Panel session 6
Measuring Performance in Geoscience Apps

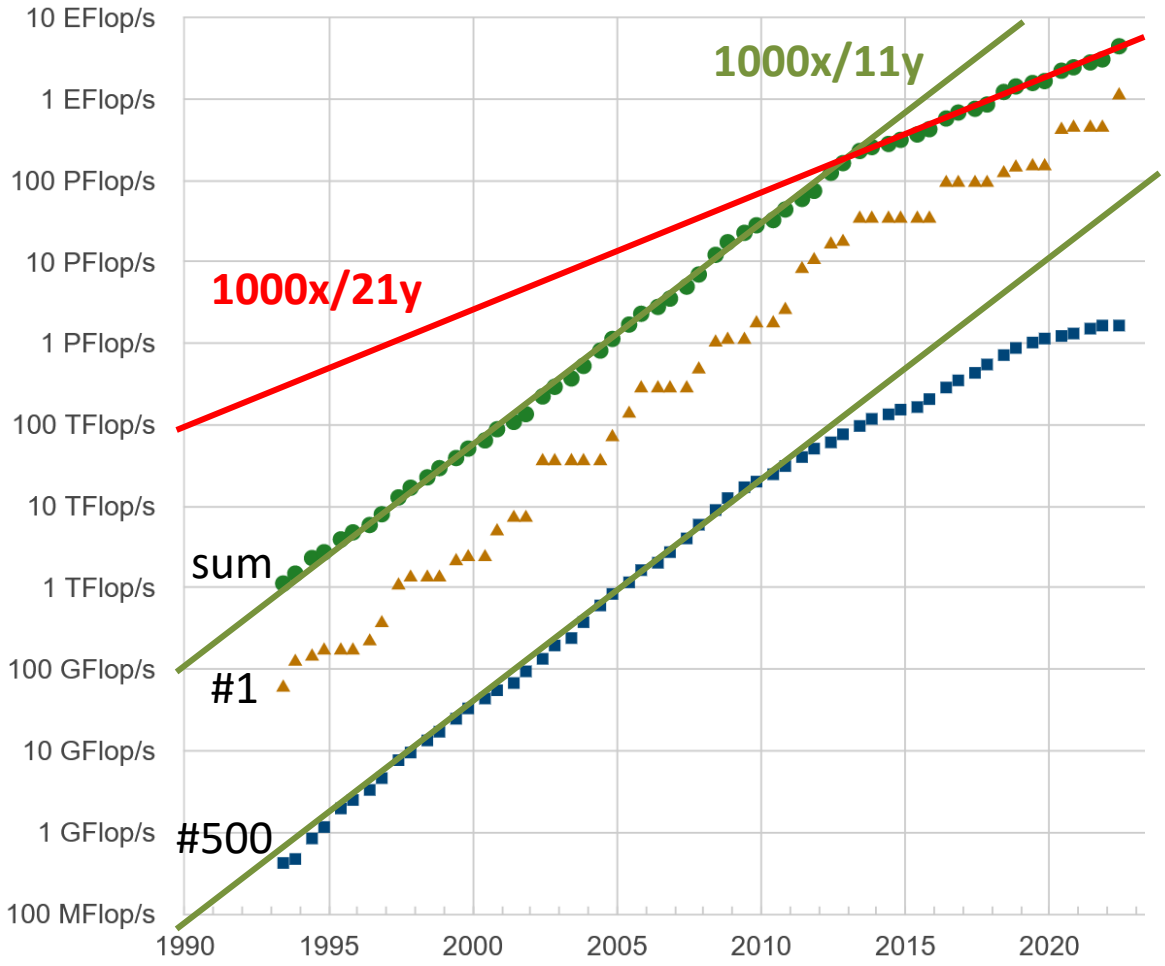
Prof. Dr. Thomas Ludwig
German Climate Computing Center (DKRZ)
University of Hamburg, Department for Computer Science (UHH/FBI)

Dr. Claudia Frauen
German Climate Computing Center (DKRZ)

TOP500 List

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

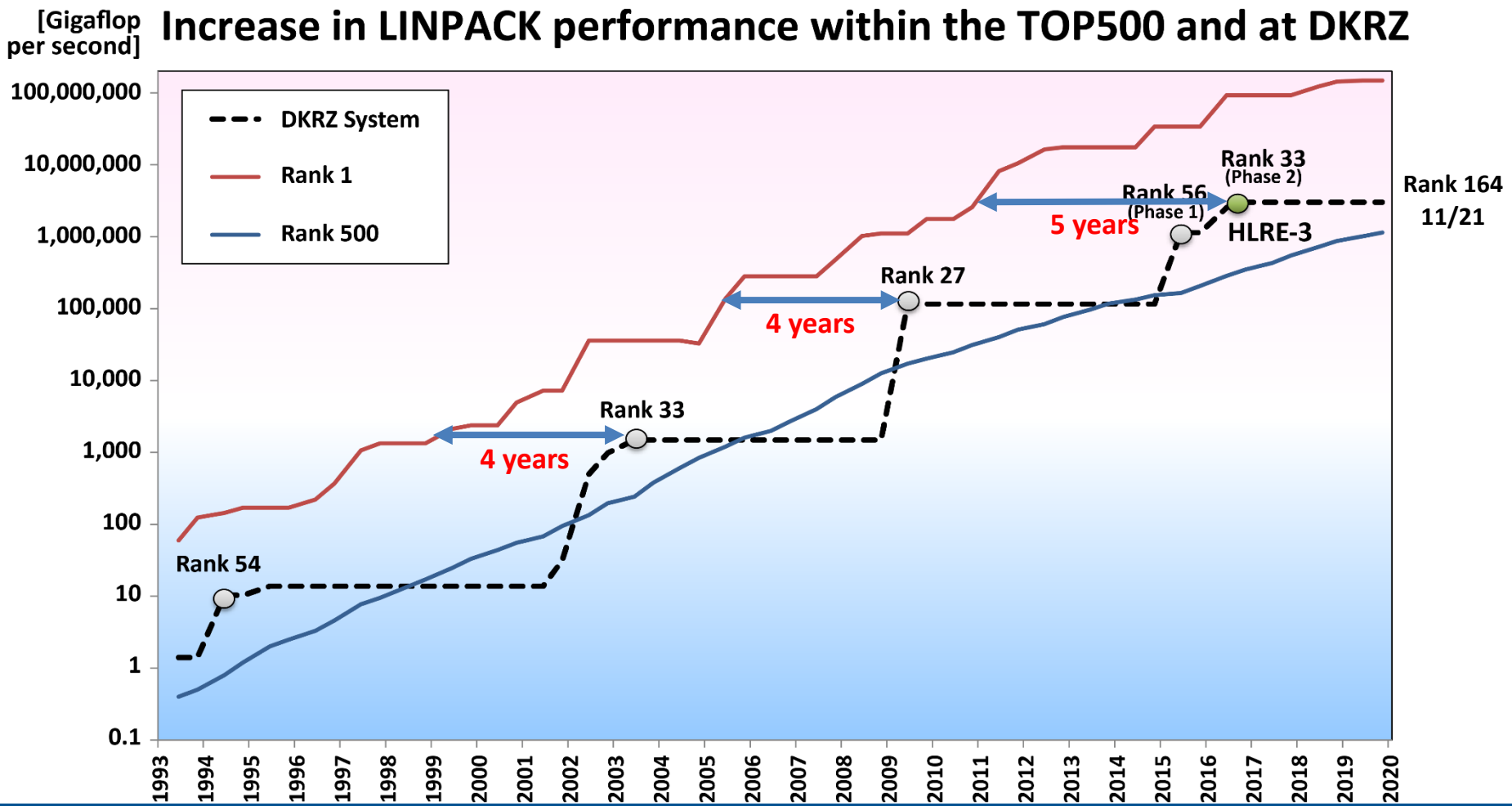
TOP500 List History



Performance Benchmarks:

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

TOP500 DKRZ List History



HLRE-4 “Levante” – #76 in TOP500 June 2022

HLRE-4 “Levante”

Levante Data Sheet 1

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

year	PFLOPS	PFLOPS Factor	GFLOPS/ MW	MW Factor	MW	System Name
2009	0.15					Blizzard
+6						
2015	3.9					Mistral
+7						
2022	17					Levante

Systems at DKRZ

year	PFLOPS	PFLOPS Factor	GFLOPS/ MW	MW Factor	MW	System Name
2009	0.15					Blizzard
+6		26				
2015	3.9					Mistral
+7		4.3				
2022	17					Levante

Systems at DKRZ

year	PFLOPS	PFLOPS Factor	GFLOPS/ MW	MW Factor	MW	System Name
2009	0.15				1.6	Blizzard
+6		26				
2015	3.9				1.4	Mistral
+7		4.3				
2022	17				2.2	Levante

Systems at DKRZ

year	PFLOPS	PFLOPS Factor	GFLOPS/ MW	MW Factor	MW	System Name
2009	0.15				1.6	Blizzard
+6		26		0.9		
2015	3.9				1.4	Mistral
+7		4.3		1.5		
2022	17				2.2	Levante

Systems at DKRZ

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

Energy Efficiency World Wide

year	PFLOPS	PFLOPS Factor	GFLOPS/ MW	MW Factor	MW	System Name
2009	0.15		0.09		1.6	Blizzard
+6		26	29x	0.9		
2015	3.9		2.8		1.4	Mistral
+7		4.3	2.8x	1.5		
2022	17		**7.7**		2.2	Levante

June 2022

#1 Frontiers (USA): **52** GFLOPS/MW with GPU accelerator

#2 Fugaku (Japan): **14** GFLOPS/MW with ARM processors

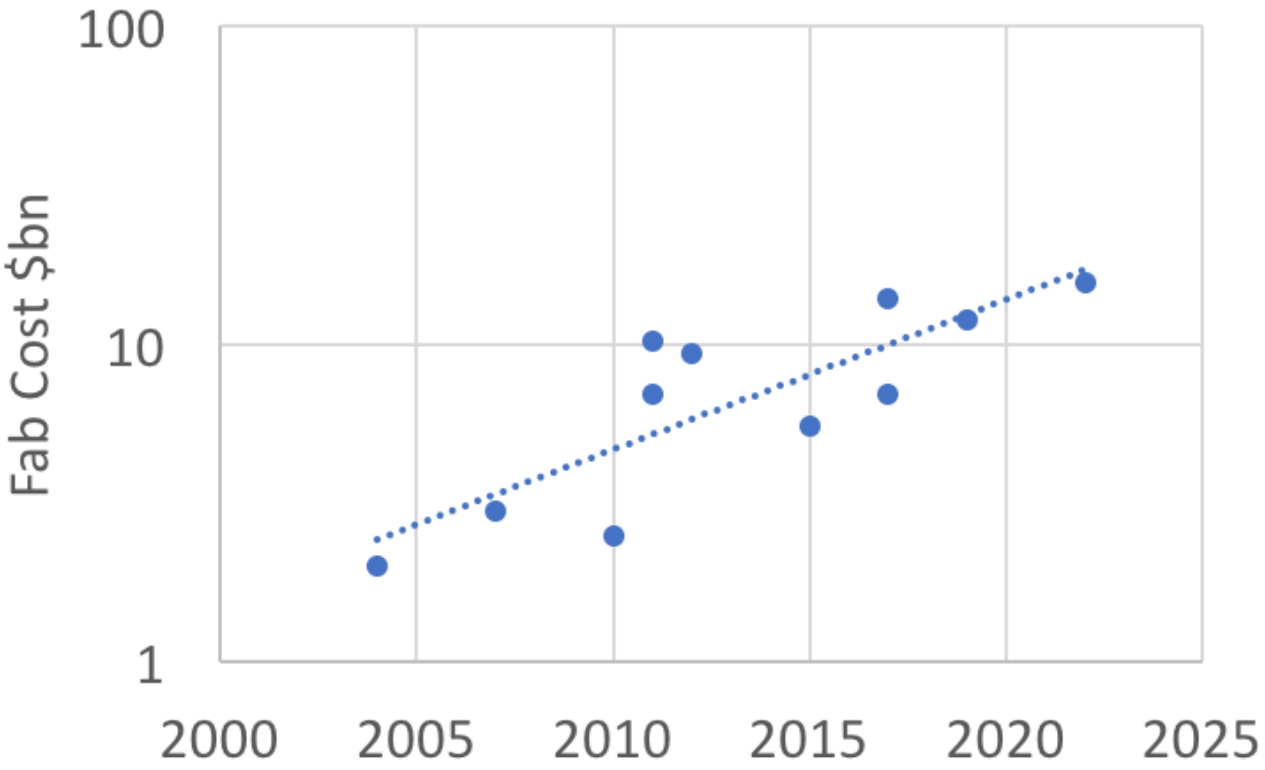
Electronic Waste? (80t per machine at DKRZ)

year	PFLOPS	PFLOPS Factor	GFLOPS/ MW	MW Factor	MW	System Name
2009	0.15		0.09		1.6	Blizzard
+6		26	29x	0.9		
2015	3.9		2.8		1.4	Mistral
+7		4.3	2.8x	1.5		
2022	17		**7.7**		2.2	Levante

scrap the old machine!

continue to use the old machine?

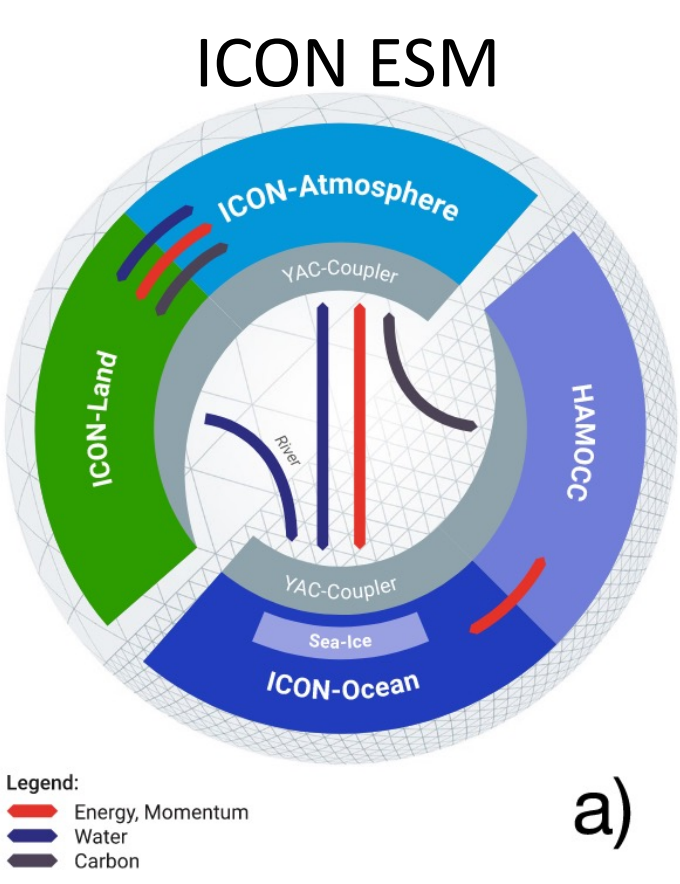
Semiconductor Fabrication Plants Costs



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



Jungclaus et al., 2022

- 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)

Current ICON performance

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

Current ICON performance

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

Coupled high res simulations

Resolution	Machine	Nodes	SDPD
5 km	Mistral	420	17
5 km	Levante	600	126
2.5 km	Levante	600	20
1.25 km	Levante	900	2.5

Hohenegger et al., 2022,
D. Klocke pers. comm.

Current ICON performance

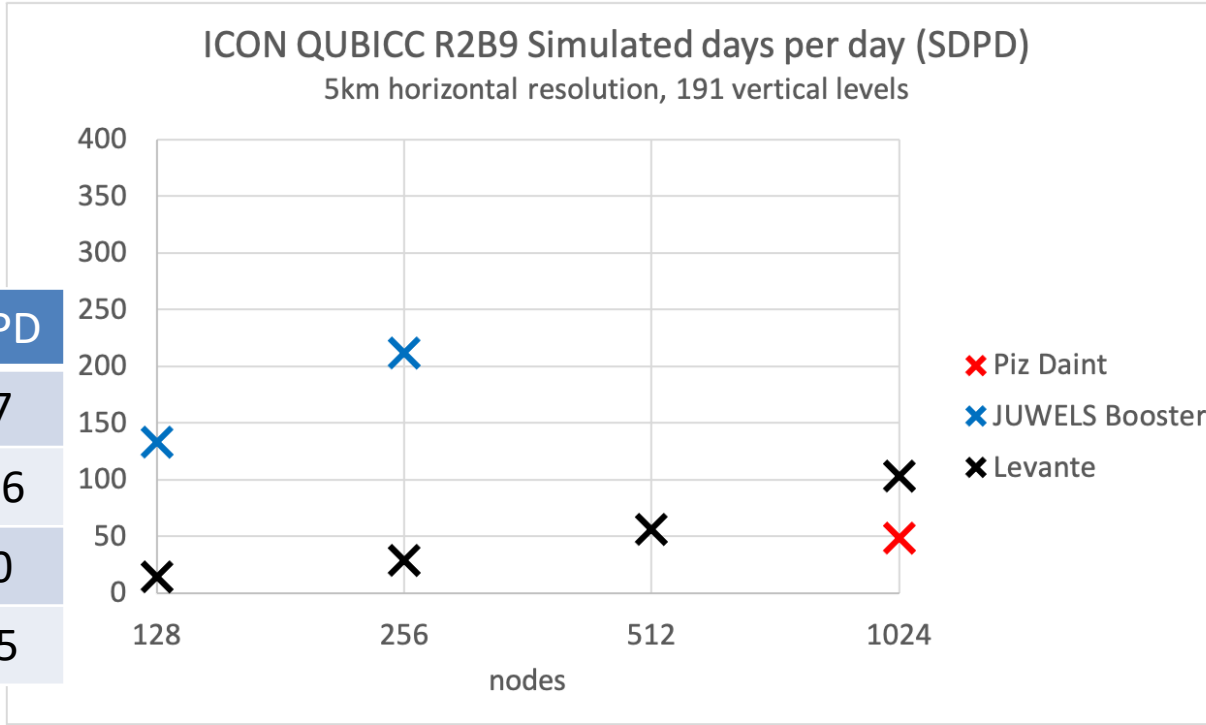
ICON ESM (Resolution:
A -> 158km, O -> 40km)
Mistral 120 nodes -> 120 SYPD

Coupled high res simulations

Resolution	Machine	Nodes	SDPD
5 km	Mistral	420	17
5 km	Levante	600	126
2.5 km	Levante	600	20
1.25 km	Levante	900	2.5

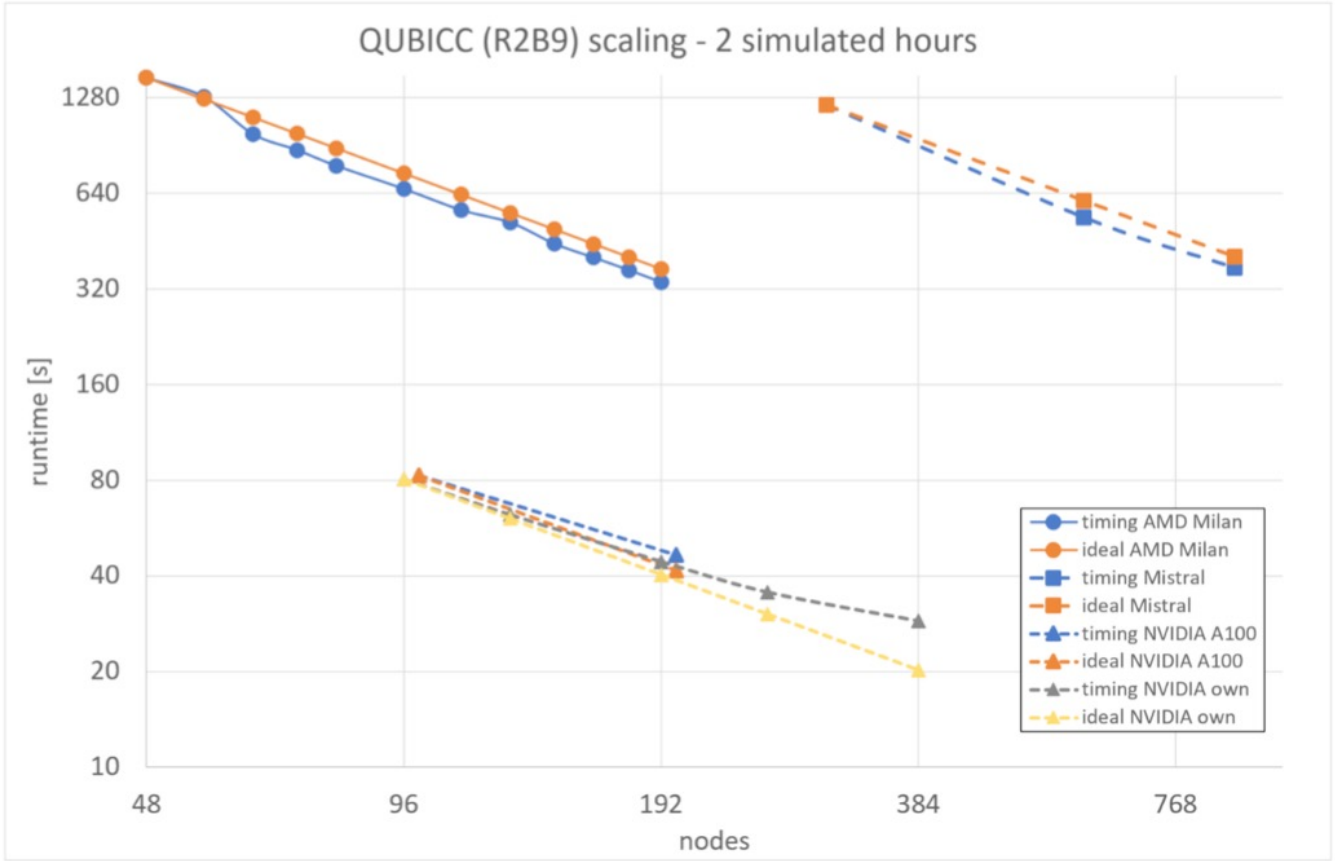
Hohenegger et al., 2022,
D. Klocke pers. comm.

Atmosphere only high res simulations

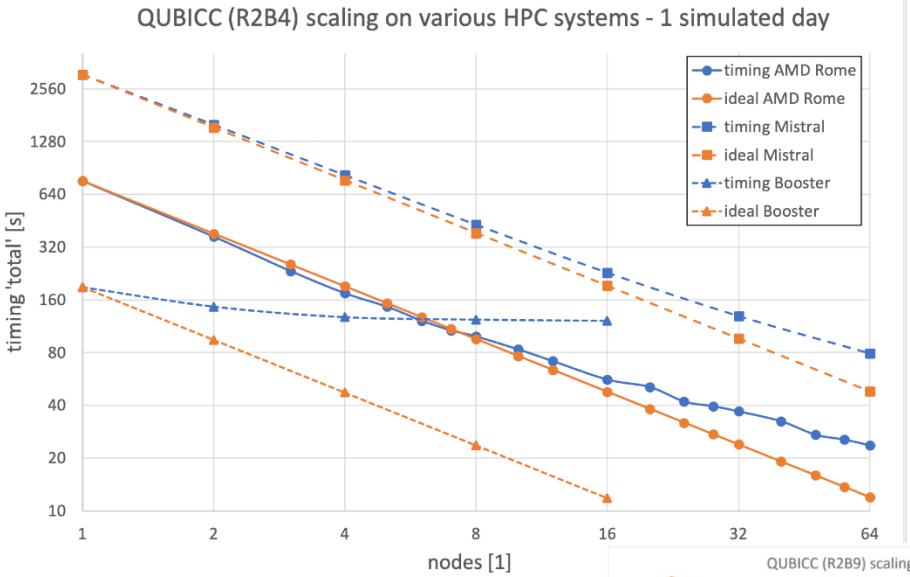


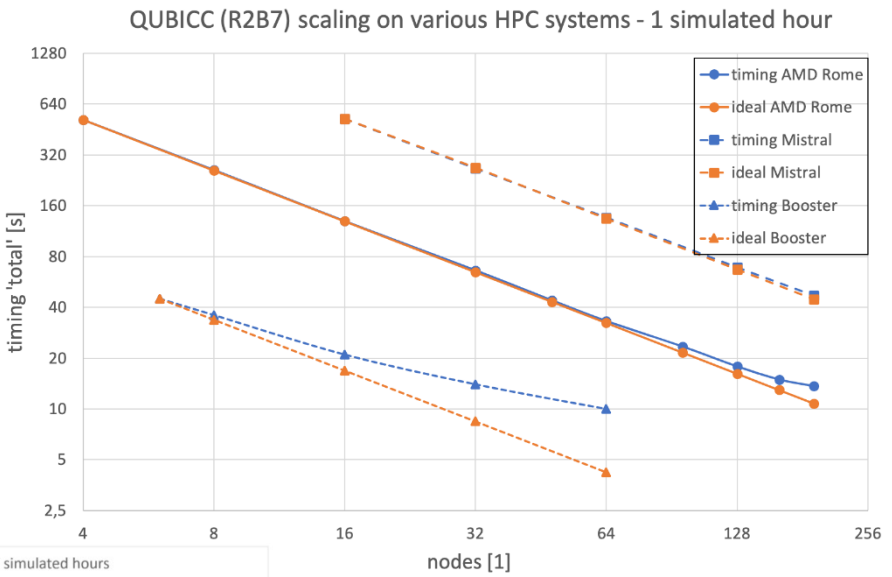
Adapted from Giorgetta et al., 2022

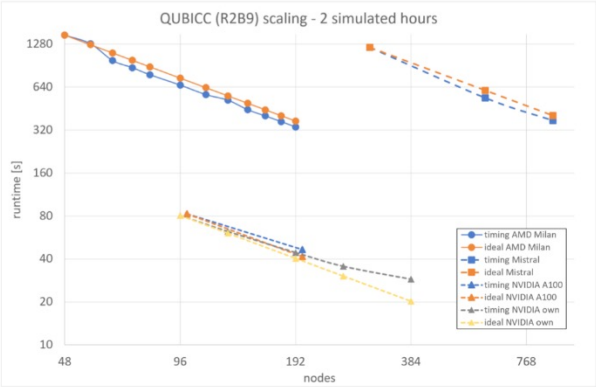
ICON scalability



ICON scalability



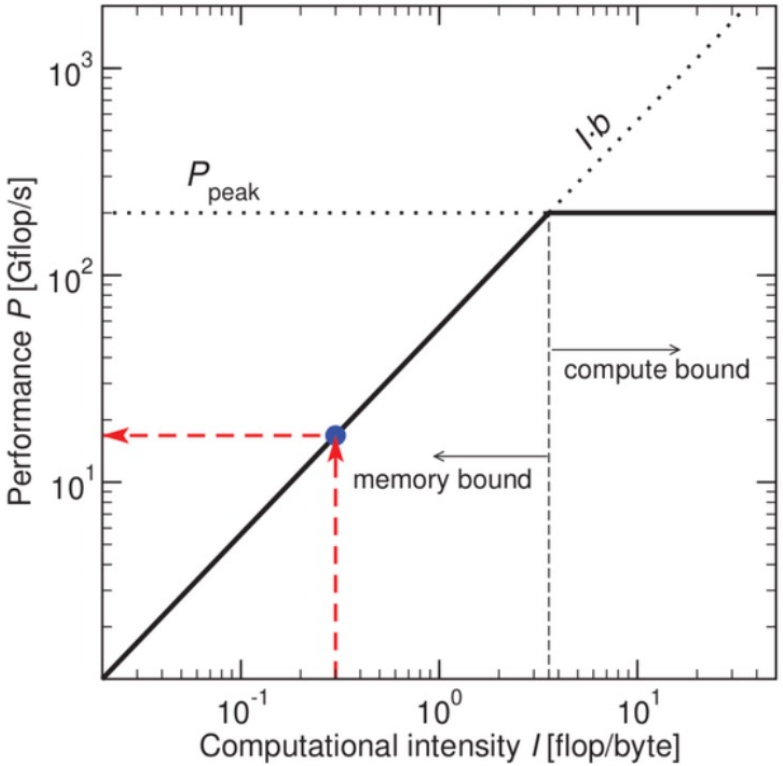


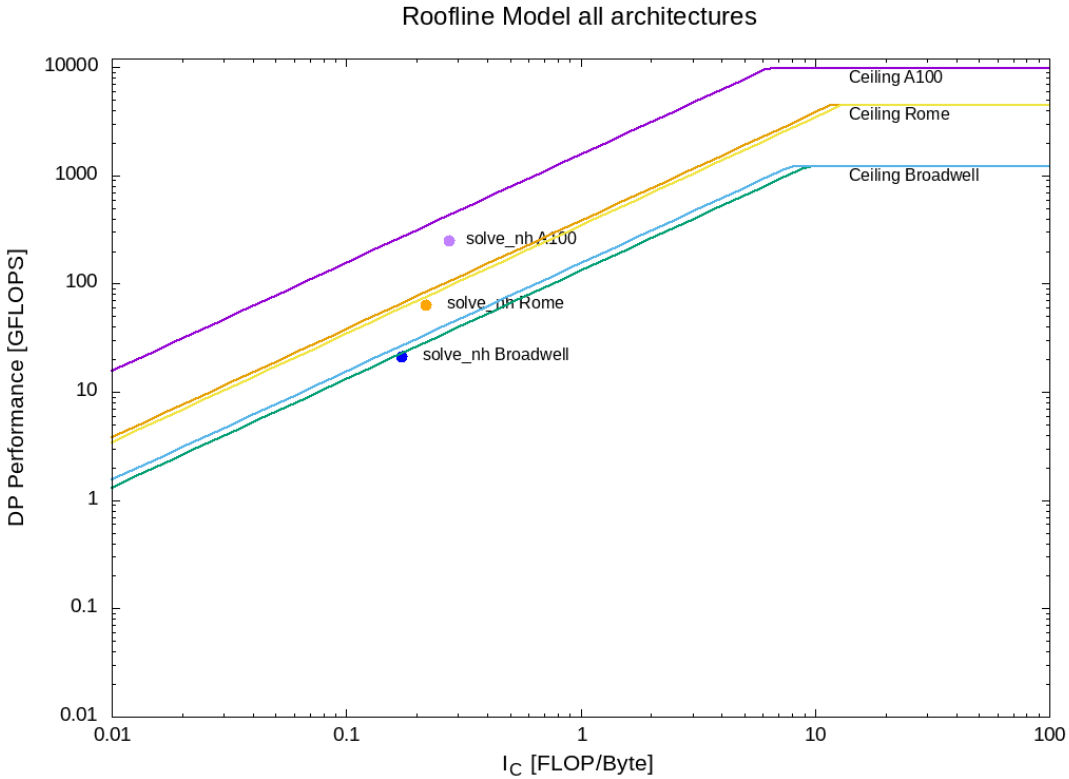


Performance: Compute-bound vs memory-bound

- 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





Exemplary presentation of the roofline model,
https://hpc-wiki.info/hpc/performance_model

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