EGU21-4762
https://doi.org/10.5194/egusphere-egu21-4762
EGU General Assembly 2021


# New communication strategies in NEMO

**Italo Epicoco**[1,2], Silvia Mocavero[1], Francesca Mele[1], Alessandro D'Anca[1], and Giovanni Aloisio[1,2]
[1]Euro-Mediterranean Centre on Climate Change, Foundation, Italy {silvia.mocavero, francesca.mele, alessandro.danca}@cmcc.it
[2]University of Salento, Dep. Engineering for Innovation, Lecce, Italy {italo.epicoco, giovanni.aloiso}@unisalento.it

One of the main bottlenecks for NEMO scalability is the time spent performing communications. Two complementary strategies are here proposed to reduce the communication frequency and the communication time: the MPI3 neighbourhood collective communications instead of multiple point to point exchanges and the increasing of the halo region size.

NEMO performs Lateral Boundaries Conditions update by using four point to point MPI communications at north, south, east and west for each MPI domain. The model completes east-west exchange before performing north-south communications. The order of the exchanges allows us to preserve both 5-point and 9-point stencils. MPI3 neighbourhood collectives provide a way to have sub-communicators used to perform collective communications. Two different sub-communicators can be defined in order to support the two different stencils. A single MPI message is needed to be built for all neighbours instead of 4 different messages before calling the collective communication, while the received message is used to update the halo region, following the order of the neighbours in the sub-communicator.

The new communication strategy has been tested on two computational kernels (i.e. one for 5-point stencil and one for 9-point stencil), selected among the main relevant routines from the computational point of view. Preliminary tests, performed on a domain size of 3000x2000x31 grid points on the Zeus Intel Xeon Gold 6154 machine, available at CMCC, show a gain in communication time for the 5-point stencil use case up to 31% on 2016 cores. The improvement is reduced when communications with processes on the diagonal are activated. However, a modest gain is still achieved, depending on the number of cores.

On the other side, the analysis of some NEMO routines shows how the exchange of more than one row/column of halo would allow to move communications outside the routine, preserving data dependencies. A wider halo size reduces the frequency of message exchanges whilst increases the message size at each exchange. It allows us to adopt some optimisation strategies (i.e. loop fusion, tiling, etc.) to improve the data locality. Nevertheless, the use of a wider halo introduces itself some improvements for some kernels like for the MUSCL advection scheme which shows a gain of ~23% in the execution time comparing the original version and the new one with halo extended to 2 lines and the communication moved outside the computing region.

The current work has been performed according to the NEMO development strategy plan, defined

by the NEMO Consortium, which establish the priorities of the design strategies to reduce the bottlenecks to the scalability and the time to solution.

Acknowledgments

This work is co-funded by the EU H2020 IS-ENES project Phase 3 (ISENES3) under Grant Agreement number 824084.