Single-precision in Climate Modelling with ICON

Dylan Kieran, DKRZ
December 17, 2025

Overview

- Overview of ICON Model
- Exascale ICON
- Reduced-Precision in ESMs
- Single-Precision in ICON
    - Preliminary Work
    - Initial Performance & Results
    - Development Tools
    - Future Work

(me today)

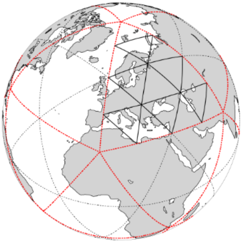
Overview of ICON Model

ICON Model

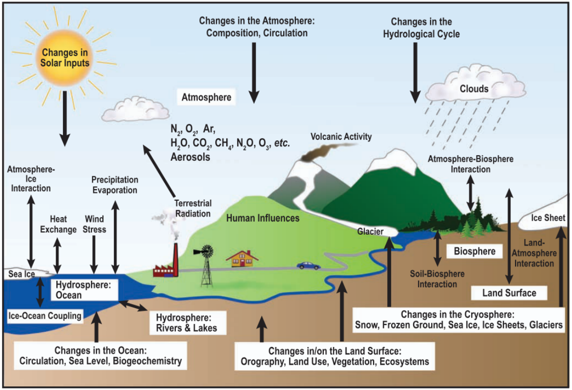
- ICOsahedral Nonhydrostatic (ICON) model
- Development 20+ years
- Collaboration: DWD, MPIM, DKRZ, KIT, MCH/ETHZ/CSCS
- Open source : https://www.icon-model.org

ICON Model

- Fortran, (C++/Kokkos in development)
- $\approx$ 500k lines
- MPI, OpenMP, OpenACC
- CPU, GPU, Vector Engines



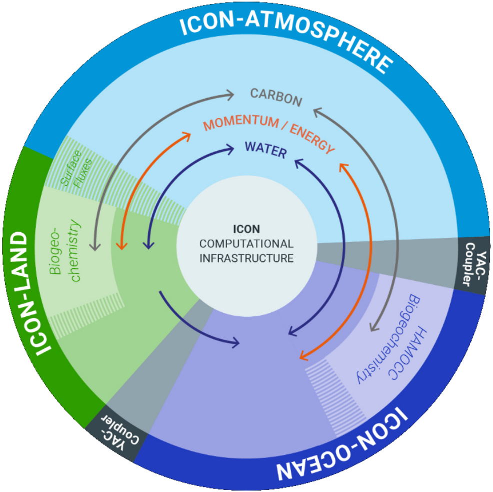
ICON Model
- Processes



PCC AR4 WG1 Ch.1 [1]

ICON Model

- Components



www.icon-model.org

Exascale ICON

Exascale ICON

'Computing the Full Earth System at 1km Resolution' [2]:

- 1.25km global horizontal resolution
- 8TiB main memory (double-precision)
- 20480 GPUs on JUPITER Booster
- Goal: 180 simulated days per day, achieved 145.7
- Model component to architecture mapping
- For reference:
  - 3.36e8 horizontal cells, 90 vertical levels.
  - 20480 Grace Hopper superchips, 5120 of 5884 GPU nodes.

Exascale ICON

'Computing the Full Earth System at 1km Resolution' [2]:

- 1.25km global horizontal resolution
- 8TiB main memory (double-precision)
- 20480 GPUs on JUPITER Booster
- Goal: 180 simulated days per day, achieved 145.7
- Model component to architecture mapping
- For reference:
  - 3.36e8 horizontal cells, 90 vertical levels.
  - 20480 Grace Hopper superchips, 5120 of 5884 GPU nodes.

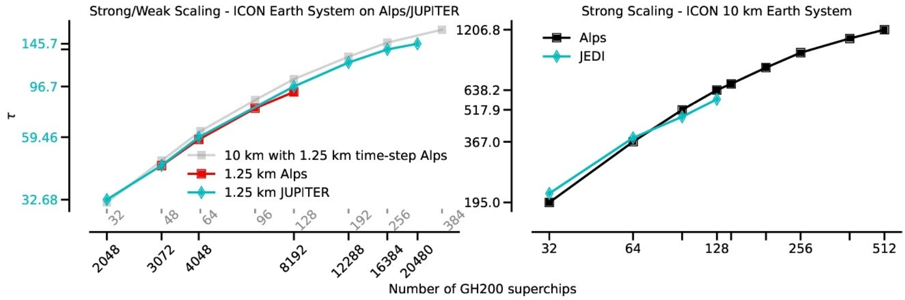
Exascale ICON

'Computing the Full Earth System at 1km Resolution' [2]:

- 1.25km global horizontal resolution
- 8TiB main memory (double-precision)
- 20480 GPUs on JUPITER Booster
- Goal: 180 simulated days per day, achieved 145.7
- Model component to architecture mapping
- For reference:
  - $3.36e8$ horizontal cells, 90 vertical levels.
  - 20480 Grace Hopper superchips, 5120 of 5884 GPU nodes.

Exascale ICON

Scaling



ICON at 1.25km horizontal resolution (left), at 10km horizontal resolution (right) [2]

Reduced-Precision in ESMs

Reduced-Precision in ESMs

Earth System Models (ESMs) utilizing single-precision:

- COSMO - 10 year climate forecasts over European region [3]
- IFS - Medium range weather forecasts [4]
- IFS - 13 month global forecasts [5]
- Met Office Unified Model - single precision iterative solver for Helmholtz equation in dycore [6]
- NICAM - mixed precision [7]
- NEMO - high resolution ocean model in DestinE [8]

Single-Precision in ICON

Single-Precision in ICON

- Why thinking about single precision

Three primary aspects for turning ICON to single precision:

- memory use for data is halved
- bandwidth for data transfer of any kind is halved
- on many architectures a lot of computations can be doubled per cycles

Single-Precision in ICON

- Why thinking about single precision

Three primary aspects for turning ICON to single precision:

- memory use for data is halved
- bandwidth for data transfer of any kind is halved
- on many architectures a lot of computations can be doubled per cycles

Previous work:

- mixed-precision atmospheric dynamical core
- 30-40% dycore speedup (5-10% total speedup) on CPU/GPUs

Preliminary Work

- Precision handling

src/shared/mo_kind.f90

```
  INTEGER, PARAMETER :: sp = SELECTED_REAL_KIND(ps,rs) !< single precision, fp32
  INTEGER, PARAMETER :: dp = SELECTED_REAL_KIND(pd,rd) !< double precision, fp64
  !
  INTEGER, PARAMETER :: wp = dp                        !< selected working precision
#ifdef __MIXED_PRECISION
  INTEGER, PARAMETER :: vp = sp                        !< selected variable precision
#else
  INTEGER, PARAMETER :: vp = dp                        !< selected variable precision
#endif
```

Preliminary Work

- Precision handling

"New" precision handling

src/shared/mo_kind.f90

```
  INTEGER, PARAMETER :: sp = SELECTED_REAL_KIND(ps,rs) !< single precision, fp32
  INTEGER, PARAMETER :: dp = SELECTED_REAL_KIND(pd,rd) !< double precision, fp64
  !
#ifdef __SINGLE_PRECISION
  INTEGER, PARAMETER :: wp = sp                        !< selected working precision
#else
  INTEGER, PARAMETER :: wp = dp                        !< selected working precision
#endif
#ifdef __MIXED_PRECISION
  INTEGER, PARAMETER :: vp = sp                        !< selected variable precision
#else
  INTEGER, PARAMETER :: vp = dp                        !< selected variable precision
#endif
```

- Demo branch: 4 months
- Master branch: 8 additional months

Preliminary Work
- Testing

1. Single-precision builders enabled for every merge request:
   - `dwd_nec_sp`
   - `lumi_cpu_sp`
   - `levante_{gpu_nvhpc,gcc,intel,nag}_sp`

Preliminary Work

- Testing

1. Single-precision builders enabled for every merge request:
   - `dwd_nec_sp`
   - `lumi_cpu_sp`
   - `levante_{gpu_nvhpc,gcc,intel,nag}_sp`
2. Testing:
   - Compilation
   - Infrastructure tests
   - Unit-testing*

Preliminary Work
- Testing

1. Single-precision builders enabled for every merge request:
   - `dwd_nec_sp`
   - `lumi_cpu_sp`
   - `levante_{gpu_nvhpc,gcc,intel,nag}_sp`

2. Testing:
   - Compilation
   - Infrastructure tests
   - Unit-testing*
3. . . . added later . . .
   - Jablonowski-Williamson: Dycore-only with high tolerance threshold
   - Tracer-Hadley: Tracer-only with analytic validation
   - Warm-Bubble: Scientific testcase (smoke-test only)

Initial Performance & Results

Initial Performance & Results

- Atmospheric Dynamical Core

$$\frac{\partial v_{\mathrm{n}}}{\partial t} + \frac{\partial K_{\mathrm{h}}}{\partial n} + (\zeta + f)v_{\mathrm{t}} + w\frac{\partial v_{\mathrm{n}}}{\partial z} = -c_{\mathrm{pd}}\theta_v \frac{\partial \pi}{\partial n} + F(v_{\mathrm{n}}),$$

$$\frac{\partial w}{\partial t} + \mathbf{v_h} \cdot \nabla w + w\frac{\partial w}{\partial z} = -c_{\mathrm{pd}}\theta_v \frac{\partial \pi}{\partial z} - g,$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{v}\rho) = 0,$$

$$\frac{\partial \rho\theta_v}{\partial t} + \nabla \cdot (\mathbf{v}\rho\theta_v) = \tilde{Q}.$$

diabatic processes

adiabatic fluid dynamics

dynamical core

parameterizations

Note: The conservation equation for tracer mass is neglected, here. It can also contain sources from parameterizations.

(Zängl et al., QJRMS, 2014)

Initial Performance & Results

- Dynamical Core

R2B6 (40km) simulation on Levante AMD 7763 CPU node.

total_avg(s)	total	integrate_nh	model_init	speedup (total time)
cpu, intel, dp	25.971	25.404	4.651	
cpu, intel, mp	18.520	17.953	5.500	1.40
cpu, intel, sp	12.687	12.420	3.477	2.05

R2B6 (40km) simulation on Levante A100 GPU node, using only 1 GPU.

total_avg(s)	total	integrate_nh	model_init	speedup (total time)
gpu, nvhpc, dp	8.124	7.875	163.420?	
gpu, nvhpc, sp	5.254	5.141	121.836?	1.55

Initial Performance & Results

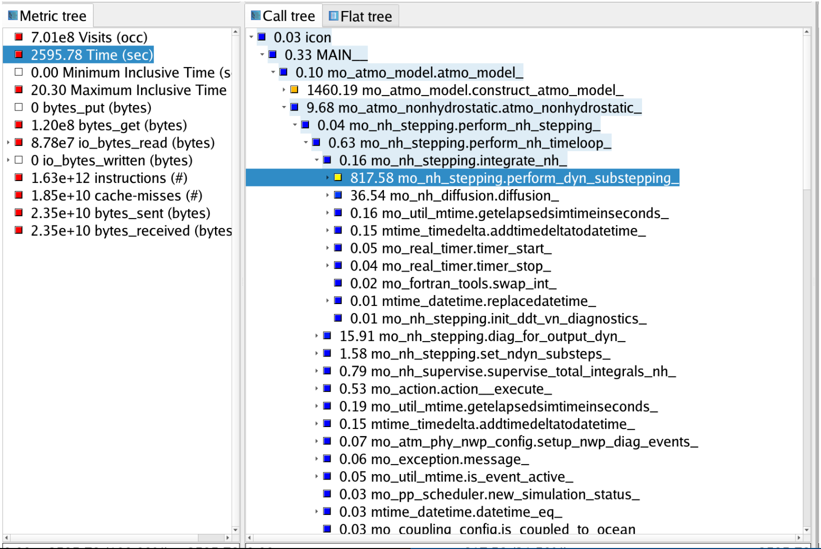
- Dynamical Core

R2B8 (10km) simulation on 8 GPU nodes for double-precision and 4 GPU nodes for single-precision.

total_avg(s)	total	integrate_nh	model_init	efficiency
gpu, nvhpc, dp	4.882	4.750	109.479?	
gpu, nvhpc, sp	6.443	6.332	152.366?	1.52

Initial Performance & Results
- Dynamical Core

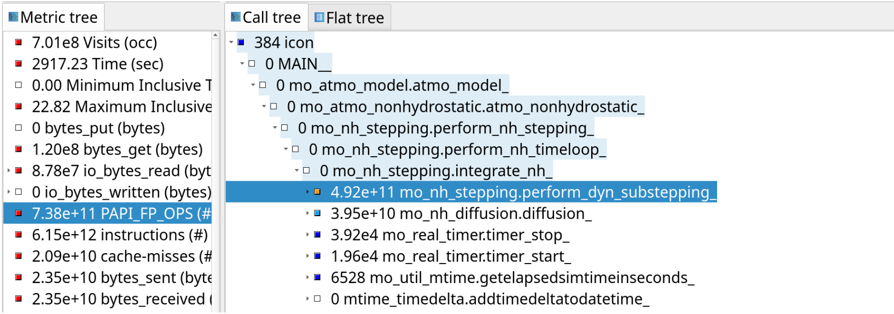
A brief look at CPU performance metrics:



Initial Performance & Results

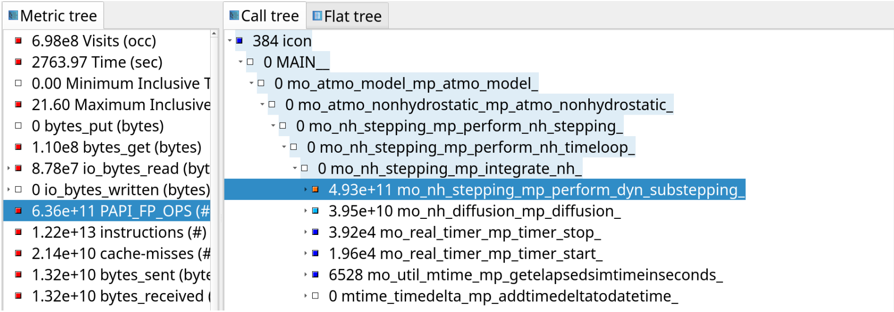
- Dynamical Core

Double-precision floating point operations:



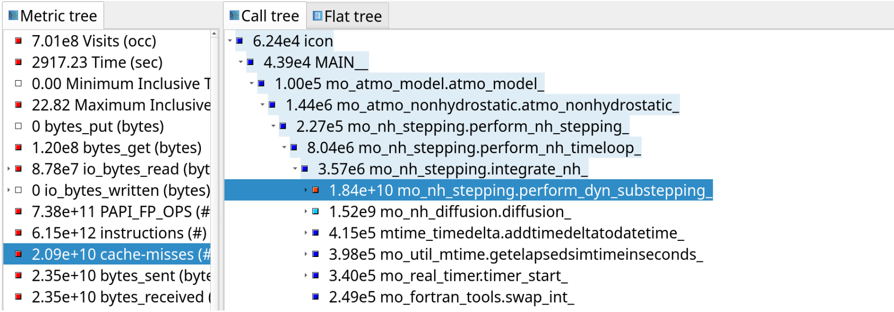
Initial Performance & Results
- Dynamical Core

Single-precision floating point operations:

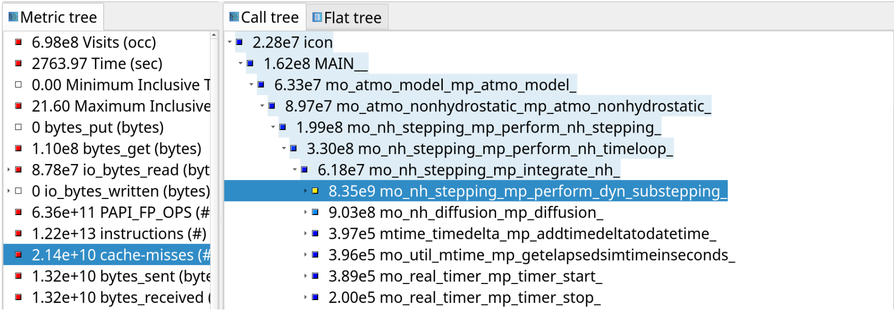


Initial Performance & Results
- Dynamical Core

Double-precision cache misses:

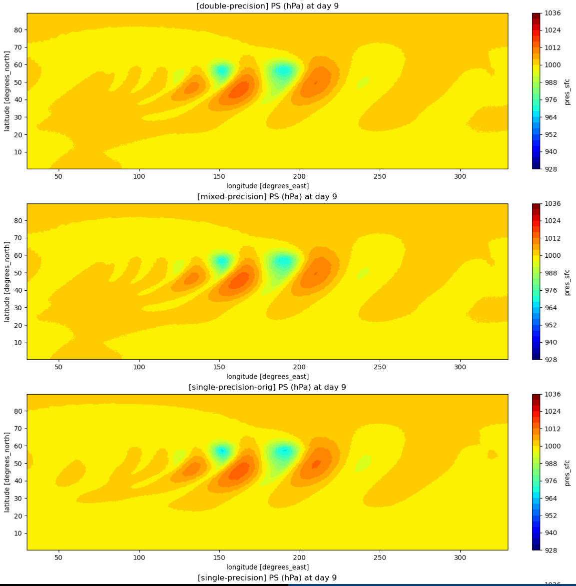


Single-precision cache misses:



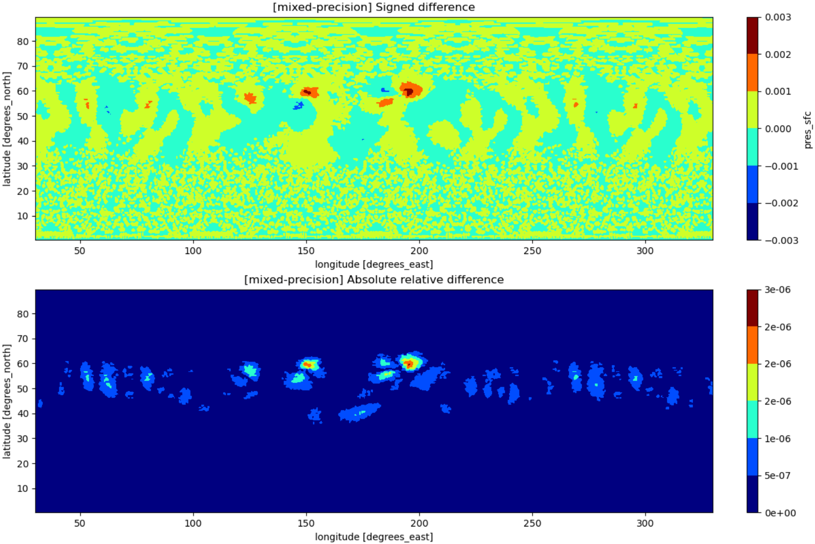
Initial Performance & Results

- Dynamical Core



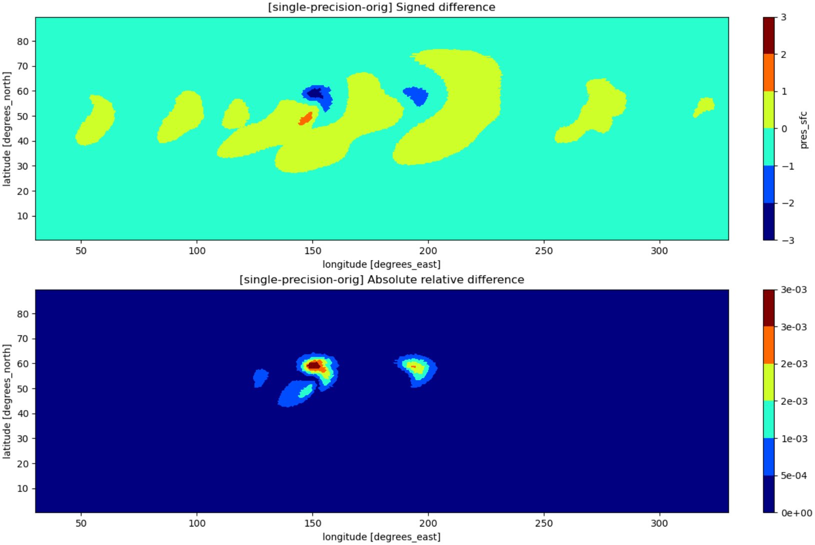
Initial Performance & Results

- Dynamical Core



Initial Performance & Results

- Dynamical Core



Development Tools

Development Tools

- Verrou

Verrou is an open-source (github link) tool for analyzing errors in floating-point arithmetic.

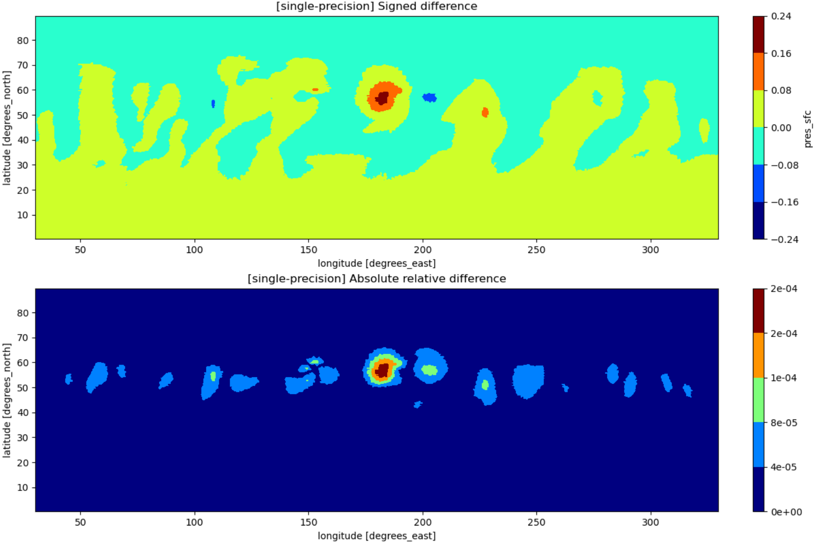
- Valgrind extension
- Rounding options: IEEE, stochastic, single-precision emulation
- Debugging tools for code bisection `verrou_dd`.

Similar tools: verificarlo, AutoRPE (later slide), ...

Initial Performance & Results

- Return to Dynamical Core

Single precision **but** with Exner pressure fields in double-precision:



Development Tools

- AutoRPE

AutoRPE (Automatic Reduced Precision Emulator) tool for optimizing numerical precision in Fortran code.

- Developed by Barcelona SuperComputing BSC
- Link to documentation.
- Used for development of single-precision NEMO ocean model.
- Provides additional debug tools similar to verrou (eg for bisection).

Initial Performance & Results

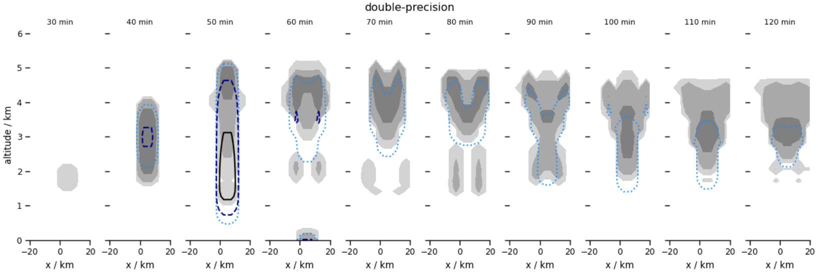
- AES Bubble

More info on the bubble experiment and plots here.

Plot on next slide is visualizing the condensate for the saturated bubble.

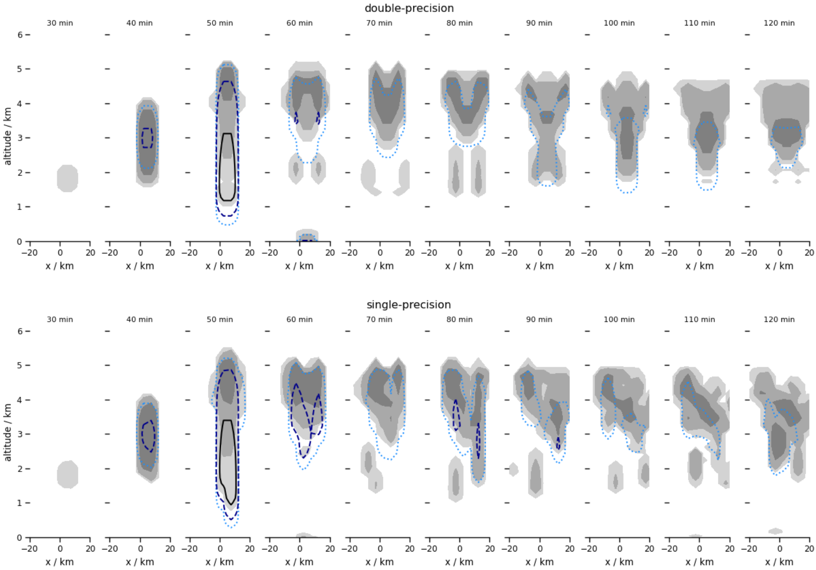
Initial Performance & Results

- AES Bubble



Initial Performance & Results

- AES Bubble



Future Work

Future Work

- Further AMD GPU development for `Alice Recoque` ?
- ...
- Tuning reduced precision dycore
- Sandbox experiments for asymetry behaviour
- New algorithm for vector angle calculations
- Technical and scientific expertise
- Robust testing (eg tolerance-based instead of bit-identical)

Supporting Slides

Supporting Slides

Memory usage on GPU using `run_nml::msg_level>=15`:

levante_gpu_nvhpc aes_amip_test r2b4
0: GPU mem usage: used GB 0.185432E+02 free GB 0.607076E+02
0: used GB 0.185432E+02 free GB 0.607076E+02

levante_gpu_nvhpc_sp aes_amip_test r2b4
0: GPU mem usage: used GB 0.993567E+01 free GB 0.693151E+02
0: used GB 0.993567E+01 free GB 0.693151E+02

Single-Precision in ICON

- Incremental single conversion

Why convert incrementally:

- allows for testing
- allows for parallel evaluation of numerical schemes
- helps in encountering and implementation of algorithmic required mixed precision

Soft requirements to convert incrementally:

- modular design (with clean interfaces)
- modular testing

. . . Which we don't have . . .

Development Tools

- Verrou: First Results

Verrou bisection report:

rddmin

```
ddmin0: Fail Ratio: 100.00%     Fail indexes: 0
    mo_solve_nonhydro.f90:566 (__mo_solve_nonhydro_MOD_solve_nh)
```

The culprit code related to Exner pressure.

mo_solve_nonhydro.f90:566

```
! non-extrapolated perturbation Exner pressure, saved in
    exner_pr for the next time step
DO jk = 1, nlev
    ...
    DO jk = 1, nlev
        DO jc = i_startidx, i_endidx
            exner_pr(jc,jk,jb) = exner(jc,jk,jb) -
                exner_ref_mc(jc,jk,jb)
        ENDDO
    ENDDO
ENDDO
```

Bibliography I

[1] Susan Solomon. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*. Vol. 4. Cambridge university press, 2007.

[2] Daniel Klocke et al. "Computing the Full Earth System at 1km Resolution". In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. SC '25. New York, NY, USA: Association for Computing Machinery, 2025, 125–136. ISBN: 9798400714665. DOI: 10.1145/3712285.3771789. URL: https://doi.org/10.1145/3712285.3771789.

[3] Hugo Banderier et al. "Reduced floating-point precision in regional climate simulations: an ensemble-based statistical verification". In: *Geoscientific Model Development* 17.14 (2024), pp. 5573–5586.

[4] Filip Váňa et al. "Single precision in weather forecasting models: An evaluation with the IFS". In: *Monthly Weather Review* 145.2 (2017), pp. 495–502.

Bibliography II

[5] Christoph Schär et al. "Kilometer-scale climate models: Prospects and challenges". In: *Bulletin of the American Meteorological Society* 101.5 (2020), E567–E587.

[6] Christopher M Maynard and David N Walters. "Mixed-precision arithmetic in the ENDGame dynamical core of the Unified Model, a numerical weather prediction and climate model code". In: *Computer Physics Communications* 244 (2019), pp. 69–75.

[7] Masuo Nakano et al. "Single precision in the dynamical core of a nonhydrostatic global atmospheric model: Evaluation using a baroclinic wave test case". In: *Monthly Weather Review* 146.2 (2018), pp. 409–416.

Bibliography III

[8] Stella V. Paronuzzi-Ticco, Gladys Utrera, and Mario C. Acosta. "Efficient variable precision reduction in chaotic climate models: Analysis of the NEMO case in the destination earth project". In: *Computers & Geosciences* 205 (2025), p. 105989. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2025.105989. URL: https://www.sciencedirect.com/science/article/pii/S0098300425001396.