LA-UR- 99-820

*Approved for public release;
distribution is unlimited.*

Title: ASCI Application Performance and the Impact of Commodity Processor Architectural Trends

Author(s): Olaf M. Lubeck, CIC-19
Adolfy Hoisie, CIC-19
Federico Bassetti, CIC-19
Kirk Cameron, CIC-19
Yong Luo, CIC-19
Harvey J. Wasserman

Submitted to: Proceeding of IEEE Workshop on Innovative Architecture

Los Alamos
NATIONAL LABORATORY

ASCI Application Performance and the Impact of Commodity Processor Architectural Trends[1]

Olaf Lubeck, Adolfy Hoisie, Federico Bassetti,
Kirk Cameron, Yong Luo, and Harvey Wasserman

Scientific Computing Group
Los Alamos National Laboratory
Los Alamos, New Mexico, U.S.A.

Abstract.

The purpose of this paper is to summarize recent performance results from an important ASCI-related application and to speculate on how trends within the computer industry and in computer architecture relate to these results.

Introduction. The DOE Accelerated Strategic Computing Initiative (ASCI) is an applications-driven program requiring use of scalable, high-performance architectures to meet aggressive engineering needs related to safety of the Nation's nuclear stockpile [1]. ASCI will accelerate development of computational methods and tools for predictive simulation and for virtual prototyping needed to re-certify the existing stockpile, assess the effects of component aging, and to evaluate accident scenarios.

There are a multitude of computational issues confronting ASCI, including software validation and verification, portability, and the use of object-oriented frameworks; network and security issues associated with distance computing; and programming environment issues related to designing and debugging codes on tens of thousands of processors [2].

Sustained computational performance is no less of a key issue. Individual programs within ASCI have well-defined simulation runtime goals and there are immediate needs for 100-fold improvements in throughput. Computationally, ASCI's safety prediction capability is considerably more challenging than that of weapon design. While current calculations use on the order of a million cells, a canonical goal of ASCI is to do 3-D billion-cell problems. Adaptive mesh refinement and unstructured grids will become increasingly important, implying both irregular memory access and non-uniform inter-processor communication patterns. Such methods coupled with current discretization schemes also imply relatively low ratios of floating-point operations to memory operations, typically less than one FLOP per memory reference, on average.

ASCI Performance Modeling Research. We are addressing performance issues by implementing a comprehensive program to characterize ASCI algorithms and application codes. The purpose of this work is to gauge the performance progress of ASCI codes, develop scaleable strategies for code development, reveal scalability issues in hardware and software, allow for technical planning for future ASCI architectures in the context of a five-year estimate of COTS technologies, and engage industry and university partners in ASCI application-driven performance problems. An additional goal is development and dissemination of a suite of "Compact Applications" in which critical ASCI performance problems are embodied in open, compact codes.

Scalability analysis must address two key questions: (1) What single-processor computational efficiency can we expect on future machines; and (2) What parallel efficiency can we expect? Insight into both of these is gained through development of models that incorporate key characteristics of both the applications and the architectures.

[1] This report was sponsored by United States Department of Energy under contract W-7405-ENG-36 to the Regents of the University of California.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ASCI Applications. Although many ASCI codes will simulate the same physics and chemistry effects, current code development projects at Los Alamos are partitioned into approaches according to mesh strategies, discretization schemes, and programming styles. For example, the CRESTONE project is oriented towards Eulerian Hydrodynamics using structured Cartesian grids with cell-by-cell AMR, with vectorizable Fortran77. In contrast, the BLANCA project uses a structured, Arbitrary Eulerian-Lagrangian (ALE) code written with an object-oriented C++ framework that separates the physics and mesh manipulation from the parallel programming implementation. Other projects use arbitrarily-connected, unstructured meshes and Fortran90.

Particle transport via both Monte Carlo and deterministic methods is an important component of the ASCI workload, accounting for upwards of 50-80% of simulation time on current DOE systems. As such, there has been a great deal of research devoted to improving the performance of codes that carry out this kind of simulation [3]. In recent years there have been reports of neutral particle transport simulations on parallel computer systems such as the CRAY T3D, nCUBE, and Thinking Machines, Inc. CM-2 and CM-5 [4-7]. This is in addition to the pioneering work done on vector supercomputers many years ago [8].

Modeling Scalability. We recently developed a performance model for algorithms consisting of multiple wavefronts partitioned and pipelined on multidimensional processor grids [9]. We applied this model to Cartesian-coordinate, deterministic particle transport, as abstracted in the ASCI Compact Application "SWEEP3D" [10]. The algorithm in SWEEP3D is inherently recursive and so wavefront processes are typically used to enable parallelism. In a 2-D MIMD domain-decomposition using a message-passing model, wavefront-like "sweeps" through the processor grid are generated, with parallel efficiency improved by logically stacking additional work from the third dimension and other (non-spatial) discretized variables. Overlap of communication and computation occurs at some (but not all) steps in the simulation, and message passing imposes additional constraints that tend to limit parallelism in the algorithm.

An important use of our model was to understand particle transport scalability as a function of per-processor sustained speed, and MPI latency and bandwidth on a future-generation system - a hypothetical, mesh-topology (i.e., non-clustered) 100-TFLOPS-peak machine with 20,000 processors that might be in existence around 2004. We considered both conservative and optimistic changes in CPU and network technology. Interestingly, the model showed that on a one billion-cell problem, this application is *compute* bound; i.e., inter-processor communication is not the primary bottleneck (although communication does become important for smaller problem sizes).

Modeling Single-CPU Memory Performance. It was interesting for us to learn that single-processor performance is the dominant factor for SWEEP3D, because we had also been studying what factors limit single-processor performance in a variety of applications. Many codes of which we are aware achieve only 5-10% of peak performance on typical RISC microprocessors, in terms of either MFLOPS or CPI [11-13].

Many recent studies have attempted to identify the microprocessor architectural features that lead to diminished performance relative to peak [12, 14]. Memory performance consistently stands out as a critical bottleneck in all these studies. Our own studies, in which we use a simplified empirical parameterization along with data from hardware event counters to obtain memory stall time [15], showed that on single processors of the MIPS R10000 memory stall time for SWEEP3D accounts for about 45% of total CPI. This means that if one could optimize the code so as to eliminate all memory stall time, the improvement would be about a factor of two in execution time, which would correspond to about 80 MFLOPS per processor, or about 20% of peak [11].

Applying this result to the scalability study of SWEEP3D (above) leads to the conclusion that machines constructed of microprocessors reasonably expected to be in existence within the next few years may be unable to satisfy ASCI performance goals. For example, our wavefront scalability model predicts that in order to run a billion-cell SWEEP3D problem in 60 hours, we would require 2,500 MFLOPS sustained per-node performance. This implies either considerably larger than 10% sustained performance relative to peak or what is probably an impossibly-high peak rate.

Furthermore, known trends in CPU speed vis-a-vis memory speed suggest that in the future, memory performance may play an even larger role than it does now, further limiting the achieved performance relative to peak [16].

More recent work in our group is oriented towards understanding processor performance in the absence of memory effects. This work shows that processor inefficiency in SWEEP3D is also probably due to a mismatch between the instruction mix in SWEEP3D and the micro architectural characteristics of the MIPS R10000, such as its allocation of functional units [17].

COTS Technology. A key ASCI strategy is to use commodity off-the-shelf (COTS) technologies to compose larger systems in an attempt reduce costs and improve price/performance ratios over traditional supercomputers. On the one hand, a potential problem with this is that scientific computation comprises only a small portion of desktop and server workloads and is therefore not considered to be an important driver for RISC microprocessor architecture. On the other hand, the needs of scientific and commercial workloads are not entirely orthogonal, since recent performance studies have shown that memory performance of commercial workloads is relatively *worse* than that of scientific workloads [18-22].

Two factors have led several prominent researchers to question whether superscalar processors will prevail as the microprocessors with the greatest commercial impact [23-25]. The first is a combination of the processing rate inefficiency often observed in today's microprocessors coupled with the likely *additional* pipeline-stall, instruction fetch, and cache hit rate affects brought about by increasing memory latency. The second is the extent to which processor size and power requirements will limit the applicability of superscalar processors to multimedia-based workloads that are emerging as the dominant application regime of the future. Many believe that these new workloads, which will result from the huge consumer market need for video, sound, speech, graphics, telephony, and network processing, will cause drastic change in the architecture of commodity systems [23, 25, 26].

Thus, an important question is what impact this architectural shift will have on ASCI. In particular, we wonder about the extent to which new features implemented to support media applications might still be able to support (or actually enhance) numerical simulation. Foremost among these features is SIMD processing. Several recent studies [27, 28] have shown that many important media processing kernels are highly vectorizable. In an effort to better support this workload shift, all major microprocessor manufacturers now have introduced short, vector-like extensions to their instruction sets [29]. These extensions have limited capability and usually operate only on narrow data types common to media applications. Recent studies have demonstrated quantitatively that a more traditional, long-vector architecture is considerably faster on some media applications than the short vector extensions [30]. An important conclusion reached in these studies is that microprocessors consisting of a superscalar core tightly coupled to a CMOS-based, multipipeline vector unit can provide a scalable, cost-effective solution for desktop computing [31]. Such an architecture might well help preserve any investment in superscalar-optimized code [32] and at the same time afford significant benefit to those applications that remain vectorizable.

Efforts to more fully understand what alignment may exist between media-based and numerical workloads are underway in our group. Subsequent publications will compare these workloads at both the algorithmic level as well as in terms of instruction level parallelism.

References

1. A. Larzelere, "Creating Simulation Capabilities," IEEE Computational Science & Engineering, IEEE Computer Society, Vol. 5, No. 1, January/March 1998, pp 27-35.

2. D. Clark, "ASCI Pathforward: 30 TFLOPS and Beyond," IEEE Concurrency, April/June 1988, pp 13-15.

3. See, for example, M. R. Dorr and C. H. Still, "Concurrent Source Iteration in the Solution of Three-Dimensional Multigroup Discrete Ordinates Neutron Transport Equations," Technical Report UCRL-JC-116694, Rev 1, Lawrence Livermore National Laboratory, Livermore, CA, May, 1995.

4. K. R. Koch, R. S. Baker and R. E. Alcouffe, "Solution of the First-Order Form of the 3-D Discrete Ordinates Equation on a Massively Parallel Processor," Trans. of the Amer. Nuc. Soc., 65, 198, 1992.

5. R. S. Baker and R. E. Alcouffe, "Parallel 3-D S_N Performance for DANTSYS/MPI on the CRAY T3D, Proc. of the Joint Intl'l Conf. On Mathematical Methods and Supercomputing for Nuclear Applications, Vol 1. page 377, 1997.

6. R. S. Baker, C. Asano, and D. N. Shirley, "Implementation of the First-Order Form of the 3-D Discrete Ordinates Equations on a T3D, Technical Report LA-UR-95-1925, Los Alamos National Laboratory, Los Alamos, NM, 1995; 1995 American

Nuclear Society Meeting, San Francisco, CA, 10/29-11/2/95.

7. R. S. Baker, K. R. Koch, "An Sn Algorithm for the Massively Parallel CM200 Computer", Nucl. Sci. and Eng., Vol 128, No. 3, p. 312, 1998.

8. R. E. Alcouffe, ``Diffusion Acceleration Methods for the Diamond-Difference Discrete-Ordinates Equations," Nucl. Sci. Eng.{64}, 344 (1977).

9. A. Hoisie, O. L. Lubeck, and H. J. Wasserman, "Performance and Scalability of Multi-Dimensional Wavefront Algorithms With Application to Particle Transport," Accepted for Publication, *Proceedings of Frontiers in Massively Parallel Computing*, IEEE Computer Society, February, 1999.

10. SWEEP3D is available on http://www.c3.lanl.gov/cic19/teams/par_arch/Codes.html.

11. H. Wasserman, O. M. Lubeck, Y. Luo, and F. Bassetti, "Performance Evaluation of the SGI Origin2000: A Memory-Centric Characterization of LANL ASCI Applications," Proc. Supercomputing '97, IEEE Computer Society.

12. Bhandarkar, D. and Cvetanovic, Z., "Performance Characterization of the Alpha 21164 Microprocessor Using TP and SPEC Workloads," Proc. Second. Int. Symp. on High-Perf.. Comp. Arch., IEEE Computer Society Press, Los Alamitos Ca., 1996.

13. H. J. Wasserman, "Benchmark Tests on the New IBM RISC System/6000 590 Workstation," Scientific Programming, Vol. 4, No. 1, Spring 1995, pp 23-34.

14. Bhandarkar, D. and Ding, J., "Performance Characterization of the Pentium Pro Processor, " Proc. Third. Int. Sypm. on High-Perf. Comp. Arch., IEEE Computer Society Press, Los Alamitos Ca., pp 288-297, 1997.

15. O. M. Lubeck, Y. Luo, H. J. Wasserman, and F. Bassetti, "Development and Validation of a Hierarchical Memory Model Incorporating CPU- and Memory-Operation Overlap," *Contributed presentation and Proc. of PDPTA*, July, 1998.

16. Wulf, W. A. and McKee, S. A. "Hitting the Memory Wall: Implications of the Obvious," Comp. Arch. News, Assoc. for Computing Mach., March, 1995.

17. Cameron, K. W., Luo, Y., and Schwarzmeier, J. , Instruction-level Microprocessor Modeling of Scientific Applications, accepted by International Symposium on High Performance Computing'99, June, 1999, Kyoto, Japan.

18. Keeton, K. and Pattersonm, Y. Q. He, R. C. Raphael, and W. E. Baker. "Performance Characterization of a Quad Pentium Pro SMP Using OLTP Workloads," Proceedings of the 25th International Symposium on Computer Architecture, Barcelona, Spain, June 1998. Extended version published as UC Berkeley Computer Science Division Technical Report UCB//CSD-98-1001, April 1998.

19. Q. Cao, P. Trancos, and J. Torrellas, *"Characterizing TPC-D on a MIPS R10K Architecture,"* Proc. 1st Workshop on Computer Architecture Evaluation using Commercial Workloads, February, 1998.

20. J. H. Moreno, M. Moudgill, J. D. Wellman, P. Bose, and L. Trevillyan, *"Trace-driven Performance Exploration of a PowerPC 601 OLTP Workload on Wide Superscalar Processors,"* Proc. 1st Workshop on Computer Architecture Evaluation using Commercial Workloads, February, 1998.

21. Z. Cvetanovic, *"Analysis of Commercial and Technical Workloads on AlphaServer Platforms,"* Proc. 1st Workshop on Computer Architecture Evaluation Using Commercial Workloads, February, 1998.

22. P. Trancoso, J-L. Larriba-Pey, Z. Zhang, and J. Torrellas, *"The Memory Performance of DSS Commercial Workloads in Shared-Memory Multiprocessors,"* Proc. 3rd Ann. IEEE Intl.Symp. High-Perf. Comp. Arch., Feb, 1997.

23. C. E. Kozyrakis and D. A. Patterson, "A New Direction for Computer Architecture Research," IEEE Computer, vol. 31, no. 11, November 1998 (p. 24-32).

24. C. Kozyrakis, S. Perissakis, D. Patterson et. al.: " Scalable Processors for the Billion Transistors Era: IRAM", IEEE Computer, vol. 30, no. 9, September 1997.

25. W. J. Dally, " Tomorrow's Computing Engines," presentation at Fourth Annual IEEE Symposium on High-Performance Computing Architecture, February 3, 1998.

26. K. Diefendorff and P. Dubey, "How Multimedia Workloads Will Change Processor Design," IEEE Computer, September, 1997, pp 43-45.

27. K. Asanovic, "Vector Microprocessors," Ph.D. Dissertation, University of California, Berkeley, Spring, 1988, pp 193-229.

28. C.G. Lee and M.G. Stoodley. Simple Vector Microprocessors for Multimedia Applications. Accepted for publication in the 31st Annual International Symposium on Microarchitecture. Dec 1998; available on http://www.eecg.toronto.edu/~stoodla/publications.html.

29. C. Lee, "Short Vector Extensions in Commercial Microprocessors," http://www.cs.berkeley.edu/~corinna/svx.

30. M. Stoodley and C. Lee, "Vector Microprocessors for Desktop Computing," submitted to 26th Annual International Symposium on Computer Architecture, 1998; available on http://www.eecg.toronto.edu/~stoodla/publications.html.

31. C. G. Lee and D. J. DeVries, "Initial Results on the Performance and Cost of Vector Microprocessors," Proc. 30th Ann. Intl. Symp. on Microarchitecture, pp171-182, December, 1997.

32. G. H. Golub and C. F. VanLoan, "Matrix Computations," Third Edition, John Hopkins University Press, 1997.