Reflecting on the Goal and Baseline of Exascale Computing

Thomas C. Schulthess
Tracking supercomputer performance over time?

Linpack benchmark solves: $Ax = b$
Tracking supercomputer performance over time?
Linpack benchmark solves: $Ax = b$
Tracking supercomputer performance over time?
Linpack benchmark solves: $Ax = b$

1,000-fold performance improvement per decade
Tracking supercomputer performance over time?
Linpack benchmark solves: $Ax = b$

1,000-fold performance improvement per decade
Tracking supercomputer performance over time?

Linpack benchmark solves: $Ax = b$

1,000-fold performance improvement per decade
Tracking supercomputer performance over time?
Linpack benchmark solves: $Ax = b$

1,000-fold performance improvement per decade

1st application at > 1 TFLOP/s sustained

1st application at > 1 PFLOP/s sustained

KKR-CPA (MST)
Tracking supercomputer performance over time?

Linpack benchmark solves: $Ax = b$

1,000-fold performance improvement per decade
Tracking supercomputer performance over time?
Linpack benchmark solves: $Ax = b$
Tracking supercomputer performance over time?

Linpack benchmark solves: $Ax = b$

1,000-fold performance improvement per decade
Tracking supercomputer performance over time?

Linpack benchmark solves: \(Ax = b\)

1,000-fold performance improvement per decade seems to hold for multiple-scattering-theory (MST)-based electronic structure for materials science.

- KKR-CPA (MST)
- LSMS (MST)
- WL-LSMS (MST)

1st application at > 1 TFLOP/s sustained
1st application at > 1 PFLOP/s sustained
1,000x perf. improv. per decade
“Only” 100-fold performance improvement in climate codes

Source: Peter Bauer, ECMWF
Has the efficiency of weather & climate codes dropped 10-fold every decade?
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Floating points efficiency dropped from 50% on Cray Y-MP to 5% on today’s Cray XC (10x in 2 decades)

Source: Peter Bauer, ECMWF
Computational power drives spatial resolution

Moore's law with a doubling time of 18 months
Moore's law with a doubling time of 24 months

Source: Christoph Schär, ETH Zurich, & Nils Wedi, ECMWF
Computational power drives spatial resolution

Source: Christoph Schär, ETH Zurich, & Nils Wedi, ECMWF
Can the delivery of a 1km-scale capability be pulled in by a decade?

Source: Christoph Schär, ETH Zurich, & Nils Wedi, ECMWF
Leadership in weather and climate

Peter Bauer, ECMWF
Leadership in weather and climate

European model may be the best - but far away from sufficient accuracy and reliability!

Peter Bauer, ECMWF
The impact of resolution: simulated tropical cyclones

HADGEM3 PRACE UPSCALE, P.L. Vidale (NCAS) and M. Roberts (MO/HC)
Resolving convective clouds (convergence?)

**Bulk convergence**

Area-averaged bulk effects upon ambient flow:
E.g., heating and moistening of cloud layer

**Structural convergence**

Statistics of cloud ensemble:
E.g., spacing and size of convective clouds

Source: Christoph Schär, ETH Zurich
Structural and bulk convergence

Statistics of cloud area

- No structural convergence

Statistics of up- & downdrafts

- Bulk statistics of updrafts converges

Source: Christoph Schär, ETH Zurich

(Panosetti et al. 2018)
What resolution is needed?

- There are threshold scales in the atmosphere and ocean: going from 100 km to 10 km is incremental, 10 km to 1 km is a leap. At 1 km
  - it is no longer necessary to parametrise precipitating convection, ocean eddies, or orographic wave drag and its effect on extratropical storms;
  - ocean bathymetry, overflows and mixing, as well as regional orographic circulation in the atmosphere become resolved;
  - the connection between the remaining parametrisation are now on a physical footing.
- We spend the last five decades in a paradigm of incremental advances. Here we incrementally improved the resolution of models from 200 to 20 km
- Exascale allows us to make the leap to 1 km. This fundamentally changes the structure of our models. We move from crude parametric presentations to an explicit, physics based, description of essential processes.
- The last such step change was fifty years ago. This was when, in the late 1960s, climate scientists first introduced global climate models, which were distinguished by their ability to explicitly represent extra-tropical storms, ocean gyres and boundary current.
## Simulation throughput: Simulate Years Per Day (SPYD)

<table>
<thead>
<tr>
<th></th>
<th>NWP</th>
<th>Climate in production</th>
<th>Climate spinup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simulation</strong></td>
<td>10 d</td>
<td>100 y</td>
<td>5'000 y</td>
</tr>
<tr>
<td>Desired wall clock time</td>
<td>0.1 d</td>
<td>0.1 y</td>
<td>0.5 y</td>
</tr>
<tr>
<td><strong>ratio</strong></td>
<td>100</td>
<td>1'000</td>
<td>10'000</td>
</tr>
<tr>
<td><strong>SPYD</strong></td>
<td>0.27</td>
<td>2.7</td>
<td>27</td>
</tr>
</tbody>
</table>
Simulation throughput: Simulate Years Per Day (SPYD)

<table>
<thead>
<tr>
<th></th>
<th>NWP</th>
<th>Climate in production</th>
<th>Climate spinup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>10 d</td>
<td>100 y</td>
<td>5'000 y</td>
</tr>
<tr>
<td>Desired wall clock time</td>
<td>0.1 d</td>
<td>0.1 y</td>
<td>0.5 y</td>
</tr>
<tr>
<td>ratio</td>
<td>100</td>
<td>1'000</td>
<td>10'000</td>
</tr>
</tbody>
</table>

```
SYPD       0.27    2.7    27
```

Minimal throughout 1 SYPD, preferred 5 SYPD
## Summary of intermediate goal (reach by 2021?)

<table>
<thead>
<tr>
<th>Category</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal resolution</td>
<td>1 km (globally quasi-uniform)</td>
</tr>
<tr>
<td>Vertical resolution</td>
<td>180 levels (surface to ~100 km)</td>
</tr>
<tr>
<td>Time resolution</td>
<td>Less than 1 minute</td>
</tr>
<tr>
<td>Coupled</td>
<td>Land-surface/ocean/ocean-waves/sea-ice</td>
</tr>
<tr>
<td>Atmosphere</td>
<td>Non-hydrostatic</td>
</tr>
<tr>
<td>Precision</td>
<td>Single (32bit) or mixed precision</td>
</tr>
<tr>
<td>Compute rate</td>
<td>1 SYPD (simulated year wall-clock day)</td>
</tr>
</tbody>
</table>
Running COSMO 5.0 at global scale on Piz Daint

Scaling to full system size: ~5300 GPU accelerate nodes available

Running a near-global (±80° covering 97% of Earths surface) COSMO 5.0 simulation & IFS
  > Either on the hosts processors: Intel Xeon E5 2690v3 (Haswell 12c).
  > Or on the GPU accelerator: PCIe version of NVIDIA GP100 (Pascal) GPU
Today’s Outlook: GPU-accelerated Weather Forecasting

John Russell

MeteoSwiss New Weather Supercomputer

World’s First GPU-Accelerated Weather Forecasting System

2x Racks
48 CPUs
192 Tesla K80 GPUs
> 90% of FLOPS from GPUs
Operational in 2016
Constant budget for investments and operations

Data assimilation

Ensemble with multiple forecasts

Grid 2.2 km → 1.1 km

Requirements from MeteoSwiss
We need a 40x improvement between 2012 and 2015 at constant cost.
Where the factor 40 improvement came from

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- **Constant budget for investments and operations**
- **Grid 2.2 km → 1.1 km**
- **Ensemble with multiple forecasts**
- **Data assimilation**
- **Requirements from MeteoSwiss**

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- Requirements from MeteoSwiss
- Data assimilation
- Ensemble with multiple forecasts
- Grid 2.2 km → 1.1 km
- 1.7x from software refactoring (old vs. new implementation on x86)
- 24x
- 6x
- Constant budget for investments and operations

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- 24x Ensemble with multiple forecasts
- 10x Grid 2.2 km → 1.1 km
- 6x Data assimilation
- 1.7x from software refactoring (old vs. new implementation on x86)
- 2.8x Mathematical improvements (resource utilisation, precision)
- 10x Constant budget for investments and operations

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- 1.7x from software refactoring (old vs. new implementation on x86)
- 2.8x Mathematical improvements (resource utilisation, precision)
- 2.3x Change in architecture (CPU → GPU)
- 6x Requirements from MeteoSwiss
- Ensemble with multiple forecasts
- 24x
- Grid 2.2 km → 1.1 km
- 10x
- Constant budget for investments and operations

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- **Requirements from MeteoSwiss**
  - 6x improvement
  - Data assimilation

- 1.7x from software refactoring (old vs. new implementation on x86)

- 2.8x from mathematical improvements (resource utilisation, precision)

- Ensemble with multiple forecasts
  - 24x improvement

- 2.3x Change in architecture (CPU → GPU)

- Grid 2.2 km → 1.1 km
  - 10x improvement

- 2.8x Moore’s Law & arch. improvements on x86

- Constant budget for investments and operations

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- **Requirements from MeteoSwiss**: 6x
- **Data assimilation**: 1.7x from software refactoring (old vs. new implementation on x86)
- **Ensemble with multiple forecasts**: 24x
- **Grid 2.2 km → 1.1 km**: 10x
- **Constant budget for investments and operations**: 10x
- **2.8x Moore’s Law & arch. improvements on x86**: 2.8x
- **2.3x Change in architecture (CPU → GPU)**: additional processors

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

- Requirements from MeteoSwiss
- Data assimilation
- Ensemble with multiple forecasts
- Grid 2.2 km → 1.1 km
- Constant budget for investments and operations

1.7x from software refactoring (old vs. new implementation on x86)
2.8x Mathematical improvements (resource utilisation, precision)
2.3x Change in architecture (CPU → GPU)
2.8x Moore’s Law & arch. improvements on x86
1.3x additional processors

Bonus: reduction in power!

We need a 40x improvement between 2012 and 2015 at constant cost
Where the factor 40 improvement came from

Investment in software allowed mathematical improvements and change in architecture

We need a 40x improvement between 2012 and 2015 at constant cost

There is no silver bullet!
Near-global climate simulation at 1km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018
Near-global climate simulation at 1km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

(c) Time compression (SYPD) and energy cost (MWh/SY) for three moist simulations. At 930 m grid spacing obtained with a full 10d simulation, at 1.9 km from 1,000 steps, and at 47 km from 100 steps
Near-global climate simulation at 1km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

Metric: simulated years per wall-clock day

2.5x faster than Yang et al.’s 2016 Gordon Bell winner run on TaihuLight!
# The baseline for COSMO-global and IFS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.93 km (non-uniform)</td>
<td>1.25 km</td>
</tr>
<tr>
<td>Shortfall</td>
<td>0.9x</td>
<td>1.25x</td>
</tr>
<tr>
<td>Horizontal resolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical resolution</td>
<td>60 levels (surface to 25 km)</td>
<td>62 levels (surface to 40 km)</td>
</tr>
<tr>
<td>Shortfall</td>
<td>3x</td>
<td>3x</td>
</tr>
<tr>
<td>Time resolution</td>
<td>6 s (split-explicit with sub-stepping)</td>
<td>120 s (semi-implicit)</td>
</tr>
<tr>
<td>Shortfall</td>
<td>-</td>
<td>4x</td>
</tr>
<tr>
<td>Coupled</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Coupled Shortfall</td>
<td>1.2x</td>
<td>1.2x</td>
</tr>
<tr>
<td>Atmosphere</td>
<td>Non-hydrostatic</td>
<td>Non-hydrostatic</td>
</tr>
<tr>
<td>Precision</td>
<td>Double</td>
<td>Single</td>
</tr>
<tr>
<td>Compute rate</td>
<td>0.043 SYPD</td>
<td>0.088 SYPD</td>
</tr>
<tr>
<td>Shortfall</td>
<td>23x</td>
<td>11x</td>
</tr>
<tr>
<td>Other (I/O, full physics, ...)</td>
<td>Limited I/O Only microphysics</td>
<td>Full physics, no I/O</td>
</tr>
<tr>
<td>Shortfall</td>
<td>1.5x</td>
<td>-</td>
</tr>
<tr>
<td>Total shortfall</td>
<td>65x</td>
<td>198x</td>
</tr>
</tbody>
</table>
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q B}{D \hat{B}} \]
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D} \frac{B}{\hat{B}} \]
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D} \hat{B} \]
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{\hat{Q}}{\hat{B}} \]

0.88

Necessary data transfers

Actual data transfers
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D} \frac{B}{\hat{B}} \]

Necessary data transfers
Actual data transfers
Max achievable BW

\(0.88\)
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D} \cdot \frac{B}{\hat{B}} \]

- Necessary data transfers
- Actual data transfers
- Achieved BW
- Max achievable BW

\[ 0.88 \]
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[
MUE = I/O \text{ efficiency} \cdot BW \text{ efficiency} = \frac{Q}{D} \cdot \frac{B}{\hat{B}}
\]

\(0.88\) \hspace{1cm} \(0.76\)

- Necessary data transfers
- Achieved BW
- Actual data transfers
- Max achievable BW
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D} \cdot \frac{B}{\hat{B}} \]

- **Necessary data transfers**: \( Q \)
- **Achieved BW**: \( \frac{Q}{D} \)
- **Actual data transfers**: \( B \)
- **Max achievable BW**: \( \hat{B} \)

\[ F_P = \text{data set size} \cdot \text{number of independent runs} \]

\[ T = \frac{F_P}{B} \]

\[ MUE = T \cdot F_P \]

\[ \text{FLOPs} = \text{data set size} \cdot \text{number of independent runs} \cdot \text{MUE} \]

- **CUDA STREAM (double)**: \( a[i] = b[i] \)
- **CUDA STREAM (float)**: \( a[i] = b[i] \)
- **COPY (double)**: \( a[i] = b[i] \)
- **COPY (float)**: \( a[i] = b[i] \)
- **AVG i-stride (float)**: \( a[i]=b[i+1]+b[i+3]+b[i+5] \)
- **5-POINT (float)**: \( a[i]=b[i] + b[i+1] + b[i-1] + b[i+j\text{stride}] + b[i-j\text{stride}] \)

\[ \frac{\text{Necessary data transfers}}{\text{Actual data transfers}} = \frac{Q}{B} = \text{MUE} \]

\[ \frac{\text{Max achievable BW}}{\text{Achieved BW}} = \frac{\hat{B}}{\frac{Q}{D}} = \frac{B}{D} \]

\[ \text{Achieved BW} = \frac{Q}{D} \]

\[ \text{Max achievable BW} = \frac{B}{\hat{B}} \]
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = I/O \text{ efficiency} \cdot BW \text{ efficiency} = \left( \frac{Q}{D} \right) \left( \frac{B}{\hat{B}} \right) \]

- Necessary data transfers
- Achieved BW
- Actual data transfers
- Max achievable BW

2x lower than peak BW

\( a[i] = b[i] \)

\( a[i] = b[i] + b[i+1] + b[i-1] + b[i+j\text{stride}] + b[i-j\text{stride}] \)

\( a[i] = b[i] \) (1D)

\( a[i] = b[i+1] + b[i] + b[i-1] + b[i+j\text{stride}] + b[i-j\text{stride}] \)
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

\[ MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D} \cdot \frac{B}{\hat{B}} = 0.67 \]

Necessary data transfers
Achieved BW
Actual data transfers
Max achievable BW

2x lower than peak BW
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

4x
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

\[ a[i] = b[i+1] + b[i-1] + b[i+j\text{stride}] + b[i-j\text{stride}] \]

\[ a[i] = b[i] \]
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

2. Improving BW efficiency
   - Improve BW efficiency and peak BW
   - (results on Volta show this is realistic)
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   
   2x fewer grid-columns
   
   Time step of 10 ms instead of 5 ms

2. Improving BW efficiency

   Improve BW efficiency and peak BW
   
   (results on Volta show this is realistic)
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

2. Improving BW efficiency
   - Improve BW efficiency and peak BW
     (results on Volta show this is realistic)

3. Weak scaling
   - 4x possible in COSMO, but we reduced available parallelism by factor 2
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

2. Improving BW efficiency
   - Improve BW efficiency and peak BW
     (results on Volta show this is realistic)

3. Weak scaling
   - 4x possible in COSMO, but we reduced available parallelism by factor 2

4. Remaining reduction in shortfall
   - Numerical algorithms (larger time steps)
   - Further improved processors / memory
How realistic is it to overcome 65-fold shortfall of a grid-based implementation like COSMO-global?

1. Icosahedral grid (ICON) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

2. Improving BW efficiency
   - Improve BW efficiency and peak BW
     (results on Volta show this is realistic)

3. Weak scaling
   - 4x possible in COSMO, but we reduced available parallelism by factor 2

4. Remaining reduction in shortfall
   - Numerical algorithms (larger time steps)
   - Further improved processors / memory

But we don’t want to increase the footprint of the 2021 system beyond “Piz Daint”
The importance of ensembles

Peter Bauer, ECMWF
The importance of ensembles

Peter Bauer, ECMWF
Remaining goals beyond 2021 (by 2024?)

1. Improve the throughput to 5 SYPD

2. Reduce the footprint of a single simulation by up to factor 10
Data challenge: what is all models run at 1 km scale?

With current workflow used by the climate community, IPPC at 1 km scale avg. would produce 50 exabytes of data!

The only way out would be to analyse the model while the simulation is running

New workflow:
1. first set of runs generate model trajectories and checkpoints
2. Reconstruct model trajectories from checkpoints and analyse (as often as necessary)
Summary and Conclusions
Summary and Conclusions

• While flop/s may be good to compare performance in history of computing, it is not a good metric to design the systems of the future
Summary and Conclusions

• While flop/s may be good to compare performance in history of computing, it is not a good metric to design the systems of the future

• Given today’s challenges use Memory Use Efficiency (MUE) instead

  MUE = “I/O Efficiency” X “Bandwidth Efficiency”
Summary and Conclusions

• While flop/s may be good to compare performance in history of computing, it is not a good metric to design the systems of the future

• Given today's challenges use Memory Use Efficiency (MUE) instead
  \[ \text{MUE} = \text{"I/O Efficiency" \times "Bandwidth Efficiency"} \]

• Convection resolving weather and climate simulations @ 1 km horizontal resolution
  • Represents a big leap for quality of weather and climate simulations
  • Aim is to pull in the milestone by a decade, in the early 2020s rather than 2030s
  • Desired throughput of 1 SYPD by 2021 and 5 SYPD by mid 2020s
  • Need a 50-200x performance improvement to reach 2021 goal of 1SYPD
Summary and Conclusions

• While flop/s may be good to compare performance in history of computing, it is not a good metric to design the systems of the future

• Given todays challenges use Memory Use Efficiency (MUE) instead

  \[
  \text{MUE} = \text{"I/O Efficiency" \times \"Bandwidth Efficiency"}
  \]

• Convection resolving weather and climate simulations @ 1 km horizontal resolution

  • Represents a big leap for quality of weather and climate simulations

  • Aim is to pull in the milestone by a decade, in the early 2020s rather than 2030s

  • Desired throughput of 1 SYPD by 2021 and 5 SYPD by mid 2020s

  • Need a 50-200x performance improvement to reach 2021 goal of 1SYPD

• System improvement needed

  • Improve “Bandwidth Efficiency” for regular but non-stride-1 memory access

  • Improve application software

  • Further improvement of scalability

  • Improve algorithms (time integration has potential)
Collaborators

Tim Palmer (U. of Oxford)
Bjorn Stevens (MPI-M)
Peter Bauer (ECMWF)
Oliver Fuhrer (MeteoSwiss)

Nils Wedi (ECMWF)
Torsten Hoefler (ETH Zurich)
Christoph Schar (ETH Zurich)