Challenges in High Performance Computing (HPC) for Climate Prediction and Projection

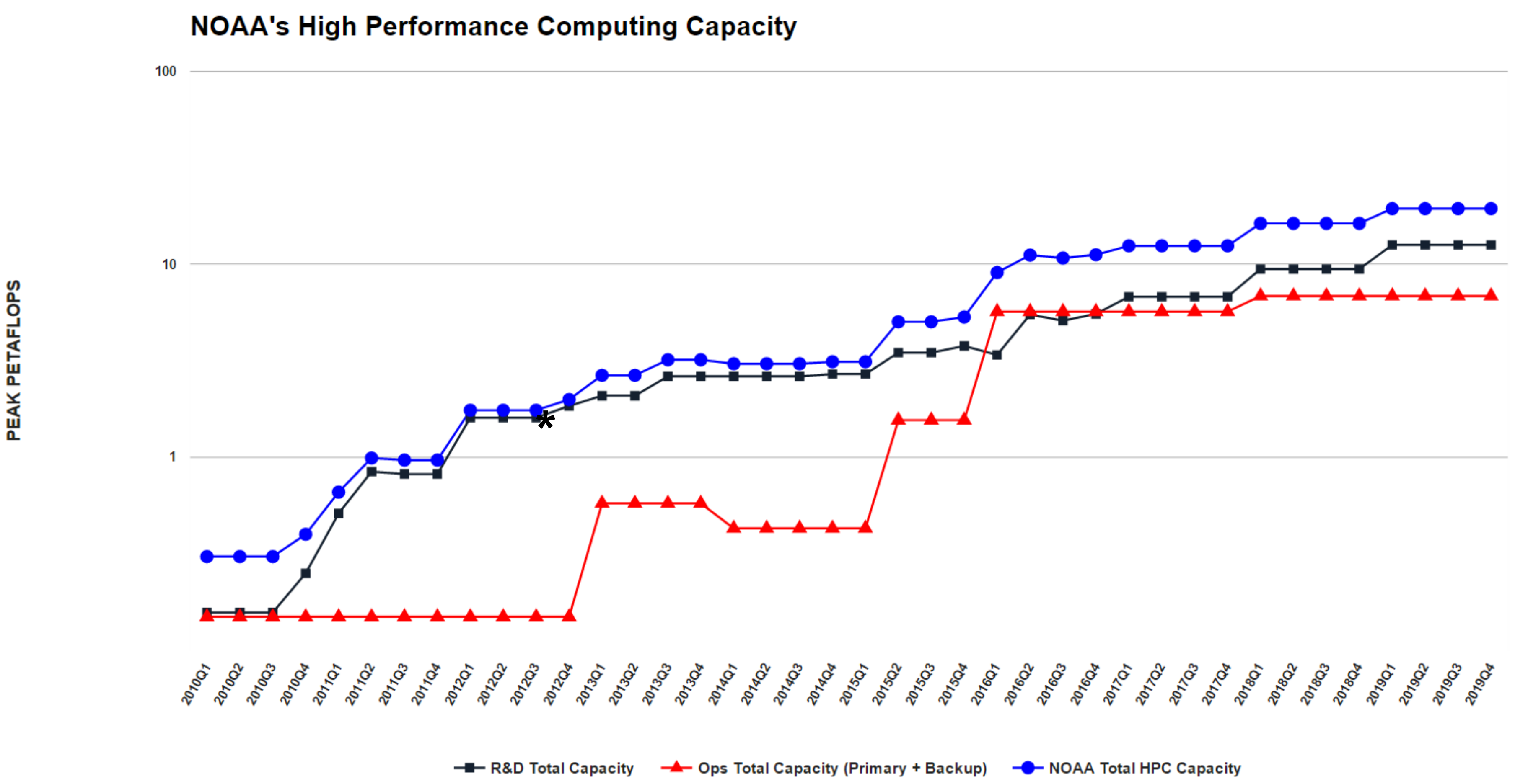
Frank Indiviglio
Production Lead, High Performance Computing
June 27, 2016

Acknowledgements: Bill Collins (DOE), Bill Putman (NASA), John Michalakes (DOD), V.Balaji (NOAA)

Driving Questions

- What horizontal and vertical resolutions are necessary to adequately resolve processes in the coupled system that drive both prediction error in short term forecasts and climate simulation bias?
- What is the computational cost of the key biogeochemical/physical processes must be included in models to address mission requirements?
- What is the ideal size of the ensemble needed for this effort both for prediction, for understanding coupled processes and biases, and quantifying uncertainty?
- What modeling improvements will most significantly impact computing and storage requirements (e.g., resolution, processes/complexity, ensemble members, etc) and system balance (between compute, networking, storage, etc)?

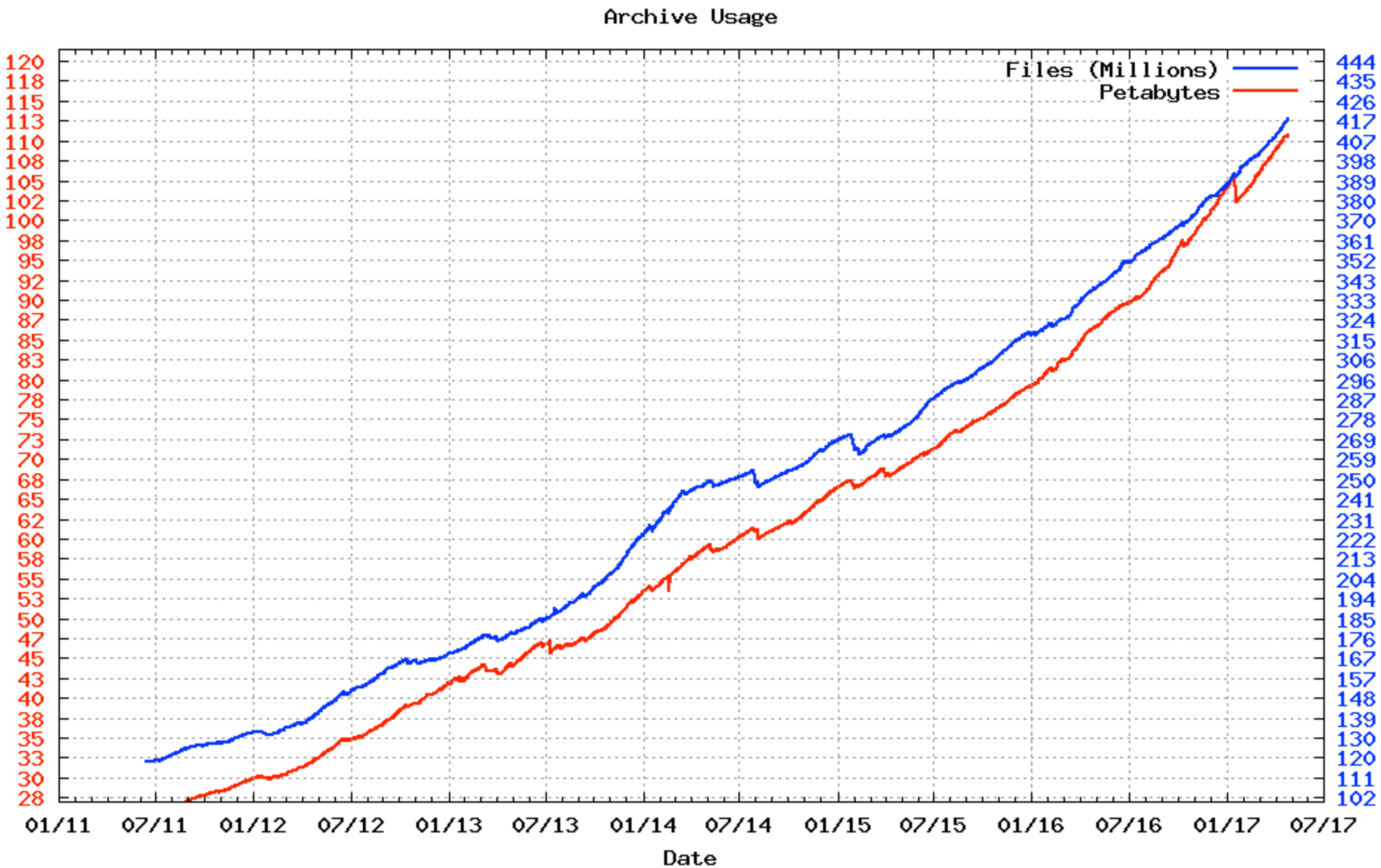
NOAA HPC Growth



Data Challenges

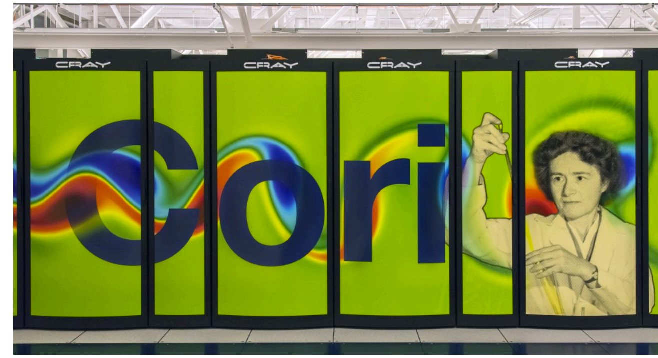
- How will increasingly high-resolution data be stored and shared for community research?
- As resolution increases, it becomes more difficult to save every bit to disk. How can we reduce the storage burden from coupled hi-res integrations?
- What must be analyzed at full resolution, and what can be evaluated at coarser spatial resolution?
- What aspects of your analysis can be in-lined during computation to reduce the required storage?
- How does increased horizontal resolution impact the necessary temporal resolution of your analysis and data storage?
- What new technologies, such as non-volatile random-access memory (NVRAM), provide the greatest potential to improve the scalability and efficiency of your coupled systems and particularly IO bottlenecks that are inevitable at high resolution?

NOAA Data Archive Growth



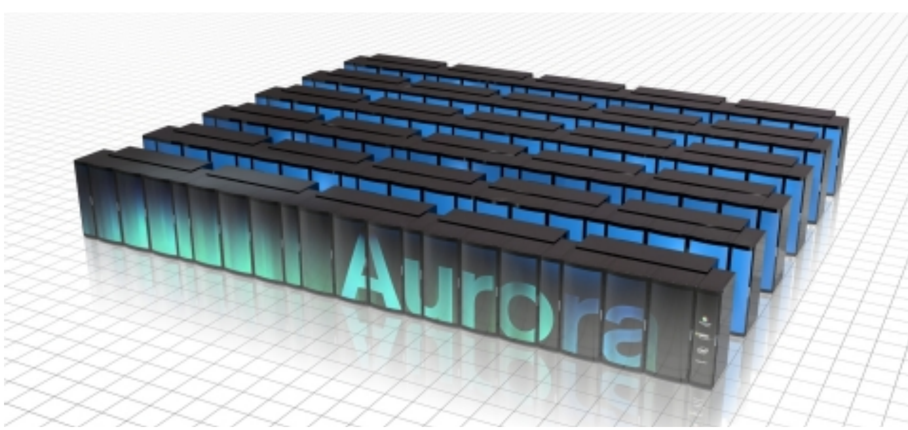
Today's DOE Leadership Systems

- NERSC Cori (Phase II)
  - >31.4 Pflops
  - 29+ Pflops Xeon Phi
  - 32 Core Hazwell and 68 Core Xeon Phi
- OLCF Titan
  - 27 Pflops
  - 16 core AMD Opteron + NVIDIA GPU
- ALCF Mira
  - 10 PFlops
  - 16 core PowerPC

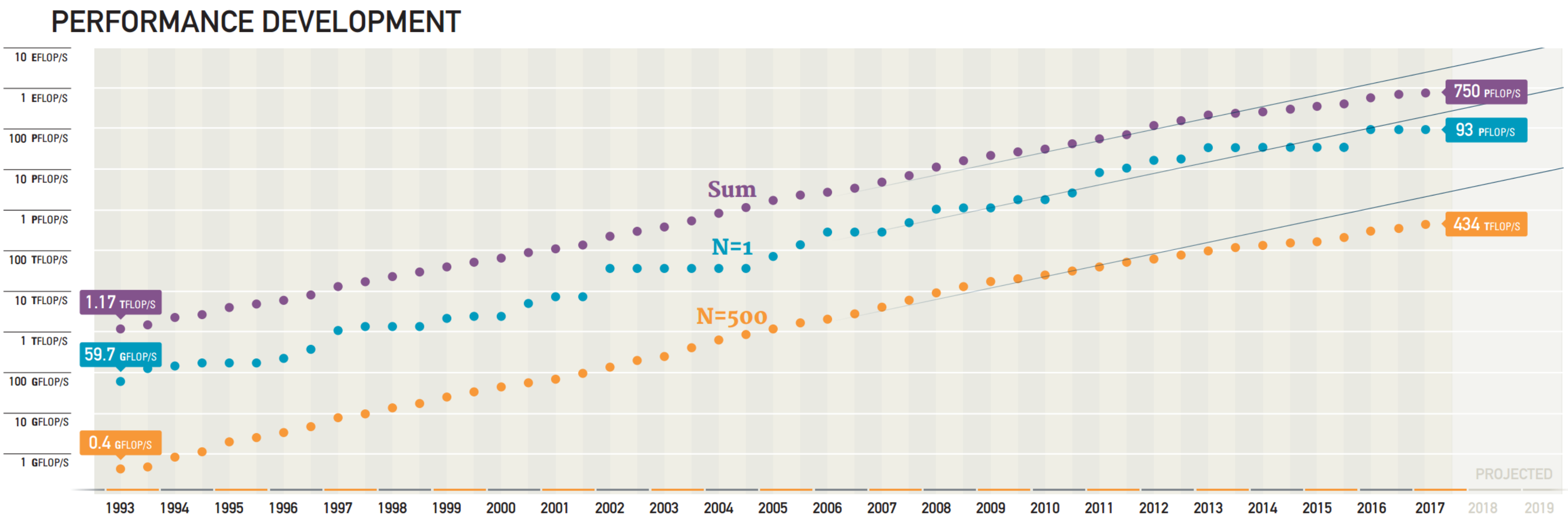


DOE Next Generation Machines

- Key goal of ACME: Run on next generation DOE machines:
- OLCF Summit 2018
  - ~200 PFlops
  - Multiple IBM power9 and NVIDIA GPUs
- ALCF Aurora 2018
  - ~180 Pflops
  - 50K Nodes, 3rd gen Intel Phi



A Look at The HPC Industry



- The top 500 systems performance continues to flatten
- Accelerated platforms now occupy almost 20% of the list
- The benchmarks in which the industry is using to evaluate performance are changing, HPCG is now being incorporated into the evaluation

Graphic – HPCwire Top 500 Results June 19, 2017 - https://www.hpcwire.com/2017/06/19/49th-top500-list-announced-isc/

JPSY Comparison Across ESMs

Model	Machine	Resolution	SYPD	CHSY	JPSY
CM4	Gaea/c2	1.2×10^8	4.5	16000	8.92×10^8
CM4	Gaea/c3	1.2×10^8	10	7000	3.40×10^8

- Comparative measures of capability (SYPD), capacity (CHSY), and energy cost (JPSY) per "unit of science".
- Can you have codes that are "slower but greener"? Algorithms that are "less accurate but more eco-friendly"?
- From Balaji et al (2016), in review at GMDD.

http://goo.gl/Nj1c2N

On software development ….

- Experience to date with fine-grained architectures: kernels can sing (~40X), but complex multi-physics codes croak (~<2X)
- Approach: code revisions for performance on conventional architectures will get us a significant way toward performance on fine-grained systems.
  - Component Concurrency
  - Offload I/O, Diagnostics
  - Performance analysis tools
  - vectorization (requires interaction with compiler vendors)
  - wide halos (to reduce comms)
  - nonmalleable executables (aka static memory)
  - direct use of coarray

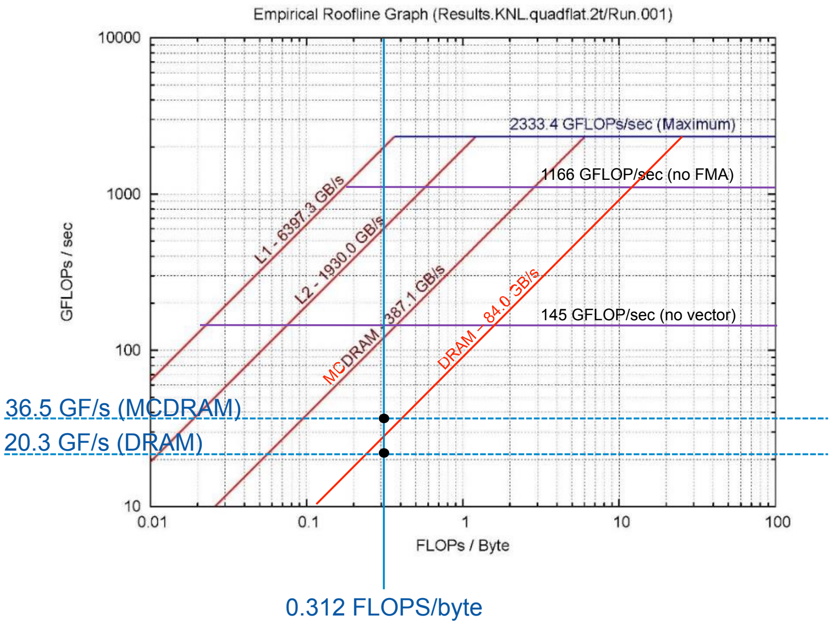
ACME Mini-App Strategy

- Recommendation of ACME/Exascale study group: *Identify key kernels/modules that are small enough so a single person can understand/refactor/rewrite to test new approaches, but that are large enough that successful results are meaningful for ACME.*
- Target: Transport mini-apps for both atmosphere and ocean to cover finite element and finite volume approaches used in ACME
- Atmosphere tracer transport is single most expensive ACME component. Ocean tracer transport is 30% of the ocean model

Neptune Example

- Roofline Model of Processor Performance
  - Bounds application performance as a function of **computational intensity**
  - If intensity is high enough, application is "compute bound" by floating point capability
  - If intensity is not enough to satisfy demand for data by the processor's floating point units, the application is "memory bound"
    - 128 GB main memory (DRAM)
    - 16 GB High Bandwidth memory (MCDRAM)
  - KNL is nominally 3 TFLOP/sec but to saturate full floating point capability, need:
    - 0.35 flops per byte from L1 cache
    - 1 flop per byte from L2 cache
    - 6 flops per byte from high bandwidth memory
    - 25 flops per byte from main memory
  - Hard to come by in real applications!
    - NEPTUNE benefits from MCDRAM (breaks through the DRAM ceiling) but realizing only a fraction of the MCDRAM ceiling

KNL



NEPTUNE E14P3L40
(U.S. Navy Global Model Prototype)

- John Michalakes – 17th Workshop on High Performance Computing in Meteorology

Other Questions to Consider

- What do workflow and machine policy add to the cost of science?
- What is the scaling "data intensity" of data with compute and how does that change with model resolution?
- How will future architectures effect how we enact these workflows?

Summary/Discussion

- Driving questions center on what HPC capability and configuration will be need needed to address science priorities.
- Climate agency HPC is growing, but it's unlikely that it's growing fast enough.
- The interaction between simulation, data analytics, and storage needs to be constantly assessed.
- Software innovations to leverage anticipated HPC architectures are hard to implement. And necessary.
- Partnerships on hardware and software have accelerated us toward our goals.

Questions?

NOAA's Science Network