Final Report for the Castle Project
1 May 1994 - 30 April 1998

www.cs.berkeley.edu/projects/parallel/castle/

Tom Anderson, David Culler, James Demmel, Jerry Feldman, Susan Graham, Paul Hilfinger, Katherine Yelick

Computer Science Division
University of California, Berkeley

16 February 2000

The goal of the Castle project was to provide a parallel programming environment that enables the construction of high performance applications that run portably across many platforms. Our approach was to design and implement a multilayered architecture, with higher levels building on lower ones to ensure portability, but with care taken not to introduce abstractions that sacrifice performance. Our layers include the following (listed from bottom to top, with highlights):

1. Interprocessor communications. This includes work on Active Messages and MPI. The work on Active Messages was part of the basis of the new industry communication standard VIA, adopted by Intel and other computer companies.

2. System support. This includes threads, scheduling, load balancing, and I/O support. This work led to the world's record disk-to-disk sort in 1997 (`now.cs.berkeley.edu/NowSort/index.html`).

3. Libraries. This includes work on ScaLAPACK, SuperLU, Multipol, PHIPAC, and studies of the impact of heterogeneous computing. ScaLAPACK has been widely adopted by many computer companies as part of their standard parallel library, and is part of the ASCI Red standard system build.

4. Languages. This includes Split-C, pSather and Titanium.

5. Applications. This includes a large number of studies, including climate modeling, circuit simulation, phylogeny tree computation, cell simulation, semiconductor modeling, computational fluid dynamics, connected components, sorting, connectionist networks, computing Gröbner bases, and web searching.

   The symmetric eigensolver in ScaLAPACK was incorporated in a material science code at SNL, which was runner up in the 1998 Gordon Bell Competition at Supercomputing.

   The distributed memory version of the SuperLU sparse direct linear system solver was used to solve an open problem in quantum chemistry at NERSC, resulting in the cover of the 24 Dec 1999 issue of Science.

DISCLAIMER

The web searching application, initially implemented on the NOW using Active Messages, was made into a commercial product called Inktomi which was worth over $1B after going public in 1998.

In addition, we made the following contributions:

1. Performance Modeling and Tuning. In addition to performance modeling in the context of the above applications, various studies were performed to identify bottlenecks on different kinds of parallel architectures.

2. Education. Two courses were developed. "Applications of Parallel Computing" has extensive on-line notes and is taught regularly at Berkeley, with the notes used at several other universities and for short courses. "Parallel Computer Architecture" is also taught at Berkeley, and resulted in a published textbook.

3. Collaborations. We collaborate extensively with NERSC and other groups at Lawrence Berkeley National Lab on various issues related to high performance computing, including linear algebra libraries, and language support for computational fluid dynamics. Some of our software (ScaLAPACK and SuperLU) is also appearing in the ACTS Toolkit, funded by DOE 2000. This software as well as the Titanium language is part of the NSF NPACI Parallel Tools and Environments effort.

   The NOW was part of the NSF NPACI SuperComputer Center, one of the supercomputers NSF users could elect to use.

We also received a $6M grant from Intel Corporation, as well as other large grants from SUN Microsystems, Microsoft, and the National Science Foundation to build a parallel computing infrastructure for 18 departments across the Berkeley campus. This system, called Millennium (www.millennium.berkeley.edu) draws heavily on our experience in Castle.

We divide our report into several sections which lists the publications for the above topics.

1 Interprocessor Communications

1. S. Lumetta, D. Culler, "Managing Concurrent Access for Shared Memory Active Messages." IPPS/SPDP 98, Orlando, Fl. March 1998

2. S. Lumetta, A. Mainwaring, D. Culler, "Multi-Protocol Active Messages on a Cluster of SMP's." Supercomputing '97, San Jose, CA, November 1997

3. B. Chun, A. Mainwaring, S. Schleimer, D. Wilkerson, "System Area Network Mapping." SPAA '97, Newport, RI, June 1997

4. S. Rodrigues, T. Anderson, D. Culler, "High-Performance Local Area Communication with Fast Sockets." USENIX '97, 1997

5. A. Mainwaring, D. Culler, "Active Message Applications Programming Interface and Communication Subsystem Organization." UC Berkeley CS Division Report CSD-96-918, 1995

6. E. Brewer, F. Chong, L. Liu, J. Kubiatowicz, S. Sharma, "Remote Queues: Exposing Message Queues for Optimization and Atomicity." Proceedings of SPAA, 1995

7. R. Martin, "HPAM: An Active Message Layer for a Network of HP Workstations." Hot Interconnects, 1994

8. L. Liu, ""An evaluation of the Intel Paragon Communication Architecture." UC Berkeley CS Dept., 1995 (MS Thesis)

2 System Support

1. S. Lumetta, D. Culler, "Mantis User's Guide Version 1.0." University of California at Berkeley CS Division, 1994 Tech Report (UCB/CSD-94-828)

2. S. Lumetta, "Mantis: A Debugger for the Split-C Language." University of California at Berkeley, Computer Science Division, 1994 (MS Thesis)

3. S. Luna, "Implementing an Efficient Portable Global Memory Layer on Distributed Memory Multiprocessors." University of California at Berkeley Computer Science Division, 1994 MS Thesis

4. T. Anderson, D. Culler, D. Patterson, NOW Team, "A Case for Networks of Workstations: NOW." IEEE Micro, Feb. 1995

5. D. Culler, A. Arpaci-Dusseau, R. Arpaci-Dusseau, B. Chun, S. Lumetta, A. Mainwaring, R. Martin, C. Yoshikawa, F. Wong, "Parallel Computing on the Berkeley NOW." JSPP '97 (9th Joint Symposium on Parallel Processing)

6. D. Patterson, K. Asanovic, A. Brown, R. Fromm, J. Golbus, B. Gribstad, K. Keeton, C. Kozyrakis, D. Martin, S. Perissakis, R. Thomas, N. Treuhaft, and K. Yelick. "Intelligent RAM (IRAM): the Industrial Setting, Applications, and Architecture." '97 Interntaional Conference on Compute Design, October 1997.

7. S. Perissakis, C. Kozyrakis, T. Anderson, K. Asanovic, N. Cardwell, R. Fromm, K. Keeton, D. Patterson, R. Thomas, and K. Yelick. "Scaling Processors to 1 Billion Transistors and Beyond: IRAM." IEEE Computer (Special Issue), September 1997.

8. D. Patterson, T. Anderson, N. Cardwell, R. Fromm, K. Keeton, C. Kozyrakis, R. Thomas, and K. Yelick. "A Case for Intelligent DRAM: IRAM." IEEE Micro, April 1997, pp. 34–44. Also appeared as an Award Paper, Hot Chips VIII , August 1996.

9. D. Patterson, T. Anderson, N. Cardwell, R. Fromm, K. Keeton, C. Kozyrakis, R. Thomas, and K. Yelick. "Intelligent RAM (IRAM): Chips that remember and compute." Proceedings of the 1997 IEEE International Solid-State Circuits Conference, February 1997, pp. 224–225

10. A. Krishnamurthy, K. E. Schauser, C. J. Scheiman, R. Y. Wang, D. E. Culler, and K. Yelick, "Evaluation of Architectural Support for Global Address-Based Communication in Large-Scale Parallel Machines." Proceedings of Architecture Support on Programming Languages and Operating Systems, 1996.

11. A. Krishnamurthy and K. Yelick, "Analyses and Optimizations for Shared Address Space Programs." Journal of Parallel and Distributed Computation, 1996.

12. A. Krishnamurthy and K. Yelick, "Optimizing Parallel Programs with Explicit Synchronization." Proceedings of the Sigplan Conference on Programming Language Design and Implementation, San Diego, California, June 1995.

13. R. Arpaci, D. Culler, A. Krishnamurthy, S. Steinberg, and K. Yelick, "Empirical Evaluation of the CRAY-T3D: A Compiler Perspective," International Symposium on Computer Architecture, Santa Margherita Ligure, Italy, June 1995.

14. C.-P. Wen, S. Chakrabarti, E. Deprit, A. Krishnamurthy and K. Yelick, "Runtime Support for Portable Distributed Data Structures," Languages, Compilers, and Runtime Systems for Scalable Computers, Boleslaw K. Szymanski and Balaram Sinharoy (Editors), Kluwer Academic Publishers, Boston, MA, 1995, pp. 111–120.

15. A. Krishnamurthy, D. Culler, and K. Yelick, "Empirical Evaluation of Global Memory Support on the Cray-T3D and Cray-T3E," UCB//CSD-98-991, 1998.

16. S. Chakrabarti, E. Deprit, J. Jones, A. Krishnamurthy, E.-J. Im, C.-P. Wen, and K. Yelick, "Multipol: A Distributed Data Structure Library." UCB//CSD-95-879, July 1995.

17. S. Chakrabarti, "Efficient Resource Scheduling in Multiprocessors." Ph.D. dissertation, University of California, Berkeley, 1996.

18. C.-P. Wen, "Portable Library Support for Irregular Applications." Ph.D. dissertation, University of California, Berkeley, 1995.

3 Libraries

1. "Using the Matrix Sign Function to Compute Invariant Subspaces," Z. Bai, J. Demmel, SIAM J. Mat. Anal. Appl., v. 19, n. 1, January 1998 (also appeared as "Design of a Parallel Nonsymmetric Eigenroutine Toolbox, Part II," with Z. Bai, Department of Mathematics Research Report 95-11, U. Kentucky, 1995. Submitted to SIMAX, Jan. 1996.)

2. "A Supernodal Approach to Sparse Partial Pivoting", J. Demmel, S. Eisenstat, J. Gilbert, X. Li and J. W. H. Liu, SIAM J. Mat. Anal. Appl., v. 20, n. 3, pp 720-755, July 1999 UC Berkeley CS Division Report UCB//CSD-95-883, Sept 1995

3. "Practical Experience in the Numerical Dangers of Heterogeneous Computing," L.S. Blackford, A. Cleary, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, A. Petitet, H. Ren, K. Stanley and R.C. Whaley, ACM Trans. Math. Soft., v. 23, n. 2, pp. 133-147, June 1997; also appeared as "Practical Experience in the Dangers of Heterogeneous Computing", chapter in Applied Parallel Computing, Industrial Computation and Optimization, Third International Workshop, PARA '96, Eds. J. Waśniewski, J. Dongarra and K. Madsen, Lecture Notes in Computer Science, Vol. 1184, Springer-Verlag, 1996.

4. "An Asynchronous Parallel Supernodal Algorithm for Sparse Gaussian Elimination," J. Demmel, J. Gilbert and X. Li. SIAM J. Mat. Anal. Appl., v. 20, n. 4, pp 915-952, October 1999 (submitted 1997)

5. "The Spectral Decomposition of Nonsymmetric Matrices on Distributed Memory Parallel Computers", Z. Bai, J. Demmel, J. Dongarra, A. Petitet, H. Robinson, and K. Stanley, SIAM J. Sci. Comp. , v. 18, n. 5, Sept 1997

6. "Execution Time of Symmetric Eigensolvers", K. Stanley, UC Berkeley PhD Thesis, 1997 (Computer Science Division Technical Report CSD-99-1039)

7. "Application of a High Performance Parallel Eigensolver to Electronic Structure Calculation", G. Henry, K. Stanley, M. Sears, Supercomputing 1998 (Gordon Bell Prize runner-up)

8. J. Bilmes, K. Asanovic, J. Demmel, D. Lam, C.W. Chin, "PHiPAC: A Portable, High-Performance, ANSI C Coding Methodology and its application to Matrix Multiply," University of Tennessee Technical Report, LAPACK Working Note 111, 1996

9. J. Bilmes, K. Asanovic, C.-W. Chin, J. Demmel, "Using PHiPAC to speed Error Back-Propagation Learning," Proceedings of ICASSP 97, Munich, April 1997

10. J. Bilmes, K. Asanović, C.-W. Chin, J. Demmel, "The PHiPAC v1.0 Matrix-Multiply Distribution," International Computer Science Institute, Techical Report 35, 1998

11. J. Bilmes, K. Asanovic, C.-W. Chin, J. Demmel, "The PHiPAC v1.0 Matrix-Multiply Distribution", Department of EECS, CS Division, University of California at Berkeley, Tech Report 1020, 1998

4 Languages

1. K. Yelick, L. Semenzato, G. Pike, C. Miyamoto, B. Liblit, A. Krishnamurthy, P. Hilfinger, S. Graham, D. Gay, P. Colella, and A. Aiken. "Titanium: A High-Performance Java Dialect ," ACM 1998 Workshop on Java for High-Performance Network Computing. To appear in Concurrency: Practice and Experience.

2. D. Stoutamire, W. Zimmermann, M. Trapp, "An Analysis of the Divergence of Two Sather Dialects", International Computer Science Institute Technical Report TR-96-037, 1996

3. D. Stoutamire, S. Omohundro, "The Sather 1.1 Specification," International Computer Science Institute Technical Report TR-96-012, August 1996

4. D. Stoutamire, M. Kennel "Sather revisited: A high performance free alternative to C++", Computers in Physics, Sept/Oct 1995.

5. D. Stoutamire, "Portable, Modular Expression of Locality" PhD Thesis, UC Berkeley, Computer Science Division, 1997

5 Applications

1. A. Dusseau, "Modeling Parallel Sorts with LogP on the CM-5." University of California at Berkeley Computer Science Division Tech Report, 1994 (UCB/CSD-94-829)

2. A. Arpaci-Dousseau, R. Arpaci-Dousseau, Culler, Hellerstein, Patterson, "Searching for the Sorting Record: Experiences in Tuning the NOW-Sort." The 1998 Symposium on Parallel and Distributed Tools (SPDT '98), Welches, Oregon, August 3-4, 1998.

3. A. Arpaci-Dousseau, R. Arpaci-Dousseau, Culler, Hellerstein, Patterson, "High-Performance Sorting on Networks of Workstations." SIGMOD '97, Tuscon, AZ, May 1997

4. S. Steinberg, J. Yang and K. Yelick, "Performance Modeling and Composition: A Case Study in Cell Simulation." International Parallel Processing Symposium, April 1996.

5. J. Jones and K. Yelick, "Parallelizing the Phylogeny Problem." Supercomputing '95, San Diego, California, December 1995.

6. C.-P. Wen and K. Yelick, "Portable Runtime Support for Asynchronous Simulation." Proceedings of the International Conference on Parallel Processing, Oconomowoc, Wisconsin, August 1995.

7. S. Chakrabarti and K. Yelick, "Distributed Data Structures and Algorithms for Gröbner Basis Computation," Lisp and Symbolic Computation, Volume 7, 1994, Pages 147–172.

8. A. Krishnamurthy, S. Lumetta, D. Culler, and K. Yelick "Connected Components on Distributed Memory Machines," Parallel Algorithms, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, 1997. (Also the proceedings of the 3rd DIMACS Parallel Implementation Challenge Workshop, October 1994.)

9. K. Yelick, C.-P. Wen, S. Chakrabarti, E. Deprit, J. Jones, A. Krishnamurthy, "Portable Parallel Irregular Applications," Workshop on Parallel Symbolic Languages and Systems, Beaune, France, October 1995. To appear in Lecture Notes in Computer Science.

10. S. Steinberg, "Parallelizing a Cell Simulation: Analysis, Abstraction, and Portability." (M.A. thesis) UCB/CSD-95-878, December, 1994.

6 Performance Modeling and Tuning

1. R. Martin, A. Vahdat, D. Culler, T. Anderson, "Effects of Communication Latency, Overhead, and Bandwidth, in a Cluster Architecture." ISCA 24, Denver, Co, June 1997.

2. D. Culler, L. Liu, R. Martin, C. Yoshikawa, "LogP Performance Assessment of Fast Network Interfaces." IEEE Micro, 1996.

3. D. Keeton, T. Anderson, D. Patterson, "LogP Quantified: The Case for Low-Overhead Local Area Networks." Hot Interconnects III: A Symposium on High Performance Interconnects, 1995.

4. L. Liu, D. Culler, "Evaluation of the Intel Paragon on Active Message Communication" Proceedings of Intel Supercomputer Users Group Conference, 1995.

5. L. Liu, D. Culler, "Measurement of Active Message Performance on the CM-5." UC Berkeley Tech. Report, 1994 (CSD-94-807)

6. R. Arpaci-Dousseau, A. Arpaci-Dousseau, D. Culler, J. Hellerstein, D. Patterson, "The Architectural Costs of Streaming I/O: A Comparison of Workstations, Clusters, and SMPs." HPCA 4, Las Vegas, NV, February 1998

7. S. Chakrabarti, J. Demmel, and K. Yelick, "Modeling the Benefits of Mixed Data and Task Parallelism," Proceedings of the Symposium on Parallel Algorithms and Architectures, Santa Barbara, California, July 1995.

8. A. Krishnamurthy and K. Yelick, "Optimizing Parallel SPMD Programs," Workshop on Languages and Compilers for Parallel Computing, August 1994.

9. A. Krishnamurthy "Optimizing Explicitly Parallel Programs." University of Berkeley Computer Science Division Tech Report, Sept. 1994 (UCB/CSD-94-835)

7 Education

We developed extensive course material in conjuction with this grant. First, we developed CS 267, "Applications of Parallel Computing." This course is taught annually at Berkeley, usually with 50% enrollment from within CS and 50% from other engineering and physical science departments (and the occasional economist). It has been taught by PIs Culler, Demmel, and Yelick. Demmel developed an extensive set of on-line notes concentrating on algorithms (www.cs.berkeley.edu/~demmel/cs267), and later augmented by Culler (now.CS.Berkeley.EDU/cs267/) and Yelick (www.cs.berkeley.edu/~dmartin/cs267/). In addition to being taught at Berkeley, it was taught in short course form at a NSF-CBMS Lecture on Parallel Numerical Linear Algebra (June 1995, Demmel as main lecturer and organizer), at a NATO Advanced Summer Institute in July 1996, at ETH Zurich in 1996, and at an NPACI Training Institute in September 1997.

Second, we wrote and published a textbook for CS 258, "Parallel Computer Architecture". Culler and J.P. Singh are principal authors, with Anoop Gupta of Stanford and Microsoft. See www.cs.berkeley.edu/~culler/book.alpha/index.html. This book explains the forces behind the convergence of shared-memory, message-passing, data parallel, and data-driven computing architectures. It then examines the design issues that are critical to all parallel architecture across the full range of modern design, covering data access, communication performance, coordination of cooperative work, and correct implementation of useful semantics. It not only describes the hardware and software techniques for addressing each of these issues but also explores how these techniques interact in the same system.