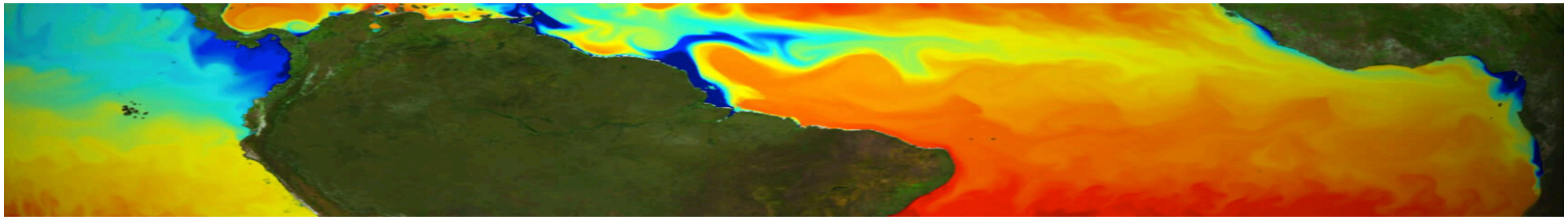
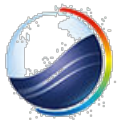


High Performance Computing
for

https://forge.ipsl.jussieu.fr/nemo/wiki/WorkingGroups/NEMO_HPC



Mercator Ocean
Ocean Forecasters

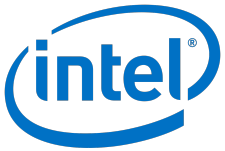
cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

Science & Technology Facilities Council

UNIVERSITY OF EXETER

National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

ECMWF



CRAY

NVIDIA

Atos

WG members

Active members:

Miroslaw Andrejczuk (Met Office/ST)
Mike Bell (Met Office)
Miguel Castrillo (BSC)
Louis Douriez (ATOS)
Matt Glover (Met Office)
David Guibert (ATOS)
Eric Maisonnave (CERFACS)
Sébastien Masson (CNRS/ST)
Francesca Mele (CMCC/ST)
Silvia Mocavero (CMCC/ST)
Andrew Porter (STFC)
Erwan Raffin (ATOS)
Oriol Tinto (BSC)

Former members:

Jeremy Appleyards (NVIDIA)
Mondher Chekki (Mercator-Ocean)
Tim Graham (Met Office)
Cyril Mazauric (ATOS)

Others members:

Lucien Anton (CRAY)
Tom Bradley (NVIDIA)
Clement Bricaud (MERCATOR/ST)
Marcin Chrust (ECMWF)
Peter Dueben (ECMWF)
Marie-Alice Foujols (CNRS)
Jason Holt (NOC)
Dmitry Kuts (Intel)
Michael Lange (ECMWF)
Claire Levy (CNRS/NEMO Project Manager)
Julien Le Sommer (CNRS)
Gurvan Madec (CNRS/NEMO Scientific Leader)
Yann Meurdesoif (CEA)
Kristian Mogensen (ECMWF)
Stan Posey (NVIDIA)
Martin Price (Met Office/ST)
Martin Schreiber (Univ of Exeter)
Kim Serradell (BSC)
Nils Wedi (ECMWF)

2018 on-going activities

NEMO benchmark (CNRS, CERFACS)

Development of a benchmark to identify the bottlenecks to NEMO scalability:

- idealised geometry (removing Ice model, IO and MPI collectives), realistic parameters at different resolutions (1, 0.25 and 1/12 degree)
- internal timings, parametrisable communications

Results:

- Scalability tests (ORCA-1 like) on Intel Broadwell Meteo-France machine; definition of the scalability limit (**~7x7** subdomain size)

Next steps:

- Scalability tests on ORCA-025 and ORCA-12
- Scalability tests on different machines (ATOS, BSC, CMCC)

2018 on-going activities

Intra-node performance (CMCC, MetO)

Analysis at kernel level to understand which kernels are limiting real (compared with peak) performance; improving utilisation of memory hierarchy; loop vectorisation. Investigation and integration of a second level of parallelism based on the shared memory paradigm

Improving utilisation of memory hierarchy:

- cache blocking application on tracers advection loops (trunk@rev8676)
- optimisation of the block size for each loop

Tests with different domain sizes and compilers:

- **~7.5%** (Intel2018, O3) and **~12.5%** (Intel2013/15, O2) gain which decreases with the MPI subdomain size

Next step:

- Refactoring of the kernel code to extend the loop size (tessellated tiling approach proposed by MetO)

2018 on-going activities

Intra-node performance (CMCC)

Improving loop vectorisation:

- Introduction of SIMD directives in the most expensive kernels

Results:

- **~8%** gain on the total elapsed time
- **~11%** gain on the total elapsed time (ocean dynamics of the GLOB16 real configuration)

Hybrid parallelization:

- Coarse-grained approach on tracers advection

Results:

- Improvement at internode level similar to the vertical physics (ORCA2/ Ice/Bio, **~7%** of the parallel efficiency on 512 cores compared with the pure MPI version)
- It is not enough to justify the code refactoring

2018 on-going activities

Inter-node performance

Limiting the communication overhead

- introducing arbitrary width haloes and thereby reduce the necessary frequency of boundary exchanges (NOC)
- reducing global communications (CMCC)
- investigation of MPI3 collective neighbours communications (CMCC)

I/O performance improvement (MetO)

Investigation of techniques to reduce the I/O overhead: extension of the use of XIOS for reading/writing restart files (available in the 4.0 beta-release)

Results:

- Tested on MetO system (eORCA025, eORCA12), no relevant gain
- Tested on CMCC system (GLOB16, only restart reading), reduced time variability in input reading routines, **6%** gain on execution time (3days simulated)

Next steps:

- Finish implementing XIOS for restart and diagnostic IO

2018 on-going activities

Performance portability on heterogeneous architectures (STFC)

Investigation of a light-DSL approach to apply DSL in NEMO without impacting on the NEMO coding structure rules:

- Processing the NEMO code to create an internal representation compliant to the PSyclone tool
- Manipulating the intermediate code to perform PSyclone transformations

Reduced precision (BSC)

Optimization of the NEMO model by using a mixed precision approach:

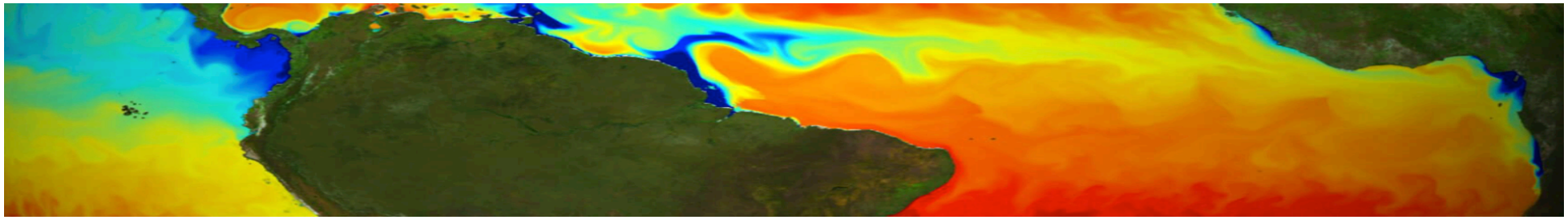
- Study on the precision needed by the different processes in NEMO (implement the emulator, design and prepare the tests, execute the tests, assimilate the results)
- Implementation of the mixed precision approach in NEMO

Results:

- Tests on GYRE1 (NEMO3.6), 89% of the variables could use reduced precision

Next steps:

- Tests on ORCA grid



THANKS!