JAN 2025

POST-EXASCALE COMPUTING IN US AND THE DOE PERSPECTIVE

J. TAYLOR CHILDERS
Deputy Technical Director for ALCF-4
Argonne Leadership Computing Facility
Argonne National Laboratory

REACHING EXASCALE

- DOE operates 25 machines on the Top500, three of which offer more than an ExaFLOP of compute
- More than 110,000 GPUs for scientific research at NERSC, OLCF, and ALCF.
  - 10x more cores than previous gen DOE HPCs
- ECP set out in 2016 to reach Exascale by 2021.
- Reaching exascale was not a simple turn of the crank:
  - Shear number of hardware components leads to stability challenges.
  - Traditional filesystems face scaling challenges.
  - Balancing networking is a challenge
  - Establishing uniform software support to form a unified development environment is lagging.
  - This is consistent with META's RAS publication

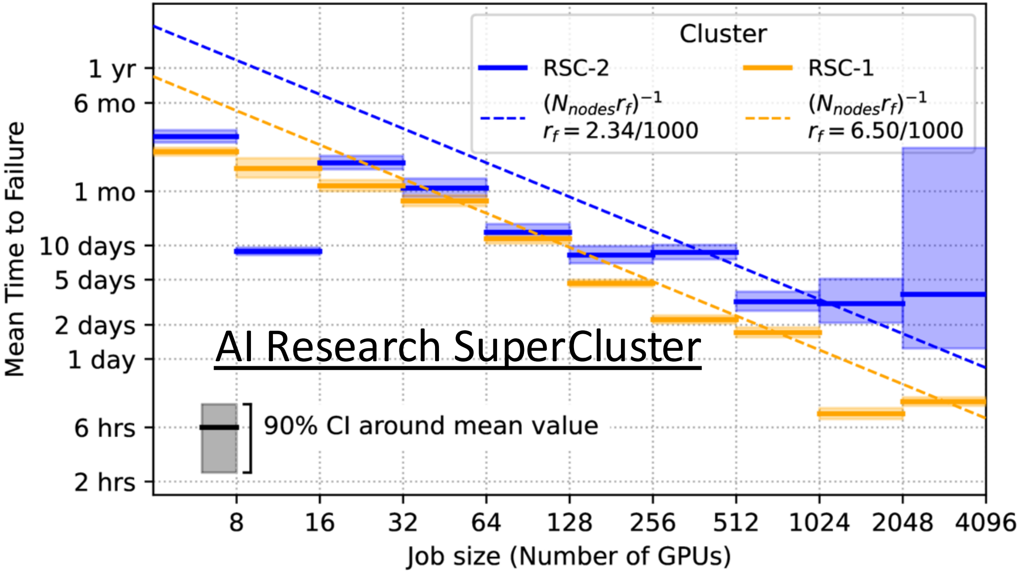
REACHING EXASCALE

- DOE operates 25 machines on the which offer more than an ExaFLOP/s
- More than 110,000 GPUs for scie NERSC, OLCF, and ALCF.
  — 10x more cores than previous ge
- ECP set out in 2016 to reach Exasca
- Reaching exascale was not a simple
  — Shear number of hardware com stability challenges.
  — Traditional filesystems face scalin
  — Balancing networking is a challen
  — Establishing uniform software s unified development environment
  — This is consistent with [META's RA](#)

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**El Capitan** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,039,616	1,742.00	2,746.38	29,581
2	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607
3	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
19	**Perlmutter** - HPE Cray EX 235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-11, HPE DOE/SC/LBNL/NERSC United States	888,832	79.23	113.00	2,945

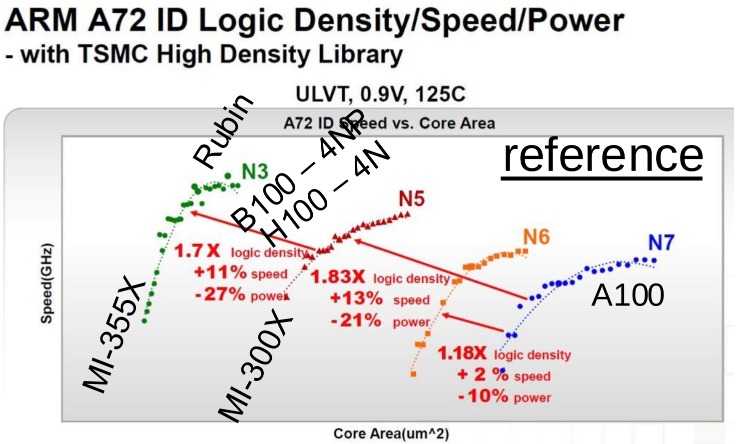
REACHING EXASCALE

- DOE operates 25 machines on the Top500, three of which offer more than an ExaFLOP/s
- More than 110,000 GPUs for scie NERSC, OLCF, and ALCF.
  - 10x more cores than previous ger
- ECP set out in 2016 to reach Exasca
- Reaching exascale was not a simple
  - Shear number of hardware com stability challenges.
  - Traditional filesystems face scalin
  - Balancing networking is a challen
  - Establishing uniform software s unified development environment
  - This is consistent with META's RAS publication

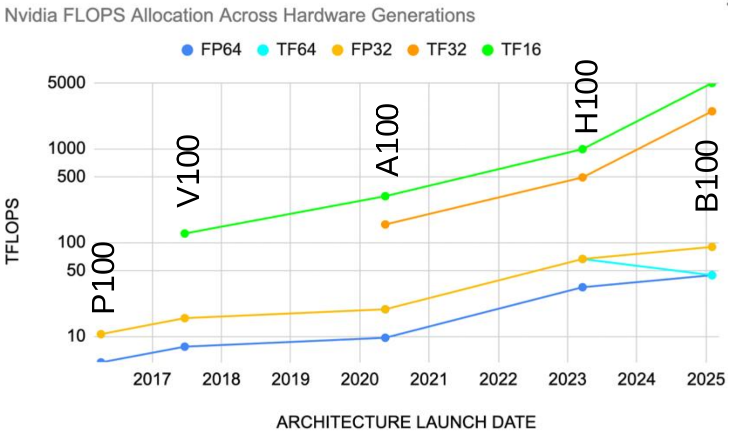


ASYMPTOTIC REGION

- GPUs remain the best compute per cost with broad software support
- Performance gains on consecutive generations of TSMC printing are 25-65%
  - Asymptotically approaching limits of lithography techniques and quantum realm
  - Additional speed up factors come from growing chip area and customization of transistor allocation to match application goals
  - Consecutive Gens taking more time to roll out
- Overall FP64 FLOPS allocation is not keeping up with scientific ambitions.
  - FP64 FLOPS increased 9x in 9 years
  - TF32 FLOPS increased 16x in 5 years
  - TF16 FLOPS increased 40x in 7.5 years

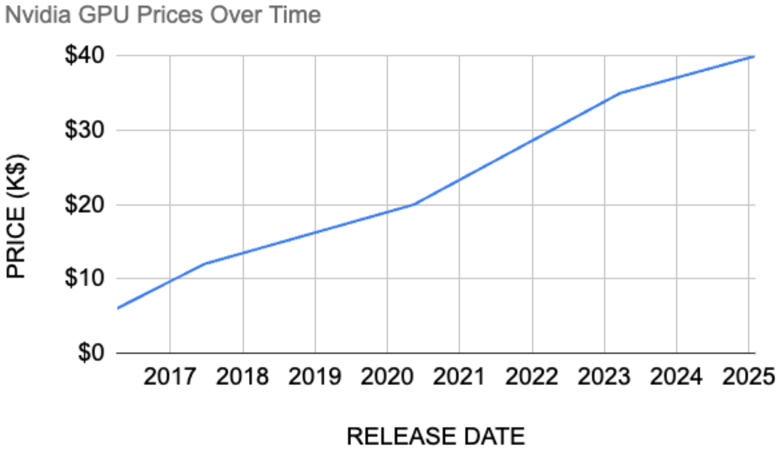


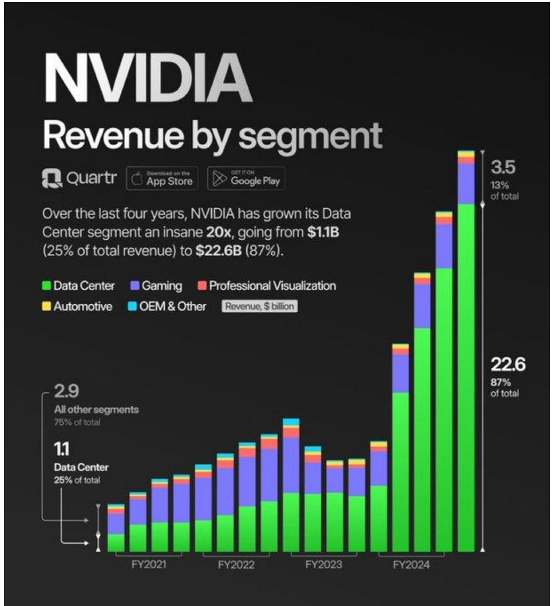
TSMC Process Node



THE COST OF AI

- AI has dramatically changed the GPU industry since the contracts for exascale machines were signed.
- Most notably, the cost of a GPU has increased by 6x in 9 years
- I probed google for the release prices of Nvidia GPUs over time (same chips as previous plot)
- The price has been increasing linearly
- This will limit future machine size unless DOE HPC budgets changes or market competition reduces prices.
  - — Some hope for competitive price reductions as Europeans, Chinese, Japanese, and others begin ”sovereign” chip production.





ROUGH EXPECTATIONS FOR THE FUTURE

- What would it take to get go 10x bigger?
- Current GPUs have about 50 TFLOPS FP64
- Let's assume by 2030 TMSC is at N2 process node yielding an approximate 2x increase in transistor density, and we'll assume dies have grown bigger by 1.5x, so a 3x increase in chip performance.
- Resulting in roughly 150 TFLOPS FP64 per GPU.
- We would need 67k GPUs (Aurora has 64k)
- At B200 Power of 1kW, that's a 67MW machine for just GPUs, assuming GPU power utilization doesn't increase between now and 2030.
- GPUs are costing about $35k, assuming a "good-will science" discount of 50%, such a system would cost $1.17B for GPUs alone.
- To compare to CPU-only: 10 EFLOP system would need 1.6M CPUs at $3.7B

GPT prompt: A supercomputer designed by Apple.

POST-TOP500 ERA

- DOE recognizes that 10 FP64 EFLOPS is an unrealistic goal.
- Application performance, or Science Throughput, on HPC architectures will be driving system metrics in the future.
  - There can still be distractions here, e.g. LLM training time, but more helpful than LINPACK.
  - But looking to drive development of mixed-precision and AI surrogates in domains for performance optimization
- Next Generation of DOE HPCs are in their procurement process with operations starting in the 2028 to 2030 window:
  - NERSC-10 (call for proposals closed)
  - OLCF-6 (call for proposals closed)
  - ALCF-4 (call for proposals end of 2026)
- There will be lots of compute available in these machines for HL-LHC.
- Systems plan for dedicated edge service nodes as well.

OpenAI's Sora:
Generate Claymation scientists making lots of discoveries

FASST: A DOE AI INITIATIVE

- Frontiers in AI for Science, Security and Technology
- Seen as the follow on to Exascale Computing Project ($1.8B over 5 years)
- Goals include:
  - deploying large scale (GigaWatt) data-centers for large scale AI advancement in science
  - curating domain science training datasets
  - facilitating scientists to apply AI in new ways in their domains
  - partnering with industry for scientific advancement
- https://www.energy.gov/fasst
- DOE is going to be a big player in driving AI in scientific workflows, via funding and computing systems.


Argonne
NATIONAL LABORATORY
ALCF