



EarthWorks: The computational and engineering challenges faced when building a global storm-resolving resolution modeling system

Sheri Mickelson¹

PIs: David Randall², Jim Hurrell²

Richard Loft³, Thomas Hauser¹, Michael Duda¹, Dylan Dickerson¹, Supreeth Suresh¹, Jian Sun¹,
Chris Fisher¹, Donald Dazlich², Gunther Huebler⁴, Raghu Raj Kumar⁵, Pranay Reddy Kommera⁵

1 National Center for Atmospheric Research

2 Colorado State University

3 AreandDee, LLC

4 University of Wisconsin, Milwaukee

5 NVIDIA Corporation



COLORADO STATE UNIVERSITY



AreandDee LLC

Come scale away...



EarthWorks:

Five-year project led by CSU, with participation by 3 NCAR labs and NVIDIA.

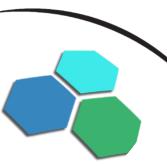
Funded by NSF CSSI.

Science Goals

- Begin to resolve storms at ~4-km grid.
- Eliminate deep convection or gravity-wave drag parameterizations.
- Include a resolved stratosphere.
- Enable new science (extreme events!) for both weather and climate.
- Provide a critical capability to the climate community for guiding adaptation at global, regional and local levels.

Computational Goals

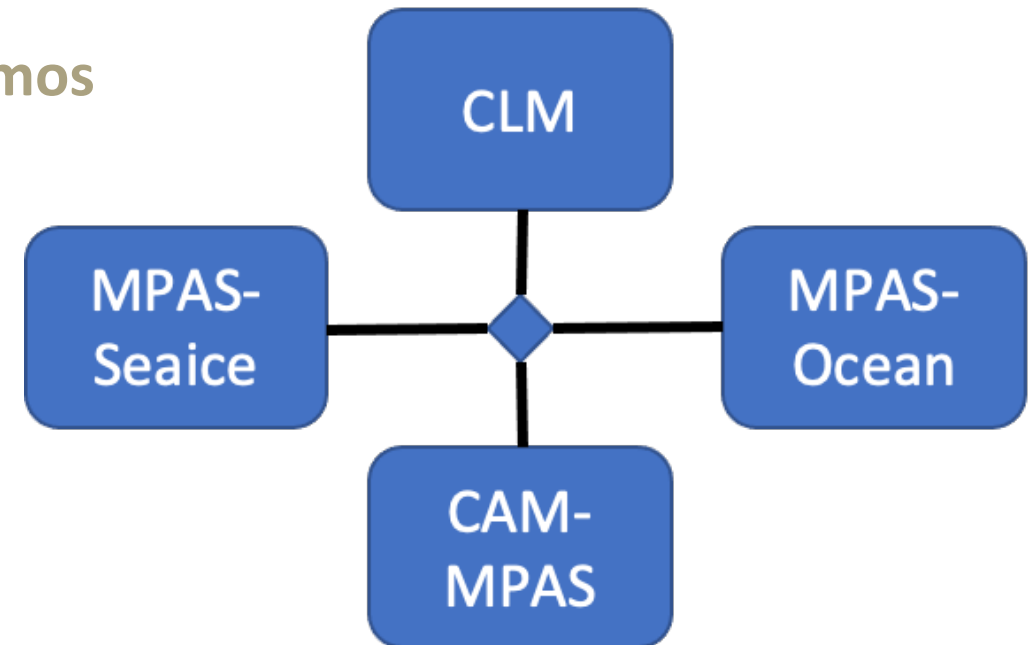
- The EarthWorks ESM will run on CPUs for low resolution experiments and for testing short ultra-high resolution setups.
- Provide a fully GPU-enabled implementation of ocean and atmosphere for tackling high resolutions.
- EarthWorks will put huge demands on computational and data systems. Thus the project incorporates infrastructure development efforts for both big data and machine learning inference.

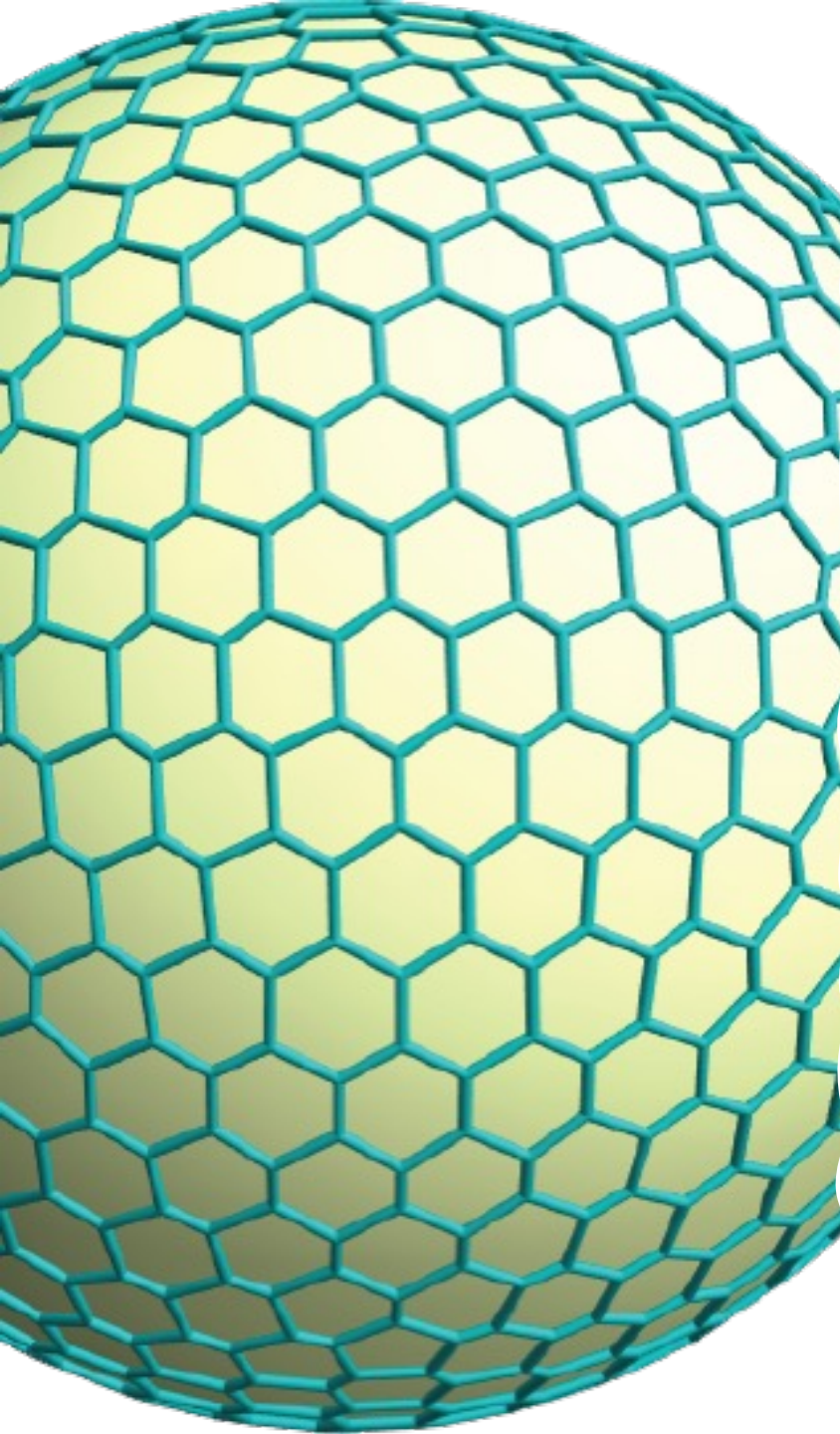


Infrastructure targets for EarthWorks

Based on the Community Earth System Model (CESM), but it is NOT CESM

- The MPAS non-hydrostatic dynamical core, with a resolved stratosphere and CAM physics
- The MPAS ocean model, developed at Los Alamos
- The MPAS sea ice model, based on CICE
- The Community Land Model (CLM)
- The Community Mediator for Earth Prediction Systems (CMEPS)





Infrastructure targets for EarthWorks

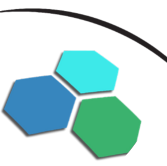


- EarthWorks uses the same *geodesic mesh and cell spacing* for all components.
- A mesh spacing of 3.75 km works well for the atmosphere, the ocean, and the land surface.
- Using a single mesh reduces the operation count, message-passing overhead, and memory requirements.



Supported model configurations

- **Resolutions (120 km, 60 km, 30 km, 15km, 7.5km, 3.75km)**
- **Five configurations with MPAS-dynamical core in CAM**
 - Idealized Held Suarez climate
 - Moist baroclinic wave + KESSLER microphysics
 - Aquaplanet - full up atmosphere with data ocean, no land
 - CAM6-MPAS atmosphere with CTSM (land) + data ocean
 - CAM-6 MPAS atmosphere + CTSM + MPAS-Ocean + MPAS Sea-Ice












Performance goals

Our 2025 performance goals for a version of EarthWorks with 3.75 km global grid spacing are

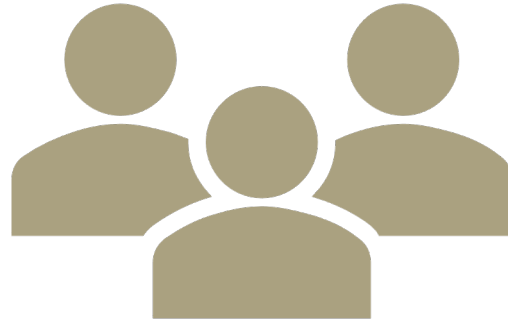
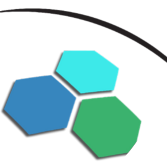
- Half a simulated year per day in atmosphere-only simulations with a resolved stratosphere
- One simulated year per day in coupled simulations with fewer stratospheric layers.



Goal: End-to-end workflow portability

Objective	Tools
Revision Control	Github 
Containers for portability	Singularity and Docker  
Performance portability	OpenACC, OpenMP, OpenMPI   
Scalable I/O	PIO
Analysis	Atmospheric Diagnostic Framework and Raijin 
Data Transfer	Globus 
Science Gateway	Containerized Gateway 

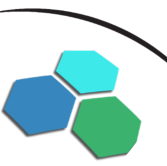




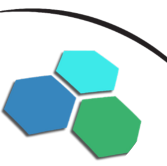
How to work with a diverse group?
Form a community
&
Create a development plan



Coordinating the effort among the software engineers



-
- Get everyone in the same room and talking to one another
 - All voices are equally important
 - Celebrate all successes
 - Create a clear vision and path
 - Everyone should understand how their contributions fit into the big picture
 - Remove any barriers to entry
 - Empower all team members



Coordinating the effort with scientists

-the moving target problem-



- The scientists have to be equally invested in the project
- The software engineers have to be aware of the science planning
- The scientists need to be involved with the software engineering planning
- How will the software be maintained after the work has completed?

EarthWorks software roadmap

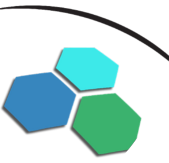


Version (Target Delivery Date)	Deliverables
Version 1.0 (Released March 2023)	120,60,30 km CPU Intel/GNU compilers AMD/Intel based systems
Version 2.0 (November 2023)	15 km resolution nvhpc compiler support GPU offload for dycore and physics ¹
Version 3.0 (March 2024)	7.5 km resolution Scalable diag tools v1 release GPU offload for physics ¹ MPAS-Ocean GPU offload
Version 4.0 (July 2024)	3.75 km resolution Scalable diag tools v2 release

•¹physics includes {RRTMGP, CLUBB, and uphysics}

Where are
we now?

We're
working
along four
fronts



Running short ultra-high resolution
simulations to flush out potential issues

Porting the atmospheric model to run on
GPUs

Maintaining the repository and making
sure it's in sync with CESM

Developing workflows for data
visualization and analysis on the native
MPAS grid and at scale



Running short ultra-high resolution simulations to flush out potential issues

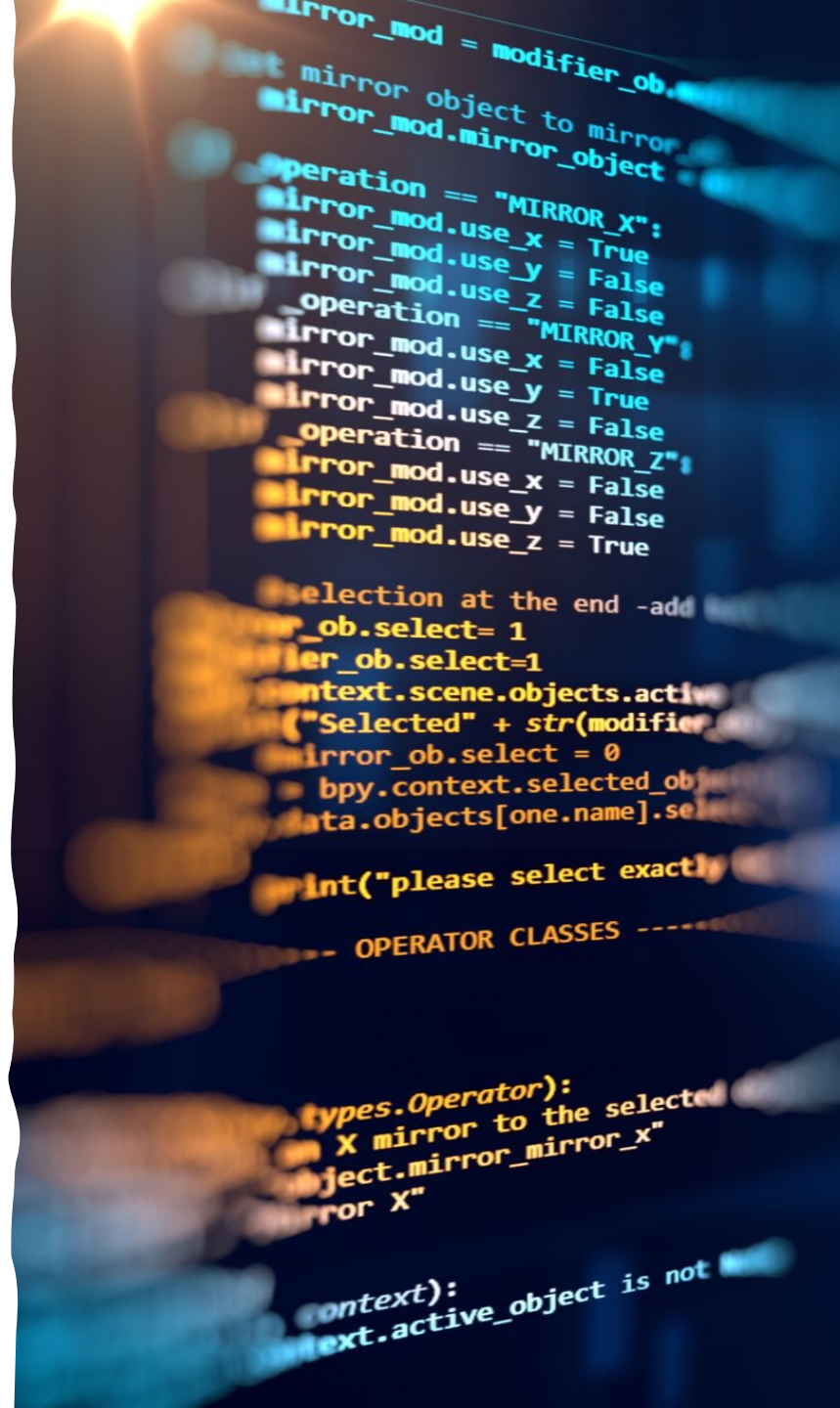
Cross Institutional Collaboration
EarthWorks Team, NCAR, TACC,
ESMF Core Team, AER

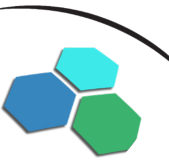
Issues encountered (and fixed!)

- **Initialization (abnormally long times, high memory use)**
 - Traced to an issue in the ESMF framework, resulted in a patch release.
 - **Impact: 5.7x speedup, 2x reduction in memory use in initialization.**
- **100x slowdown in history I/O bandwidth**
 - Traced to the ROMIO MPI-IO implementation in PnetCDF, resulted in a problem report and workaround.
 - **Impact: expected history I/O performance restored**
- **Run after restart errors**
 - Traced to an issue with the PIO2 (parallel I/O) infrastructure in CESM, resulted in a patch release.
 - **Impact: correct model restarts restored**

Our approach to porting to GPUs

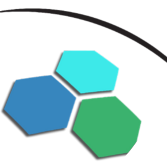
- We are porting using the OpenACC directive approach
 - We see the best performance with OpenACC
 - All of our code is written in Fortran and porting to a new language is not currently an option
- We have worked with Intel on their OpenACC to OpenMP offload conversion tool and we can pivot to OpenMP offload if necessary
 - Intel(r) Application Migration Tool written by Harald Servat
<https://github.com/intel/intel-application-migration-tool-for-openacc-to-openmp>





EarthWorks porting efforts onto GPUs

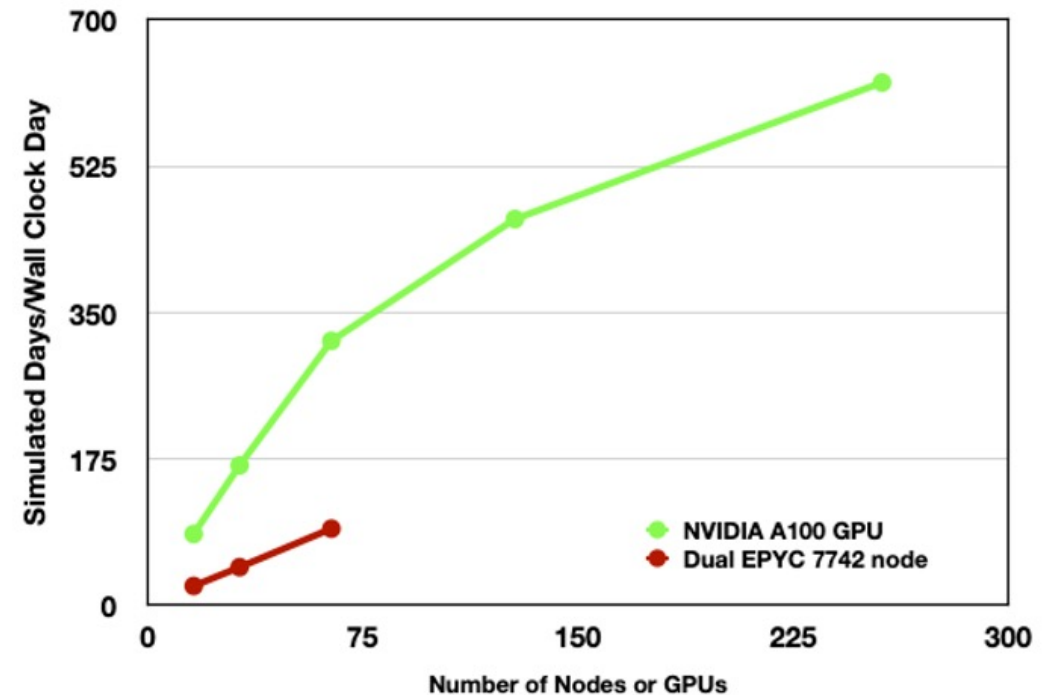
Component	Subcomponent	Package	GPU Porting Status	Offload Paradigm	Version Skew?	Optimized?	Status
Atmosphere			in progress			In progress	
	Dycore	MPAS-7.x	completed	OpenACC	merge with MPAS-7 code base.	yes	Awaiting results of upgraded MPAS-7 version scaling tests.
	Physics	CAM-6	in progress			In progress	
		PUMAS	completed	OpenACC & OpenMP (in separate repository)	Sync'd with CAM physics changes	yes	Add new μ processes: e.g. graupel.
		RRTMGP	completed	OpenACC and OpenMP		Deferred until validated in multi-month tests	
		CLUBB	completed	OpenACC & OpenMP	Merging changes for hi-res cloud physics may be required.	In progress	
Ocean	MPAS-O		completed	OpenACC	Yes (E3SM 2.0 vs 2.1)	yes	Need to test GPU version and merge with CPU version.
Sea-Ice	MPAS-SI		deferred	OpenACC	TBD	no	Investigating feasibility of GPU port



MPAS dycore performance

- **Experiment:** MPAS-7 (5.9M cell mesh; 56 levels; FP32) ran dry baroclinic test case for 10 simulated days
- **Equipment:** Selene supercomputer; nodes = AMD Dual socket EPYC 7742 “Rome” CPUs with 8x NVIDIA A100 GPUs; 10 HDR links/node.
- **Resources:** Benchmark of 128-core ROME CPU node vs A100 GPU
- **Takeaways:**
- **Early scaling looks impressive - and 3.5x faster than CPU node.**
- **Slowdown of MPAS-7 compute (m) was recently isolated to **not declaring new variables GPU resident.****

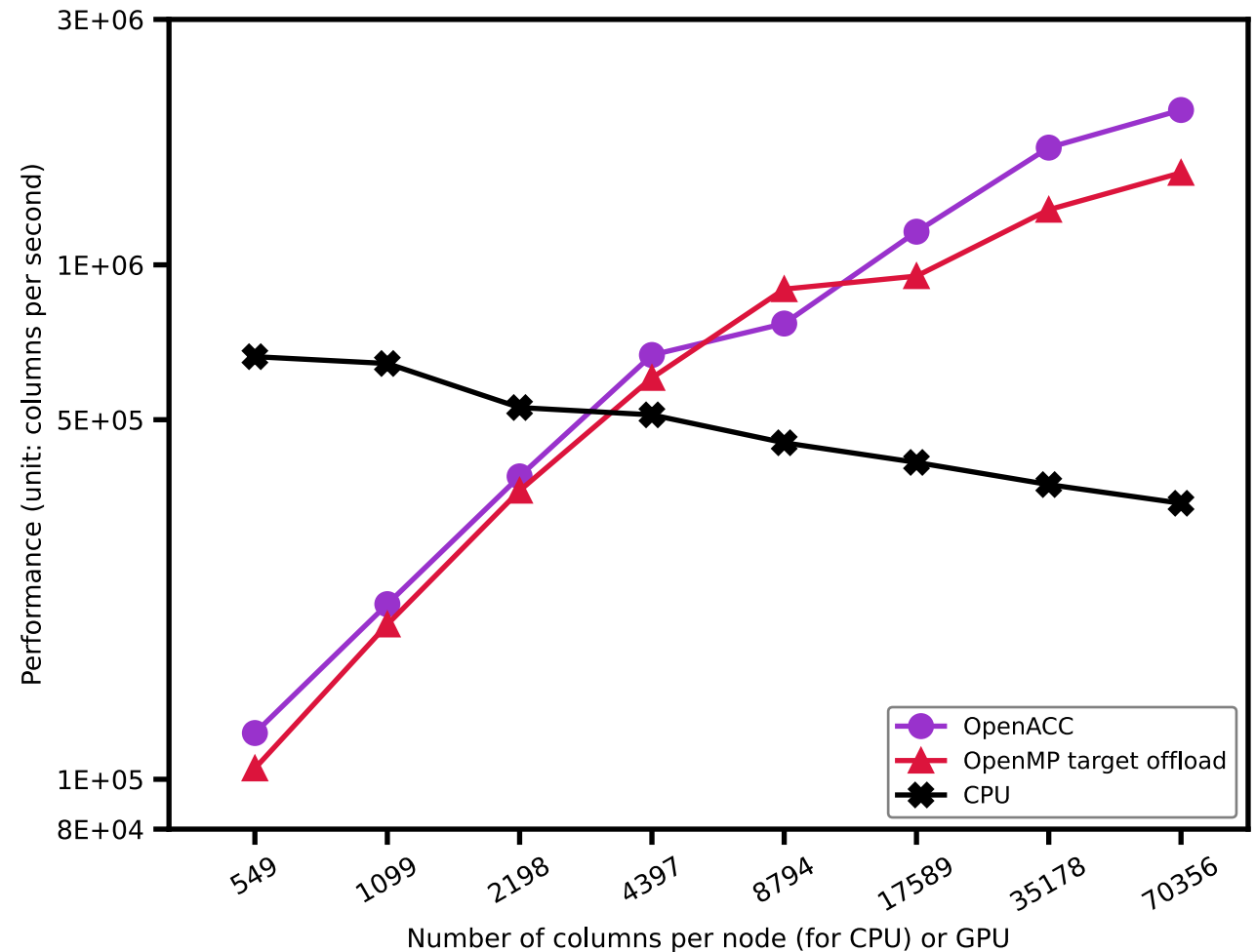
MPAS 7.3 Dynamical Core Scaling: 10 km (5.6M cells) 56 level, FP32, Selene Cluster



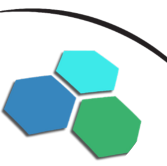


PUMAS/MG3

- Code has been fully ported onto GPUs and has been integrated into the CAM development code
 - There exists an OpenACC version and OpenMP offload version
- Comparison done between one EPYC 7763(AMD Milan) and one A100 in the integrated within CAM (timing results contain data transfer time)
- Maximum speedup over CPUs is 5.8x (for OpenACC) and 3.4x (for OpenMP)

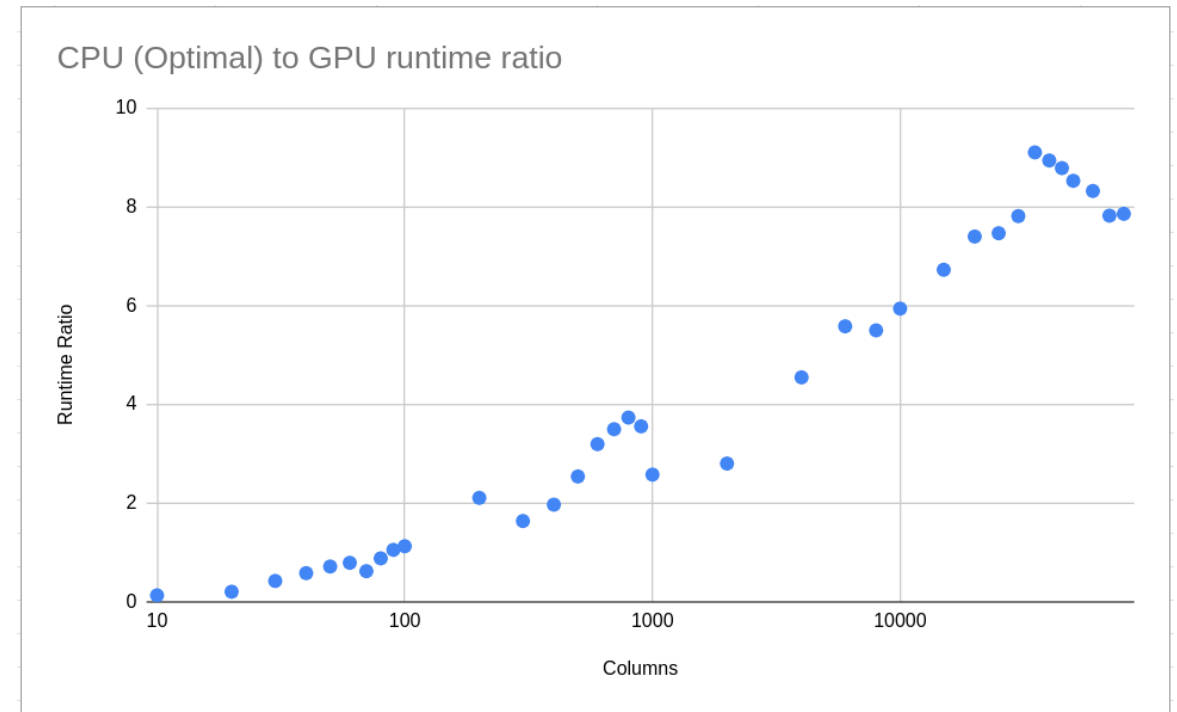
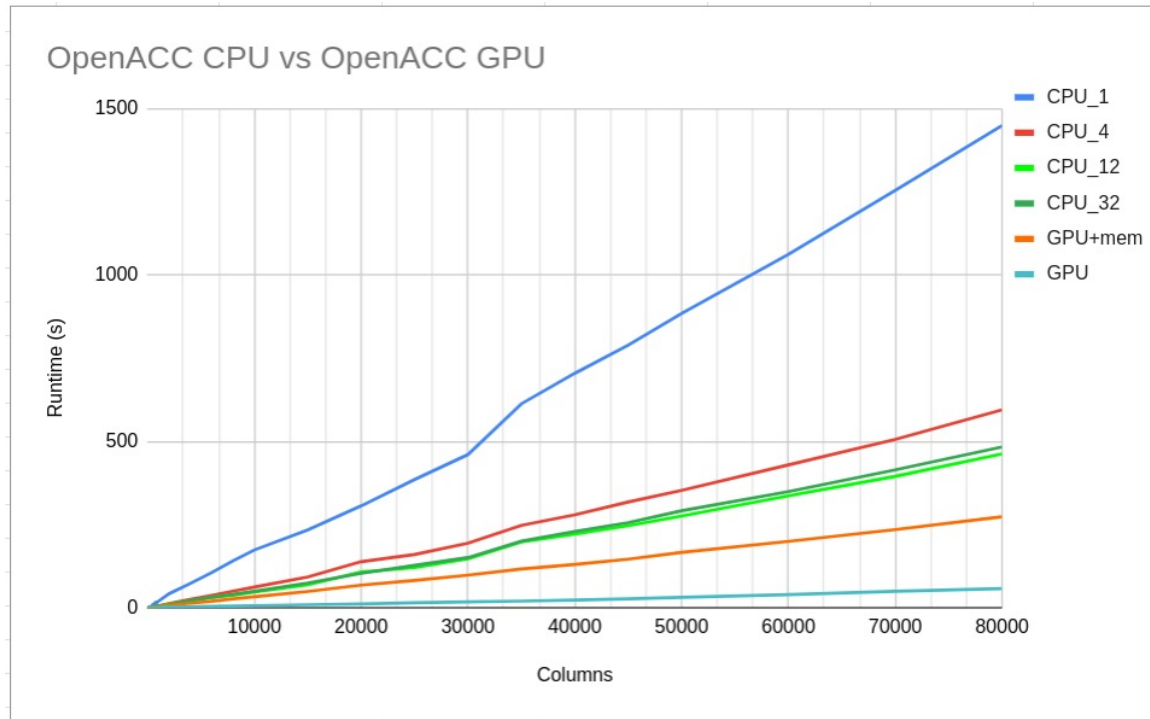


Results courtesy of Jian Sun, NCAR

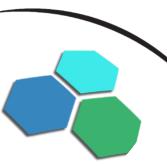


CLUBB

- CLUBB has been fully ported onto GPUs using OpenACC. There is also a version that uses OpenMP offload.
- We are currently working on incorporating CLUBB-GPU into CAM



*Results are Preliminary: Comparisons were run on one EPYC 7763(AMD Milan) and one A100, comparisons done in standalone CLUBB



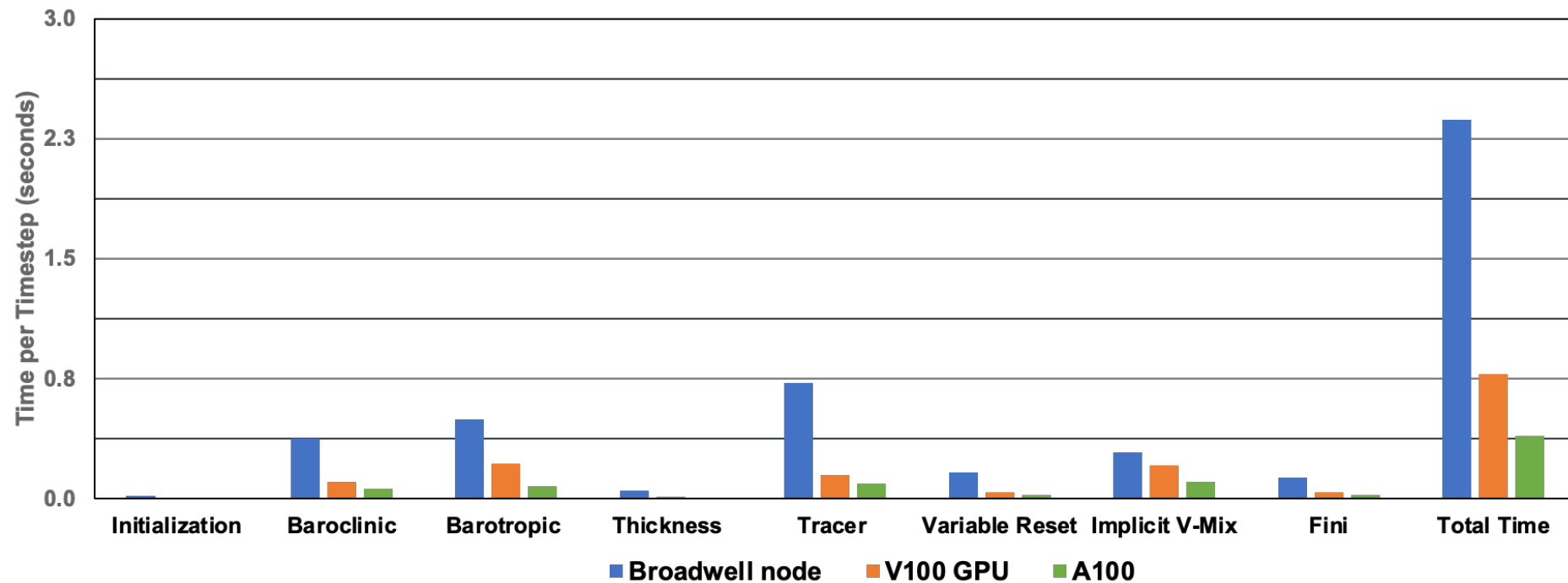
RRTMGP

- Utilizing code from Robert Pincus, et al
 - <https://github.com/earth-system-radiation/rte-rrtmgp>
- We have incorporated RRTMGP into the latest CAM development version
 - It took some time to debug
 - Verified that answers match between CPU and GPU
- Very preliminary results show about a 10x speedup on GPUs in the stand alone version (without data transfers)

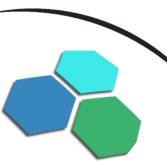


MPAS Ocean

Performance comparison between two 40c Broadwell nodes, and two V100 (Prometheus) and two A100 (Selene) GPUs



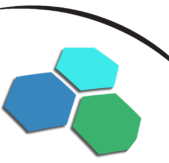
Experiment: Testcase = "EC60to30"; dt = 30 min; 118K cells/GPU; 60 Levels; FP64; 1 MPI Rank/GPU



What is left to do on the GPU front?

- Use profiles of CAM with the GPU-ized physics packages and dynamical core to drive further GPU refactoring.
- We are not planning to port the land or sea-ice model to GPUs as part of the EarthWorks project.
- We *might* need to coordinate with CGD and NOAA to port CCPP coupling framework to GPUs.

QUESTIONS?



Code Availability

<https://github.com/EarthWorksOrg/EarthWorks>

Contact information

mickelso@ucar.edu

Thanks to the full EarthWorks team for their efforts and guidance

David Randall¹, James Hurrell¹, Donald Dazlich¹, Lantao Sun¹, Andrew Feder¹, William Skamarock², Andrew Gettelman², Brian Medieros², Xingying Huang², Sheri Mickelson², Supreeth Suresh², Thomas Hauser², Ming Chen², Dylan Dickerson², Brian Dobbins, Michael Duda², Jim Edwards², Chris Fisher², Jihyeon Jang², Mariana Vertenstein², Richard Loft³, Phil Jones⁴, Luke Van Roeckel⁴, John Cazes⁵, Gunther Huebler⁶

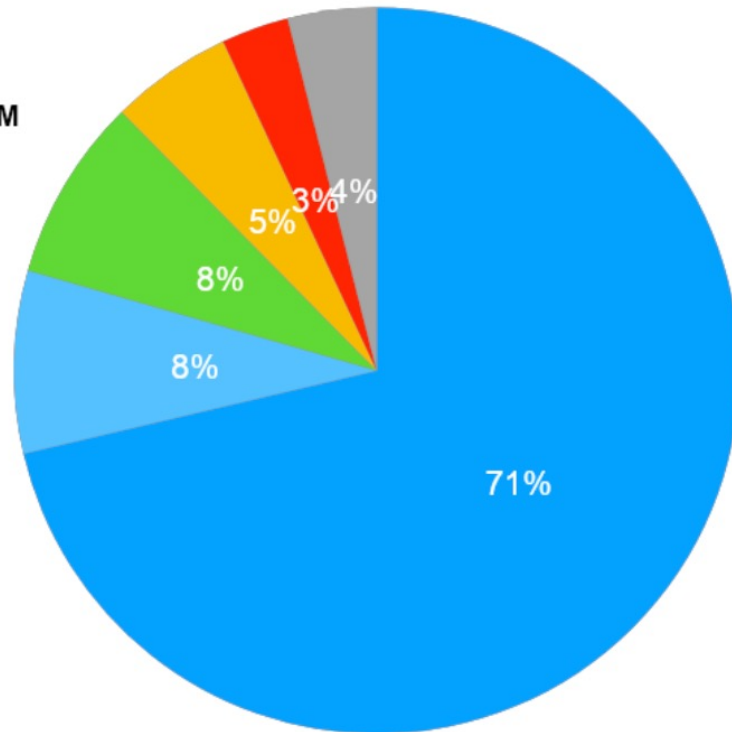
¹Colorado State University, ²National Center for Atmospheric Research, ³AreandDee, LLC,

⁴Los Alamos National Laboratory, ⁵University of Texas at Austin, ⁶University of Wisconsin, Milwaukee

CPU CESM-AMIP Timing Breakdown

AMIP 15km/58 levels. Derecho 10752
ranks/48 I/O ranks

■ ATM RUN
■ ATM I/O
■ LND
■ CPL RUN
■ CPL COMM
■ ICE



•Derecho Supercomputer

- 84 dual-socket AMD-Milan nodes
- HPE Slingshot V11 interconnect

•1 day simulation

- I/O: full history +1 restart
- OCN: climatological
- Throughput: 0.24 myears/day

•Model config:

- ATM: 128 ranks/node
- LND, CPL, ICE, OCN: 1 rank/node
- I/O: 48 ranks total