

Enabling Climate Simulation at Extreme Scale - Performance Analysis and Modeling -Monika Lücke, Alexandru Calotoiu, Prof. Felix Wolf

Performance Analysis

- Parallel efficiency and scalability crucial for climate simulations that run for months on high core counts
- Usage of performance analysis tools, e.g. Scalasca

Automated Performance Modeling

- Applications may entail latent performance limitations, which would show only on larger scale
- Performance models enable sophisticated analysis and extrapolation of performance properties, but their manual generation is very laborious
- Identified computational load imbalance as root cause for poor scalability of some communication routines within the ocean and sea ice model

| bsolute | Absolute | Peer distribution |
|--|--|---------------------------------------|
| Metric tree | Call tree Flat view | System tree Box Plot Topology 0 earth |
| 0.00 Time 3.40e7 Execution 0.00 MPl 0.00 Communication 1.29e5 Point-to-point 2.63e4 Collective 1.10e6 Init/Exit 2.02e4 Overhead 4.54e10 Visits 0 Synchronizations 0 Communications 0 Point-to-point 1.87e9 Sends 1.87e9 Receives 5.88e7 Collective 1.59e13 Bytes transferred 0.00 Computational imbalance 4.89e5 Overload 5.12e5 Underload 1.77e5 Non-participation 0.00 Singularity | □ 0.00 MAIN(0.00%) □ 0.00 ccsm_comp_mod.ccsm_pre_init_(0.00%) □ 0.00 ice_comp_mct.ice_run_mct_(0.00%) □ 5483.71 ice_comp_mct.ice_import_mct_ □ 5057.45 ice_step_mod.step_therm2_ □ 0.00 ice_step_mod.step_dynamics_(0.00%) □ 5922.44 ice_boundary.ice_haloupdate2di4_ □ 68049.93 ice_grid.t2ugrid_vector_ □ 3325.61 ice_boundary.ice_haloupdate2dr8_ □ 0.00 ice_boundary.ice_haloupdate3dr8_ □ 1449.66 MPI_Isend □ 2.38e4 MPI_Waitall □ 30.74 ice_boundary.ice_haloupdate2dr8_ □ 0.00 ice_grid.u2tgrid_vector_(0.00%) □ 0.00 ice_boundary.ice_haloupdate2dr8_ □ 1.294 MPI_Irecv □ 49.87 MPI_Isend □ 1.32e4 ice_transport_driver.transport_remap_ □ 1.13e4 ice_state_bound_state_ □ 3757.25 cice_runmod.coupling_prep_ | |
| .00 1.29e5 (0.37%) 3.52e7 | 0.00 1.72e4 (13.34%) 1.29e5 | 0.00 0.00 100 |

Scalasca performance analysis of a sea ice simulation run on Jugene using 23'404 cores. Communication time spent in function MPI_Waitall indicates long wait time at the border to sea ice regions. Courtesy: John M. Dennis (NCAR)

- Tool developed to automate model-building process ^[1]
- Automated modeling revealed two scalability issues in HOMME, the dynamical core of the Community Atmospheric Model (CESM-CAM)



Automatically generated performance model of the atmospheric model reveals two latent scalability bottlenecks, one of them previously unknown. Experiments on Juqueen.

Load-Balancing Simulator

- Implementation and test of different load balancing strategies usually not possible without major code surgery
- Development of a software engineering tool that facilitates comparison of different load-balancing strategies via simulation
- An initial simulation estimates that an alternative partitioning of sea ice would result in a speedup of 2.7



[1] Alexandru Calotoiu, Torsten Hoefler, Marius Poke, Felix Wolf: Using Automated Performance Modeling to Find Scalability Bugs in Complex Codes. In Proc. of the ACM/ IEEE Conference on Supercomputing (SC13), Denver, CO, USA, ACM, 11/2013.

Project Funding

G8 Research Councils Initiative on Multilateral Research, Interdisciplinary Program on Application Software towards Exascale Computing for Global Scale Issues



Partners

- INRIA (France)
- Barcelona Supercomputing Center (Spain)
- National Center for Atmospheric Research (USA)
- National Center for Supercomputing Applications (USA)

Initial partitioning of the sea ice model shows high load imbalance (front), while alternative partitioning improves load balance (back). Experiments on Juropa.

- German Research School for Simulation Sciences (Germany)
- University of Illinois at Urbana Champaign (USA)
- University of Tennessee (USA)
- University of Victoria (Canada)
- Tokyo Tech and University of Tsukuba (Japan)



