Snowmass Computing Frontier I2: Distributed Computing and Facility Infrastructures

Ken Bloom
Richard Gerber
July 31, 2013

Who we are

- **Ken Bloom**, Associate Professor, Department of Physics and Astronomy, University of Nebraska-Lincoln
  - Co-PI for the Nebraska CMS Tier-2 computing facility
  - Tier-2 program manager and Deputy Manager of Software and Computing for US CMS
  - Tier-2 co-coordinator for CMS
  - Leader of effort to develop and deploy data federations
- **Richard Gerber**
  - User Services Deputy Group Lead, NERSC (National Energy Research Scientific Computing Center), Berkley Lab
  - NERSC Senior Science Advisor
  - Co-Convener "Large Scale Production Computing and Storage Requirements for Science" series of requirements reviews for DOE.
  - NERSC-7 (Edison) Deputy Project Manager

The Big Picture

- Will adequate computing facilities be available to support HEP science for the next:
  - 5 Years?
  - 10 Years?
  - Beyond?

Our charge

1. How do the various computational problems that are posed by HEP science map onto various types of computing facilities?
2. Given the computational needs of various HEP efforts in both experimental and theoretical work, will computing resources of the required size be available over the appropriate timescales without any new targeted efforts?
3. Will the existing distributed computing models for particle physics, largely based around grid infrastructures and access to distributed data, scale up to meet future needs without any new targeted efforts?
4. What role will national computing centers play in computations for HEP? Will there be a role for clouds?
5. What coordination will be required across facilities and research teams, and are new models of computing required for it?

Our approach

- Leveraged our work off three forays into the community:
  - NERSC was already scheduled to do an assessment of program needs for high-energy physics [ref], which gave us information about the "facility infrastructures"
    - Requested case studies from experimentalists and theorists on all three "frontiers" that described computing needs of upcoming efforts, both in terms of type and scale
    - Report currently in draft form, expected to be released this fall.
  - Took advantage of the Open Science Grid All-Hands Meeting in March [ref] to convene a discussion panel on the future of the grid, which gave us information about "distributed computing"
    - Recruited panelists from different parts of the grid world: operations, technology, security, big thinking
    - Snowmass report will summarize the discussion
  - Listened carefully to Tuesday presentations from CpF E, T groups

I: Computing problems and facilities

- The Worldwide LHC Computing Grid (WLCG) efficiently handles data analysis and detector simulation for CMS and ATLAS, the dominant energy-frontier experiments
  - LHC computational problems are well suited to high-throughput computing paradigm
  - WLCG workflow model is working well for experiments, expected to continue to do so in the future
  - There is work yet to be done on how to use existing (and future) resources efficiently, but improvements can be implemented in an evolutionary manner

I: Computing problems and facilities

- National High Performance Computing (HPC) centers are used and required by a number of projects
  - Lattice QCD (EF)
  - Accelerator design and R&D (EF and IF)
  - Data analysis and synthetic maps (CF)
  - N-body and hydro-cosmology simulation (CF)
  - Supernova modeling (CF)
  - Efforts underway to perform theory computations (e.g. perturbative QCD) directly related to experiment
- Currently looking for more information on how IF experiments might need and use HPC resources
  - NERSC already hosting (and did host) efforts from Daya Bay (Tier 1), KAMLAND, Ice Cube, BaBar, SNO – experience with data

I: HEP Computing

- 2013 DOE & NSF Allocations for HEP
  - DOE Production (NERSC): 168 M Hours
    - LQCD 50 M (113 M included NP allocation)
    - Cosmology 53 M
    - Accel 23 M (32 M including BES & NP)
  - DOE INCITE (ALCF, OLCF): 820 M Hours
    - LQCD 400 M
    - Supernova 230 M
    - Cosmology 80 M
  - NSF XSEDE: 120 M
    - LQCD: 90 M
    - HEP Theory: 12 M

Distributed and HPC Computing in HEP:

CMS + ATLAS in 2012:

~2.6 Billion Hours

National DOE & NSF HPC Centers 2013:

~1.4 Billion Hours

2: Can we meet expected needs?

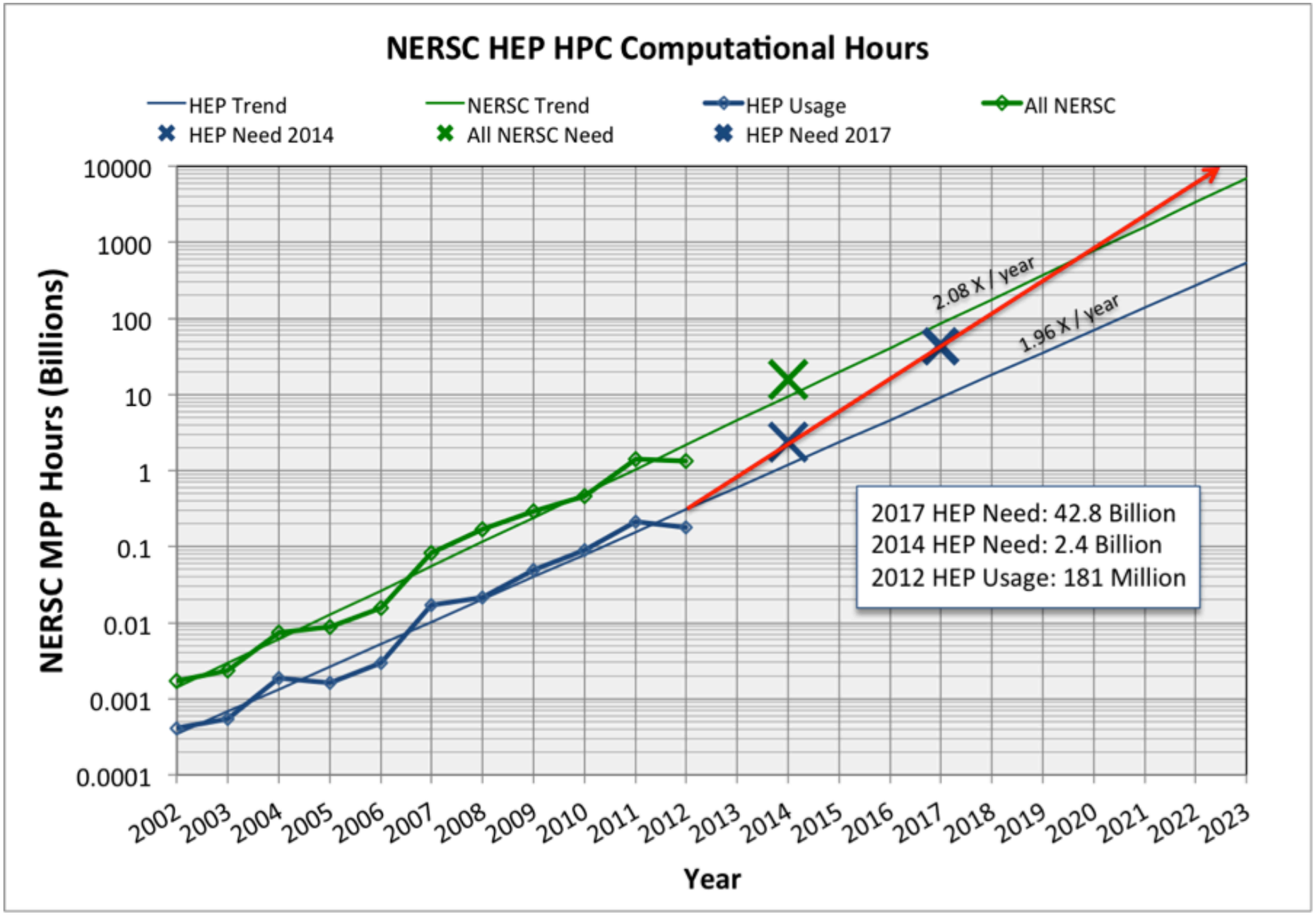
- EF should have sufficient resources for foreseeable (~10 year) future if WLCG funding remains approximately flat-flat
  - Have running experience which helps with predicting needs
  - Operational efficiency gains will probably be needed for this
  - Note that WLCG resources significantly bigger than what is available at HPC facilities, which can provide only so much to HEP (and LHC); need to maintain these resources and infrastructure
- IF processing and computing needs are relatively small compared to EF needs, even in aggregate, and thus it should be straightforward to meet them
  - But could be organizational challenges in finding efficiencies of scale.

2: HEP HPC Needs

- Demand by traditional HEP HPC community will outstrip expected availability by 2017 at NERSC by a factor of 4

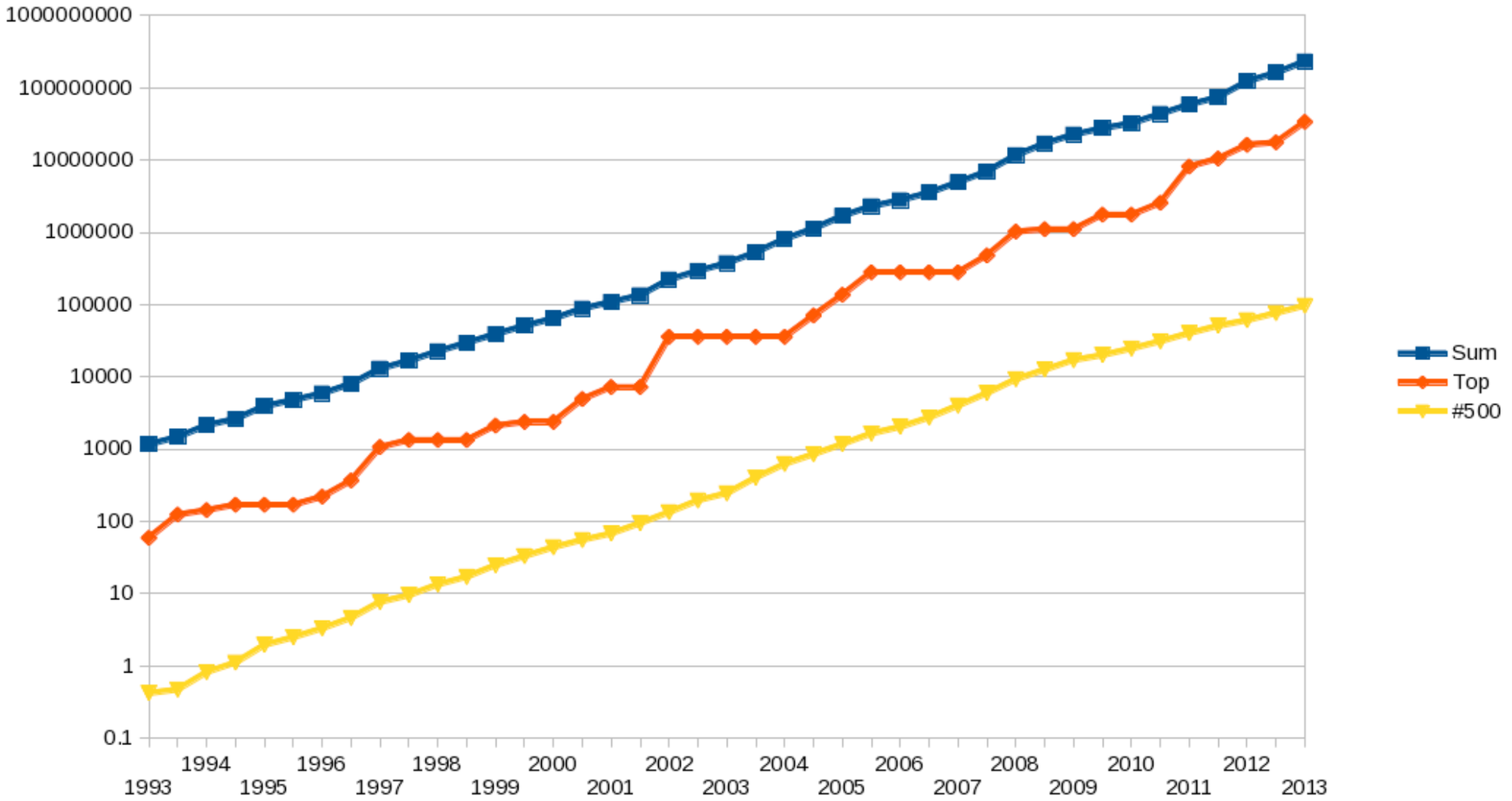
Even this is optimistic wrt funding

Driven by LQCD, accelerator, astrophysics



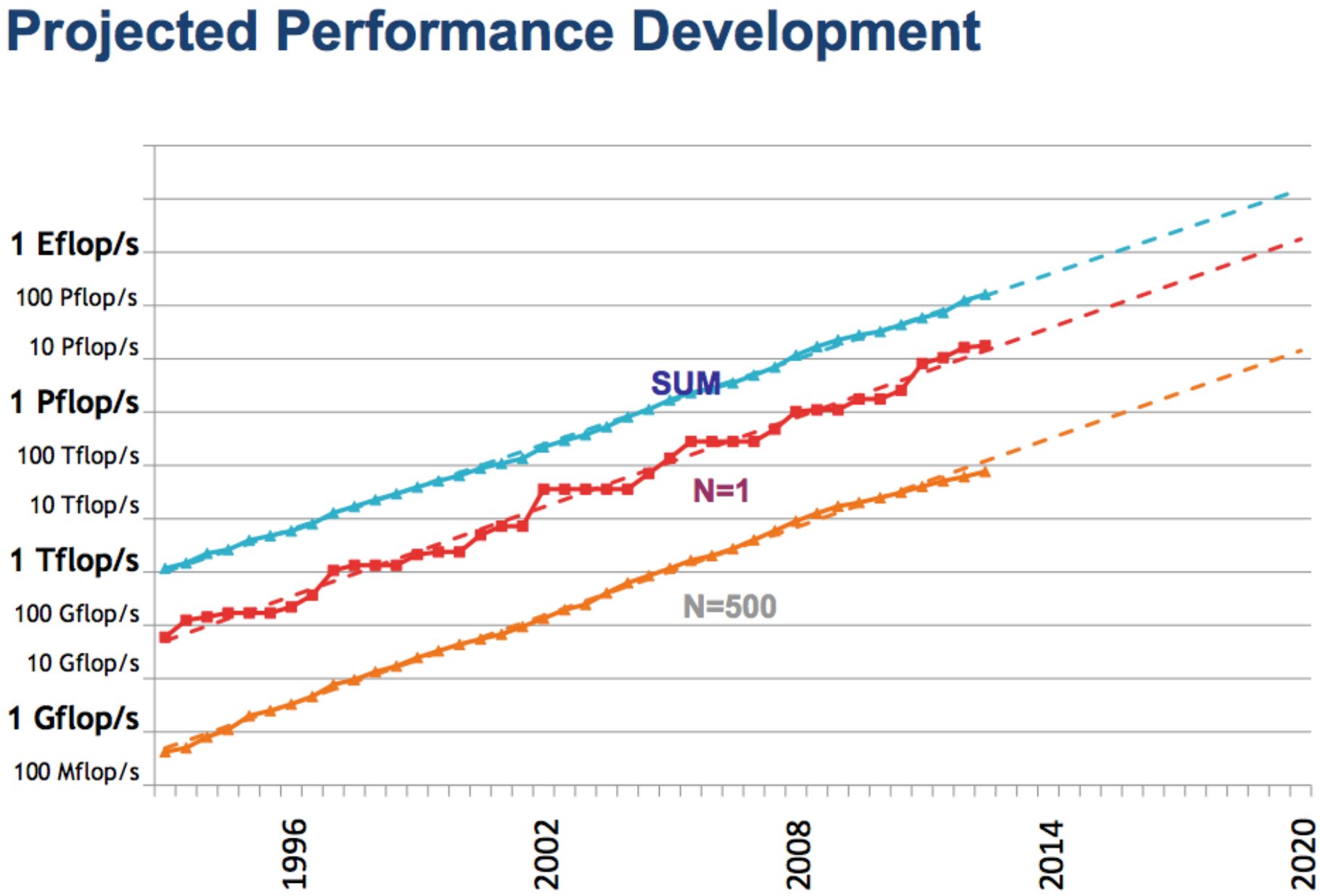
Traditional HPC

- The HPC facilities in DOE (NERSC & LCFs) and NSF hope to stay on the Top 500 Moore's Law slope, but it depends on
  - Funding
  - Technology improvements in processors and systems: DOE Fast Forward and Design Forward efforts with vendors



Revolution in Energy Efficiency Needed

Even though energy efficiency is increasing, today's top supercomputer (N=1) uses ~9 MW or roughly $9M/year to operate. Even if we could build a working exaflop computer today, it would use about 450 MW and cost $450M/year to pay for power.



3: Distributed computing models

- Grid infrastructures well-suited to work done by large collaborations of experimental particle physicists
- Experiments, especially EF, are making good use of the grid, which has been a key technology for physics discovery
- No show-stoppers seen for long-term scaling of high-throughput grid computing, but various developments should be pursued to improve efficiency and ease of use
  - Simplification/scaling of job submission, identity management, streamlined operations, storage management and federated data access, dynamic scheduling, readiness for cloud infrastructures
- HEP is the largest user of the grid and must take a leadership role in its continuing development

4: Role of national facilities

- National HPC centers already play a significant role in some areas.
  - LQCD
  - Cosmology & large scale structure
  - Accelerator research and design
  - Supernova physics
  - Data-driven science in the Intensity and Cosmic Frontiers
- WLCG-based tasks at HPC centers?
  - At the moment it is hard to see that the centers can provide a large fraction of needed resources, but this should be explored as part of a program of diversifying the LHC computing architectures

4: National facilities for "non-traditional" HPC

- Advantages of HPC centers:
  - Access to large resources, state-of-the-art architectures, strong operations and support, consulting and training, centralized software/data repositories, good growth rate, good for data-intensive projects needing large storage, I/O, world-best networking.
- HPC centers have challenges too:
  - Integration with WLCG workflows, job scheduling, designed for parallel rather than serial, virtualization for validated environments, formalities for allocation fo resources, transition to multicore.
  - Funding needed to support additional computing.
- Successes at NERSC already with PDSF (KAMLAND, Ice Cube, BaBar, SNO, ALICE), Daya Bay, PTF
  - NERSC data strategy, web portals, e.g. https://portal-auth.nersc.gov/dayabay/odm

4: Role of cloud computing

- [Commercial] Cloud facilities not currently suited for HEP, mostly due to cost issues at the moment.
- But concerns about data might be remedied by the "content provider" model, not much needs to be stored in the cloud.
- Cloud computing provides many advantages, including customized environments that enable users to bring their own software stack.
- Clouds have the ability to quickly surge resources to address larger problems.
- Significant gaps and challenges exist in managing virtual environments, workflows, data, cyber-security, and other areas.
- There are efforts by traditional HPC platforms to combine the flexibility of cloud models with the performance of HPC systems.

5: Coordination and new models

- Have found a limited amount to say about this
- However, yesterday's IF presentation suggested that those experiments could use some larger organization that can advocate for it with respect to computing resources
  - Given that IF is a DOE funding area, is there a way to fund the IF Computing Consortium?
- Can we find a way to make more seamless transitions between working in HPC and HTC environments?
  - Technical: Identity management, project management,...?
  - "Political": within/across DOE (Production, INCITE) and NSF (XSEDE) programs?

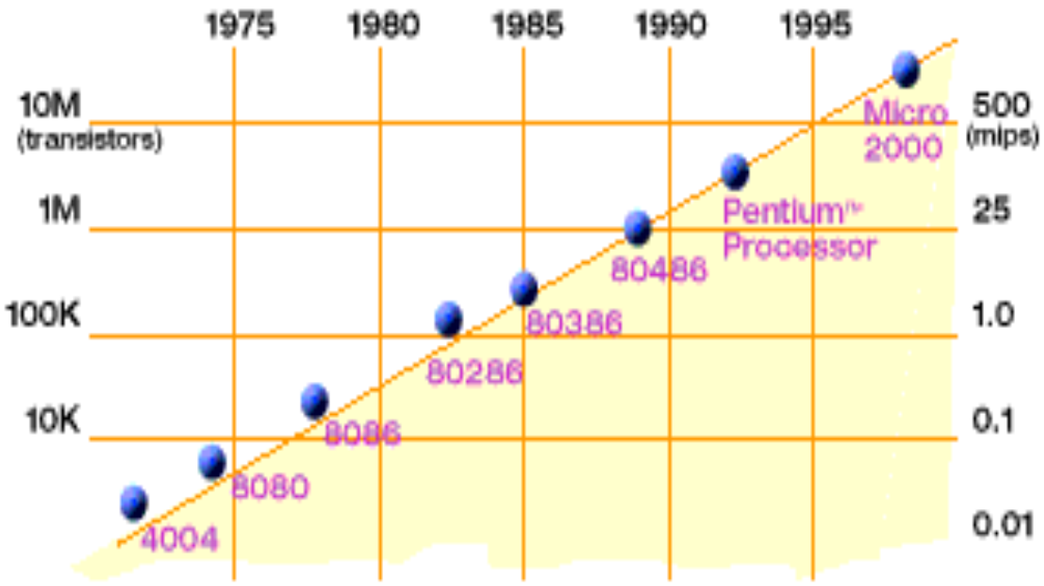
Outlook

- Yesterday's session was very useful for understanding the needs of the E and T groups
- Within the next ten years, things seem OK on the HTC front for e.g. EF and IF experiments
- A greater challenge with the demand for HPC resources
- Any conclusions depend on assumptions about funding, technology
- Have a good outline of the report and plenty of text, further to go to finish things off

Additional Materials

Computing Models

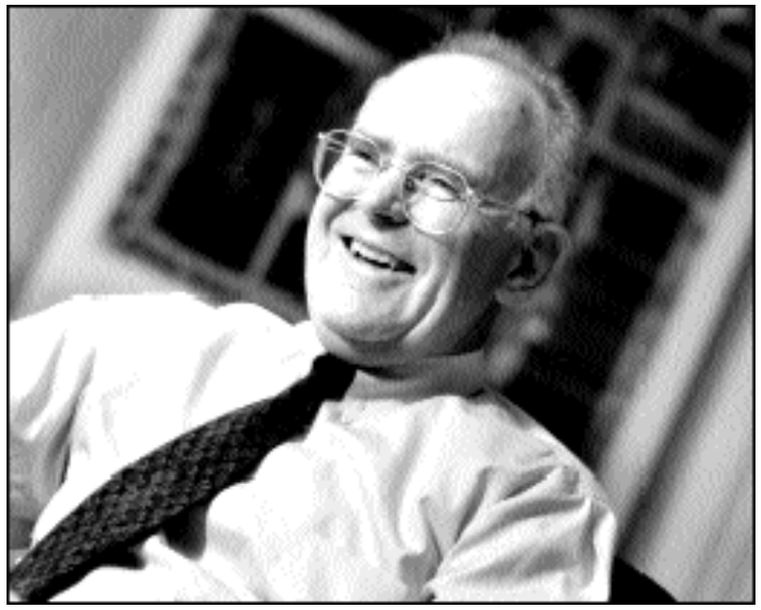
Why You Need Parallel Computing: The End of Moore's Law?



2X transistors/Chip Every 1.5 years

Called "Moore's Law"

Microprocessors have become smaller, denser, and more powerful.

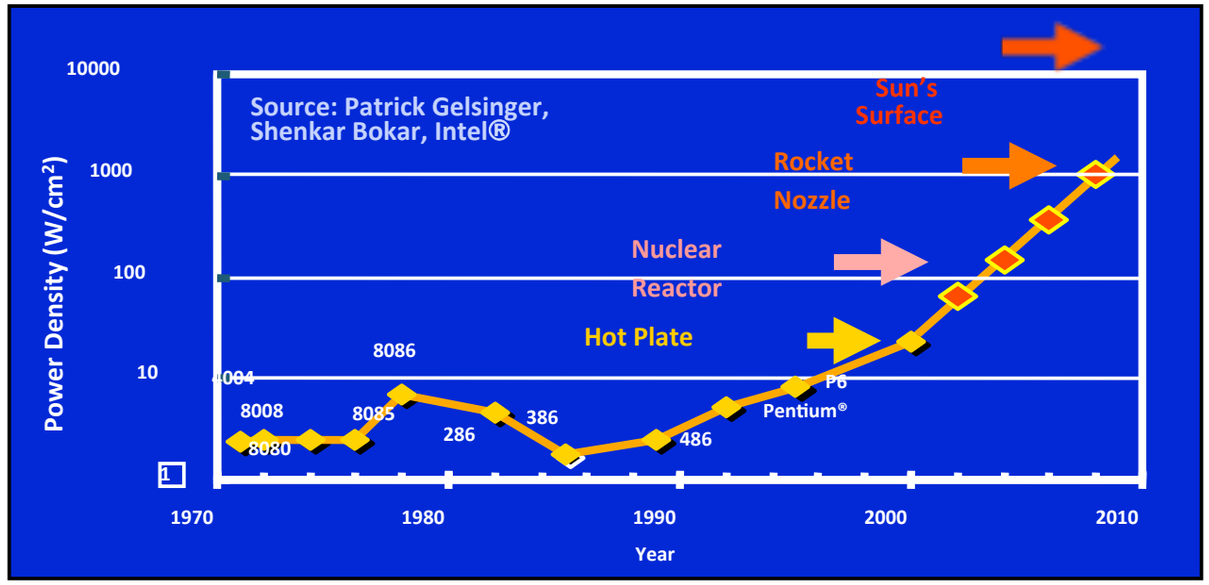


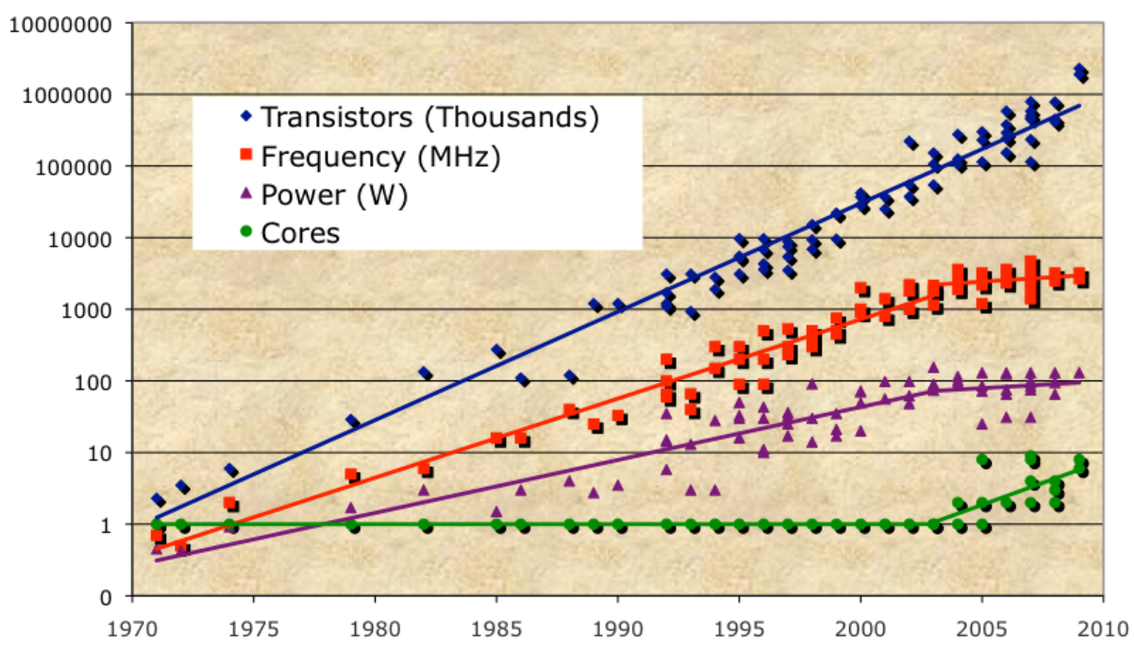
Gordon Moore (co-founder of Intel) predicted in 1965 that the transistor density of semiconductor chips would double roughly every 18 months.

Slide source: Jack Dongarra

Power: Biggest Architectural Challenge

If we just kept making computer chips faster and more dense, they'd melt and we couldn't afford or deliver the power.





Now compute cores are getting slower and simpler, but we're getting lots more to maintain the performance curve.

GPUs and Intel Phi have hundreds of "light-weight cores"

Programming Challenge

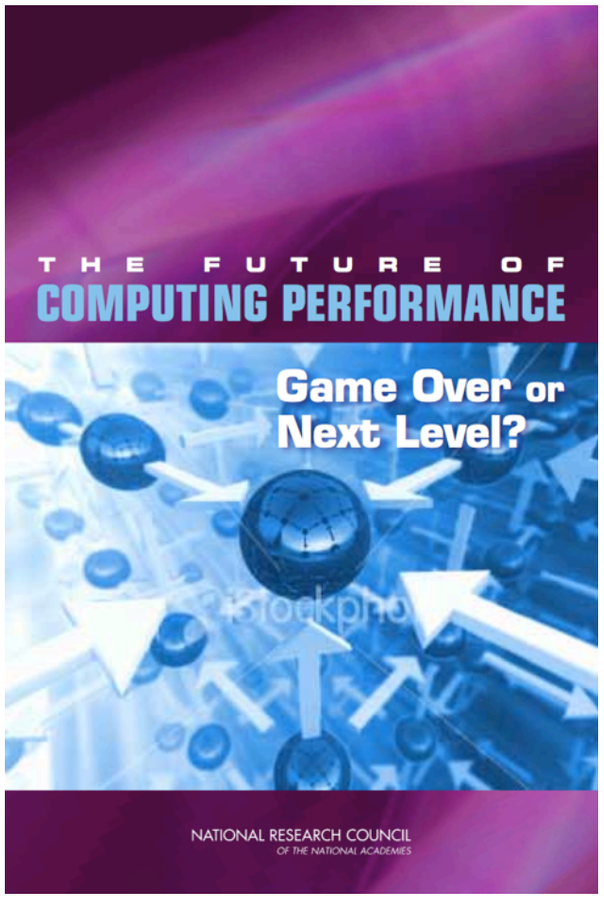
- **To effectively use many-core processors, programs must exploit 100K – 1M way parallelism.**
- **Traditional programming paradigms won't work**
  - Too resource intensive per MPI task
  - Data movement is extremely expensive
- **Current programming methods for accelerators (GPUs) are difficult**
  - Need one "fat core" (at least) for running the OS
  - Data movement from main memory to GPU memory kills performance
  - Programmability is very poor
  - Most codes will require extensive overhauls

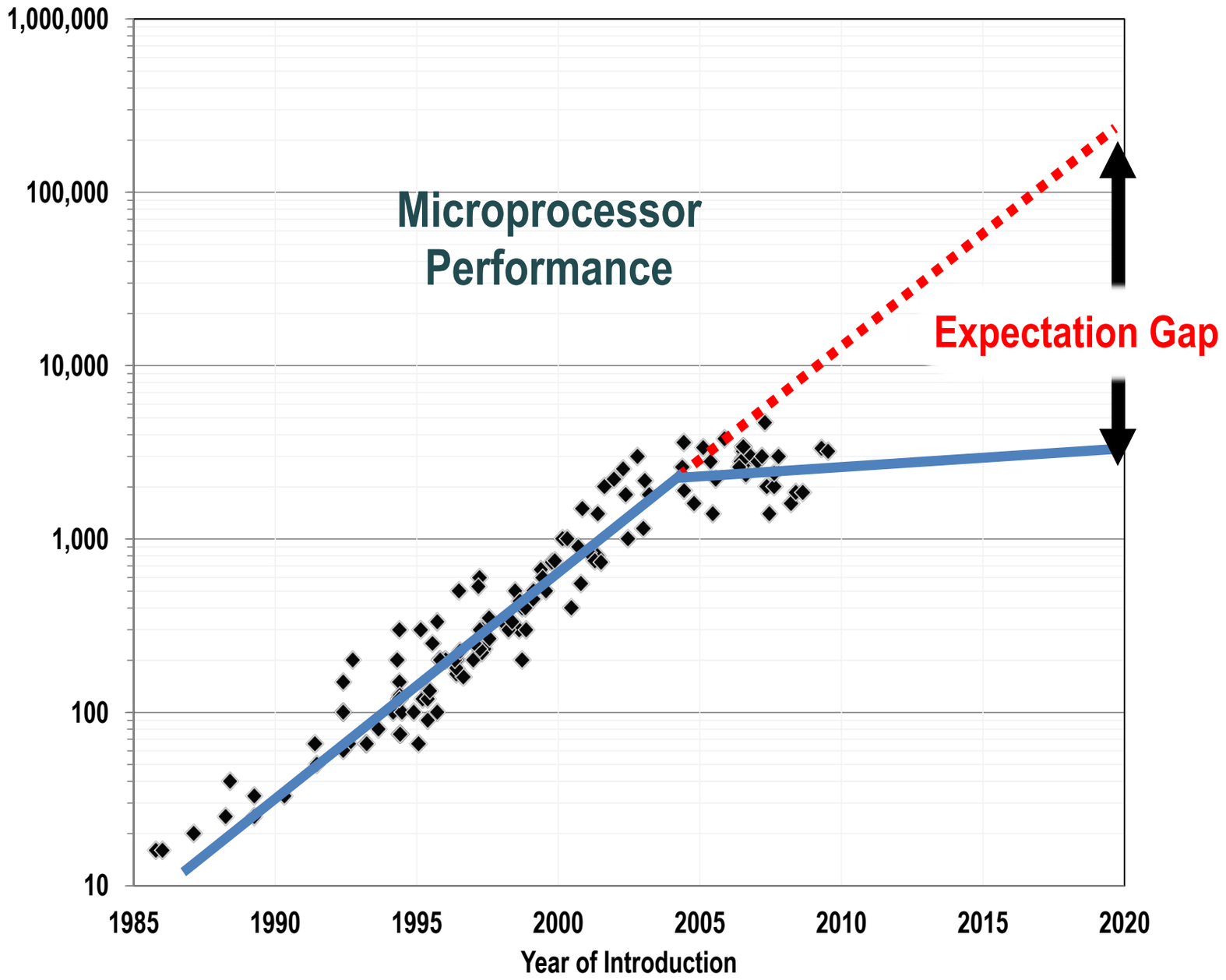
Future Performance

Moore's Law Reinterpreted

- Number of cores per chip will increase
- Clock speed will not increase (~~possibly~~ probably decrease)
- Need to deal with systems with millions of concurrent threads
- Need to deal with inter-chip parallelism (OpenMP threads) as well as intra-chip parallelism (MPI)
- Any performance gains are going to be the result of increased parallelism, not faster processors

Serial Processing = Left Behind





Are GPUs the Answer?

- **GPUs show promise for some applications**
  - Many small, energy-efficient cores (GPUs)
  - Accelerators are theoretically very fast
  - Much better theoretical Flop/Watt
- **Challenges are considerable**
  - GPU have private memory space
  - Attached to motherboard via PCI interface currently
  - Need one fat core (at least) for running the OS
  - Data movement from main memory to GPU memory kills performance
  - Programmability is very poor
  - Most codes will require extensive overhauls

My Perspective

- "Many core" is here to stay
- You will have to find fine-grained parallelism in your code or you will be left behind
- OpenMP or a similar threading model (OpenACC?) is the most likely viable long-term (5-10 years) programming model
- GPU accelerators have a lot of momentum in the short term and can be useful for certain applications
- Simulation and data analysis will become even more intertwined and will need to share close data spaces