**Beyond Petascale Computing -- The End of the Beginning or the Beginning of The End**

- **Document Number**: 5225272
- **SAND Number**: 2004-4328 C
- **Review Type**: Electronic
- **Status**: Approved
- **Sandia Contact**: DEBENEDICTIS, ERIK P.
- **Submit Date**: 08/30/2004
- **Requestor**: DEBENEDICTIS, ERIK P.
- **Event (Conference/Journal/Book) Info**
  - **Name**: Conference on Computational Physics 2004
  - **City**: Genoa
  - **State**:
  - **Country**: Italy
  - **Start Date**: 09/01/2004
  - **End Date**: 09/04/2004
- **Partnership Info**
  - **Partnership Involved**: No
  - **Agreement Number**: 
- **Patent Info**
  - **Scientific or Technical in Content**: Yes
  - **Technical Advance**: No
  - **TA Form Filed**: No
- **Classification and Sensitivity Info**
  - **Title**: Unclassified-Unlimited
  - **Abstract**: 
  - **Document**: Unclassified-Unlimited
  - **SD Number**: 
- **DUSA**: None.

### Routing Details

<table>
<thead>
<tr>
<th>Role</th>
<th>Routed To</th>
<th>Approved By</th>
<th>Approval Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivative Classifier Approver</td>
<td>YARRINGTON, PAUL</td>
<td>YARRINGTON, PAUL</td>
<td>08/30/2004</td>
</tr>
<tr>
<td>Classification Approver</td>
<td>WILLIAMS, RONALD L.</td>
<td>WILLIAMS, RONALD L.</td>
<td>08/30/2004</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Conditions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manager Approver</td>
<td>PUNDIT, NEIL D.</td>
<td>PUNDIT, NEIL D.</td>
<td>08/30/2004</td>
</tr>
<tr>
<td>Conditions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrator Approver</td>
<td>LUCERO, ARLENE M.</td>
<td>FARRELLY, JEREMIAH</td>
<td>05/30/2007</td>
</tr>
</tbody>
</table>

Created by WebCo  Problems? Contact CCHD: by email or at 845-CCHD (2243).

For Review and Approval process questions please contact the Application Process Owner
Beyond Petascale Computing – The End Of The Beginning Or The Beginning Of The End?

Erik P. DeBenedictis & William J. Camp
Sandia National Laboratories, USA

Conference on Computational Physics 2004
Genoa, Italy
Next Generation

Design Parameters
- True MPP, Designed to be a single system
- Fully connected high performance 3-D mesh interconnect
- Topology - 27 X 16 X 24 compute nodes and 2 X 8X 16 service and I/O nodes
- 108 compute node cabinets and 10,368 compute node processors (AMD Sledgehammer @ 2.0 GHz)
- ~10 TB of DDR memory @ 333 MHz (1.0 GB per processor)
- Red/Black switching - ~1/4, ~1/2, ~1/4
- 8 Service and I/O cabinets on each end (256 processors for each color)
- 240 TB of disk storage (120 TB per color)
- Functional hardware partitioning - service and I/O nodes, compute nodes, and RAS nodes
- Functional system software partitioning - LINUX on service and I/O nodes, LWK (Catamount) on compute nodes, stripped down LINUX on RAS nodes
- Separate RAS and system management network (Ethernet)
- Router table based routing in the interconnect
- Less than 2 MW total power and cooling
- Less than 3,000 square feet of floor space

Performance
- Peak of ~ 40 TF
- Expected MP-Linpack performance >20 TF
- Aggregate system memory bandwidth ~55 TB/s
- Interconnect
  - Aggregate sustained interconnect bandwidth > 100 TB/s
  - MPI Latency - µs neighbor, 5 µs across machine
  - Bi-Section bandwidth ~2.3 TB/s
  - Link bandwidth ~3.0 GB/s in each direction
- I/O System
  - Sustained 50 GB/s disk I/O bandwidth for each color
  - Sustained 25 GB/s external network bandwidth for each color

ETA: January 2005
Applications and Computer Technology

Technologies

1. Quantum Dots/Reversible Logic mP (green)
2. Best-case logic (red) → Best-case logic – 100k_B T limit
3. Full Global Climate

Applications

No schedule provided by source

Technology

- 100× 1000× [SCaLeS 03]
- Compute as fast as the engineer can think [NASA 99]
- Geodata Earth Station Range [NASA 02]
- Best-case logic (red) → Best-case logic – 100k_B T limit
- μP – 125× below 100k_B T limit

System Performance

- 1 Zettaflops
- 100 Exaflops
- 1 Exaflops
- 100 Petaflops
- 10 Petaflops
- 1 Petaflops
- 100 Teraflops

Year

2000 2010 2020


Outline

• The Computing of Physics: The Need for Zettaflops
• Limits of Moore’s Law
  Today’s Technologies
• An Expert System/Optimizer for Supercomputing
• The Physics of Computing:
  Reaching to Zettaflops
• Roadmap and Future Directions
Global Climate

• Objective
  – Collect data about Earth
  – Model climate into the future
  – Provide “decision support” and ability to “mitigate”

• Approaches
  – Climate models exist, but need they more resolution, better physics, and better initial conditions (observations of the Earth)

• Computer Resources Required
  – Increments over current workstation on next slide
<table>
<thead>
<tr>
<th>FLOPS Increases for Global Climate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Issue</strong></td>
</tr>
<tr>
<td>1 Zettaflops</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>100 Exaflops</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1 Exaflops</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>10 Petaflops</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>100 Teraflops</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>10 Gigaflops</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Requirements for Plasma Simulation

• Very high peak performance requirements
  – but seeking algorithmic improvements

• Two methods
  – Red regions very scalable, Monte Carlo
  – Green regions $N^4$ scaling (FEM)

• Long term objective
  – Merge methods into a single code

Ref. “Plasma Science Contribution to the SCAleS Report,”
S.C. Jardin, October 2003
Based on these inputs, various portions of the Modeling and Data Assimilation System will require anywhere from $10^7$ to $10^{13}$ GFLOPS of computational resources. In other words, the range of computational resources needed is $10^{16}$ to $10^{21}$ Floating Point Operations per Second. For the curious, the range can also be stated as 10 PetaFLOPS to 1 ZettaFLOPS.

4.1.2. Anticipated Computing Technology Capabilities

At first glance, the numbers discussed in the previous section appear so high as to be impossibly ludicrous. However, with the expected growth in computing capabilities, the lower end of this spectrum actually falls within the domain of possibility.

“...the ultimate goal of making the computing underlying the design process so capable that it no longer acts as a brake on the flow of the creative human thought...”

Requirement 3 Exaflops

Note: In the context of this report, this requirement is for one or a few engineers, not a supercomputer center!
Outline

• The Computing of Physics: The Need for Zettaflops
• Limits of Moore’s Law
  Today’s Technologies
• An Expert System/Optimizer for Supercomputing
• The Physics of Computing: Reaching to Zettaflops
• Roadmap and Future Directions
### Best-Case Logic | Microprocessor Architecture | Physical Factor | Source of Authority
--- | --- | --- | ---
100 Exaflops | 800 Petaflops | Reliability limit 750KW/(80k_B T) | Esteemed physicists (T=60°C junction temperature)
25 Exaflops | 200 Petaflops | Derate 20,000 convert logic ops to floating point | Floating point engineering (64 bit precision)
4 Exaflops | 32 Petaflops | Derate for manufacturing margin (4×) | Estimate
1 Exaflops | 8 Petaflops | Uncertainty (6×) | Gap in chart

### Assumption: Supercomputer is size & cost of Red Storm:
US$100M budget; consumes 2 MW wall power; 750 KW to active components

- 2×10^{24} logic ops/s
- Projected ITRS improvement to 22 nm (100×)
- Lower supply voltage (2×)
- Red Storm contract
Thermal Noise Limit

This logical irreversibility is associated with physical irreversibility and requires a minimal heat generation, per machine cycle, typically of the order of $kT$ for each irreversible function.
– R. Landauer 1961

$kT$ “helper line,” drawn out of the reader’s focus because it wasn’t important at the time of writing
– Carver Mead, Scaling of MOS Technology, 1994
### Semiconductor Roadmap

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2013</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YEAR OF PRODUCTION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DRAM ½ PITCH (nm)</strong></td>
<td>45</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td><strong>MPU / ASIC ½ PITCH (nm)</strong></td>
<td>50</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td><strong>MPU PRINTED GATE LENGTH (nm)</strong></td>
<td>25</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td><strong>MPU PHYSICAL GATE LENGTH (nm)</strong></td>
<td>18</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td><strong>Physical gate length high-performance (HP) (nm)</strong> [1]</td>
<td>18</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td><strong>Equivalent physical oxide thickness for high-performance $T_{ox}$ (EOT) (nm)</strong> [2]</td>
<td>0.5 0.8</td>
<td>0.4 0.6</td>
<td>0.4 0.5</td>
</tr>
<tr>
<td><strong>Gate depletion and quantum effects electrical thickness adjustment factor (nm)</strong> [3]</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>$T_{ox}$ electrical equivalent (nm)</strong> [4]</td>
<td>1.2</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td><strong>Nominal power supply voltage ($V_{dd}$) (V)</strong> [5]</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Nominal high-performance NMOS sub threshold leakage current, $I_{ds,th}$ (at 25°C) ($\mu A/\mu m$)</strong> [6]</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td><strong>Nominal high-performance NMOS saturation drive current, $I_{dd}$ (at $V_{dd}$ at 25°C) ($\mu A/\mu m$)</strong> [7]</td>
<td>1200</td>
<td>1500</td>
<td>1500</td>
</tr>
<tr>
<td><strong>Required percent current-drive &quot;mobility/transconductance improvement&quot;</strong> [8]</td>
<td>30%</td>
<td>70%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Parasitic source/drain resistance (Rsd) (ohm/µm)</strong> [9]</td>
<td>110</td>
<td>90</td>
<td>80</td>
</tr>
<tr>
<td><strong>Parasitic source/drain resistance (Rsd) percent of ideal gate resistance 2</strong></td>
<td>25%</td>
<td>30%</td>
<td>35%</td>
</tr>
<tr>
<td><strong>Parasitic capacitance percent of ideal gate capacitance 3</strong></td>
<td>31%</td>
<td>36%</td>
<td>42%</td>
</tr>
<tr>
<td><strong>High-performance NMOS device $t$ ($C_{gate}$ * $V_{dd}$ / $I_{ds-NMOS}$) [10]</strong></td>
<td>0.39</td>
<td>0.22</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>Relative device performance [13]</strong></td>
<td>4.5</td>
<td>7.2</td>
<td>10.7</td>
</tr>
<tr>
<td>*<em>Energy per ($W/L_{gate}=3$) device switching transition ($C_{gate}<em>(3</em>L_{gate})<em>V^2$) ($f/J$ / Device) [14]</em></em></td>
<td>0.015</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Static power dissipation per ($W/L_{gate}=3$) device (Watts/Device) [15]</strong></td>
<td>9.7E-08</td>
<td>1.4E-07</td>
<td>1.1E-07</td>
</tr>
</tbody>
</table>

White—Manufacturable Solutions Exist, and Are Being Optimized
Yellow—Manufacturable Solutions are Known
Red—Manufacturable Solutions are NOT Known

1,000 $k_B T$/transistor
### Scientific Supercomputer Limits

<table>
<thead>
<tr>
<th>Expert Opinion</th>
<th>Microprocessor Architecture</th>
<th>Physical Factor</th>
<th>Source of Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best-Case Logic</td>
<td>100 Exaflops 800 Petaflops</td>
<td>Reliability limit</td>
<td>Esteemed physicists</td>
</tr>
<tr>
<td>125:1</td>
<td>750KW/(80k_BT)</td>
<td>(T=60°C junction temperature)</td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td>25 Exaflops 200 Petaflops</td>
<td>Derate 20,000 convert logic ops to floating point</td>
<td>Floating point engineering</td>
</tr>
<tr>
<td>4 Exaflops 32 Petaflops</td>
<td></td>
<td>(64 bit precision)</td>
<td></td>
</tr>
<tr>
<td>1 Exaflops 8 Petaflops</td>
<td>Derate for manufacturing margin (4x)</td>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>Assumption: Supercomputer is size &amp; cost of Red Storm: US$100M budget; consumes 2 MW wall power; 750 KW to active components</td>
<td>Uncertainty (6x)</td>
<td></td>
<td>Gap in chart</td>
</tr>
<tr>
<td></td>
<td>Improved devices (4x)</td>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td></td>
<td>Projected ITRS improvement to 22 nm (100x)</td>
<td></td>
<td>ITRS committee of experts</td>
</tr>
<tr>
<td></td>
<td>Lower supply voltage (2x)</td>
<td></td>
<td>ITRS committee of experts</td>
</tr>
<tr>
<td></td>
<td>Red Storm</td>
<td></td>
<td>contract</td>
</tr>
</tbody>
</table>
Outline

• The Computing of Physics: The Need for Zettaflops
• Limits of Moore’s Law
  Today’s Technologies
• An Expert System/Optimizer for Supercomputing
• The Physics of Computing: Reaching to Zettaflops
• Roadmap and Future Directions
Supercomputer Expert System

Application/Algorithm run time model as in applications modeling

Logic & Memory Technology design rules and performance parameters for various technologies (CMOS, Quantum Dots, C Nano-tubes …)

Interconnect Speed, power, pin count, etc.

Physical Cooling, packaging, etc.

Expert System & Optimizer (looks for best 3D mesh of generalized MPI connected nodes, µP and other)

Time Trend Lithography as a function of years into the future

Results
1. Block diagram picture of optimal system (model)
2. Report of FLOPS count as a function of years into the future
Sample Analytical Runtime Model

- Simple case: finite difference equation
- Each node holds $n \times n \times n$ grid points

- Volume-area rule
  - Computing $\propto n^3$
  - Communications $\propto n^2$

$$T_{step} = 6 n^2 C_{bytes} T_{byte} + n^3 F_{grind/floprate}$$

Volume $n^3$ cells

Face-to-face $n^2$ cells
Expert System for Future Supercomputers

• Applications Modeling
  – Runtime
    \[ T_{\text{run}} = f_1(n, \text{design}) \]
• Technology Roadmap
  – Gate speed = \( f_2(\text{year}) \),
  – chip density = \( f_3(\text{year}) \),
  – cost = \$(n, \text{design}), \ldots
• Scaling Objective Function
  – I have \$C_1 \& can wait \( T_{\text{run}}=C_2 \) seconds. What is the biggest \( n \) I can solve in year \( Y \)?

• Use “Expert System” To Calculate:
  \[ \text{Max} \quad n: \$<C_1, \quad T_{\text{run}}<C_2 \]
  All designs

• Report:
  \underline{Floating operations}
  \[ T_{\text{run}}(n, \text{design}) \]
  and illustrate “design”
Outline

• The Computing of Physics: The Need for Zettaflops
• Limits of Moore’s Law
  Today’s Technologies
• An Expert System/Optimizer for Supercomputing
• The Physics of Computing: Reaching to Zettaflops
• Roadmap and Future Directions
Candidate Technologies for Zettaflops

- CMOS per Moore’s Law
  - Cluster/μP solution exceeds limits by 10,000×
    - Trillion US$ cost
    - 10 × Hoover Dam for power supply
  - Custom logic solution exceeds limits by 100×
    - US$10 billion cost
    - 100 MW power
  - ∴ worth our while to consider alternatives

- Limiting search for Alternatives to CMOS
  - Digital not Analog
  - Floating-point friendly
  - Controllable by something recognizable as “programming”
  - Mature enough for above issues to be addressed in published papers
  - Rules out coherent quantum, neural nets, DNA computing, optical interference, …
Alternatives to CMOS for Zettaflops

• New Devices
  – Superconducting: RSFQ (a. k. a. nSQUID, parametric quantrons)
  – Quantum Dots/QCA
  – Rod Logic
  – Helical Logic
  – Single Electron Transistors
  – Carbon Nanotube Y Junctions
  – ...

• Logic and Architecture
  – “Reversible logic” will be unfamiliar to today’s engineers but has been shown to be sufficient
  – Arithmetic elements and microprocessors have been demonstrated
  – Leading architecture:
    • Reversible ALU/CPU
    • Irreversible memory
1 Zettaflops Scientific Supercomputer

- How could we increase “Red Storm” from 40 Teraflops to 1 Zettaflops?
- Answer
  - >2.5×10^7 power reduction per operation
  - Faster devices × more parallelism >2.5×10^7
  - Smaller devices to fit existing packaging

An Exemplary Device: Quantum Dots

• Pairs of molecules create a memory cell or a logic gate

Ref. “Clocked Molecular Quantum-Dot Cellular Automata,” Craig S. Lent and Beth Isaksen
IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 50, NO. 9, SEPTEMBER 2003
Not Specifically Advocating Quantum Dots

- A number of post-transistor devices have been proposed.
- The shape of the performance curves have been validated by a consensus of reputable physicists.
- However, validity of any data point can be questioned.
- Cross-checking appropriate; see →

Reversible Multiplier Status

- 8×8 Multiplier Designed, Fabricated, and Tested by IBM & University of Michigan
- Power savings was up to 4:1
**QCA Microprocessor Status**

- M. Niemier Ph. D. Thesis, University of Notre Dame
- 12 Bit µP
- CAD design tool principles
  - $10 \times$ circuit density of CMOS at same $\lambda$
- Applies to various devices
  - Metal dot 4.2 nm$^2$
  - Molecular 1.1 nm$^2$

---

**Figure 4.6.** A 2-bit QCA Simple 12 ALU with registers
Reversible Microprocessor Status

• Status
  – Subject of Ph. D. thesis
  – Chip laid out (no floating point)
  – RISC instruction set
  – C-like language
  – Compiler
  – Demonstrated on a PDE
  – However: really weird and not general to program with +=, -=, etc. rather than =
CPU Design

• Leading Thoughts
  – Implement CPU logic using reversible logic
    • High efficiency for the component doing the most logic
  – Implement state and memory using conventional logic
    • Low efficiency, but not many operations
  – Permits programming much like today
Supercomputer is 211K chips, each with 70.7K nodes of 5.77K cells of 240 bytes; solves 86T=44.1Kx44.1Kx44.1K cell problem.

System dissipates 332KW from the faces of a cube 1.53m on a side, for a power density of 47.3KW/m². Power: 332KW active components; 1.33MW refrigeration; 3.32MW wall power; 6.65MW from power company.

System has been inflated by 2.57 over minimum size to provide enough surface area to avoid overheating.

Chips are at 99.22% full, comprised of 7.07G logic, 101M memory decoder, and 6.44T memory transistors. Gate cell edge is 34.4nm (logic) 34.4nm (decoder); memory cell edge is 4.5nm (memory).

Compute power is 768 EFLOPS, completing an iteration in 224µs and a run in 9.88s.
Outline

• The Computing of Physics: The Need for Zettaflops
• Limits of Moore’s Law
  Today’s Technologies
• An Expert System/Optimizer for Supercomputing
• The Physics of Computing: Reaching to Zettaflops
• Roadmap and Future Directions
Where to Go Next: You Can Help

• What is the largest FLOPS rate that can be justified on the basis of scientific discovery?
  – Not exactly for today’s applications, but for scaled up problems of the same type
  – If your answer is
    • < 1 Zettaflops: you will be in good company
    • > 1 Zettaflops, you can be the high performance leader!
• This information would be helpful in creating increasingly powerful supercomputers to enable scientific discoveries
The Future: Architecture and Software

• Software lasts a long time
  – Code written today will be debugged later this year
  – ...but may not run at full scale for decades

• What will the supercomputer be like that runs today’s code at a scale sufficient to complete the mission?
  – In many cases, the supercomputer will be of the “next generation”
  – Gross attributes of the “next generation” can be known
Will Supercomputers Grow Forever?

- Will supercomputer simulations scale up forever, or will there be a maximum?
  - Zettaflops simulates the Earth, and the Earth is the largest thing that we care about in detail

- Will progress in science always come through “simulating physics on a computer”?
  - Perhaps future problems could be formulated as a combination of symbolic reasoning (artificial intelligence) and floating point