Frontiers of Extreme Computing 2007
Zettaflops Workshop

Erik P. DeBenedictis
Acknowledgements

- David Womble to introduce workshop
  - Thank Erik DeBenedictis for being general chair
  - Thank Thomas Sterling for being deputy chair and also for developing the Petaflops workshops 1994 onward
  - Thank Horst for also founding of Zettaflops
  - Acknowledge Sandia/CSRI for producing the workshop
  - Thank organizing committee
    - In alphabetical order, Bill Camp, Candy Culhane, Mike Foster, Jag Shah, Horst Simon, Rick Stevens, Tom Theis, Stan Williams, David Womble
  - Thank financial contributors
    - In order of contribution, Sandia, HP, Intel, DOE/Sc, DARPA, LBL, Cray
  - Thank Caltech for hosting the Website
  - Thank Yeen Mankin of LBL and Deanna Ceballos and Bernadette Watts of Sandia for their support
  - Thank speakers and group leads for their extra effort
  - Thank all participants for expending the effort of travel and their time and assure them that we hope to give them a positive return
  - Thank Thomas Sterling in advance for writing the monograph
History
History and Book

- 1994 Petaflops I, Pasadena
- 1999 Petaflops II, Santa Barbara
- 2002 WIMPS, Bodega Bay
- 2005 Zettaflops, Santa Cruz
- 2007 Zettaflops, Santa Cruz

[Note: there were other activities]
Petaflops/Zettaflops Format

• These are interdisciplinary workshops on computation in the future
  – Technology is best sold for the benefit of its use to society
    • This is an objective of the workshop
  – We assemble people representing the self-organized “technology stack” that benefits society through computation, reinforcing our team
From Petaflops Workshop 1994

The objectives of the [1994] workshop were to:

- Identify applications that require PetaFLOPS performance and determine their resource demands
- Determine the scope of the technical challenge to achieving effective PetaFLOPS computing
- Identify critical enabling technologies that lead to PetaFLOPS computing capability
- Establish key research issues
- Recommend elements of a near-term research agenda
Continuity and Changes
Petaflops (94→03) to Zettaflops (04→07)

• We all lived out the last decade, buying ever faster PCs
• A decade ago, the vision of computation was limited by our imagination because the technology was set to grow exponentially in power over time (Moore’s Law)
• Moore’s Law delivered easy-to-exploit clock rate increases, as well as density increases

• In this workshop, we will see somewhat of a reversal
• Moore’s Law delivers through more parallelism
• We will see compelling applications that exceed the ability of technology to execute
• Thus, we have this workshop to organize efforts to improve the technology
What Can We Accomplish?
(Erik’s Suggestion, need you help)
What Can We Accomplish?
(Erik’s Suggestion, need you help)

- We have a unique group
  - Broader: Devices through applications
- There are several post-petaflops initiatives approaching Congress
  - Note: I will recommend broadening beyond FLOPS
- Zettaflops is not a part of any such initiative, but we are funded by DOE, DARPA, and have participation by several other Government agencies, and industry

- Action: Leverage our unique breadth by thinking through the key, broad cross-cutting issue of the day
- See if we can support one or more of the advanced computing initiatives, increasing the likelihood of their getting funded
- The cross-cutting issue: how much value to society will result from different computational technology investments
Objective of Workshop: Fill In Blanks Here

<table>
<thead>
<tr>
<th>Supercomputer Performance (5 MW)</th>
<th>Applications</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Zettaflops</td>
<td>Full Global Climate Phil Jones</td>
<td></td>
</tr>
<tr>
<td>100 Exaflops</td>
<td>NASA Computing needs Tom Cwik</td>
<td></td>
</tr>
<tr>
<td>10 Exaflops</td>
<td>Personalized Medicine Nathan Price</td>
<td></td>
</tr>
<tr>
<td>1 Exaflops</td>
<td>Limit of CMOS with different architecture</td>
<td></td>
</tr>
<tr>
<td>100 Petaflops</td>
<td>Limit of CMOS with legacy compatible core</td>
<td></td>
</tr>
<tr>
<td>10 Petaflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Petaflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 Teraflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Teraflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Teraflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 Gigaflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Gigaflops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 Megaflops</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mobile Performance (5 W)

2000 2010 2020 2030 Year →
Erik’s Suggestion for Workshop Output

- Make proposals for short, medium, long term computation
- Suggest applications, software, architecture, and technology that everybody believes is feasible for each timeframe
- Such that the technology’s capabilities fit the application’s requirements
- This would appear to be a unique contribution...
The March to Moscow and Back
The March to an Application

Watts or Megawatts defined by application (5 MW/5W)

Logic: CMOS, etc.

Raw Logic, Memory, or Communications

Architecture: 32/64 bit floating point, Multi-core, ...

Ops useful to user (OPs/FLOPS)

Software: (algorithms, languages, …)

Sustained performance

Is this enough?

N Gates f frequency
BIT = Nf
Binary Throughput

X megaops or X megaflops per core, n cores: nX peak throughput

Overhead for easy programming (cache)

Parasitic + Interconnect Loss

Unrealized “fraction of peak”

Note: Not to scale, loss through the pipeline is around 99.9%
Monday Speakers
Tom Theis

- Talk Title: *Prospects for Computing Beyond CMOS Logic*
- Speaker Title: Director of Physical Sciences, IBM Yorktown
- History: Participated in 2005 Workshop
- Upside Potential: Could extend Moore’s Law by transition to another device concept, making it easy going for downstream participants
Horst Simon

- Talk Title: *E3 Exascale Initiative*
- Speaker Title: Acting community representative
- History with Workshop: Zettaflops Founder
- Upside Potential: E3 is an exascale initiative currently seeking funding. Both Horst and DOE/Sc are funding this workshop. If the workshop produces a useful output, E3 could see increased prospects of funding.
George Bourianoff

- **Talk Title:** *More Moore, More than Moore, beyond CMOS and the ITRS*
- **Speaker Title:** Manager of Emerging Research Technologies, Intel
- **History:** New to Workshop
- **Upside Potential:** Tracking CMOS to ultimate limits (ITRS) and discussion of other devices that may exceed CMOS capabilities
Phil Jones

• Talk Title: *Climate Modeling on Future Architectures*

• Speaker Title: Scientist, Los Alamos National Laboratory

• History: Presented at 2004 Workshop; David Bader presented climate modeling at 2005 workshop

• Upside Potential: Science validated by Nobel Prize, address “planetary emergency”
• Talk Title: Operating Systems for Exascale Computing
• Speaker Title: Endowed Professor, LSU
• History: Founded workshop in 1994, continuous participant
• Upside Potential: Addresses a layer in the technology stack with a hard job to perform, but with upside potential. Addresses system reliability.
• Talk Title: *Challenges to Reaching Exascale Computing Levels*
• Speaker Title: Associate Laboratory Director, LBL
• History: In 2004, helped found Zettaflops workshop as a follow-on to the Petaflops workshops
• Upside Potential: