E3SM's GPU Strategy & Performance

Mark Taylor
Sandia National Laboratories

E3SM Nonhydrostatic Atmosphere NGD (Caldwell)
E3SM-MMF Exascale Computing Project. (Taylor)
E3SM Performance Group. (Jones, Sreepathi)
CMDV-SM project (Salinger)

ESMD PI Meeting (Virtual)
October 26, 2020
SAND2020-10336 C

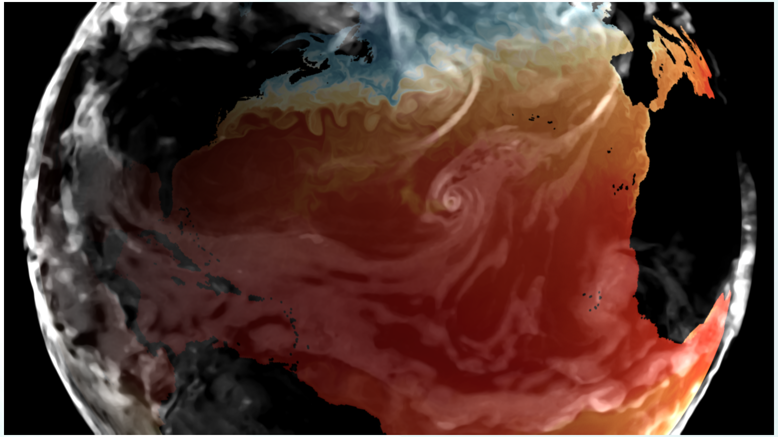
EXASCALE COMPUTING PROJECT

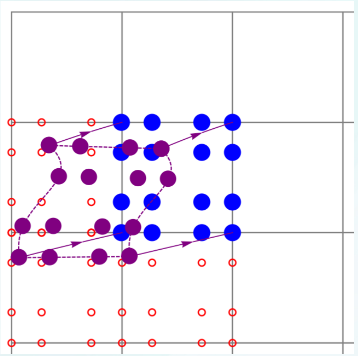
E3SM Energy Exascale Earth System Model

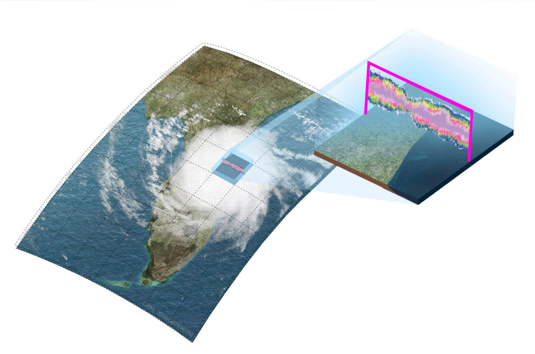
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

U.S. DEPARTMENT OF ENERGY

E3SM Model Development







- **BER: E3SM Project**
- ~70 FTEs, 8 labs + Universities
- **E**nergy **E**xascale **E**arth **S**ystem **M**odel
- DOE-SC science mission: Energy & water issues looking out 40 years
- Ensure E3SM will run well on upcoming DOE exascale computers

- **ASCR/BER SciDAC**
- ~10 FTEs over multiple projects
- Large focus on new algorithms

- **ASCR ECP Project**
- ~10 FTEs
- E3SM-MMF: ”superparameterization”

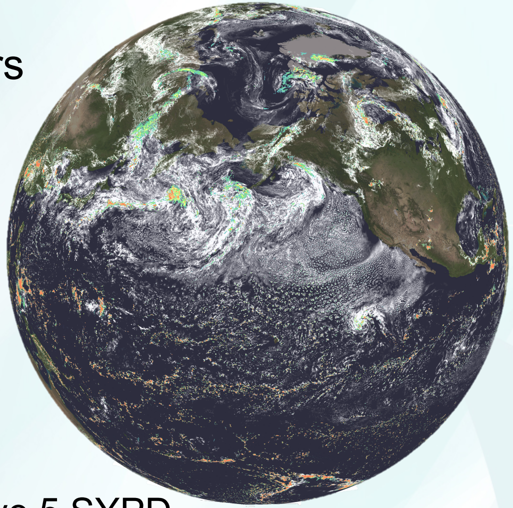
E3SM Exascale strategy: Running on GPUs

- All new DOE SC machines will be GPU based
- 2021: NERSC Perlmutter
  - ~1500 nodes with 4 GPUs each
  - ~3000 CPU nodes
- 2021/2022: OLCF Frontier
  - 30MW
  - Each node with 4 GPUs
- 2021/2022: ALCF Aurora
  - Intel GPU nodes

Key Points for Earth System Models

- CPU performance (per watt) has nearly stagnated
  - 2x speedup over the last 6 years
- GPUs: 3x speedup (per watt) over today's CPUs
  - But only in the high-workload regime
  - Need major code rewrite or refactor
- Traditional climate simulation campaigns are run in the low-workload regime
- E3SM: Focus on several new types of simulation campaigns where GPUs will allow us to run simulations not possible on CPU systems

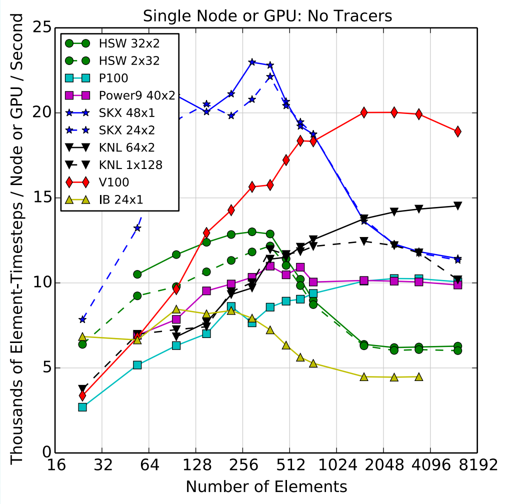
High Workload Simulations for GPUs



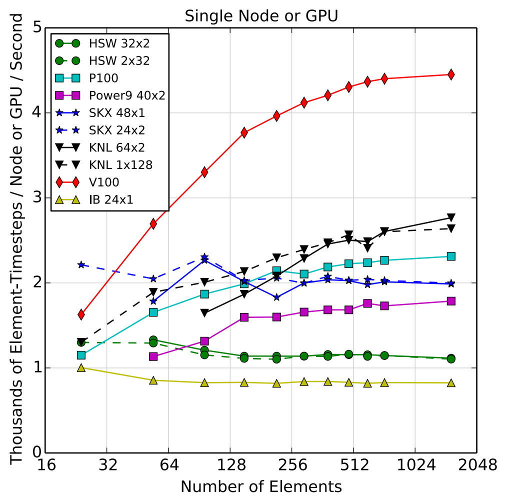
- Earth System Model running at 5 SYPD for ~300 simulated years
  - A $5M commodity CPU cluster is most efficient
- **Ultra high resolution “Cloud Resolving” model**
  - BER: E3SM’s ”SCREAM” project
  - On track for E3SM V3 (2021-2022)
  - Typical INCITE award: 10-30 simulated year
- **Increase “local” complexity**
  - ECP: E3SM-MMF project: “superparameterization”
  - Achieve many aspects of a cloud resolving simulation and also achieve 5 SYPD
  - Running on Summit since OLCF Early Access Program
- **Large Ensembles**
  - Each ensemble member is running slower but more efficiently on GPU systems.
  - E3SM V4 capability: large ensembles on GPU systems

Detailed single node comparisons

- Atmosphere dynamical core with and without tracers (~1/2 the atmosphere mode)
- Shows slow improvement in CPUs going back to 2012
- Gives upper bound on possible performance: doesn't include. MPI and other overheads
- Dynamical core only: minimal benefit from GPUs
- Dynamical core + tracers: Large benefit from GPU, but only in high work load regime.
- Hope to soon add newer CPUs (Epyc, Fugaku ARM) and new GPUs



Atmosphere dynamical core



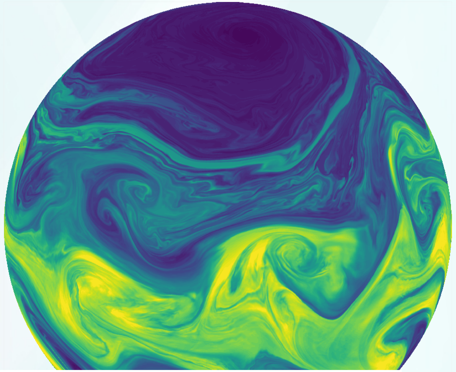
Atmosphere dynamical core with 40 tracers

Bertagna et al., *HOMMEXX 1.0: A Performance Portable Atmospheric Dynamical Core for the Energy Exascale Earth System Model*, GMD 2019

Timeline and GPU Readiness

- V2: 2019 code freeze, 2020 begin simulations
  - CPU: All major simulation campaigns, Fortran code base
  - GPU: NH atmosphere dycore. partial: MPAS-O, ATM physics
- V3: 2022 code freeze, 2023 begin simulations
  - CPU: All major simulation campaigns, Fortran code base
  - GPU: MPAS components (Fortran/openMP)
  - GPU: SCREAM (prescribed aerosols) NH atmosphere
  - GPU: MMF ("super-parameterization") atmosphere
- V4: 2025 code freeze, 2026 begin simulations
  - Full Earth System Model running on both CPU and GPU systems

NGGPS cloud resolving (3km) benchmark
Scaling to all of Summit



- Standardized benchmark from the National Weather Service
- Atmosphere model with realistic configuration and idealized physics
- Highlights from several generation of computers and Global clould resolving models (GRCMs)
- Double precision results (reported real*4 results ~1.6x faster)
- Results inline with Linpack/HPGC comparison

GCRM Model	Computer (Linpack rating)	NGGPS 3km Benchmark
NOAA FV3	Edison (2.6PF)	0.16 SYPD
HOMME (CESM)	TaihuLight (125 PF)	0.34 SYPD
E3SM's HOMMEXX_NH	Summit (200 PF)	0.97 SYPD (1.14 hydrostatic)

L. Bertagna et. al., SC '20: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2020

Closing thoughts

- CPU systems are getting better (Fugaku)
- GPU systems are getting better (Frontier, Aurora)
- E3SM V3: Two approaches to cloud resolving simulations made possible by GPU architectures
- E3SM V4 will run efficiently on *both* architectures
  - hopefully well prepared to adapt to post-Exascale hardware

Backup Slides

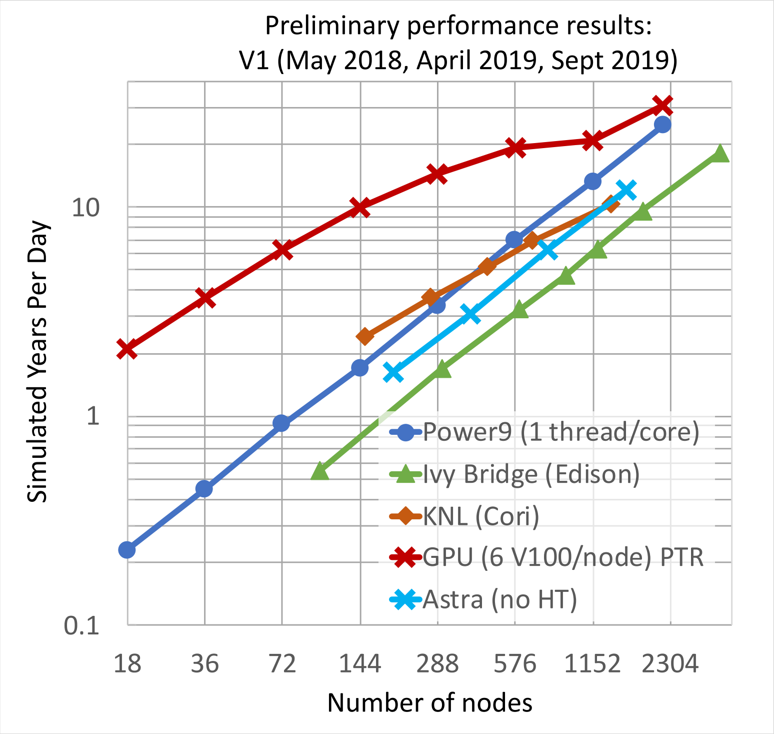
Performance & Portability Strategy

- Large investments in two approaches
  - Both strategies require complete code refactor or rewrite, with careful coding to obtain competitive performance
  - Takes us several years per major component
- C++/Kokkos:
  - Rewriting from scratch allows us to take the opportunity to: replace legacy code, introduce low level unit and property tests
- Fortran/openMP
  - Death of Fortran continues to be predicted
  - GPU Support now lagging substantially
  - Vectorization (for CPUs) remains superior

Benchmarks: High-Res Climate, Strong scaling

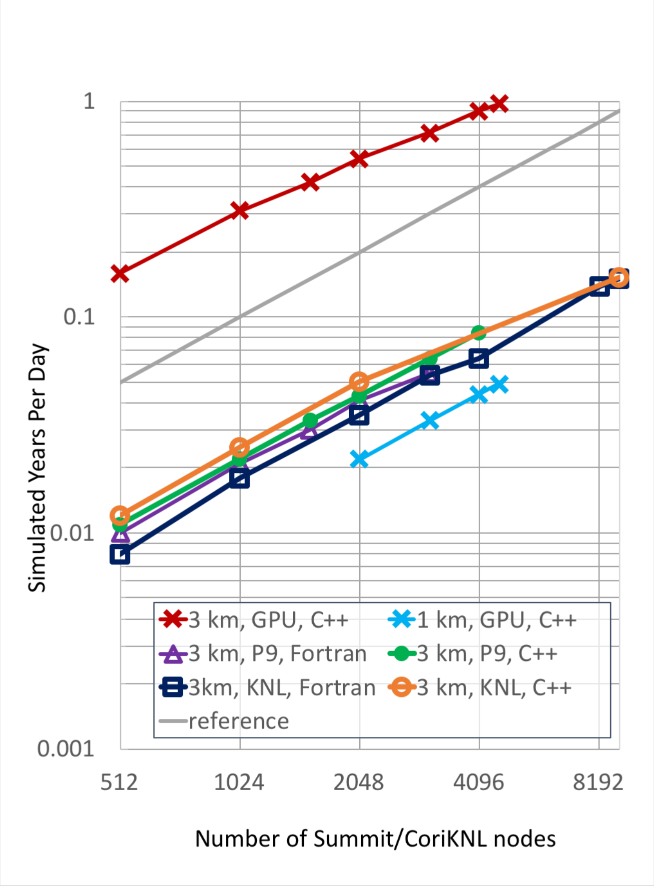
Consider the E3SM v1 model running on Summit:

- Atmosphere dycore benchmark (25km, 72L, 40 tracers).
- In strong scaling limit, GPU systems cannot outperform CPUs
- If you are willing to run slower, GPUs provide significant advantage
- E3SM high-res coupled model (projections)
  - 5 SYPD: no GPU benefit
  - 0.5 SYPD: GPUs more efficient
- How will this change in the future?



NGGPS 3km Benchmark: Strong scaling (per node)

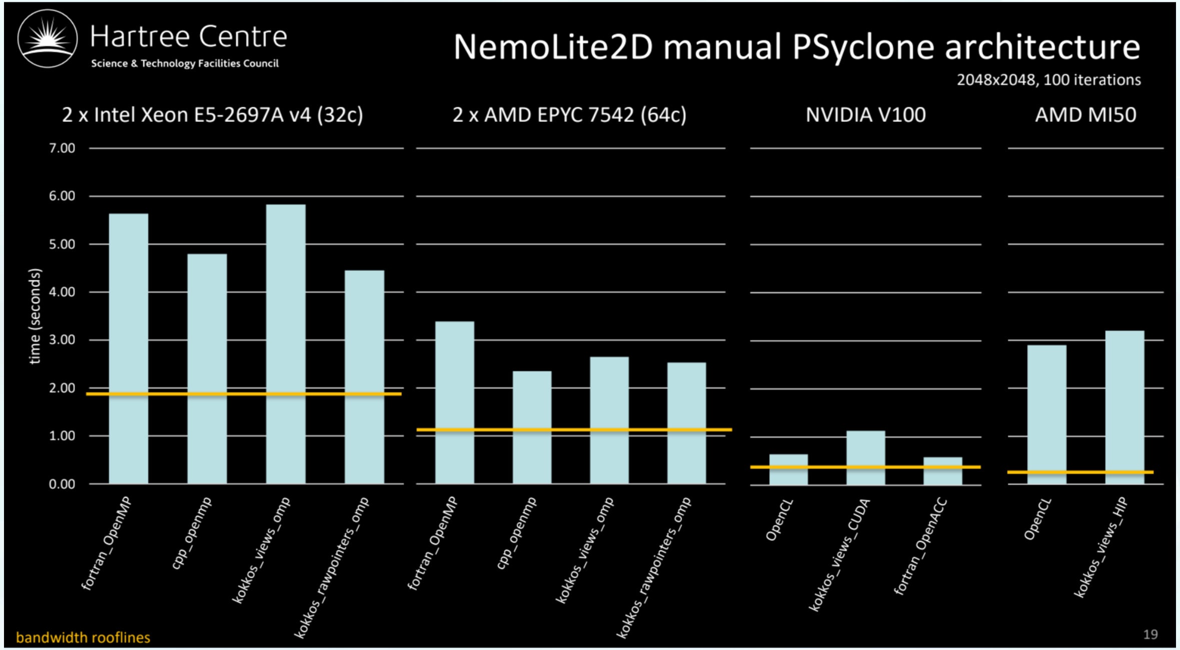
- C++/kokkos and Fortran codes competitive IBM P9
- C++/kokkos code has hand-vectorized every loop, leading to excellent KNL performance.
- Summit node with 6 V100s obtains ~ 12x speedup (for ~6x more power?)
- 1km resolution also running well, but throughput is impractically low



L. Bertagna et. al., SC '20: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2020

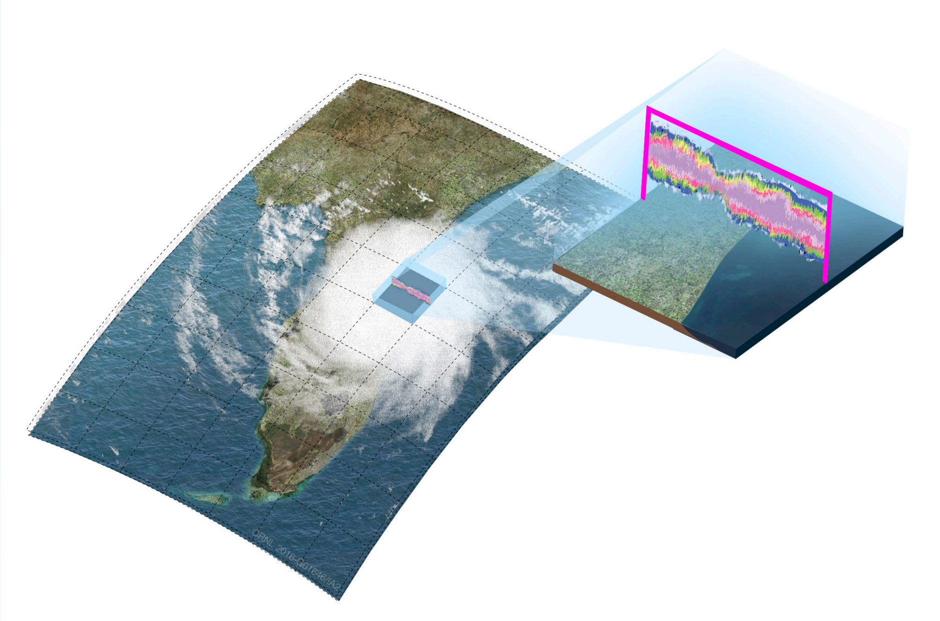
NemoLite 2D results (single node/GPU)

- From Sergi Siso, Multicore 10 Workship PSyclone presentation.
- Showing performance of several programming model backends
- Roughly 2x speedup in CPU nodes, from Intel Ivy Bridge to AMD EPYC
- V100 about 3x faster than EPYC node



ECP Project: E3SM-MMF

- E3SM-MMF approach addresses structural uncertainty in cloud processes by replacing traditional parameterizations with cloud resolving "**superparameterization**" within each grid cell of global climate model
- Super-parameterization dramatically increases arithmetic intensity, making the MMF approach ideal for GPU acceleration.
- Exascale + MMF approach will make it possible for the first time to perform climate simulation campaigns with *some aspects* of cloud resolving resolutions.



Three overarching science drivers

U.S. energy sector is vulnerable to:

- Decreasing water availability
- More intense storm events and flooding
- Increasing temperatures
- Sea level rise

- **Water cycle:** How does the hydrological cycle interact with the rest of the human-Earth system on local to global scales to determine water availability and water cycle extremes?
- **Biogeochemistry:** How does the biogeochemical cycle interact with other Earth system components to influence energy-sector decisions?
- **Cryosphere systems:** How do rapid changes in cryospheric systems evolve with the Earth system and contribute to sea level rise and increased coastal vulnerability?

Challenge: water cycle, biogeochemistry, and cryosphere systems interactions cannot be ignored for predictions or projections at longer time scales