Toolboxes and Templates for Large Scale Linear Algebra Problems
DOE Award # DE-FG03-94ER25219
Final Report
1 Jul 98 through 31 Oct 02
\$67 in unexpended funds

Zhaojun Bai
Computer Science Dept.
University of California
Davis, CA 95616

James Demmel
Mathematics Dept. and
Computer Science Div. of EECS
University of California
Berkeley, CA 94720

Jack Dongarra
Computer Science Dept.
University of Tennessee
Knoxville, TN 37966

Participants

There are 3 PIs and co-PIs, 78 other participants, and 10 other industrial contacts. The first list below contains the PIs. The second list below contains unfunded participants, The third list below contains participants funded by this grant. The fourth list contains “contacts”, people or organizations with whom we have been in significant contact about this work.

Principal Investigators

1. James Demmel, UC Berkeley
2. Jack Dongarra, U Tennessee and ORNL
3. Zhaojun Bai, UC Davis (formerly U Kentucky)

Unfunded collaborators

1. Beresford Parlett, UC Berkeley, Unfunded faculty collaborator, Developer of new eigenvalue algorithms, worked 160 hours
2. William Kahan, UC Berkeley, Unfunded faculty collaborator, collaborator on complexity and floating point issues, worked 160 hours
3. Osni Marques, NERSC/Lawrence Berkeley Natl. Lab., unfunded senior scientist, Developer of new eigenvalue algorithms, worked 160 hours
4. Xiaoye Li, NERSC/Lawrence Berkeley Natl. Lab., unfunded senior scientist, SuperLU developer and tester, coauthor of Eigentemplates book, worked 160 hours
5. Inderjit Dhillon, UT Austin, Unfunded faculty collaborator, Developer of new eigenvalue algorithms, worked <160 hours
6. Ming Gu, UCLA, Unfunded faculty collaborator, Developer of new eigenvalue algorithms and coauthor on Eigentemplates book, worked <160 hours
7. RenCang Li, U. Kentucky, Unfunded faculty collaborator, Developer of new eigenvalue algorithms, coauthor of Eigentemplates book, worked <160 hours
8. Benedikt Grosser, U. Wuppertal, German, unfunded postdoctoral contributor, developer of new eigenvalue algorithms
9. Gregorio Malajovich, Federal U. of Rio de Janeiro, Brazil, Unfunded visiting faculty, worked on complexity of condition estimation, worked <160 hours
10. Plamen Koev, UC Berkeley, grad student, funded, coauthor on Eigentemplates book, Developer of new high accuracy linear algebra, computational astrophysics, and computational algebra algorithms worked 160 hours
11. David Bindel, UC Berkeley, grad student, funded, worked on CLAPACK and on applications of reduced order modeling to MEMS CAD, worked 160 hours.

12. David Garmire, UC Berkeley, grad student, funded, worked on applications of reduced order modeling to MEMS CAD, worked 160 hours.
13. Jason Clark, UC Berkeley, grad student with other funding, working on reduced order modeling in SUGAR, worked <160 hours
14. David Day, Sandia Natl. Lab., Unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
15. Alan Edelman, MIT, Unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
16. Thomas Ericsson, Chalmers U., Sweden, unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
17. Roland Freund, Bell Labs, unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
18. Bo Kågström, U. of Umeå, Sweden, unfunded faculty collaborator, coauthor on Eigentemplates book, LAPACK contributor, worked <160 hours
19. T. Kowalski, MSC Software Inc., unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
20. Andrew Knyazev, U. of Colorado, Denver, unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
21. Richard Lehoucq, Sandia Natl. Lab., unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
22. Ross Lippert, Sandia Natl. Lab., unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
23. Kristi Maschoff, Cray Computer Co., unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
24. Karl Meerbergen, Rutherford Appleton Lab, England, unfunded senior collaborator, coauthor on Eigentemplates book, worked <160 hours
25. Ronald Morgan, Baylor U., unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
26. Axel Ruhe, Chalmers U., Sweden, unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
27. Yousef Saad, U. of Minnesota, unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
28. Gerard Sleijpen, U. of Utecht, the Netherlands, unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours

29. Dan Sorensen, Rice U., unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
30. Henk van der Vorst, U. of Utecht, the Netherlands, unfunded faculty collaborator, coauthor on Eigentemplates book, worked <160 hours
31. Linda Kaufman, Bell Labs, unfunded senior collaborator, contributions to LAPACK band routines and coauthor in BLAS Technical Forum, worked <160 hours
32. Sven Hammarling, NAG Ltd, United Kingdom, unfunded senior collaborator, contributions to LAPACK and coauthor in BLAS Technical Forum, worked 160 hours
33. Zohair Maany, NAG Ltd, United Kingdom, unfunded senior collaborator, contributions to LAPACK and coauthor in BLAS Technical Forum, worked <160 hours
34. Iain Duff, Rutherford Lab, United Kingdom and CERFACS, France, unfunded senior collaborator, coauthor in BLAS Technical Forum, worked <160 hours
35. Mike Heroux, Sandia Natl. Lab., unfunded senior collaborator, coauthor in BLAS Technical Forum, worked <160 hours
36. Fred Krogh, JPL, unfunded senior collaborator, coauthor in BLAS Technical Forum, worked <160 hours
37. Roldan Pozo, NIST, unfunded senior collaborator, coauthor in BLAS Technical Forum, worked 160 hours
38. Karin Remington, NIST, unfunded senior collaborator, coauthor in BLAS Technical Forum, worked <160 hours
39. George Corliss, Marquette U., unfunded faculty collaborator, coauthor in BLAS Technical Forum, worked <160 hours
40. Chenyi Hu, U. Houston, unfunded faculty collaborator, coauthor in BLAS Technical Forum, worked <160 hours
41. Baker Kearfott, U. Southwestern Louisiana, unfunded faculty collaborator, coauthor in BLAS Technical Forum, worked <160 hours
42. Bill Walster, Sun Microsystems, unfunded senior collaborator, coauthor in BLAS Technical Forum, worked <160 hours
43. Greg Henry, Intel Corp., unfunded industry collaborator, coauthor in BLAS Technical Forum
44. J. Wolff v. Gudenberg, U. Würzburg, Germany, unfunded faculty collaborator, coauthor in BLAS Technical Forum, worked <160 hours
45. Christof Vömel, CERFACS, France, unfunded graduate student, programmer in BLAS Technical Forum, worked 160 hours
46. Marcelin Youan, CERFACS, France, unfunded undergrad, programmer in BLAS Technical Forum, worked 160 hours

47. Michael Martin, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
48. Jimmy Iskandar, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
49. Daniel Jhin Yoo, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
50. Yozo Hida, UC Berkeley, unfunded undergrad, became unfunded grad student (holds NSF Fellowship), BLAS Technical Forum, XBLAS, high accuracy arithmetic, worked > 160 hours
51. Teresa Tung, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
52. Yulin Li, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
53. Henry Chen, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
54. Brandon Thompson, UC Berkeley, unfunded undergrad, programmer in BLAS Technical Forum, worked <160 hours
55. Tzu-Yi Chen, UC Berkeley, grad student on fellowship, sparse matrix balancing, coauthor on Eigentemplates book worked 160 hours
56. Richard Vuduc, UC Berkeley, unfunded grad student, worked on tuning BLAS implementations
57. Eduardo D'Azevedo, Oak Ridge National Lab, unfunded senior collaborator, coauthor on out-of-core algorithms paper
58. Andy Cleary, Lawrence Livermore Natl. Lab., unfunded senior collaborator, Coauthor on band solver paper, worked <160 hours
59. Markus Hegland, Australian Natl. U., unfunded senior collaborator, Coauthor on band solver paper, worked <160 hours
60. Peter Arbenz, Swiss Federal Inst. Tech., Zurich, Switzerland, unfunded senior collaborator, Coauthor on band solver paper, worked <160 hours
61. David Watkins, Washington State University, unfunded senior collaborator, Coauthor on parallel nonsymmetric QR paper, worked <160 hours
62. Anne Trefethen, NAG Ltd, unfunded senior collaborator, coauthor on cluster computing paper
63. Keith Moore, U. Tennessee, unfunded senior collaborator, coauthor on cluster computing paper
64. Soni Mukherjee, unfunded undergrad, programmer on XBLAS, worked 160 hours

Funded: (either by DOE or NSF)

1. Pitor Luszczek, U. Tennessee, graduate student coauthor on recursive sparse LU paper, worked 160 hours
2. Ben Liblett, UC Berkeley, grad student
3. Antoine Petit, U. Tennessee, postdoc, worked on LAPACK in Java and BLAS Technical Forum, worked <160 hours
4. Benjamin Diamant, UC Berkeley, grad student, worked on complexity of condition estimation, worked 160 hours
5. Daniel Skoogh, UCDavis, postdoc, graduated from Chalmers Technical University, Sweden, and partially supported by The Foundation Blanceflor Boncompagni-Ludovisi, Sweden. worked on sparse eigenvalue problems and reduced-order modeling of dynamical systems.
6. Clint Whaley, U. Tennessee, unfunded senior collaborator, coauthor in BLAS Technical Forum, worked 160 hours
7. Monika Jankun-Kelly, UC Davis, grad student, funded, worked on sparse eigenvalue problems and applications, worked 160 hours.
8. Liang Huang, UC Davis, worked on model reduction, funded, worked 160 hours
9. Farial Shanaz, U Tennessee, undergrad, worked <160 hours
10. Victor Eijkhout, U. Tennessee, Res. Asst. Prof., unfunded senior collaborator, Coauthor on survey paper, worked <160 hours
11. Kenneth Roche, U Tennessee, grad student, worked on ScaLAPACK and cluster computing, worked 160 hours
12. Ben Wanzo, UC Berkeley, undergrad, worked on BLAS, worked <160 hours
13. Anil Kapur, UC Berkeley, undergrad, worked on BLAS, worked <160 hours
14. Susan Blackford, U. Tennessee, senior staff member, responsible for overall LAPACK and BLAS Technical Forum releases, worked 160 hours

Other Contacts:

1. Patrick Amestoy, ENSEEIHT, Toulouse, France
2. Cleve Moler, The Mathworks
3. Fred Gustavson, IBM
4. Paresh G. Pattani, NEC
5. Hiroshi Ina, Fujitsu

6. Satoshi Hasakawa, Hitachi
7. Brian Whitney, Sun Microsystems
8. Todd Needham, Microsoft Research
9. Greg Astfalk, HP
10. Mimi Cella, SGI
11. Louis Komzsik, MSC software Inc.
12. Thomas Kowalski, MSC software Inc.

Activities and Findings

The following list of topics includes those in the original proposal and one added later. Each section of the report (Activities, Findings, Publications, Products, Contributions) is organized using the same list of topics.

- Activities

1. **Direct Sparse Solvers.** X. Li continued to perform performance tuning and detailed performance comparisons with other direct solvers. Patrick Amestoy, coauthor of one of the best alternative solvers MUMPS spent a year visiting to perform these comparisons, which led to a paper and talks at a SIAM conference. (this visit and work also had other funding).

We consulted with numerous users of SuperLU and helped them incorporate it into their applications and commercial products. We continued to develop the parallel distributed version of SuperLU and tune its performance (this work also had other funding, including NPACI).

We explored using a recursive formulation, similar to that introduced by Toledo and Gustavson for dense factorizations, for sequential sparse factorization, and evaluated its performance in comparison to other sparse factorization routines which have more complicated data structures designed for memory locality.

2. **Large Scale Eigenvalue Problems.** We completed work on the book “Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide,” or “Eigentemplates” for short. This book describes the state-of-the-art in solving eigenvalue problems. The 25 coauthors include many if not most of the world’s experts on this problem. This 410 page book was published by SIAM in October 2000.

We also developed new algorithms for computing functionals of large sparse matrices (eg determinant, trace etc), developed methods to precondition (“balance”) large scale eigenproblems to improve the accuracy of their solution, and collaborated with engineers to apply our algorithms to design of disk brakes that do not squeal.

3. **LAPACK and ScaLAPACK.**

We have continued to develop the “Holy Grail” of Parlett and Dhillon, propagating this optimal symmetric eigensolver to all places in LAPACK and ScaLAPACK where it can be used. We established a collaboration with Bendikt Grosser of U. Wuppertal, Germany, in whose PhD thesis (under Bruno Lang) he figured out how to extend the Holy Grail to the bidiagonal SVD, dealing with the difficult case of clustered singular values near zero. This was a major technical hurdle keeping us from propagating the Holy Grail to the singular value decomposition algorithms in LAPACK and ScaLAPACK.

We continue to work on accurate parallel variants of the “Holy Grail” that are most efficient when only computing a subset of the eigenvalues and eigenvectors.

We completed the production of CLAPACK 3.0, the C version of LAPACK 3.0, which was completed in 1999.

We completed the production of J LAPACK, the Java version of LAPACK, which was completed in 1999.

We also did a significant amount of consulting with Cleve Moler of the Mathworks (producers of Matlab) during the incorporation of LAPACK into Matlab.

We are investigating the use of new high precision BLAS (see under BLAS Technical Forum) in making certain LAPACK routines for iterative refinement of the solution of linear systems more accurate.

We developed more efficient parallel algorithms for the nonsymmetric eigenvalue problem and for data redistribution in ScaLAPACK.

4. **Automatic Performance Tuning.** We continue to work on the automatic tuning of BLAS and other computational kernels. This activity originated in a graduate parallel computing class at UC Berkeley several years ago, when the students were given one week to write a matrix multiplication routine and tune it to run as fast as possible on an IBM RS6000/590, a fast machine at that time. The point was to show how difficult it was to go near the machine peak speed because of the need to make the algorithm match the details of the memory hierarchy. At the end of the week we had a “race” against the vendor-tuned BLAS from IBM (in IBM’s ESSL library). Most student teams did nowhere near as fast as ESSL, as expected, but two teams beat ESSL. This inspired us to write a program, called PHIPAC, for automatically tuning matrix multiplication on any machine architecture by searching over a “space” of algorithms, running the algorithms, picking the fastest one (or ones, since it could depend on problem size).

Since then there has been an explosion of such tuning efforts, most successfully FFTW from MIT (for FFTs) and our own ATLAS (for the BLAS). ATLAS, inspired by the limitations of PHIPAC (inapplicability to Intel machines, too long search times), produced a widely used and much more complete set of tuned BLAS. This work is now expanding to tune many other computational and communication primitives. ATLAS’s impact will be described under “Contributions.”

5. **BLAS Technical Forum** The BLAS Technical Forum is a standards committee with representatives from academia, industry, and national labs. Its mandate is to extend the widely used BLAS to include
 - (a) useful dense and banded routines not included in the original BLAS1, BLAS2 and BLAS3 standards,
 - (b) sparse BLAS, and
 - (c) extended and mixed precision BLAS.

After years of regular meetings and votes, documented at www.netlib.org/blas/blast-forum, the final standard was approved in late 1999. Subsequently we completed the final 305-page standard document, several papers describing design details and justification, and a reference implementation. We are continuing to work on incorporating these new routines into improved versions of existing LAPACK routines.

6. **High Accuracy Numerical Linear Algebra.** We continued our investigations into the inherent complexity of accurate floating point computation. In the previous year we explored the classes of matrices that permit accurate floating point computations of basic linear algebra computations, such as inversion, and finding eigenvalues and singular values. We continued by exploring the complexity classes to which certain

floating point computations belong, asking whether certain known hard problems could be reduced to floating point expression evaluation, and whether weakening the accuracy requirement from forward to backward stability lowered the complexity. We made significant progress on several kinds of structured matrices: diagonally dominant M-matrices, polynomial Vandermonde matrices, and certain totally positive matrices. The work on totally positive generalized Vandermonde matrices involved devising new fast and accurate algorithms for the numerical evaluation of Schur and Jack polynomials, which are complicated combinatorial objects.

7. **Accuracy of Floating Point Computation.** We investigated the effect of roundoff on floating point summation, and showed that by summing floating point numbers in decreasing order of their absolute values, that extra bits of precision in the accumulator would lead to correct sums despite all intermediate cancellation and roundoff errors. We extended Harry Diamond's Theorem on how accurately two functions $f(x)$ and its inverse $g(x) = f^{-1}(x)$ still satisfy $f(g(x)) = x$ if $f(x)$ and $g(x)$ are rounded to the nearest floating point number.
8. **Structured matrices in computational astrophysics.** We collaborated with Julian Borrill of Lawrence Berkeley National Laboratory on exploiting structure in matrices arising in data analysis of Cosmic Microwave Background (CMB) radiation, in particular map making and power spectrum estimation. This resulted in parallel software that permits larger data sets to be analyzed than ever before, in preparation for upcoming PLANCK satellite experiments involving tens of millions of pixels.
9. **Computational Algebra.** We developed a new algorithm for efficiently evaluating the complete symmetric polynomial in the algebra of $n \times n$ matrices, and used this algorithm to prove part of an open conjecture from ring theory regarding polynomial identities in this matrix algebra.
10. **Reduced-order modeling of dynamical systems.** We have extended our study of Krylov-subspace based techniques for reduced-order modeling, from linear dynamical systems to non-linear dynamical systems, with applications in MEMS (microelectromechanical systems) and other areas. We have examined the so-called "nonlinear dynamics using linear modes" approach for nonlinear dynamical systems arising from MEMS applications, and studied reduced-order modeling scheme that preserve desirable properties of the original system, like stability and passivity.

- **Findings**

1. **Direct Sparse Solvers.** In “Analysis and Comparison of Two General Sparse Solvers for Distributed Memory Computers”, by P. Amestoy, J.-Y. L’Excellent, and X. Li, we compared the parallel sparse direct solvers SuperLU and MUMPS, comparing efficiency, accuracy, sparsity preservation and memory usage, interprocessor communication and scalability. We concluded that MUMPS is faster on the problem sizes and machines (Cray T3E) tested, but uses more memory, and would not scale as well to larger numbers of processors.

In “Recursive Approach in Sparse Matrix LU Factorization,” by J. Dongarra, V. Eijkhout, P. Luszczyk, we described and analyzed a recursive implementation of sparse LU factorization, and concluded that it can frequently be as efficient as specialized sparse matrix codes that use complicated data structures to achieve memory locality.

2. **Large Scale Eigenvalue Problems.** We completed work on the book “Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide,” or “Eigentemplates” for short. This book describes the state-of-the-art in solving eigenvalue problems. The 25 coauthors include many if not most of the world’s experts on this problem. This 410 page book was published by SIAM in October 2000.

We also developed new algorithms for computing functionals of large sparse matrices (eg determinant, entries or trace of the inverse, values of a transfer function etc.), combining the Lanczos method, theory of moments, orthogonal polynomials and approximation theory to devise efficient algorithms. These were exploited in our work on computational astrophysics described below.

We developed methods to precondition (“balance”) large scale eigenproblems to improve the accuracy of their solution, by find a method relying only on matrix-vector multiplications by a sparse matrix A to find a diagonal matrix D minimizing the norm of DAD^{-1} . Numerical examples showed that this can significantly improve the accuracy of subsequently computed eigenvalues.

We collaborated with engineers to apply our algorithms to design of disk brakes that do not squeal. Our ability to compute a few desired eigenvalues quickly let engineers quickly assess finite element models of proposed designs for disk brakes to see whether they had any eigenvalues (resonant frequencies, or “squeals”) in an annoying audible range.

3. **LAPACK and ScaLAPACK.**

In June 1999 we released version 3.0 of LAPACK (about 940K lines of code) and a new 407 page User’s Guide; both are substantially larger than their predecessors. LAPACK was incorporated into Matlab 6.0.

In August 2001 we released version 1.7 of ScaLAPACK. This release included new or improved routines for the nonsymmetric eigenvalue problem, the divide-and-conquer symmetric eigenvalue problem, generalized symmetric definite eigenvalue problem, and reduction to tridiagonal form. It also included a new parallel BLAS (PBLAS 2.0).

We published a large number of papers completing our understanding of the Holy Grail algorithm and how to ensure the accuracy and orthogonality of computed eigenvectors no matter how clustered the eigenvalues may be.

4. **Automatic Performance Tuning.**

ATLAS is widely used to produce tuned BLAS for many platforms. It won a 1999 R&D 100 Award. ATLAS and LAPACK on top of it were incorporated into Matlab 6.5.

5. **BLAS Technical Forum**

We completed the design and reference implementation of the new BLAS Standard.

6. **High Accuracy Numerical Linear Algebra.** J. Demmel was honored to be an invited speaker at the 2002 International Congress of Mathematicians in Beijing to describe this work.

Floating point numbers can be thought of as a pair of integers, the fraction and the exponent, and when doing complexity of floating point computation one must consider the dependence of the complexity on the size of both integers. If the size of the exponent is bounded, or grows slowly compared to the size of the fraction, then floating point numbers may be converted to integers without significant increase in complexity and most problems in numerical linear algebra may be solved in polynomial time using known algorithms. But if the size of the exponent is not small compared to the size of the fraction, then one can show that the complexity of accurate floating point computation is essentially equivalent to symbolic manipulation, and so has much higher complexity. For example, we determined that the problem of evaluating the “middle bits” of the very simple floating point expression $\prod_{i=1}^n (a_i + b_i)$ is PP-Hard (much harder than NP-Hard) because one can reduce computing the permanent to this problem. The complexity remains high for some expressions even if one settles for mere backward stability.

We devised new fast and accurate algorithms for computing the singular value decomposition of weakly diagonally dominant M-matrices, devised new fast and accurate algorithms for the evaluation of Schur functions, used these Schur function algorithms for accurately and efficiently solving totally positive generalized Vandermonde systems of equations. These results appear in the PhD thesis of Plamen Koev and subsequent publications. These results are being extended to finding accurate eigenvalue and singular value decompositions of a large class of totally positive matrices, and to the fast and accurate evaluation of Jack polynomials.

7. **Accuracy of Floating Point Computation.** We investigated the effect of roundoff on floating point summation, and showed that by summing n floating point numbers with f -bit fractions in decreasing order of their absolute values, and using an accumulator with $F > f$ bits in its fraction, then as long as $n \leq \text{floor}(2^{F-f}/(1 - 2^{-f})) + 1$, the sum will be computed with nearly all correct bits despite all intermediate cancellation and roundoff errors. We applied this to improve the speed of certain algorithms in computational geometry and mesh generation.

We extended Harry Diamond’s Theorem on how accurately two functions $f(x)$ and its inverse $g(x) = f^{-1}(x)$ still satisfy $f(g(x)) = x$ if $f(x)$ and $g(x)$ are rounded to the nearest floating point number. Specifically, if $F(x)$ is the rounded version of $f(x)$, $G(x)$ is the rounded version of $g(x)$, and $H(x) = G(F(x))$, then Harry Diamond’s Theorem asserts that if f is strictly convex or concave, then we can only expect that $H(H(H(x))) = H(H(x))$, rather than $H(x) = x$. We show that if rounding is done as in the IEEE floating point standard, then we can eliminate the requirement that convexity/concavity be strict.

8. **Structured matrices in computational astrophysics.** We lowered the complexity of map making and power spectrum estimation for the Cosmic Microwave Background (CMB) radiation from $O(n^3)$ to $O(n^{3/2})$, and provided a scalable parallel implementation, making it possible for the first time to analyze the enormous data sets being collected now and in the future in the PLANCK experiment. This work was presented as invited talks at the SIAM Annual Meeting in San Diego, and at the PLANCK 2001 Workshop in Pisa, Italy, both in July 2001. This work is described in Koev's PhD thesis.
9. **Computational Algebra.** We developed a new algorithm for efficiently evaluating the complete symmetric polynomial in the algebra of $n \times n$ matrices. This algorithm exploits dynamic programming to run exponentially faster than the traditional algorithm. We used this algorithm to prove part of an open conjecture from ring theory regarding polynomial identities in this matrix algebra, namely that all polynomial identities in this algebra of degree $2n + 2$ follow from the Standard Identity of degree $2n$, for $n = 4$ and $n = 5$. This work is described in Koev's PhD thesis.
10. **Reduced-order modeling of dynamical systems.** We have developed a symmetric band Lanczos process for reduced-order modeling that is significantly more accurate on large scale models of circuits, and preserves passivity of the original system both in theory and in the presense of roundoff.
 We have developed partial Padè approximations using the Lanczos method to simultaneously match certain poles and zeros, and closely match the original transfer function. This lets us preserve stability and passivity in the reduced order model.
 We have developed reduced order models for non-linear systems by starting with the Carleman bilinearization and then producing a reduced bilinear system using Krylov subspace techniques.
 Finally, we describe how we introduced reduced order modeling techniques into SUGAR, a Computer Aided Design (CAD) tool for MicroElectroMechanical Systems (MEMS).

- **Training and Development.** Of the 78 other participants, there were 10 current graduate students, 3 post docs, 5 former graduate students, and 12 undergraduates trained by working on this project. Former graduate students include the following.

Plamen Koev got his PhD in Mathematics from UC Berkeley in 2002 and is currently a postdoc in mathematics at MIT, working with Alan Edelman. His PhD thesis title was “Accurate and efficient computations with structured matrices.” His adviser was J. Demmel.

Tzu-Yi Chen got her PhD in Computer Science from UC Berkeley in 2002 and is currently an assistant professor at Pomona College. Her PhD thesis title was “Preconditioning sparse systems for computing eigenvalues and solving linear systems.” Her adviser was J. Demmel.

Tom Kowalski got his PhD in Mathematics from the University of Kentucky in 2000. His thesis title was “Extracting a Few Eigenpairs of Symmetric Indefinite Matrix Pencils.” After graduation he went to work for MSC Software Inc., the producers of NASTRAN. His adviser was Z. Bai.

Benjamin Diamant got his Master’s in Computer Science from UC Berkeley in 1999, and is currently working at Google. His thesis title was “On the Complexity of Computing Error Bounds.” His adviser was J. Demmel.

Kenneth Roche got his Master’s in Computer Science from U Tennessee in 2002. His thesis title was “Deploying parallel numerical library routines to cluster computing in a self-adapting fashion: A case study.” His adviser was J. Dongarra.

- **Outreach Activities.** We have presented tutorials on our software at training sessions organized by NPACI at their yearly All Hands Meetings.

Publications and Products

• Journal Publications

1. Direct Sparse Solvers.

- “Recursive Approach in Sparse Matrix LU Factorization,” J. Dongarra, V. Eijkhout, P. Luszczek, *Scientific Programming*, V. 9, n. 1, pp 51-60, 2001
- “Analysis and Comparison of Two General Sparse Solvers for Distributed Memory Computers,” P. R. Amestoy, I. Duff, J.-Y. L’Excellent, X. S. Li, *ACM Trans. Math. Soft.*, v. 27, n. 4, Dec 2001, pp. 388-421
- “Electron-impact ionization of atomic hydrogen,” M. Baertschy, T. N. Rescigno, W. A. Isaacs, X. S. Li and C. W. McCurdy, *Physical Review A*, vol. 63, Jan 18, 2001

2. Large Scale Eigenvalue Problems

- Z. Bai and R. Freund, Eigenvalue-based Characterization and Test for Positive Realness of Scalar Transfer Functions, *IEEE Trans. Automat. Control.*, Vol. 45, No.12, Dec. 2000, pp.2396 – 2402.
- G. Lou, T. W. Wu, L. Lee, K. Xu, and Z. Bai, Disk Brake System Squeal Prediction Using the ABLE Algorithm, *Journal of Sound and Vibration*, to appear
- Z. Bai and G. Golub, Computation of large-scale quadratic forms and transfer functions using the theory of moments, quadrature and Pade approximation. pp.1-30, in “Modern Methods in Scientific Computing and Applications”, edited by A. Bourlioux and M. J. Gander, Kluwer Academic Publisher, Netherlands, 2002 NATO Science Series, II Mathematics, Physics and Chemistry, Vol. 75.
- “Balancing Sparse Matrices for Computing Eigenvalues,” James Demmel, Tzu-Yi Chen, *Lin. Alg. Appl.*, 309(1-3) 2000, pp. 261-287

3. LAPACK and ScaLAPACK

- “A Comparison Of Parallel Solvers For General Narrow Banded Linear Systems”, P. Arbenz, A. Cleary, J. Dongarra, and M. Hegland, *Parallel and Distributed Computing Practices*, v. 2, 1999, pp 385-400.
- “Algorithmic redistribution methods for block-cyclic decompositions,” A. Petit, J. Dongarra, *IEEE Trans. Par. Distr. Systems*, v. 10, n. 12, pp 201-220, 1999.
- “Numerical Linear Algebra Algorithms and Software”, Jack Dongarra and Victor Eijkhout, *Journal of Computation and Applied Mathematics*. Volume 123, 2000, pp 489-514.
- “A Parallel Implementation of the Nonsymmetric QR Algorithm for Distributed Memory Architectures”, G. Henry, D. Watkins, J. Dongarra, *SIAM Journal on Scientific Computing*, v. 24, n. 1, pp 284–311, 2003.
- “Numerical Libraries and Tools for Scalable Parallel Cluster Computing”, J. Dongarra, K. Moore, and A. Trefethen, *International Journal of High performance Applications and Supercomputing*, Volume 15, No. 2, Summer 2001, pp 175-180.
- “Deploying parallel numerical library routines to cluster computing in a self-adapting fashion”, K. Roche and J. Dongarra, in *Parallel Computing: Advances and Current Issues*, Proceedings of the International Conference ParCo2001, Eds. Joubert, Murli, Peters, Vanneschi; Imperial College Press, London, England, Jan 2002.

- “The design and implementation of the parallel out-of-core ScaLAPACK LU, QR, and Cholesky factorization routines”, Eduardo D’Azevedo and Jack Dongarra, *Concurrency: Practice and Experience*, Volume 12, Issue 15, 2000. Pages: 1481-1493, ISSN: 1040-3108.
- “An Implementation of the dqds Algorithm (positive case),” B. Parlett and O. Marques, *Lin. Alg. Appl.*, v. 309, 2000, pp. 217–259
- “For tridiagonals T : Replace T with LDL^T ”, B. Parlett, *J. Computational and Applied Math.*, v. 123, 2000, pp 117-130.
- “Relatively Robust Representations of Symmetric Tridiagonals,” I. Dhillon and B. Parlett, *Lin. Alg. Appl.* v. 309, 2000, pp 121-151
- “Orthogonal Eigenvectors and Relative Gaps,” I. Dhillon and B. Parlett, submitted to *SIAM J. Mat. Anal. Appl.*, 2001; also LAPACK Working Note 154
- “Multiplied Representations for Orthogonality,” I. Dhillon and B. Parlett, submitted to *Lin. Alg. Appl.*, 2002
- “Perturbation of eigenpairs of factored symmetric tridiagonal matrices,” B. Parlett, to appear in *Found. Comput. Math.* 2003

4. Automatic Performance Tuning

- “Automated Empirical Optimization of Software and the ATLAS Project”, R. Whaley, A. Petitet, and Jack Dongarra, *Parallel Computing*, v. 27, n. 1-2, pp 3-25, 2001.
- “Self adapting numerical algorithm for next generation applications,” J. Dongarra, V. Eijkhout, to appear in *Intern. J. High Performance Appl. and Supercomputing*, 2002

5. BLAS Technical Forum

- “An Updated Set of Basic Linear Algebra Subprograms (BLAS)”, S. Blackford, J. Demmel, J. Dongarra, I. Duff, S. Hammarling, G. Henry, M. Heroux, L. Kaufman, A. Lumsdaine, A. Petitet, R. Pozo, K. Remington, C. Whaley, *ACM Trans. Math. Soft.*, v. 28, i. 2, pp 135-151, June 2002
- “Design Implementation and Testing of Extended and Mixed Precision BLAS”, J. Demmel, X. Li, D. Bailey, G. Henry, Y. Hida, W. Kahan, M. Martin, J. Iskandar, A. Kapur, T. Tung, D. J. Yoo, *ACM Trans. Math. Soft.*, v. 28, issue 2, June 2002. (also LBNL Technical Report LBNL-45991, also available at www.netlib.org/blas/blast-forum)
- “On computing Givens rotations reliably and efficiently,” D. Bindel, J. Demmel, W. Kahan, O. Marques, *ACM Trans. Math. Soft.*, v. 28, issue 2, June 2002; also available at www.netlib.org/blas/blast-forum BLAS”, 1999;

6. High Accuracy Numerical Linear Algebra

- “The complexity of accurate floating point computation,” J. Demmel, *Proc. ICM 2002*, Beijing
- “Accurate SVDs of weakly diagonally dominant M-matrices,” P. Koev and J. Demmel, submitted to *Numerische Mathematik*, 2002
- “Accurate SVDs of polynomial Vandermonde matrices involving orthonormal polynomials,” P. Koev and J. Demmel, submitted to *Lin. Alg. Appl.*, 2002

- “Necessary and Sufficient Conditions for Accurate and Efficient Singular Value Decompositions of Structured Matrices,” J. Demmel and P. Koev, (invited) Contemporary Mathematics Series # 281, Special Issue on Structured Matrices in Mathematics, Computer Science and Engineering, Vol. II (V. Olshevsky, editor), Proceedings of an AMS-IMS-SIAM Joint Summer Research Conference at UC Boulder, June 1999, American Mathematical Society, 2001, pp 117-145
- “Accurate SVDs of Structured Matrices,” J. Demmel, SIAM J. Mat. Anal. Appl., v. 21, n. 2, 1999, pp 562-580
- “Computing the Singular Value Decomposition with High Relative Accuracy,” J. Demmel, Z. Drmac, S. Eisenstat, M. Gu, I. Slapnicar, K. Veselic, Linear Algebra and its Applications, v. 299, i. 1-3, pp 21-80, Nov 1999
- “On the Complexity of Computing Error Bounds,” J. Demmel, B. Diament and G. Malajovich, Found. Comput. Math. 1:101-125 (2001)

7. Accuracy of Floating Point Computation.

- “Accurate Floating Point Summation,” J. Demmel, Y. Hida. UC Berkeley Computer Science Division Technical Report UCB/CSD-02-1180, May 2002. Submitted to SIAM J. Sci. Comp. 2002 (accepted 2003)
- “Fast and Accurate Floating Point Summation with Application to Computational Geometry,” presented at 10th GAMM-IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics (SCAN 2002)

8. Computational Algebra.

- “Computational Approach to Polynomial Identities of Matrices – A Survey”, with F. Benanti, V. Drensky, P. Koev, to appear in “Ring Theory: Polynomial Identities and Combinatorial Methods”, A. Giambruno, A. Regev and M. Zaicev, eds., Marcel Dekker, 2002

9. Reduced-order modeling of dynamical systems.

- Z. Bai, D. Bindel, J. Clark, J. Demmel, K. S. J. Pister, N. Zhou, New Numerical Techniques and Tools in SUGAR for 3D MEMS Simulations, to appear in the proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems, March, 2001.
- Z. Bai and R. Freund, A symmetric band Lanczos process based on coupled recurrence and some applications, SIAM J. of Sci. Comput. v. 23, pp 542–562, 2001
- Z. Bai and R. Freund, “A partial Pade-via-Lanczos method for reduced-order modeling,” Linear Algebra and its Applications, v. 332-334, Aug 2001, pp 139-164
- Z. Bai and D. Skoogh, Krylov subspace techniques for reduced-order modeling of nonlinear dynamical systems (6 pages), in Electronic proceedings of 15th international symposium on the Mathematical Theory of Networks and Systems, D. Gilliam and J. Rosenthal, eds., Univ. of North Dame, Indiana, Aug.12-16, 2002
- Z. Bai, Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems, Applied Numerical Mathematics, Vol.43, Issues 1-2, Pages 9-44, Oct. 2002.

- **Books or Other one-time Publications**

1. **Sparse Direct Solvers.**

- “Recursive Approach in Sparse Matrix LU Factorization” Jack Dongarra, Victor Eijkhout, Pitor Luszczek, SGI Users’ Conference Proceedings, Edited by M. Bubak, J. Moscinski, M. Noga, Publisher Academic Computer Center CYFRONET AGH, Krakow, Poland, ISBN 83-902363-9-7, 2000, pp. 409-418

2. **Large Scale Eigenvalue Problems.**

- “Templates for the Solution of Algebraic Eigenvalues Problems: A Practical Guide”, Edited by Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, H. van der Vorst, with contributions from Z. Bai, T-Z. Chen, D. Day, J. Demmel, J. Dongarra, A. Edelman, T. Ericsson, R. Freund, M. Gu, B. Kågström, A. Knyazev, T. Kowalski, R. Lehoucq, R-C. Li, R. Lippert, K. Maschoff, K. Meerbergen, R. Morgan, A. Ruhe, Y. Saad, G. Sleijpen, D. Sorensen, H. van der Vorst; 410 pages, SIAM, 2000
- “Preconditioning sparse systems for computing eigenvalues and solving linear systems.” Tzu-Yi Chen, UC Berkeley, PhD Thesis, Computer Science Division, 2002.
- “Extracting a Few Eigenpairs of Symmetric Indefinite Matrix Pencils.” T. Kowalski, U. Kentucky, PhD Thesis, Mathematics Department, 2000.

3. **LAPACK and ScaLAPACK.**

- “LAPACK Users’ Guide”, E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov and D. Sorensen, SIAM, Third edition, October 1999, 407 pages

4. **BLAS Technical Forum.**

- “Basic Linear Algebra Subprograms Technical (BLAST) Forum Standard”, with J. Dongarra, S. Hammarling, S. Blackford, A. Lumsdaine, L. Kaufman, Z. Maany, A. Petitet, I. Duff, M. Heroux, R. Pozo, K. Remington, G. Henry, W. Kahan, X. Li, W. C. Whaley, F. Krogh, G. Corliss, C. Hu, B. Kearfott, W. Walster, J. Wolff v. Gudenberg, 305 pages, Intern. J. High Performance Comput. Appl. v. 16, n. 1, Spring 2002; also available at www.netlib.org/blas/blast-forum

5. **High Accuracy Numerical Linear Algebra.**

- “Accurate and Efficient Computations with Structured Matrices,” P. Koev, UC Berkeley, PhD Thesis, Mathematics Department, 2002.
- “On the Complexity of Computing Error Bounds,” B. Diamant, UC Berkeley, Master’s Thesis, Computer Science Division, 1999.

- **Internet Dissemination**

1. **Sparse Direct Solvers.** Our SuperLU software and documentation is available at www.nersc.gov/~xiaoye/SuperLU/index.html It has also been installed at various NPACI sites.
2. **Large Scale Eigenvalue Problems.** Software described in our “Eigentemplates” book may be found at www.cs.ucdavis.edu/~bai/ET/contents.html.

3. LAPACK and ScaLAPACK.

- **LAPACK.** www.netlib.org/lapack
- **CLAPACK.** C translation of LAPACK. www.netlib.org/clapack
- **JLAPACK.** Java translation of LAPACK. www.cs.utk.edu/f2j
- **ScaLAPACK.** www.netlib.org/scalapack

4. Automatic Performance Tuning. Atlas, see www.netlib.org/atlas

5. BLAS Technical Forum.

- www.netlib.org/blas/blast-forum
- **XBLAS.** Extended and mixed precision BLAS reference implementation. www.nersc.gov/~xiaoye/XBLAS

6. Courseware. www.cs.berkeley.edu/~demmel/cs267_Spr99 This course has been offered yearly by other faculty, with continuous updating by other faculty of the on-line notes. For the on-line version of the latest offerings, see inst.eecs.berkeley.edu/~cs267

• Other Products

– Software and Documentation

1. **Sparse Direct Solvers.** Our SuperLU software and documentation is available at www.nersc.gov/~xiaoye/SuperLU/index.html It has also been installed at various NPACI sites.
2. **Large Scale Eigenvalue Problems.** Software described in our “Eigentemplates” book may be found at www.cs.ucdavis.edu/~bai/ET/contents.html.
3. **LAPACK and ScaLAPACK.**

- * **LAPACK.** www.netlib.org/lapack
- * **CLAPACK.** C translation of LAPACK. www.netlib.org/clapack
- * **JLAPACK.** Java translation of LAPACK. www.cs.utk.edu/f2j
- * **ScaLAPACK.** www.netlib.org/scalapack

4. Automatic Performance Tuning. Atlas, see www.netlib.org/atlas

5. BLAS Technical Forum.

- * www.netlib.org/blas/blast-forum
- * **XBLAS.** Extended and mixed precision BLAS reference implementation. www.nersc.gov/~xiaoye/XBLAS

- **Courseware.** Graduate course in Applications of Parallel Computing, see www.cs.berkeley.edu/~demmel/cs267_Spr99. This course has been offered yearly by other faculty, with continuous updating by other faculty of the on-line notes. For the on-line version of the latest offerings, see inst.eecs.berkeley.edu/~cs267

Contributions

- **Contributions within discipline.**

1. **Direct Sparse Solvers.** Our sparse solver SuperLU is widely used, although the diversity of sparse matrices means that no single algorithm will be best in all cases. Versions of SuperLU have been included (or are planned for inclusion) in the DOE scientific programming tools Omega3P (for particle accelerator design), Dspice (for circuit simulation), Trilinos, NIKE, PETSc and Hypre, the NPACI Engineering Thrust code OpenSees (for earthquake simulation), industrial libraries from HP, Matlab, Sun, Boeing, NAG, FEMLAB and Python.
2. **Large Scale Eigenvalue Problems.** By summarizing the state-of-the-art in software, describing a “decision tree” for choosing the best algorithm, and collecting the best software at a single location, we have made it possible for many more people to solve their problem efficiently, as well as identified what problems remain open for researchers in the field to work on.
3. **LAPACK and ScaLAPACK.** LAPACK and ScaLAPACK are widely used dense and band linear algebra libraries. LAPACK forms the basis of the mathematical libraries of HP/Convex, IBM, Compaq/Digital, Sun, Intel, NEC, Fujitsu, Hitachi, SGI/Cray, and NAG. Matlab 6 now uses LAPACK for its matrix computation. ScaLAPACK has been also incorporated into the parallel libraries of NAG, IBM, and SGI/Cray, and is being incorporated into the IMSL, Fujitsu, HP/Convex, Hitachi and NEC libraries. The ScaLAPACK library has become an official release for DOE’s ASCI Red operating system. Altogether, this forms a large fraction, if not the majority, of the dense linear algebra user community.
4. **Automatic Performance Tuning.** The ATLAS library for producing tuned BLAS has been very influential. It does not simply lift the burden on a few programmers at machine vendors to produce tuned BLAS, it makes the dissemination of good software much easier. For example, the Mathworks did not want to incorporate LAPACK into Matlab for many years because LAPACK’s speed depends on having tuned BLAS, and some vendors charged more for their tuned BLAS than all of Matlab cost, making it very unattractive for customers. ATLAS fixed this problem by making *public domain* tuned BLAS available on any platforms, and so letting the Mathworks distribute Matlab with LAPACK and tuned BLAS on any platform.
5. **BLAS Technical Forum.** The popularity of the BLAS (Basic Linear Algebra Subroutines) as a standard interface for high performance software motivated the community to extend its functionality to a wider set of mathematical operations, data structures, and precisions. This Forum provided a standard setting body to find a consensus among academic and industrial researchers as to what this standard should contain, and to encourage reference implementations. This process has been successful, but wide acceptance of the new standard may ultimately depend on tuned implementations being made available using approaches analogous to those described above under “Automatic Performance Tuning.”
6. **High Accuracy Numerical Linear Algebra.** Our investigations into the underlying complexity of solving linear algebra problems with floating point data has been ongoing

for many years, and was recognized with invitations to present the work both at the 2002 International Congress of Mathematicians (ICM) in Beijing and the 2003 International Congress on Industrial and Applied Mathematics (ICIAM) in Sydney.

7. **Accuracy of Floating Point Computation.** Our algorithm for accurate floating point summation is much simpler and faster than any previous algorithm for this purpose.
8. **Reduced-order modeling and applications.** Our methods for reduced order modeling that preserve desirable properties of the original system like stability and passivity, and use efficient Krylov subspace techniques for nonlinear systems, are among the first to solve these important problems.

- **Contributions to other disciplines.**

The cover article of Nature (April 27, 2000) featured the BOOMERanG collaboration, which is analyzing cosmic microwave background radiation using the MADCAP code, which in turn depends on ScaLAPACK for some of its data analysis.

The cover article of Science (Dec 24, 1999) described the first accurate first-principles computation of three-body scattering. This large scale eigenvalue problem depended on the distributed memory version of SuperLU (see also the paper listed above entitled “Electron-impact ionization of atomic hydrogen.”)

SuperLU has been incorporated in codes (see above for names) for particle accelerator design, circuit simulation and earthquake simulation.

Our “Eigentemplates” book makes it easier for non-experts to find the best algorithms and software for their large scale eigenvalue problems.

Our work in computational astrophysics is currently one of the best available codes for analyzing cosmic microwave background radiation data.

Our work on fast evaluation of certain matrix polynomials has made it possible to computationally explore a number of ring theoretic conjectures that have been open for a long time, such as the structure of polynomial identities in matrix algebras, and which ones follow from the Standard Identity proven by Amitsur and Levitzki in 1950. By using our exponentially faster algorithm and a large parallel computer, we were able to prove some special cases of a conjecture that all identities must follow from the Standard Identity.

Our work on the MEMS CAD tool SUGAR has made this tool more widely useful to the growing community of MEMS designers.

- **Contributions to Human Resource Development.** Of the 78 other participants, there were 10 current graduate students, 3 post docs, 5 former graduate students, and 12 undergraduates trained by working on this project.

Our on-line course on parallel computing is available at www.cs.berkeley.edu/~demmel/cs267_Spr99. This course has been offered yearly by other faculty, with continuous updating by other faculty of the on-line notes. For the on-line version of the latest offerings, see inst.eecs.berkeley.edu/~cs267

- **Contributions to Resources for Research and Education.** Software described in our book “Eigentemplates” is available on-line at www.cs.ucdavis.edu/~bai/ET/contents.html,

providing a resource for practitioners in all fields to get state-of-the-art software for solving their large scale eigenvalue problems.

- **Contributions beyond Science and Engineering.** Our software (LAPACK, ScaLAPACK, ATLAS, SuperLU) is widely used in commercial software products from many companies as listed in earlier sections.