IFCES2 SCALEXA

Exploring Functional Concurrency and Load Balancing for ICON

Fabian Senf, Panagiotis Adamidis, Estela Suarez, Nobert Eicker,
Daniel Klocke, Carsten Clauss, Matthias Lieber, Wolfgang E. Nagel

Roxana Cremer, Jens Stoll, Xingran Wang, Manoel Römmer, Fatemeh Chegini,
Simon Pickartz, Sonja Happ, Johann Biedermann

natESM Workshop, Hamburg
15 November 2023.

ParTec MODULAR SUPERCOMPUTING
JÜLICH Forschungszentrum
MAX-PLANCK-INSTITUT FÜR METEOROLOGIE
Leibniz Association
DKRZ DEUTSCHES KLIMARECHENZENTRUM
SPONSORED BY THE Federal Ministry of Education and Research
Funded by the European Union NextGenerationEU
TECHNISCHE UNIVERSITÄT DRESDEN
TROPOS Leibniz Institute for Tropospheric Research

Motivation

Vision

- **Is current ESM ready for upcoming hardware heterogeneity?**
- **Can ESM deal with adaptive process complexity?**

TROPOS

Motivation

Vision

- Is current ESM ready for upcoming hardware heterogeneity?
- Can ESM deal with adaptive process complexity?

Climate modeling needs to become more flexible!

TROPOS

Motivation

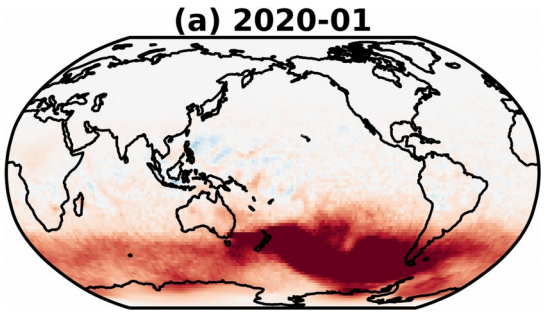
Hardware heterogeneity

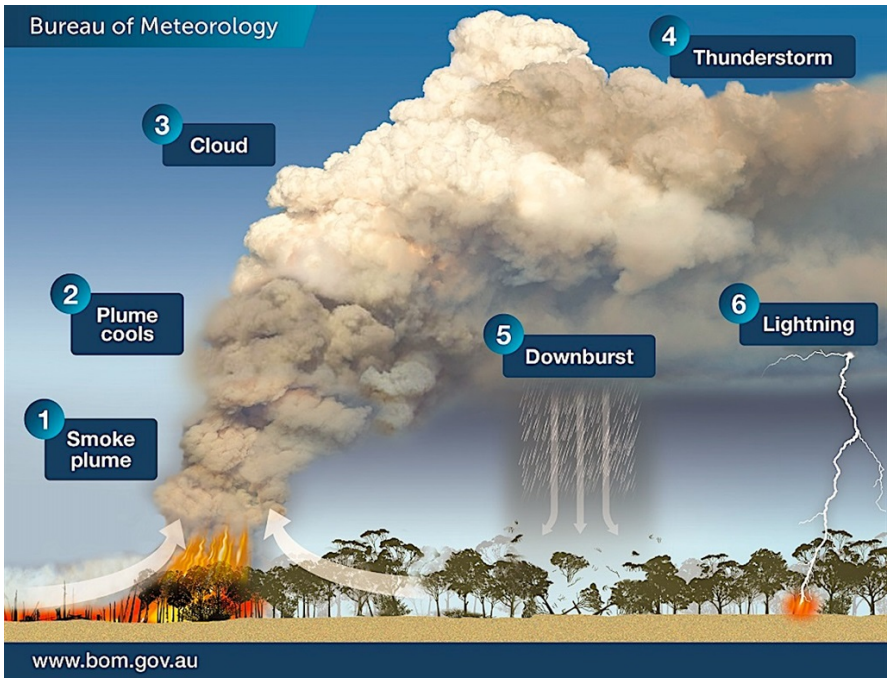
- Jülich will build the **first exascale system**
  - JUPITER: Europe‘s first exascale computer
  - target applications: KI, medicine, material science, physics and climate modeling
  - modular architecture with cluster and booster modules

Motivation

Arguments for adaptive process complexity

- E.g. we do not know how the **threat of extreme wildfires** will change in the future.
  - explosive pyro-convection needs localized description with high process detail
  - global impacts due to long-range smoke transport



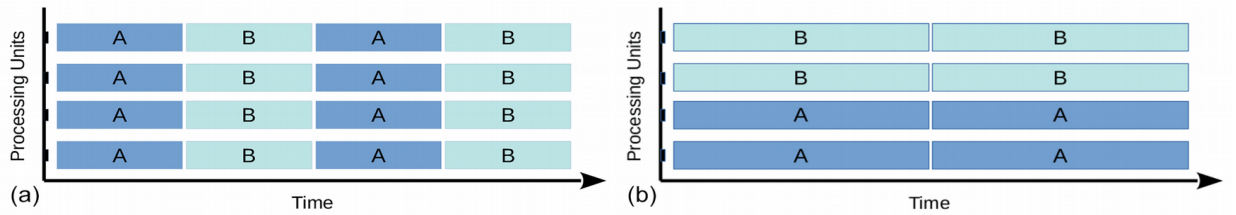


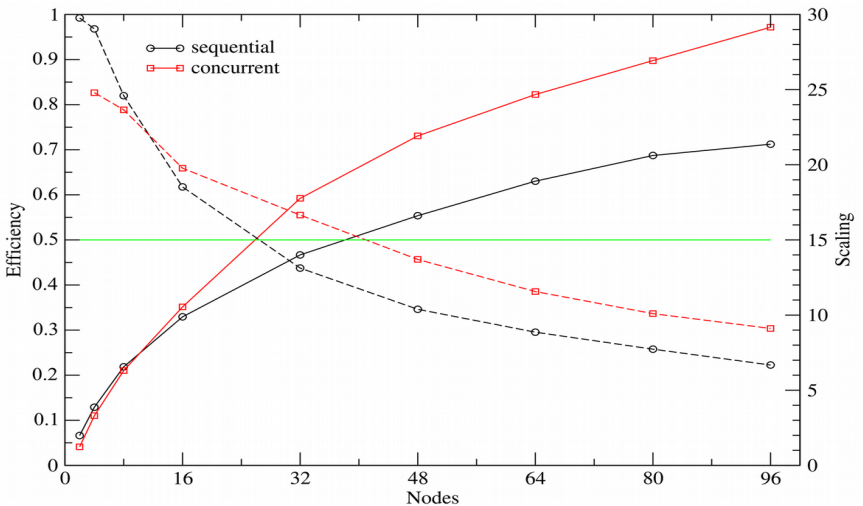
Senf et al. (2023). **How the extreme 2019–2020 Australian wildfires affected global circulation and adjustments**, ACP, https://doi.org/10.5194/acp-23-8939-2023

TROPOS

Concepts

Functional Concurrency





Scaling (solid line) and efficiency (dashed line) for the 160 km sequential and concurrent setups.

- **multi-level and multi-dimensional parallelism**
  - coarse-grained component concurrency as additional parallelism dimension
  - complements methods such as domain decomposition and loop-level shared-memory approaches
  - a function-parallel technique

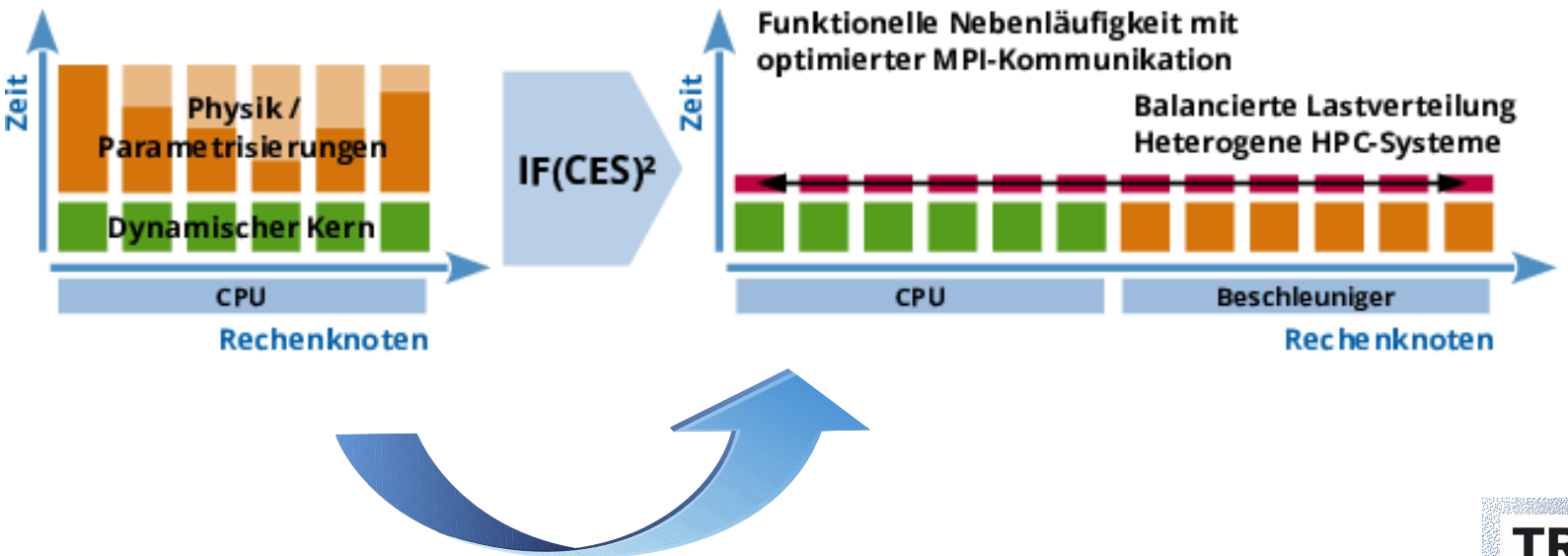
Linardakis et al.: **Improving scalability of Earth system models through coarse-grained component concurrency – a case study with the ICON v2.6.5 modelling system**, GMD, https://doi.org/10.5194/gmd-15-9157-2022, 2022.

TROPOS

Concepts

Functional Concurrency

- **extending** functional concurrency for high-resolution simulation and for different platforms (hardware heterogeneity)
- **invent** new functionalities to deal with inbalances

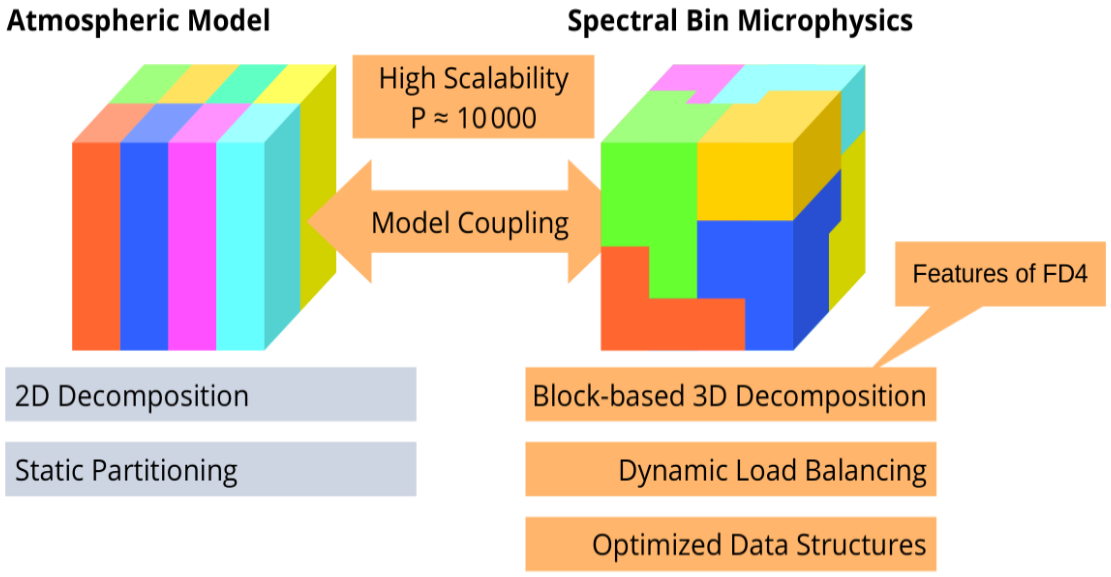


TROPOS

Concepts

Load Balancing

FD4 Concept: Load-balanced Coupling



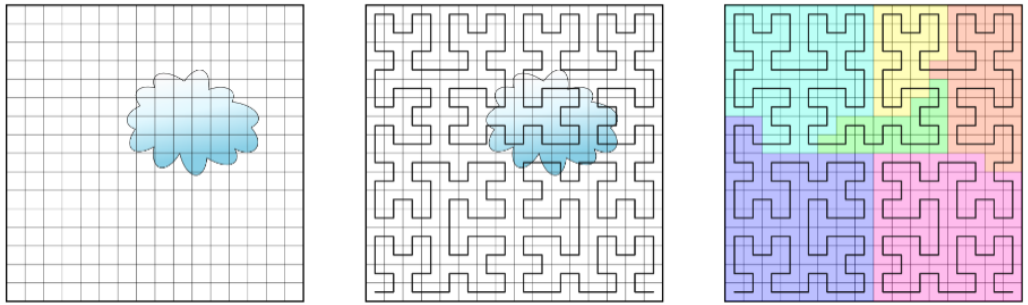
- **flexibility via load balancing**
  - revises grid decomposition based on dynamical load calculations
  - optimized data structures
  - needs smart and efficient communication

M. Lieber, W.E. Nagel, and H. Mix, **Scalability Tuning of the Load Balancing and Coupling Framework FD4**, in "NIC Symposium 2014", NIC Series, Vol. 47, pp. 363-370, 2014 (full paper, full proceedings)

TROPOS

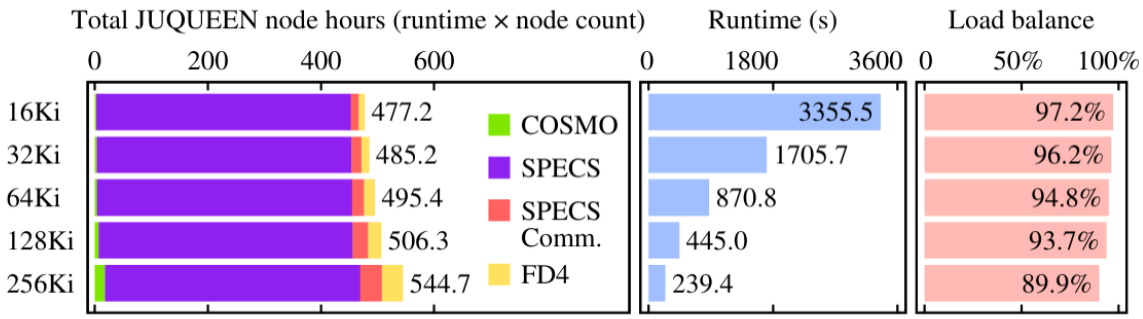
Concepts

Load Balancing



- **space-filling curve (SFC) for partitioning**
  - need to be extended to triangular ICON grid

- **scaling**
  - exceptional performance of predecessor library FD4
  - challenge for new ICON infrastructures



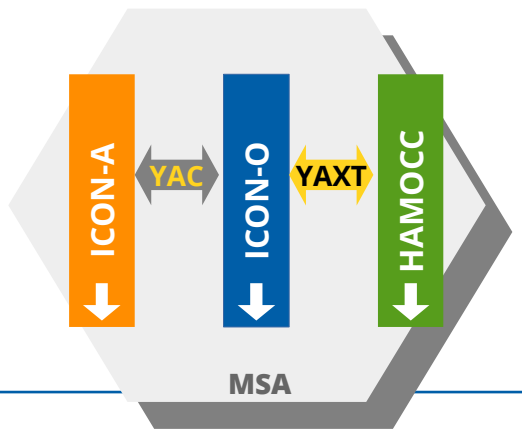
M. Lieber, W.E. Nagel, and H. Mix, **Scalability Tuning of the Load Balancing and Coupling Framework FD4**, in "NIC Symposium 2014", NIC Series, Vol. 47, pp. 363-370, 2014 (full paper, full proceedings)

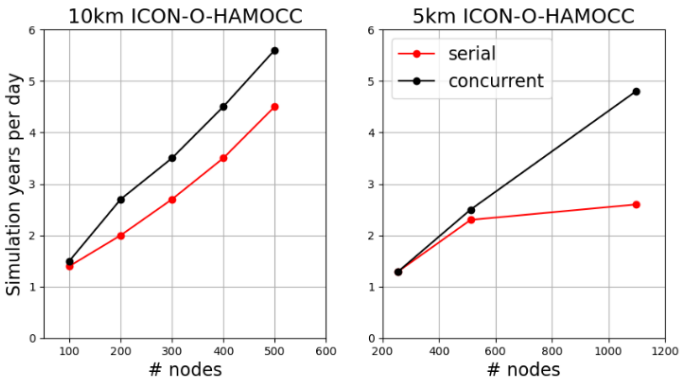
TROPOS

First Successes

Concurrent HAMOCC

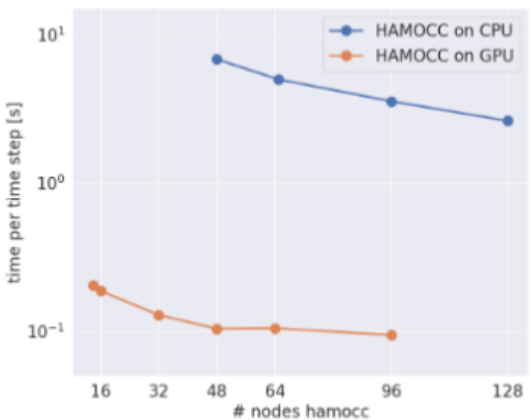
- **functional concurrency**
  - HAMOCC coupled to ICON-O with YAXT
  - across CPU - GPU concurrency





- **concurrency on Levante**
  - esp. beneficial for high resolutions

- **concurrency on JUWELS**
  - porting is promising



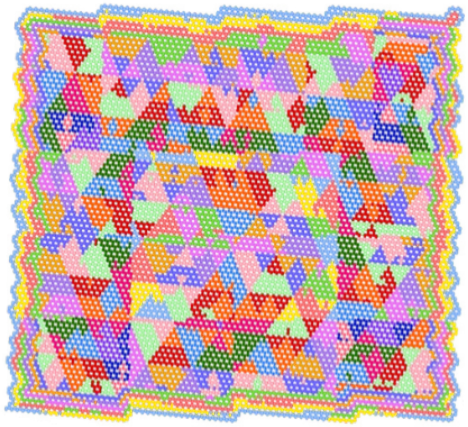
work by Manoel Römmer & Fatemeh Chegini

many thanks to Leonidas Linardakis and to natESM for sprint support

TROPOS

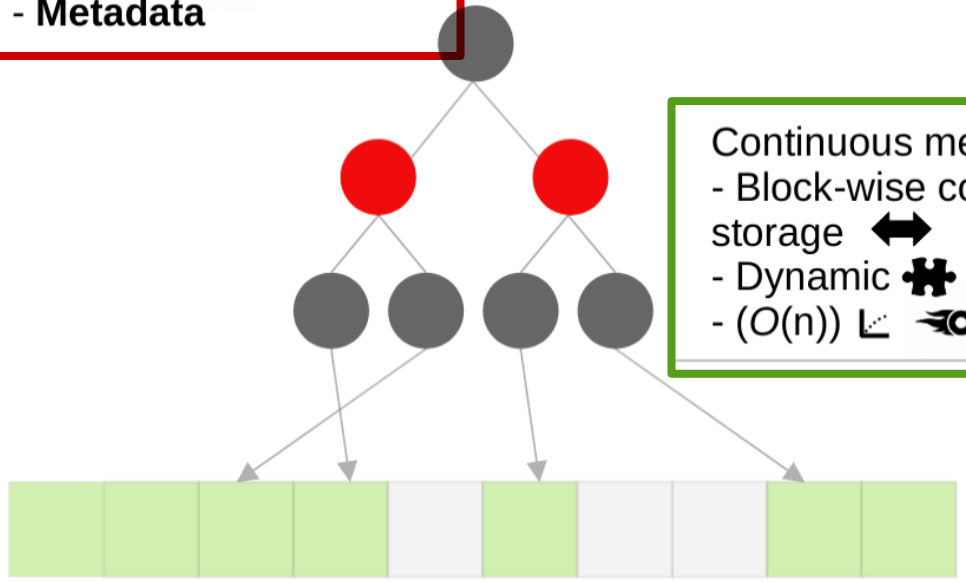
First Successes

LOBSTR



- **explore new flexible data structures for ICON**
  - small blocks
  - vertical decomposition

Red Black Tree
- *O*(log n) search
- Scattered in Memory
- Dynamic
- Iterable
- **Metadata**

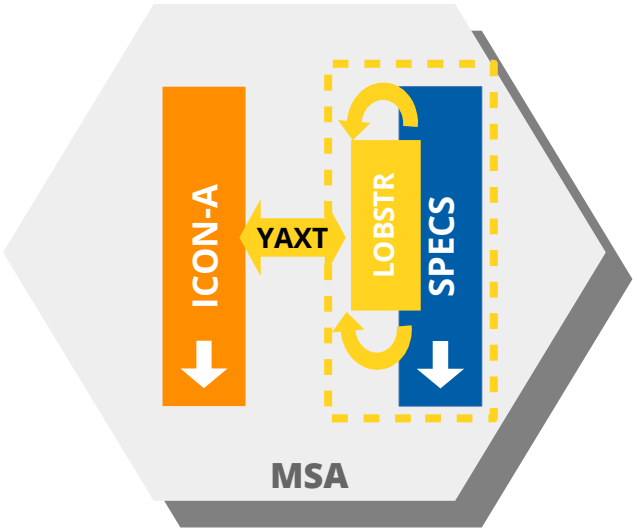


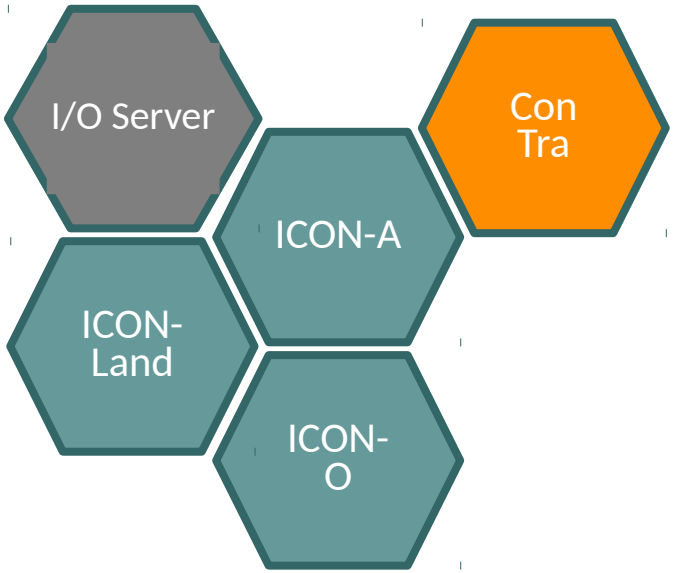
Continuous memory
- Block-wise continuous **Data storage**
- Dynamic
- (*O*(n))

First Successes

ICON ConTra

- **flexible tracer infrastructure**
  - unites concurrency with load balancing





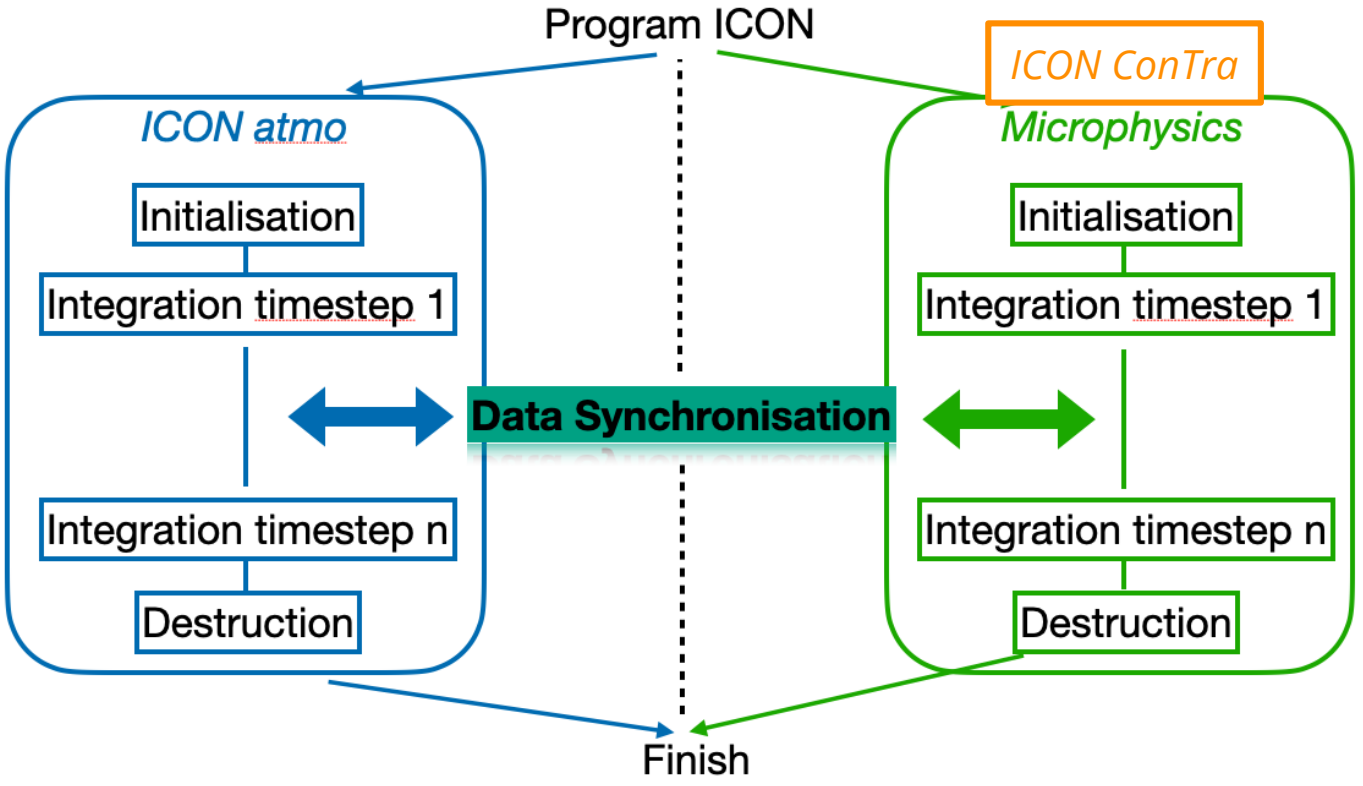
- **ConTra design**
  - treated as separate MPI groups

work by Panos Adamidis, Xingran Wang & Jens Stoll

TROPOS

First Successes

ICON ConTra



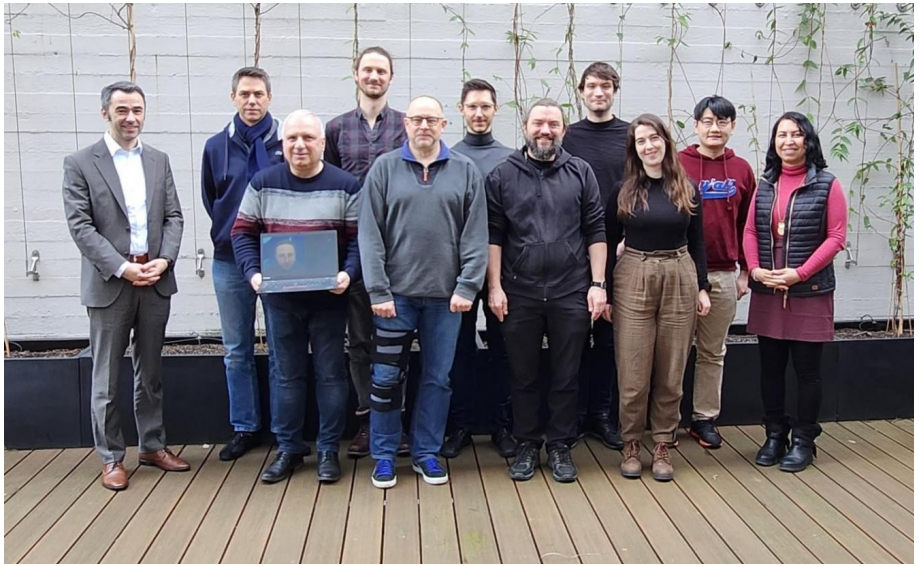
work by Panos Adamidis, Xingran Wang & Jens Stoll

Outlook

- **Concurrency**
  - improve scalability of concurrent HAMMOC at JSC
  - further develop the concurrent tracer abstraction layer **`ConTra`**
- **Load Balancing**
  - finalize new data handling
  - solve vertical decomposition challenge
- **Use Cases**
  - hires hurricane with complex cloud microphysics
  - hires coupled ICON-A/O/HAMOCC

Outlook

- **Concurrency**
  - improve scalability of concurrent HAMMOC at JSC
  - further develop the concurrent tracer abstraction layer **`ConTra`**
- **Load Balancing**
  - finalize new data handling
  - solve vertical decomposition challenge
- **Use Cases**
  - hires hurricane with complex cloud microphysics
  - hires coupled ICON-A/O/HAMOCC



A big thank you to the team!

Interest ? → Get in touch!

TROPOS