Joint European Exascale Projects Workshop

Innovative Algorithms for Exascale

WP4: Algorithms and libraries

Overview

Dmitry Khabi, High Performance Computing Center Stuttgart (USTUTT-HLRS)

Work package overview

WP 4: : Algorithms and libraries	
Lead beneficiary	USTUTT-HLRS
Partners:	UEDIN, USTUTT, KTH, Cray UK, DLR, ECMWF
Status	M30 of 39 months

- Objectives:

The objective of this work package is to address the limitations of existing algorithms and libraries for exascale computing systems on various levels…

Our focus is on developing new implementations, which will yield step-wise improvements in the application codes at all stages from pre-processing, simulation and post-processing.

We will also investigate how the work produced can be taken forward beyond the lifetime of CRESTA.

CRESTA

Algorithms in CRESTA WP4 (Algorithms and libraries)

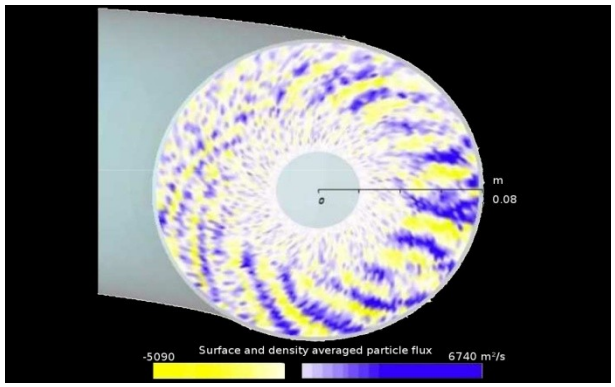
- Sparse linear systems solver
  - Data sets (linear systems) from real applications
  - Test of the existing libraries (PETsC , Trilinos ...)
  - Implementation of a linear solver(s) using the disruptive and innovative strategies

- Collective operations
  - Blocking and Non-Blocking Collectives
  - Multi-precision software for reduction operations
  - Microbenchmark suite to analyse the characteristics of the implementation of collective operations

- Multi-dimensional FFTs
  - Support changes in data decomposition to quickly optimization of the FFT strategy for the available hardware

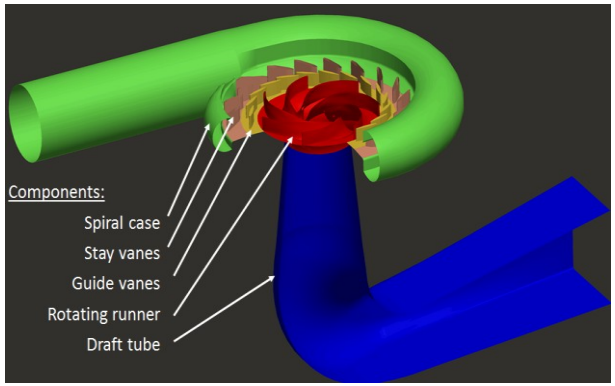
CRESTA

Sparse linear systems from real scientific applications



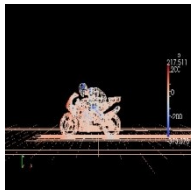
Elmfire (Åbo Akademi):
Global plasma simulation to simulate burning plasma transport

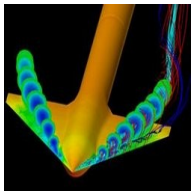
Sparse matrix size: 10Mx10M

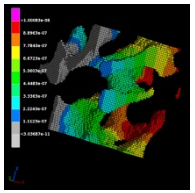


OpenFOAM (USTUTT):
Simulation of the flow in an entire hydraulic turbine using a Large Eddy Simulation

Sparse matrix size: 100M x 100M



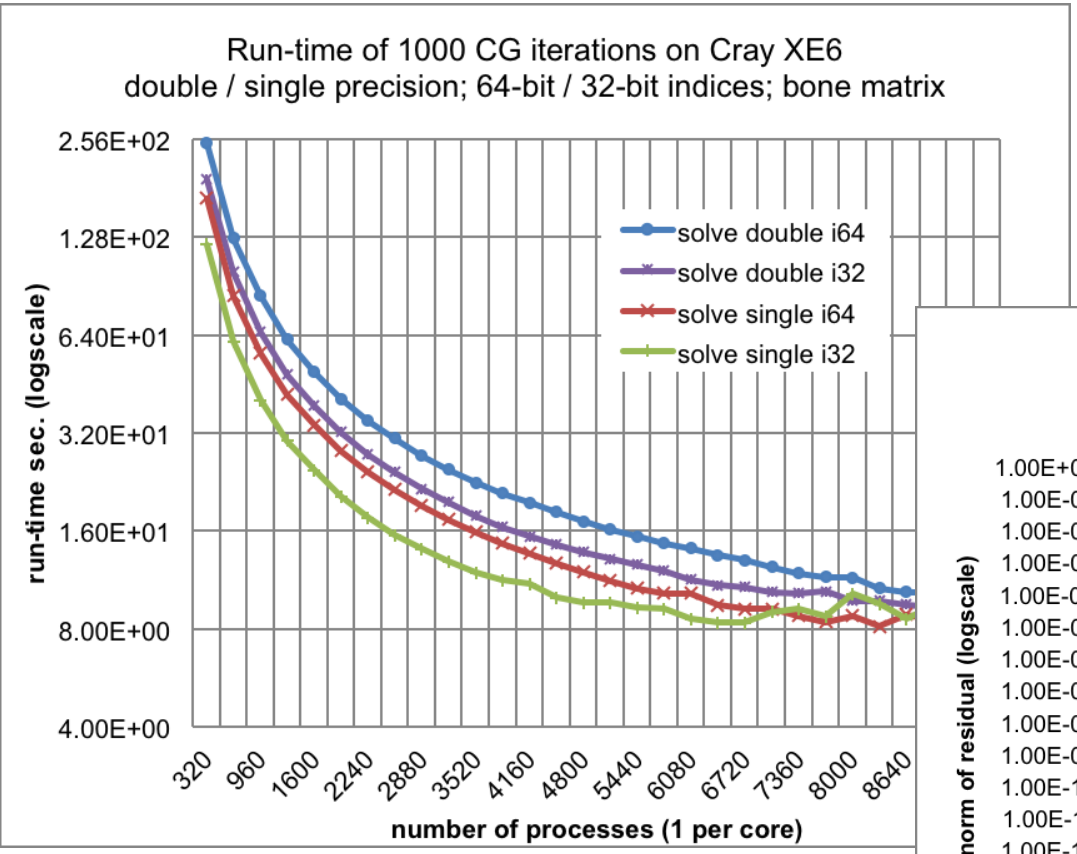




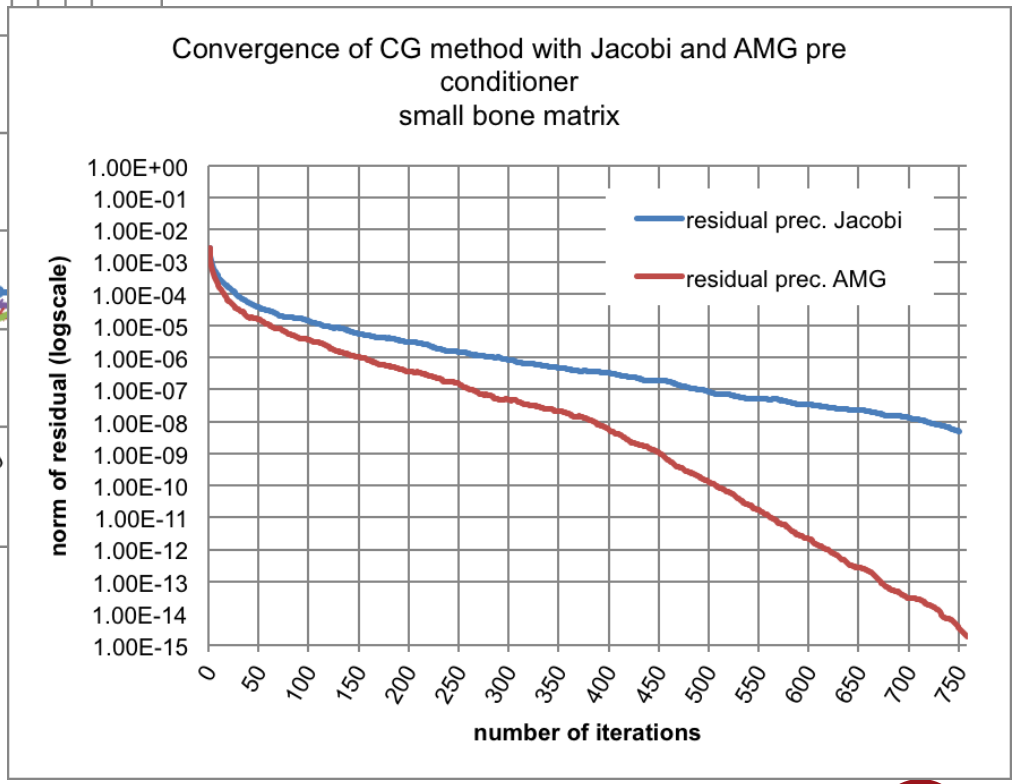
… further test cases

CRESTA

Existing algorithms and libraries



Tested linear solver:
PETsC, BoomerAMG,
Trilinos, OpenFoam

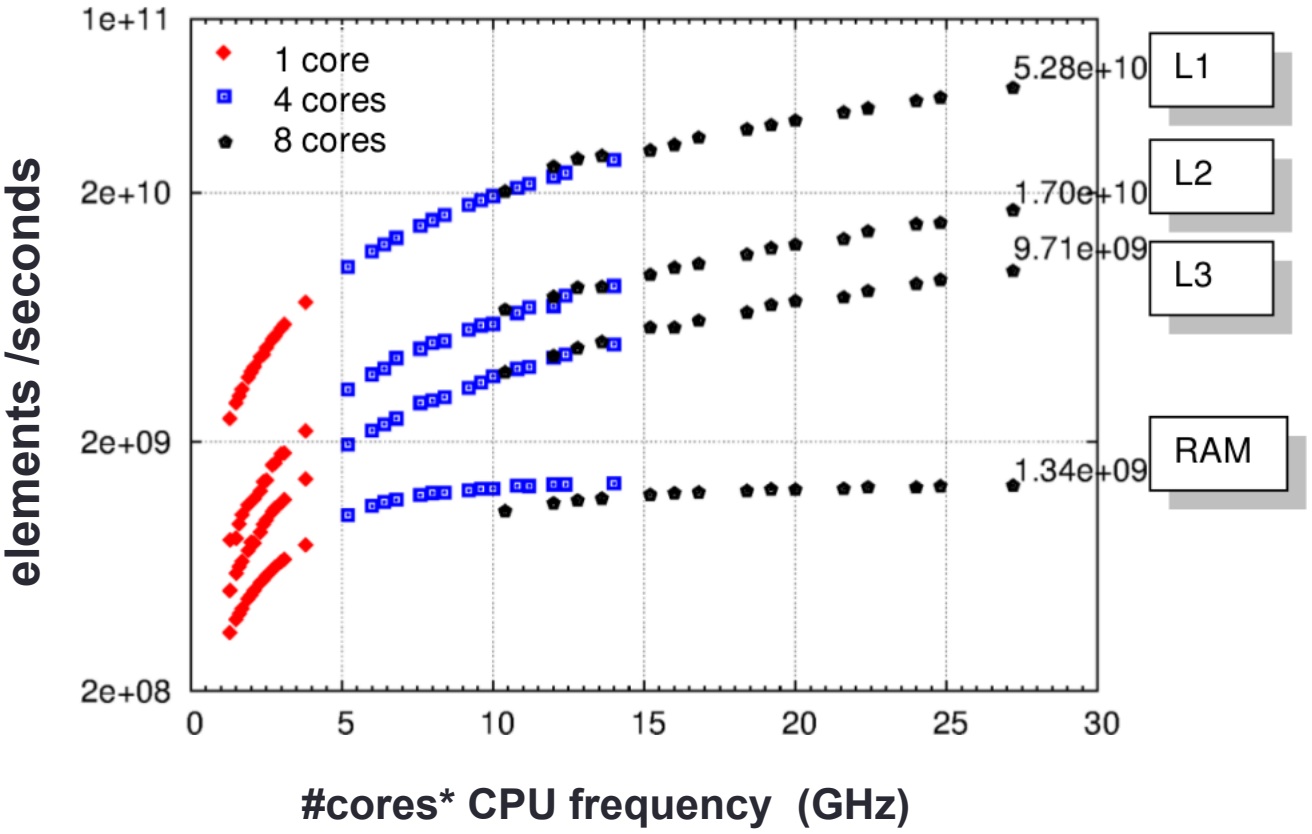


Overall performance ~ 1% of the peak

CREST

Hardware limiting factors (CPU)

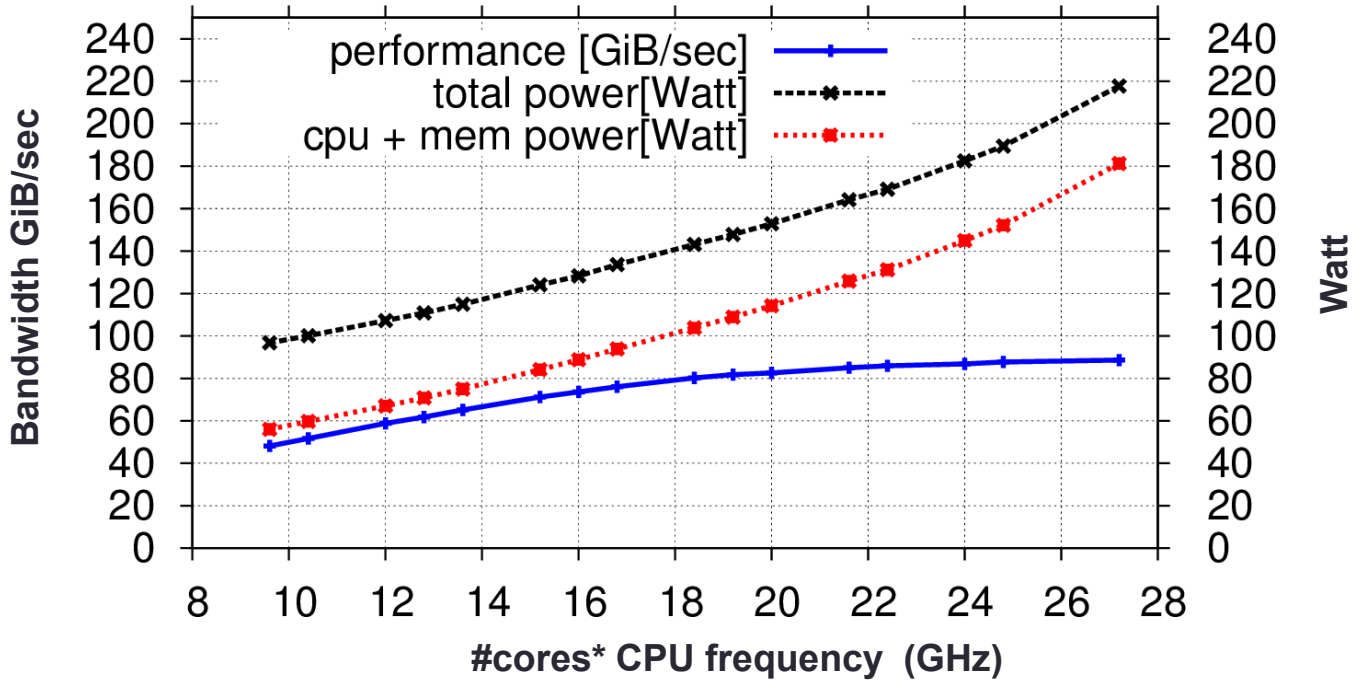
Performance of a[i]=b[i]+c[i]; data in L1, L2, L3 and RAM; 1, 4 and 8 cores
Intel E5-2687W 3.1 GHz, Turbo 3.4-3.8 GHz 8 cores, 4 mem. channels 1600 MHz



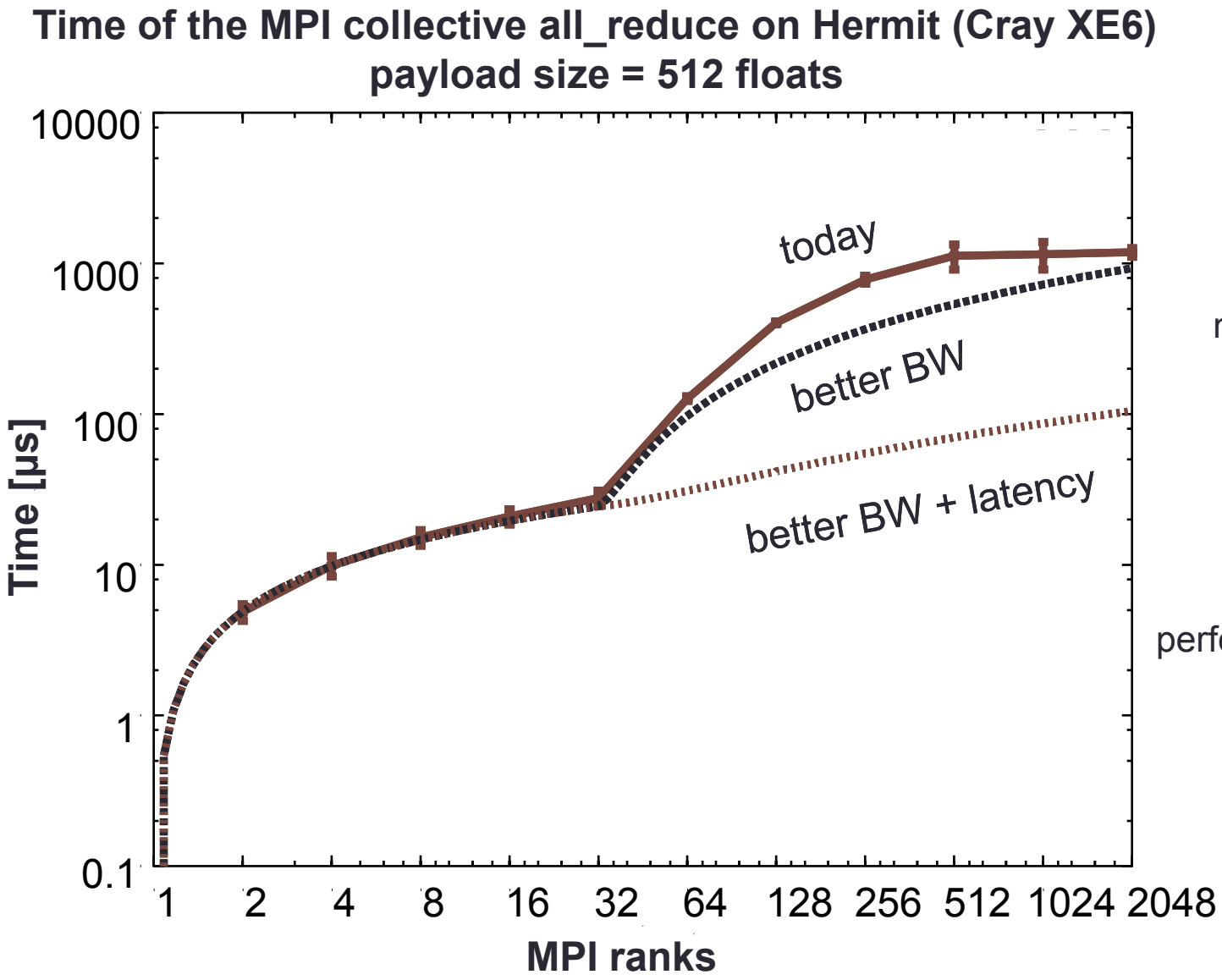
Hardware limiting factors (CPU)

Dependencies of Power, Performance and CPU Frequency sparse matrix vector multiplication

Profile of CG for sparse matrix of size (~ 10 GiB data)
Intel E5-2687W 3.1 GHz, Turbo 3.4-3.8 GHz 8 cores, 4 mem. channels 1600 MHz



Hardware limiting factors (network) – Prediction models



Increasing the network bandwidth

Only modest performance improvements

CRESTA

Hardware limiting factors – our conclusions

•The performance of each node in an HPC system will continue to rise. This is realized by the increasing number of cores and their vector units (AVX).
•An increase in the gap in performance between cache and memory can also be expected in the foreseeable future.
•The efficient use of Exascale hardware should include active fault handling in message passing.

•Vector components of the processors (by using of suitable data structures and controlling of the compiler). This increases the performance and energy efficient of today's processors. It is also essential for the next generation of the hardware.
•We must clearly distinguish between the calculations in memory and in cache.
•It's recommended to use two (by dual socket system) communication processes (mpi process) per node. The threads must be pinned to the cores.

•Use opportunities to overlap the communication and computation:
Definition of overlap availability ***O*** = (T ***C***omputation + T ***N***etwork) – ***T*** measured

$$\text{ovl\%} = \frac{(C+N)-T}{C+N-\max(C,N)} * 100 = \frac{O}{\min(C,N)} * 100$$

CRESTA

Non-Blocking and Blocking Collectives

CRESTA Collectives Microbenchmark Suite

The approach is to initially investigate the limitations of existing collective communication libraries in order to identify the key areas where completely new approaches are necessary and where optimisation can be realized focused on the implementation level.

CRESTA Collective Communication Library (CCCL)

It is a library that allows replacement of the most typical MPI collective operations with alternative CRESTA collectives, requiring only very minimal intrusion to the source code, e.g. the API arguments are exactly the same as with the original MPI collective. These CRESTA collectives perform exactly the same communication pattern and yield the same outcome as the original corresponding MPI collective.

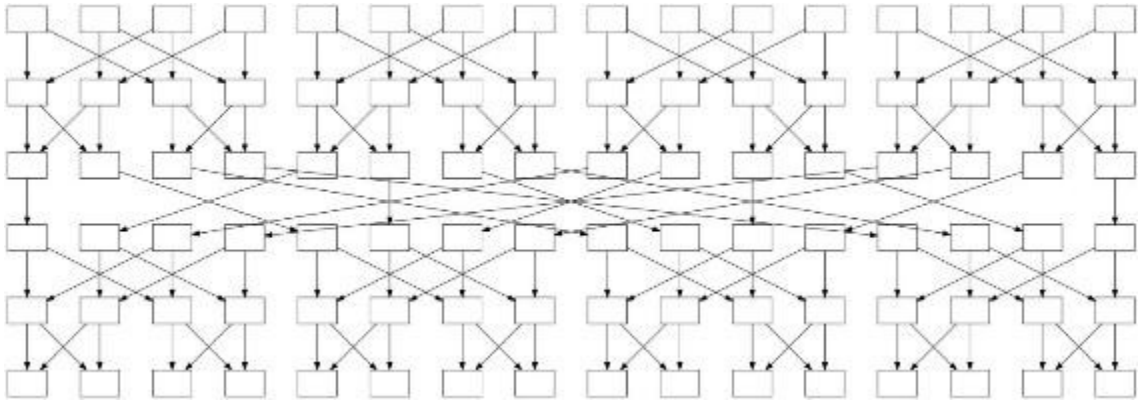
CRESTA Optimised Reduction approach

Multi precision software for collective reduction operations (long double)

CRESTA

Multi-dimensional FFT

- The most common use of Fourier transforms in HPC applications are as multi-dimensional FFTs.
- Generic library (reshape) to support changes in data decomposition that can then be used to quickly ***optimize the FFT strategy for the available hardware*** (far more general set of decompositions than most parallel libraries).



Graph representation of a (2^2x 2^2) 2D FFT

- The same code can be used to build MPI-IO file-view data types to support parallel IO to the different decompositions .
- Developed by Stephen P Booth (UEDIN)

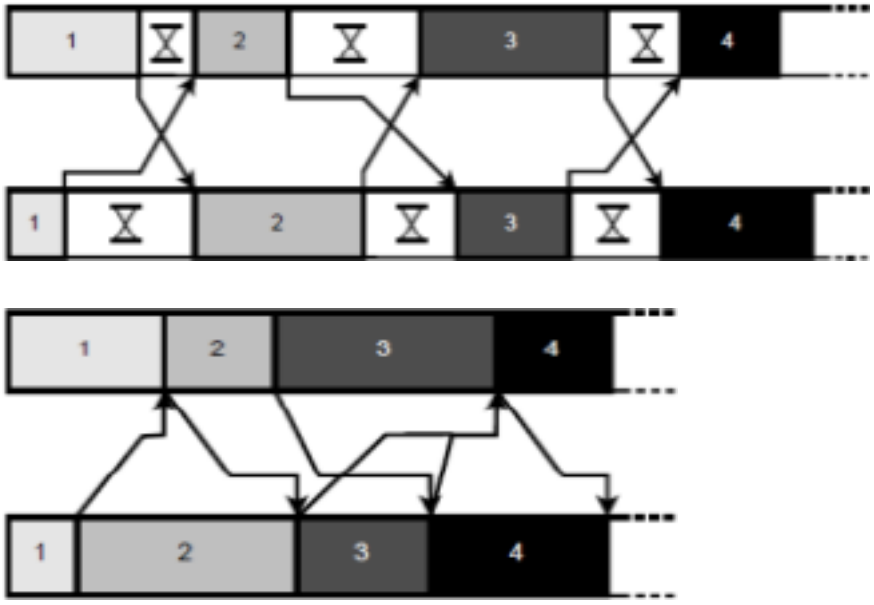
CRESTA

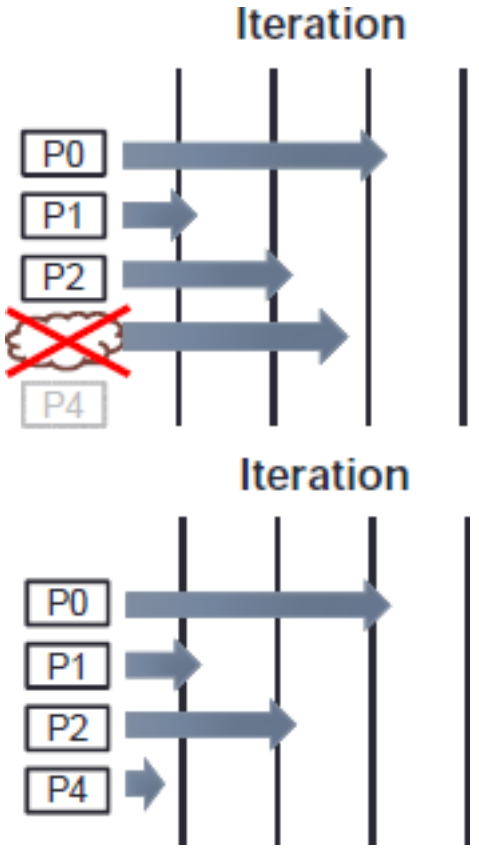
Asynchronous optimized Schwarz methods

Classical parallel iterative methods are synchronous:

- Load balancing
- Global communication
- Fault tolerance

Parallel asynchronous iterative methods don't have such problems





Developed by Frederic Magoulès (ECP)

CREST

Initial prototype of exascale algorithms and solvers

CEL-Linear-Solver

- Parallel paradigm: One communication + N worker threads (e.g. Hybrid MPI/OpenMP)
- Matrix vector multiplication with an emphasis on ***overlapping of computation and communication***
- Conjugate gradient method is implemented

Current and Further work

- Improve the ***efficiency*** (performance, scalability, power consumption)
- ***AMG preconditioner*** (according K. Stüben)
- ***Dynamic load balancing*** through varying of the number of active cores and CPU frequencies
- Linking of the library into the CRESTA co-design applications (OpenFOAM , ELMFIRE)

CRESTA

Thank you for your attention !

Magoules ,Frederic(ECP); Henty, David (EPCC); Stephen Booth(EPCC); Matura, Gregor(DLR), Niethammer, Cristoph(HLRS); Jose Gracia(HLRS); Pekka Manninen (Cray); Alistair Hart(Cray), Harvey Richardson (Cray), Dmitry Khabi(HLS) , Bastian Koller(HLRS) …

CRESTA