Day 2 Session 8

- Michael Horner (Moderator)

- Experts
- **Dr Marie-Christine Sawley**
- **Dr Thomas Ludwig**
- **Dr Matthias Fouquet-Lapar**

Exascale Computing is here - will HPC and AI/ML converge and improve our climate & weather skills?

Dr Marie-Christine Sawley

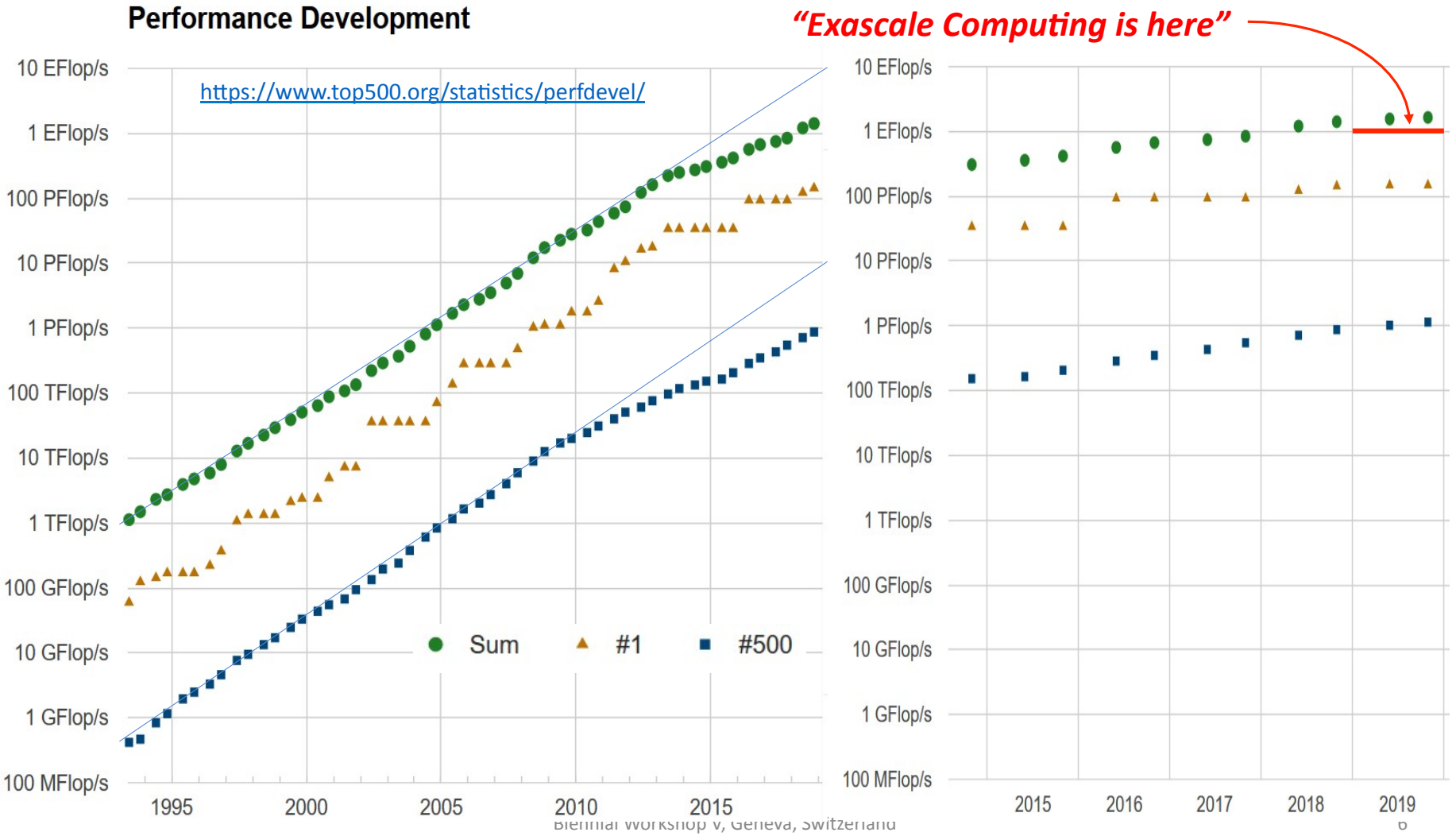
During her career, Marie-Christine has built an experience of more than 20 years in HPC and IT at very large scale. She presently works at Intel, managing technical collaborations between Intel and leading research institutions in order to break some of the exascale challenges such as node performance, power efficency, memory hierarchies, non volatile memory and extreme parallelism. Prior to joining Intel, Marie-Christine worked at ETH Zurich as director of CSCS for 5 years, and for 3 years as CMS scientist at CERN. She held various positions at EPFL during the period 1988-2003, mostly to manage ventures with industrial partners at the core of leading edge innovation in HPC. Marie-Christine holds a EPFL PhD in Plasma Physics and an Exec MBA from IMD. Married with two daughters, she has been travelling between France, Italy, Switzerland and Australia during her life.

Dr Thomas Ludwig

"Thomas Ludwig works in HPC for over 30 years. He started his career at the Technische Universität München where he finished his PhD and habilitation theses. He had a first professor position in Heidelberg for 8 years. In 2009 he moved to Hamburg where he is the director of the German Climate Computing Center. He also holds a position as a professor for HPC at the Universität Hamburg. His research focus is in the field of parallel I/O and file systems. Recently he looks into reproducibility and science organization issues."

Exascale Computing is here*
Can HPC and AI/ML converge to improve our climate & weather skills?

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


Performance Development
https://www.top500.org/statistics/perfdevel/
10 EFlop/s
1 EFlop/s
100 PFlop/s
10 PFlop/s
1 PFlop/s
100 TFlop/s
10 TFlop/s
1 TFlop/s
100 GFlop/s
10 GFlop/s
1 GFlop/s
100 MFlop/s
1995
2000
2005
2010
2015
Sum
#1
#500
"Exascale Computing is here"
2015
2016
2017
2018
2019

The Deceleration of Acceleration

- The end of Moore´s Law
  - Hardly any smaller transistors, i.e.
    - no processors with more performance per cm^2
    - no increased energy efficiency
- Systems at DKRZ (same investment in € plus inflation)
  - 2009: NEC -> IBM with 100x improvement
  - 2015: IBM -> Bull/Atos with 24x improvement
  - 2021: Bull/Atos -> N.N. with 4x improvement
    - 1.7x power consumption
- Partner centers recently
  - LRZ (Garching/München): 4x
  - HLRS (Stuttgart): 3.5x

2019-12-03

AI/ML Exceeds Moore´s Law

Anthony Sarkis: “Why AI progress is faster than Moore’s Law—the age of the algorithm” (Sep 2018)

https://medium.com/@anthony_sarkis/the-age-of-the-algorithm-why-ai-progress-is-faster-than-moores-law-2fb7d5ae7943

- deep learning algorithms improvement
  - region proposal network
  - image classification
- specialized hardware (GPGPU,…)

2019-12-03

Climate Science with Machine Learning Gains Momentum!

Exascale Deep Learning for Climate Analytics

Kurth, Treichler, Romero, Mudigonda, Luehr, Phillips, Mahesh,

2018

We extract pixel-level masks of extreme weather patterns using variants of Tiramisu and DeepLabv3+ neural networks. We describe improvements to the software frameworks, input pipeline, and the network training algorithms necessary to efficiently scale deep learning on the Piz Daint and Summit systems. The Tiramisu network scales to **5300 P100 GPUs** with a sustained throughput of **21.0 PF/s** and parallel efficiency of 79.0%. DeepLabv3+ scales up to **27360 V100 GPUs** with a sustained throughput of **325.8 PF/s** and a parallel **efficiency of 90.7%** in **single precision**. By taking advantage of the **FP16 Tensor Cores**, a half-precision version of the DeepLabv3+ network achieves a peak and sustained throughput of 1.13 EF/s and 999.0 PF/s respectively.

A Comment on Weather, Climate, and ML

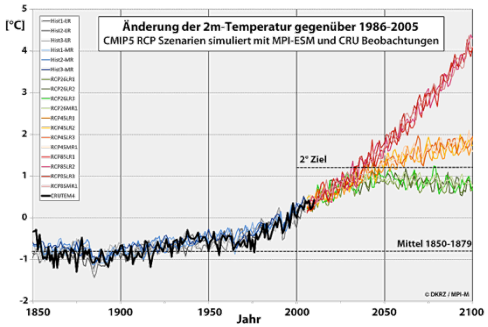
- Weather

  Weather is the **state of the atmosphere**

- Climate

  Climate is the **statistics of weather** over a usually 30 years interval

- Computational weather prediction
  - Make it quick
- Computational climate projection
  - Make it exact



Dr Matthias Fouquet-Lapar

“Matthias started his career in supercomputing with Cray Research at the University of Stuttgart who installed the first Cray-2 supercomputer running a UNIX based Operating Systems. He spend 3 years in the US as part of the Cray T3D SW development team working on Cray’s first massively parallel system. He then joined SGI’s core engineering team to work on scaling SGI’s ccNUMA platform up to 1024 nodes and leading SGI’s Compute Accelerator team. More recently Matthias worked in Huawei 2012 Labs R&D division on long-term strategic planning before joining Lenovo this year as technical lead for HPC and AI/ML"

Exascale Computing Is Here

- Current TOP500 list : Summit Supercomputer #1
  - 2,414,592 cores , 146 TFLOP/s , 10 MW Power
- 45.6% of the systems in China
- Based on industry roadmaps and trends 2021 will see Exascale systems
- However :
  - Compare this to what for example AWS is doing (data from 2017)
    - One data center 45,000 – 50,000 servers , power 25 – 32 MW
    - One Availability Zone re-groups 4-5 DC, > 100 MW, > $1billion investment
  - Typically not focused on energy-efficiency
  - Kolos Data Center, Norway, 600,000 m2, 1000 MW (!)
- With the political will to prioritize climate research over Google or FB recommender systems to influence elections or recommend the best tooth-paste we could have exa-scale system for climate research today, there are no technical hurdles

Exascale Computing Is Here – and ?

- NASA sponsored study of Supercomputer systems.
  … « But many weather and climate models struggle to run efficiently in their HPC environments, with users reporting "fundamentally inappropriate" node and CPU designs and job efficiencies below one percent. »
  [Nov 20th 2019 HPCWire]
  Thank you NASA !

- *Study included «* 15 different weather and climate organizations in the U.S. and Europe, including the European Centre for Medium-Range Weather Forecasts (ECMWF), Los Alamos National Laboratory (LANL), the National Oceanic and Atmospheric Administration (NOAA), Oak Ridge National Laboratory (ORNL), the University Corporation for Atmospheric Research (UCAR), the University of Delaware and more. »
- CPU not perceived as the bottleneck, memory, network and I/O bandwidth are main concerns
- "If you have a data-starved CPU," he said, "it doesn't matter how fast your processor is."
- Peter Bauer (ECMWF) at ISC2019 :
  - For Bauer, weather and climate prediction isn't just *an* application for exascale — it might be *the* application: the use case with the greatest challenges – and the greatest potential to showcase the power of exascale computing
  - " For the most part, the two categories shared solutions – "do less," "do it cheaper," "do it on new technology" and "do it yourself."

Exascale Computing Is Here – what if

- AI/ML will augment science, not replace
- AI/ML has created > 50 startups with new processor / accelerator technologies : what can be leveraged for scientific codes (lower precision arithmetic) ?
- For example:
  - GRAPE-PFN2 523 TFLOP half-precision, MN-3 system with 300 racks and 4800 accelerators : > 2 Exaflops / 3.36 MW in 2020
  - Cerebras Wafer Scale engine, 400,000 cores one on chip, 9 PByte/sec memory bandwidth

Personal thoughts
Are we good citizens ?

- In 2018 online video viewing generated > 300 $MtCO_2$ similar to Spain's greenhouse gas emissions [The Shift Project]
- Digital Technologies now emit 4% of all GHG, more than civil aviation
- By 2025 projected 8% of all GHG, same as current emissions from cars
- New technology and applications are increasing the environmental foot-print
  - 5G driving new applications (more on-line streaming, AR/VR), IoT
  - AI/ML everywhere (smart cities, video surveillance, analytics)
  - More & more data-centers, communications storage
- Climate Research should play a role model in energy efficient computing

Questions?

Thank you!