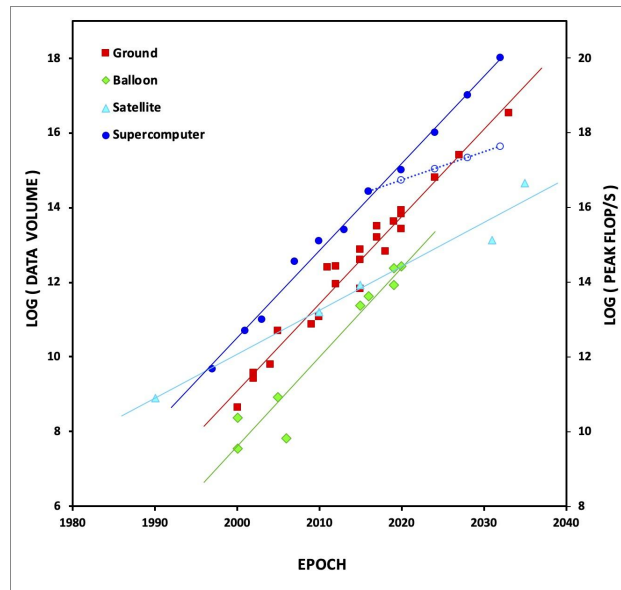
Julian Borrill - Berkeley Lab & UC Berkeley

CMB-S4 Project Data Scientist

NERSC CMB Community Repo PI

Context

- CMB datasets track Moore's Law for ~50 years; computing dominated by number & complexity of mock datasets
- The CMB community has had a shared repo at NERSC for 25 years
 - ~1% of NERSC cycles each year shared by O(200) data analysts & O(10) experiments
 - Long used to retooling our software stack every 2-3 years
- We have a strong collaborative relationship
 - Prototyped many services
 - Spawned only long-term allocation (Planck)
 - NASA award to NERSC management
 - Direct and indirect FTE support



- Also a strong HTC program; CMB-S4 is working to align these.

Major Notes

- Thank you! More notes (in progress) in the google doc.
- Access to sufficient resources (cycles, storage, bandwidth), guaranteed for the project lifetime (construction + operations), is still our most critical priority.
 - How do we efficiently use multiple/disparate HPC/HTC facilities ... superfacility?
- With the end of Moore's Law, not just maintaining but *improving* our code efficiency on the next 20+ years (7+ generations) of architecture is essential.
 - What comes after GPUs and how can we be preparing for it?
- The balance between generic (cost/human-efficient) and specific (domain-efficient) algorithm & implementation solutions is challenging.
- How do we truly realize the necessary applied math + computer science + computational science + domain(s) science collaboration, especially with imperfectly matched hybrid inter- and intra-agency funding
 - Eg. HEP/ASCR - limitations of SciDAC, base vs grant vs project funding, targeted vs cross-cutting research, etc