Sparse Days 2019

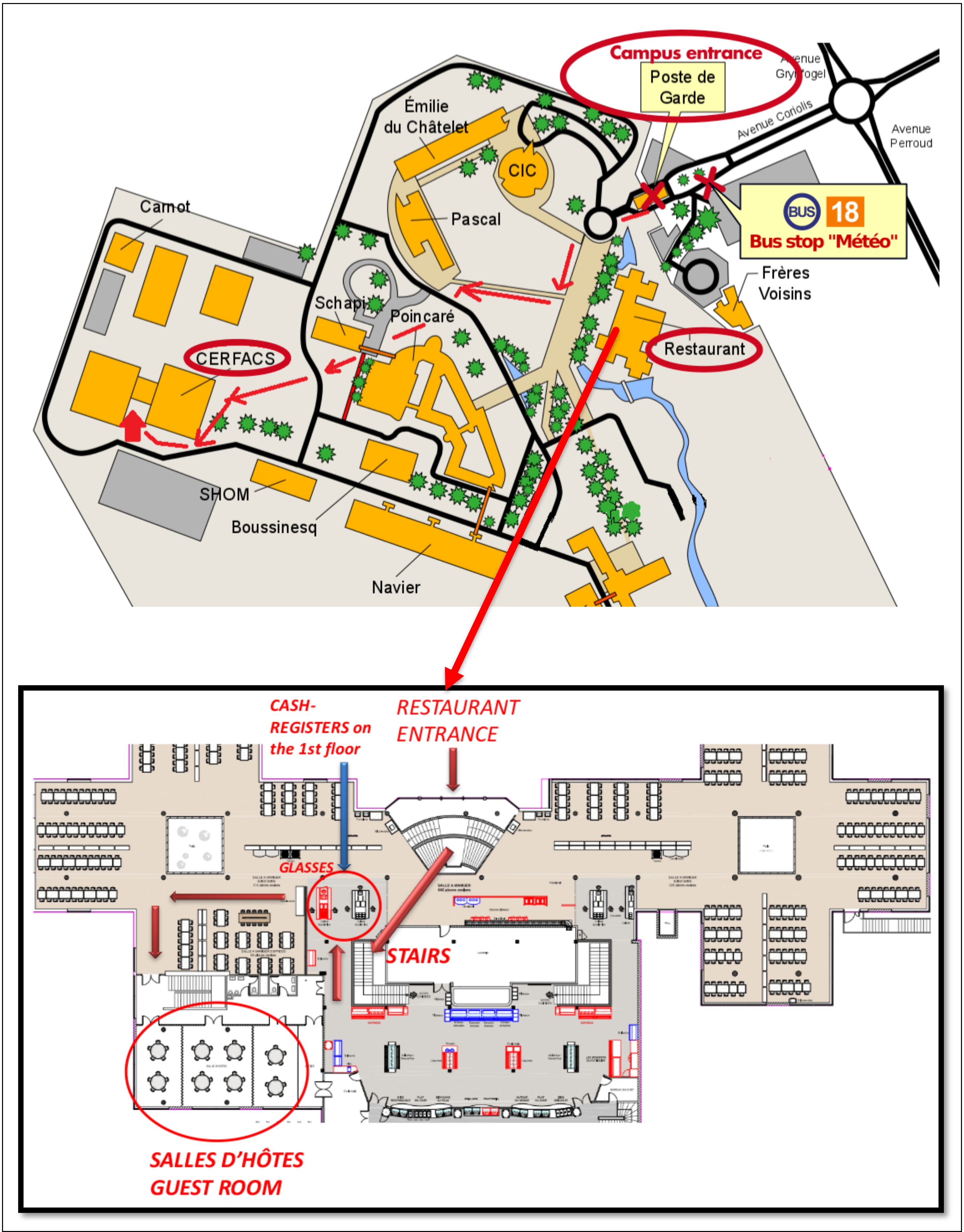
July 11th – 12th, 2019

Cerfacs, Toulouse, France

CERFACS

CENTRE EUROPÉEN DE RECHERCHE ET DE FORMATION AVANCÉE EN CALCUL SCIENTIFIQUE

Welcome to Cerfacs



Sparse Days programme
Thursday, July 11th 2019

09:00 – 09:45 **Registration and coffee**

09:45 – 10:00 **Welcome message**

10:00 – 12:00 **Session 1 – chair: D. Ruiz**
10:00 – 10:30 Highly robust iterative, domain decomposition solvers and load balancing aspects in the setup phase
Martin KÜHN (Cerfacs, France)
10:30 – 11:00 Sparse Matrices in Molecular Simulation and in Physiological Models
Marcus WEBER (Zuse Institute Berlin (ZIB), Germany)
11:00 – 11:30 Data structures to implement the Sparse Vector in Crout ILU Preconditioner
Raju RAM (Fraunhofer ITWM, Germany)
11:30 – 12:00 Multigrid-based augmented block-Cimmino method
Philippe LELEUX (Cerfacs, France)

12:00 – 13:30 **Lunch break**

13:30 – 15:30 **Session 2 – chair: I. Duff**
13:30 – 14:00 Design and implementation of a parallel Markowitz threshold algorithm
Stojce NAKOV (STFC, United Kingdom)
14:00 – 14:30 A massively-parallel algorithm for Bordered Almost Block Diagonal systems on GPUs
Monica DESSOLE (Università di Padova, Italy)
14:30 – 15:00 Implicit Analysis of Jet Engine Models on Thousands of Processors
Robert LUCAS (LSTC, United States)
15:00 – 15:30 Rank revealing QR methods for sparse block low rank solvers
Esragul KORKMAZ (Inria, France)

15:30 – 16:00 **Coffee break**

16:00 – 17:30 **Session 3 – chair: L. Giraud**
16:00 – 16:30 High memory bandwidth architectures for very sparse matrices
Frank HÜLSEMANN (EDF R&D, France)
16:30 – 17:00 Parallel Algorithms through Approximation: Vertex Weighted Matching
Alex POTHEN (Purdue University, United States)
17:00 – 17:30 On the scheduling of sparse matrix factorizations: is it time for a change?
Mathias JACQUELIN (Lawrence Berkeley National Laboratory, United States)

20:00 **Reception dinner**
Les Caves de la Maréchale, 3 rue Jules Chalande, Toulouse

Friday, July 12th 2019

09:30 – 10:30 **Session 4 – chair: U. Rüde**
09:30 – 10:00 Algebraic Multigrid for Hypersonic Simulations
Raymond TUMINARO (Sandia National Laboratories, United States)
10:00 – 10:30 FEAST v4.0 with Applications
Eric POLIZZI (UMass Amherst, United States)

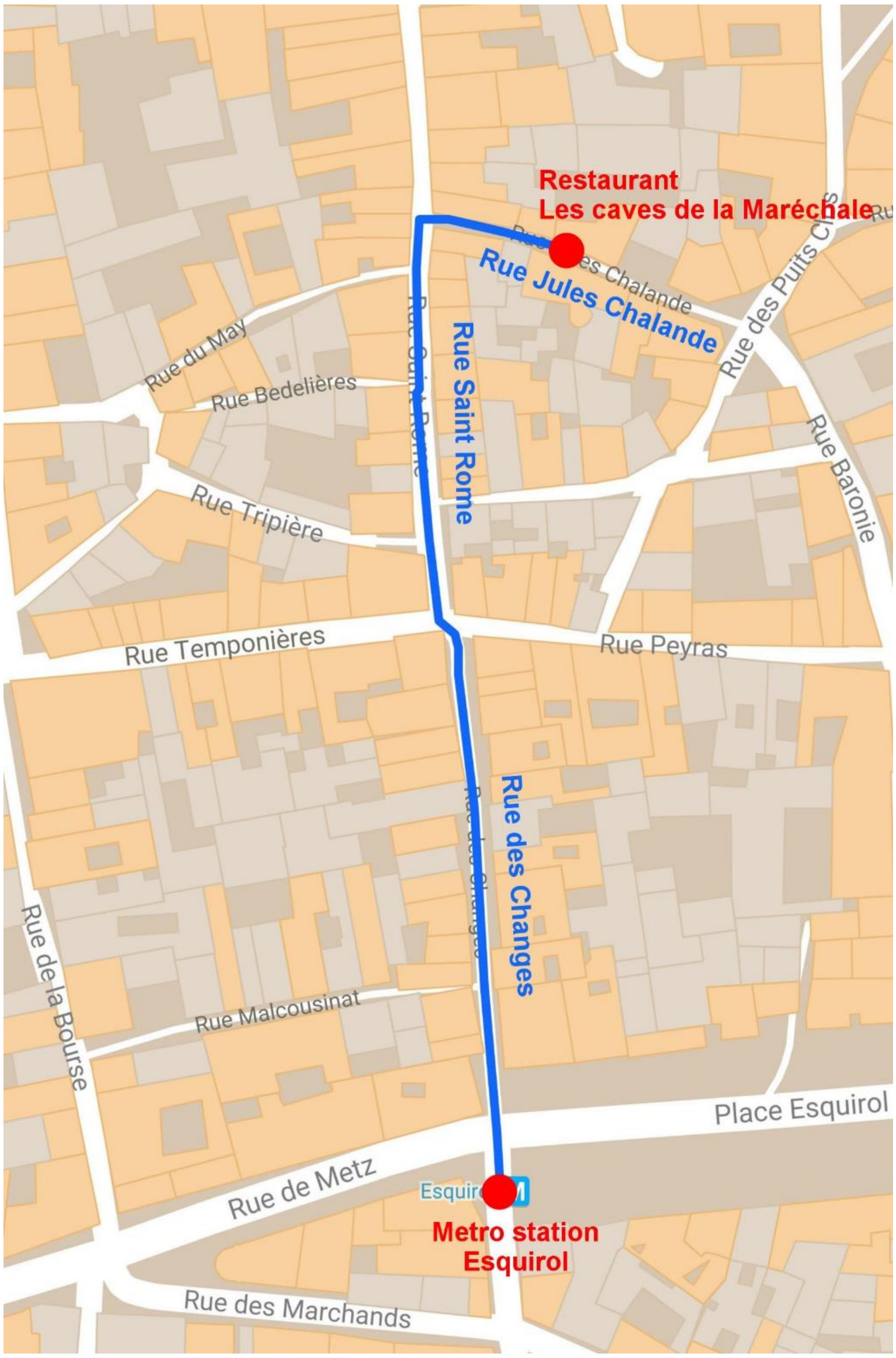
10:30 – 11:00 **Coffee break**

11:00 – 12:30 **Session 5 – chair: C. Kruse**
11:00 – 11:30 Preconditioning for nonsymmetric systems
Andy WATHEN (Oxford University, UK, United Kingdom)
11:30 – 12:00 Block Preconditioners for Incompressible Magnetohydrodynamics Problems
Michael WATHEN (STFC Rutherford Appleton Laboratory, United Kingdom)
12:00 **End of the meeting**

Reception dinner at Les Caves de la Maréchale
3 rue Jules Chalande, Toulouse
Thursday, July 11th 2019, 20:00

Directions from Météo-France to the restaurant:

Get on **bus 18 towards "Basso Cambo"** at stop "Météo" (close to the entrance of Météo France).
Get off at **stop "Basso Cambo"** (end of the line).
Take **metro A** (towards "Balma Gramont") and get off at **station "Esquirol"**.
Walk (5 min) from station "Esquirol" to the restaurant.



Abstracts

Highly robust iterative, domain decomposition solvers and load balancing aspects in the setup phase

Martin KÜHN (Cerfacs, France)

Domain decomposition methods are highly scalable parallel solvers for large sparse systems obtained from the discretization of partial differential equations (PDEs). However, the convergence behavior of well-known iterative domain decomposition solvers such as FETI-DP and BDDC with a standard coarse spaces deteriorates significantly if composite materials are considered. In such cases, problem-dependent (or adaptive) coarse spaces offer a crucial remedy. In adaptive methods, difficulties arisen from highly heterogeneous materials are detected automatically by solving local generalized eigenvalue problems and an adaptive coarse space is set up. On the one hand, adaptive methods are thus characterized by great robustness. On the other hand, the unequally distributed material heterogeneities generally lead to very bad load balancings in the setup phase of the adaptive coarse space. For an efficient parallel implementation, load balancing has to be achieved and unnecessary eigenvalue problems should be discarded to reduce the computational overhead of the setup phase.
Joint work with Axel Klawonn (Universität zu Köln) and Oliver Rheinbach (Technische Universität Bergakademie Freiberg).

Sparse Matrices in Molecular Simulation and in Physiological Models

Marcus WEBER (Zuse Institute Berlin (ZIB), Germany)

Molecular Simulation and Systems Biology is not only based on solving ordinary differential equations. Functional analysis and the discretization of special operators leading to sparse matrices plays a major role in modern applications. Also in cell biology it is known that the cell signaling network is represented by a sparse matrix. The talk will discuss ideas, how to deal with and how to preserve sparsity of those matrices. We will present an application of sparse matrices for a disease where tissue irregularly turns into bones.

Data structures to implement the Sparse Vector in Crout ILU Preconditioner

Raju RAM (Fraunhofer ITWM, Germany)

The solution of large sparse linear systems is a ubiquitous problem in chemistry, physics, and engineering applications. Krylov subspace methods are preferred to solve the large scale linear systems instead of direct methods as they are faster and use less memory. Parallel performance of Krylov Solvers depends upon the implementation of underlying basic operations (such as SpMV and global reduction etc.) and properties of the underlying matrix. In many cases, an effective preconditioner is required to improve the properties of an ill-conditioned matrix. We are developing efficient parallel preconditioners in a linear solver library GaspiLS, that is based on distributed parallel programming model GASPI. The iterative methods frequently incorporate incomplete LU (ILU) preconditioner because of its robustness, accuracy, and usability as a black-box preconditioner. The ILU Preconditioner does not have inherent parallelism. To introduce parallelism, Crout version of ILU is used as a preconditioner in state of the art package such as ILUPACK. Additionally, Crout ILU also controls the growth of error in the preconditioning operation by setting a bound on the inverse of the triangular (L and U) factors. The Crout ILU method uses multilevel nested dissection approach to decompose the large matrix into independent sub-matrices. Crout ILU decomposition of these submatrices requires dynamic insertion and deletion of entries in sparse vectors. Efficient implementation of the sparse vectors is crucial for the performance of the Crout ILU preconditioner. We have investigated the time and memory complexity of different data structures that implement the sparse vector. In this talk, we show the advantages and disadvantages of various data structures. To support our claim, we additionally show various profiler plots.

Multigrid-based augmented block-Cimmino method

Philippe LELEUX (Cerfacs, France)

The Augmented Block Cimmino Distributed Solver (ABCD Solver) is a hybrid method designed to solve large sparse unsymmetric linear systems of the form: Ax=b, where A is a full row rank m×n sparse matrix, x is a vector of size n and b is a vector of size m. The approach is based on the block Cimmino row projection method (BC) [2]. BC is applied on the system which is partitioned in row blocks. Convergence rate of BC is known to be slow and we rather solve the symmetric semi-positive definite and consistent system Hx=k obtained when considering the fix point of the iterations. To accelerate the convergence of the block Cimmino method, we solve instead this system using a stabilized block Conjugate Gradient (BCG) [3].

The solver also offers the possibility to construct a larger system with an enlarged matrix where the numerical orthogonality between partitions is enforced. As a result, the augmented block Cimmino method converges in one single iteration. This results in a pseudo-direct method (ABCD) [1] with the solution depending on projections, as in BC, and on the direct solution of a condensed system. Implementation of both ABCD and BC are available in ABCD Solver package.

For large PDE problems, we investigate extensions of this augmentation approach, in which we relax the strict orthogonality between blocks so as to reduce the size of the augmentation blocks. The purpose is to control better the memory needs and computations. To do so, we exploit ideas from the multigrid framework. Assuming that we have at hand several levels of grids for the system, the original system is then augmented very similarly to the ABCD method, by enforcing the orthogonality between partitions on a coarse level. The result is an augmented system with approximated orthogonality on the fine grid level discretized operator. While in the exact ABCD approach the size of the augmentation can be large for highly connected subdomains, the new approach gives a way to control explicitly this size by choosing coarser levels of grids. This system can be solved with the classical block-Cimmino with fast linear convergence for a wide range of systems coming from PDE problems. We demonstrate the efficiency of the method on Helmholtz and Convection-Diffusion 2D problems.

Joint work with Dr. Daniel Ruiz (IRIT, France) & Prof. Ulrich Rüde (FAU, Germany; Cerfacs, France)

[1] Duff, I. S., Guivarch, R., Ruiz, D., & Zenadi, M. (2015). The augmented block Cimmino distributed method. SIAM Journal on Scientific Computing, 37(3), A1248-A1269.

[2] Elfving, T. (1980). Block-iterative methods for consistent and inconsistent linear equations. Numerische Mathematik, 35(1), 1-12.

[3] Ruiz, D. (1992). Solution of large sparse unsymmetric linear systems with a block iterative method in a multiprocessor environment. CERFA CS TH/PA/9, 6.

Design and implementation of a parallel Markowitz threshold algorithm

Stojce NAKOV (STFC, United Kingdom)

We develop a novel algorithm for the parallel factorization of an unsymmetric sparse matrix using a Markowitz threshold algorithm. The main idea behind the algorithm is to find a set of independent pivots. For this, we have extended Luby's algorithm for finding a maximum independent set in an undirected graph and have adapted it for directed graphs. Then we update the Schur complement, and repeat the process on the trailing matrix until the matrix is factorized. We have developed a parallel multithreaded open-source library, ParSHUM, based on OpenMP. Furthermore, in order to cope with distributed platforms, we first apply a Singly Bordered Block Decomposition (SBBD). Each diagonal block is then factorized using ParSHUM. We thus exploit two levels of parallelism on a distributed level by performing the factorization of each diagonal block concurrently, and in shared memory through the multithreaded ParSHUM solver applied to each block.

A massively-parallel algorithm for Bordered Almost Block Diagonal systems on GPUs

Monica DESSOLE (Università di Padova, Italy)

We present PARASOF (PARAllel Structured Orthogonal Factorization), a new algorithm for the solution of linear systems with BABD (Bordered Almost Block Diagonal) matrices on massively parallel computing systems like GPUs. BABD matrices are particular sparse matrices with structure that mainly arise in the numerical solution of Boundary Value Problems for Ordinary Differential Equations. We study existing parallel solvers for BABD systems, and analyze the weaknesses of their straightforward implementation on a GPU. The proposed PARASOF algorithm is compared with the state-of-the-art algorithms, in particular SOF (Structured Orthogonal Factorization), from which it is inspired and takes the same stability properties. The proposed PARASOF solver achieves up to 60x speedup over the umfpack CPU general solver. Its parallel computational complexity and execution times agree in declaring this algorithm as a substantial improvement against the state-of-the-art existing algorithms.

Implicit Analysis of Jet Engine Models on Thousands of Processors

Robert LUCAS (LSTC, United States)

Cray, LSTC, the University of Illinois's National Center for Supercomputing Applications (NCSA), and Rolls-Royce formed a partnership to explore the future of implicit computations as both the finite element models and the systems they run on increase in scale. To facilitate this work, Rolls-Royce created a family of dummy engine models, using solid elements, the largest of which has almost 200 million degrees of freedom. After two years of work, we are now able to run these engine models on NCSA's Blue Waters machine and Oak Ridge National Laboratory's Titan, using tens of thousands of their AMD cores. This talk will discuss the linear algebra related challenges encountered, including the deployment and performance optimization of new linear constraint, nested dissection and symbolic factorization algorithms. This is ongoing work, and we will conclude by discussing our plans for the future, and some challenges we foresee looking forward towards hundreds of thousands of cores.

Rank revealing QR methods for sparse block low rank solvers

Esragul KORKMAZ (Inria, France)

Solving linear equations of type Ax=b for large sparse systems frequently emerges in science and engineering applications, which creates the main bottleneck. In spite that the direct methods are costly in time and memory consumption, they are still the most robust way to solve these systems. Nowadays, increasing the amount of computational units for the supercomputers became trendy, while the memory available per core is reduced. Therefore, when solving these linear equations, memory reduction becomes as important as time reduction. For this purpose, compression methods of dense blocks appearing inside sparse matrix solvers have been introduced to reduce the memory consumption, as well as the time to solution. While looking for the lowest possible compression rank, Singular Value Decomposition (SVD) gives the best result. It is however too costly as the whole factorization is computed to find the resulting rank. In this respect, rank revealing QR decomposition variants are less costly, but can introduce larger ranks. Among these variants, column pivoting or matrix rotation can be applied on the matrix A, such that the most important information in the matrix is gathered to the leftmost columns and the remaining unnecessary information can be omitted. For reducing the communication cost of the classical QR decomposition with column pivoting, blocking versions with randomization are suggested as an alternative solution to find the pivots. In these randomized variants, the matrix A is projected on a much lower dimensional matrix by using an independent and identically distributed Gaussian matrix so that the pivoting/rotational matrix can be computed on the lower dimensional matrix. In addition, to avoid unnecessary updates of the trailing matrix at each iteration, a truncated randomized method is suggested and shown to be more efficient for larger matrix sizes. Thanks to these methods, closer results to SVD can be obtained and the cost of compression can be reduced. In this presentation, a comparison of all these methods in terms of complexity, numerical stability and performance will be presented.

High memory bandwidth architectures for very sparse matrices

Frank HÜLSEMANN (EDF R&D, France)

The arithmetic intensity of PDE discretization schemes with compact support, in particular on unstructured meshes, is so low that on current architectures these schemes are limited by the available bandwidth more than by the computational power. Over the last few years, a small number of high bandwidth memory (HBM) systems have become available. This talk concentrates on one of these architectures, the NEC vector engine SX-Aurora Tsubasa, and compares obtained performance results to the prediction of the roofline model.

Parallel Algorithms through Approximation: Vertex Weighted Matching

Alex POTHEN (Purdue University, United States)

Maximum vertex-weighted matching (MVM) is a problem that arises in internet advertising, where advertisers pay a price to place their ads on a webpage that shows the result of a keyword search. These have led to some of the largest matching problems ever solved, both in terms of the size of the graphs and money spent. We describe new exact and 2/3-approximation algorithms to solve MVM. These "Direct" algorithms designed by our group in earlier work did not have much concurrency since they required the vertices to be matched in non-increasing order of weights. We show that "Iterative" algorithms could be designed that relax this ordering requirement, and hence algorithms could be parallelized. The parallel algorithm needs to be carefully designed to avoid "livelock", a state of affairs where none of the threads is able to augment the matching due to cyclic dependences in thread-synchronization. We have implemented this algorithm on shared memory multi-processors, and show that it is currently the fastest parallel (and serial) algorithm in practice on a collection of test problems.

Joint work with Ahmed Al-Herz, Computer Science department, Purdue University.

On the scheduling of sparse matrix factorizations: is it time for a change?

Mathias JACQUELIN (Lawrence Berkeley National Laboratory, United States)

For the past decades, sparse matrix factorizations have been heavily relying upon the elimination tree and heuristics based on the subtree-to-subcube or proportional mapping strategy to assign data to processing elements. . These techniques have significantly improved performance of sparse matrix direct solvers. Although these methods are taking the computational load into account before making scheduling decisions, they rely on an incomplete view of the full task graph. As we have entered the era of manycore processors and GPUs, the ever growing amount of parallelism available raises an important question: should this crucial step of sparse direct solvers be revised?

Algebraic Multigrid for Hypersonic Simulations

Raymond TUMINARO (Sandia National Laboratories, United States)

With an ever growing variety of space vehicles, the simulation of hypersonic flight gives rise to increasingly important research challenges associated with the the efficient solution of highly non-symmetric systems. While the use of multigrid within sub-sonic and transonic flow simulations has been well studied, the hypersonic flow case has received little attention and is still very much an open research topic. In this talk, we investigate the application of algebraic multigrid to solve the linear systems that arise from the use of Newton's method in conjunction with adaptive pseudo-time stepping to solve steady state hypersonic flow problems. The overall approach gives rise to a series of linear systems associated with a Jacobian approximation of the nonlinear equations that is then solved by a multigrid method. Unfortunately, the application of algebraic multigrid can be problematic for highly non-symmetric systems. Specifically, the coarse grid equations produced by a standard Petrov-Galerkin approximation may not be stable, though the fine grid equations are stable. Thus, standard smoothing algorithms such as ILU and line Gauss-Seidel often diverge when applied directly to coarse level equations. Further, solutions from a coarse grid do not necessarily reduce error (as the underlying projections are not orthogonal). We discuss a few algebraic multigrid algorithms to address the above-mentioned concerns. This includes the adaptation of specialized grid transfers that attempt to mimic upwinding ideas used for highly convective flows and the use of non-Galerkin or non-Petrov-Galerkin coarse grid operators. The basic idea behind a non-Petrov-Galerkin approach is to adds a perturbation to the coarse level operator produced by a standard Petrov-Galerkin projection. Two natural possibilities for this matrix perturbation include a diagonal term (motivated by considering a reduced pseudo-time step on the coarse level) as well as a projection portion of the symmetric part of the matrix (loosely motivated by artificial dissipation considerations). In both cases, the aim is to add the smallest perturbation such that the standard smoothing algorithms will converge on the resulting system. Numerical results will be given on HPC systems using test problems arising from a hypersonic flow simulation capability, SPARC, that is being developed at Sandia.

FEAST v4.0 with Applications

Eric POLIZZI (UMass Amherst, United States)

The FEAST eigensolver uses complex contour integration and subspace iterations to calculate the eigenvectors whose eigenvalues that are located inside some user-defined region. The algorithm can be interpreted as a generalization of shift-and-invert iterations that uses multiple shifts in the complex plane leading to an optimal filter projector. In the FEAST v4 package, the solver has been reimplemented to make use of residual inverse iterations. Although, the new filter form is mathematically equivalent to the original FEAST linear projector, it is numerically more efficient and more appealing in a number of new situations. We will demonstrate the effectiveness of the FEAST residual inverse iterations for addressing (i) the inexact inner-outer iterative approach (IFEAST or FEAST without factorization for large sparse systems), (ii) the mixed precision arithmetic iterative procedure, and (iii) the non-linear eigenvalue problem.

Preconditioning for nonsymmetric systems

Andy WATHEN (Oxford University, UK, United Kingdom)

Preconditioning for nonsymmetric matrices is much less solidly founded in theory than in the case of real symmetric matrices. In this talk I will discuss some preconditioning strategies for nonsymmetric matrices for which some theoretical guarantees may exist.

Block Preconditioners for Incompressible Magnetohydrodynamics Problems

Michael WATHEN (STFC Rutherford Appleton Laboratory, United Kingdom)

We consider preconditioning techniques for a mixed finite element discretization of an incompressible magnetohydrodynamics (MHD) problem. Upon discretization and linearization, a 4-by-4 nonsymmetric block-structured linear system needs to be (repeatedly) solved. One of the principal challenges is the presence of a skew-symmetric term that couples the fluid velocity with the magnetic field. We propose a preconditioner that exploits the block structure of the underlying linear system, utilizing and combining effective solvers for the mixed Maxwell and the Navier–Stokes subproblems. We perform a spectral analysis for an ideal version of the preconditioner, and develop and test a practical version of it. Large-scale numerical results in two and three dimensions validate the effectiveness of our approach.