Panel session 6
Measuring Performance in Geoscience Apps

Prof. Dr. Thomas Ludwig
German Climate Computing Center (DKRZ)
University of Hamburg, Department for Computer Science (UHH/FBI)

Dr. Claudia Frauen
German Climate Computing Center (DKRZ)
## TOP500 List

<table>
<thead>
<tr>
<th>Rank</th>
<th>System</th>
<th>Cores</th>
<th>Rmax (PFlop/s)</th>
<th>Rpeak (PFlop/s)</th>
<th>Power (kW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Frontier</strong> - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States</td>
<td>8,730,112</td>
<td>1,102.00</td>
<td>1,685.65</td>
<td>21,100</td>
</tr>
<tr>
<td>2</td>
<td><strong>Supercomputer Fugaku</strong> - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, <strong>Fujitsu</strong> RIKEN Center for Computational Science Japan</td>
<td>7,630,848</td>
<td>442.01</td>
<td>537.21</td>
<td>29,899</td>
</tr>
<tr>
<td>3</td>
<td><strong>LUMI</strong> - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, <strong>HPE EuroHPC/CSC</strong> Finland</td>
<td>1,110,144</td>
<td>151.90</td>
<td>214.35</td>
<td>2,942</td>
</tr>
<tr>
<td>4</td>
<td><strong>Summit</strong> - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, <strong>IBM</strong> DOE/SC/Oak Ridge National Laboratory United States</td>
<td>2,414,592</td>
<td>148.60</td>
<td>200.79</td>
<td>10,096</td>
</tr>
</tbody>
</table>
TOP500 List History

Performance Benchmarks:

- High Performance Linpack
  - standard, problematic
  - used for TOP500
- High-Performance Conjugate Gradient
  - alternative benchmark
- Green500
  - evaluates energy efficiency
TOP500 DKRZ List History

Increase in LINPACK performance within the TOP500 and at DKRZ

- DKRZ System
- Rank 1
- Rank 500

Years:
- 1993
- 1994
- 1995
- 1996
- 1997
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015
- 2016
- 2017
- 2018
- 2019
- 2020

Ranks:
- Rank 164 (11/21)
- Rank 54
- Rank 33 (Phase 1)
- Rank 56 (Phase 2)
- HLRE-3

Performance Levels:
- 10,000,000
- 100,000
- 10
- 1
- 0.1

Legend:
- Red line: Rank 1
- Blue line: Rank 500
- Dashed black line: DKRZ System

Timeline:
- 4 years
- 5 years

© Thomas Ludwig, Claudia Frauen
ICES 2022
2022-09-30
HLRE-4 “Levante” – #76 in TOP500 June 2022
HLRE-4 “Levante”
Installation: 2021-2022

Producer: Atos

Model: Atos BullSequana XH2000

No. of cores: 370,000

Network: HDR-Infiniband, NVIDIA Mellanox InfiniBand HDR 100G/200G

Disk system: 130 Petabyte (Lustre) von DDN

HSM system: Cristie / StrongBox Data Solutions / Huawei
Levante Data Sheet 2

**CPU-Partition**
- 2,832 compute nodes
  - 2,520 nodes with 2 processors AMD 7763 (256 GB memory)
  - 294 nodes with 2 processors AMD 7763 (512 GB memory)
  - 18 nodes with 2 processors AMD 7763 (1024 GB memory)
- Peak performance: 14 PetaFLOPS
- Main memory: 815 TB

**GPU-Partition**
- 60 GPU nodes with
  - 2 processors AMD 7713 (512 GB memory)
  - 4 Nvidia A100 GPUs (56 nodes with 80 GB, 4 nodes with 40 GB local memory)
- Peak performance: 2.8 PetaFLOPS
- Main memory: 30 TB
## Systems at DKRZ

<table>
<thead>
<tr>
<th>year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/ MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Blizzard</td>
</tr>
<tr>
<td></td>
<td>+6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mistral</td>
</tr>
<tr>
<td></td>
<td>+7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Levante</td>
</tr>
</tbody>
</table>
# Systems at DKRZ

<table>
<thead>
<tr>
<th>year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Blizzard</td>
</tr>
<tr>
<td></td>
<td>+6</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mistral</td>
</tr>
<tr>
<td></td>
<td>+7</td>
<td>4.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Levante</td>
</tr>
</tbody>
</table>
# Systems at DKRZ

<table>
<thead>
<tr>
<th>Year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td>1.6</td>
<td>Blizzard</td>
</tr>
<tr>
<td></td>
<td>+6</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
<td>1.4</td>
<td>Mistral</td>
</tr>
<tr>
<td></td>
<td>+7</td>
<td>4.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td>2.2</td>
<td>Levante</td>
</tr>
</tbody>
</table>
## Systems at DKRZ

<table>
<thead>
<tr>
<th>year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td>+6</td>
<td></td>
<td></td>
<td>1.6</td>
<td>Blizzard</td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td>+7</td>
<td>26</td>
<td>0.9</td>
<td>1.4</td>
<td>Mistral</td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td>4.3</td>
<td>1.5</td>
<td>2.2</td>
<td>Levante</td>
</tr>
</tbody>
</table>
### Systems at DKRZ

<table>
<thead>
<tr>
<th>year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td></td>
<td>0.09</td>
<td></td>
<td>1.6</td>
<td>Blizzard</td>
</tr>
<tr>
<td></td>
<td>+6</td>
<td>26</td>
<td>29x</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td></td>
<td>2.8</td>
<td></td>
<td>1.4</td>
<td>Mistral</td>
</tr>
<tr>
<td></td>
<td>+7</td>
<td>4.3</td>
<td>2.8x</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td>7.7</td>
<td></td>
<td>2.2</td>
<td>Levante</td>
</tr>
</tbody>
</table>
## Energy Efficiency World Wide

<table>
<thead>
<tr>
<th>year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td></td>
<td>0.09</td>
<td></td>
<td>1.6</td>
<td>Blizzard</td>
</tr>
<tr>
<td></td>
<td>+6</td>
<td>26</td>
<td>29x</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td></td>
<td>2.8</td>
<td></td>
<td>1.4</td>
<td>Mistral</td>
</tr>
<tr>
<td></td>
<td>+7</td>
<td>4.3</td>
<td>2.8x</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td>7.7</td>
<td></td>
<td>2.2</td>
<td>Levante</td>
</tr>
</tbody>
</table>

June 2022

- #1 Frontiers (USA): 52 GFLOPS/MW with GPU accelerator
- #2 Fugaku (Japan): 14 GFLOPS/MW with ARM processors
## Electronic Waste? (80t per machine at DKRZ)

<table>
<thead>
<tr>
<th>year</th>
<th>PFLOPS</th>
<th>PFLOPS Factor</th>
<th>GFLOPS/MW</th>
<th>MW Factor</th>
<th>MW</th>
<th>System Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.15</td>
<td></td>
<td>0.09</td>
<td></td>
<td>1.6</td>
<td>Blizzard</td>
</tr>
<tr>
<td>+6</td>
<td></td>
<td>26</td>
<td>29x</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>3.9</td>
<td></td>
<td>2.8</td>
<td></td>
<td>1.4</td>
<td>Mistral</td>
</tr>
<tr>
<td>+7</td>
<td>4.3</td>
<td></td>
<td>2.8x</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>17</td>
<td></td>
<td>7.7</td>
<td></td>
<td>2.2</td>
<td>Levante</td>
</tr>
</tbody>
</table>

- scrap the old machine!
- continue to use the old machine?
Semiconductor Fabrication Plants Costs

![Graph showing the cost of semiconductor fabrication plants from 2000 to 2025. The y-axis represents the cost in billions of dollars, ranging from 1 to 100, and the x-axis represents the years from 2000 to 2025. The cost shows a general upward trend over the years.]
Summary for Hardware

- Improvements with transistors come to an end
  - Decreasing gains in energy efficiency
  - Decreasing gains in performance
- Production costs and complexity escalate
- No alternative in the next (many?) years
Example: The weather and climate model ICON

- ICON is a weather and climate model that can be used for very different use cases:
  - Numerical weather forecast
  - Earth system model for CMIP type simulations (lower resolution, long time scales)
  - Storm and ocean eddy resolving model (km-scale resolution, only possible for short time scales)

Jungclaus et al., 2022
Current ICON performance

ICON ESM (Resolution: A -> 158km, O -> 40km)
Mistral 120 nodes -> 120 SYPD
Current ICON performance

ICON ESM (Resolution: \(A \rightarrow 158\text{km}, O \rightarrow 40\text{km}\))
Mistral 120 nodes \(\rightarrow\) 120 SYPD

Coupled high res simulations

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Machine</th>
<th>Nodes</th>
<th>SDPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 km</td>
<td>Mistral</td>
<td>420</td>
<td>17</td>
</tr>
<tr>
<td>5 km</td>
<td>Levante</td>
<td>600</td>
<td>126</td>
</tr>
<tr>
<td>2.5 km</td>
<td>Levante</td>
<td>600</td>
<td>20</td>
</tr>
<tr>
<td>1.25 km</td>
<td>Levante</td>
<td>900</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Hohenegger et al., 2022,
D. Klocke pers. comm.
Current ICON performance

ICON ESM (Resolution: A -> 158km, O -> 40km)
Mistral 120 nodes -> 120 SYPD

Coupled high res simulations

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Machine</th>
<th>Nodes</th>
<th>SDPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 km</td>
<td>Mistral</td>
<td>420</td>
<td>17</td>
</tr>
<tr>
<td>5 km</td>
<td>Levante</td>
<td>600</td>
<td>126</td>
</tr>
<tr>
<td>2.5 km</td>
<td>Levante</td>
<td>600</td>
<td>20</td>
</tr>
<tr>
<td>1.25 km</td>
<td>Levante</td>
<td>900</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Adapted from Giorgetta et al., 2022
D. Klocke pers. comm.
ICON scalability

QUBICC (R2B9) scaling - 2 simulated hours

- Runtime [s]
- Nodes

- timing AMD Milan
- ideal AMD Milan
- timing Mistral
- ideal Mistral
- timing NVIDIA A100
- ideal NVIDIA A100
- timing NVIDIA own
- ideal NVIDIA own
ICON scalability

QUBICC (R2B4) scaling on various HPC systems - 1 simulated day

QUBICC (R2B7) scaling on various HPC systems - 1 simulated hour

QUBICC (R2B9) scaling - 2 simulated hours
Performance: Compute-bound vs memory-bound

- Performance limiting factors:
  - Compute-bound:
    - Compute-bound kernels spend most of their time doing calculations and are limited by the peak performance of the hardware (HPL benchmark is compute-bound)
  - Memory-bound:
    - Memory-bound kernels are limited by the bandwidth of the memory interface (HPCG benchmark is memory-bound)
Roofline analysis of ICON non-hydrostatic solver

ICON on GPUs

- ICON-A is the only code running at DKRZ that can run on GPUs
- ICON is memory-bound. The speedup on GPU vs CPU is mostly due to wider memory bandwidth
- Low resolution ICON setups, like used for ICON-ESM, don’t provide enough computational load for GPUs
- Need to improve the computational intensity in order to better exploit the hardware capabilities
New Project: WarmWorld - Goals

Assess the detailed trajectory of global warming and the quantitative implications of this warming for human and natural systems

- Coupled ICON running with an acceptable simulation quality on km scale > 0.5 SYPD by 2026
- ICON-C: A free and open source software implementation of the fully (land, ocean, atmosphere) coupled ICON to enable scalable development
- Integrated workflow to expose information of ICON alongside ECMWF’s IFS-based solutions and observational data