Review & Approval

Prepare Request
Search Requests
Generate Reports
Approvals
Help
Wizard

Search Requests
New Search
Refine Search
Search Results

Clone Request
Edit Request
Cancel Request

Search Detail

Submittal Details

Document Info

Title : Maintaining Scalability of Information Processing Systems
Document Number : 5234992 SAND Number : 2005-4794 C
Review Type : Electronic Status : Approved
Sandia Contact : DEBENEDICTIS,ERIK P. Submittal Type : Conference Paper
Requestor : DEBENEDICTIS,ERIK P. Submit Date : 07/29/2005
Comments : The meeting information says "these meetings are not structured like traditional conferences, but rather as highly-participatory meetings (like 'Gordon' conferences)..." This paper is formatted like others that I find on the Web.
Peer Reviewed? : N

Author(s)

DEBENEDICTIS,ERIK P.

Event (Conference/Journal/Book) Info

Name : CRA Conference on Grand Research Challenges "Revitalizing Computer Architecture Research"
City : Monterey Bay State : CA Country : USA
Start Date : 12/04/2005 End Date : 12/07/2005

Partnership Info

Partnership Involved : No
Partner Approval : Agreement Number :

Patent Info

Scientific or Technical in Content : Yes
Technical Advance : No TA Form Filed : No
SD Number :

Classification and Sensitivity Info

Title : Unclassified-Unlimited Abstract : Document : Unclassified-Unlimited
Additional Limited Release Info : None.
DUSA : DIS-CS

Routing Details

Role	Routed To	Approved By	Approval Date
Manager Approver	PUNDIT,NEIL D.	PUNDIT,NEIL D.	07/29/2005
Conditions:			
Administrator Approver	LUCERO,ARLENE M.	KRAMER,SAMUEL	07/17/2007

Scaling Information Processing Systems to the Post-Transistor Era

Erik DeBenedictis, Sandia National Laboratories, P. O. Box 5800 MS 1110, Albuquerque NM, 87112-1110, epdeben@sandia.gov, (505) 284-4017.

The submitted Grand Research Challenge is to maintain the technological progress colloquially known as "Moore's Law" beyond the limits of transstors. In this short note, we are taking Moore's Law to represent information processing capability and are using Supercomputing FLOPS as the measurable quantity.

The figure illustrates some opportunities to improve life through computing, as well as the upcoming Grand Research Challenge of figuring out how to make a computer powerful enough to successfully address the opportunity. The left part of the figure shows projections of computer performance required to address certain important problems in society, such as the problem of generating infinite electricity from seawater via plasma fusion, simulating the full global climate to understand and mitigate global warming, and various problems in engineering. Published reports show requirements to about 1 Zettaflops (10^{21} FLOPS).

Curve ① on the right hand side shows the problem. Curve ① is a projection of Moore's Law as might apply to a $100M microprocessor-based supercomputer (i. e. the supercomputer architecture prior to the "revitalization" under discussion in this workshop). Performance rises with "Moore's Law" until the curve become asymptotic with a thermodynamic limit set by the need to have the energy in the zeros and ones inside the computer be sufficiently greater than thermal noise to prevent spontaneous errors. The problem is that the asymptotic level of curve 1 is below the level of the most ambitious applications.

Curve ② (dashed because there is no systematic effort to realize it) shows possible improvements through computer architecture, achievement of which I will deem a "medium challenge." For a fixed device logic technology, it is possible to improve the efficiency of a computer through "architectural" changes. In essence, these changes come through the rearrangement of data paths and functional units so that more FLOPS directly attributable to the algorithm occur for each logic or memory operation not countable as useful work. However, efficiency can never exceed 100%. A solidly packed array of floating point units (e. g. a special purpose hardware device such as a radar signal processor) is about 2 orders of magnitude more efficient than a microprocessor, as illustrated in curve ②. This type of architecture improvement is

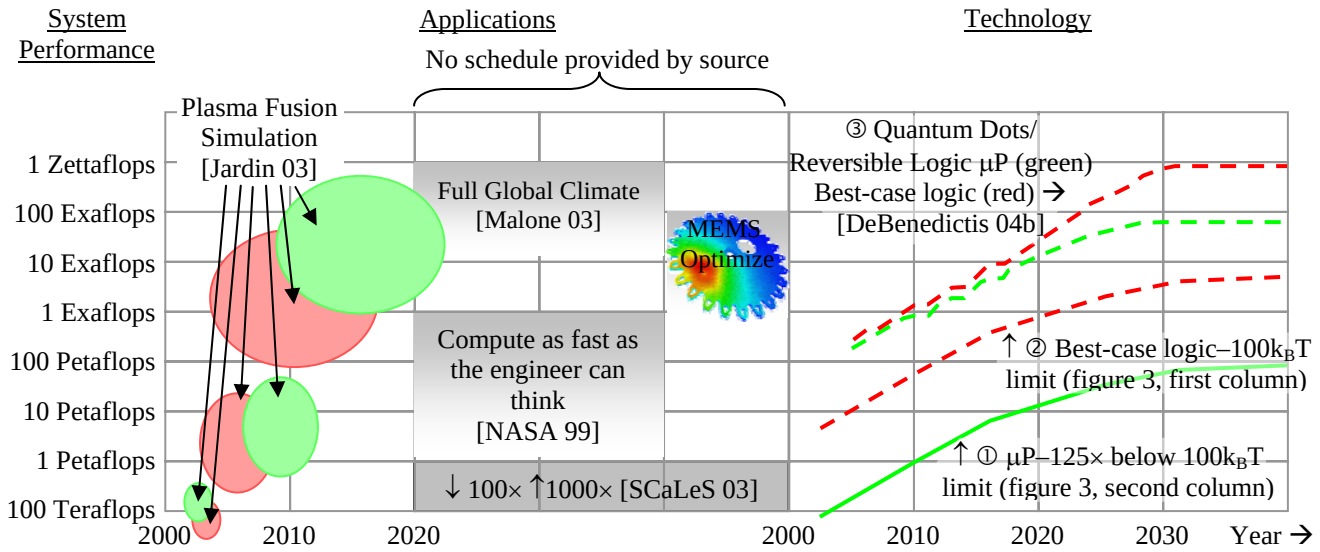

Figure. Supercomputer applications and technology projections. For references in figure, see DeBenedictis, E. P. 2004. Will Moore's Law Be Sufficient? In *Proceedings of the 2004 ACM/IEEE Conference on Supercomputing* (November 06 - 12, 2004). Conference on High Performance Networking and Computing. IEEE Computer Society, Washington, DC, 45.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

This document has been approved for Unclassified Unlimited release as SAND2005-4794C

important, yet the asymptote of curve ② is also below the performance of the most ambitious applications.

To reach above curve ② will involve Grand Challenges in both device physics and computer architecture. The thermodynamic power limit defining curves ① and ② was identified by Landauer in 1961 as applying to "irreversible logic," or logic where a thermodynamically irreversible destruction of information occurs after each gate operation. Since in the ubiquitous AND-OR-NOT digital logic, AND and OR gates destroy information as part of each gate operation (i. e. it is not possible to recover the inputs of an AND or OR gate from knowledge of its output), this power limit applies to nearly every digital logic system in existence today as well as the design tools and skills of designers for creating new systems. However, there has been continuous progress (since 1961) in ways to organize devices (including transistors) into reversible gates that do not destroy information. Reversible gates include gate sets that are logically complete (e. g. Toffoli gates). Since the physical size and operating cost of a computer is more or less defined by its heat and power characteristics, reducing information destruction can permit more processing per unit cost. Thus, the recipe for exceeding curve ② is to construct devices at the performance level of "end of the roadmap" transistors (a Grand Challenge) and then organizing them as a "reversible logic" supercomputer (another Grand Challenge).

Based on my review of the field, the architectural Grand Research Challenge will require addressing a number of substantial sub-issues:

1. Computer architects must live with the devices they are given. Device physicists must find devices where parasitic losses (heat generation) are very low compared to the energy of ubiquitous thermal motion. Finding these devices will be difficult, and the ones that are found will probably <u>not</u> have the same speed of operation, ratio of memory-to-logic-to-communications performance, reliability, manufacturability, etc. as current CMOS.
2. Physics guarantees that AND-OR-NOT logic can never exceed curve ②. While there are provably sufficient methods that can exceed curve ②, these methods are unquestionably more burdensome and not widely known.
3. New computer architectures must be invented to capitalize on thermodynamic reversibility. Reversible microprocessor architectures have been demonstrated in several Ph. D. theses, showing that this task is possible. However, these theses show the architectures have much room for improvement. Also, reversible microprocessors are likely to be several orders of magnitude lower in performance than reversible "best case logic" – for the same reason curve ① is below curve ② in the figure.
4. Most algorithms will need rethinking due to an increase in parallelism mandated by physics. Even today, "Moore's Law" is not raising the speed of individual threads very much. While flat lining of speed is due to both fundamental and engineering issues (e. g. the speed of light and diffusion delay) there is little doubt that further increases will be hard to come by. This implies that future performance increases will come mostly through more parallelism. However, any specific algorithm has only so much parallelism available to be exploited. To get further performance increases more parallelism will need to be exploited. When the parallelism in algorithms is exhausted, more highly parallel algorithms will be needed if progress is to continue.

Curve ③ in the figure (from the published paper, as noted in the caption) represents a plausibility demonstration that resolutions to the issues above might permit addressing the grand challenges of science. I gathered published information about proposed research devices (Quantum Dots, and others), logic design methods, reversible architectures, and algorithms and applied them to the problem of simulating of the global climate for the purpose of understanding and mitigating global warming. While there are major uncertainties in my input data, combining them mathematically yielded curve ③, showing about enough computing power to solve the most ambitious problems.

I offer two notes before concluding: First, the discussion above has been framed in terms of a $100M supercomputer. Similar arguments for less pricey, lower power, and smaller systems (e. g. mobile robots) would lead to similar research challenges. Second, it is popularly believed that coherent quantum computing will raise computing performance. This may be true for some problems, yet coherent quantum computing is only narrowly applicable and the challenge described here will remain.

Biography

Erik DeBenedictis is a member of technical staff at Sandia National Laboratories.

Erik has a Ph. D. degree from Caltech (1983) where he designed and constructed the Cosmic Cube computer (which is the architectural basis for many supercomputers today). Erik worked at Bell Laboratories, Holmdel NJ (1983-1989), where he won a Gordon Bell award during 1988 (its first year). Erik also worked at Ansoft (a finite element software company) and nCUBE (a parallel supercomputer supplier). Erik founded and ran a wireless Internet company named NetAlive.

While at Sandia, Erik has led and participated in efforts in batch scheduling for supercomputers, the Red Storm supercomputer, quantum computing, and advanced supercomputer architectures. Erik is a member of the International Technology Roadmap for Semiconductors (ITRS) Emerging Research Device (ERD) panel.