Present and Future Computing Requirements for **P**redicting **I**ce **S**heet and **C**limate **E**volution at **E**xtreme **S**cales (PISCEES)

Stephen Price

Los Alamos National Laboratory

NERSC BER Requirements for 2017
September 11-12, 2012
Rockville, MD

*** Support from SciDAC through DOE's Office of Science BER and ASCR Programs ***

(1) DOE lab participants

- **LANL**
  - Bill Lipscomb, overall PI (Phil Jones, acting PI)
  - Steve Price, Science Team Lead - Dycore Development
- **LBNL**
  - Esmond Ng, Computational Team Lead (**FASTMath, SUPER**)
  - Dan Martin, BISICLES dycore development (**Chombo**)
  - Sam Williams, dycore performance (**SUPER**)
- **ORNL**
  - Kate Evans, Science Team Lead - V&V
  - Pat Worley, dycore performance (**SUPER**)
  - Matt Norman, V&V, advection
- **SNL**
  - Andy Salinger, Computational Team Lead (**FASTMath**, **QUEST**, **Trilinos**)
  - Mike Eldred, UQ (**QUEST**, Dakota)
  - Ray Tuminaro, scalable preconditioners (Trilinos)

(1) University participants

- **Florida State U.**
  - Max Gunzburger, FELIX dycore development
  - Mauro Perego, FELIX dycore development
- **MIT**
  - Patrick Heimbach, adjoint methods
- **U. South Carolina**
  - Lili Ju, FELIX dycore development
- **U. Texas at Austin**
  - Charles Jackson, Science Team Lead - Optimization and UQ
  - Georg Stadler, initialization and optimization
  - Omar Ghattas, coordination with PARADICE project (**QUEST**)
- **NCAR**
  - Mariana Vertenstein, CESM integration
  - Bill Sacks, CESM integration

(1) Project Goals

- To develop and apply robust, accurate, and scalable dynamical cores ("dycores") for ice sheet modeling on structured and unstructured meshes with adaptive refinements
- To evaluate ice sheet models using new tools and data sets for Verification and Validation (V&V) and Uncertainty Quantification (UQ)
- To integrate these models and tools in the Community Ice Sheet Model (CISM) and Community Earth System Model (CESM)

(1) PISCEES Now vs. 2017

NOW:

- New dycore development, model coupling (dycores into CISM & CISM into CESM), initial V&V efforts (test suite automation, dataset gathering/massaging for use in model-obs comparison), initial UQ (low param. space / sampling-based approaches & exploratory use of adjoint-capable codes), and baseline performance evaluations for analysis by SUPER

2017:

- Stand-alone and CESM fully-coupled (ocean-atmos-ice) runs with optimized initial conditions; ensembles of fwd model runs for UQ on model outputs (e.g., sea-level rise)

(2) Computational Strategies

- Computation time is dominated by repeated solution of large, nonlinear, ill-conditioned, sparse elliptic system of equations
- Codes: Community Ice Sheet Model (CISM) and Community Earth System Model (CESM)
- Algorithms: Krylov-based methods on linear systems, Newton-based methods on nonlinear systems; FEM & FVM discretizations using **Trilinos** and **Chombo** libraries; MPAS unstructured meshing framework
- Computational challenges: robustness of nonlinear solver over range of input datasets/resolutions; problem-specific solver convergence; performance variability related to preconditioning
- Scaling limited by: problem size, hardware issues on *LCF (e.g. cpu layout)
- Work with SUPER to identify computational kernels that can be attacked w/ GPUs / threading
- NOTE: most of work is done within **Trilinos** or **Chombo** – so accelerator improvements to those libraries are "free" to our codes

(2) Computational Strategies (cont.)

- Anticipated changes by 2017:
  - Reliable scalability to O(10^4) cpus
  - Baseline >2x speed-up over current prototype codes (through basic optimizations, e.g. PGI→GNU gives ~2x)
  - ~6x increase in problem size due to higher resolution & stability requirements for explicit time stepping
  - Improvements over existing fwd Euler time stepping (semi- or fully-implicit) allowing for larger time steps (non-CFL limited)
  - Acceleration through GPUs, threading, etc. (as allowed for by improvements to solver libraries)

Assumptions

- Ice sheet codes for future use are BISICLES & FELIX
- FELIX scales approx. like SEACISM code, which we've used for most production scale runs (5 km res Greenland ice sheet)
- BISICLES, when accounting for move to 3d and other improvements, performs within ~2x of SEACISM for similar sized problems: BISICLES ~ FELIX ~ SEACISM w.r.t. performance
- Antarctica problems are ~8x larger than Greenland problems
- Doubling resolution on a regular mesh increases comp. time by ~2^2=8x (2x2 for spatial increase in 2d, 2x for halving of explicit time step)
- Anticipated problems require <1 km res in some areas, >5 km res in others
- Unstructured and/or refined mesh savings is ~6-10x
- Increase in problem size from current to anticipated spatial resolution (and accounting for mesh changes) gives problem size increase factor of ~6x

(3) Current HPC use

- Facilities: NERSC & OLCF
- Hours used in 2012*:
  - NERSC: 900e3 hrs (ice sheet), 1400e3 hrs (ocean)
  - OLCF: 3,275e3 hrs (ice sheet)

  total: 4.1 million (ice only) + 1.4 million (ocean)
- Concurrency, run time, runs per year:

  Typical production run**
  - Core hours used: ~22 k hrs
  - Wall clock hours: ~22 hrs (1 k cpus for 22 hrs ~200 model yrs)
  - No. runs per year: ~30 runs @ NERSC + 30 runs @ OLCF (remainder = devel)

* 2011-2012 for ice sheet modeling as part of ASCR ISICLES and BER IMPACTS projects

** 5 km res Greenland = 650 k grid cells and ~1.3e6 DOF (~8x for Ant.)

(3) Current HPC use (cont.)

- I/O per run:
  - Typical production run**
    - Checkpoint = 60 Mb
    - Output = 6 Gb
- Memory used per node: currently no special requirements
- Software, services & infrastructure:
  - Trilinos & Chombo solver libraries; netCDF; standard MPI & compiler libraries
- Data resources used / data stored:
  - Currently only small amount of project data stored on HPSS (e.g. ~100 Gb at OLCF)

** 5 km res Greenland = 650 k grid cells and ~1.3e6 DOF (~8x for Ant.)

(3) HPC use in 2017

Estimated hours needed in 2017:

Ice sheet model only runs

100, 100 yr Greenland: 100 * 55e3 hrs = **5.50e6 hrs** (11e6 hrs)

100, 100 yr Antarctica: 100 * 435e3 hrs = **43.50e6 hrs** (87e6 hrs)

Coupled runs

3 ice/ocean Greenland: 3 * (0.5e6 hrs + 55e3 hrs) = **1.67e6 hrs**

3 ice/ocean Antarctica: 3 * (5e6 hrs + 435e3 hrs) = **16.31e6 hrs**

2 ice/ocean/atmos Greenland: 2 * (2*0.5e6 hrs + 55e3 hrs) = **2.11e6 hrs**

2 ice/ocean/atmos Antarctica: 2 * (2*5e6 hrs + 435e3 hrs) = **20.87e6 hrs**

1 ice/ocean/atmos/sea-ice Greenland: = 3*0.5e6 hrs + 55e3 hrs = **1.56e6 hrs**

1 ice/ocean/atmos/see-ice Antarctica: 3*5e6 hrs + 435e3 hrs = **15.44e6 hrs**

total: 107 million hrs (156 million hrs)

**For coupled runs, ice sheet model is small fraction of overall cost (3-10%)

CESM component (ocean, atmos, sea-ice) estimates from M. Maltrud (LANL)

(3) HPC use in 2017 (cont.)

I/O per run:

Stand alone runs (neglecting restarts)

100, 100 yr stand-alone Greenland: 100 * 33 Gb = **3,300 Gb**

100, 100 yr stand-alone Antarctica: 100 * 270 Gb = **27,000 Gb**

Coupled runs (neglecting restarts)

3 ice-sheet/ocean Greenland: 3 * (33 Gb + 120 Gb) = **460 Gb**

3 ice sheet/ocean Antarctica: 3 * (270 Gb + 1,000 Gb) = **3,810 Gb**

2 ice-sheet/ocean/atmos Greenland: 2 * (33 Gb + 2 * 120 Gb) = **545 Gb**

2 ice-sheet/ocean/atmos Antarctica: 2 * (270 Gb + 2 * 1,000 Gb) = **4540 Gb**

1 ice-sheet/ocean/atmos/sea-ice Greenland: 33 Gb + 3 * 120 Gb = **390 Gb**

1 ice-sheet/ocean/atmos/see-ice Antarctica: 270 Gb + 3 * 1,000 Gb = **3270 Gb**

total: 43,315 Gb ~ 43 Tb

Restarts: ~0.4 - 4 Gb (Greenland, Antarctica), ~80-800 Gb for ocean, or atmos, or sea ice

(3) HPC use in 2017 (assumptions)

- Assumptions that ~20 model optimization (deterministic inverse) problems are done for both Antarctica and Greenland using an efficient adjoint-based code, which costs ~100x one forward model solve (for a 1 yr time step, 100 forward model solves is approx. equal in cost to a 100 yr forward model run).

** Climate-coupled runs are assumed to be conducted using MPAS-ocean/atmos/sea-ice, at high spatial resolution and on a regional domain, allowing for ~10x savings over current hi-res (10^{th} degree) POP. Cost of atmos and sea ice are assumed similar. Original estimates were made for Antarctic sub-domain. A Greenland sub-domain is assumed to be ~10x cheaper.

*** Estimate for the total number of stand-alone ice sheet simulations is optimistic in terms of addressing UQ aspects of the proposed work. In particular, it assumes success in characterizing, and forward propagation of uncertainties using "linearized UQ" (linear adjoint approaches) AND/OR the use of efficient emulators to sample the parameter space efficiently without large numbers of expensive forward model runs. Neither of these approaches are currently in common use or have been adequately tested on analogous problems. For these reasons, it should be noted that sampling based methods of (forward propagation) UQ could easily increase the cpu requirements noted here by 1-2 orders of magnitude (and possibly more). Numbers in parentheses above assume a modest ~2x increase in the number of stand-alone ice sheet model runs in order to allow for a more expansive UQ approach.

(3) HPC use in 2017 (cont.)

- Memory used per node: no anticipated changes
- Software, services & infrastructure: no anticipated changes
- Data resources used / data stored:

HPSS archive for all production runs and fraction of restart files
= 100 Tb

(5) Strategies for new architectures

- Strategy for running on new many core architectures:
  - PISCEES project members (Worley & Williams) are part of SUPER institute, and are tasked with scoping code for kernels that may be amenable to GPUs, etc. Also, improvements to **Trilinos** and **Chombo** libraries should come along to our codes for free.
- To date we have prepared for many core by … (see above)
- We are already planning to … (see above)
- To be successful on many core systems we will need help with:

(6) Summary

- New science that might be afforded:
  - hi-res, fully coupled simulations of ice sheet & climate evolution (e.g. sea-level rise) with uncertainty quantification
- Architecture recommendations:
  - queue option for physically co-located processor blocks (?)
- What could you achieve w/ 32x your current allocation:
  - above given CPU estimates are ~30x current total allocations for ice sheet modeling on NERSC and OLCF during the past ~2 yrs
- What expanded HPC resources are important to your project?
- General discussion