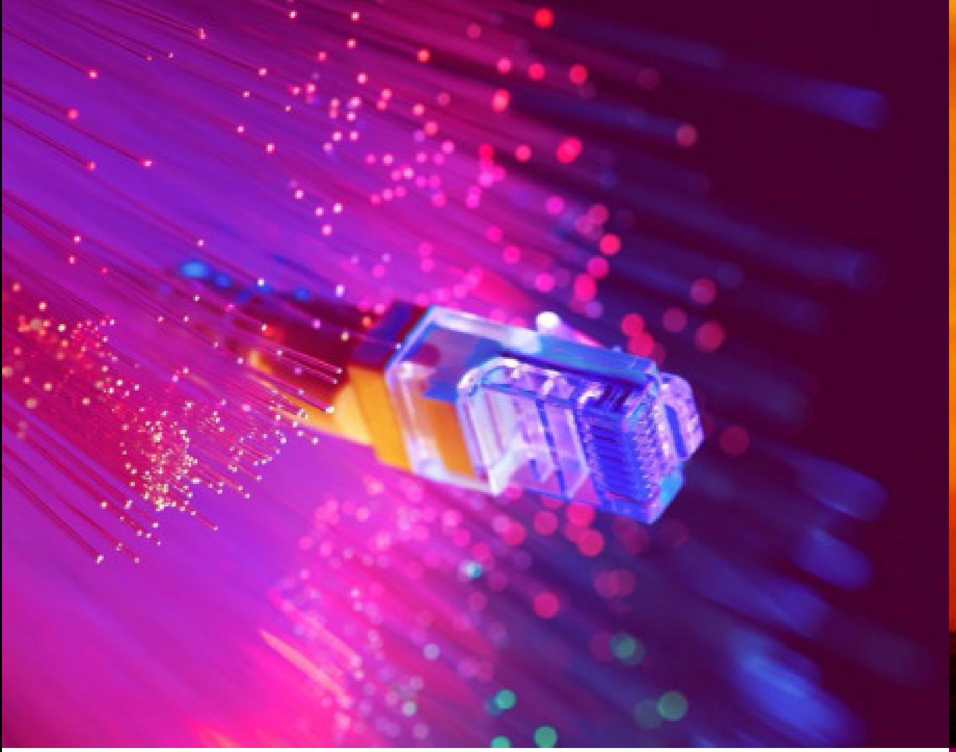
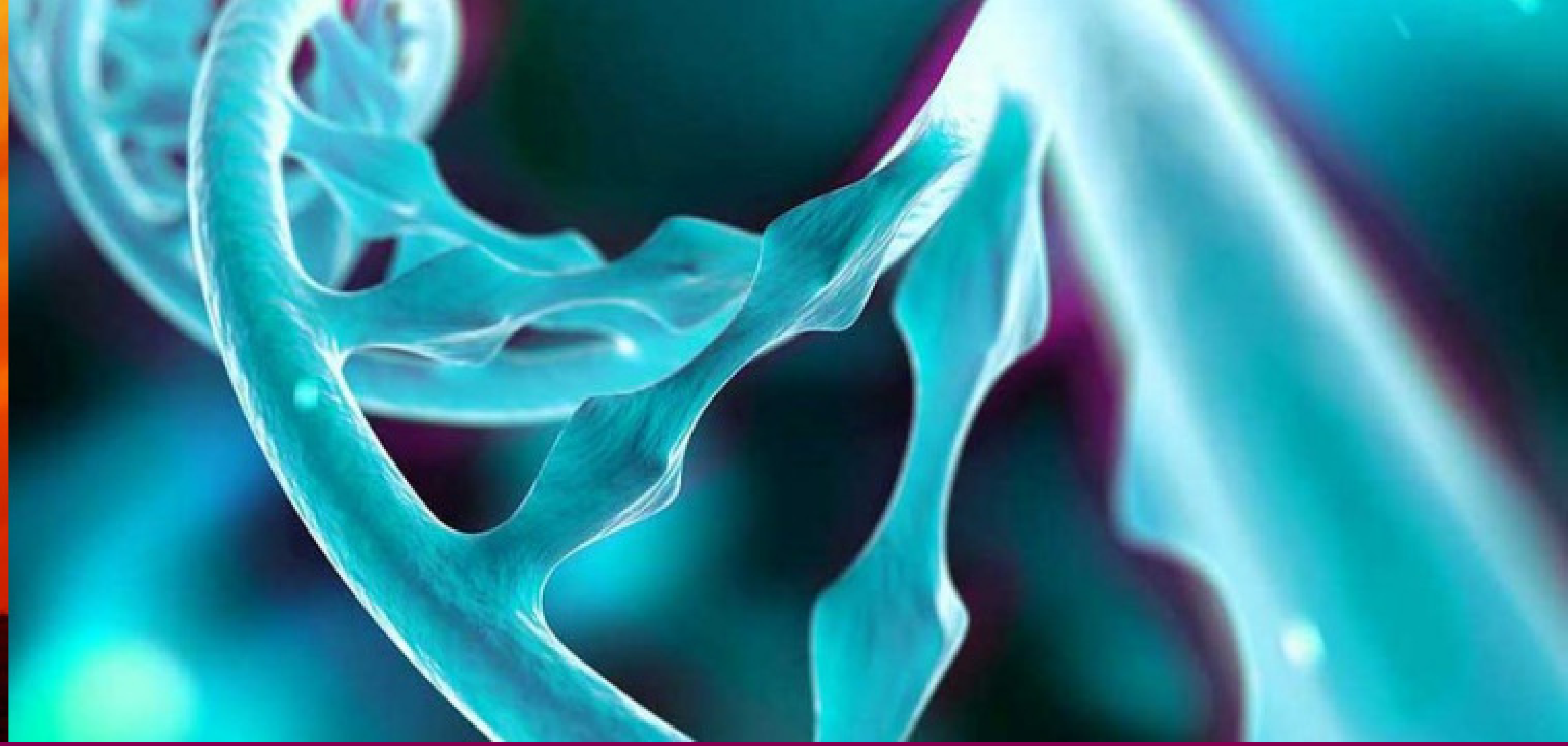




CSC

Enabling km-scale coupled climate simulations with ICON on AMD GPUs

CUG 2025

Jussi Enkovaara, CSC - IT Center for Science


CSC – Finnish expertise in ICT for research, education and public administration

Collaborators

- CSC: Tuomas Lunttila
- Deutsches Klimarechenzentrum: Claudia Frauen, Jan Frederik Engels
- Max-Planck-Institut für Meteorologie: Luis Kornblueh, Lukas Kluft, Rene Redler, Sergey Kosukhin, Reiner Schnur, Daniel Klocke
- HPE: Alfio Lazarro, John Levesque
- AMD: Samuel Antao

Acknowledgements

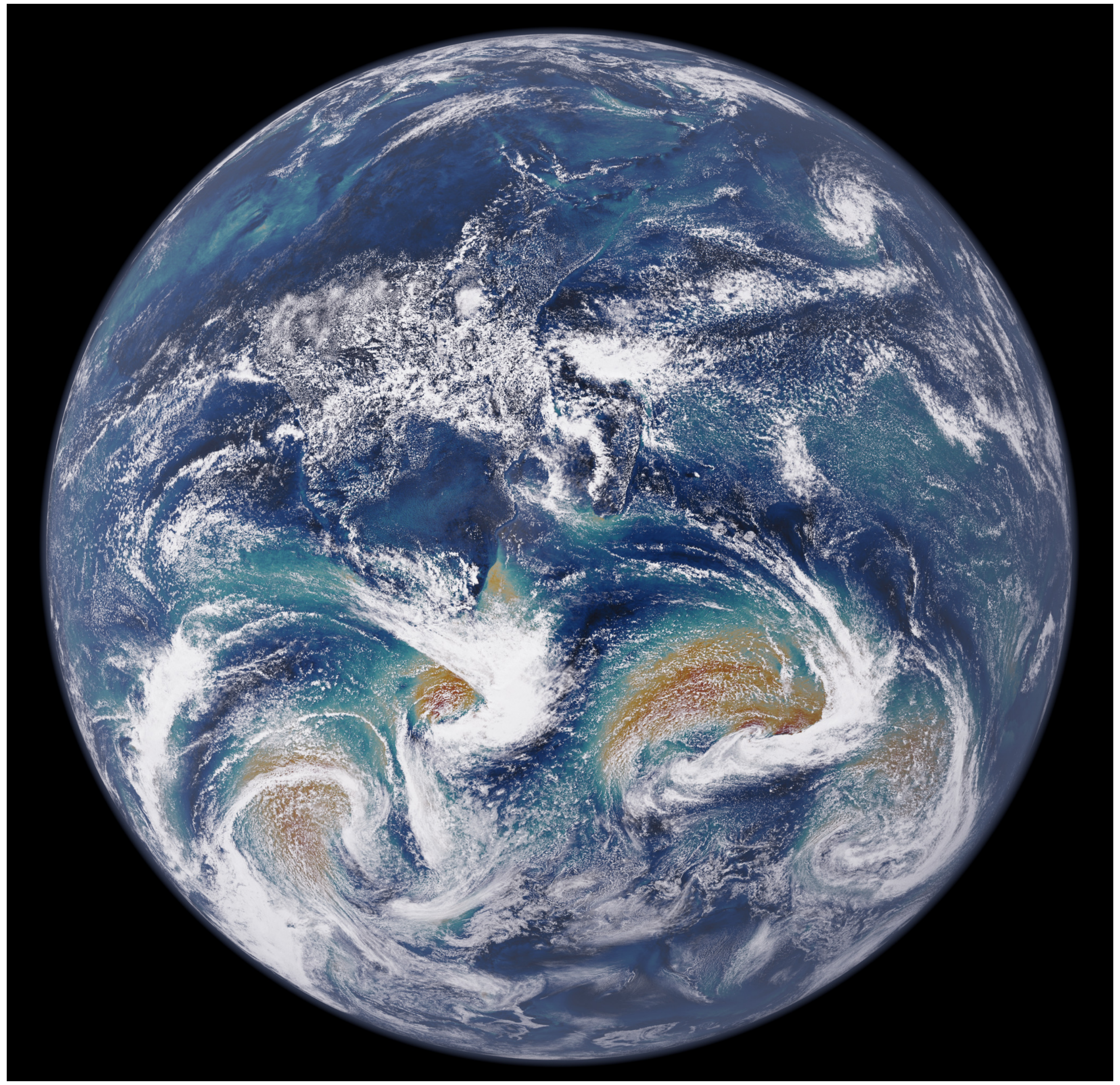
- NVIDIA: Dmitri Alexeev
- The whole ICON developer community
- Destination Earth program https://destination-earth.eu/

Outline

- Climate adaptation digital twin and LUMI supercomputer
- ICON climate model
- Porting ICON to LUMI
- Performance results
- Future prospects

Climate adaptation digital twin

- A new type of climate information system for assessing impacts of climate change and different adaptation strategies
  - https://platform.destine.eu/
- High resolution global climate simulations are key component
  - Eddy-rich, storm-resolving models (10 km or higher)

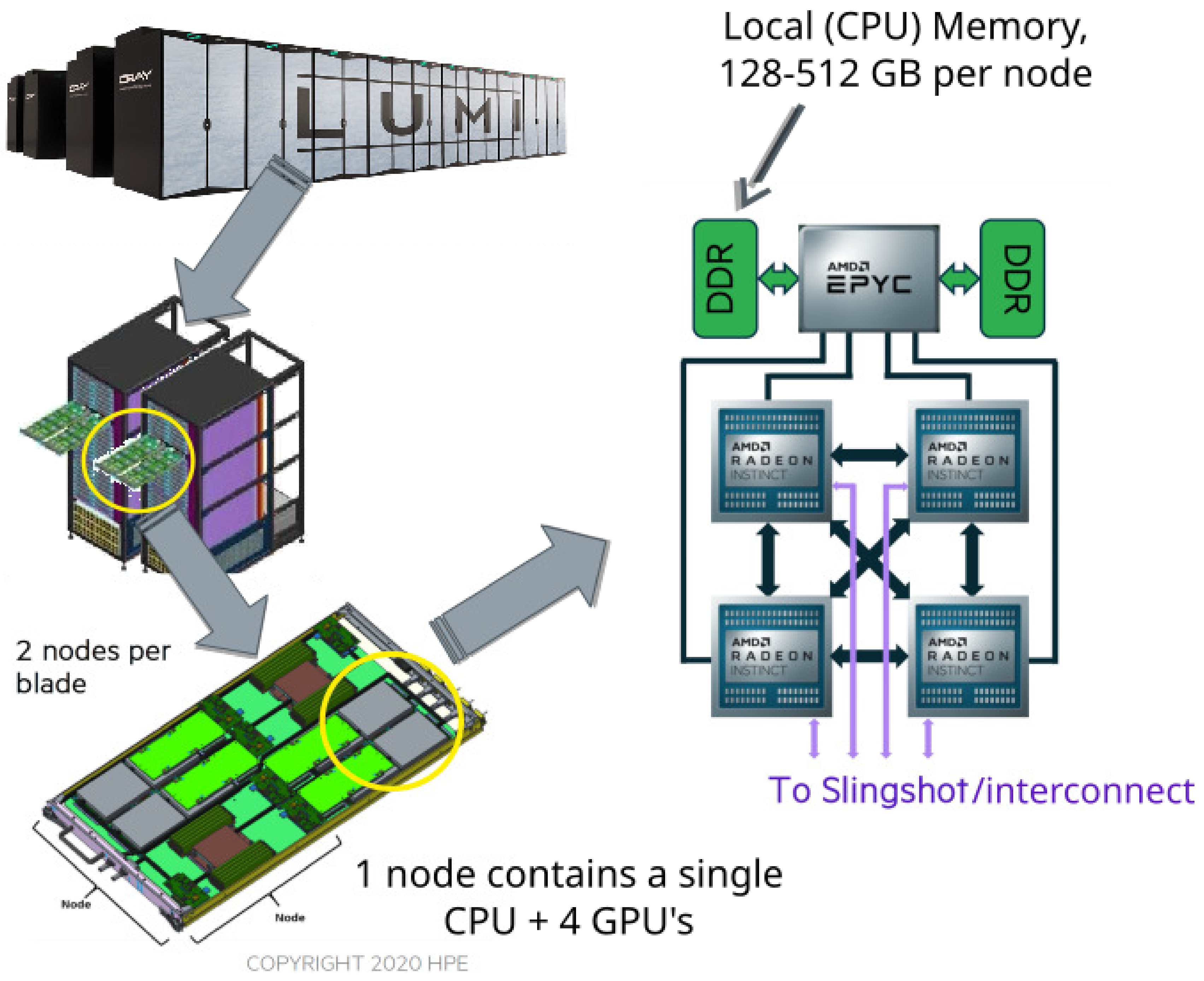


Simulation by ICON

Visualization Niklas Röber, NVIDIA

LUMI supercomputer

- Funded by EuroHPC and consortium of eleven countries
  - Hosted in CSC datacenter
- #3 powerful in Europe, #8 in world (380 PFlop/s)
- ~3000 nodes
  - Each node has 4 AMD MI250X GPUs (logically 8 GPUs per node)
- Supplementary CPU partition with ~260 000 cores (8.4 PFlop/s)



Introduction to ICON

- Icosahedral Nonhydrostatic Weather and Climate Model
- Next generation modeling system capable of global simulations down to 1km resolutions
- Rewrite of earlier weather model in in the early 2000s
  - 1 600 000 lines of modern Fortran
- Developed my several meteorology institutes
  - DWD, MPI-M, MeteoSwiss, ...
  - ~200 developers
- Operational weather forecasts since 2015
- Open source since 2024 (https://www.icon-model.org/)

ICON on GPUs

- First GPU developments in 2010
- GPU version using OpenACC and few CUDA kernels
  - ~1000 Fortran source files, ~300 contain OpenACC directives
  - ~15 000 lines with OpenACC directives
- (Almost) all computations in GPUs
  - Ocean currently in CPUs
- Production runs in NVIDIA GPUs since 2020
- Part of LUMI procurement benchmarks

Porting ICON to LUMI

- Wide range of problems
  - Build time errors
  - Runtime errors with clear diagnostics
  - Runtime crashes (segfaults etc.)
  - Incorrect results
  - Performance problems
- Generally, in all error types there has been a mixture of compiler bugs, non-standard compliant behaviour, and bugs in the application code

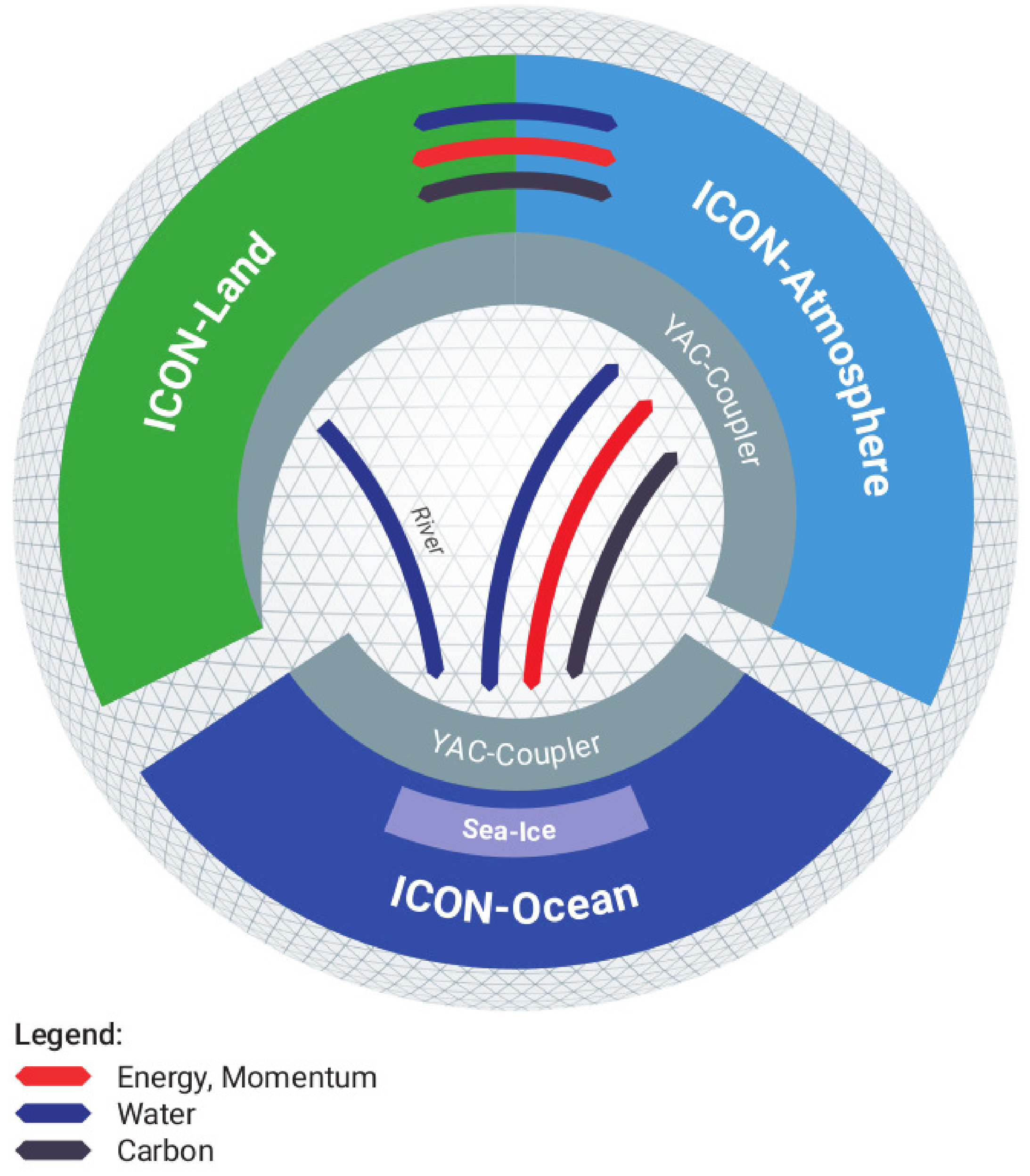
OpenACC support in CCE with ICON

- Early (CCE <= 15) versions had multiple issues
  - Lots of workarounds in the source code
- CCE 16.0.1 : relatively stable, still quite a few workarounds in the source
- CCE 17.0.1 : regressions, new workarounds in the source
- CCE >= 18.0.0 : even more regressions, runtime failures
  - No workarounds yet
  - No simple reproducible bug known at the moment
- CCE team has been provided full source code and small (1-2 node) input data set

Optimizations and performance results

Model setup

- Terrain-following vertical coordinates
  - 90 vertical levels, top 80 km
- Radiative Transfer for Energetics RRTM (RTE-RRTMGP)
- 1-moment (graupel) microphysics and 3-d Smagorinsky turbulence scheme
- JSBACH land model
- ICON ocean model
  - 72 vertical levels



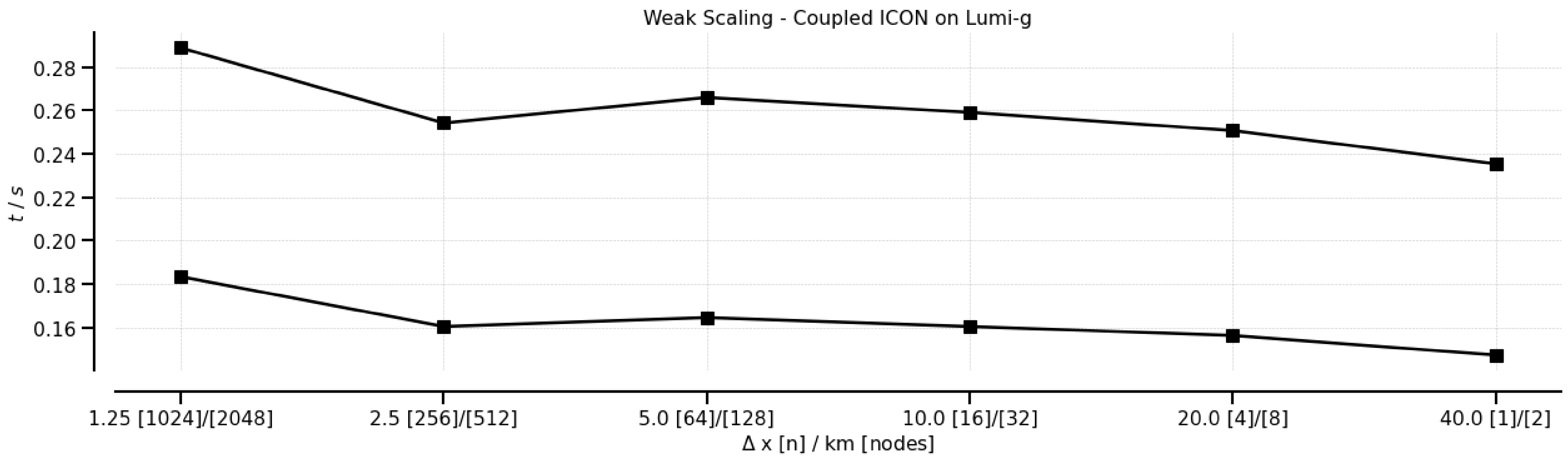
Hybrid CPU - GPU setup for coupled simulations

- Each GPU node has a 64 core CPU
- Atmosphere and land components run on GPUs
- Ocean component runs on the CPUs of the GPU nodes
  - With 32 cores per node, ocean runs faster than atmosphere
- Coupling adds ~10 % overhead
- Remaining idle CPU cores can be used also for I/O servers

Optimizations

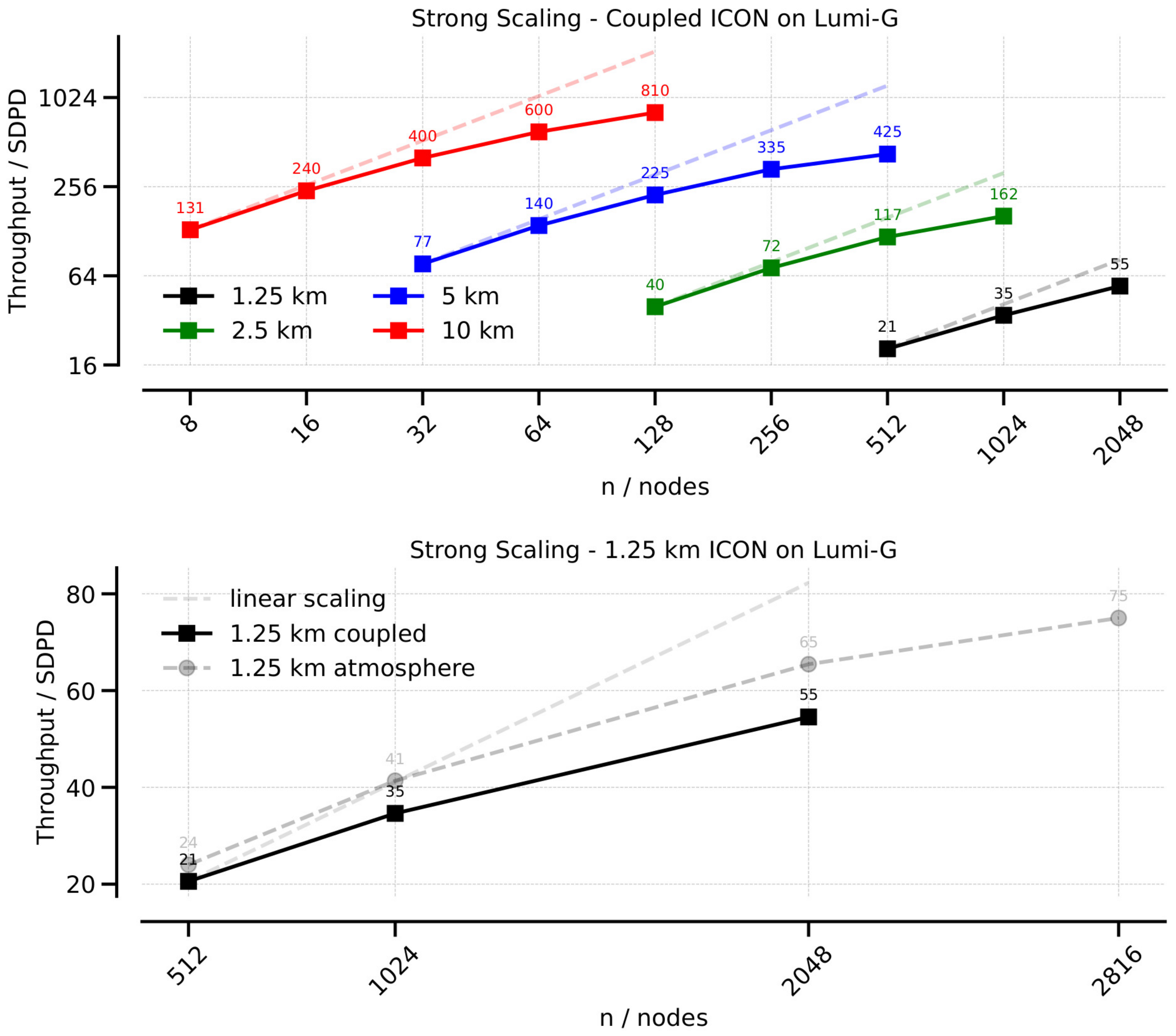
- ICON has very flat performance profile
  - Most time consuming GPU kernels take 3-5 % of the total time
- Optimizing kernel launch latencies has given 20-30 % speedup
  - OpenACC `ASYNC` clause
  - “overwriting” unnecessary HIP calls inserted by OpenACC compiler
- Padding arrays for optimal alignment gives up to 10 % improvement
  - 256 B “magic” alignment
- Some loop interchanges provide few percent speedup
- Few percent improvements from CCE 17.0.1

Weak scaling



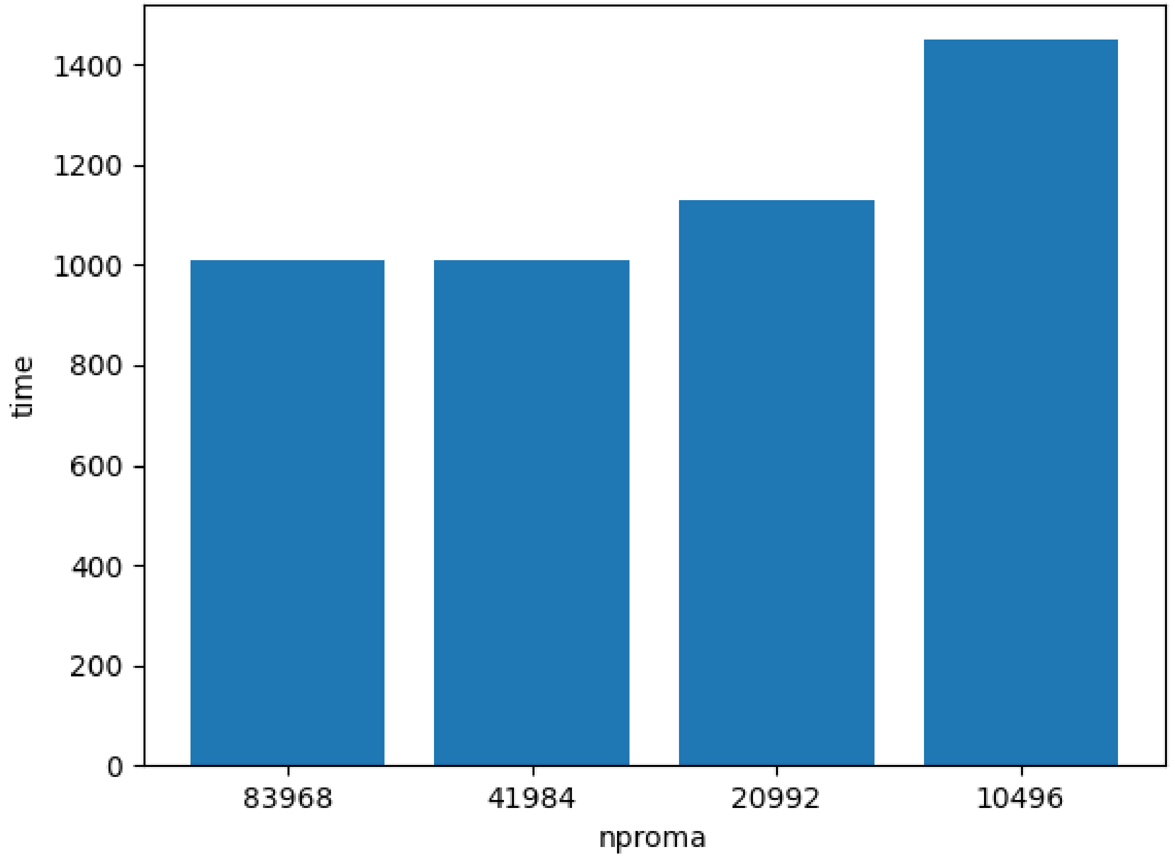
- Technical setup: same number of time steps in all resolutions
  - model init excluded
- Halving resolution increases computational and communication cost by factor of four
- 85 % parallel efficiency from 1 (2) nodes to 1024 (2048) nodes

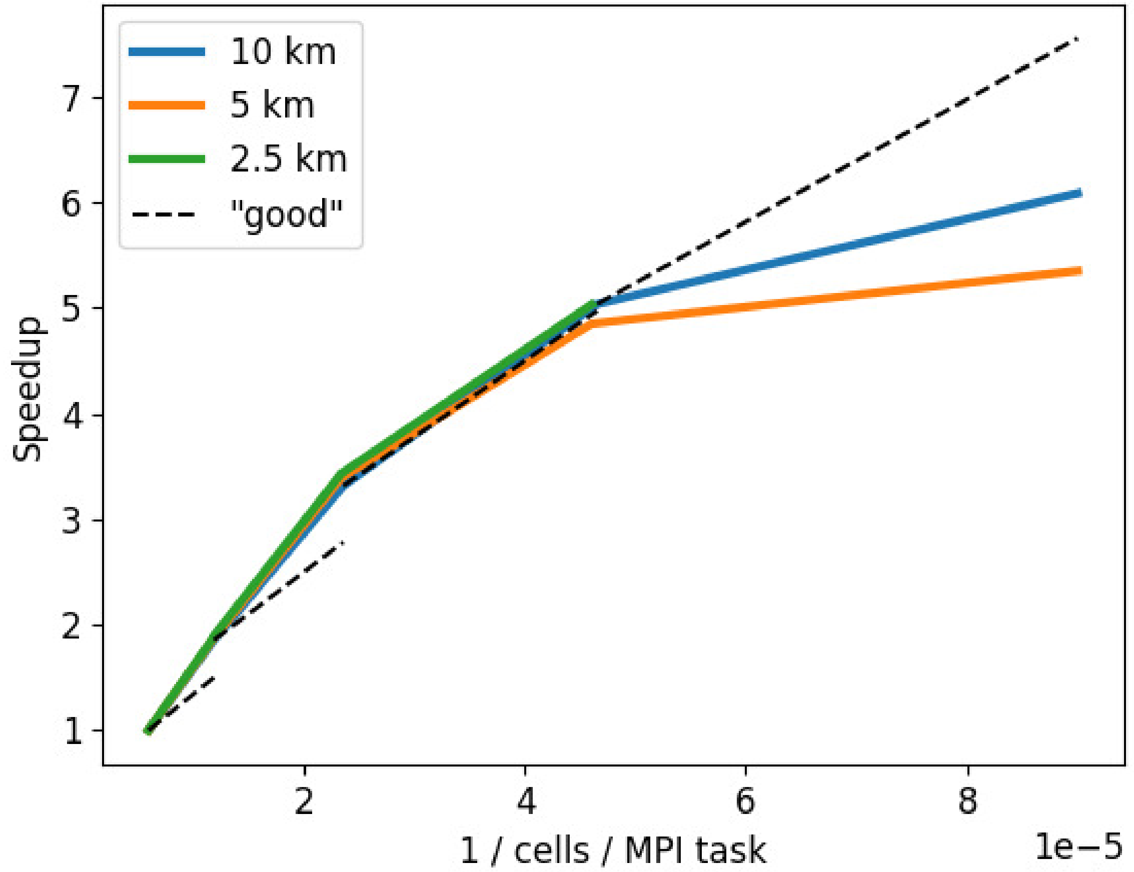
Strong scaling



What limits strong scaling?

- Communication is mostly halo exchange, not major bottleneck
  - Coupling step involves more complex/expensive communication
- With too few horizontal cells per GPU performance starts to suffer
  - For AMD MI250X the limit is about 20 000 cells per GCD





Energy efficiency of ICON with GPUs

- Same number of GPU nodes (four AMD MI250x's per node) and CPU nodes (two 64 core AMD Milans)
  - Atmosphere + land is ~10 times faster on GPUs
  - Coupled simulation is ~13 times faster with GPUs
  - (pure CPU has better strong scaling behaviour)
- GPU nodes consume ~1700 W and CPU nodes ~600 W
- ICON is 3.5 - 4.5 times more energy efficient with GPUs

Potential performance improvements

- Currently, GPU-aware MPI in LUMI is unstable (random crashes)
  - Small scale tests indicate ~15 % performance improvement with GPU-aware communication
- Using single precision for some parts of code
  - Current mixed precision implementation shows no performance benefits on LUMI
- CUDA graphs improve performance in land model and in turbulent mixing
  - Non-standard OpenACC extension available only in NVIDIA systems

Summary

- Porting large codes is non-trivial
  - LUMI port of ICON has required 3-5 FTEs
- Programming model is only as good as compilers supporting it
  - Some ICON development is moving away from Fortran + OpenACC
- ICON exhibits good parallel scalability
  - Strong scalability is limited by data per GPU
- ICON is significantly more energy efficient with GPUs