In Honour of Bernard Philippe

PROGRAMME AND ABSTRACTS

5th International Workshop on Parallel Matrix Algorithms and Applications (PMAA'08)

http://www.dcs.bbk.ac.uk/pmaa08/

Department of Computer Science, University of Neuchâtel, Switzerland
June 20-22, 2008

Address:

Universite de Neuchâtel
UniMail
Rue Emile-Argand 11
CH-2007 Neuchâtel
Switzerland

ERCIM
European Research Consortium
for Informatics and Mathematics

PMAA08 Co-Chairs

E.J. Kontoghiorghes (CY), P. Arbenz (CH), Y. Saad (USA) and A. Sameh (USA).

PMAA08 Programme committee:

M. Adams (USA), M. Arioli (UK), M. Bollhofer (DE), O. Beaumont (FR), C. Bekas (CH), M. Berry (USA), E. Boman (USA), P. D'Ambra (IT), M. Dayde (FR), T. Drummond (US), E. Elmroth (SE), N. Emad (FR), J. Erhel (FR), E. Gallopoulos (GR), L. Giraud (FR), A. Grama (USA), L. Grigori (FR), M. Gutknecht (CH), F. Guyomarch (FR), T. Huckle (DE), E. Jeannot (FR), A. Karaivanova (BG), J. Kraus (AT), D. Kressner (CH), P. Kropf (CH), N. Missirlis (GR), S. Margenov (BG), E. Ng (USA), G. Oksa (SK), S. Petiton (FR), J. Roman (FR), J.E. Roman (ES), M. Sosonkina (USA), O. Schenk (CH), D. di Serafino (IT), D. Szyld (USA), D. Trystram (FR), M. Tuma (CZ), P. Yanev (CH), M. Vajtersic (AT), P. Vasconcelos (PT), J. Wasniewski (DK), Z. Zlatev (DK).

Local organizing committee:

J. Argyrou, C. Barna, C. Bekas, G. Botalla, E. Constandinide, L. Dolamic, C. Fautsch, C. Gatu, M. Hofmann, J. Misserli, P. Kropf, M. Lyra, A. Matei, J. Savoy, E. Tsinti and P. Yanev.

Dear Friends and Colleagues,

Welcome to the 5th International Workshop on Parallel Matrix Algorithms and Applications (PMAA'08). The workshop Co-chairs are happy to host this international conference in Neuchâtel. Tow days of the meeting take place jointly at the same venue as the First Workshop of the ERCIM Working Group on Computing & Statistics. The first two workshops PMAA00 and PMAA02 took place in Neuchâtel while the last two PMAA04 and PMAA06 in France.

The PMAA08 is organized in honour of Bernard Philippe.

The workshop aims to be a forum for an exchange of ideas, insights and experiences in different areas of parallel computing in which matrix algorithms are employed. The Workshop will bring together experts and practitioners from diverse disciplines with a common interest in matrix computation. The PMAA08 programme consists of 17 regular sessions, 2 plenary talks and around 80 presentations. There are over 100 participants. The ERCIM meeting provides additional 300 presentations and participants.

Peer reviewed papers presented at the PMAA08 will be considered for publication in a special issue of the Parallel Computing journal and the journal of Applied and Numerical Mathematics.

The Co-chairs have endeavored to provide a balanced and stimulating programme that will appeal to the diverse interests of the participants. The local organizing committee hopes that the conference venue will provide the appropriate environment to enhance your contacts and to establish new ones.

The conference is a collective effort of many individuals and organizations. The Co-chairs, the scientific programme committee, the local organizing committee and volunteers have contributed substantially to the organization of the conference. We are acknowledging the support of our sponsors and particularly the host Department of Computer Science, University of Neuchâtel.

We hope that you enjoy the conference and your stay in Neuchâtel.

The conference Co-chairs:
Erricos John Kontoghiorghes (Chair), *Peter Arbenz*, *Yousef Saad* and *Ahmed Sameh*.
The local organizers: *Costas Bekas, Cristian Gatu, Peter Kropf* and *Petko Yanev*.

SCHEDULE

All lectures take place at the UniMail building (Rue Emile-Argand 11), University of Neuchâtel.

Thursday, 19th June 2008

18:00 - 19:30	Reception

Friday, 20th June 2008

11:10 - 11:20	Opening (Room: GGA)
11:20 - 12:15	Plenary Talk (Bernard Philippe)
12:15 - 14:00	Lunch Break
14:00 - 16:00	Parallel Sessions G
16:00 - 16:25	Coffee Break
16:25 - 18:30	Parallel Sessions H
20:00	Conference Dinner

Saturday, 21st June 2008

09:00 - 11:00	Parallel Sessions I
11:00 - 11:20	Coffee Break
11:20 - 13:20	Parallel Sessions J
13:20 - 15:00	Lunch Break
15:00 - 17:00	Parallel Sessions K
17:00 - 17:20	Coffee Break
17:20 - 18:15	Plenary Talk (Michael Berry)
19:30	Fondue Dinner

Sunday, 22nd June 2008

08:40 - 10:40	Parallel Sessions M
10:40 - 11:00	Coffee Break
11:00 - 13:00	Parallel Sessions N
13:00 - 14:30	Lunch Break

SOCIAL EVENTS

- The coffee breaks will last one hour each (which adds fifteen minutes before and after to the times that are indicated in the programme). Weather permitting the coffee breaks will take place on the terrace by the cafeteria of UniMail, otherwise they will take place in the first and second floor of UniMail.
- Welcome Reception, Thursday 19th June, 18:00. The reception is open to all registrants. It will take place in the Neuchâtel Castle (Salle de Chavalieres, Château de Neuchâtel). You must have your conference badge in order to attend the reception.
- Lunches will be served at the *Restaurant Le Romarin* which is 10 minutes walk from the venue. Anyone not registered for the lunch can have meals at the UniMail cafeteria (except the weekend) and at the various restaurants of the shopping centre which is 15 minutes walk from the venue.
- Conference Dinner, Friday 20th June, 20:00. The Conference Dinner will take place at the gastronomic restaurant Hotel DuPeyrou, Avenue DuPeyrou 1, CH-2000 Neuchâtel. The restaurant is 10-15 minutes walk from UniMail and the town centre (Detailed information will be available at the conference registration desk). The conference dinner is optional and registration is required.

  You must have your Conference Dinner ticket and your conference badge in order to attend the conference dinner.
- Fondue Dinner, Saturday 21st June, 19:30. The Fondue Dinner will take place at two different places: at the *Brasserie Le Cardinal* (Rue de Seyon 9, Neuchâtel) which is at the centre of the town, and *La Taverne Neuchâteloise* (Rue de l'Orangerie 5, Neuchâtel) which is 10-15 minutes walk from the centre. The tickets indicate the name of the restaurant which you should attend. Please note that the restaurants are fully booked and you should attend the restaurant indicated in your ticket. The fondue dinner is optional and registration is required.

  You must have your Fondue Dinner ticket and your conference badge in order to attend the conference dinner.

GENERAL INFORMATION

Lecture Rooms

The paper presentations will take place at the UniMail, University of Neuchâtel. There are ten lecture rooms. Three of them (GGA, GPA and GB1) are in the Chemistry building, while the other seven are in the main building of UniMail. There will be signs indicating the location of the lecture rooms. Please ask for assistance and directions at the registration desk.

The plenary talks will take place in the lecture room GGA (Chemistry building), and will last 55 minutes including questions. Each session will be 2 hours long. Chairs are requested to keep the session on schedule. Papers should be presented in the order in which they are listed in the programme for the convenience of attendees who may wish to switch rooms mid-session to hear particular papers. In the case of a no-show, please use the extra time for a break or a discussion so that the remaining papers stay on schedule.

Presentation instructions

The lecture rooms will be equipped with a PC, a computer projector and in most cases an overhead projector. The session chairs should obtain copies of the talks in a USB stick before the session starts (use the lecture room as the meeting place), or obtain the talks by email prior to the conference beginning. Presenters must deliver the files with the presentation in PDF (Acrobat) or PPT (Powerpoint) format on a USB memory stick to the session chair ten minutes before each session.

The PC in the lecture rooms should be used for presentations. The session chairs should have a laptop for backup.

Swiss plugs/power outlets are different from those in the rest of Europe, including Germany. We cannot provide adapters, so please do not forget to take your adapters if needed.

Internet

There will be access to PCs connected to the Internet at the main entrance of the UniMail. The wireless Internet connection is also freely available at UniMail.

Messages

You may leave messages for each other on the bulletin board by the registration desks.

SUPPORTERS

Department of Computer Science, University of Neuchâtel, Switzerland

ERCIM (European Research Consortium for Informatics and Mathematics)

International Association for Mathematics and Computers in Simulation

Society for Industrial and Applied Mathematics

The International Linear Algebra Society

Journal of Parallel Computing

Journal of Applied and Numerical Mathematics

Elsevier

Banque Cantonale Neuchâteloise.

University of Salerno, Italy

PUBLICATIONS OUTLETS

Journals of Parallel Computing

Papers containing strong parallel computing component will be considered for publication in a special peer-reviewed issue of the Parallel Computing journal (PARCO). The guest editors of the special issue are C. Bekas, P. D'Ambra, A. Grama, Y. Saad and P. Yanev.

The deadline for paper submissions is the 30th of July 2008.

For further information please contact: Costas Bekas (E-mail: BEK@zurich.ibm.com).

Journals of Applied Numerical Mathematics

Papers containing strong numerical linear algebra component will be considered for publication in a special peer-reviewed issue of the Applied Numerical Mathematics journal (APNUM). The guest editors of the special issue are E. Kontoghiorghes, B. Philippe and A. Sameh.

The deadline for paper submissions is the 15th of September 2008.

For further information please contact: Erricos J. Kontoghiorghes: (E-mail: erricos@dcs.bbk.ac.uk).

Papers will go through the usual review procedures and will be accepted or rejected based on the recommendations of the editors and referees. However, the review process will be streamlined in every way possible to facilitate the timely publication of the papers. As always, papers will be considered for publication under the assumption that they contain original unpublished work and that they are not being submitted for publication elsewhere.

Contents

Friday, 20.06.2008 11:20-12:15 Room: GGA Plenary talk 1

A parallel GMRES method preconditioned by a multiplicative Schwarz iteration

Speaker: **Bernard Philippe, INRIA Rennes, France** Chair: Y. Saad

Domain decomposition provides a class of methods suitable for the solutions of linear or nonlinear systems of equations arising from the discretization of partial differential equations. For linear problems, domain decomposition methods are often used as preconditioner for Krylov subspace iterations. Traditionally, there are two classes of iterative method which derive from domain decomposition with overlap: say Additive Schwarz and Multiplicative Schwarz. When using those two methods as solvers, the convergence rates are very slow and the convergence is just proved for symmetric positive definite matrices and M-matrix. For that reason, the particular interest of Schwarz methods is as preconditioner of Krylov subspace methods. The additive version is usually preferred because it is easily implemented on a parallel computer although it is usually a less efficient preconditioner than its multiplicative version. The challenge of this work was to derive a fully automatic parallel GMRES method preconditioned through a Multiplicative Schwarz iteration based on algebraic domain decomposition. For that purpose the following results have been obtained:

- Construction of an automatic 1D partitioner of a sparse matrix,
- Derivation of an explicit expression of the preconditioner,
- Parallel pipeline to build a basis of the Krylov subspace which is then orthogonalized,
- Control of the dimension of the basis trough an estimation of the involved roundoff errors,
- Design of a code in the PETSc format.

Numerical experiments illustrate the results.

Saturday, 21.06.2008 17:20-18:15 Room: GGA Plenary talk 2

Exploiting nonnegativity in matrix and tensor factorizations for improved text mining

Speaker: **Michael Berry, University of Tennessee, Knoxville, USA** Chair: A. Sameh

Automated approaches for the identification and clustering of semantic features or topics are highly desired for text mining applications. Using low rank nonnegative matrix factorizations (NNMFs) to retain natural data nonnegativity, one can eliminate subtractive basis vector and encoding calculations present in techniques such as principal component analysis for semantic feature abstraction. Moving beyond two-way factorizations, we demonstrate how nonnegative tensor factorizations (NNTFs) can be used to capture temporal and semantic proximity and thereby enable the tracking of both targeted and latent (previously unknown) discussions or communication patterns. Demonstrations of NNMF and NNTF algorithms for topic (or discussion) detection and tracking using the Enron Email Collection and documents from the Airline Safety Reporting System (ASRS) are provided. Recent with alternative (one and infinity) norm formulations of the objective functions associated with NNMF factorization will also be presented.

Friday 20.06.2008 14:00-16:00 Parallel Session G

PS07 Room: GGA ROBUST MULTILEVEL METHODS AND PARALLEL ALGORITHMS - 1 Chair: Johannes Kraus

#25: **Compatible relaxation in parallel algebraic multigrid**

Presenter: Ulrike Meier Yang@Lawrence Livermore National Laboratory, USA
Co-authors: Robert Falgout, James Brannick

The development of high performance, massively parallel computers and the increasing demands of computationally challenging applications have necessitated the development of scalable solvers and preconditioners. One of the most effective ways to achieve scalability is the use of multigrid or multilevel techniques. Algebraic multigrid (AMG) can be a very effective algorithm for solving large problems on unstructured grids. Its efficiency depends strongly on the choice of its components, particularly the coarsening algorithm and the construction of the interpolation operator. Compatible relaxation, a fairly recently developed coarsening algorithm, uses various smoothing steps to identify good candidates for coarse points, i.e. points where the smooth error is large. This approach often leads to grids with lower complexities as well as a smaller number of grid levels. The implementation of this approach on a parallel computer requires parallel smoothers as well as a parallel independent set algorithm to determine good candidates. This presentation investigates the issues mentioned above, describes the resulting algorithms, and presents numerical results obtained on a massively parallel computer.

#19: **A scalable multi-level preconditioner for matrix-free micro-finite element analysis of human bone structures**

Presenter: Cyril Flaig@ETH Zurich, Switzerland
Co-authors: Peter Arbenz, Costas Bekas, Harry van Lenthe, Uche Mennel, Ralph Muller, Marzio Sala

The recent advances in microarchitectural bone imaging are disclosing the possibility to assess both the apparent density and the trabecular microstructure of intact bones in a single measurement. Coupling these imaging possibilities with microstructural finite element (micro-FE) analysis offers a powerful tool to improve bone stiffness and strength assessment for individual fracture risk prediction. Many elements are needed to accurately represent the intricate microarchitectural structure of bone; hence, the resulting micro-FE models possess a very large number of degrees of freedom. In order to be solved quickly and reliably on state-of-the-art parallel computers, the micro-FE analyses require advanced solution techniques. For the solution of the resulting linear system we applied the conjugate gradient algorithm, preconditioned by an aggregation-based multigrid method. We introduce a variant of the preconditioner that does not need assembling the system matrix but uses element-by-element techniques to build the multilevel hierarchy. The preconditioner exploits the voxel approach that is common in bone structure analysis, it has modest memory requirements, while being at the same time robust and scalable. Using the proposed methods, problems of size up to 1.5 billion dofs have been solved on massively parallel system like CRAY XT3 and IBM BlueGene.

#27: **On the strength of nodal dependence in AMG for vector-field problems**

Presenter: Erwin Karer@Austrian Academy of Science, Austria
Co-authors: Johannes Kraus

An important task in the classical AMG framework is the determination of the strength of connectivity of 'algebraic' vertices. Despite to the case of scalar problems it is not quite obvious how to measure the nodal dependence for vector-field problems. Based on a concept of so-called computational molecules (introduced by J. Kraus in 2006) we study different strategies with a main focus on linear elasticity problems. A natural measure is defined via the local CBS constant associated with the angle between the two subspaces spanned by the basis functions corresponding to the respective algebraic vertices. The presented numerical results provide insight to the advantages and disadvantages of various approaches.

#24: **An algebraic multigrid (AMG) solver for a finite element (FEM) discretization of the Stokes/Navier-Stokes system on hybrid meshes and its parallelization**

Presenter: Huidong Yang@Johannes Kepler University Linz, Austria
Co-authors: Walter Zulehner

In this talk, we will present an algebraic multigrid method for a finite element discretization of the Stokes/Navier-Stokes equations. The method we used is mainly based on Wabro's previous work on AMG solvers for the incompressible Navier-Stokes problem. The standard stabilized P1-P1 element has been extended to hybrid meshes containing several different elements types. Different smoothers for the AMG solver applied to the saddle point problem will be discussed. We will present a simple parallel implementation for the Stokes problem using MPI communication Library Toolbox (by Manfred Liebmann). Finally, some numerical results will be shown.

PS11 Room: GPA KRYLOV SPACE METHODS AND APPLICATIONS Chair: Mario Arioli

#47: **On the numerical behavior of Simpler GMRES and GCR**

Presenter: Pavel Jiranek@Technical University of Liberec, Czech Republic
Co-authors: Miroslav Rozloznik, Martin H. Gutknecht

We analyze the numerical behavior of several minimum residual methods, which are mathematically equivalent to the GMRES method. Two main approaches are compared: the one that computes the approximate solution in terms of a Krylov space basis from an upper triangular linear system for the coordinates, and the one where the approximate solutions are updated with a simple recursion formula. We show that a different choice of the basis can significantly influence the numerical behavior of the resulting implementation. While Simpler GMRES and ORTHODIR are less stable due to the ill-conditioning of the basis used, the residual basis is well-conditioned as long as we have a reasonable residual norm decrease. These results lead to a new implementation, which is conditionally backward stable, and, in a sense they explain the experimentally observed fact that the GCR (or full ORTHOMIN) method delivers very accurate approximate solutions when it converges fast enough without stagnation.

#65: **Backward stability of FGMRES**

Presenter: Mario Arioli@STFC-RAL, UK
Co-authors: Iain Duff

We are concerned with the solution of large sparse systems $Ax = b$ with A nonsingular but ill-conditioned, using a combination of direct and iterative methods in order to compute an accurate solution x. We will give a review of stability results for preconditioned GMRES and for Flexible GMRES (FGMRES) methods. In particular, we will focus on the backward stability of FGMRES when the preconditioning process is performed using a direct solver based on Gaussian factorization specialized for sparse matrices. In the case when the classical Iterative Refinement(IR) is not converging FGMRES can recover full backward stability. We show how the selective use of double precision post-processing can enable solutions with a backward error (scaled residual) of double precision accuracy (around 1.0D-16) even when the factorization is computed in single precision.

#64: **On the use of mixed precision for the fast and robust solution of sparse symmetric linear systems.**

Presenter: Jonathan Hogg@RAL, UK
Co-authors: Jennifer Scott

The main bottleneck for emerging computer architectures is memory bandwidth. Through the use of single precision arithmetic, the amount of data moved around within a sparse direct solver can be approximately halved. However, the cost of this is a potential loss of accuracy in the solution of the linear systems. Double precision iterative methods preconditioned by a single precision factorization can enable the recovery of high precision solutions more quickly than a sparse direct solver run using double precision arithmetic. In particular, if iterative refinement fails FGMRES can be used with an expectation of regaining precision in many cases. We report on our investigations into the practical success of these methods, providing a basis for the heuristics used in a new code HSL_MA79 - a fast, robust mixed precision sparse symmetric solver that will be included in the mathematical software library HSL. Numerical results for a wide range of problems from practical applications obtained are presented.

#68: **Block algorithms for computing PageRank by sites**

Presenter: Saint-Jean A. O. Djungu@Faculte Polytechnique de Mons, Belgium
Co-authors: Pierre Manneback

The current size of the Web, estimated at several billion pages, slows down the progression of the crawlers by significantly increasing the necessary time for the completion of one exploration's cycle. It implies a considerable time for the index base's update, as well as for the PageRank calculation. In this contribution, we propose to exploit the natural Web decomposition into blocks (hosts, domains, repertories, etc) for approximating the PageRank vector. One advantage is that the index base's update can be performed without requiring the crawling of the whole Web. We have designed two parallel algorithms, ParBlowRank and SpeedSiteRank, for computing distributed block PageRanks. These algorithms rely on classical iterative matrix algorithms. They allow the transformation of the PageRank problem into the solving of a linear system defined by a M-matrix. We present in detail these algorithms and their complexity. Experimentations carried out on a cluster showed the effectiveness of these algorithms. They offer a very attractive speed while preserving the top-*k* pages classification of the classical Pagerank algorithm.

Friday 20.06.2008 16:25-18:30 Parallel Session H

PS03 Room: GGA PARALLEL PRECONDITIONERS Chair: Ahmed Sameh

#17: **Parallel banded preconditioners for non-symmetric linear system solvers**

Presenter: . . . . . Ananth Grama@Purdue University, USA
Co-authors: . . . . . Murat Manguoglu, Mehmet Koyuturk, Ahmed Sameh

The emergence of multicore architectures and highly scalable platforms motivates novel algorithms and techniques that emphasize concurrency and are tolerant to deep memory hierarchies, as opposed to minimizing raw FLOP counts. In this talk, we present a novel class of banded preconditioners and solvers that have excellent concurrency characteristics, while delivering high aggregate FLOP counts. These methods are shown to achieve excellent scalability on various architectures. In this talk, we present (i) reordering schemes that allow extraction of a narrow central band that can be used as a banded preconditioner, (ii) a parallel solver, Spike used as the inner banded solver, and (iii) a parallel iterative outer solver. Our results demonstrate that (i) banded preconditioners are more robust than the a broad class of incomplete factorization based methods, (ii) they deliver better convergence results (iteration counts) than incomplete factorization methods, (iii) they deliver higher processor performance, resulting is faster time to solution, and (iv) they deliver excellent parallel performance and scalability on diverse parallel platforms. We also show how we can derive models of performance that characterize the performance of our solvers accurately. We demonstrate these results experimentally on a large class of problems selected from diverse application domains.

#10: **Block incomplete LU preconditioning on modern hardware platforms**

Presenter: . . . . . Achim Basermann@NEC Europe Ltd, Germany
Co-authors: . . . . . Jens Georg Schmidt

NLE-IT's PILUTS (Parallel Incomplete LU with Threshold preconditioned Solvers) library includes parallel iterative sparse solvers for real symmetric positive definite, general real symmetric and real non-symmetric matrices preconditioned by scaling methods, symmetric or non-symmetric incomplete block factorisations with threshold and Distributed Schur Complement algorithms. On modern hardware platforms like PC clusters or General Purpose Graphical Processing Units (GPGPUs), the incomplete factorisations are the most crucial operations for performance. Block versions of these operations are promising since they avoid indirect addressing and permit the use of optimised BLAS routines. In the recent years the performance of GPGPUs has increased much faster than the performance of standard CPUs. While programming general applications on a GPU was quite awkward and complicated in the past, the new generation of GPGPUs comes with programming tools like language extensions (CUDA from NVIDIA, Ct from Intel) or easy-to-use libraries (cuFFT, cuBLAS) which tremendously reduce the effort for the average user to utilize the compute power of these devices. Tests with several BLAS routines on a 64bit AMD Opteron CPU, combined with an NVIDIA 8800 GTS GPU, showed that the BLAS routines perform about ten times faster on the NVIDIA GPU than on the Opteron CPU. Here, in particular numerical and performance investigations of variable block incomplete factorisations with automatic block detection (developed by Yousef SaadÂ´s group) will be presented in the context of the PILUTS library on PC clusters and GPGPUs.

#9: **Algebraic preconditioners for parallel hybrid solvers**

Presenter: . . . . . Azzam Haidar@INPT-IRIT, France
Co-authors: . . . . . Luc Giraud, Stephane Pralet

In this work we investigate the parallel scalability of variants of additive Schwarz preconditioners for three dimensional non-overlapping domain decomposition methods. To alleviate the computational cost, both in terms of memory and floating-point complexity, we investigate variants based on a sparse approximation. The robustness of the preconditioners is illustrated on a set of linear systems arising from the finite element discretization of academic convection-diffusion problems (un-symmetric matrices), and from real-life structural mechanical problems (symmetric indefinite matrices). Parallel experiments on up to a thousand processors on some problems will be presented. The efficiency from a numerical and parallel performance view point are studied on problem ranging from a few hundred thousands unknowns up-to a few tens of millions.

#22: **Parallel multilevel ILU preconditioners**

Presenter: . . . . . Alberto Fco. Martin-Huertas@University Jaume I, Spain
Co-authors: . . . . . Jose I. Aliaga, E Quintana-Orti, Matthias Bollhoefer

This contribution reviews the design of a parallel algorithm that computes multilevel ILU preconditioners on shared-memory platforms. In order to reveal the parallelism inherent to sparse factorizations, we employ a MLND re-ordering of the graph underlying the linear system $Ax = b$, together with the corresponding elimination tree. From this tree, we construct a coarse-grain structure, called task-dependency tree, which splits the computation into (a modest number of) tasks and identifies the dependencies between them. Our approach then employs a dynamic scheduling strategy in order to balance the workload among threads. This strategy maps tasks to threads during execution while preserving the order imposed by the dependencies. The preconditioner thus computed aims at accelerating the convergence of a Krylov subspace method for the solution of the (preconditioned) linear system. Besides, the task tree presented here can also be used to split the computations involved in the preconditioned solver, yielding a parallel version of the solver. Experimental results are reported using an OpenMP implementation of our parallel preconditioner

on a shared-memory platform with 16 processors.

#13: **A parallel hybrid banded solver and its generalization to sparse linear systems**

Presenter: . . . . . Ahmed Sameh@Purdue University, USA
Co-authors: . . . . . Maxim Naumov

A new parallel algorithm for the solution of banded linear systems is presented. The scheme tears the coefficient matrix into several overlapped independent blocks in which the size of the overlap is equal to the system's band- width. A corresponding splitting of the right-hand side is also provided. The resulting independent, and of smaller size, linear systems are solved under the constraint that the solutions corresponding to the overlap regions are identical. This results in a linear system whose size is proportional to the sum of the over- lap regions which we refer to as the *balance* system. We propose a solution strategy that does not require obtaining this *balance* system explicitly. Once the balance system is solved, retrieving the rest of the solution can be realized with almost perfect parallelism. The generalization of the algorithm to sparse linear systems is also proposed. It is based on a recursive approach and takes advantage of sparse LU- and UL-factorizations. Numerical experiments will be presented.

PS04 Room: GPA PARALLEL COMBINATORIAL SCIENTIFIC COMPUTING Chair: C. Bekas

#80: **Efficient and scalable parallel graph partitioning**

Presenter: . . . . . Jun-Ho Her@ENSEIRB, LaBRI & INRIA Bordeaux - Sud-Ouest, France
Co-authors: . . . . . Francois Pellegrini

The realization of efficient parallel graph partitioners requires the parallelization of the multi-level framework which is commonly used in sequential partitioners to improve quality and speed. While parallel matching algorithms are now efficient and un-biased enough to yield coarsened graphs of good quality, the local optimization algorithms used in the refinement step of the uncoarsening process are still an issue. This talk will present the results obtained to date in the PT-Scotch project regarding k-way graph partitioning and parallel static mapping. We will show how parallel diffusive method can advantageously replace classical (and purely sequential) Fiduccia-Mattheyses-like algorithms for local optimization, as well as the specific algorithmic problems posed by static mapping.

#44: **A hybrid two-dimensional method for sparse matrix partitioning**

Presenter: . . . . . Rob Bisseling@Utrecht University, Netherlands
Co-authors: . . . . . Tristan van Leeuwen, Umit Catalyurek

Two-dimensional partitioning of sparse matrices has been shown to be highly effective in distributing the work and minimising communication for parallel computations in a wide range of applications. We present a hybrid method which combines two types of 2D methods: (i) the fine-grained approach by Catalyurek and Aykanat which recursively splits the current subset of nonzero matrix elements looking at individual elements, thereby allowing for the most general solution; (ii) the Mondriaan approach by Vastenhouw and Bisseling which recursively splits the sparse matrix into two submatrices, trying both row and column directions and each time choosing the best. Both methods use a multilevel hypergraph partitioner as their splitting engine. This greedily minimises the exact metric of communication volume. We present results for a variety of sparse matrices, including web link matrices originating in PageRank computations. We use both the PaToH and Mondriaan hypergraph partitioners in our numerical experiments. Furthermore, we present several new features incorporated in version 2 of the Mondriaan package, such as a visualisation interface with Matlab, specialised partitioning of fine-grained hypergraphs, and dynamic adjustment of the load balance as the partitioning proceeds.

#42: **Hypergraph partitioning for sparse linear systems: a case study with discontinuous PDEs**

Presenter: . . . . . Masha Sosonkina@Iowa State University, USA
Co-authors: . . . . . Yousef Saad

Beyond managing the parallel overhead incurred in performing the matrix-vector multiplications and preconditioning operations, one must partition the problem carefully at the outset, so that fast iterative convergence is attained. Graph partitioners typically consider the nonzero structure of a matrix to balance the number of unknowns and to decrease communication volume among the parts. However, it is known that this objective is not sufficient for many realistic practical problems, such as when the coefficients in the underlying Partial Differential Equation (PDE) are discontinuous or when modeling complex multi-physics phenomena. Hence, in the partitioning process, it is desirable to consider certain matrix numerical properties or physical information about the problem in addition to the matrix adjacency graph. The present work exploits the flexibility of hypergraph partitioning to incorporate numerical information. We propose a modification of a standard hypergraph model and several weight schemes that use problem description and enable hypergraph partitioning to target good iterative convergence. Numerical experiments are performed on two- and three-dimensional elliptic problems with discontinuous coefficients on rectangular meshes. We compare the proposed approaches with standard hypergraph models by examining the resulting number of iterations and time to converge.

#14: **Very large scale graph partitioning problems in micro finite element analyses of human bone structure**

Presenter: ..... Costas Bekas@IBM Research, Switzerland
Co-authors: ..... Alessandro Curioni, Peter Arbenz

High-resolution in vivo peripheral quantitative computed tomography provides detailed information on bone structure. The analysis of bone density (using other, more commonly available technology) is today's approach of predicting bone strength and fracture risk in diseases like osteoporosis that is, according to the WHO, second only to cardiovascular disease as a leading health care problem. Coupling recent imaging capabilities with microstructural finite element (micro-FE) analysis offers a powerful means to determine bone stiffness and strength. The intricate microarchitectural structure of bone entails that these micro-FE models possess a very large number of degrees of freedom. The computational model is based on linear elasticity and the linear solver is a matrix-free conjugate gradient algorithm preconditioned by aggregation-based AMG. The method is implemented in a software package called ParFE which is parallelized using MPI, and is based on the public-domain software Trilinos, ParMETIS and HDF5. We targeted the IBM BG/L Supercomputer in order to exploit its excellent scale-out potential. We conducted the largest simulation of its kind so far (1.5 billions dof) and calculated the effective strain of a vertebral bone specimen. We will discuss that the major bottleneck in extreme scale-out of this application (and micro-FE applications in general) is in graph repartitioning: a) The intricate structure of bone causes significantly imbalanced partitions that have a strong negative effect when thousands of processors are used. b) The scalability of parallel graph partitioning tools (such as ParMETIS, Zoltan and others) on tens of thousands of processors appears to be a formidable task. Clearly, in anticipation of the petaflop machines, efficient mapping of applications on millions of processing elements will require next generation algorithms and efficient mapping models.

PS09 Room: GB1 LARGE-SCALE SPARSE MATRIX COMPUTATIONS Chair: Pasqua D'Ambra and Daniela di Serafino

#28: **A matrix partitioning interface to PaToH in MATLAB**

Presenter: ..... Bora Ucar@CERFACS, France
Co-authors: ..... C. Aykanat, U.V. Catalyurek

Recently, we have developed a MATLAB interface to PaToH. The interface offers sparse matrix partitioning tools built according to the hypergraph models proposed in a series of papers by Catalyurek and Aykanat. In addition to the matrix partitioning methods, the interface also offers partitioning visualization tools and supplementary codes to measure the quality of a given matrix partition. Using the interface, we have performed extensive tests on the matrices from University of Florida Sparse Matrix collection. The presentation will cover essential parts of the PaToH's MATLAB interface and our observations deduced from the extensive tests.

#30: **MLD2P4: A package of scalable algebraic multilevel Schwarz preconditioners**

Presenter: ..... Salvatore Filippone@University of Rome "Tor Vergata", Italy
Co-authors: ..... Pasqua D'Ambra, Daniela di Serafino

The solution of sparse linear systems arises in the numerical simulation of many scientific and engineering applications. Obtaining robustness of convergence of iterative solvers across a range of difficult problems is a challenge; the use of parallel computing architectures adds to this challenge because the convergence properties of an iterative method/preconditioner pair are often dependent on the degree of parallelism. Multilevel preconditioners are appropriate tools to enhance the convergence robustness of Krylov subspace solvers in a parallel setting, and thus may be used in applications tackling challenging problems. MLD2P4 is a preconditioner package specifically designed to provide scalable and easy-to-use multilevel preconditioners in the context of the PSBLAS parallel computing framework. MLD2P4 enables the user to easily specify different aspects of a generic algebraic multilevel Schwarz preconditioner, thus allowing the user to search for the best method/preconditioner combination for the problem at hand. The package has been designed employing object-oriented techniques, using Fortran 95 and MPI, with interfaces to additional external libraries such as UMFPACK, SuperLU and SuperLU_Dist, and is easily extensible. Experimental results from model problems and from a demanding CFD application have shown the effectiveness of MLD2P4.

#32: **Towards a parallel analysis phase for a multifrontal sparse solver.**

Presenter: ..... Alfredo Buttari@LIP, ENS-Lyon, France
Co-authors: ..... Patrick Amestoy, Jean-Yves L'Excellent

The analysis phase in a direct, multifrontal solver for sparse matrices is in charge of computing a pivot ordering that reduces the fill-in at factorization time, building an elimination tree that allows exploitation of parallelism, estimating the structure of the factors by means of a symbolic factorization and performing supervariables detection and nodes amalgamation for exploitation of Level-3 BLAS operations during the system factorization. Despite not being critical for performance purposes since no floating point operation is performed at this phase, the problem analysis may pose serious memory requirements for large scale problems. This work focuses on the parallelization of the analysis phase in the MUMPS direct, multifrontal solver package aiming at reducing the memory consumption. Some performance improvements may be expected as well though scalability may be not optimal. The proposed approach is based on the usage of parallel graph partitioning tools like PT-SCOTCH or ParMETIS for the computation of the pivot ordering and elimination tree. The distributed symbolic factorization is achieved by means of a method which is derived from an Approximate Minimum Degree algorithm and is based on the usage of quotient graphs.

#37: **Communication avoiding Gaussian elimination**

Presenter: ..... Laura Grigori@INRIA, France
Co-authors: ..... James W. Demmel, Hua Xiang

We present a Communication Avoiding algorithm for the LU factorization (CALU) of dense and sparse matrices distributed in a two-dimensional cyclic layout. The algorithm is based on a new pivoting strategy, referred to as ca-pivoting, that is stable in practice. This strategy leads to significant decrease in the number of messages exchanged during the factorization of a block-column relatively to conventional algorithms, and CALU overcomes the latency bottleneck of the LU factorization as in current implementations like ScaLAPACK for the dense case. In particular CALU is found to be significantly faster than PDGETRF in the cases when the latency time is an important factor of the overall time, as for example when a small matrix is executed on large number of processors. The parallel algorithm for the sparse LU factorization is based on HUND, a hypergraph-based unsymmetric nested dissection ordering algorithm for reducing the fill-in, developed by L. Grigori, E. Boman, S. Donfack and T. Davis. HUND is suitable for performing Gaussian elimination in parallel, with partial pivoting, and provides a robust reordering algorithm, in the sense that it is close to, or the best (often within 10%) of all the other methods on our matrix test set.

#58: **Programming models and techniques for sparse matrix kernels on multicore platforms**

Presenter: ..... Richard Vuduc@Georgia Institute of Technology, USA
Co-authors: ..... Sam Williams, Lenny Oliker, John Shalf, James Demmel, Katherine Yelick

I present recent work for optimizing the performance of sparse matrix-vector multiplication for current multicore systems. I emphasize (1) alternative programming models, including a comparison of *off-the-shelf* hybrid OpenMP+MPI against multicore-specific optimizations, and (b) the design and implementation of OSKI, an autotuned sparse BLAS-like library, discussing how its internal data structure representation supports optimizing transformations relevant to multicore.

Saturday 21.06.2008 09:00-11:00 Parallel Session I

PS13 Room: GGA ROBUST MULTILEVEL METHODS AND PARALLEL ALGORITHMS - 2 Chair: Svetozar Margenov

#23: **Hierarchical multilevel splittings for discontinuous Galerkin approximations of elliptic problems with high-frequency-high-contrast coefficients**

Presenter: ..... Johannes Kraus@Austrian Academy of Sciences, Austria
Co-authors: ..... Ivan Georgiev, Svetozar Margenov

In this talk we present a new framework for multilevel preconditioning of large sparse systems of linear algebraic equations arising from discontinuous Galerkin discretization of second-order elliptic partial differential equations (PDE). Though our focus is on a particular family of rotated bilinear finite elements in two space dimensions (2D) here, the proposed rather general setting is neither limited to this particular choice of elements nor to 2D problems. Two basic problems are studied: the scalar elliptic equation and the Lame system of linear elasticity. The presented results are in the spirit of algebraic multilevel iteration (AMLI) methods. The innovation concerns the construction of robust methods for problems with large jumps in the PDE coefficients that can only be resolved on the finest finite element mesh, which is assumed to be a result of multilevel refinement of a given coarse mesh. In the well-established theory of hierarchical basis multilevel methods one basic assumption is that the PDE coefficients are smooth functions on the elements of the coarsest mesh partition. The newly developed generalized hierarchical basis, however, shows a well expressed robustness of the related multilevel splitting with respect to coefficient jumps between elements on the finest mesh.

#26: **Multilevel preconditioning in H(div) and applications to a posteriori error estimates**

Presenter: ..... Satyendra Tomar@Austrian Academy of Sciences, Austria
Co-authors: ..... Johannes Kraus

In this talk we will present an algebraic multilevel iterative method for solving linear systems arising from the finite element discretization of certain boundary value problems that have their weak formulation in the space H(div). Apart from the primary interest in an efficient numerical solution of practical problems, e.g., in continuum mechanics, such iterative schemes become important when using advanced (functional-type) a posteriori error estimates developed by Repin (e.g. 2003). As recently observed, when using these estimates for discontinuous Galerkin approximations of elliptic problems (see Lazarov, Repin, Tomar, 2007), the computation of such guaranteed a posteriori error estimates can be even more expensive than computing the solution of the original problem, and thus, demands fast iterative solvers. Preconditioners for such linear systems within the framework of domain decomposition and multigrid techniques have been studied by several authors, see e.g. Arnold, Falk and Winther (1997, 2000) and Hiptmair (1997). Our work is based on a different approach, namely, algebraic multilevel iterative methods.

#35: **Parallel Schwarz type solvers for THM modelling**

Presenter: ..... Radim Blaheta@Institute of Geonics AS CR, Czech Republic
Co-authors: ..... Petr Byczanski, Roman Kohut, Jiri Stary

The contribution concerns modelling thermal, mechanical and hydrogeological (THM) processes in rocks, which are important for assessment many geotechnical projects. This modelling can be extremely demanding, when one consider large time and spatial scale as well as some spatial details and want to consider all the processes with some of their couplings. It gives a motivation for investigation of efficient numerical methods suitable for parallel computing. We use Schwarz type technique based on overlapping domain decomposition. It is applied separately to T, H, M processes. For mechanical problems, we use Schwarz-type methods with domain decomposition and additionally a coarse grid space constructed algebraically by aggregation. For evolution heat transfer problems solved by a time stepping algorithm, we are able to show, that for time steps appearing in adaptive time stepping schemes, the Schwarz-type methods are very efficient even without any additional coarse space contribution. Further results are obtained for solving Darcy flow problems by mixed finite elements. Then the finite element matrices are not further positive definite and application of the Schwarz method is not direct and there are several ways how to overcome this situation. We shall show one of such techniques with grad-div preconditioner.

#36: **Micro FEM analysis of geocomposites**

Presenter: ..... Jiri Stary@Institute of Geonics AS CR, Czech Republic
Co-authors: ..... Radim Blaheta, Roman Kohut, Alexej Kolcun, Svetozar Margenov

This contribution concerns the analysis of mechanical behaviour of geocomposites arising from injection of polyurethane resin into coal environment. Such procedure can be used to reinforce coal pillars during mining. The questions to be solved are: what are the upscaled elastic properties of coal geocomposites, how sensitive are these properties on the quality of filling of the coal matrix by the polyurethane, etc. The homogenized properties are determined by numerical upscaling. To this end, the structure of a sample is digitalized by X-ray CT and the upscaled properties are obtained via numerical implementation of loading tests. We shall consider strain and stress driven tests implemented numerically by means of the FEM analysis of the microstructure. The standard conforming linear tetrahedral finite elements are used in this implementation. The contribution addresses also two kinds of parallel iterative solution methods for the FEM system. They are CG with parallel displacement decomposition – MIC(0) factorization preconditioning and CG with two level Schwarz type preconditioning methods. In the latter case we use coarse subproblems created by aggregation. Finally, we compare the FE solution with conforming linear tetrahedral FE with the nonconforming

rotated trilinear one.

#21: **Parallel PCG algorithms for numerical homogenization of voxel structures**

Presenter: ..... Yavor Vutov@Bulgarian Academy of Scienses, Bulgaria
Co-authors: ..... Radim Blaheta, Svetozar Margenov

Numerical homogenization technique is used for upscaling of the linear elasticity tensor of strongly heterogeneous composite materials. The implemented 3D algorithm is described in terms of six auxiliary elastic problems for the reference element. Rotated trilinear Rannacher-Turek finite elements are used for numerical solution of the involved subproblems. The PCG method is implemented for efficient solution of the arising large-scale systems with sparse, symmetric positive definite matrices. The performance of two parallel solvers is studied. Both are based on displacement decomposition. The first one uses modified incomplete Cholesky factorization MIC(0) and the other - algebraic multigrid. The comparative analysis includes the parallel times as well as the related number of PCG iterations. Numerical upscaling results for two representative test problems are shown: (a) trabecular bone tissue, taking into account the elastic response of both, solid and soft (fluid) phases, and (b) geocomposite: polyurethane is injected in a coal matrix. In both cases, high resolution computer tomography images are used to get the voxel microstructure of the composite materials.

PS14 Room: GPA SVD AND JACOBI METHODS Chair: Gabriel Oksa

#16: **Parallel versions of one-sided bidiagonalization**

Presenter: ..... Nela Bosner@University of Zagreb, Croatia
Co-authors: ..... Jesse L. Barlow, Zlatko Drmac

We observe an algorithm for computing the complete singular value decomposition (SVD) of a matrix based on bidiagonalization as the first step. Recently proposed a new bidiagonalization algorithm, the so-called one-sided bidiagonalization, which is well adapted to parallel computation. A numerically stable version of this algorithm has already been proposed. When incorporated into singular value decomposition software, the numerically stable one-sided bidiagonalization is faster than the corresponding ScaLAPACK routine in most cases. Actually, we present here several versions of the parallel one-sided bidiagonalization: the original algorithm implemented by PBLAS and ScaLAPACK routines, the block parallel version, and a modification of the parallel bidiagonalization which reduces communications events. All this versions were tested on two platforms: on a cluster with slow interprocessor communication consisting of 20 computers connected in a local 1Gb network; and on a cluster with fast interprocessor communication. Timings and analysis are presented for all the tests.

#57: **Parallel implementations of the one–sided indefinite block Jacobi methods**

Presenter: ..... Sanja Singer@University of Zagreb, Croatia
Co-authors: ..... Sasa Singer, Vjeran Hari, Kresimir Bokulic, Davor Davidovic, Marijan Juresic, Aleksandar Uscumlic

The indefinite one-sided Jacobi algorithm for computing the hyperbolic singular value decomposition of a rectangular matrix is known to be accurate in the relative sense. However, it is far too slow for serial computation, when compared with some other accurate algorithms. It is well-known that the Jacobi eigenvalue algorithm is an ideal for parallelization, but, if we respect the modern design of computer clusters with multi-level memory hierarchy, there appear many open questions how to implement it in an efficient way. This memory hierarchy structure can be efficiently exploited by blocking of the algorithm. We describe two different approaches for obtaining block generalizations of the simple indefinite Jacobi algorithm: a block oriented and a full block approach. Then we construct a parallel, locally fast Jacobi algorithm which respects a three-level memory hierarchy consisting of: distant data, locally stored data, and local cache memory. This is done by using two levels of blocking. The outer level of blocking uses the modulus block strategy for data independence to achieve proper load balancing between different processors. The inner level of blocking uses the cyclic block strategy to reuse data that is already in the local cache memory of each processor. Since both of the above approaches can be used for blocking at each level, we obtain many variants of the algorithm. We describe implementation details that affect the speedup, and present a series of numerical results that show an almost ideal speedup of properly implemented versions of the indefinite block Jacobi algorithm.

#52: **Convergence to diagonal form of general Jacobi-type processes**

Presenter: ..... Vjeran Hari@University of Zagreb, Croatia
Co-authors: ..... Vida Zadelj-Martic

The standard fast SVD solvers for a general matrix first reduce it to bidiagonal form. This initial reduction can deteriorate the relative accuracy of the smallest singular values even if the initial matrix is well-behaved for accurate SVD computation. For such matrices, the one-sided Jacobi method has proved to be very accurate and fast, especially on parallel computers. Similar arguments holds for the solvers of other matrix problems: hyperbolic SVD, eigenvalue problem for non-Hermitian matrices, generalized Hermitian eigenvalue problem and generalized singular value problem. The global convergence of one-sided Jacobi-type processes reduces to the convergence of their two-sided counterparts. This report considers convergence to diagonal form of a general two-sided Jacobi-type process $A^{(k+1)} = [P^{(k)}]^* A^{(k)} Q^{(k)}$, $k \geq 0$, where $P^{(k)}$ and $Q^{(k)}$ are regular elementary matrices, which differ from the identity matrix in one principal submatrix. The modulus pivot strategy is assumed since it is weakly equivalent to the most common cyclic strategies for sequential and parallel processing. The technique uses the theory of Jacobi annihilators

which is due to Henrici and Zimmermann. The main result provides sufficient conditions for the convergence of such a process to diagonal form. Recent research includes the block Jacobi type processes.

#48: **Optimal data distribution in the preconditioned parallel two-sided block Jacobi SVD algorithm**

Presenter: . . . . . Gabriel Oksa@Slovak Academy of Sciences, Slovakia
Co-authors: . . . . . Martin Becka, Laura Grigori, Marian Vajtersic

Recently, we have proposed the preconditioned version of the parallel two-sided block Jacobi SVD algorithm. When computing all singular values together with all right and left singular vectors of a rectangular matrix A, the preconditioning step consists of the parallel computation of the QR decomposition with column pivoting followed by the optional LQ factorization of the R-factor. The parallel two-sided block Jacobi method is then applied to the R-factor (or L-factor). The purpose of preconditioning is to concentrate the Frobenius norm near the matrix diagonal so that the Jacobi method needs substantially less parallel steps for convergence than in the case without preconditioning. However, to perform optimally, the parallel QR (or LQ) decomposition and the parallel two-sided block Jacobi method need various data layouts. Having p processors, the Jacobi algorithm performs very well when implemented on the processor grid $1 \times p$ with matrix block columns local to processor, because the majority of computations can be done locally. But this data distribution is not well suited for the column-oriented parallel QR decomposition because of a low efficiency in using the computational power. Instead, some sort of a cyclic matrix distribution on a processor grid $r \times c$ with $p = rc, , r, c > 1$, is needed so that all processors remain busy during the whole parallel QR (or LQ) factorization. We have designed, implemented and tested the preconditioned parallel two-sided block Jacobi algorithm with two various data distributions for both phases of computation. We report first test results with chosen distribution of singular values. They show that we can decrease a portion of the parallel execution time T_p occupied by preconditioning, which leads also to a decrease of T_p when compared with the case of the whole computation performed on the processor grid $1 \times p$.

Saturday 21.06.2008 11:20-13:20 Parallel Session J

PS01 Room: GGA ALGEBRAIC PRECONDITIONING OF ITERATIVE METHODS Chair: Miroslav Tuma

#8: **Multilevel preconditioning for large-scale nonconvex PDE-constrained optimization**

Presenter: Olaf Schenk@University of Basel, Switzerland

Fast nonlinear programming methods following the all-at-once approach usually employ Newton's method for solving linearized Karush-Kuhn-Tucker (KKT) systems. In nonconvex problems, the Newton direction is only guaranteed to be a descent direction if the Hessian of the Lagrange function is positive definite on the nullspace of the active constraints, otherwise some modifications to Newton's method are necessary. This condition can be verified using the signs of the KKT's eigenvalues (inertia), which are usually available from direct solvers for the arising linear saddle point problems. Iterative solvers are mandatory for very large-scale problems, but in general do not provide the inertia. We propose an algebraic multilevel preconditioning technique using maximum weighted matchings for nonconvex optimization problems to be used in interior point methods. The preconditioning approach for the symmetric indefinite Karush-Kuhn-Tucker systems is based on maximum weighted matchings and algebraic multi-level inverse-based incomplete LBL^T factorizations. The largest nonconvex optimization problem from three-dimensional PDE-constrained optimization with the multilevel preconditioning approach has more than 30 million state variables, hundred of thousands control variables with both lower and upper bound and 300 million nonzeros in the Jacobian.

#6: **Recent advances in preconditioning large-scale symmetric indefinite systems**

Presenter: Matthias Bollhoefer@TU Braunschweig, Germany

We discuss preconditioning methods for Helmholtz equations in two and three spatial dimensions with high wave number in inhomogeneous media. In particular we present an algebraic multigrid algorithm that is based on three major algorithmic components: Symmetric maximum weight matchings to increase the block diagonal dominance of the system, algebraic inverse–based pivoting to drive the algebraic coarsening process and, filtering based on selected sample frequencies derived from the wave number of the underlying Helmholtz equation. These components can be seen as an efficient black-box approach for the ill-conditioned symmetric indefinite Helmholtz matrices. We present theoretical results and demonstrate the algorithmic effectiveness through inhomogeneous model problems and two- and three dimensional seismic subsurface geology simulations.

#4: **Symmetric permutations for I-matrices to avoid small pivots during incomplete factorization**

Presenter: Jan Mayer@Universitat Karlsruhe, Germany

Incomplete LU-factorizations have been very successful as preconditioners for solving sparse linear systems iteratively. However, for unsymmetric, indefinite systems small pivots (or even zero pivots) are often very detrimental to the quality of the preconditioner. A fairly recent strategy to deal with this problem has been to permute the rows of the matrix and to scale rows and columns to produce an I-matrix, a matrix having elements of modulus one on the diagonal and elements of at most modulus one elsewhere. These matrices are generally more suited for incomplete LU-factorization. Clearly, I-matrices are preserved by symmetric permutation, i.e. by applying the same permutation to rows and columns of a matrix. In this talk, we will discuss different approaches to constructing such permutations. These strategies generally aim at improving the sparsity and diagonal dominance of an initial block. Numerical results will underline that these additional permutations greatly improve the quality of the factorization often resulting in much shorter total computation time.

#7: **On the usage of triangular preconditioner updates in matrix-free environment.**

Presenter: Jurjen Duintjer Tebbens@Czech Academy of Sciences, Czech Republic
Co-authors: Miroslav Tuma

Efficient solution of sequences of linear systems is a task arising in numerous applications in engineering and scientific computing. Depending on the linear solver and the properties of the system matrices, several techniques to share part of the computational effort throughout the sequence may be used. Our contribution considers a new black-box approximate update scheme for factorized preconditioners that was recently introduced by the authors. It is designed for general nonsymmetric linear systems solved by arbitrary iterative solvers and the basic idea is to combine an incomplete reference factorization with a Gauss-Seidel type of approximation of the difference between the current and the reference matrix. The updated factorizations may be particularly beneficial when preconditioner computations from scratch are expensive, like in matrix-free environment where the matrix has to be estimated first. In this talk we give a brief description of the basis update technique and then discuss their usage in matrix-free environment. We present, among others, a new implementation strategy which is based on mixed matrix-free/explicit triangular solves.

#3: **Frobenius norm minimization and probing**

Presenter: Thomas Huckle@TU Munchen, Germany

We present the augmented or modified SPAI method for defining preconditioners for the iterative solution of a system of linear equations $Ax = b$. This method can be seen as a generalization of the class of preconditioners derived by Frobenius norm

minimization by minimizing the more general problem

$$\min \left\| \begin{pmatrix} C \\ f^T \end{pmatrix} M - \begin{pmatrix} B \\ e^T \end{pmatrix} \right\|_F$$

for given sparsity pattern in M, e.g. with $C = A$, $B = I$, and $f^T = e^T A$. To this aim we augment the Frobenius minimization by an additional norm minimization improving the preconditioner relative to a chosen subspace. This new method is closely related to preconditioners related to weighted Frobenius norm minimization (e.g. FSAI, SPAI, SPAI with target matrix), and also to modified preconditioners like MILU or MICC (where the chosen subspace is related to the vector $(1,1,\ldots,1)$). Furthermore, this augmented Frobenius norm minimization can be seen as a regularization of the interface probing for Schur complement systems allowing any set of probing vectors. The new approach can be used to derive improved preconditioners e.g. for Schur complement problems, but also for sparsification of dense or thick matrices where we are looking for a sparse approximation of the original problem. All the computations are embarrassingly parallel. Many numerical examples, e.g. from PDE applications such as domain decomposition and Stokes problem, show that these new preconditioners often lead to faster convergence and smaller condition numbers.

PS06 Room: GPA PARALLEL DENSE NUMERICAL LINEAR ALGEBRA Chair: Peter Arbenz

#18: **The envelope method**

Presenter: ..... Beresford Parlett@ETH Zurich, Switzerland

The task is to compute orthogonal eigenvectors (without Gram-Schmidt) of symmetric tridiagonals for isolated clusters of close eigenvalues. We review an old method, the Submatrix method, and describe an extension which significantly enlarges the scope to include several mini-clusters within the given cluster. An essential feature is to find the envelope of the associated invariant subspace.

#59: **Dynamic node-scheduling of a multishift QR sweep algorithm**

Presenter: ..... Lars Karlsson@Umea University, Sweden
Co-authors: ..... Bo Kagstrom

Recent multishift QR-like algorithms for the nonsymmetric eigenvalue problem introduce level 3 operations in the bulge-chasing subalgorithm (a QR Sweep algorithm) by accumulating basic transformations inside a diagonal block window (using level 1 and 2 operations) into a small dense orthogonal matrix which is subsequently applied (by level 3 operations) to the off-diagonal blocks in the associated block row and column. Distributed memory parallelizations of multishift QR Sweep algorithms face a trade-off between using a large distribution block size to reduce communication and a small distribution block size to improve load balance within each iteration. One approach used in similar algorithms is to choose the block size to reduce communication while introducing multiple windows to improve load balance. We have applied our new experimental and minimalistic framework for dynamic node-scheduling to a QR Sweep algorithm. Without changing the structure of the code, we dynamically reorder the execution of subcomputations within the nodes and thereby exploiting the inherent load balance of a complete QR Sweep. We show how this might enable implementations where the distribution block size can be chosen to reduce communication without requiring the use of multiple windows. Our dynamic node-scheduling can also be combined with multiple windows.

#81: **Parallel factorization of band matrices**

Presenter: ..... Remon Alfredo@Universidad Jaime I, Spain
Co-authors: ..... Enrique S. Quintana-Orti, Gregorio Quintana-Orti, Robert van de Geijn

We pursue the scalable parallel implementation of the factorization of band matrices with medium to large bandwidth targeting SMP and multi-core architectures. Our approach decomposes the computation into a large number of fine-grained operations exposing a higher degree of parallelism. The FLAME methodology, combined with the support for hierarchical storage provided by FLASH, and the SuperMatrix run-time system which allows an out-of-order dynamic scheduling of operations improves the programmability of the solution. Experimental results for the Cholesky factorization of band matrices on a CC-NUMA platform with sixteen processors demonstrate the performance and scalability of the approach, which clearly outperform the parallel implementations in LAPACK based on the use of a multithreaded BLAS.

#60: **Gaussian elimination based algorithms on the GPU**

Presenter: ..... Aydin Buluc@UC Santa Barbara, USA
Co-authors: ..... John R. Gilbert

We implemented and evaluated several Gaussian elimination (GE) based algorithms on the Graphic Processing Units (GPUs). These algorithms, LU decomposition without pivoting, all-pairs shortest paths, and transitive closure, have all similar data access patterns. The impressive computational power and memory bandwidth of the GPU make it an attractive platform to run such computationally intensive algorithms. Although improvements over CPU implementations have been achieved for those algorithms in terms of raw speed, the utilization of the underlying computational resources was quite low. We implemented recursive versions of GE based algorithms that harness the power of GPUs better than existing implementations. We discovered that divide-and-conquer algorithms that use recursion map better to the GPU hardware than iterative algorithms. This is mostly due to the premature satu-

ration of bandwidth resources of the GPU when using iterative algorithms. By increasing temporal locality, our implementations run more than 10 times as fast as iterative ones when implemented on an NVIDIA 8800 Ultra. Our results, showing that algorithms that minimize data transfers perform better on the GPU, might lead to rethinking the algorithm design process for general purpose computing on GPUs.

#83: **Tridiagonalizing complex symmetric matrices**

Presenter: . . . . . Wilfried Gansterer@University of Vienna, Austria
Co-authors: . . . . . Hannes Schabauer, Andreas Gruber

We discuss methods for tridiagonalizing dense complex symmetric (non-Hermitian) matrices. This reduction process constitutes a central building block for solving dense complex symmetric eigenproblems, which arise in various applications. The standard approach for such problems employs methods for non-Hermitian problems, which cannot exploit the structural symmetry. This paper investigates the potential for improving performance through symmetry-preserving non-Hermitian tridiagonalization processes. We investigate two approaches: a splitting method, which has been discussed earlier, separates real and imaginary parts, whereas a novel non-splitting method operates on the complex matrix as a whole. Three central aspects are discussed: numerical accuracy (which is potentially compromised when using non-unitarian transformations for preserving symmetry), sequential performance, and parallelization (which has hardly been addressed so far for dense complex symmetric eigenproblems). Experimental evaluations are summarized. Numerical accuracy and execution times are compared to the standard routine for non-Hermitian eigenproblems, LAPACK/zgeev. Parallelization and scaling properties are investigated. It turns out that the splitting method has advantages in terms of numerical accuracy, whereas the non-splitting method has the potential for achieving higher performance and parallelizes better. In combination, the two methods are particularly interesting in situations with reduced accuracy requirements, because they allow for a trade-off between accuracy and performance.

Saturday 21.06.2008 15:00-17:00 Parallel Session K

PS10 Room: GGA PARALLEL EIGENSOLVERS AND APPLICATIONS Chair: Jose E. Roman and Paulo Vasconcelos

#45: **Towards a parallel code without communication for the eigenvalues of symmetric tridiagonals**

Presenter: . . . . . Carlos Campos@IPLeiria / ESTG, Portugal
Co-authors: . . . . . Rui Ralha, Vicente Hernandez, David Guerrero

The bisection method is a very reliable method for the computation of eigenvalues of a real symmetric tridiagonal matrix T. The method relies on the function Count(x) which uses a simple recurrence to compute the number of eigenvalues of T that are smaller than x. In exact arithmetic, Count(x) is a monotonic increasing step function of x (i.e., if $x < y$, then Count(x) $<$ Count(y)) but in practice, for several reasons, monotonicity can fail. In LAPACK working note #70 (LAPACK#70), several bisection-like parallel algorithms were presented and they all employ communication between the processors to guarantee correctness. In fact, in LAPACK#70 one may read *We have not succeeded in devising a correct, load balanced algorithm that does not require communication; this remains an open problem*. Recently, a load balanced parallel algorithm which does not require any communication to guarantee the correctness of the results (even if monotonicity of Count(x) can not be ensured) has been proposed by one of the authors. The present work goes in this direction. In order to achieve a better assymptotic convergence rate, we use a variant of Newton's iteration, to refine approximations of isolated eigenvalues. We have developed sequential and parallel codes along these lines. We will benchmark such codes, comparing their execution times with those of the routines DSTEBZ and PDSTEBZ.

#34: **Experiences in the computation of interior eigenvalues for electronic structure calculations**

Presenter: . . . . . Osni Marques@Lawrence Berkeley National Laboratory, USA

The solution of the single particle Schr"odinger equation that arises in electronic structure calculations often requires solving for interior eigenstates of a large Hamiltonian. The states at the top of the valence band and at the bottom of the conduction band determine the band gap that relates to important physical characteristics such as optical or transport properties. In order to avoid the explicit computation of all eigenstates, a folded spectrum method has been usually employed to compute only the eigenstates near the band gap. In this talk, we compare the conjugate gradient minimization, the optimal block preconditioned conjugate gradient, the implicit restarted Lanczos, and variants of the (Jacobi-)Davidson algorithms applied to the folded spectrum matrix for the computation of eigenstates of interest.

#73: **Eigenvalue system for the scattering from rough surfaces saving in computation time by a physical approach**

Presenter: . . . . . Richard Dusseaux@University of Versailles, France
Co-authors: . . . . . Karim Ait Braham, Nahid Emad

The Curvilinear Coordinate Method is an efficient theoretical tool for analysing rough surface. It consists on solving Maxwell's equations written in a nonorthogonal coordinate system. The CM leads to eigenvalue systems and the scattered fields can be expanded as a linear combination of eigensolutions. The boundary conditions allow the combination coefficients to be determined. The dominant computational cost for the C method is the eigenvalue problem solution which is of the order of N3 where N is the size of eigenvalue systems. We can associate the C Method with the Short Coupling Range Approximation (SCRA) in order to reduce the computational time. The adopted procedure consists of two stages. First, the surface fields are obtained by the CM associated with the SCRA. The Short Coupling Range Approximation shows that the surface fields at a given point of a rough surface only depend on the shape of the profile inside an interval centered at this point and that has a width of one or two wavelengths of the incident light. According to the SCRA, the whole surface is represented by several elementary ones. For each elementary surface, the surface current densities are derived from the C Method. The total surface field is deduced from a concatenation of elementary surface current densities. Second, the far-field and the scattering coefficients are derived from the Huygens principle applied to the total surface fields. In the paper, we show that the SCRA applied with the CM is suitable for parallel computing and allows a significant saving in computation time.

#33: **Fast eigenvalue calculations in a massively parallel plasma turbulence code**

Presenter: . . . . . Jose E. Roman@Universidad Politecnica de Valencia, Spain
Co-authors: . . . . . Matthias Kammerer, Florian Merz, Frank Jenko

Magnetic fusion aims at providing CO2 free energy for the 21st century and well beyond. However, the success of the international fusion experiment ITER (currently under construction) will depend to a large degree on the value of the so-called energy confinement time. One of the most advanced tools describing the underlying physical processes is the highly scalable (up to at least 32,768 cores) plasma turbulence code GENE. The GENE solves a set of nonlinear partial integro-differential equations in five-dimensional phase space by means of the method of lines, with a 4th-order explicit Runge-Kutta scheme for time integration. To maximize its efficiency, the code computes the eigenspectrum of the linearized equation to determine the largest possible timestep which maintains the stability of the method. This requires the computation of the largest (in terms of its magnitude) eigenvalue of a complex, non-Hermitian matrix whose size may range from a few millions to even a billion. SLEPc, the Scalable Library for Eigenvalue Problem Computations, is used to effectively compute this part of the spectrum. Additionally, eigenvalue computations

can provide new insight into the properties of plasma turbulence. The latter is driven by a number of different unstable modes, including dominant and subdominant ones, that can be determined employing SLEPc. This computation is more challenging from the numerical point of view, since these eigenvalues can be considered interior, and also because the linearized operator is available only in implicit form. We analyze the feasibility of different strategies for computing these modes, including matrix-free spectral transformation as well as harmonic projection methods.

#82: **A parallel code for computing eigenvalues of integral operators**

Presenter: . . . . . Paulo Vasconcelos@CMUP - FEP, Portugal
Co-authors: . . . . . Filomena Almeida, Jose E. Roman

In this work, we address the parallel implementation of the Multipower Defect Correction method for the computation of a cluster of eigenvalues of a compact Fredholm integral operator and associated invariant subspace basis. The integral operator is discretized by a projection method on a subspace of moderate dimension. The spectral elements of its representing matrix are computed. These spectral elements are then iteratively refined, by a defect correction type procedure accelerated by a few power iterations, to yield an approximation to the spectral elements of the operator. The algorithm is rich in matrix multiplications involving matrices whose construction is distributed among the available processors. A careful distribution of data for a balanced computation is, therefore, crucial. The defect correction phase requires the solution of a numerically quasi-singular linear system of moderate dimension. Computational approaches to solve the problem will be discussed, using the Portable Extensible Toolkit for Scientific Computation (PETSc) and the Scalable Library for Eigenvalue Problem Computations (SLEPc) packages. Numerical results on an astrophysics application, whose mathematical modelling involves a weakly singular integral operator, will be presented and discussed.

PS15 Room: GPA PARALLELIZING ITERATIVE METHODS Chair: Matthias Bollhoefer

#55: **Highly scalable multigrid method for circulant and Toeplitz matrices using non-Galerkin coarse grid operators**

Presenter: . . . . . Matthias Bolten@Research Centre Juelich, Germany

Toeplitz and circulant matrices have been developed by various authors in the past years. All theses methods use the Galerkin operator on the coarse grid, as the associated theory relies on the variational property of this operator. Motivated by the fact that the usage of the Galerkin operator is not mandatory in geometric multigrid and by the growth of the bandwidth of banded matrices, we derived sufficient conditions for non-Galerkin coarse grid operators that still keep the optimal convergence of multigrid methods, while the convergence rate is worse. Based on these conditions, we constructed schemes for the construction of non-Galerkin coarse grid operators that keep the sparsity pattern for certain classes of banded circulant matrices. These patterns are applicable to Toeplitz matrices, as well. The schemes were implemented in a parallel solver for 3-level Toeplitz and circulant matrices that is highly scalable. In this talk the extension of the multigrid theory will be outlined and the schemes mentioned above as well as the scaling behavior of the solver on the Blue Gene/P at the Juelich Supercomputing Centre will be presented.

#40: **AMG for equation systems in commercial fluid dynamics software**

Presenter: . . . . . Maximilian Emans@AVL List GmbH, Austria

A wide variety of AMG variants have been previously described. Despite widely used in commercial software, apart from parameter tuning still a few questions are open, in particular with respect to parallelisation. As far as computational fluid dynamics with finite volume methods and SIMPLE pressure velocity coupling is concerned, the most advantageous application of AMG methods is its employment as a preconditioner of a Krylov-solver of the pressure correction equation. We focus on a comparison of a very lean classical Ruge-Stueben-type AMG and a more expensive variant with an enhanced interpolation scheme, both applied as preconditioners. We first test the algorithms with the Poisson equation on standard cubic geometry and examine convergence and scaling properties. Later we examine the convergence by testing the solvers in typical CFD applications. Finally results of run-time measurements for parallel applications are presented. Conclusion is that lean Ruge-Stueben-type AMG variants are inferior to more advanced schemes with respect to convergence if high accuracy is required. However, for typical applications within the SIMPLE pressure velocity coupling scheme, they are competitive due to low setup cost and cheap application of elementary multigrid operations.

#41: **Symmetric low rank updates of ISM based preconditioners**

Presenter: . . . . . Jose Mas@Universitat Politecnica de Valencia, Spain
Co-authors: . . . . . Juana Cerdan, Jose Marin

Let A be a symmetric and positive definite matrix and let B be the matrix obtained after applying a symmetric low rank update to A, i.e., $B = A + CC'$. In this work we consider the problem of updating a previously computed preconditioner for A in order to solve the updated linear system $Bx = b$ by preconditioned iterations. In particular we consider the application of the recently proposed BIF preconditioner. This preconditioner computes both the LDL' and the inverse factorization at the same time. The method use the approach based on the Sherman-Morrison formula. In this algorithm the direct and inverse factors directly influence each other throughout the computation, and a balance in the factors is achieved using suitable dropping strategies, helping to control the conditioning of the factors. We analyze how to apply the same dropping strategy for updating of the preconditioner.

#54: **Exploiting constant Jacobian entries in seed matrix construction**

Presenter: ..... Ebadollah Varnik@RWTH Aachen, Germany
Co-authors: ..... Uwe Naumann

A sparse Jacobian of a vector function $F : R^n \to R^m, \quad Y = F(X)$ at a given point **x** can be computed in forward mode *automatic differentiation* as $B = F' \cdot S$, where the *seed matrix* $S \in R^{n \times p}$ yields $B \in R^{m \times p}$ with $p \in 1, \ldots, n$. The complexity of computing the compressed Jacobian B is $O(p)$, where p represents the number of columns of S. The goal is to minimize p, which is known to be NP-complete. Therefore we consider the Jacobian $F' = F'_v + F'_c$ as the sum of two matrices F'_v and F'_c containing the variable and constant Jacobian entries, respectively. Substitution yields $B = F' \cdot S = (F'_v + F'_c) \cdot S = F'_v \cdot S + F'_c \cdot S$. We compute F'_c and obtain S by application of some coloring technique to a graph representation of F'_v. Having B, F'_c, and S one can obtain F'_v by solving the linear system $B - F'_c \cdot S = F'_v \cdot S$. A similar technique is applied to the computation of Hessians.

Sunday 22.06.2008 08:40-10:40 Parallel Session M

PS17 Room: B103 LINEAR SYSTEMS - 1 Chair: Denis Trystram

#71: **A NUMA aware scheduler for a parallel sparse direct solver**

Presenter: Mathieu Faverge@INRIA Bordeaux - Sud-Ouest, France
Co-authors: Xavier Lacoste, Pierre Ramet

Over the past few years, parallel sparse direct solvers made significant progress and are now able to solve efficiently industrial three-dimensional problems with several millions of unknowns. An hybrid MPI-thread implementation of our direct solver PaStiX is already well suited for SMP nodes or new multi-core architectures and drastically reduced the memory overhead and improved scalability. In the context of distributed NUMA architectures, a dynamic scheduler based on a work-stealing algorithm has been developed to fill in communication idle times. On these architectures, it is important to take care of NUMA effects and to preserve memory affinity during the work-stealing. The scheduling of communications also needs to be adapted, especially to ensure the overlap by computations. Experiments on numerical test cases will be presented to prove the efficiency of the approach on NUMA architectures. If memory is not large enough to treat a given problem, disks must be used to store data that cannot fit in memory (out-of-core storage). The idle-times due to disk access have to be managed by our dynamic scheduler to prefetch and save datasets. Thus, we design and study specific scheduling algorithms in this particular context.

#63: **A mixed precision iterative method with fast quadruple precision arithmetic operation tuned by SSE2**

Presenter: Hidehiko Hasegawa@University of Tsukuba, Japan
Co-authors: Hisashi Kotakemori

To improve the convergence of Krylov iterative methods, we implemented *double-double* precision arithmetic operation, and tuned it by SSE2 instruction. To have same user interface as that of the Lis library, we used quadruple variables only for all internal variables in each solvers. The computation time of our tuned quadruple operations is at most 3.7 times as long as of the double precision operations on Pentium 4 Xeon 2.8GHz. In distributed parallel environment, the double-double arithmetic operation has a large parallelization ratio, 3.88 for 4 PEs, and 7.61 for 8 PEs. The reason is much more computations with almost same data transfer between PEs. Also the ratio of the double-double arithmetic operations is decreased to 2.86 from 3.7. To reduce total computation time we proposed a mixed precision iterative method: *SWITCH* algorithm based on a combination of the double and the quadruple precision operations. In the *SWITCH* algorithm, the iteration begins with double precision operation, then switches to quadruple precision operation by the detection of stagnation or slow convergence. The switching has almost no extra cost in our implementation. Numerical results and comparison will be shown at the conference.

#56: **Interface-splitting algorithm for parallelisation of diagonal dominant tridiagonal systems with multiple right-hand sides.**

Presenter: Arpiruk Hokpunna@TU Munchen, Germany
Co-authors: Michael Manhart

Solving tridiagonal systems is an important kernel of scientific computing. Tridiagonal systems appear in many approximation problems such as spline interpolation, wavelets or numerical schemes for differential equations. These systems usually are solved repeatedly for an enormous number of different right-hand sides. In three-dimensional time-dependent problems, this number can easily reach $2^3 2$. Therefore an efficient parallel algorithm solving tridiagonal systems for multiple right-hand sides is essential. This can be accomplished by minimising communications and synchronisations. Divide and conquer algorithms are well suited for the current trend in scientific computing. In this paper, we present a novel interface-splitting algorithm for diagonal dominant tridiagonal matrices. This algorithm requires $(5N/p + 4J - 8)M$ operations and two overlapping communications. The idea is to decrease the communication and reduce the data dependency. This is achieved by computing the solution at the interfaces between processors before computing the solution on the inner points. This algorithm exploits an exponential decay of the inverse of diagonal dominant matrices. The purposed scheme is highly scalable and competitive. It has less complexity and more accurate than a previous algorithm. It requires one less synchronisation phase, therefore it is less sensitive to load balancing problems. This scheme is applicable for symmetric and nonsymmetric matrices as well as periodic systems. The accuracy and the stability of the proposed scheme are studied. The performance of the proposed algorithm on commodity Gigabit Ethernets and a high performance interconnection NUMALink(TM) is studied and compared. The scalability study up to 2000 processors on Altix 4700 at the Leibniz-Rechenzentrum will be presented at the conference.

#78: **Combining numerical iterative solvers**

Presenter: Yanik Ngoko@Universite de Yaounde I - Laboratoire d'Informatique de Grenoble, Cameroon/France
Co-authors: Alfredo Goldman, Denis Trystram

Given a linear system there are several solvers which can be used to solve it. However, according to the properties of the linear system different solvers may have different conver- gence speeds, or may even not converge at all. Nevertheless, it can be difficult to verify these properties in practice, mainly due to rounding errors, and there are also some cases where no direct property can be used. In this special situations there is no easy choice on the best solver, so instead of determining it, we are interest in finding

good combinations of the solvers. We are interested by the resolution of sparse systems with three solvers based on three different iterative methods. The numerical methods used are the Conjugate Gradient (CG) method, the BiConjugate Gradient Stabilized (BiCGSTAB) method and the Transpose Free Quasi Minimal Residual Method (TFQMR). To combine numerical solvers, we use an approach based on algorithm portfolio. The basic idea is to interleave iterations of numerical solvers in cycles which are executed until one solver finds a solution. We first study the combination of numerical solvers in an offline setting. In this setting we suppose that a representative set of all linear systems is available. The goal is to combine the set of numerical solvers in order to minimize the average completion time. Then, we study the combination in an on-line setting. In this setting, we do not suppose any previous knowledge. Some heuristics are presented. The first heuristic periodically executes a same cycle of iterations for each numerical solvers. The other heuristics adapts their cycles of iterations from convergence informations gathered from previous cycles. The main difficulty in these latter cases is to define metrics and rules that will be used to evaluate the convergence of previous cycles and to define next cycles. We experiment our approaches using the SPARSKIT library. We present comparisons among heuristics and solvers, and we also study the impact of the cycle size on the execution times.

PS18 Room: B104 AUTOMATIC PARALLELIZATION TUNING AND GPU PROGRAMMING Chair: Costas Bekas

#46: **Automatic generation of multithreaded vectorized adaptive libraries for matrix multiplication**

Presenter: ..... Frederic de Mesmay@Carnegie Mellon University, USA
Co-authors: ..... Franz Franchetti, Yevgen Voronenko, Markus Pueschel

The development of high performance matrix-matrix multiplication (MMM) libraries for current off-the-shelf microprocessors (e.g., Intel Core 2 Duo) is a challenging task. Typically it requires the implementation of a hand-scheduled and hand-vectorized MMM kernel for small sizes and, for larger sizes, cache-aware multi-threaded blocking strategies to achieve maximum speed across the spectrum. These optimizations are usually platform-dependent and hence have to be repeated upon porting. To overcome these problems, our goal is to fully automate the development of MMM libraries without major losses in performance. At the core of our library generator is a declarative mathematical domain-specific language that describes blocked access patterns. Using this language, we express the knowledge about different MMM blocking strategies as rules. The library generator now proceeds autonomously by a) generating efficient vectorized MMM kernels; b) deriving the set of mutually recursive functions that are needed to compute MMM using arbitrary blocking strategies; and c) assembling the components into a multithreaded adaptive library infrastructure that decides on the algorithm at runtime depending on the matrix sizes. Vectorization and parallelization are performed using rewriting systems, and, in contrast to ATLAS (another MMM generator, which can produce MMM kernel code) are also fully automated.

#72: **GPU supercomputing case study: the Kaczmarz algorithm**

Presenter: ..... Joseph Elble@University of Illinois, USA
Co-authors: ..... Panagiotis Vouzis, Nikolaos Sahinidis

Until recently supercomputing was a high-cost investment, not in terms of TFLOPS/USD, but in terms of the nonrecurring building cost of a TFLOPS computer, and the recurring cost of power consumption. The advent of Graphical Processing Units (GPUs) has changed that, and today it is possible to build a USD 3000 workstation with over one TFLOPS of peak-performance power. Recently, there has been an increased interest in using GPUs for attacking computationally intensive problems in areas such as bioinformatics, linear programming, and encryption. This work demonstrates the use of GPUs for the implementation of Kaczmarz's algorithm for solving linear systems of equations, $Ax = b$. This iterative algorithm has found many applications, ranging from computer tomography to digital signal processing. This algorithm lends itself to parallelization, because the required norms and dot products can be computed using the fine grain parallelism available to a GPU. For the GPU implementation of Kaczmarz's algorithm, the Compute Unified Device Architecture (CUDA) of nVidia, Inc is used, along with a Tesla C870 and a GeForce 9800 GX2 to measure the speed up of the algorithm compared to a dual-core AMD Opteron 2.4 GHz CPU.

#79: **Integrating performance models and adaptive approaches for efficient parallel algorithms**

Presenter: ..... Wahid Nasri@ESSTT, Tunisia
Co-authors: ..... Sami Achour, Luiz Angelo Steffenel

Today, with the spreading of complex architectures, we face a great challenge to develop and optimize applications for high performance computing. Unlike dedicated parallel systems, new architectures are inherently heterogeneous. Due to the increasing diversity of existing parallel systems consisting on collections of heterogeneous machines, it is very difficult – and mostly impossible – for a user to choose an adequate algorithm because the execution supports are continuously evolving. This problem of portability becomes crucial with present architectures and applications. These different elements require to revise the classical parallel algorithms and to develop more powerful approaches to efficiently use this type of platforms. In this paper, we present a framework to integrate performance models with adaptive techniques in order to design efficient parallel algorithms in arbitrary environments. To illustrate our approach, we apply the methodology on the matrix multiplication problem, where we compare different parallel (standard and fast) algorithms on heterogeneous environments. We demonstrate that accurate performance predictions can be obtained from analytical performance models, which allow us to adaptively select the most appropriate algorithm to use for given problem and environment without creating heavy additional overhead.

#39: **Research trends on automatic tuning methods for matrix computations and proposal of a new run-time automatic tuning method**

Presenter: Ken Naono@Hitachi, Ltd, Japan
Co-authors: Takao Sakurai, Masashi Egi

Recently, methods of automatic tuning (AT) are required for high performance matrix computation, in order to reduce the cost of hand-tuning of complex algorithms on high performance computer platforms. This paper describes the trend on the AT methods for matrix computations. Two views of matrix software hierarchy and of software development cycle are used in order to classify the methods and to exploit the ongoing research trends. With the views, it is found that the AT research will approach the areas of higher matrix software semantics and that more data of evaluations will be necessary for software development cycle. This paper also proposes a new method of run-time automatic tuning for sparse matrix computations, with respect to the AT research trend. Semantics level of sparse matrix computation is treated differently from that of dense matrix computation. A statistical method that captures fluctuations of convergence is introduced for run-time automatic tuning. The method is regarded as micro software development cycle. Numerical experiments show that the method achieves about 5 times faster performance in the most effective case. The results of the proposed method imply that higher semantics with monitoring is critical for automatic tuning methods.

PS19 Room: B013 HPC APPLICATIONS IN PHYSICS Chair: Nahid Emad

#62: **Solving many-body Schrodinger equation using density functional theory and finite elements**

Presenter: Ondrej Certik@Czech Academy of Sciences, Czech Republic
Co-authors: Jiri Vackar, Miroslav Tuma

In the first part of our talk we show how to solve one particle Schrodinger equation using the finite element method, we describe in detail how to assemble the global matrices and how to solve the generalized eigenvalue problem using our program in Python and C. We use pysparse which uses the Jacobi-Davidson method. We illustrate the solver on 3D examples of an infinite potential well (particle in a box), linear harmonic oscillator (particle on a spring) and the hydrogen atom. In the next part of our talk, we'll show how to solve the many-body Schrodinger equation using the density functional theory, how to setup the self-consistency cycle in our finite elements code and we'll discuss the technical challenges we had to overcome. As a part of the self-consistency cycle, one also needs to solve a poisson equation and we do that either using sparse direct solver Umfpack or the Scipy's conjugate gradient iterative solver, that needs less memory. An example calculation of the electronic ground state of the Boron atom will be shown.

#61: **High-performance finite-element simulations of seismic wave propagation in 3D inelastic media**

Presenter: Fabrice Dupros@BRGM, France
Co-authors: Florent De Martin, Evelyne Foerster, Dimitri Komatitsch, Jean Roman

Simulations of seismic wave propagation are an important tool for risk mitigation and estimation of damage in future earthquake scenarios. From the numerical point of view, several methods have been successfully used for these simulations in three-dimensional geological structures. In the case of inelastic media, the finite-element method is a suitable approach to handle nonlinear soil behavior. As the numerical scheme for time integration is implicit in this case, large linear systems arising from the assembly of the stiffness matrix need to be solved. For the nonlinear procedure, a modified Newton-Raphson algorithm is implemented. This avoids the global assembly of the stiffness matrix during the nonlinear iterative step and therefore saves CPU time. Direct linear solvers are then the best option because of their ability to compute the factorized matrix once and for all and store the coefficients to later perform forward and backward substitutions. We will report on the use of the PaStiX solver in the context of inelastic seismic wave simulation and present an analysis of the parallel performance of the code. The test case for our simulations is a geological model of the French Riviera region with an increasingly larger mesh size in order to evaluate performance.

#67: **A domain decomposition method applied to large eigenvalue problems in neutron physics**

Presenter: Bruno Lathuiliere@EDF R&D - INRIA, France
Co-authors: Maxime Barrault, Pierre Ramet, Jean Roman

The simulation of the neutron transport inside a nuclear reactor leads to the computation of the lowest eigen pair of a simplified transport operator. This computation is done by a power inverse algorithm accelerated by a Chebyshev polynomials based process. At each iteration, a large linear system is solved inexactly by a block Gauss-Seidel algorithm. For our applications, one Gauss-Seidel iteration is already sufficient to ensure the right convergence of the inverse power algorithm. For the approximate resolution of the linear system at each inverse power iteration, we propose a non overlapping domain decomposition based on the introduction of Lagrange multipliers in order to: - get a parallel algorithm, which allows to circumvent memory consumption problem and to reduce the computational time; - deal with different numerical approximations in each subdomain; - minimize the code modifications in our industrial solver. When the Chebyshev acceleration process is switched off, the method performs well on up to 100 processors for an industrial test case. It exhibits a good efficiency which allows us to realize some computations beyond the reach of standard workstations. Besides, we study the efficiency of the Chebyshev acceleration process in our domain decomposition method.

#5: **Implicit domain decomposition algorithm for multiphase flow through random porous formation**

Presenter: . . . . . Tanuja Sheorey@PDPM IIITDM Jabalpur, India

Domain decomposition (DD) algorithm fits naturally into parallel computing and can run efficiently on distributed memory machines. Efficiency and accuracy of DD lies in its interface treatment. A fully implicit domain decomposition algorithm is presented for solving multiphase flow through random porous formation. Naturally occurring porous formations have high degree of permeability variation affecting transport behavior of multiphase flow through such formations. Emphasis of the present study was to test domain decomposition algorithm on physical regions that closely approximate natural porous formations. Two such formations are considered in the present study, 1. Random permeability distribution in physical domain, 2. Physical domain divided into blocks of Homogeneous permeability regions. Partitioning of the physical domain is carried out along one of the coordinate direction. Numerical results show that the algorithm is stable and efficient in reducing the computational time through reduction in subdomain sizes. Moreover, owing to better conditioning of the matrix, it also reduces the floating point operations, without disturbing multiphase flow physics through porous formation having variable permeability distribution. The difference in water front movement in the present simulation with random distribution of permeability and homogeneous permeability blocks can be seen to be quite prominent.

#38: **Towards optimum diffusion for load balancing in heterogeneous torus**

Presenter: . . . . . Nikolaos Missirlis@University of Athens, Greece

In the present work we determine the optimum Diffusion (DF) method for heterogeneous networks and in particular for 2d-torus. Heterogeneous networks consist of processors with different computing speeds and different inter-processor communication links. In previous studies the optimum values of the involved parameters were either computed numerically via the eigenvalues of the Laplacian matrix of the communication graph or they used empirical formulae. However, computing numerically eigenvalues is a time consuming process. It is therefore of vital importance to determine optimum values for the parameters involved in DF using closed form formulae in order to (i) maximize its rate of convergence and (ii) make efficient the process of redistributing the load due to changes in the communication graph. By following a similar approach to our previous work, we are able to find a closed form formula for the set of the parameters of local DF in the sense that its rate of convergence is maximized for heterogeneous torus networks. These optimum values depend only upon the speed of neighbor processors and the incident communication edge weights hence their computation requires only local communication. Further, the optimum value of the convergence factor of heterogeneous local DF is a function of the speed of neighbor processors only.

Sunday 22.06.2008 11:00-13:00 Parallel Session N

PS02 Room: B103 NONNEGATIVE MATRIX FACTORIZATION ALGORITHMS AND APPLICATIONS Chair: Michael Berry

#77: **Step up 2 NNDSVD: better initial approximations for NMF**

Presenter: Efstratios Gallopoulos@University of Patras, Greece
Co-authors: Vassilis Kalofolias, Giorgos Kollias, Dimitrios Zeimpekis, Christos Boutsidis

Low-rank factorization of data matrices has become an important algorithmic tool in modern statistical data analysis. There are application areas in which the data is nonnegative and Nonnegative Matrix Factorization (NMF) appears to offer enhanced intepretability of results over the SVD. Computing NMF's is very hard, however, so we rely on approximations. Specifically, given an m*n nonnegative matrix A, and a positive integer $k < min(m,n)$, NMF approximates A as a product of two nonnegative, rank-k matrices. This is a nonconvex optimization problem with inequality constraints and iterative methods become necessary for its solution. We recently proposed NNDSVD, a fast initialization for such iterative algorithms that uses as starting point the k leading singular factors of A and builds the initial column and row dyads from a unit rank approximation of the positive section of each factor. We describe novel techniques that rely in part on better approximation of rank-2 factors and thus offer NMF estimates of higher quality at little additional cost. Numerical experiments illustrate the effectiveness of the proposed techniques.

#76: **Using the non-smooth NMF model in bioinformatics**

Presenter: Alberto Pascual Montano@Complutense University of Madrid, Spain

High throughput experiments are changing the way research is done in biology, switching the focus from the reductionist approach of studying one gene or protein at a time to studying collections of components and how the interact. The advent of the so-called Systems Biology has brought new challenges to the scientific community. Microarray experiments, for instance, produce as a result lists of genes that are significantly differentially expressed within collections often containing thousands of genes. It is then up to the researcher to extract biological meaning from this raw information. When more elements come into play researchers must handle a broader scope of knowledge to understand their results. There is therefore a big interest in developing resources that help biologist in their interpretations. Non-negative Matrix Factorization (NMF) is a factorization techniques that has gained a lot of popularity in the bioinformatics community due to its capability of providing new insights and relevant information about the complex latent relationships in high-dimensional biological data sets. The comprehensible properties of the NMF model and the intuitivism of the results it provides have centered the attention of many researches in different fields of bioinformatics where it has been applied to the analysis of gene expression data, protein sequences, functional categorization of genes and text mining. In this presentation we will review different applications of non-negative matrix factorization technique, and more in particular, the use of a sparse variant of NMF based on non-smoothness constraints (nsNMF). In addition we will discuss computational efficiency of these algorithms that seriously limits their application in real case scenarios, and some solutions to this problematic using parallel and grid computing. We will focus our attention on two efficient web-based analysis tools that we have developed using NMF and some of its variants in the field of gene expression analysis (bioNMF: http://bionmf.dacya.ucm.es) and the extraction of semantic features in scientific text analysis in molecular biology (SENT: http://sent.dacya.ucm.es).

#12: **Using a literature-based NMF model for discovering gene functional relationships**

Presenter: Elina Tjioe@University of Tennessee, Knoxville, USA
Co-authors: Michael Berry, Ramin Homayouni, Kevin Heinrich

The rapid growth of the biomedical literature and genomic information present a major challenge for determining the functional relationships among genes. Several bioinformatics tools have been developed to extract and identify gene relationships from various biological databases. In this study, we develop a Web-based bioinformatics tool called Feature Annotation Using Nonnegative matrix factorization (FAUN) to facilitate both the discovery and classification of functional relationships among genes. Both the computational complexity and parameterization of nonnegative matrix factorization (NMF) for processing gene sets is discussed. FAUN is first tested on a small manually constructed 50 gene collection that we, as well as others, have previously used. We then apply FAUN to analyze several microarray-derived gene sets obtained from studies of the developing cerebellum in normal and mutant mice. FAUN provides utilities for collaborative knowledge discovery and identification of new gene relationships from text streams and repositories (e.g., MEDLINE). It is particularly useful for the validation and analysis of gene associations suggested by microarray experimentation.

PS05 Room: B013 PARALLELIZING MATRIX COMPUTATIONS Chair: Thomas Huckle

#15: **A lower bound for sparse-matrix dense-vector multiplication using packet communication**

Presenter: Michael Schnupp@TU Munchen, Germany
Co-authors: Riko Jacob

We consider the problem of multiplying a huge sparse matrix with a dense vector. Our model of computation are parallel processor that communicate in large packets. More precisely, we assume that every processor has a memory of size M and sends packets of size B, with the limitation that at most M/B different packets can be sent or received by a single processor at any time. As it

turns out, this parallel model is closely related to the serial Disk-Access-Model (aka I/O-Model) with the same parameters M and B. This allows us to modify a known lower bound for sparse-matrix vector multiplication to also hold in this parallel model. We will also address the question of worst case optimal algorithms as well as variants of the model and the problem: different layout or access to the matrix, non-square matrices, simultaneously multiplying with different vectors.

#69: **Parallel dense and sparse matrix multiplication in a quantum control problem**

Presenter: ..... Michael Bader@TU Munchen, Germany

We present efficient algorithms and parallel implementations for sparse and dense matrix multiplication for two matrix problems that occur in a quantum control application. The first problem is to compute the exponential function of a sparse matrix. There, the exponential function is approximated by a Csebyshev polynomial; the evaluation of the polynomial for a given sparse matrix then requires a sequence of sparse-dense matrix multiplications. The second problem is the parallel prefix computation of a sequence of dense matrices, which result from the computation of the expontential functions. In that problem, both a fine-grain approach (parallel matrix multiplication) and a coarse-grain approach (concurrent computation of matrix products in a tree-like scheme) exist, plus respective hybrid approaches. Both, matrix multiplication and parallelisation is based on the use of space-filling curves: in particular, efficient data structures based on Peano curves are used to improve the cache efficiency, and to efficiently parallelise matrix multiplication. We present performance results on multicore platforms as well as on compute clusters.

#70: **Chebyshev methods for the matrix exponential**

Presenter: ..... Konrad Waldherr@TU Munchen, Germany

When dealing with special quantum control problems the computation of the exponential of a matrix plays a key role. Thereby appearing matrices are sparse and their dimensions grow exponentially in the system size. In concrete applications the size of the occurring matrices lies between $2^{10} \times 2^{10}$ and $2^{20} \times 2^{20}$. The matrix exponential is in general one of the most studied matrix functions. There are different ways to compute it such as eigen decomposition or Pad'e methods. We will have a look at a polynomial expansion based on Chebyshev polynomials. In combination with the scaling & squaring technique this ansatz leads to a very accurate and efficient method for the computation of the matrix exponential. Since only elementary matrix operations are required, it can be parallelized very easily. The Chebyshev method is very appropriate for the mentioned quantum control problems, as the evaluation of the occurring matrix polynomials can be organized in such a way that only matrix products of the form dense $\times$ sparse appear. Several numerical examples reveal the effectiveness of our methods.

#29: **Parallel matrix computations arising in two-phase flow problems**

Presenter: ..... Oliver Fortmeier@RWTH Aachen, Germany

We consider the parallel solution of computational fluid dynamics problems involving two different phases. Examples include falling films and rising oil droplets in water. Here, there is need to solve a large number of systems of linear equations. A two-phase flow problem can be modeled by coupled Navier-Stokes equations describing the dynamics of the fluids with a level set equation describing the boundary of the two phases. Discretising these equations by a finite element methods leads to a large number of structured systems of linear equations. Using an adaptive grid that changes over time, the structure of these linear systems changes over time as well. To solve the complex numerical formulation, advanced numerical algorithms and parallel computing are urgently needed. We present recent techniques for the parallel solution of two-phase flow problems and report on numerical experiments on SUN Fire parallel computer. This is done by an joint effort in collaboration with the chair of numerical mathematic.

PS16 Room: B104 LINEAR SYSTEMS - 2 Chair: Pascal Henon

#49: **Solution of three-dimensional Helmholtz equation in the frequency domain, using Krylov methods preconditioned by multigrid**

Presenter: ..... Xavier Pinel@CERFACS, France
Co-authors: ..... Henri Calandra, Iain S. Duff, Serge Gratton, Xavier Vasseur

The topic of our work is the solution of the three-dimensional Helmholtz equation with absorbing boundary conditions in the frequency domain, with application to inverse problems. We consider a finite difference discretization scheme that leads to linear systems with up to one billion degrees of freedom. For such linear systems, the use of Krylov methods preconditioned by multigrid techniques is not new. The originality and the success of our approach rely in the particular smoother, coarse problem solver and cycle type we use. Indeed, in our case, geometric multigrid with traditional relaxation methods cannot be used, because the Helmholtz problem becomes indefinite for high frequencies. To overcome this problem we limit the number of levels and use preconditioned Krylov methods both as a smoother and as the coarsest grid solver. A consequence of these choices is that our preconditioner, for which the memory cost growths linearly with the problem size, varies along the iterations. Therefore we have to use flexible Krylov subspace methods. Spectra and theoretical Fourier analyses will be presented for two-dimensional problems, and three-dimensional scalability experiments will be shown on massively parallel computers.

#51: **A proposal of preconditioning based on AGM (arithmetic-geometric mean) of eigenvalues and its estimation**

Presenter: Seiji Fujino@Kyushu University, Japan
Co-authors: Kazuhito Somehara

In this paper, we propose a new Preconditioning technique based on AGM (Arithmetic-Geometric Mean) of Eigenvalues by means of approximation of its determinant to 1. We refer to it as PAGME method. Through numerical experiments, we verify parallelism of CG method with PAGME on SMP machine compared with the conventional preconditioned CG methods.

#53: **Efficient iterative solution of large sparse linear systems on a cluster of geographically separated clusters**

Presenter: Tijmen Collignon@Delft University of Technology, Netherlands
Co-authors: Martin van Gijzen

Efficiently solving large sparse linear systems on cluster and Grid computers is an active field of research. The induced heterogeneity and volatile nature of the aggregated computational resources present numerous algorithmic challenges. Parallel asynchronous iterative algorithms exhibit features that are extremely well–suited for Grid computing, such as lack of synchronisation points and coarse–graininess. Unfortunately, they also suffer from slow (block Jacobi–like) convergence rates. In this talk we propose using said asynchronous methods as a coarse–grain preconditioner in a flexible iterative method, where the preconditioner is allowed to change in each iteration step. By combining a slowly converging asynchronous method on distributed hardware and a fast converging flexible method on parallel hardware, we aim to reap the benefits and awards of both techniques. We will present a full implementation of the algorithm using mature Grid middleware that allows for both synchronous and asynchronous communication. Advantages and disadvantages of the approach are discussed. Numerical experiments are performed utilising the Dutch DAS–3 national supercomputer, which consists of five geographically separated clusters spread over four academic institutions, connected by fiber optic technology. Promising results are obtained with application to large 3D bubbly flow problems, which is an important problem from computational fluid dynamics in two–phase fluid flow. This research is funded by the Delft Centre for Computational Science and Engineering. The Netherlands Organisation for Scientific Research (NWO) is gratefully acknowledged for the use of the DAS–3.

#66: **HIPS: a parallel hybrid direct/iterative solver based on a Schur complement approach.**

Presenter: Pascal Henon@INRIA, France
Co-authors: Jeremie Gaidamour

The resolution of large sparse linear systems is often the most consuming step in scientific applications. Direct methods are very robust but they need a prohibitive amount of memory and CPU time to deal with very large systems arising from 3D problems. On the other hand, generic preconditioners for iterative methods are often based on incomplete factorizations. The robustness of these preconditioners depends on the number of non-zero allowed in the factors (fill-in). They usually rely on scalar implementation (no use of BLAS routine) and thus, on difficult problems, the implementation of this kind of preconditioner is not efficient. HIPS is a parallel solver that combines the most of these two classes of method. Our method relies on a partition of the matrix adjacency graph into small domains with overlap. The interior of the domains are factorized by a direct method; thus solving the whole system amounts to solve the Schur complement system on the interface. We solve the Schur complement by a preconditioned Krylov method. Using this "hybrid" approach, we will present some original reordering and incomplete blockwise factorization algorithms that allow to obtain a good trade-off between memory consumption, robustness and parallel performance in many cases.

Author index