Sandia National Laboratories

SAND2020-11083C

A Performance-Portable Nonhydrostatic Atmospheric Dycore for the Energy Exascale Earth System Model Running at Cloud-Resolving Resolutions

L.Bertagna[1], O.Guba[1], M.Taylor[1], J.Foucar[1], J.Larkin[2], A.Bradley[1], S.Rajamanickam[1], A.Salinger[1]

[1]Sandia National Laboratories, Albuquerque, NM
[2]Oak Ridge National Laboratory, Oak Ridge, TN

SC20, Nov 9-19th, 2020

SAND 2020-10220 C

a National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly ubsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA00035

Energy **E**xascale **E**arth **S**ystem **M**odel

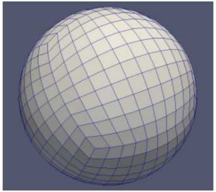
Sandia National Laboratories

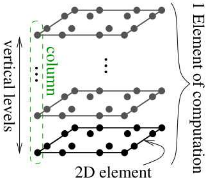
What is E3SM?

- DOE effort for a high resolution earth model.
- Branched from **C**ommunity **E**arth **S**ystem **M**odel (CESM) in 2014.
- Modular library, with several components: atmosphere dynamics/physics, land, land-ice, ocean, sea-ice, biogeochemistry, ...
- All component can run with variable-resolution, unstructured grids.
- Mostly written in Fortran 90.
- Broad variety of time and space scales.
- 2018: E3SM version 1 is released in April.
- 2021: E3SM version 2 code freeze (estimate: June).

High-Order Methods Modeling Environment

- Component of E3SM (and CESM) for dynamics and transport in the atmosphere.
- Accounts for 20-25% of total run time of typical fully-coupled simulation.
- Highly optimized for MPI and OpenMP parallelism.
- Horizontal (2D) and vertical (1D) differential operators are decoupled.
- Spectral Element Method (SEM) in the horizontal direction.
- Eulerian or Lagrangian schemes for vertical operators.





The Kokkos library

What is Kokkos?

- Developed at Sandia National Labs, written in C++ (with C++11 required).
- Provides templated constructs for on-node parallel execution: execution space (host vs device), execution policy (range vs team), parallel operation (for, scan, reduce).
- Provides template abstraction for a multidimensional array: data type, memory space (host, device, UVM), layout (left, right, ...), memory access/handling (atomic, unmanaged, ...).
- Supports several back-ends: Serial, OpenMP, Cuda, HIP, ...
- Available at `http://github.com/kokkos/kokkos`.

Simple Cloud-Resolving E3SM Atmosphere Model

Sandia National Laboratories

Cloud-Resolving Models (CRMs):

- Have horizontal resolution of $\lesssim$3km.
- Can do away with parametrizations of some physical process (deep convection), but needs more complex models for state equations.

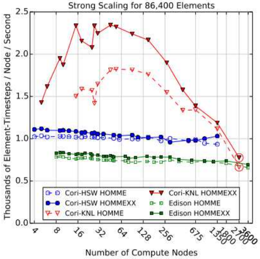
SCREAM:

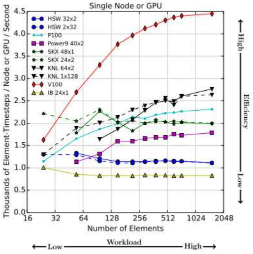
- Atm model for E3SM at ultra-high resolution.
- Targets 3km horizontal resolution, and 128 vertical levels (7.2 billion grid points).

HOMMEXX: hydrostatic dycore (preqx)

Task completed in 2018[1].

- Incremental F90 → C++ conversion.
- Heavily tested.
- Bit-for-bit with F90 implementation.
- Minimization of architecture-specific code.
- Primary design goals:
  - expose parallelism,
  - maximize vectorization,
  - minimize memory movement.





L. Bertagna et al., HOMMEXX 1.0: a performance-portable atmospheric dynamical core for the Energy Exascale Earth System Model, Geo. Model Dev., 2019

HOMMEXX: non-hydrostatic dycore (theta-l)

Same design goals as hydrostatic dycore:

- Expose parallelism and vectorization, minimize memory movement.
- Bit-for-bit with F90 implementation.
- Minimization of architecture-specific code.

With respect to hydrostatic dycore:

- Model differences: adds two state variables, uses potential temperature, can run in both hydrostatic and non-hydrostatic mode.
- Main additional challenges: nonlinear solver, larger memory footprint.

Results: tested architectures

Single node architectures specs:

(KNL) Intel Xeon Phi: 68 cores/node, 4 threads/core, HBM+DDR4

(P9) IBM Power9: 2 sockets/node, 10 cores/socket, 4 threads/core, DDR4

(V100) NVidia Volta: 2 sockets/node, 3 GPUs/socket, 2560 DP cores/GPU

Full cluster specs:

(KNL) Cori: 9688 compute nodes (max used: 9216), located at NERSC.

(P9) Summit: 4608 compute nodes (max used: 4600), located at ORNL.

(V100) Summit: 27648 GPUs (max used: 27600), located at ORNL.

Results: strong scaling

Sandia National Laboratories

- Solve for state and 10 tracers (NGPPS benchmark).
- Run at 3km and 1km horizontal resolution.
- Achieves 0.97 SYPD on GPU on full Summit system.
- Excellent scaling at 3km, perfect scaling at 1km.
- CPU performance comparable (or slightly better) to original F90.
- Approximately 10x speedup when using GPUs vs P9 on Summit.
- 1km resolution features approximately 1 trillion DOFs.

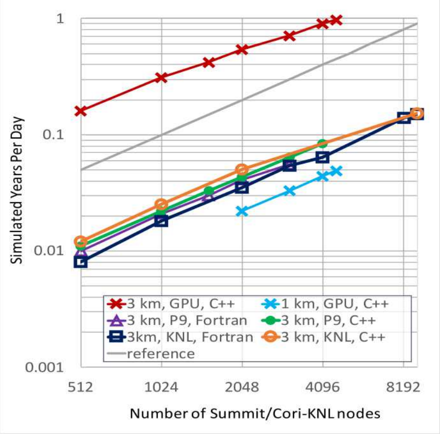


Figure: Achieved SYPD for different implementations and resolutions.

Results: nonlinear solver

- Homme Timestepping: Horizontally Explicit, Vertically Implicit (HEVI).
- Vertical Implicit: Diagonally Implicit Runge Kutta (DIRK).
- Solves one nonlinear eqn per GLL point (coupling vertical levels).
- Newton step requires tridiag solve.
- Strictly diagonally dominant matrix.
- No pivoting allows packing equations.
- Uses Thomas on CPU, Cyclic Reduction on GPU.
- Independent residual check for each column.
- Only two kernels per solve (initial guess + Newton).

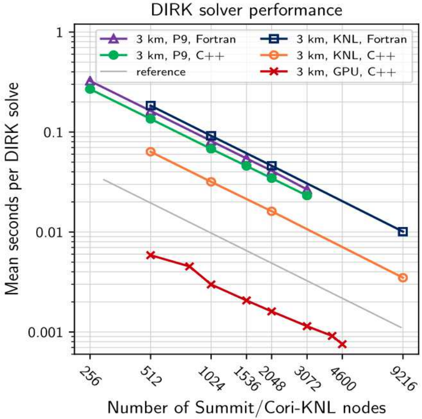


Figure: Performance of the DIRK solver on Cori KNL, Summit Power9, and Summit V100 GPU, for the 3 km benchmark.

Results: workspace manager

- Allows temporary buffers management within single kernel.
- Simplifies code development in temporary-heavy kernels.
- Only active on GPU builds, at high workload/node.
- Limited overhead when WSM is active, negligible if inactive.
- Necessary to fit 1km problem in memory for less than 4096 nodes on Summit.

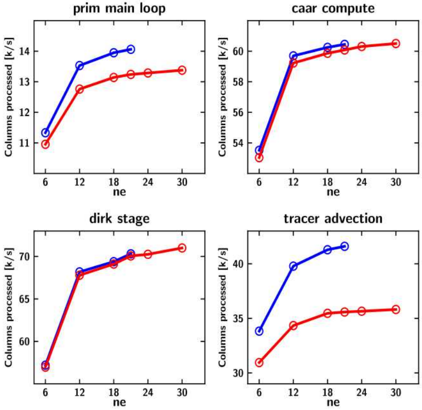


Figure: Processing rate of key functors as well as main time loop timer for the workspace manager (WSM) enabled (red) and disabled (blue).

Conclusions

Sandia National Laboratories

- Completed C++/Kokkos rewrite of HOMME non-hydrostatic dycore.
- New implementation faster than original on CPU.
- New implementation achieves 0.97 SYPD at 3km resolution on GPU on full Summit system.
- Non-hydrostatic dycore required efficient nonlinear solver and tempoary buffers management.