The study is still being written up

This is only my perspective

This material is based on work sponsored by DARPA, AFRL, and GTRI
HPCS Program targets Petascale systems circa 2010

What research is needed for Exascale, circa 2018?

I know of two efforts to address this question:
NSA Advanced Computing Systems
Gary Hughes
DARPA ExaScale Computing Study
Bill Harrod
Determine what research the government needs to fund to enable its computer vendors to credibly decide, circa 2011, to initiate product development for Petascale systems that would be available later in the next decade.
<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Name</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shekhar Borkar</td>
<td>Intel</td>
<td>Dean Klein</td>
<td>Micron</td>
</tr>
<tr>
<td>Dan Campbell</td>
<td>GTRI</td>
<td>Peter Kogge</td>
<td>Notre Dame</td>
</tr>
<tr>
<td>Bill Carlson</td>
<td>IDA</td>
<td>Bob Lucas</td>
<td>USC/ISI</td>
</tr>
<tr>
<td>Bill Dally</td>
<td>Stanford</td>
<td>Mark Richards</td>
<td>Georgia Tech</td>
</tr>
<tr>
<td>Monty Denneau</td>
<td>IBM</td>
<td>Al Scarpelli</td>
<td>AFRL</td>
</tr>
<tr>
<td>Paul Franzon</td>
<td>NC State</td>
<td>Steve Scott</td>
<td>Cray</td>
</tr>
<tr>
<td>Bill Harrod</td>
<td>DARPA</td>
<td>Allan Snavely</td>
<td>SDSC</td>
</tr>
<tr>
<td>Kerry Hill</td>
<td>AFRL</td>
<td>Thomas Sterling</td>
<td>LSU</td>
</tr>
<tr>
<td>Jon Hiller</td>
<td>STA</td>
<td>Stan Williams</td>
<td>HP</td>
</tr>
<tr>
<td>Sherman Karp</td>
<td>STA</td>
<td>Kathy Yelick</td>
<td>LBNL &amp; UCB</td>
</tr>
<tr>
<td>Steve Keckler</td>
<td>University of Texas</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bill Harrod is the DARPA Program Manager
Peter Kogge is the Principle Investigator
Participants
Countless Other Contributors

David {Bailey, Koester} LBNL and MITRE
Keren Bergman Columbia
Loring Craymer NSA ACS

Lots of people from each host institution.
Four Meetings

Meeting #1: May 30, STA
Meeting #2: June 26-27, HP
Topic #1: Packaging: July 17-18, Georgia Tech
Meeting #3: July 24-25, Intel
Meeting #4 Memory Roadmap and issues August 16-17, Micron
Topic #2: Architectures and Programming August 30, Stanford University
Topic #3: Applications, Storage, and I/O September 6-7, UC Berkeley
Topic #4: Optical interconnects Sept. 25-26, Stanford University
Meeting #5: October 10-11, USC/ISI
Meeting #6: November 15, SC|07
Some Stuff We’ve Discussed

Power
Memory volume
Programming
Reliability
Packaging
Pervasive problem, requires range of solutions!

Architecture:
- Intel Polaris spent most power issuing instructions

Process and circuit technology
- Sub threshold devices?

Optics:
- Interconnect (DARPA MTO program)
- Clock distribution (save 10 W per socket?)

Memory:
- ~5W for a GByte DRAM
Memory Capacity Challenge

1,000,000 DRAM chips, circa 2014

Compliments of Alan Charlesworth (Sun) and David Koester (MITRE)

ExaBytes

Peak Performance (TFlops/s)

Memory Size (TiBytes)

Less balanced

More balanced

Pelascale Leadership-Class System HEC Revitalization Task Force

Vector

MPP

Cluster

Fat node
A PetaByte main memory won’t be useless
   There are applications that look like Linpack
   Others like sPPM have small footprints
For scaled speedup, we’ll need more main memory
   Novel technology for main memory
   EDRAM for L2
   DRAM as L3
What about scratch and the file system?
What about archives?
Programming Challenge

Perhaps ten billion threads!
What would Mr. Amdahl think?
MPI will suffice for a few stunts.
   MPI + OpenMP per socket isn’t much better.

I’m hoping for something like UPC.
   UPC has impact because applications adopted it

God forbid, but is CUDA the model for the future?
   Multi-core
   Multi-threaded
   SIMD extensions
   Explicit memory hierarchy
Fault Tolerance

Some say it’s “only a factor of two away”.
(i.e., build two systems and compare results)
Actually, for much of the system, it’s better than that today.
Memories and transmission lines already protected
Effectively protecting logic still an issue.
Packaging

Don’t want to measure computers by the acre.
In the best case, distance equals latency.
Minimize power.

It's not clear that this is a key bottleneck to achieving Exascale.
This was been a really fun, enlightening exercise 😊

Remarkably conservative! Exotic technology may not be required.

E.g., SiGe or SFQ
Nor cooling the system to 70K, much less 4K

The space of applications is getting smaller
How many will run effectively at O(1B) threads?
Five orders-of-magnitude from today’s extreme.
Impact on DOE SC

Good news:
Growth in raw computing power will continue unabated
Enables scientific discoveries beyond imagination today

Bad news:
Even after twenty years, we’re still not done porting codes to parallel systems.
Concurrency will increase 4-5 orders-of-magnitude.
System balance will change dramatically.
Number of successful codes (even whole fields) will decline.
Facilities will have to transform (again!) to adapt (e.g., power).
An Exascale computing system within a decade is plausible.

There are a number of significant problems that will need to be overcome. The DOE should look towards addressing them now, while there’s time. DOE should continue its partnership with DOD (DARPA & NSA).