High Performance Computing: Where We Are Today And A Look Into The Future

Jack Dongarra
University of Tennessee
Oak Ridge National Laboratory
University of Manchester
Overview

• How I got started in high performance computing
• Supercomputing today
• Some of the trends in the field
• A look at the future
An Accidental Benchmarker

LINPACK was an NSF Project w/ ANL, UNM, UM, & UCSD
We worked independently and came to Argonne in the summers

Appendix B of the Linpack Users’ Guide
Designed to help users estimate the run time for solving systems of equation using the Linpack software.
First benchmark report from 1977; Cray 1 to DEC PDP-10

Top 23 List from 1977
Performance of solving Ax=b using LINPACK software

<table>
<thead>
<tr>
<th>Facility</th>
<th>N=100 micro-</th>
<th>Computer</th>
<th>Type</th>
<th>Compiler</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAR</td>
<td>14.6</td>
<td>0.14</td>
<td>CRAY-1</td>
<td>S  CFT, Assembly BLAS</td>
</tr>
<tr>
<td>LASL</td>
<td>20.6</td>
<td>0.43</td>
<td>CDC 7600</td>
<td>S  FTN, Assembly BLAS</td>
</tr>
<tr>
<td>NCAR</td>
<td>15.1</td>
<td>0.26</td>
<td>CRAY-1</td>
<td>S  CFT</td>
</tr>
<tr>
<td>LASL</td>
<td>25.7</td>
<td>0.61</td>
<td>CDC 7600</td>
<td>S  FTN</td>
</tr>
<tr>
<td>Argonne</td>
<td>2.31</td>
<td>0.86</td>
<td>IBM 370/195</td>
<td>D  H</td>
</tr>
<tr>
<td>NGAR</td>
<td>161.3</td>
<td>1.05</td>
<td>CDC 7600</td>
<td>S  Local</td>
</tr>
<tr>
<td>Argonne</td>
<td>2177.3</td>
<td>1.33</td>
<td>IBM 3033</td>
<td>D  H</td>
</tr>
<tr>
<td>NASA Langley</td>
<td>100.6</td>
<td>1.42</td>
<td>CDC Cyber 175</td>
<td>S  FTN</td>
</tr>
<tr>
<td>U. Ill. Urbana</td>
<td>196.5</td>
<td>1.47</td>
<td>CDC Cyber 175</td>
<td>S  Ext. 4.6</td>
</tr>
<tr>
<td>LLL</td>
<td>189.5</td>
<td>1.61</td>
<td>CDC 7600</td>
<td>S  CHAT, No optimize</td>
</tr>
<tr>
<td>SIAC</td>
<td>119.4</td>
<td>1.69</td>
<td>IBM 370/168</td>
<td>D  H Ext., Fast mult.</td>
</tr>
<tr>
<td>Michigan</td>
<td>109.6</td>
<td>1.84</td>
<td>AMDahl 470/V6</td>
<td>D  H</td>
</tr>
<tr>
<td>Toronto</td>
<td>77.4</td>
<td>2.59</td>
<td>IBM 370/165</td>
<td>D  H Ext., Fast mult.</td>
</tr>
<tr>
<td>Northwestern</td>
<td>87.1</td>
<td>4.20</td>
<td>CDC 6600</td>
<td>S  FTN</td>
</tr>
<tr>
<td>Texas</td>
<td>105.0</td>
<td>5.63</td>
<td>CDC 6600</td>
<td>S  RN</td>
</tr>
<tr>
<td>China Lake</td>
<td>251.4</td>
<td>5.69</td>
<td>Univac 1110</td>
<td>S  V</td>
</tr>
<tr>
<td>Yale</td>
<td>252.7</td>
<td>7.33</td>
<td>DEC KL-20</td>
<td>S  F20</td>
</tr>
<tr>
<td>Bell Labs</td>
<td>197.3</td>
<td>10.1</td>
<td>Honeywell 5080</td>
<td>S  V</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>1073.4</td>
<td>11.9</td>
<td>Univac 1110</td>
<td>S  V</td>
</tr>
<tr>
<td>Iowa State</td>
<td>103.5</td>
<td>10.2</td>
<td>Intel 85/5 mod3</td>
<td>D  H</td>
</tr>
<tr>
<td>U. Ill. Chicago</td>
<td>314.10</td>
<td>11.9</td>
<td>IBM 370/158</td>
<td>D  C1</td>
</tr>
</tbody>
</table>
Top500 Since 1993

- Hans Meuer and Erich Strohmaier had a list of fastest computers ranked by peak performance.
- I had a list of benchmark results and we put the two lists together.
- Listing of the 500 most powerful computers in the World.
- Yardstick: Performance for $Ax=b$, dense problem

Maintained and updated twice a year:
- SC‘xy in the States in November
- Meeting in Germany in June
PERFORMANCE DEVELOPMENT OF HPC OVER THE LAST 28 YEARS FROM THE TOP500

- Thinking Machine CM-5 with 1024 Processors at Los Alamos Nat Lab used for nuclear weapons design
June 2021: The TOP 10 Systems *(38% of the Total Performance of Top500)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RIKEN Center for Computational Science</td>
<td>Fugaku, ARM A64FX (48C, 2.2 GHz), Tofu D Interconnect</td>
<td>Japan</td>
<td>7,299,072</td>
<td>82</td>
<td>29.9</td>
<td>14.8</td>
</tr>
<tr>
<td>2</td>
<td>DOE / OS Oak Ridge Nat Lab</td>
<td>Summit, IBM Power 9 (22C, 3.0 GHz), NVIDIA GV100 (80C), Mellonox EDR</td>
<td>USA</td>
<td>2,397,824</td>
<td>74</td>
<td>10.1</td>
<td>14.7</td>
</tr>
<tr>
<td>3</td>
<td>DOE / NNSA L Livermore Nat Lab</td>
<td>Sierra, IBM Power 9 (22C, 3.1 GHz), NVIDIA GV100 (80C), Mellonox EDR</td>
<td>USA</td>
<td>1,572,480</td>
<td>75</td>
<td>7.44</td>
<td>12.7</td>
</tr>
<tr>
<td>4</td>
<td>National Super Computer Center in Wuxi</td>
<td>Sunway TaihuLight, SW26010 (260C) + Custom</td>
<td>China</td>
<td>10,649,000</td>
<td>74</td>
<td>15.4</td>
<td>6.05</td>
</tr>
<tr>
<td>5</td>
<td>DOE / OS NERSC - LBNL</td>
<td>Perlmutter HPE Cray EX235n, AMD EPYC 64C 2.45GHz, NVIDIA A100, Slingshot-10</td>
<td>USA</td>
<td>706,304</td>
<td>69</td>
<td>2.53</td>
<td>25.5</td>
</tr>
<tr>
<td>6</td>
<td>NVIDIA Corporation</td>
<td>Selene NVIDIA DGX A100, AMD EPYC 7742 (64C, 2.25GHz), NVIDIA A100 (108C), Mellonox HDR Infiniband</td>
<td>USA</td>
<td>555,520</td>
<td>80</td>
<td>2.64</td>
<td>23.9</td>
</tr>
<tr>
<td>7</td>
<td>National Super Computer Center in Guangzhou</td>
<td>Tianhe-2A NUDT, Xeon (12C) + MATRIX-2000 (128C) + Custom</td>
<td>China</td>
<td>4,981,760</td>
<td>61</td>
<td>18.5</td>
<td>3.32</td>
</tr>
<tr>
<td>8</td>
<td>JUWELS Booster Module</td>
<td>Bull Sequana XH-2000 , AMD EPYC 7402 (24C, 2.8GHz), NVIDIA A100 (108C), Mellonox HDR Infiniband/ParTec ParaStation ClusterSuite</td>
<td>Germany</td>
<td>448,280</td>
<td>62</td>
<td>1.76</td>
<td>25.0</td>
</tr>
<tr>
<td>9</td>
<td>Eni S.p.A in Italy</td>
<td>HPC5, Dell EMC PowerEdge C4140, Xeon (24C, 2.1 GHz) + NVIDIA V100 (80C), Mellonox HDR</td>
<td>Italy</td>
<td>669,760</td>
<td>69</td>
<td>2.25</td>
<td>15.8</td>
</tr>
<tr>
<td>10</td>
<td>Texas Advanced Computing Center / U of Texas</td>
<td>Frontera, Dell C6420, Xeon Platinum, 8280 (28C, 2.7 GHz), Mellonox HDR</td>
<td>USA</td>
<td>448,448</td>
<td>61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fugaku Total System Config & Performance

- **Total # Nodes:** 158,976 nodes (1 CPU/node)
  - 384 nodes/rack x 396 (full) racks = 152,064 nodes and
  - 192 nodes/rack x 36 (half) racks = 6,912 nodes

- **Theoretical Peak Compute Performances**
  - Normal Mode (CPU Frequency 2GHz)
    - **64 bit** Double Precision FP: 488 Petaflops
    - **32 bit** Single Precision FP: 977 Petaflops
    - **16 bit** Half Precision FP (AI training): 1.95 Exaflops
    - **8 bit Integer** (AI Inference): 3.90 Exaops
  - **Theoretical Peak Memory BW:** 163 Petabytes/s

Fugaku represents 16% of the Top500 systems.

### Current #2 System Overview

#### System Performance
- Peak performance of 200 Pflop/s for modeling & simulation
- Peak performance of 3.3 Eflop/s for 16 bit floating point used in for data analytics, ML, and artificial intelligence

#### Each node has
- 2 IBM POWER9 processors
  - Each w/22 cores
  - 2.3% performance of system
- 6 NVIDIA Tesla V100 GPUs
  - Each w/80 SMs
  - 97.7% performance of system
- 608 GB of fast memory
- 1.6 TB of NVMe memory

#### The system includes
- 4608 nodes
  - 27,648 GPUs
  - Street value $10K each
- Dual-rail Mellanox EDR InfiniBand network
- 250 PB IBM Spectrum Scale file system transferring data at 2.5 TB/s
TOP500 Highlights From June 2021

- Japanese’s Fugaku continues as #1 in the TOP500
  - 16% of the sum of the TOP500 performance
  - It performed at over 2 Exaflop/s on the HPL-AI using mixed precision algorithm (16-bit floating point arithmetic)
- TOP10 has one new system, Perlmutter at LBNL from HPE/Cray, AMD, & NVIDIA
  - 38% of the Top500 performance in the Top10
- The entry level to the list moved up to the 1.52 Pflop/s mark on the Linpack benchmark.
- China: Top consumer and producer overall.
- Intel processors largest share, 86% followed by AMD, 10%.
Countries Share

Number of Systems in a Country

Top 10 Performance

China: 186
US: 123
Japan: 35
Germany: 23
France: 16
Canada: 11
UK: 11
Italy: 6
Russia: 3
Rumored 2 Exascale Systems Up and Running in Chinese

- Qingdao Marine Sunway Pro “OceanLight” supercomputer (national lab) (Shandong Prov)
  - Completed March 2021, ~1.3 EFlops Rpeak, ~1.05 EFlops Rmax full Linpack run
  - ShenWei post-Alfa CPU ISA architecture with big & small core structure
  - Est 96 cabinets x 1024 SW39010 390-core CPU with pan-tree next-gen connect. 35MW +/- 10%
  - Access for outside institutions slowly rolling out

- NSCC Tianjin Tianhe-3 supercomputer (recall 2010 Tianhe-1A)
  - Dual-chip FeiTeng ARM and Matrix accelerator node architecture
  - Full completion expected Oct 2021
  - Est ~1.7 EFlops Rpeak and 1.3 EFlops Rmax, full system Linpack results unknown at this time
  - Likely to be more open to outside access than OceanLight
**HPCG Results; The Other Benchmark**

- High Performance Conjugate Gradients (HPCG).
- Solves $Ax=b$, $A$ large, sparse, $b$ known, $x$ computed.
- An optimized implementation of PCG contains essential computational and communication patterns that are prevalent in a variety of methods for discretization and numerical solution of PDEs.

**Patterns:**
- Dense and sparse computations.
- Dense and sparse collectives.
- Multi-scale execution of kernels via MG (truncated) V cycle.
- Data-driven parallelism (unstructured sparse triangular solves).
- Strong verification (via spectral properties of PCG).

hpcg-benchmark.org  With Piotr Luszczek and Mike Heroux
HPCG Details

3D Laplacian discretization

Preconditioned Conjugate Gradient solver

\[ p_0 \leftarrow x_0, \quad r_0 \leftarrow b - Ap_0 \]

\[ \text{for } i = 1, 2, \text{ to } \text{max\_iterations} \text{ do} \]

\[ z_i \leftarrow M^{-1}r_{i-1} \]

\[ \text{if } i = 1 \text{ then } \quad \text{Multigrid and Gauss-Seidel} \]

\[ p_i \leftarrow z_i \]

\[ \alpha_i \leftarrow \text{dot\_prod}(r_{i-1}, z_i) \]

\[ \text{else} \]

\[ \alpha_i \leftarrow \text{dot\_prod}(r_{i-1}, z_i) \]

\[ \beta_i \leftarrow \alpha_i / \alpha_{i-1} \]

\[ p_i \leftarrow \beta_i p_{i-1} + z_i \]

\[ \text{end if} \]

\[ \alpha_i \leftarrow \text{dot\_prod}(r_{i-1}, z_i) / \text{dot\_prod}(p_i, Ap_i) \]

\[ x_{i+1} \leftarrow x_i + \alpha_i p_i \]

\[ r_i \leftarrow r_{i-1} - \alpha_i Ap_i \]

\[ \text{if } \|r_i\|_2 < \text{tolerance} \text{ then} \]

\[ \text{STOP} \]

\[ \text{end if} \]

\[ \text{end for} \]
<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Computer</th>
<th>Cores</th>
<th>HPL Rmax (Pflop/s)</th>
<th>TOP500 Rank</th>
<th>HPCG (Pflop/s)</th>
<th>Fraction of Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RIKEN Center for Computational Science Japan</td>
<td><strong>Fugaku</strong>, Fujitsu A64FX 48C 2.2GHz, Tofu D, Fujitsu</td>
<td>7,630,848</td>
<td>442.0</td>
<td>1</td>
<td>16.0</td>
<td>3.0%</td>
</tr>
<tr>
<td>2</td>
<td>DOE/SC/ORNL USA</td>
<td><strong>Summit</strong>, AC922, IBM POWER9 22C 3.7GHz, Dual-rail Mellanox FDR, NVIDIA Volta V100, IBM</td>
<td>2,414,592</td>
<td>148.6</td>
<td>2</td>
<td>2.93</td>
<td>1.5%</td>
</tr>
<tr>
<td>3</td>
<td>DOE/SC/LBNL USA</td>
<td><strong>Perlmutter</strong>, HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10</td>
<td>761,856</td>
<td>64.6</td>
<td>5</td>
<td>1.91</td>
<td>2.0%</td>
</tr>
<tr>
<td>4</td>
<td>DOE/NNSA/LLNL USA</td>
<td><strong>Sierra</strong>, S922LC, IBM POWER9 20C 3.1 GHz, Mellanox EDR, NVIDIA Volta V100, IBM</td>
<td>1,572,480</td>
<td>94.6</td>
<td>3</td>
<td>1.80</td>
<td>1.4%</td>
</tr>
<tr>
<td>5</td>
<td>NVIDIA USA</td>
<td><strong>Selene</strong>, DGX SuperPOD, AMD EPYC 7742 64C 2.25 GHz, Mellanox HDR, NVIDIA Ampere A100</td>
<td>555,520</td>
<td>63.5</td>
<td>6</td>
<td>1.62</td>
<td>2.0%</td>
</tr>
<tr>
<td>6</td>
<td>Forschungszentrum Juelich (FZJ) Germany</td>
<td><strong>JUWELS Booster Module</strong>, Bull Sequana XH2000, AMD EPYC 7402 24C 2.8GHz, Mellanox HDR InfiniBand, NVIDIA Ampere A100</td>
<td>449,280</td>
<td>44.1</td>
<td>8</td>
<td>1.28</td>
<td>1.8%</td>
</tr>
<tr>
<td>7</td>
<td>Saudi Aramco Saudi Arabia</td>
<td>Dammam-7, Cray CS-Storm, Xeon Gold 6248 20C 2.5GHz, Infiniband HDR 100, NVIDIA Volta V100, HPE</td>
<td>672,520</td>
<td>22.4</td>
<td>11</td>
<td>0.88</td>
<td>1.6%</td>
</tr>
<tr>
<td>8</td>
<td>Eni S.p.A. Italy</td>
<td><strong>HPCS</strong>, PowerEdge, C4140, Xeon Gold 6252 24C 2.1 GHz, Mellanox HDR, NVIDIA Volta V100, Dell</td>
<td>669,760</td>
<td>35.5</td>
<td>9</td>
<td>0.86</td>
<td>1.7%</td>
</tr>
<tr>
<td>9</td>
<td>Information Technology Center, The University of Tokyo, Japan</td>
<td><strong>Wisteria/BDEC-01 (Odyssey)</strong>, PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu D</td>
<td>368,640</td>
<td>22.1</td>
<td>13</td>
<td>0.82</td>
<td>3.2%</td>
</tr>
<tr>
<td>10</td>
<td>Japan Agency for Marine-Earth Science and Technology</td>
<td><strong>Earth Simulator -SX-Aurora TSUBASA</strong> , A412-8, Vector Engine Type20B 8C 1.6GHz, Infiniband HDR200</td>
<td>43,776</td>
<td>0.01</td>
<td>41</td>
<td>0.75</td>
<td>5.6%</td>
</tr>
</tbody>
</table>
Comparison between Peak and HPL for June 2021
Comparison between Peak, HPL, and HPCG for June 2021
# DOE HPC Roadmap to Exascale Systems

<table>
<thead>
<tr>
<th>FY 2012</th>
<th>FY 2016</th>
<th>FY 2018</th>
<th>FY 2021</th>
<th>FY 2022</th>
<th>FY 2023</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORNL</strong></td>
<td><strong>ORNL</strong></td>
<td><strong>ORNL</strong></td>
<td><strong>ORNL</strong></td>
<td><strong>ORNL</strong></td>
<td><strong>ORNL</strong></td>
</tr>
<tr>
<td>Titan</td>
<td>Summit</td>
<td>Summit</td>
<td>Summit</td>
<td>ORNL</td>
<td>ORNL</td>
</tr>
<tr>
<td>Cray/AMD/NVIDIA</td>
<td>IBM/NVIDIA</td>
<td>IBM/NVIDIA</td>
<td>IBM/NVIDIA</td>
<td>HPE/AMD</td>
<td>HPE/AMD</td>
</tr>
<tr>
<td><strong>ANL</strong></td>
<td><strong>LANL/SNL</strong></td>
<td><strong>LANL/SNL</strong></td>
<td><strong>LANL/SNL</strong></td>
<td><strong>LANL/SNL</strong></td>
<td><strong>LANL/SNL</strong></td>
</tr>
<tr>
<td>IBM BG/Q</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
</tr>
<tr>
<td><strong>LLNL</strong></td>
<td><strong>LLNL</strong></td>
<td><strong>LLNL</strong></td>
<td><strong>LLNL</strong></td>
<td><strong>LLNL</strong></td>
<td><strong>LLNL</strong></td>
</tr>
<tr>
<td>Sequoia</td>
<td>Trinity</td>
<td>Sierra</td>
<td>Sierra</td>
<td>Sierra</td>
<td>Sierra</td>
</tr>
<tr>
<td>LLNL BG/Q</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
<td>Cray/Intel Xeon/KNL</td>
</tr>
<tr>
<td><strong>Exascale Systems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exascale architectures will excel at HPC and HPC + AI problems
Exascale is costing DOE $3.6B in total, over 5 years
What do you get for $3.6B

- 3 computers
  - $600M each

- 21 Applications

- A bunch of software (84 projects)
What’s Next After Exascale? - AI for Science

- Over 1,300 scientists participated in four town halls during the summer/fall of 2019
- Research Opportunities in AI
  - Biology, Chemistry, Materials,
  - Climate, Physics, Energy, Cosmology
  - Mathematics and Foundations
  - Data Life Cycle
  - Software Infrastructure
  - Hardware for AI
  - Integration with Scientific Facilities
- Modeled after the Exascale Series in 2007
- DOE’s Office of Science Advisory Subcommittee Report Sept 2020

https://www.anl.gov/ai-for-science-report

https://doi.org/10.2172/1734848
Modern Hardware: Lower Precision for Deep Learning

- Hardware (company)
  - GPU Tensor Cores (NVIDIA)
  - TPU MXU (Google)
  - Zion (Facebook)
  - DaVinci (Huawei)
  - Dot-product engine (HPE)
  - Eyeriss (Amazon)
  - Wafer Scale Engine (Cerebras)
  - Nervana (Intel)
  - Deep Learning Boost (Intel AI)
  - Graph Core
  - ...

- Lower-precision benchmarks
  - Baidu
  - Dawn
  - mlperf
  - Deep500
  - ...
  - HPL-AI

60+
WHY MIXED PRECISION? (Less is Faster)

- There are many reasons to consider mixed precision in our algorithms...
  - **Less Communication**
    - Reduce memory traffic
    - Reduce network traffic
  - **Reduce memory footprint**
  - **More Flop per second**
    - Reduced energy consumption
    - Reduced time to compute
  - Accelerated hardware in current architecture.
  - Suitable numerical properties for some algorithms & problems.

---

HPL-AI Benchmark Utilizing 16-bit Arithmetic

1. Generate random linear system $Ax = b$
2. Represent the matrix $A$ in low precision (16-bit floating point)
3. Factor $A$ in lower precision into $LU$ by Gaussian elimination
4. Compute approximate solution with $LU$ factors in low precision
5. Perform up to 50 iterations of refinement, e.g., GMRES to get accuracy up to 64-bit floating point
6. Use $LU$ factors for preconditioning
7. Validate the answer is correct: scaled residual small
   \[
   \frac{||Ax - b||}{||A||||x|| + ||b||} \times \frac{1}{n\epsilon} \leq O(10)
   \]
8. Compute performance rate as
   \[
   \frac{2}{3} \times \frac{n^3}{\text{time}}
   \]

Iterative refinement for dense systems, $Ax = b$, can work this way.

\[
LU = lu(A) \quad \text{lower precision} \quad O(n^3)
\]

\[
x = U \backslash (L \backslash b) \quad \text{lower precision} \quad O(n^2)
\]

GMRes preconditioned by the $LU$ to solve $Ax = b$

\[
\text{FP64 precision} \quad O(n^2)
\]
<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>Computer</th>
<th>Cores</th>
<th>HPL Rmax (Eflop/s)</th>
<th>TOP500 Rank</th>
<th>HPL-AI (Eflop/s)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RIKEN Center for Computational Science, Japan</td>
<td>Fugaku, Fujitsu A64FX, Tofu D</td>
<td>7,630,848</td>
<td>0.442</td>
<td>1</td>
<td>2.0</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>DOE/SC/ORNLE USA</td>
<td>Summit, AC922 IBM POWER9, IB Dual-rail FDR, NVIDIA V100</td>
<td>2,414,592</td>
<td>0.149</td>
<td>2</td>
<td>1.15</td>
<td>7.7</td>
</tr>
<tr>
<td>3</td>
<td>NVIDIA USA</td>
<td>Selene, DGX SuperPOD, AMD EPYC 7742 64C 2.25 GHz, Mellanox HDR, NVIDIA A100</td>
<td>555,520</td>
<td>0.063</td>
<td>6</td>
<td>0.63</td>
<td>9.9</td>
</tr>
<tr>
<td>4</td>
<td>DOE/SC/LBNL/NERSC USA</td>
<td>Perlmutter, HPE Cray EX235n, AMD EPYC 7763 64C 2.45 GHz, Slingshot-10, NVIDIA A100</td>
<td>761,856</td>
<td>0.065</td>
<td>5</td>
<td>0.59</td>
<td>9.1</td>
</tr>
<tr>
<td>5</td>
<td>Forschungszentrum Juelich (FZJ) Germany</td>
<td>JUWELS Booster Module, Bull Sequana XH2000, AMD EPYC 7402 24C 2.8GHz, Mellanox HDR InfiniBand, NVIDIA A100, Atos</td>
<td>449,280</td>
<td>0.044</td>
<td>8</td>
<td>0.47</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>University of Florida USA</td>
<td>HiPerGator, NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR</td>
<td>138,880</td>
<td>0.017</td>
<td>23</td>
<td>0.17</td>
<td>9.9</td>
</tr>
<tr>
<td>7</td>
<td>Information Technology Center, The University of Tokyo, Japan</td>
<td>Wisteria/BDEC-01 (Odyssey), PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu D, Fujitsu</td>
<td>368,640</td>
<td>0.022</td>
<td>13</td>
<td>0.10</td>
<td>4.5</td>
</tr>
<tr>
<td>8</td>
<td>National Supercomputer Centre (NSC), Sweden</td>
<td>Berzelius, NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, A100, Infiniband HDR, Atos</td>
<td>59,520</td>
<td>0.005</td>
<td>84</td>
<td>0.05</td>
<td>9.9</td>
</tr>
<tr>
<td>9</td>
<td>Information Technology Center, Nagoya University, Japan</td>
<td>Flow Type II subsystem, PRIMERGY CX2570 M5, Xeon Gold 6230 20C 2.1GHz, NVIDIA Tesla V100 SXMM2, Infiniband EDR</td>
<td>79,560</td>
<td>0.0049</td>
<td>87</td>
<td>0.03</td>
<td>4.3</td>
</tr>
<tr>
<td>10</td>
<td>#CloudMTS Russia</td>
<td>MTS GROM, NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, A100 40GB, Infiniband</td>
<td>19,840</td>
<td>0.0023</td>
<td>245</td>
<td>0.015</td>
<td>7</td>
</tr>
</tbody>
</table>
Comparison between HPL-AI, Peak, HPL, and HPCG for June 2021
2026 and 2030 Planning Targets

2026 – 10 Eflop/s (fp64) and >100 Eflop/s (AI bfp16)

2030 – 50 Eflop/s (fp64) and > 1000 Eflop/s (1 Zflop/s) (AI bfp16)

A few questions:
• How achievable are these targets given the roadmaps and vendor plans?
• Will AI accelerators (distinct from GPUs) make sense to integrate into future nodes or as sub-clusters?
• When will quantum computing accelerators intersect mainstream supercomputing?
## Zettascale System Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak performance</td>
<td>1 Zflops</td>
</tr>
<tr>
<td>Power consumption</td>
<td>100 MW</td>
</tr>
<tr>
<td>Power efficiency</td>
<td>10 Tflops/W</td>
</tr>
<tr>
<td>Peak performance per node</td>
<td>10 Pflops/node</td>
</tr>
<tr>
<td>Bandwidth between nodes</td>
<td>1.6 Tb/s</td>
</tr>
<tr>
<td>I/O bandwidth</td>
<td>10–100 PB/s</td>
</tr>
<tr>
<td>Storage capacity</td>
<td>1 ZB</td>
</tr>
<tr>
<td>Floor space</td>
<td>1000 m²</td>
</tr>
</tbody>
</table>

### Chinese proposes Zettascale by 2035

- 600x Frontier (58% CAGR)
- 3.4x Frontier (9% CAGR)
- 200x Frontier (46% CAGR)
- 66x Frontier (=> 10x nodes)
- 16x Frontier (22% CAGR)
- 1000x Frontier (64% CAGR)
- 1000x Frontier (64% CAGR)
- 2x Frontier (5% CAGR)

https://doi.org/10.1631/FITEE.1800494 Front Inform Technol Electron Eng
Emergence of AI-Specific Hardware Ecosystem

MYTHIC  DEEPHi  GRAPHCORE  NVIDIA
thinci  RAIN  WAVE COMPUTING
aws  NEUROMORPHICS  Google  intel
flexlogix  cerebras  XILINX
Baidu  SambaNova SYSTEMS
What’s Next Summary

• **US Exascale deployments 2021-2023: Frontier, Aurora and El Capitan**
  • Roadmap for 2025 and 2030 increasingly challenging as fab hit Angstrom nodes

• **AI for Science and AI Grand Challenges will require Exascale and More**
  • Inverse materials design and improved climate models among many

• **AI driven surrogates have the potential for “effective” Zetta and Yotta scale**
  • Improvements in accuracy, UQ, network search, automatic transformations

• **AI hardware accelerators 2nd and 3rd generation in flight**
  • Open question is how to integrate into HPC architectures and ecosystems
  • Need tight integration in memory space and
  • Integration of programming models Python/Julia, C++, OneAPI, etc.

• **Quantum Computing could breakout in the next decade**
  • Initial opportunities in quantum simulation, dependent on scalability and error correction
  • Will be tightly coupled to classical and AI systems for control, sampling, programming, and optimization

• Interesting article on AI: [https://arxiv.org/pdf/2104.12871.pdf](https://arxiv.org/pdf/2104.12871.pdf)
The Take Away

• HPC Hardware is Constantly Changing
  • Scalar
  • Vector
  • Distributed
  • Accelerated
  • Mixed precision
• Three computer revolutions
  • High performance computing
  • Deep learning
  • Edge & AI
• Algorithm / Software advances follows hardware.
  • And there is “plenty of room at the top”
• We will need additional benchmarks to measure performance for all this.