Exascale Computing and Big Data

BY DANIEL A. REED AND JACK DONGARRA

Communications of the ACM | July 2015 | Vol. 58 | No. 57

Avery Dunn
“Nothing tends so much to the advancement of knowledge as the application of a new instrument. The native intellectual powers of men in different times are not so much the causes of the different success of their labors, as the peculiar nature of the means and artificial resources in their possession.”

– Humphrey Davey
The interoperability and scaling convergence of these two ecosystems is crucial to the future.
History of high-performance computing

- Vector supercomputing
- Massively Parallel Processing (MPPs)
- Shared Memory Multiprocessors (SMPs)
- Clusters of commodity (Intel/AMD x86)
- Purpose-built processors
- Clusters augmented with computational accelerators (Intel coprocessors, Nvidia GPUs)

- “Performance-first hardware”

1980 1990 2000 2010

History of data analytics

- Largest data storage systems contain terabytes of secondary disk storage
- Commercial and research cloud-computing systems contain many petabytes of secondary storage

2000 2010
Today's smartphone computes as fast as yesterday's supercomputer.

*HPL = High Performance LINPACK

First successful supercomputer; Apple iPhone 6 substantially exceeds this performance
Today's personal music collection is as large as yesterday's enterprise-scale storage.

Figure 3. Growth of Amazon S3 objects.
To the future and beyond

- **Exascale** \(10^{18}\) operations per second) is the next proxy in long trajectory of exponential performance increases
- The frontiers of “big data” are made up of:
  - Large-scale data preservation
  - Sustainability within and across disciplines
  - Metadata creation and multidisciplinary fusion
  - Digital privacy and security
- One of the ultimate goals for advanced computing is practical and delivered capability to engineering and scientific researchers
With great power comes great responsibility...

- Designing and delivering exascale computing systems brings technical, organizational, cultural, and economic challenges.
- Exascale computing systems cannot be produced in an affordable and reliable way.
- Research-and-development costs to create an exascale computing system have been estimated to exceed $1 billion with annual operating costs at tens of millions.
- U.S. support for basic research is at a decadal low.
- The shift from PCs to mobile devices has raised competition among U.S. dominated x86 and ARM.
- There is competition surrounding national sovereignty, data security, Internet governance, data services, cloud computing operations, etc.
Scientific & Engineering Opportunities

Biology and biomedicine
High-energy physics
Climate science
Cosmology and astrophysics
Astronomy

Cancer treatment
Experimental and computational materials science
Steel production
Text and data mining
## Technical Challenges in Advanced Computing

The U.S. Department of Energy has identified 10 research challenges in developing a new generation of advanced computing systems based on studies conducted over the past five years (approx. 2010 - 2015).

<table>
<thead>
<tr>
<th></th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Energy-efficient circuit, power, and cooling technologies</td>
</tr>
<tr>
<td>2.</td>
<td>High-performance interconnect technologies</td>
</tr>
<tr>
<td>3.</td>
<td>Advanced memory technologies to improve capacity</td>
</tr>
<tr>
<td>4.</td>
<td>Scalable system software that is power and failure aware</td>
</tr>
<tr>
<td>5.</td>
<td>Data management software that can handle the volume, velocity, and diversity of data</td>
</tr>
<tr>
<td>6.</td>
<td>Programming models to express massive parallelism, data locality, and resilience</td>
</tr>
<tr>
<td>7.</td>
<td>Reformulation of science problems and refactoring solution algorithms</td>
</tr>
<tr>
<td>8.</td>
<td>Ensuring correctness in the face of faults, reproducibility, and algorithm verification</td>
</tr>
<tr>
<td>9.</td>
<td>Mathematical optimization and uncertainty quantification for discovery, design, and decision</td>
</tr>
<tr>
<td>10.</td>
<td>Software engineering and supporting structures to enable productivity</td>
</tr>
</tbody>
</table>
Technical Challenges in Advanced Computing Cont’d

These challenges were identified by the U.S. Department of Defense (as opposed to the U.S. DoE) in a number of hardware and software studies.

1. Energy-efficient operation
2. Memory capacity
3. Concurrency and locality
4. Resilience
5. Application scaling
6. Managing parallelism
7. Software tools

Moreover, a 2011 study done by the National Academy of Sciences (NAS) has suggested the exponential increases in performance from shrinking semiconductors is coming to an end.
Hardware & Architecture Challenges

Post-Dennard Scaling

- **Moore’s Law** has rested on the principle of Dennard scaling (as transistors get smaller, power density stays constant)
- Transistor size \(\downarrow\), power consumption \(\uparrow\)*
- Semiconductor challenges → rethinking design with power consumption at the forefront
- Chip power limits → stronger interest in ARM ecosystem (made for this!)

Resilience & Energy Efficiency at Scale

- Growth of advanced computing and data analysis systems → reliable operation starting to crumble
- Individual component failures increasingly less frequent but large component count → more frequent failure
- Scale brings challenges such as energy management and thermal dissipation
- Advanced computing and data analysis systems consume megawatts of power → limits on placement (cooling/peak power loads)

*no longer decreases, not to say that it increases
To reiterate: with great power comes great responsibility. 

There are a plethora of challenges in different realms introduced by the massive scale of advanced computing and data analytics:

- **Software and algorithmic challenges**
  - Come as a consequence to extreme system scale

- **Locality and scale**
  - Billion-way computational concurrency, aggressive parallelism; load balance is key

- **Adaptive system software**
  - Extreme scale, hardware heterogeneity, system power, and heat dissipation constraints influence system software as much as applications

- **Parallel programming support**
  - Programming models and tools are considered the biggest point of divergence between the scientific computing and big data ecosystems

- **Algorithmic and mathematics challenges**
  - Need for increasing amounts of data locality and need to obtain higher levels of fine-grain parallelism → adaptation from parallel algorithms to exploit hardware
They say politics is the art of looking for trouble.

There is an ongoing shift in consumer preferences, creating international competition:

- PCs and local servers
- Mobile devices and cloud computing services
- Intel/AMD x86 chips
- Energy-efficient microprocessors and systems-on-a-chip (SoCs), ARM architecture
International Exascale Projects

- **European Union (EU)**
  - Collaborative Research into Exascale Systemware, Tools, and Applications (CRESTA)
  - Dynamic Exascale Entry Platform (DEEP)

- **Japan**
  - Japan Ministry of Education, Culture, Sports, Science, and Technology (MEXT) selected RIKEN to develop and deploy an exascale system by 2020 (which has since happened, FUGAKU is now the world’s fastest supercomputer!)

- **China**
  - Tianhe-2 was the world’s faster supercomputer in 2015, which has since been overtaken by Japan and U.S. designs

- **United States**
  - U.S. Networking and Information Technology Research and Development program has spanned several research missions and agencies
  - U.S. Department of Energy is the most active deployer of high-performance computing systems and developer of plans for exascale computing
“It’s the hope that kills you”

- End of Dennard scaling, mobile and cloud computing takeover, and explosive growth of data are both challenges and opportunities
- Exascale computing (i.e. high-end computing) and big data (i.e. high-end data analytics) are essential to the research-and-development agenda
- Next-generation algorithm, software, and application development is as important as semiconductor and hardware development
- Private-sector competition and global-research collaboration are necessary to design, develop, and deploy exascale and big data computing systems
1. What are three of the technical challenges (identified by the U.S. Department of Energy) in advanced computing?

2. What is Dennard scaling and why is post-Dennard scaling deemed an attention-worthy issue for exascale computing?

3. What has the shift in PCs and local servers to mobile devices and cloud computing also introduced in terms of the underlying architecture of the systems?

Bonus: What is the world’s fastest supercomputer? 😎