Leveraging the New CESM1 CPL7 Architecture - Current and Future Challenges

Mariana Vertenstein

NCAR Earth System Laboratory
CESM Software Engineering Group (CSEG)

NCAR is sponsored by the National Science Foundation
CESM1 Provides a Seamless End-to-End Cycle of Model Development, Integration and Prediction with One Unified Model Code Base
Challenges

- Ultra-high resolution
  - Parallel I/O, Memory and Performance Scaling
- Ocean Data Assimilation
  - Extension of CPL7
- New Components
  - e.g., Addition of Land Ice Sheet
- Component Nesting and New Grids
  - e.g., NRCM and coastal upwelling
- New Flexible Mechanism of Passing Fields Through Coupler
  - e.g., support of isotopes
- New Physics
  - e.g., MMF (Superparameterization)
Challenge – Ultra-high resolution
Parallel I/O, Memory and Performance Scaling

Approach – add memory and performance scalability to all model components

- Parallel I/O in all model components and minimization of global arrays
- Effective use of both MPI and OpenMP in all model components
PIO rearranges data from model decomposition to I/O-friendly decomposition. It serves as an interface between the model and the I/O library. Supports:

- Binary
- NetCDF3 (serial netcdf)
- Parallel NetCDF (pnetcdf) (MPI/IO)
- NetCDF4

No imprint of model decomposition in output file.
PIO in CESM1

- Implemented in *ALL* model components
- No imprint of model decomposition in output
- Usage is critical for high resolution, high processor count simulations
  - Pure old-style serial I/O is one of the largest sources of global memory in CCSM - *will eventually always run out of memory*
  - New serial NetCDF I/O with PIO eliminates memory bottleneck – but not performance bottleneck
- New asynchronous capability currently being added
  - Challenges due to buffering
- Other Parallel I/O libraries
  - e.g., Adios: new format (bp) requires translation and imprint of model decomposition
CESM1 Cray XT Scalability

0.25° atm / 0.1° ocn

(Courtesy of Pat Worley)
CCSM/HOMME Scalability

0.125° atm / 0.25° Ind / 0.1° docn

Work of Mark Taylor, Jim Edwards and Brian Eaton

- CCSM times include ALL CCSM components (PIO use was critical)
- Scalability of the dynamical core is preserved by CAM and scalability of CAM is preserved by CCSM
- Scale out to over 128K cores get 5 SYPD (JaguarPF)
Challenge: Ocean Data Assimilation

Extension of CPL7

Why? Obtain better ocean initial conditions for CMPI5 decadel prediction runs

Approach - extend coupling framework to permit multiple instances of component models - use DART for ocean
Data Assimilation Using DART

- Leverage NSF PetaApps IE code base (NCAR, COLA, NERSC, U. Miami)
- All instances run concurrently on non-overlapping processor sets
- Setting up different number of instances requires editing ONLY 1 line of an xml file

Currently being used to perform ocean data assimilation using DART for POP2

48 instances of both DATM and POP2

DART -> creates new restarts for each POP instance
Benefit of data assimilation

(work of Steve Yeager, Jeff Anderson, Tim Hoar, DART team)
Data Assimilation Next Steps

- Optimize performance of coupler for data assimilation
- Add data assimilation capability for all model components
New Components
e.g., Addition of Land Ice

Why? Provide critical predictions of land-ice retreat and sea-level rise

Approach – CPL7 and CLM extended to permit new coupling
Glimmer in CESM 1.0

- CESM 1.0 includes Glimmer, the Community Ice Sheet Model - first publicly released IPCC class model to include a dynamic ice sheet model
- Glimmer is serial code with dynamic Greenland ice sheet – new parallel code with higher-order dynamics to be added soon
- Only ice sheet-land coupling currently done – a new surface mass balance scheme for ice sheets in CLM - computed on the global land grid, then sent to Glimmer-CISM and downsampled to the local ice sheet grid (ice sheet time step is 1 year)
- Ice sheet-ocean coupling (important for Antarctic ice sheet) under development
Component Nesting and New Grids

e.g., NRCM and coastal upwelling

Why? Determine effects of upwelling on air temperature, precipitation and wind patterns and investigate atmosphere/ocean feedbacks in upwelling regions and how these patterns evolve with climate change

Approach – *nested* regional climate models (*NRCM*) *in global model*
Eastern boundary current biases in CCSM3

One approach - use nested, regional model(s) to increase resolution at eastern boundaries to better resolve upwelling
Possibly as an interim step before unstructured grids

Increasing atmospheric resolution fixes some but not all of the bias problem

(courtesy of Justin Small)

CCSM3.5 finite volume atmosphere (Gent et al. 2010)
NRCM and Upwelling Regions

Atmospheric Surface Data

Ocean Lateral Boundary Data
CPL7 Implementation

- Create a new Hybrid OCN component that is itself a driver for POP and ROMS (regional ocean model)
- OCN driver passes fluxes and state information to POP and ROMS and controls communication between global and regional oceans - merged SST from POP and ROMS (on POP grid) passed back to driver
- Atm/Ocn fluxes calculated in CPL for POP – but now in ROMS using new extra atmosphere states time series sent by CPL
New Flexible Mechanism of Passing Fields Through Coupler
e.g., support of isotopes

Why? Provides additional diagnostics for assessing model performance of both present and past climates

Approach – extend coupler to have constituents transferred between components from only runtime specifications
Benefits and Requirements

- **Scientific Benefits**
  - Water isotopes
    - track moisture sources of air masses
    - provide insight into global hydrological cycle and cloud processes
  - Carbon isotopes
    - diagnostics tracers in ocean and land carbon cycle
    - infer deep ocean ventilation rates and validate ocean model circulation

- **Requirements**
  - New flexible scheme of passing tracers between components - must also support tagging (source/flavor) of isotopes
  - Coupling code should not change when new isotopes are added – only component code must change to support addition
  - Run or compile time specification of isotopes transferred between components
New Physics

e.g., Superparameterization

Why? Represent the effects of clouds more accurately – one of the major sources of uncertainty

Approach – implement superparameterization in CAM - replace parameterized moist physics by using nested cloud resolving models (CRMs) in each grid column.
Superparameterization

Each gridcell has a 2D cloud resolving model (CRM)

CRMs do not communicate with each other—embarrassingly parallel

Radiation runs on CRM grid, once per GCM time step—passed to GCM

SP “horizontal” grid does not have physical significance

Surface fluxes computed in GCM and passed to CRM - need higher moments sent that describe spatial variability*

Idea from W. Grabowski
Other Upcoming Challenges

- Completion of AR5/CMIP5 simulations
- Regional Refined Grids (static regionally refined meshes - MPAS)
- Incorporation of hooks for human dimensions
- Improved verification metrics
- Examination of possible uses of GPUs
- Parallelizing post-processing utilities and workflow
Improvements Were Impossible Without the Contributions of

- CCSM Software Engineering Group
  - David Bailey, Tony Craig, Brian Eaton, Jim Edwards, Chris Fischer, Diane Feddema, Brian Kauffman, Erik Kluzek, Andy Mai, Nancy Norton, Mathew Rothstein, Francis Vitt

- Software Engineering Working Group
  (DOE/SciDAC, CISL, ESMF)
  - John Dennis, Nathan Hearn, Rob Jacob, Fei Liu, Sheri Mickelson, Art Mirin, Ray Loy, Mark Taylor, John Truesdale, Pat Worley, ESMF Core Team

- CESM Scientists