Method for Mapping FEM Computation onto Cluster Grid Architectures

Tomasz Olas, Roman Wyrzykowski

olas|roman@icis.pcz.pl

Institute of Computer & Information Sciences

Czestochowa University of Technology

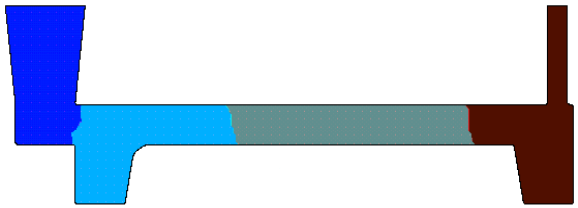
Outline

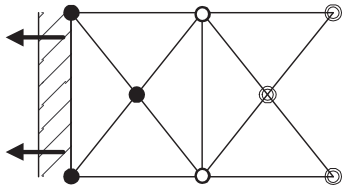
- System NuscaS for Finite Element Modeling
- Clusterix Grid
- Grid-aware parallel application
- Performance model
- Conclusions

NuscaS: Application areas

- Software package NuscaS has been developed at Czestochowa University of Technology
- NuscaS is dedicated to finite element simulation of different thermo-mechanical phenomena:
  - heat transfer
  - kinetics of solidification in castings
  - general thermo-elasto-plastic stress
  - thermo-elasto-plastic stress in solidifing castings
  - assessment of tendency to hot-tearing in castings
  - etc.

Domain decomposition aproach





internal node

boundary node

external node

constructing local system of equations

$$A_j \quad x_j \quad b_j$$

	internal nodes	boundary nodes	external nodes			
internal nodes	$A^{ii}{}_j$	$A^{ib}{}_j$	0		$x^i{}_j$	$b^i{}_j$
boundary nodes	$A^{bi}{}_j$	$A^{bb}{}_j$	$A^{be}{}_j$	X	$x^b{}_j$	$b^b{}_j$
external nodes					$x^e{}_j$	

Details of NuscaS

- To solve system of equations, the Conjugate Gradient method is used
- A version of the CG method with one synchronization point is implemented to reduce idle time of processors
- Computational kernel of the CG algorithm is the matrix-vector multiplication with sparse matrices
- Overlapping of computation and communication is exploited to reduce execution time of the algorithm

CLUSTERIX

- Aimed at developing mechanisms and tools that allow the deployment of a production Grid environment
- Basic infrastructure consists of local PC-clusters with 64-bit Linux machines located in geographically distant independent centers connected by the fast backbone provided by the Polish Optical Network PIONIER
- Existing PC-clusters, as well as new clusters with both 32- and 64-bit architecture, can be dynamically connected to the basic infrastructure
- CLUSTERIX Home Page: http://clusterix.pcz.pl

Scenarios for Grids

- For execution on Grids the following scenarios are possible:
  - Grid as the resource pool - an apropriate computational resource (local cluster) is found via resource management system, and the parallel application is started on it
  - Parallel execution on grid resources (meta-computing application):
    - Single parallel application being run on geographically remote resources
    - Grid-aware parallel application - the problem is geometrically decomposed taking into account Grid architecture

MPICH-G2

- The MPICH-G2 tool is used as a grid-enabled implementation of the MPI standard (version 1.1)
- It is based on the Globus Toolkit used for such purposes as authentification, authorization, process creation, process control, ...
- MPICH-G2 allows to couple multiple machines, potentially of different architectures, to run MPI applications
- To improve performance, it is possible to use other MPICH-based vendor implementations of MPI in local clusters (e.g. MPICH-GM)

Grid as a heterogenous system

- Hierarchical architecture of CLUSTERIX:

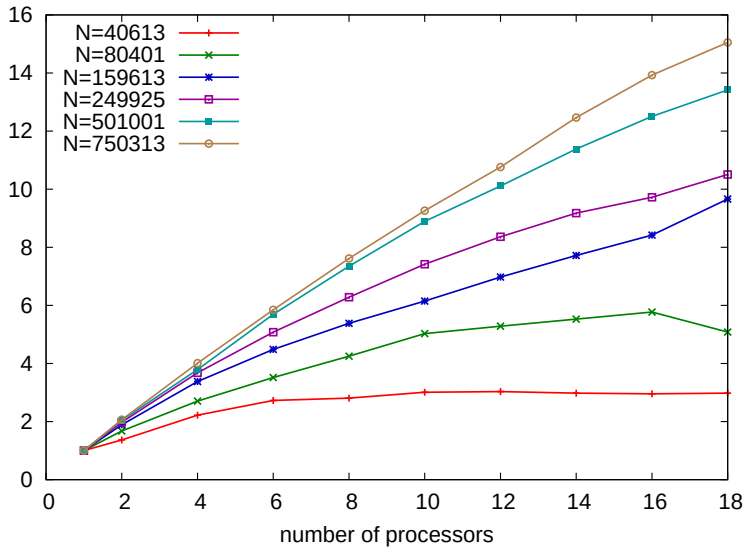
	latency	bandwitch	number of processors
single node (MPICH-G2)		5.4 $\frac{Gb}{s}$	2
local (MPICH-G2)	124 μs	745 $\frac{Mb}{s}$	6 – 32
global (MPICH-G2)	10 ms	33 $\frac{Mb}{s}$	up to 250

- It is not a trivial issue to adopt an application for its efficient execution in the CLUSTERIX environment
- Communicator construction in MPICH-G2 can be used to represent hierarchical structures of heterogenous systems, allowing applications for adaptation of their behaviour to such structures

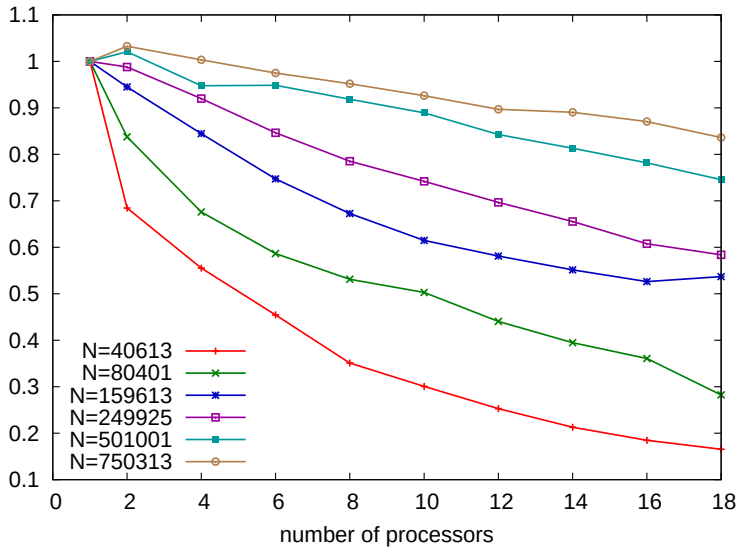
Performance results

Speedup and efficiency for the solidification of casting for diffent mesh size versus number of processors on two local clusters from CLUSTERIX (Poznan and Czestochowa)

speedup

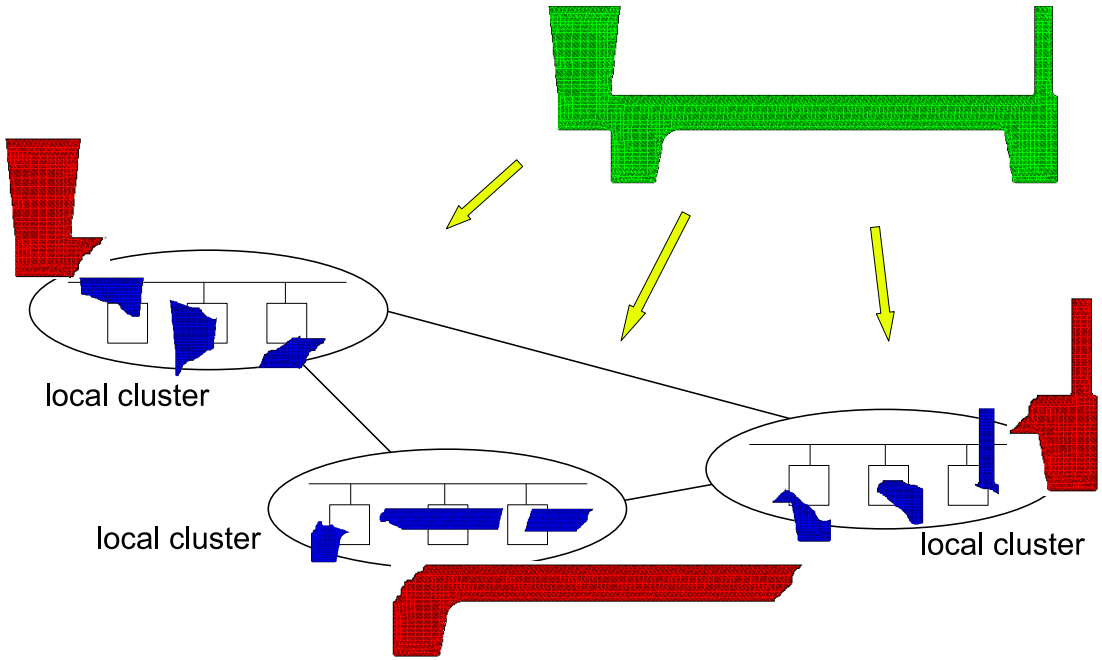


efficiency



Grid-aware parallel application

- The proposed method is based on using a two-level scheme of partitioning of FEM computational tasks, that allows for matching a local clusters engaged in computations
- This is achieved by a suitable decomposition of FEM meshes into submeshes assigned to separate clusters in the grid
- These submeshes are then divided into smaller parts which are distributed among nodes in the same site



Performance model I

- Parallel execution time:

$$T_p = \frac{1}{p}\left(\sum_{i=0}^{p-1} T_{comp}^i + \sum_{i=0}^{p-1} T_{comm}^i + \sum_{i=0}^{p-1} T_{idle}^i\right)$$

- Estimation of computation time:

$$T_{comp} = n_{op} T_f$$

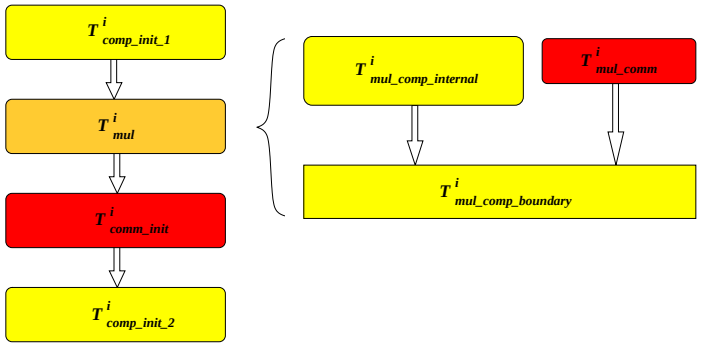
- Estimation of communication time:

$$T_{comm} = \alpha + \beta n$$

Performance model II

- Parallel execution time for solving linear system using CG method:

$$T_p = \frac{1}{p}(n_{iter}+1)\left(\sum_{i=0}^{p-1} T^i_{comp\_init} + \sum_{i=0}^{p-1} T^i_{comm\_init} + \sum_{i=0}^{p-1} T^i_{mul} + \sum_{i=0}^{p-1} T^i_{idle}\right)$$

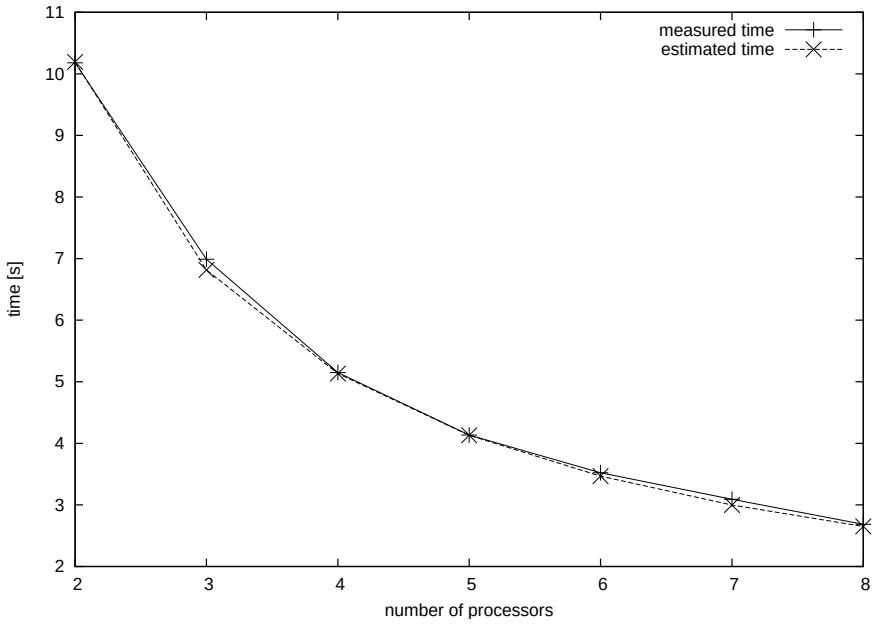


- Final formula:

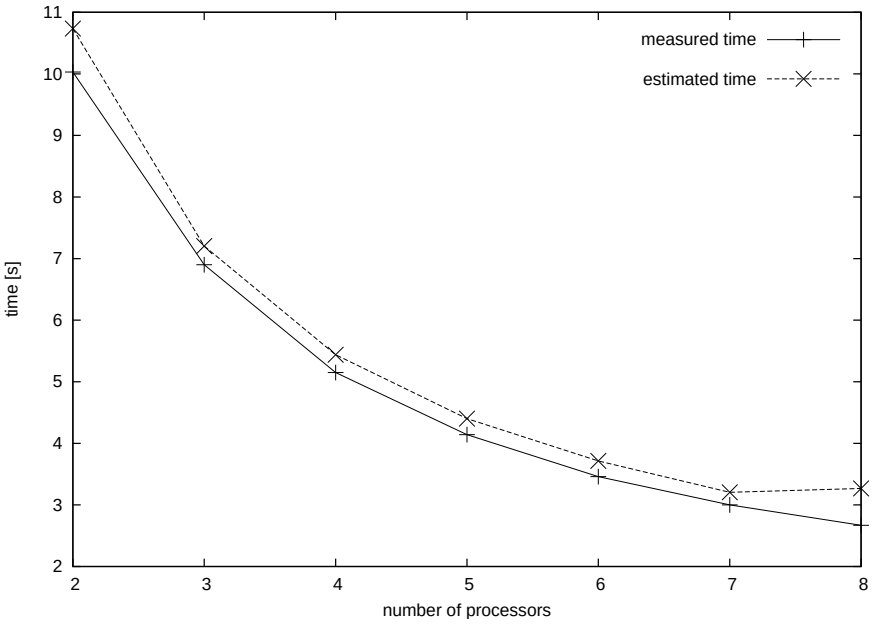
$$T_p = (n_{iter}+1)\Big((11n_l + 4n + 10)T_f + 2p(\alpha + 3\beta) + max(2(s_w + s_o)\alpha + (n_e + n_b)\beta, 2nz_iT_f) + 2nz_bT_f\Big)$$

Validation of performance model

Intel Pentium III



AMD Athlon



Performance model for grids I

- The main difference between the standard approach to FEM parallelization and grid-aware method is in matrix-vector multiplication phase
- For the standard approach, communication time of matrix-vector multiplication running on grid architecture is:

$$T^{i}_{mul\_comm} = k\alpha_g + zk\beta_g,$$

where:

- k - number of messages sent between nodes in separate local clusters
- z - average size of transferred message (it depends on problem size)
- α_g - message start-up time (WAN)
- β_g - per-word transfer time (WAN)

Performance model for grids II

- In the proposed method for mapping FEM computations onto cluster grid architecture, communication time is reduced to:

$$T^{i}_{mul\_comm} = 2\alpha_g + zk\beta_g + \frac{k}{2}\alpha + zk\beta$$

- Communications inside local clusters are much more efficient, and can be ommitted
- In that case potential time reduction can be written as:

$$T_r = k\alpha_g - 2\alpha_g$$

Conclusions

- A method for mapping FEM computations onto cluster gird architecture was presented
- The proposed metod for mapping FEM computations onto cluster grid architectures allows for the efficient execution of tasks on grids containing much more processors than in case of using the traditional approach
- Alternatively, using a fixed number of processors is it becomes possible to run efficiently even tasks with a smaller size
- The proposed performance model makes possible to estimate benefits of using the proposed approach, in an analytic way
- Implementation of the proposed method in the NuscaS package is under development