What is the Largest Finite Element System That Can be Solved Today?

U. Rüde (FAU Erlangen, ulrich.ruede@fau.de)
joint work with B. Gmeiner, M. Huber, H. Stengel (FAU)
C. Waluga, M. Huber, L. John, B. Wohlmuth (TUM)
M. Mohr, S. Baumann, J. Weismüller, P. Bunge (LMU)

Lehrstuhl für Simulation
FAU Erlangen-Nürnberg

Ferienakademie 2015
Sarntal
Overview

- Motivation
- Optimal is not always fast
- Fast multigrid solvers
- Earth mantle convection
- Textbook efficiency
The Two Principles of Science

Theory
mathematical models, differential equations, Newton

Experiments
observation and prototypes, empirical sciences

Three Computational methods open the path to Predictive Science

Computational Science
simulation, optimization
(quantitative) virtual reality
Predictive Science: the scientific discipline concerned with assessing the predictability of mathematical and computational models of reality. It embraces the processes of model selection, calibration, validation, verification, and their use in forecasting the relevant features of reality with quantified uncertainty.

The quantities of interest (QI): the goals of the simulation

UQ: quantitative analysis of the uncertainties in the predicted QI

Predictability requires knowledge of the physical laws that are proposed to explain realities and requires recognizing and quantifying uncertainties.
Two Multi-PetaFlops Supercomputers

**JUQUEEN**
- Blue Gene/Q architecture
- 458,752 PowerPC A2 cores
- 16 cores (1.6 GHz) per node
- 16 GiB RAM per node
- 5D torus interconnect
- 5.8 PFlops Peak
- TOP 500: #8

**SuperMUC**
- Intel Xeon architecture
- 147,456 cores
- 16 cores (2.7 GHz) per node
- 32 GiB RAM per node
- Pruned tree interconnect
- 3.2 PFlops Peak
- TOP 500: #14

What is the problem?
Designing Algorithms!
with four strong jet engines

Would you want to propel a Superjumbo

Large Scale Simulation Software

or with 1,000,000 blow dryer fans?

Moderately Parallel Computing

Massively Parallel MultiCore Systems
Mega=$10^6$, Giga=$10^9$, Tera=$10^{12}$, Peta=$10^{15}$, Exa=$10^{18}$

- World has a population of $7 \times 10^9$ humans
- Earth is $4.6 \times 10^9$ years old
  - the oceans together have ca. $1.3 \times 10^9$ km$^3$
  - the mantle has $0.91 \times 10^{12}$ km$^3$
- $10^{12}$ finite elements can resolve the volume of the mantle with ca. 1 km resolution
- Number of stars in the galaxy: $10^{11}$
- Avogadro’s constant: $6 \times 10^{23}$ mol$^{-1}$
- The recirculatory system contains $2.5 \times 10^{13}$ red blood cells
- The *brain* has ca. $10^{11}$ Neurons
- Processor chip has
  - $5 \times 10^9$ transistors
  - 3 GHz clock rate: $3 \times 10^9$
  - can perform $10^{11}$ (= 100 Giga) Flops
  - Supercomputer at 2020: $10^{18}$ Flops

An exa-scale system with $10^{18}$ Flops should suffice to resolve the meso-scale
Application to Earth Mantle Convection Models
Mantle Convection

Why Mantle Convection?
• driving force for plate tectonics
• mountain building and earthquakes

Why Exascale?
• mantle has $10^{12}$ km$^3$
• inversion and UQ blow up cost

Why TerraNeo?
• ultra-scalable and fast
• sustainable framework

Challenges
• computer sciences: software design for future exascale systems
• mathematics: HPC performance oriented metrics
• geophysics: model complexity and uncertainty
• bridging disciplines: integrated co-design
HHG Solver for Stokes System
Motivated by Earth Mantle convection problem


\[-\nabla \cdot \left( 2\eta \varepsilon(u) \right) + \nabla p = \rho(T)g, \]
\[\nabla \cdot u = 0, \]
\[\frac{\partial T}{\partial t} + u \cdot \nabla T - \nabla \cdot (\kappa \nabla T) = \gamma. \]

\[\begin{align*}
\textbf{u} & \quad \text{velocity} \\
\rho & \quad \text{dynamic pressure} \\
T & \quad \text{temperature} \\
\nu & \quad \text{viscosity of the material} \\
\varepsilon(u) = \frac{1}{2} (\nabla u + (\nabla u)^T) & \quad \text{strain rate tensor} \\
\rho & \quad \text{density} \\
\kappa, \gamma, g & \quad \text{thermal conductivity, heat sources, gravity vector}
\end{align*}\]

Scale up to \( \sim 10^{12} \) nodes/ DOFs
\[\Rightarrow \text{resolve the whole Earth Mantle globally with 1km resolution} \]

Stokes equation:
\[-\text{div}(\nabla u - pI) = f, \quad \text{div} u = 0\]

FEM Discretization:
\[a(u_l, v_l) + b(v_l, p_l) = L(v_l) \quad \forall v_l \in V_l,\]
\[b(u_l, q_l) - c(p_l, q_l) = 0 \quad \forall q_l \in Q_l,\]

with:
\[a(u, v) := \int_\Omega \nabla u : \nabla v \, dx, \quad b(u, q) := -\int_\Omega \text{div} u \cdot q \, dx\]

Schur-complement formulation:
\[
\begin{bmatrix}
A_l & B_l^T \\
0 & \mathbf{C}_l + B_lA_l^{-1}B_l^T
\end{bmatrix}
\begin{bmatrix}
u_l \\
p_l
\end{bmatrix}
= 
\begin{bmatrix}
f_l \\
B_lA_l^{-1}f_l
\end{bmatrix}
\]
# The Curse of Size

<table>
<thead>
<tr>
<th></th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>computer generation</strong></td>
<td><strong>gigascale: 10^9 FLOPS</strong></td>
</tr>
<tr>
<td><strong>terascale</strong></td>
<td><strong>10^{12} FLOPS</strong></td>
</tr>
<tr>
<td><strong>petascale</strong></td>
<td><strong>10^{15} FLOPS</strong></td>
</tr>
<tr>
<td><strong>exascale</strong></td>
<td><strong>10^{18} FLOPS</strong></td>
</tr>
<tr>
<td><strong>desired problem size</strong></td>
<td><strong>DoF=N</strong></td>
</tr>
<tr>
<td><strong>10^6</strong></td>
<td><strong>10^9</strong></td>
</tr>
<tr>
<td><strong>10^{12}</strong></td>
<td><strong>10^{15}</strong></td>
</tr>
<tr>
<td><strong>energy estimate (kWh)</strong></td>
<td><strong>1 NJoule x N^2</strong></td>
</tr>
<tr>
<td><strong>0.278 Wh</strong></td>
<td><strong>278 kWh</strong></td>
</tr>
<tr>
<td><strong>10 min of LED light</strong></td>
<td><strong>278 GWh</strong></td>
</tr>
<tr>
<td><strong>2 weeks</strong></td>
<td><strong>2 months electricity for Munich</strong></td>
</tr>
<tr>
<td><strong>blow drying hair</strong></td>
<td><strong>278 PWh</strong></td>
</tr>
<tr>
<td><strong>electricity for Munich</strong></td>
<td><strong>100 years world electricity production</strong></td>
</tr>
<tr>
<td><strong>all-to-all communication</strong></td>
<td></td>
</tr>
<tr>
<td><strong>TerraNeo prototype (kWh)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>0.13 Wh</strong></td>
<td><strong>0.03 kWh</strong></td>
</tr>
<tr>
<td><strong>27 kWh</strong></td>
<td><strong>?</strong></td>
</tr>
</tbody>
</table>

At extreme scale: optimal complexity is a must!
Optimal ⇔ Fast

UR, New Mathematics for Extreme-scale Computational Science?
SIAM News, Volume 48, Number 5, June 2015

- Accuracy ⇔ Computational Cost
- Higher order ⇔ Complexity of discrete system
- Mesh:
  - Unstructured ⇔ Structured
  - Adaptive refinement ⇔ Superconvergence
  - Efficiency on modern computer architectures
- Algebraic Solver:
  - Operation count ⇔ Parallel scalability
- Cost metric:
  - time to solution ⇔ energy consumption
- Asymptotic estimates ⇔ quantified constants
- Worst case ⇔ Average case
- Bitwise reproducibility ⇔ Algorithmic Fault Tolerance

What's the largest FE system solved? - Uli Rüde
Multigrid: Algorithms for $10^{12}$ unknowns

Goal: solve $A^h u^h = f^h$ using a hierarchy of grids

- Relax on $A^h u^h = f^h$
- Residual $r^h = f^h - A^h u^h$
- Restrict $r^H = I^H_h r^h$
- Correct $u^h \leftarrow u^h + e^h$
- Interpolate $e^h = I^h_H e^H$

Multigrid uses coarse grids to accomplish the inevitable global data exchange in the most efficient way possible

What’s the largest FE system solved?  - Uli Rüde
What's the largest FE system solved? - Uli Rüde

HHG Data Structures

Geometrical Hierarchy: Volume, Face, Edge, Vertex
**Why Co-Design?**

No exascale simulation without **maximal efficiency on all levels**
- model formulation - discretisation - solution algorithm
- core - node - cluster

<table>
<thead>
<tr>
<th></th>
<th><strong>TERRA</strong>prototype**</th>
<th><strong>JUQUEEN</strong> Rank 8, TOP 500, Nov 2014</th>
<th><strong>SuperMUC</strong> Rank 14, TOP 500, Nov 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOF</strong></td>
<td>8.4 ( \cdot 10^7 )</td>
<td>32 km</td>
<td>6.4 ( \cdot 10^8 ) 16 km</td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>1</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td><strong>Threads</strong></td>
<td>4</td>
<td>240</td>
<td>2</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>30 s</td>
<td>38 s</td>
<td>16 s</td>
</tr>
<tr>
<td><strong>DOF</strong></td>
<td>5.2 ( \cdot 10^9 )</td>
<td>8 km</td>
<td>4.4 ( \cdot 10^{10} ) 4 km</td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>30</td>
<td>1920</td>
<td>120</td>
</tr>
<tr>
<td><strong>Threads</strong></td>
<td>4</td>
<td>15 260</td>
<td>1 920</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>40 s</td>
<td>44 s</td>
<td>27 s</td>
</tr>
<tr>
<td><strong>DOF</strong></td>
<td>4.4 ( \cdot 10^{10} )</td>
<td>4 km</td>
<td>3.4 ( \cdot 10^{11} ) 2 km</td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>1 920</td>
<td>122 880</td>
<td>960</td>
</tr>
<tr>
<td><strong>Threads</strong></td>
<td>6</td>
<td>15 360</td>
<td>15 360</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>48 s</td>
<td>44 s</td>
<td>34 s</td>
</tr>
<tr>
<td><strong>DOF</strong></td>
<td>3.4 ( \cdot 10^{11} )</td>
<td>2 km</td>
<td>2.8 ( \cdot 10^{12} ) 1 km</td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>15 360</td>
<td>983 040</td>
<td>7 680</td>
</tr>
<tr>
<td><strong>Threads</strong></td>
<td>6</td>
<td>122 880</td>
<td>122 880</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>54 s</td>
<td>44 s</td>
<td>41 s</td>
</tr>
</tbody>
</table>

**HiQ**

**High-Q Club**
Highest Scaling Codes on JuQUEEN
Coupled Flow-Transport Problem

\[-\Delta \mathbf{u} + \nabla p = - \text{Ra} T \hat{\mathbf{r}}\]

\[\text{div} \, \mathbf{u} = 0\]

\[\partial_t T + \mathbf{u} \cdot \nabla T = \text{Pe}^{-1} \Delta T\]

- 6.5×10^9 DoF
- 10000 time steps
- run time 7 days
- Mid-size cluster: 288 compute cores in 9 nodes of LSS at FAU

For the sake of presentation, let us simplify the model by neglecting compressibility, nonlinearities and inhomogeneities in the material. In dimensionless form, we obtain:

\[-\Delta \mathbf{u} + \nabla p = - \text{Ra} T \hat{\mathbf{r}}\]

\[\text{div} \, \mathbf{u} = 0\]

\[\partial_t T + \mathbf{u} \cdot \nabla T = \text{Pe}^{-1} \Delta T\]
Conclusions and Outlook

- $2.8 \times 10^{12}$ unknowns in fully implicit solve with Peta-Scale machines demonstrated
- But optimal algorithms are not necessarily fast
- Multigrid scales to Peta and beyond
- HHG: lean and mean implementation: excellent time to sol.

We can now do $10^{13}$ DoF!