

Urychlení návrhu a optimalizace teplotních senzorů s využitím HPC technologií

Tomáš Kozubek
IT4Innovatins, VŠB-TUO

Konference e-infrastruktury CESNET, 29-30. ledna 2019

Contract research 2018



SKODA



BORCAD®

K2
atmitec



bringing life to maps



brose

itica

Continental 

T-Mobile

BAYNCORE



inset

SIEMENS
New R&D centre in Ostrava

Motivation for Exascale - Digital Twin Technology

Complex nonlinear multiphysical and multiscale problem – electric motor

- Electric fields
- Electromagnetism
- Heat transfer
 - heat generated by magnetism
 - cooling system
- Structural Mechanics
 - structural integrity
 - vibration from motion
 - high speed motors
 - influenced by electromagnetism
- Active cooling system
 - fluid flow
- Acoustic
 - generated by fluid flow
 - generated by electromagnetism
 - generated by vibrations



Multiphysics simulations

Simulation results

- Complex velocity and pressure fields
 - Ventilation losses
 - Cooling efficiency

One simulation

- 35 hours / 1200 Cores on Salomon
- “Standard” powerful workstation
- 256GB RAM 30 Cores 5 ~2 months!

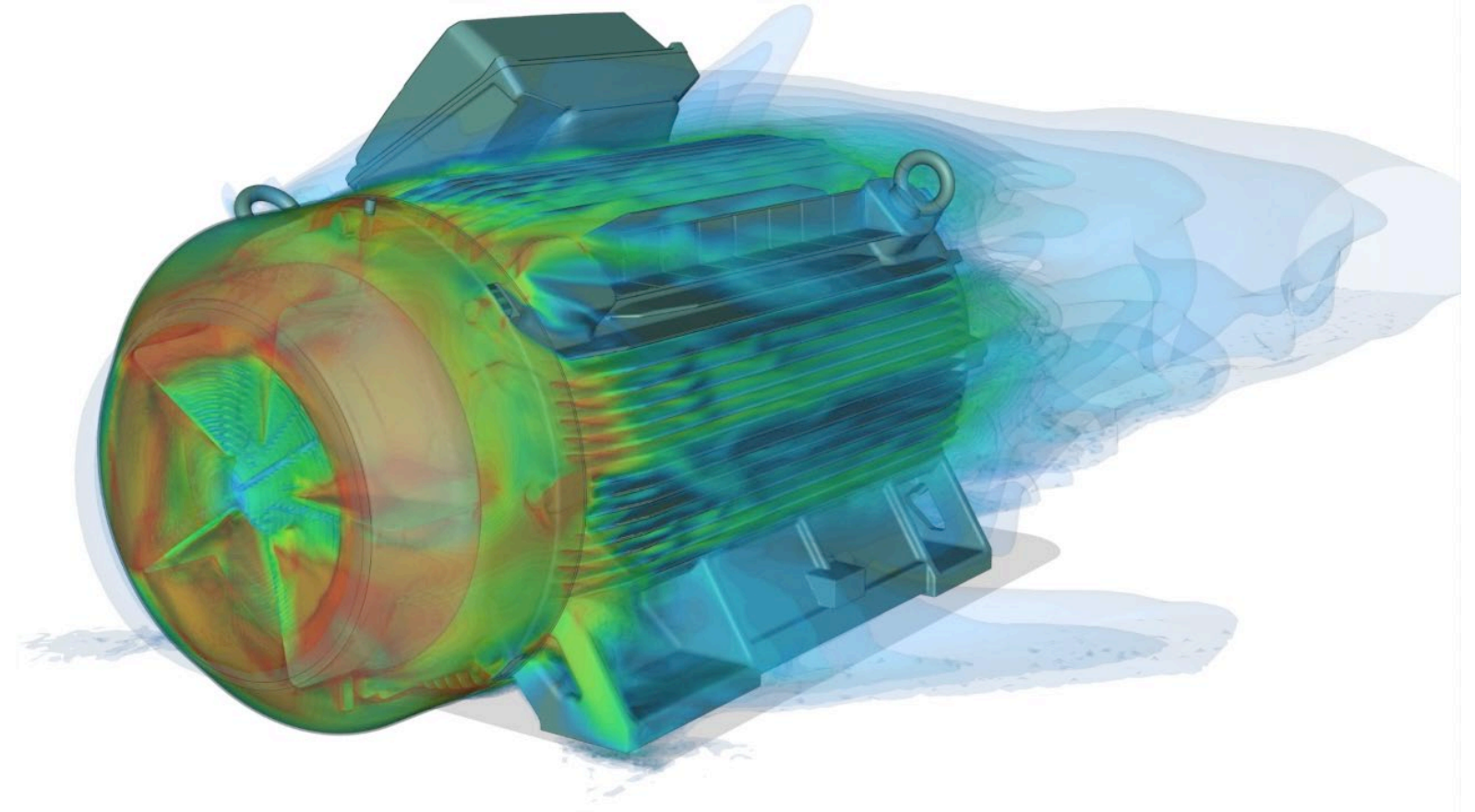
What we need for optimization?

- Geometry morphing
- different number of ribs
- different fan shape
- boundary conditions variation
- ...

Lead's to 1000's simulations

One cluster isn't enough

Full transient CFD simulation of the active cooling system



Software for engineering simulations

- Creating complex model, computational mesh, boundary condition definition, material models – pre-processing
- Solution by numerical methods
- Results analysis – post-processing

1. Open source: OpenFOAM, Code Saturne, SU², Elmer, Dune, ...

https://www.cfd-online.com/Wiki/Codes#Free_codes

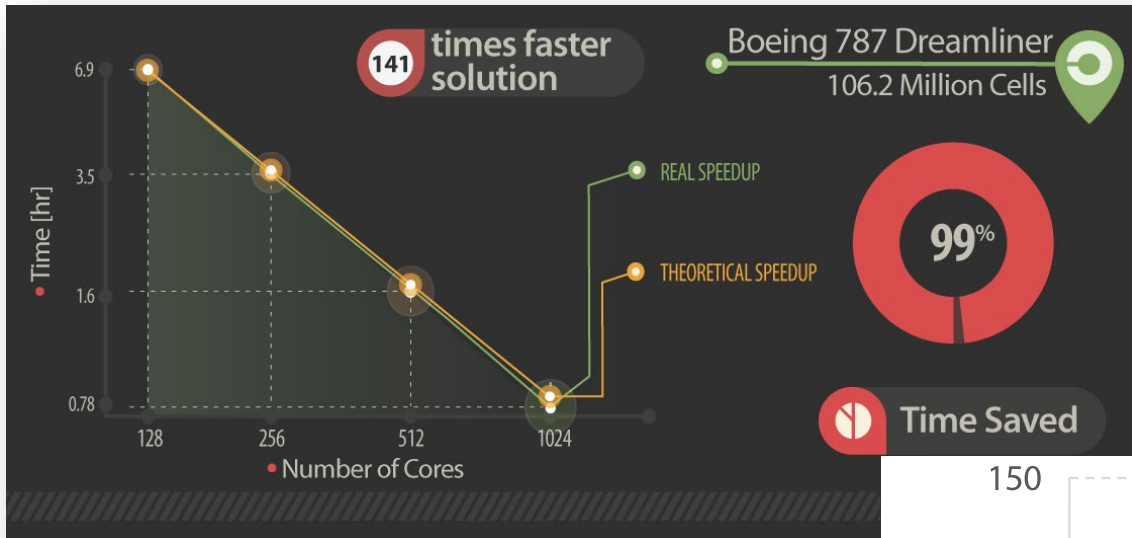
2. Commercial codes: ANSYS CFD, FLUENT, CFX, StarCCM+, COMSOL, ...

https://www.cfd-online.com/Wiki/Codes#Commercial_codes

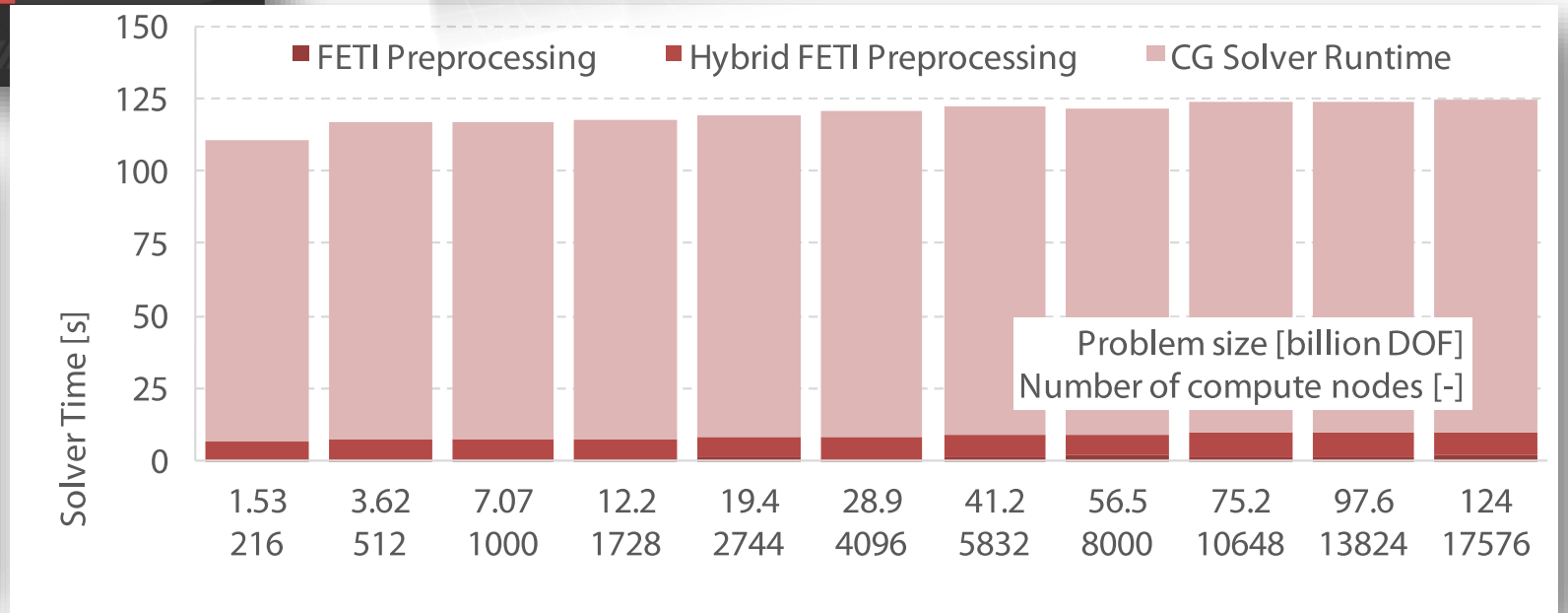
3. Or try to write it in your own way

Matlab, R, Octave, Python, C, C++, Fortran, ...

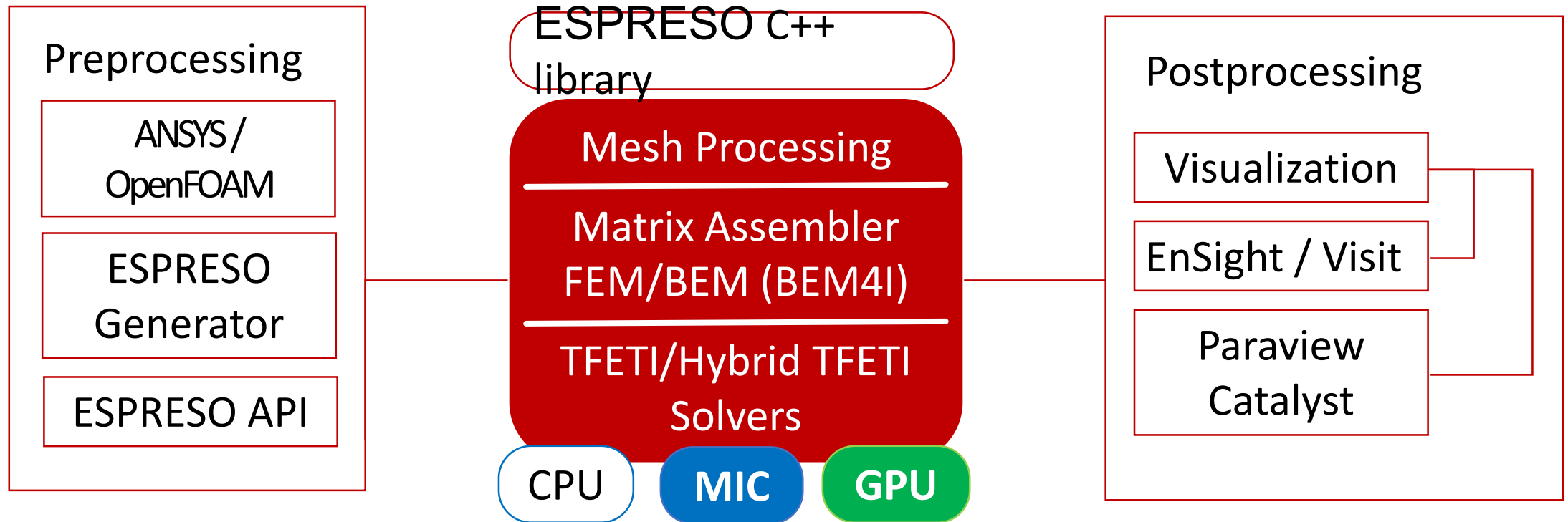
Scalable algorithms development



TOP 500 #9
The List.



ESPRESO parallel solver



MPI, cilk++(or openMP), MKL, METIS, PARDISO(in MKL) or PARDISO SC for MIC/GPU version
Optional: SCOTCH, MUMPS, CuSolver, VTK, HYPRE interface

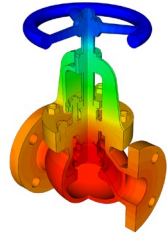
ESPRESO FEM Highly parallel FEM package for engineering simulations

Heat Transfer Module Capability List:

Load steps definition for combination of multiple steady-state and time dependent analyses

Transient solvers

- Generalized trapezoidal rule
- Automatic time stepping based on response frequency approach

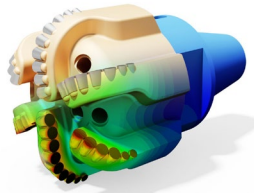


Nonlinear solvers

- Newton Raphson – full and symmetric
- Newton Raphson with constant tangent matrices
- Line search damping
- Sub-steps definition
- Adaptive precision control for iterative solvers

Linear and quadratic finite element discretization

Gluing nonmatching grids by mortar discretization techniques



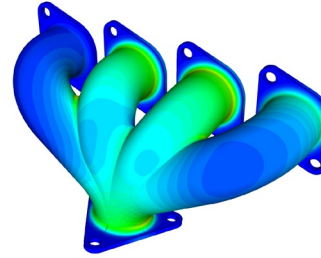
Full-fledged material models

- nonlinear materials
- isotropic, orthotropic and anisotropic material models
- materials for phase change

Element coordinate system definition – cartesian, polar and spherical

Temperature and time dependent boundary conditions

- linear convection
- nonlinear convection
- heat flow
- heat flux
- diffuse radiation
- heat source
- translation motion



Consistent SUPG and CAU stabilization for Translation Motion (advection), Inconsistent stabilization

Phase Change based on apparent heat capacity method

Boundary element discretization for selected physical applications

Highly parallel multilevel FETI domain decomposition based solver for billions of unknowns for symmetric and non-symmetric systems with accelerators support and combination of MPI and OpenMP techniques

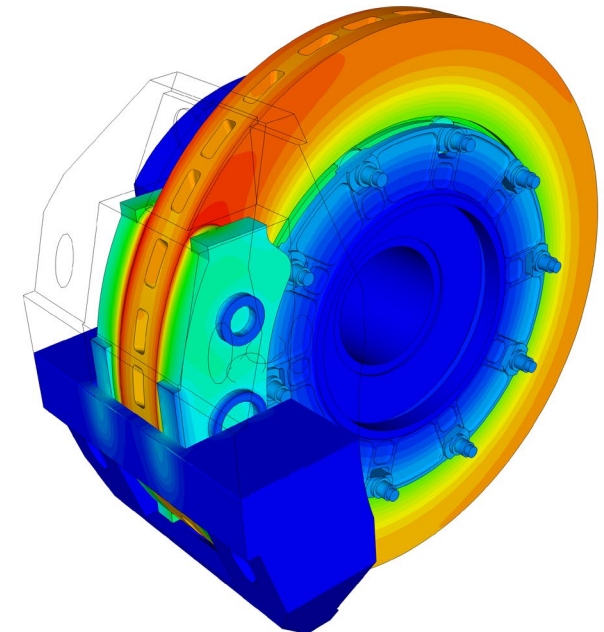
Asynchronous parallel I/O

Input mesh format from popular open source and commercial packages like OpenFOAM, ELMER or ANSYS

Output to commonly used post-processing formats, VTK and EnSight

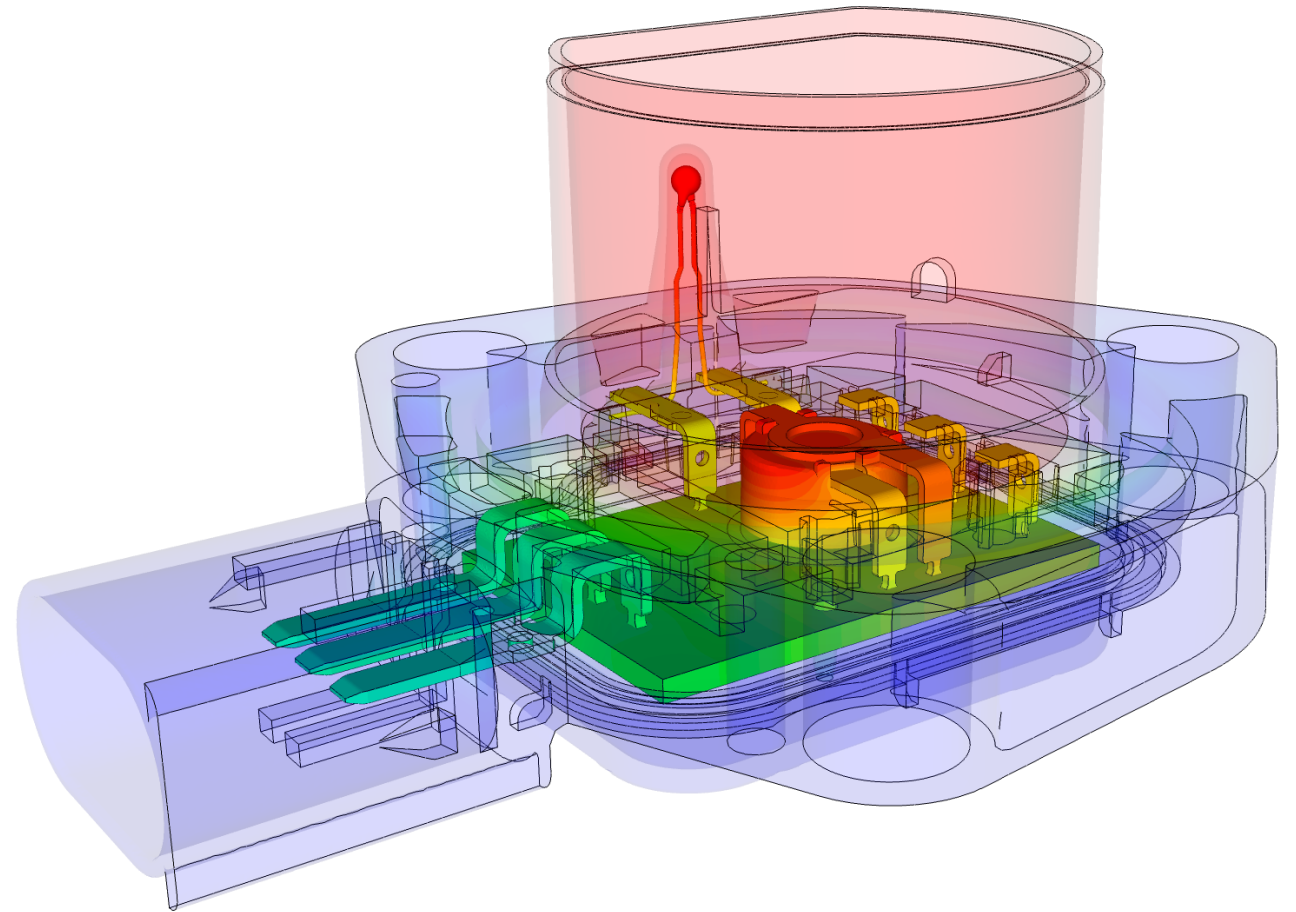
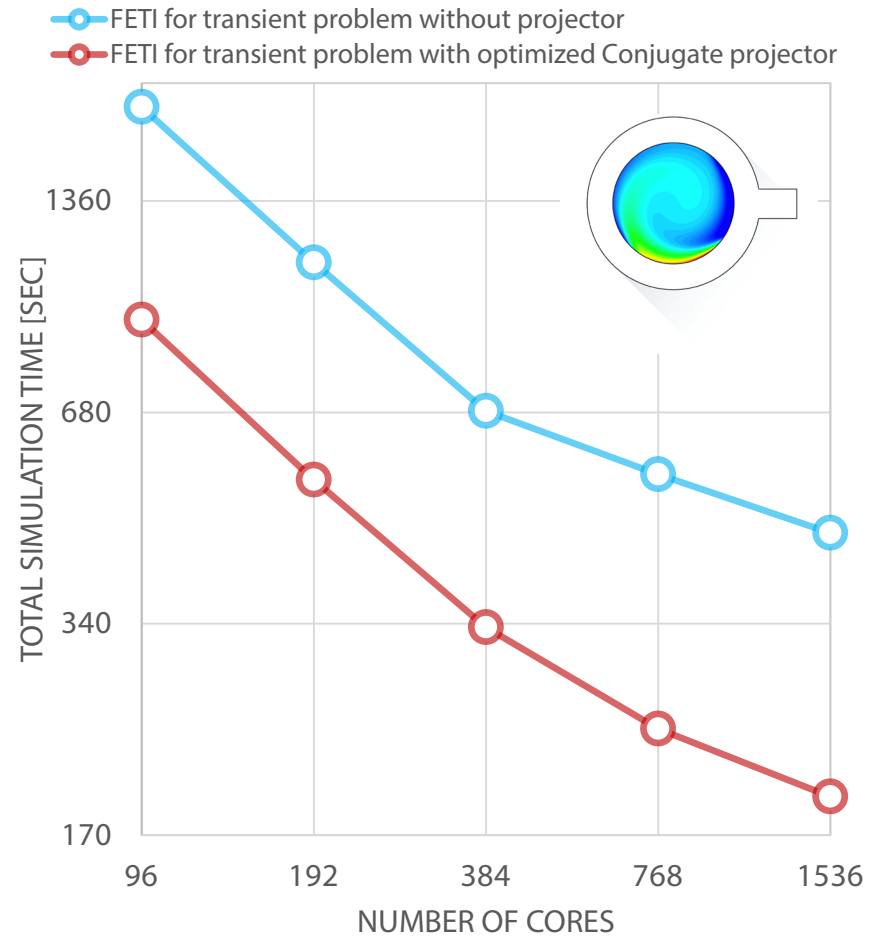
Monitoring results on selected regions for statistic and optimization toolchain

Simple text Espresso Configuration File (ecf) for setting all ESPRESO FEM solver parameters without GUI. Control each parameter in ecf file from command line



Heat transfer

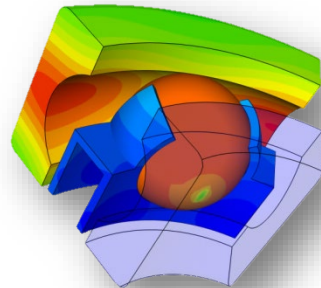
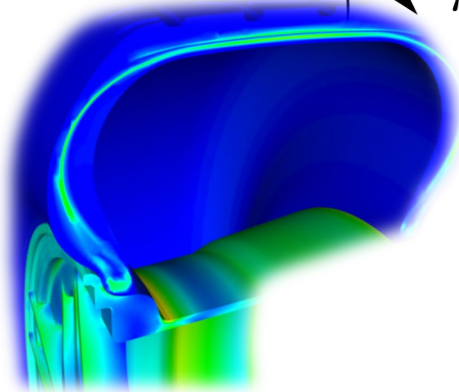
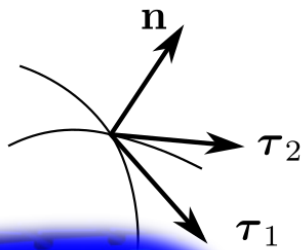
Response time optimization of the USL sensor - nonlinear transient simulation



Structural mechanics

Elasticity BVP

$$\begin{aligned}
 -\operatorname{div} \boldsymbol{\sigma}^{(k)} &= \mathbf{f}^{(k)} && \text{in } \Omega^{(k)} \\
 \sigma_{ij}^{(k)} &= c_{ijkl} \varepsilon_{kl}^{(k)} && \text{in } \Omega^{(k)} \\
 \varepsilon^{(k)} &= \frac{1}{2} (\nabla \mathbf{u} + \nabla^\top \mathbf{u}) && \text{in } \Omega^{(k)} \\
 \mathbf{u}^{(k)} &= \mathbf{0} && \text{on } \Gamma_D^{(k)} \\
 \boldsymbol{\sigma}^{(k)} \mathbf{n}^{(k)} &= \mathbf{t}^{(k)} && \text{on } \Gamma_N^{(k)}
 \end{aligned}$$



Terms on contact boundary

$$\begin{aligned}
 -g(\mathbf{X}) &= u_n(\mathbf{X}) - d(\mathbf{X}) \\
 \mathbf{t}_C^m d\gamma_C^m &= -\mathbf{t}_C^s d\gamma_C^s \\
 \mathbf{t}_C^s &= t_n \mathbf{n} + t_{1T} \boldsymbol{\tau}_1 + t_{2T} \boldsymbol{\tau}_2
 \end{aligned}$$

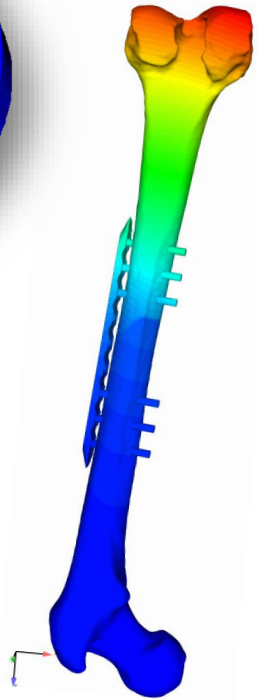
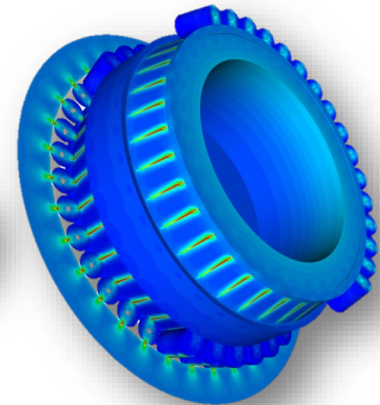
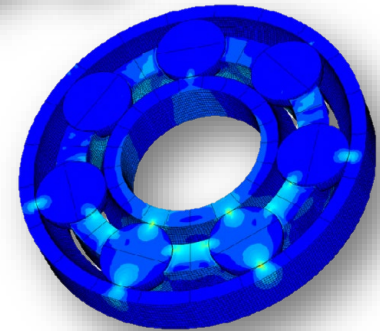
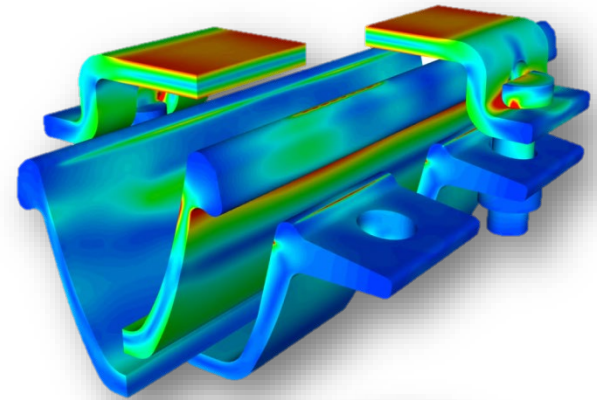
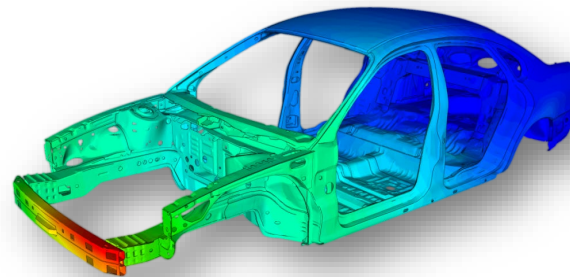
Unilateral contact (non-penetration)

$$u_n - d \leq 0, \quad t_n \leq 0, \quad t_n(u_n - d) = 0$$

Friction (Tresca or Coulomb)

$$\|\mathbf{t}_T\|_2 \leq F, \quad \begin{cases} \|\mathbf{t}_T\|_2 < F & \Rightarrow \mathbf{u}_T = \mathbf{0} \\ \|\mathbf{t}_T\|_2 = F & \Rightarrow \mathbf{u}_T = -c \mathbf{t}_T, \quad c \geq 0 \end{cases}$$

$$F = \begin{cases} sb & \dots \text{Tresca} \\ \mathcal{F} |t_n| & \dots \text{Coulomb} \end{cases}$$

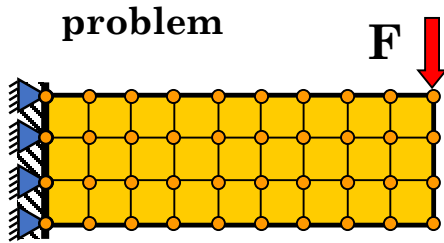


A blue circle with a white number '1' inside, positioned at the top center of the slide. The circle has a slight drop shadow.

1

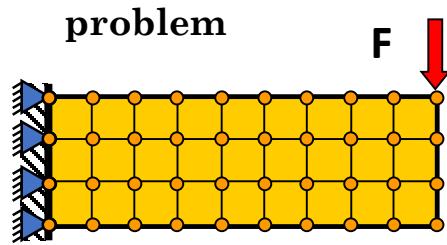
Parallelization by DDM/FETI

FETI approaches

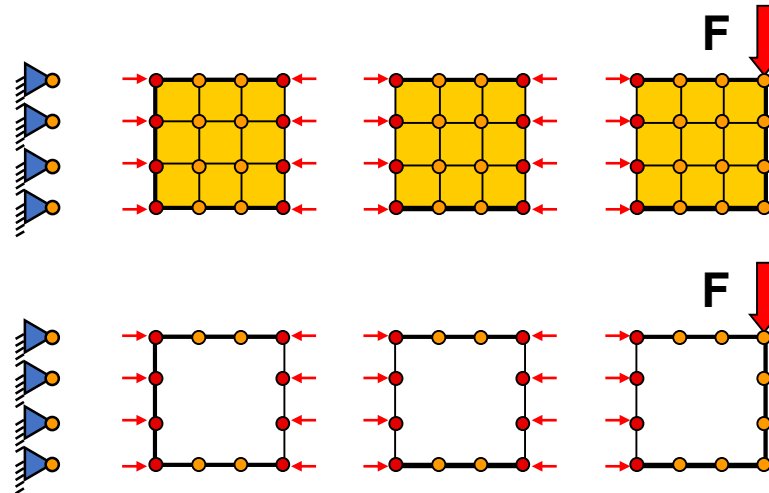


<p>1. FETI</p>		<p>subdomains are fixed or free but with different defects</p>
<p>2. FETI-DP</p>		<p>FETI-DP (partial splitting, nonsingular)</p>
<p>3. TFETI</p>		<p>all subdomains are free with the same defect</p>

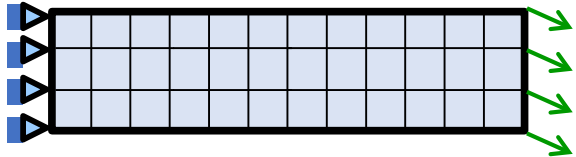
BETI approach



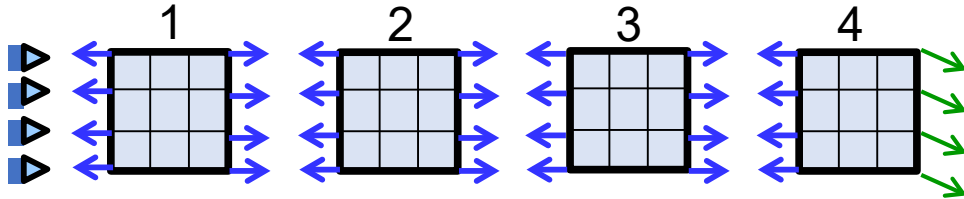
TFETI/TBETI



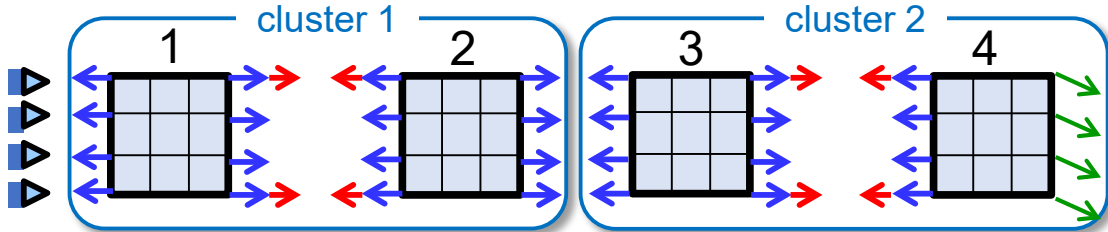
Hybrid Total FETI Method



FEM discretization

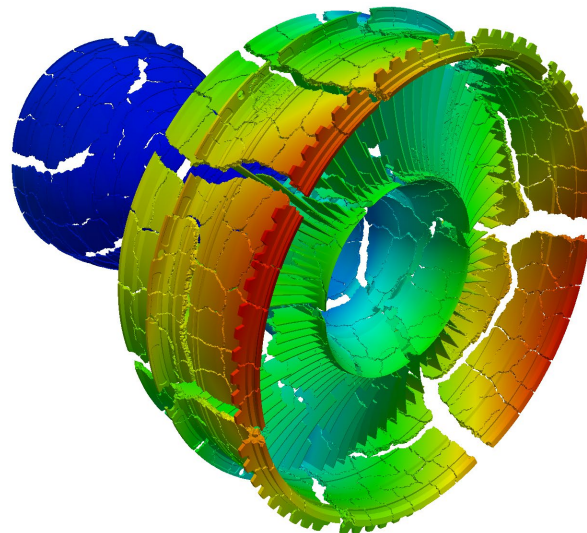
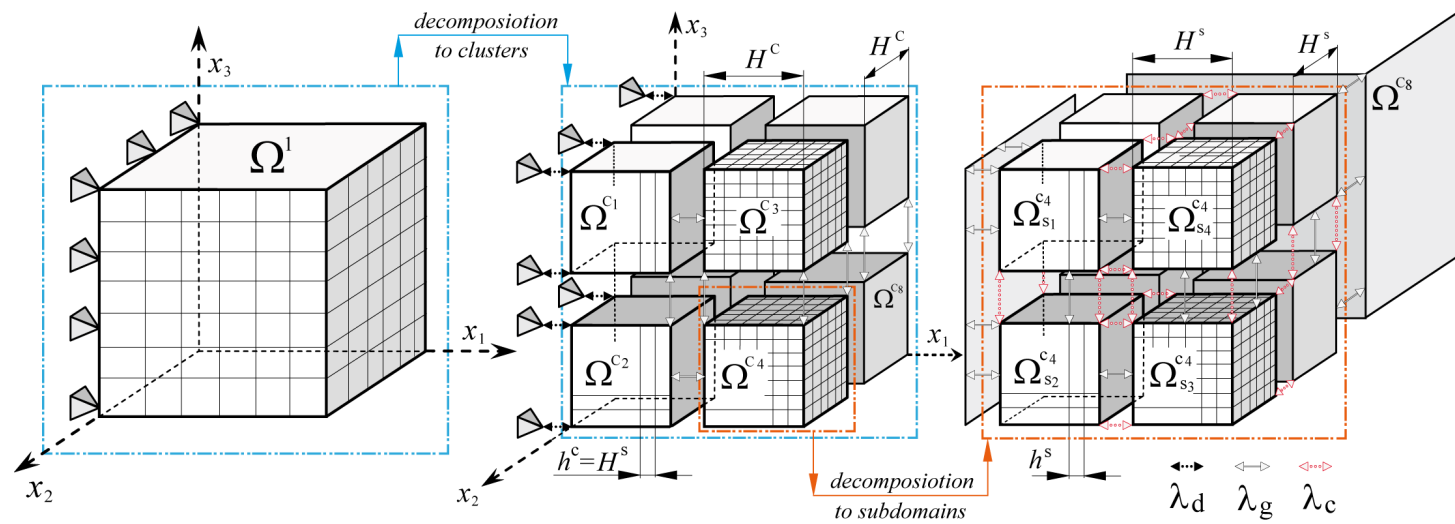


FETI method

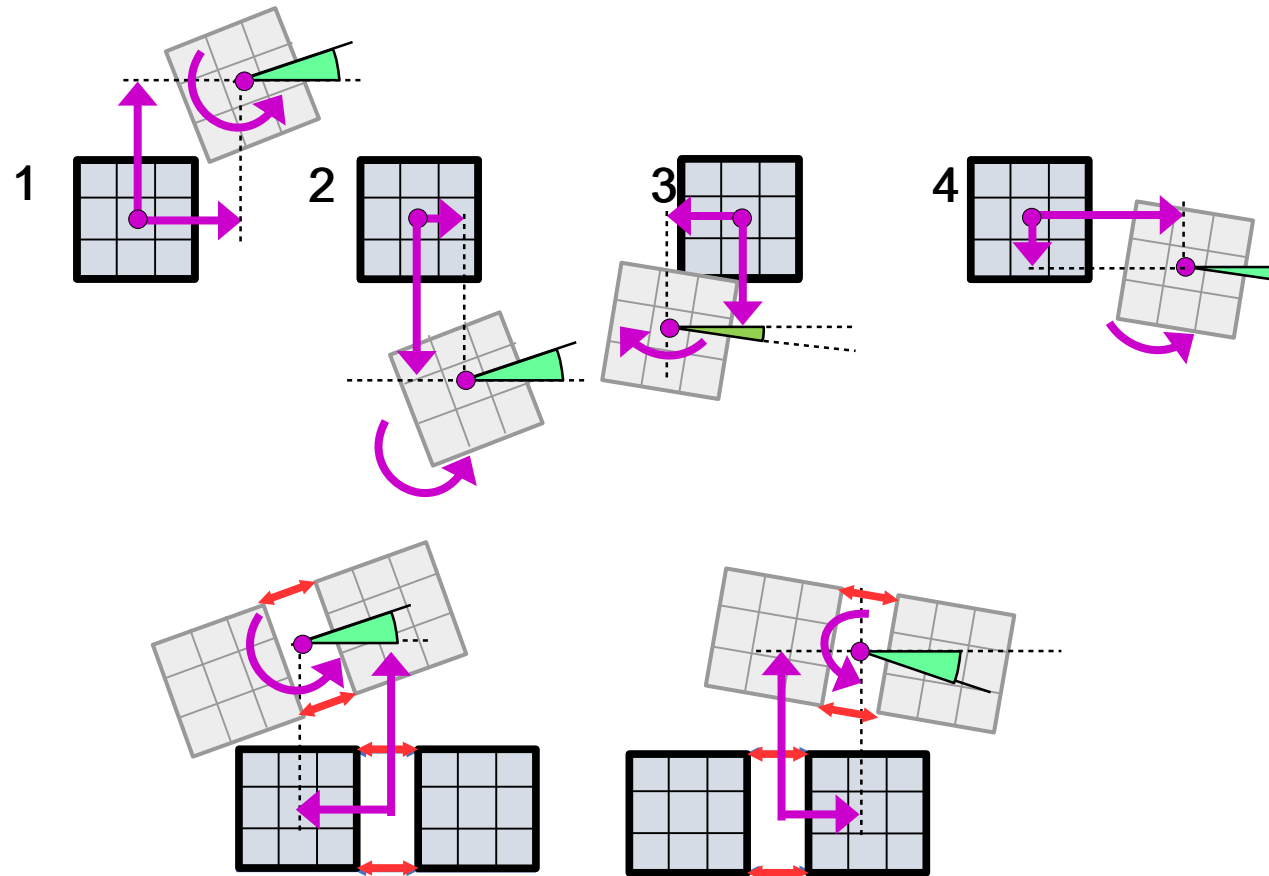


Hybrid FETI method

Hybrid Total FETI Method



Hybrid Total FETI Method



In Hybrid Total FETI method the dimension of GG^T is smaller (dim=6) compared to Total FETI (dim=12) due to additional constraints which remove local rigid body modes.

Hybrid Total FETI Method

Implementation

$$\mathbf{B}_c = \begin{pmatrix} \mathbf{B}_{c,1} & \mathbf{B}_{c,2} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,3} & \mathbf{B}_{c,4} \end{pmatrix}, \quad \mathbf{B} = (\mathbf{B}_1 \ \mathbf{B}_2 \ \mathbf{B}_3 \ \mathbf{B}_4)$$

additional constraints:
duplication of 'corners' Lagrange multipliers

$$\begin{pmatrix} \mathbf{K}_1 & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,1}^\top & \mathbf{O} & \mathbf{B}_1^\top \\ \mathbf{O} & \mathbf{K}_2 & \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,2}^\top & \mathbf{O} & \mathbf{B}_2^\top \\ \mathbf{O} & \mathbf{O} & \mathbf{K}_3 & \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,3}^\top & \mathbf{B}_3^\top \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{K}_4 & \mathbf{O} & \mathbf{B}_{c,4}^\top & \mathbf{B}_4^\top \\ \hline \mathbf{B}_{c,1} & \mathbf{B}_{c,2} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,3} & \mathbf{B}_{c,4} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \hline \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \mathbf{B}_4 & \mathbf{O} & \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \mathbf{u}_4 \\ \lambda_{c,1} \\ \lambda_{c,2} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \\ \mathbf{o} \\ \mathbf{o} \\ \mathbf{c} \end{pmatrix}$$

augmented KKT system
 by the matrix \mathbf{B}_c and λ_c

$$\begin{pmatrix} \mathbf{K}_1 & \mathbf{O} & \mathbf{B}_{c,1}^\top & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{B}_1^\top \\ \mathbf{O} & \mathbf{K}_2 & \mathbf{B}_{c,2}^\top & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{B}_2^\top \\ \mathbf{B}_{c,1} & \mathbf{B}_{c,2} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{K}_3 & \mathbf{O} & \mathbf{B}_{c,3}^\top & \mathbf{B}_3^\top \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{K}_4 & \mathbf{B}_{c,4}^\top & \mathbf{B}_4^\top \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,3} & \mathbf{B}_{c,4} & \mathbf{O} & \mathbf{O} \\ \hline \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{O} & \mathbf{B}_3 & \mathbf{B}_4 & \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \lambda_{c,1} \\ \mathbf{u}_3 \\ \mathbf{u}_4 \\ \lambda_{c,2} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{o} \\ \mathbf{f}_3 \\ \mathbf{f}_4 \\ \mathbf{o} \\ \mathbf{c} \end{pmatrix}$$

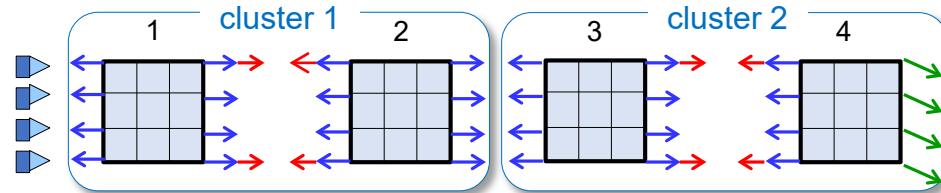
reordering according to clusters

Hybrid Total FETI Method

Implementation

$$\mathbf{B}_c = \begin{pmatrix} \mathbf{B}_{c,1} & \mathbf{B}_{c,2} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{B}_{c,3} & \mathbf{B}_{c,4} \end{pmatrix}, \quad \mathbf{B} = (\mathbf{B}_1 \ \mathbf{B}_2 \ \mathbf{B}_3 \ \mathbf{B}_4)$$

additional constraints (red arrows)



The matrix \mathbf{B}_c is a copy of specific rows from the matrix \mathbf{B} corresponding to components of λ acting on the corners between subdomains 1,2, and 3,4, respectively (red arrows).

Hybrid Total FETI Method

Modified KKT system is solvable by the same family of algorithms like FETI (Preconditioned Conjugate Projected Gradient method, ...).

Size of $\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T$ (or defect of global stiffness matrix) can be **significantly smaller** than in original FETI approach ...

$$\begin{pmatrix} \tilde{\mathbf{F}} & \tilde{\mathbf{G}} \\ \tilde{\mathbf{G}}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \tilde{\lambda} \\ \tilde{\alpha} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{d}} \\ \tilde{\mathbf{e}} \end{pmatrix}$$

$$\tilde{\mathbf{K}} = \text{diag}(\tilde{\mathbf{K}}_1, \tilde{\mathbf{K}}_2)$$

$$\tilde{\mathbf{R}}^T = (\tilde{\mathbf{R}}_1^T, \tilde{\mathbf{R}}_2^T)$$

$$\tilde{\mathbf{B}} = (\tilde{\mathbf{B}}_1, \tilde{\mathbf{B}}_2)$$

$$\tilde{\mathbf{F}} = \tilde{\mathbf{B}}\tilde{\mathbf{K}}^+\tilde{\mathbf{B}}^T$$

$$\tilde{\mathbf{d}} = \tilde{\mathbf{B}}\tilde{\mathbf{K}}^+\tilde{\mathbf{f}} - \tilde{\mathbf{c}}$$

$$\tilde{\mathbf{G}} = -\tilde{\mathbf{B}}\tilde{\mathbf{R}}$$

$$\tilde{\mathbf{e}} = -\tilde{\mathbf{R}}^T\tilde{\mathbf{f}}^T$$

A blue circle with a white number '2' inside, casting a soft shadow to the right and bottom.

2

Pipelining

Bi-Conjugate Gradients Stabilized (BiCGStab)

Algorithm 4 Standard BiCGStab

```
1: function BICGSTAB( $A, b, x_0$ )
2:    $r_0 := b - Ax_0; p_0 := r_0$ 
3:   for  $i = 0, \dots$  do
4:      $s_i := Ap_i$ 
5:     compute  $(r_0, s_i)$ 
6:      $\alpha_i := (r_0, r_i) / (r_0, s_i)$ 
7:      $q_i := r_i - \alpha_i s_i$ 
8:      $y_i := Aq_i$ 
9:     compute  $(q_i, y_i) ; (y_i, y_i)$ 
10:     $\omega_i := (q_i, y_i) / (y_i, y_i)$ 
11:     $x_{i+1} := x_i + \alpha_i p_i + \omega_i q_i$ 
12:     $r_{i+1} := q_i - \omega_i y_i$ 
13:    compute  $(r_0, r_{i+1})$ 
14:     $\beta_i := (\alpha_i / \omega_i) (r_0, r_{i+1}) / (r_0, r_i)$ 
15:     $p_{i+1} := r_{i+1} + \beta_i (p_i - \omega_i s_i)$ 
16:  end for
17: end function
```

dot-prod
SpMV
axpy

Traditional BiCGStab:

(non-preconditioned)

Global communication

- ▶ 3 global reduction phases

Semi-local communication

- ▶ 2 non-overlapping SpMVs

Local communication

- ▶ 4 axpy(-like) operations

General two-step framework for deriving pipelined Krylov methods:

*Step 1. **Avoiding communication:** merge global reductions*

*Step 2. **Hiding communication:** overlap SpMVs & global reductions*

Pipelined BiCGStab (p-BiCGStab)

Algorithm 6 Pipelined BiCGStab

```
1: function PIPE-BICGSTAB( $A, b, x_0$ )
2:    $r_0 := b - Ax_0$ ;  $w_0 := Ar_0$ ;  $t_0 := Aw_0$ ;
3:   for  $i = 0, \dots$  do
4:      $p_i := r_i + \beta_{i-1} (p_{i-1} - \omega_{i-1} s_{i-1})$ 
5:      $s_i := w_i + \beta_{i-1} (s_{i-1} - \omega_{i-1} z_{i-1})$ 
6:      $z_i := t_i + \beta_{i-1} (z_{i-1} - \omega_{i-1} v_{i-1})$ 
7:      $q_i := r_i - \alpha_i s_i$ 
8:      $y_i := w_i - \alpha_i z_i$ 
9:     compute  $(q_i, y_i)$  ;  $(y_i, y_i)$ 
10:     $\omega_i := (q_i, y_i) / (y_i, y_i)$ 
11:    overlap  $v_i := Az_i$ 
12:     $x_{i+1} := x_i + \alpha_i p_i + \omega_i q_i$ 
13:     $r_{i+1} := q_i - \omega_i y_i$ 
14:     $w_{i+1} := y_i - \omega_i (t_i - \alpha_i v_i)$ 
15:    compute  $(r_0, r_{i+1})$  ;  $(r_0, w_{i+1})$  ;  $(r_0, s_i)$  ;  $(r_0, z_i)$ 
16:     $\beta_i := (\alpha_i / \omega_i) (r_0, r_{i+1}) / (r_0, r_i)$ 
17:     $\alpha_{i+1} := (r_0, r_{i+1}) / ((r_0, w_{i+1}) + \beta_i (r_0, s_i) - \beta_i \omega_i (r_0, z_i))$ 
18:    overlap  $t_{i+1} := Aw_{i+1}$ 
19:  end for
20: end function
```

dot-prod
SpMV
axpy

p-BiCGStab:

(non-preconditioned)

Global communication

- ▶ 2 global red. phases (vs. 3)

Semi-local communication

- ▶ 2 overlapping SpMVs

Local communication

- ▶ 8 axpy(-like) operations (vs. 4)

Status after Step 2: both global comm. phases are overlapped with SpMV computations ('hidden'), at the cost of 4 additional axpys

Pipelined BiCGStab vs. Improved BiCGStab

 L.T. Yang and R.P. Brent. The improved BiCGStab method for large and sparse unsymmetric linear systems on parallel distributed memory architectures. In Proceedings of the Fifth International Conference on Algorithms and Architectures for Parallel Processing, pp. 324–328, IEEE, 2002.

	GLRED	SPMV	Flops (AXPY + DOT-PROD)	Time (GLRED + SPMV)	Memory
BiCGStab	3	2	20	3 GLRED + 2 SPMV	7
IBiCGStab	1	2	30	1 GLRED + 2 SPMV	10
p-BiCGstab	2	2*	38	2 max(GLRED, SPMV)	11

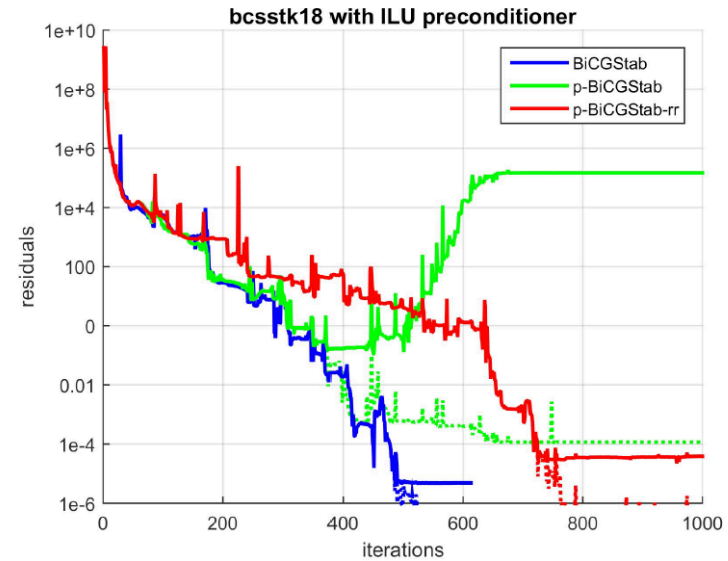
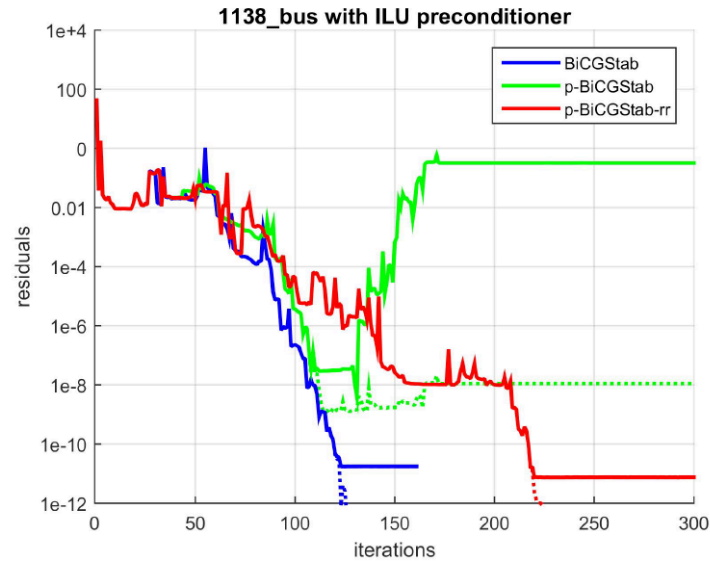
If $\text{time}(\text{GLRED}) \approx \text{time}(\text{SPMV})$:

- ▶ speed-up factor $\text{p-BiCGStab}/\text{BiCGStab} = 2.5$
- ▶ speed-up factor $\text{IBiCGStab}/\text{BiCGStab} = 1.66$

If $\text{time}(\text{GLRED}) \gg \text{time}(\text{SPMV})$:

- ▶ speed-up factor $\text{p-BiCGStab}/\text{BiCGStab} = 2.5$
- ▶ speed-up factor $\text{IBiCGStab}/\text{BiCGStab} = 3.0$

Robustness and attainable accuracy: p-BiCGStab-rr



Residual replacement every rr-th iteration

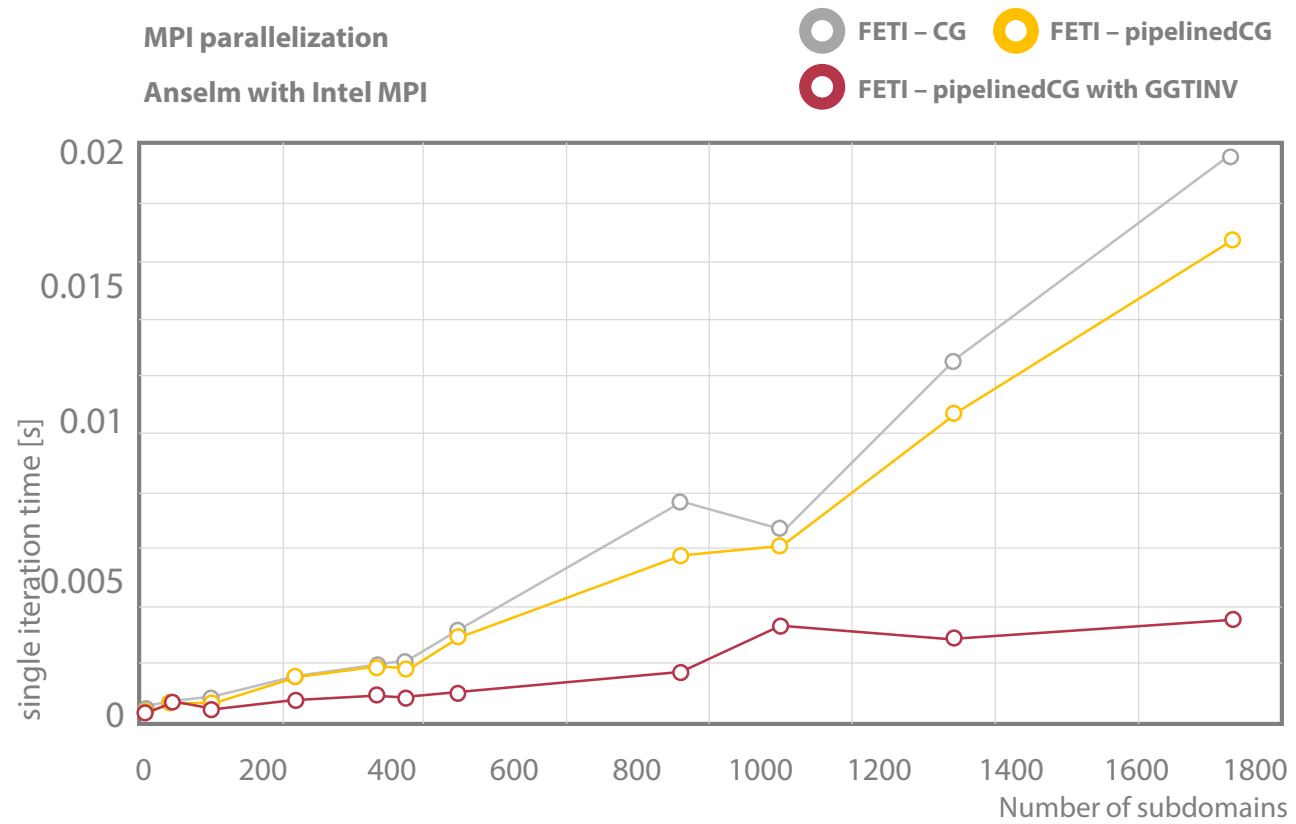
(non-automated, i.e. rr is a parameter of the method, but chosen large s.t. no. res. repl. is small)

$$\begin{aligned} r_i &:= b - Ax_i, & \hat{r}_i &:= M^{-1} r_i, & w_i &:= A\hat{r}_i, \\ s_i &:= A\hat{p}_i, & \hat{s}_i &:= M^{-1} s_i, & z_i &:= A\hat{s}_i. \end{aligned}$$

- ⊕ increased **maximal attainable accuracy**: comparable to BiCGStab level
- ⊕ increased **robustness**: negates instable true residual behaviour
- ⊖ increased **number of iterations** possible

ESPRESO FETI

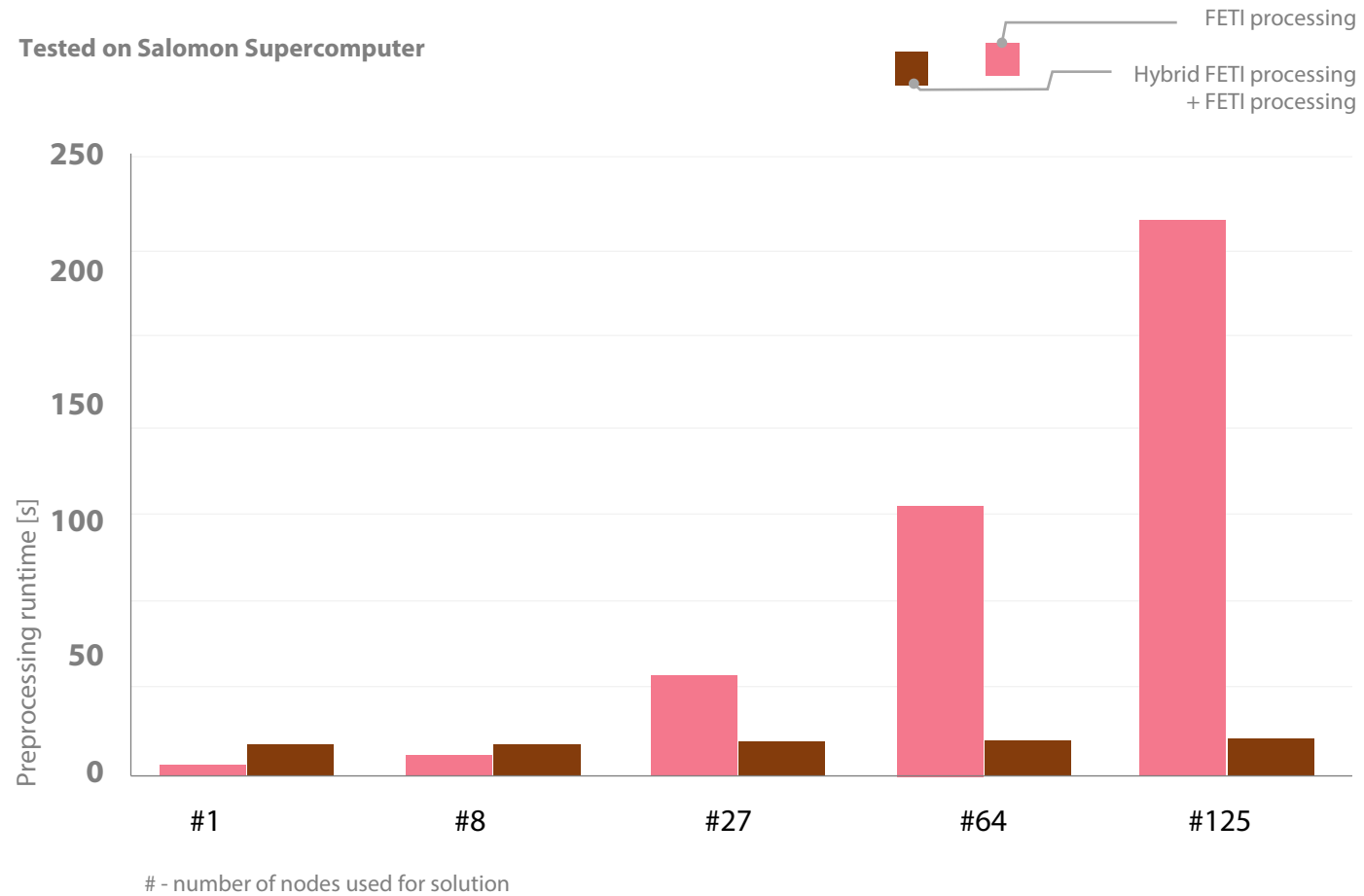
Optimization of global communication for FETI Solvers



ESPRESO FETI

Hybrid FETI vs. FETI preprocessing

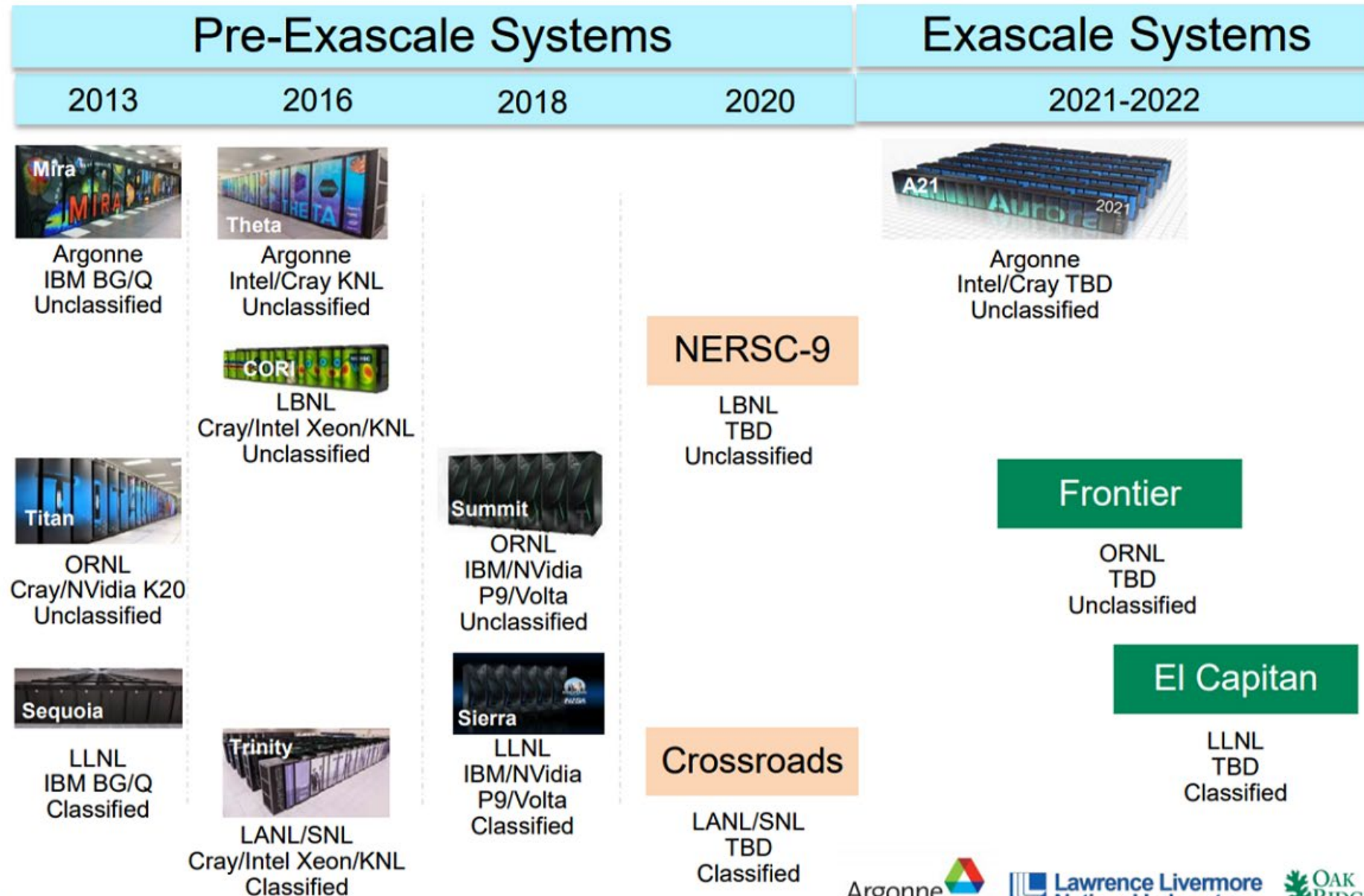
Tested on Salomon Supercomputer





Accelerators

DOE HPC Facilities Systems



ESPRESO CG in FETI

Projected Conjugate Gradient in FETI

```
1:  $r_0 := b - Ax_0; u_0 := M^{-1}r_0; p_0 := u_0$ 
2: for  $i = 0, \dots, m - 1$  do
3:    $s := Ap_i$ 
4:    $\alpha := \langle r_i, u_i \rangle / \langle s, p_i \rangle$ 
5:    $x_{i+1} := x_i + \alpha p_i$ 
6:    $r_{i+1} := r_i - \alpha s$ 
7:    $u_{i+1} := M^{-1}r_{i+1}$ 
8:    $\beta := \langle r_{i+1}, u_{i+1} \rangle / \langle r_i, u_i \rangle$ 
9:    $p_{i+1} := u_{i+1} + \beta p_i$ 
10: end for
```

Pre-processing - $S_c = B_1 K^{-1} B_1^T \rightarrow$ GPU/MIC

1.) $\lambda \rightarrow$ GPU/MIC - **PCIe transfer from CPU**

2.) $\lambda = S_c \cdot \lambda$ - **DGEMV, DSYMV on GPU/MIC**

3.) $\lambda \leftarrow$ GPU/MIC - **PCIe transfer to CPU**

4.) stencil data exchange in λ

- MPI – Send and Recv

- OpenMP – shared mem. vec

Pre-processing – K factorization

1.) $x = B_1^T \cdot \lambda$ - **SpMV**

2.) $y = K^{-1} \cdot x$ - **solve**

3.) $\lambda = B_1 \cdot y$ - **SpMV**

4.) stencil data exchange in λ

- MPI – Send and Recv

- OpenMP – shared mem. Vec

Pre-processing - $S_c = B_1 K^{-1} B_1^T$

1.) - **nop**

2.) $\lambda = S_c \cdot \lambda$ - **DGEMV, DSYMV**

3.) - **nop**

4.) stencil data exchange in λ

- MPI – Send and Recv

- OpenMP – shared mem. Vec

90 – 95% of runtime spent in Ap_i

GPU acceleration of the ESPRESO Solver

0.3 - 300 million DOF Hybrid FETI CG Solver Runtime

Linear elasticity

ORNL Titan 5rd in TOP500 LIST

speedUp

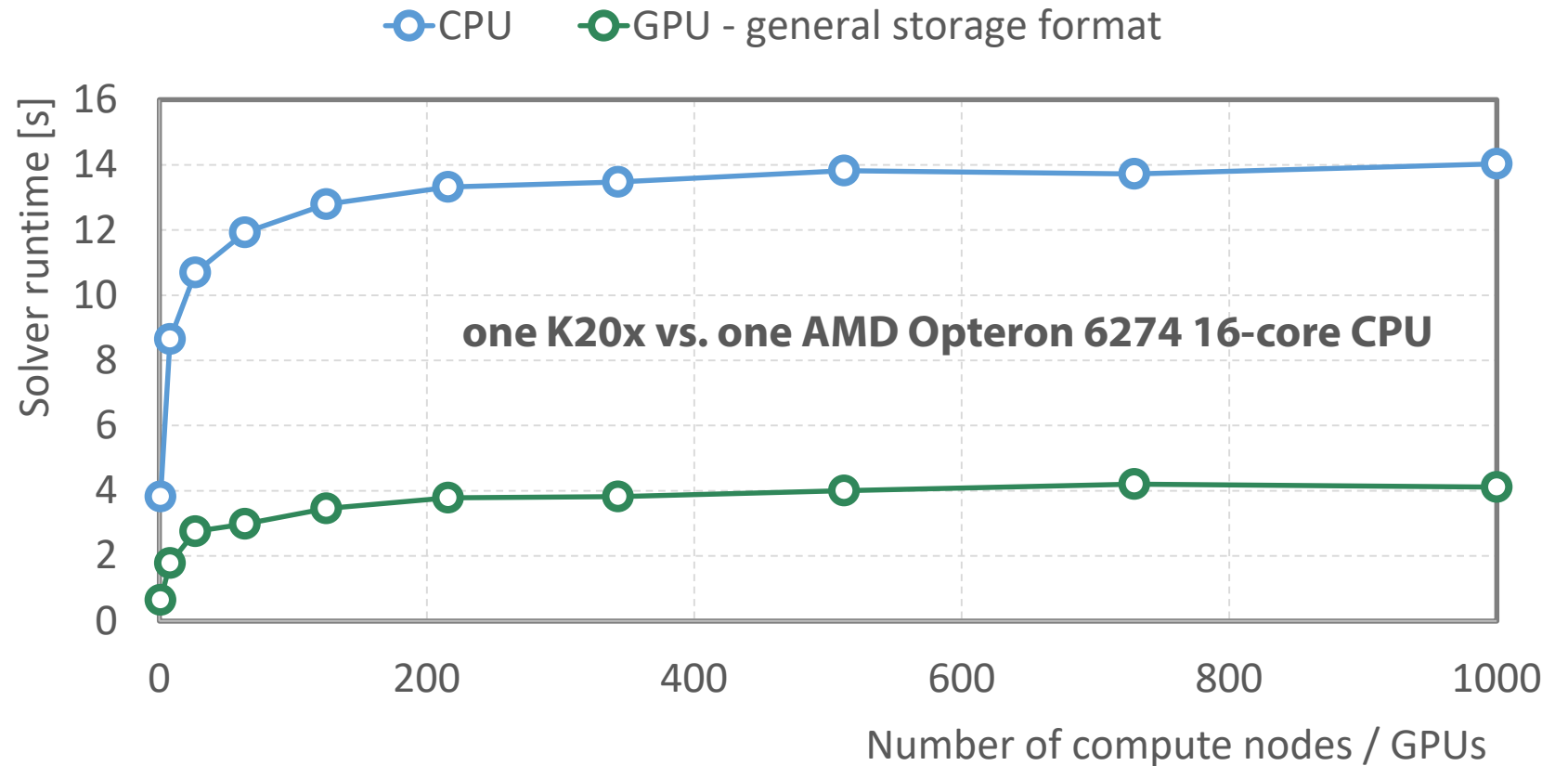
3.4

•Current work:

- support for cuDense and cuSparse solvers from CUDA toolkit
- memory optimization for symmetric local schur complements
- implementation of memory efficient methods

•Future work:

- Support for Power8/9 platforms with Pascal/Volta GPUs with NVLink

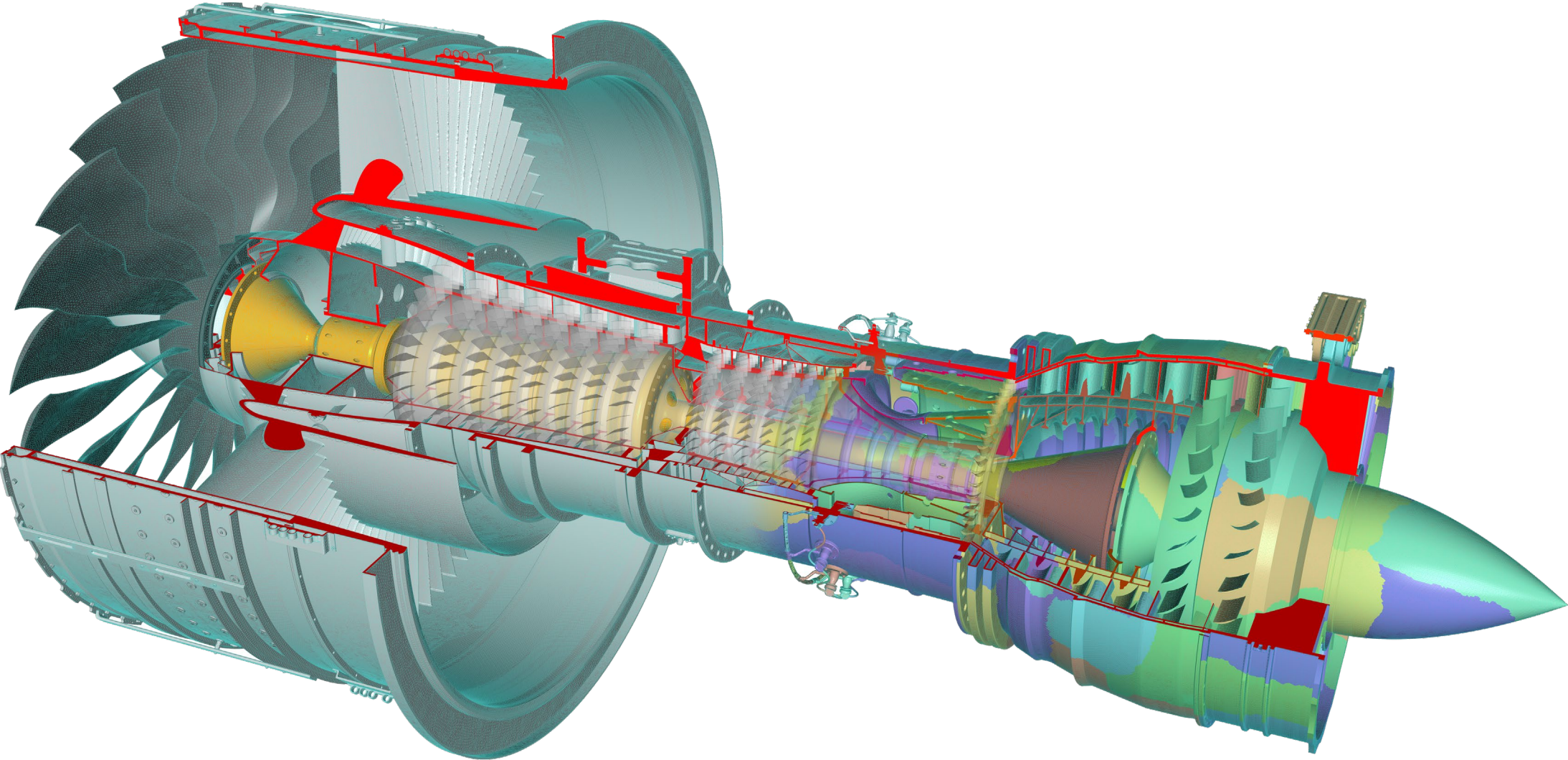


A blue circle with a white number '4' inside, casting a soft shadow to the right and bottom.

4

Parallel pre-
processing

ESPRESSO FEM – pre-processing



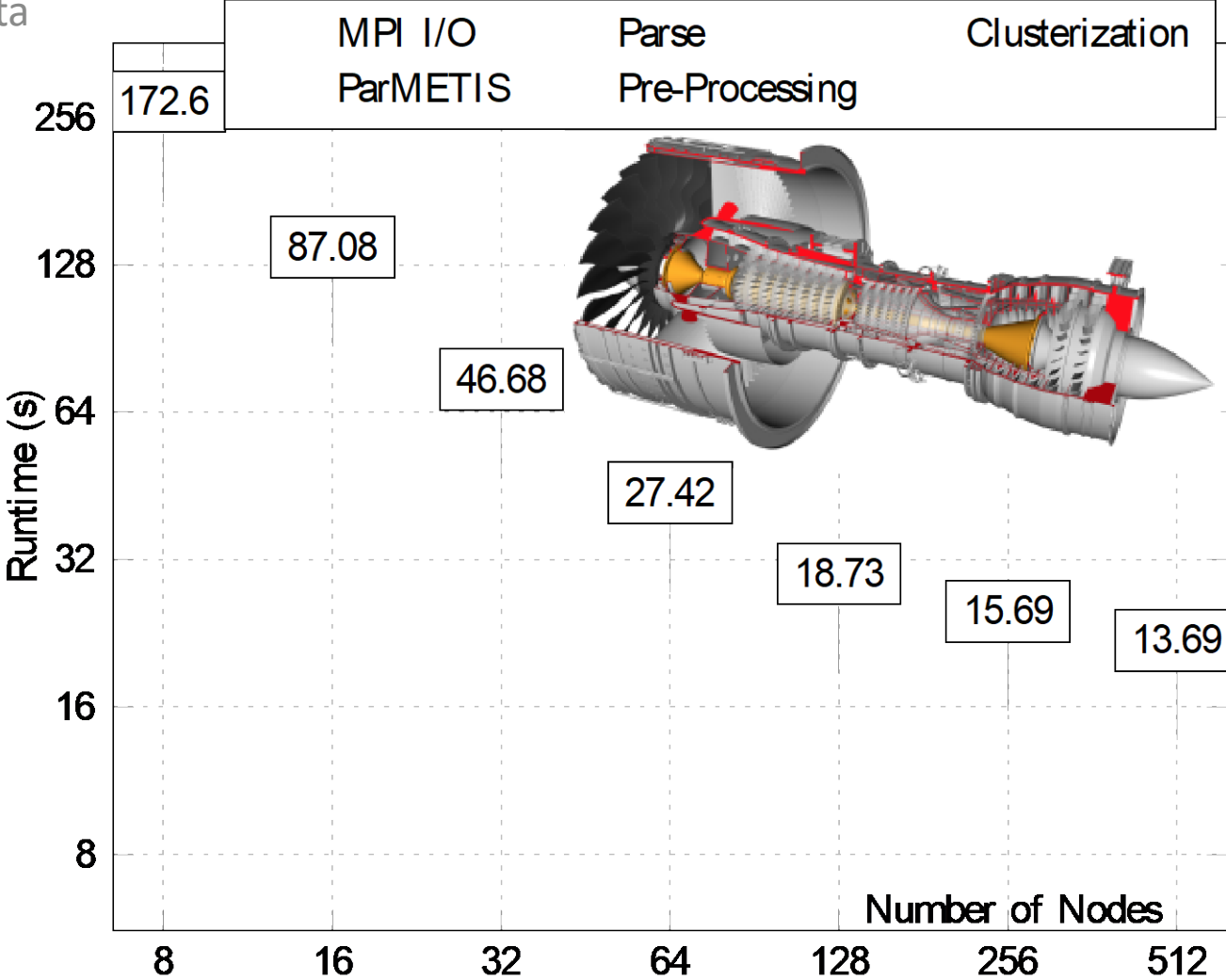
Efficient parallel loading of seq. stored unstructured mesh databases

Direct connection with commercial pre-processing tools

Highly parallel I/O of external sequentially stored data with domain decomposition techniques



Jet Engine	
file size	142 GB
Nodes	822 M
Elements	484 M
seq. processing	7200s
par. Processing	13.69 s

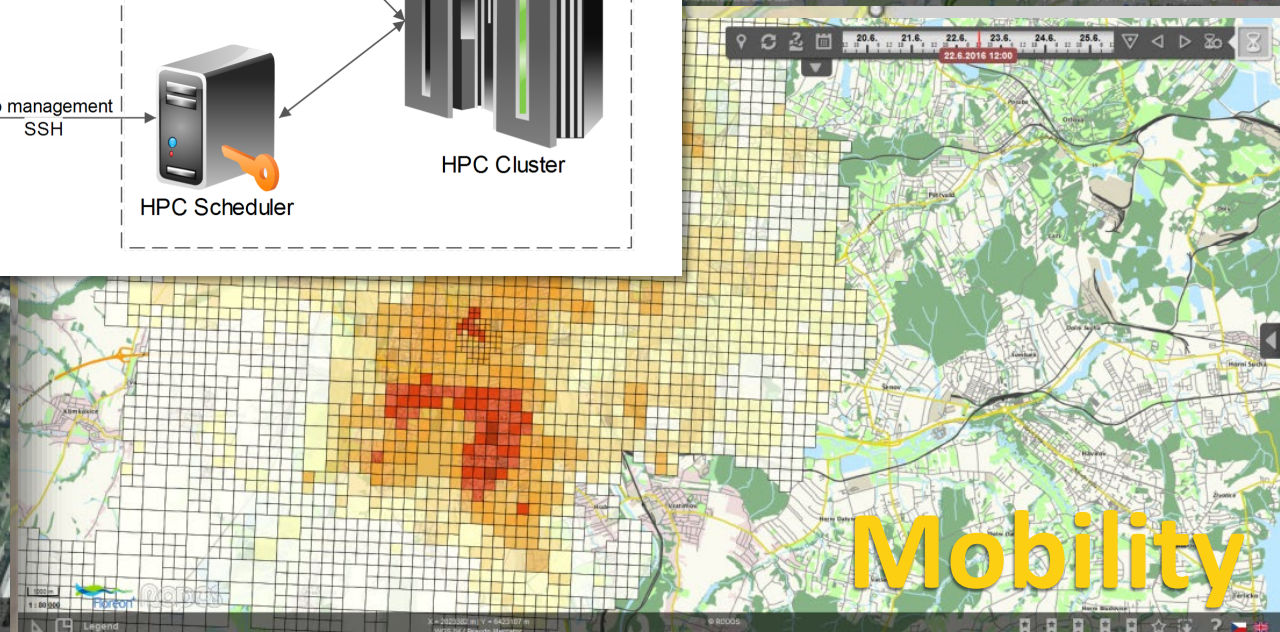
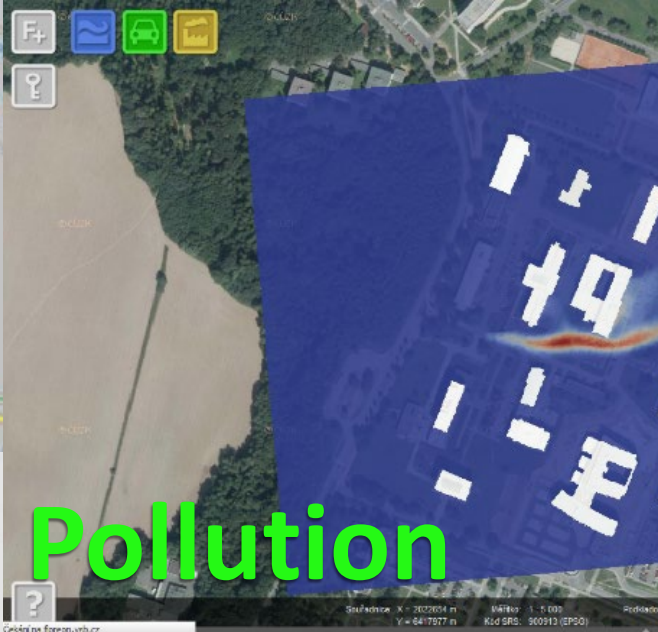
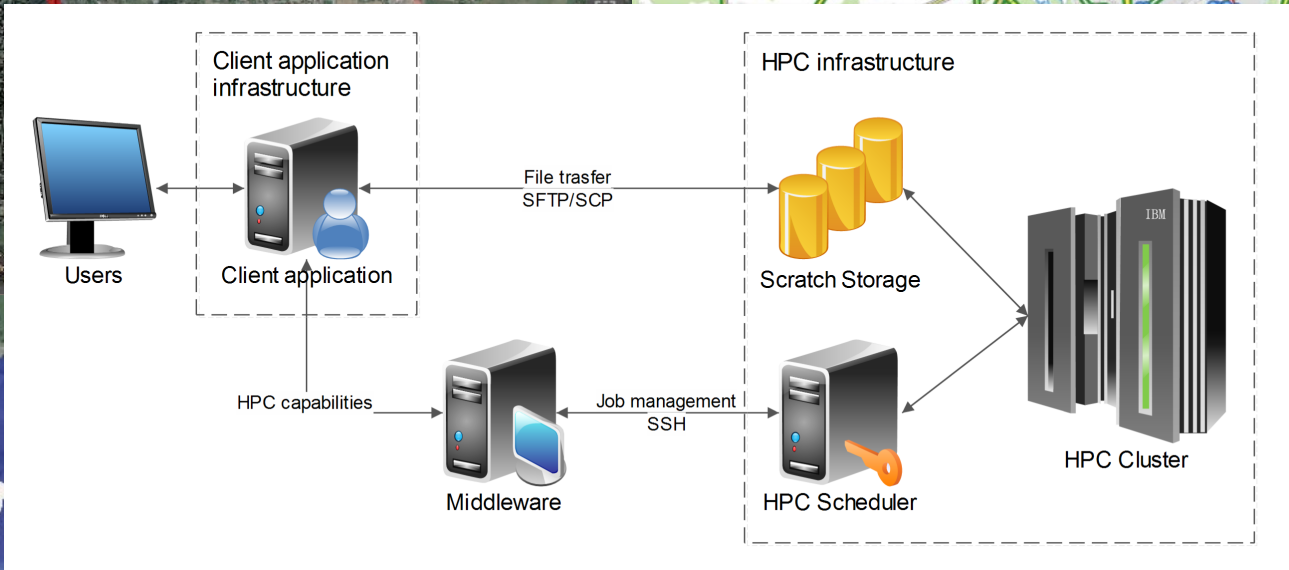
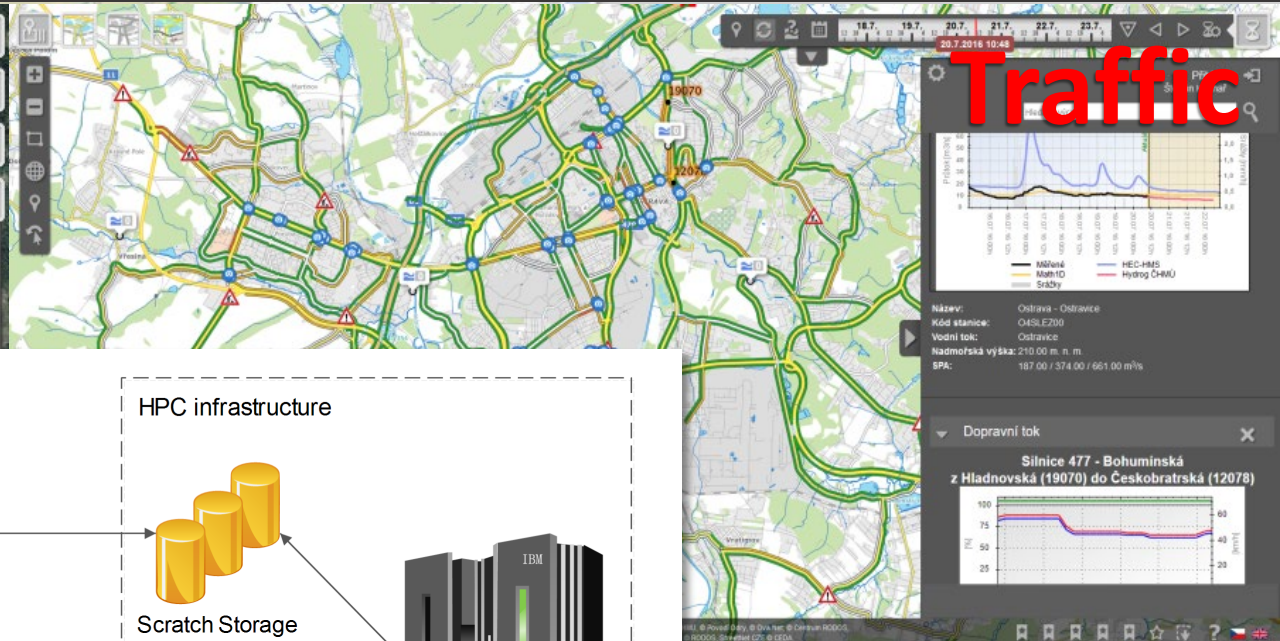
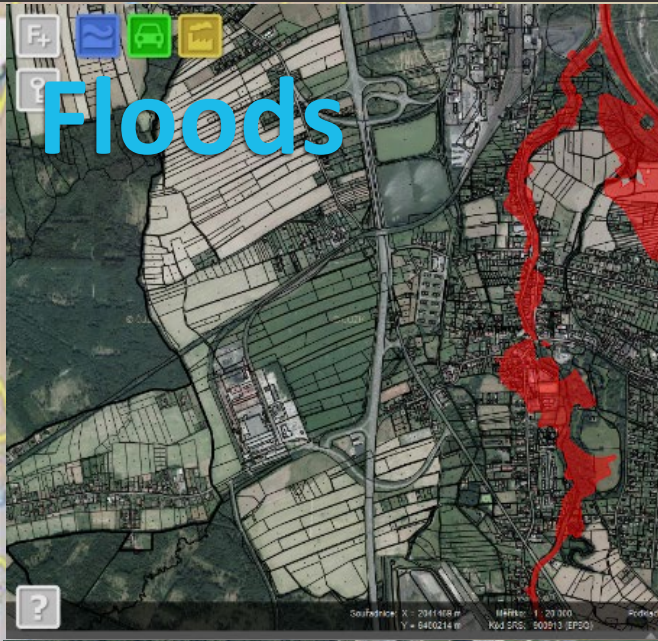




5

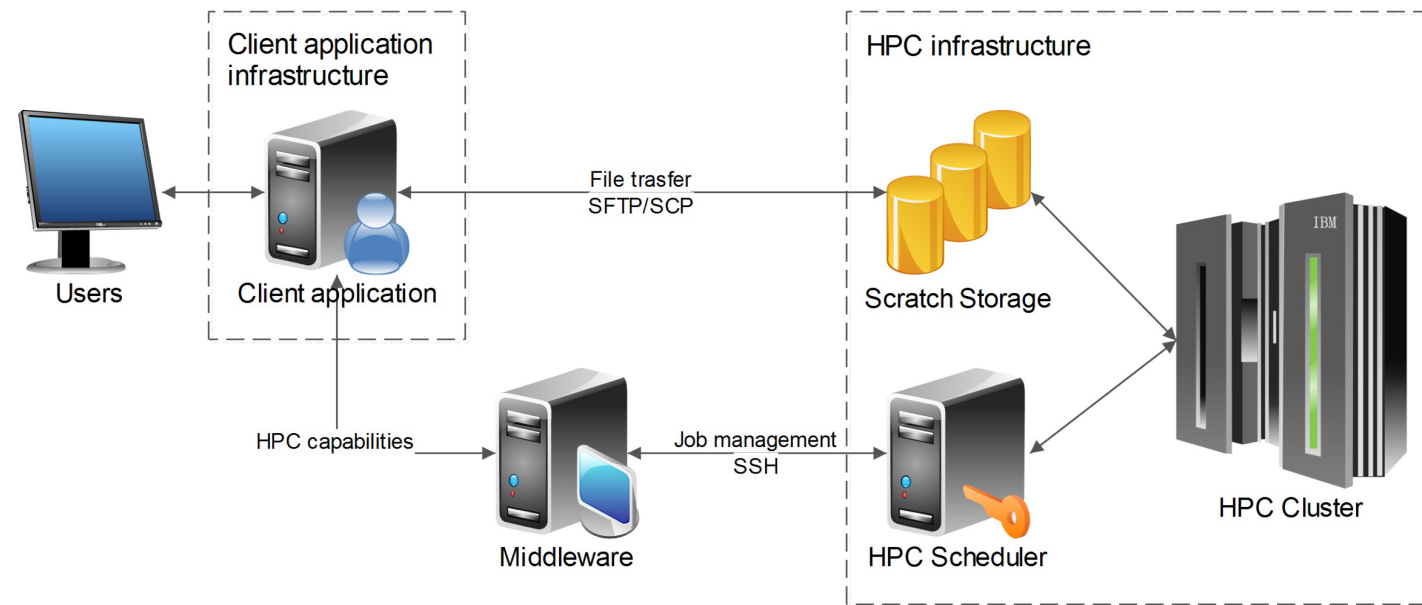
GUI and Solver as a Service

Online Solver as a Service (HPC as a Service)



HPC as a Service (HEAppE) Overview

- Providing HPC capabilities as a service to client applications and their users
- Unified interface for different operating systems and schedulers
- Authentication and authorization to provided functions
- Monitoring and reporting of executed jobs and their progress
- Current information about the state of the cluster



Thank you for attention