



CRESTA

CRAY HELPS PREPARE ECMWF'S INTEGRATED FORECAST SYSTEM FOR EXASCALE



European Centre for Medium-Range Weather Forecasts
Reading, UK
www.ecmwf.int

About ECMWF

The European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental organization supported by 34 states. It is both a research institute and a 24/7 operational service, producing and disseminating numerical weather predictions to its member states. ECMWF's supercomputer facility is one of the largest of its type in Europe.



Collaborative Research into Exascale Systemware, Tools & Applications
cresta-project.eu

About CRESTA

CRESTA (Collaborative Research into Exascale Systemware, Tools & Applications) is a collaborative research effort funded by the European Union exploring how to meet the exaflop challenge. The project has two integrated strands: one focused on enabling a key set of co-design applications for exascale, the other focused on building and exploring systemware for exascale platforms.



EPCC
University of Edinburgh
Edinburgh, Scotland
www.epcc.ed.ac.uk

About EPCC

The University of Edinburgh is one of the world's leading research universities. Through its supercomputer centre, EPCC, it is the project coordinator of the CRESTA project.



Cray Inc.
www.cray.com

About Cray

As a global leader in supercomputing, Cray provides highly advanced systems and solutions and world-class service and support to government, industry and academia. Cray technology enables scientists and engineers to not only meet existing and future simulation and analytics challenges but achieve remarkable breakthroughs by accelerating performance, improving efficiency and extending the capabilities of their most demanding applications.

The comprehensive Earth-system model developed at the European Centre for Medium-Range Weather Forecasts (ECMWF) in cooperation with Météo-France forms the basis for the centre's data assimilation and forecasting activities. All the main applications required are available through one computer software system called the Integrated Forecasting System (IFS).

IFS consists of several components: an atmospheric general circulation model, an ocean wave model, a land surface model and perturbation models for the data assimilation and forecast ensembles, producing forecasts from days to week and months ahead.

When running any high resolution model, keeping the computational costs within operational limits is an ongoing challenge. In particular, thermal radiation computations in IFS are expensive, taking approximately 10 percent of total compute time. Even this current level of cost is achieved by imposing limitations. ECMWF runs the radiation scheme only once every forecast hour and uses a radiation grid resolution which is coarser than that of the model grid.

As scalability will become more and more challenging on future supercomputers, ECMWF wanted to find a new approach for handling radiation computations. Their investigation into improving the scalability and performance of these computations was part of the CRESTA project's effort to develop the techniques and solutions that will address the challenges of computing at exascale.

"With the help of CRESTA, ECMWF continues to push the limit in state-of-the-art global numerical modeling and analysis software under the constraints of Earth-system model complexities on the one hand and the constraints and opportunities of efficient, massively parallel computing on the other," says Dr. Nils Wedi, head of the numerical aspects section at ECMWF.

ECMWF researchers identified that reorganizing radiation transfer calculations into a radiation-in-parallel scheme offered the potential for a sizable increase in efficiency for whatever parameterisation is called outside of the main stream. (Parameterisation in a weather or climate model refers to the method of replacing processes that are too smallscale or complex to be physically represented in the model by a simplified process.)



In the radiation-in-parallel configuration the radiation transfer calculations execute in parallel with the rest of the model using separate MPI tasks. From an MPI point of view, the radiation and model tasks have separate MPI communicators, while both can also use a global communicator for exchanging data between model and radiation tasks. An important and necessary requirement for this configuration is that the product of the radiation transfer calculations are returned to the model shifted by one radiation time-step.

The shift is necessary to allow the radiation transfer calculations to execute completely independently during the time the model executes a full time-step.

The main purpose of the radiation-in-parallel scheme is to improve the computational performance of an IFS model. To see how it performed, ECMWF researchers looked at two cases – a small TL159 model and a large TL3999L137 case.

Results

The ECMWF team used the Cray® XK7™ “Titan” system at Oak Ridge National Laboratory and a Cray® XC30™ system to test the performance of their radiation-in-parallel scheme. Testing the TL159 model on Titan using the radiation-in-parallel scheme (with radiation called every time-step), the team achieved 58,858 forecast days per day when using 1,024 MPI tasks for a 25 percent improvement. (In this study each MPI task uses eight OpenMP threads.) The TL3999 model used 22,624 tasks and saw a smaller performance improvement of 6 percent.

Additionally, the team tested the performance to be gained by colocating a model and radiation thread on the same core. As this task is only possible on processor architectures that support Hyper-Threading, the team used a Cray® XC30™ system for the test. Hyper-Threading provides the best performance when threads executing on the same core are using different resources of the core – for example, while one thread might be doing memory intensive operations, another on the same core is doing floating point operations. ECMWF enlisted Cray to provide the necessary scripting to produce the file needed to specify the detailed thread placement.

They ran the TL159 case on a single node using eight tasks with six OpenMP threads per task. Colocating model and radiation threads resulted in a 5 percent performance improvement for the radiation-in-parallel scheme. The TL3999 case ran on 512 nodes using 4,096 tasks with six OpenMP threads per task and saw an 8 percent performance improvement.

Overall, the team concluded that the radiation-in-parallel configuration shows strong potential for minimizing the cost of radiation transfer calculations.

“The need to run an operational high-resolution 10-day forecast model in under an hour means we must find a balance where the computations are used over the dynamics, physics, radiation, wave-model etc,” says ECMWF researcher George Mozdzyński.

“In short we need a balance which provides the optimal forecast skill within the one hour wall time limit.”



Reference

Technical Memorandum: 721, Mozdzyński, G. and J-J Morcrette.

“Reorganization of the Radiation Transfer Calculations in the ECMWF IFS”, April 2014.

http://old.ecmwf.int/publications/library/ecpublications/_pdf/tm/701-800/tm721.pdf

NEW APPROACH: Graphical representation of current (left) and radiation-in-parallel (right) configurations. In the current configuration radiation transfer calculations (shown in black) are performed every NRADFR model time-steps (shown in light grey), in this example case every 4 timesteps. In the radiation-in-parallel configuration radiation transfer calculations are performed at every timestep in parallel with the rest of the model.

