www.bsc.es

Barcelona, 25th October 2016

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

THE MULTIFACETED ASPECTS OF MODEL PERFORMANCE

IS-ENES2: Crossing the Chasm Workshop

Francisco Doblas-Reyes
Kim Serradell

MODEL PERFORMANCE

“Castells” Contest at Tàrraco Arena Plaça de Tarragona (October 2016)

Small and “trivial” modifications

Big Projects and challenges

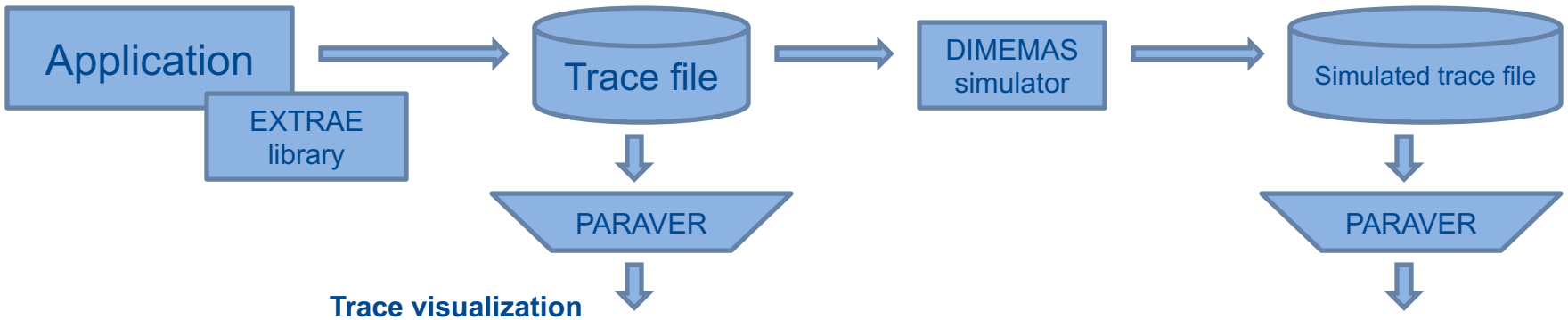
BSC Performance tools

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

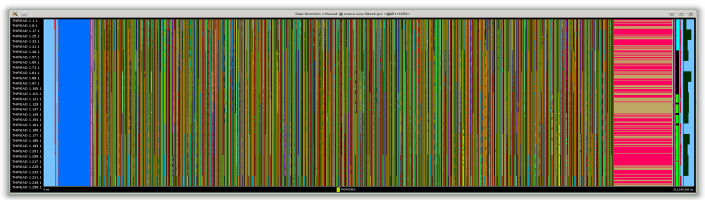
EXCELENCIA SEVERO OCHOA

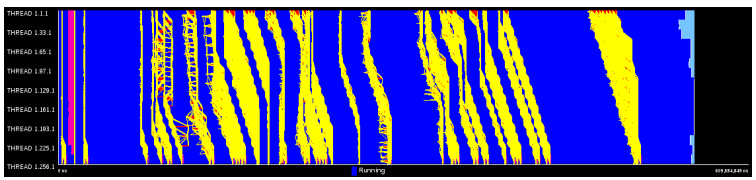
BSC-CNS performance tools:

- EXTRAE: Trace generation package.
- PARAVER: Trace visualization and analysis tool.
- DIMEMAS: Simulation tool for analysis on a configurable target platform.

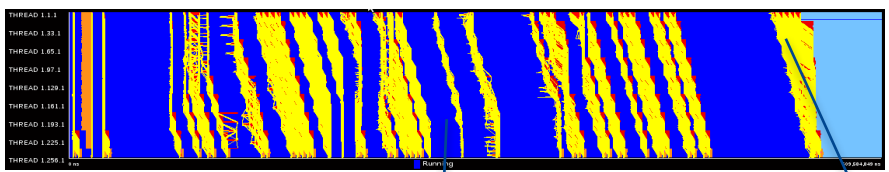


Trace visualization





DIMEMAS generated trace. Target = ideal machine

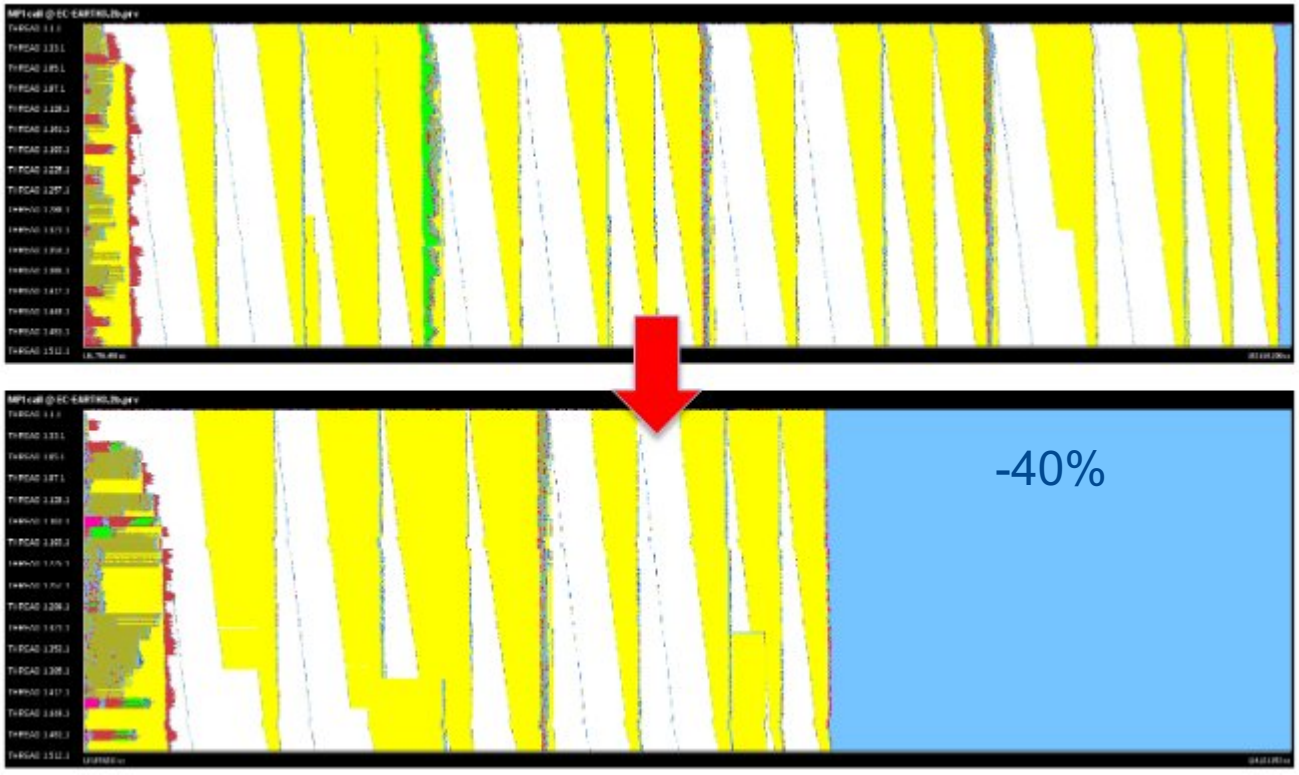


Blue regions = computation

Yellow lines = comunication

Performance analysis

- Optimize coupling strategy between IFS and NEMO
  - Analysing the coupling strategy between IFS and NEMO Aggregation of message and calculations (by default IFS passes variables to NEMO in several instances)



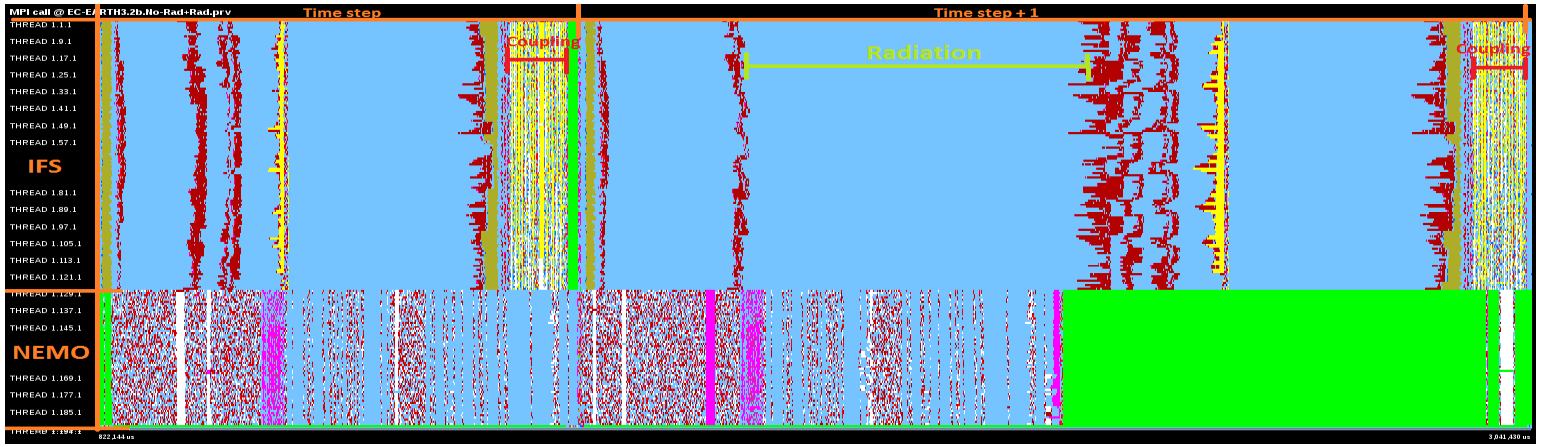
4 groups of variables

2 groups of variables

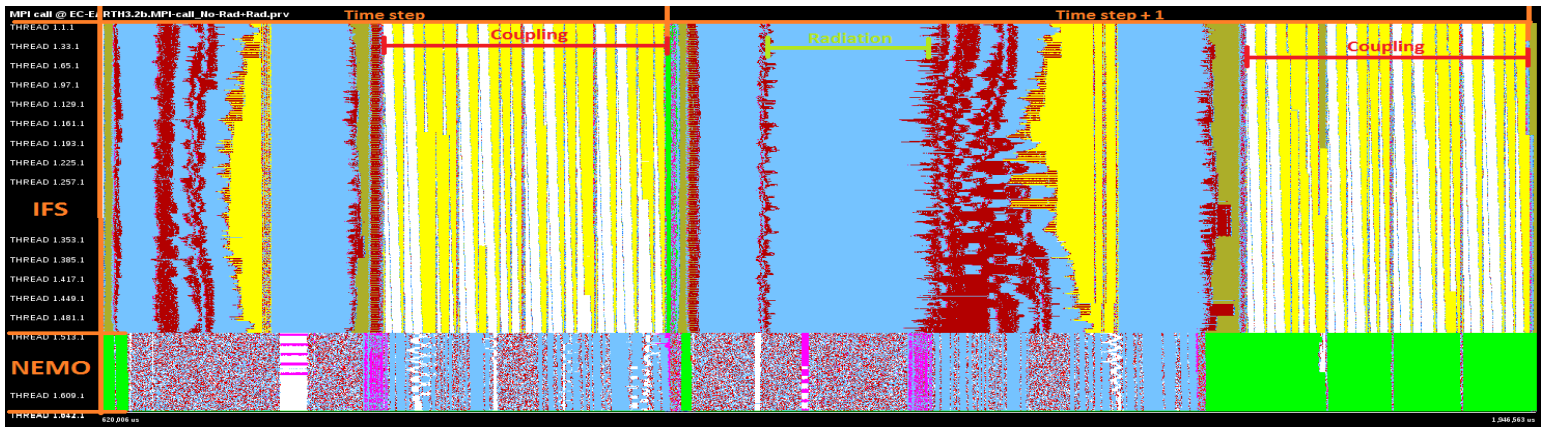
Top: using default namcouple file
Bottom: using an optimized namcouple file (pack variable groups with similar type of coupling together)

Coupling in high resolution

128 MPI processes for IFS

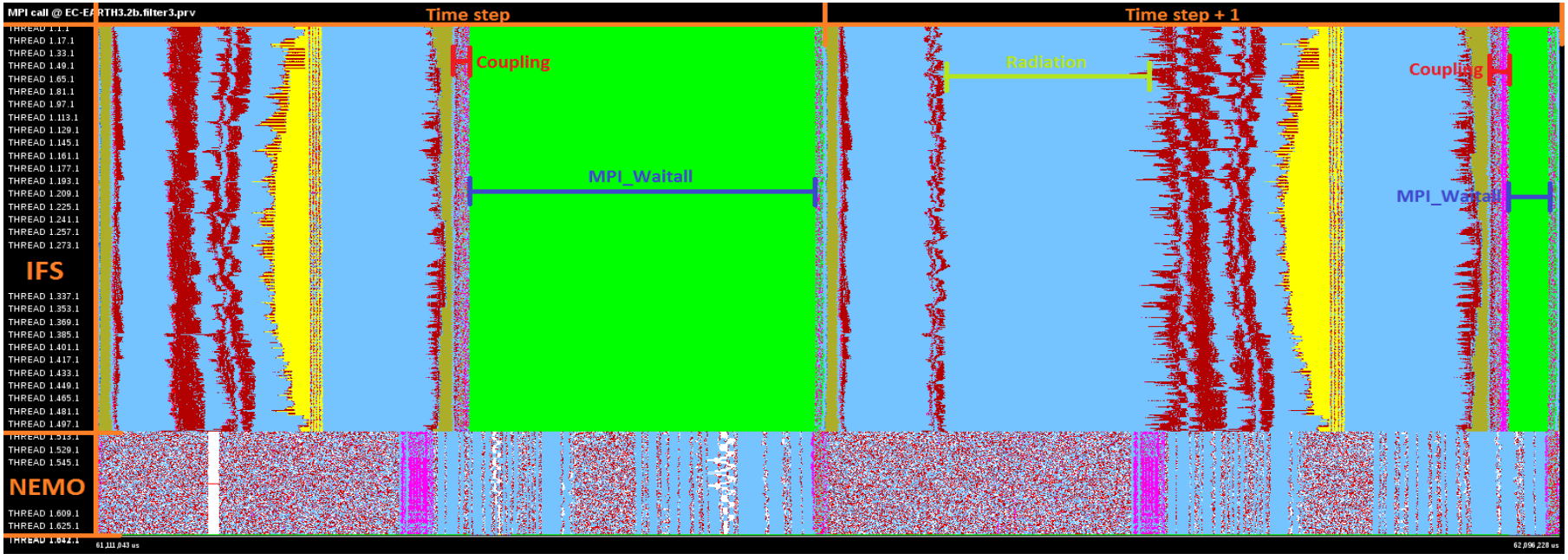


512 MPI processes for IFS



- Different traces to see the impact when scaling the number of cores
  - We can see with more cores the coupler phase is not scaling and takes longer to complete

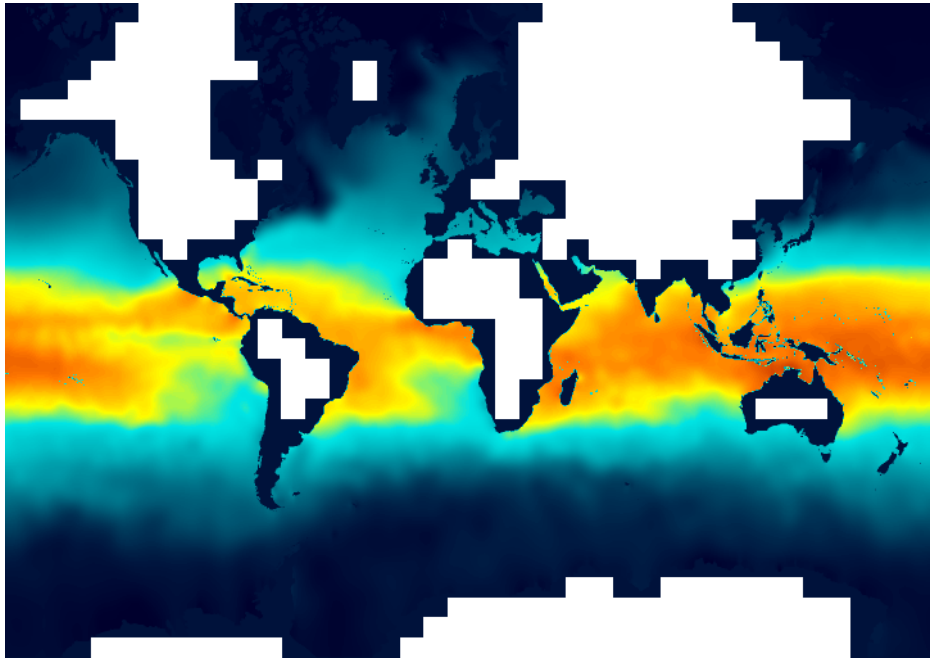
Optimizing coupling in high resolution



- Option for CONSERV operation has two implementations
  - communications all-to-one/one-to-all using MPI_send/MPI_irecv
  - collective and non-blocking communications MPI_igather/MPI_iscatter

Exclude land-only processes in NEMO

- ELPiN: A tool that allows finding adequate namelist parameters to exclude land-only processes in NEMO simulations
- NEMO decomposes automatically the domain
  - Computes and communicates in land-only processes and then discards the result
- Currently performing an evaluation of the results produced

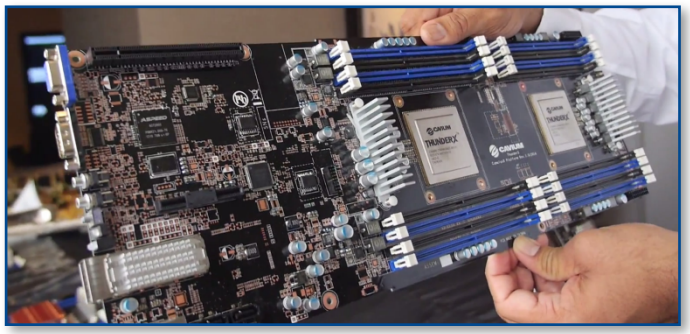


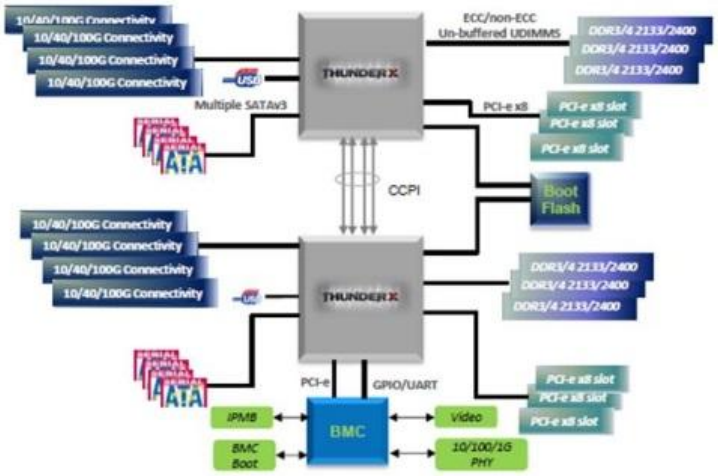
- ORCA025 domain decomposed in 1287 sub-domains
- 312 are land-only and therefore removed (24% of the total grid)

EC-Earth in ThunderX

- One of the Mont-Blanc mini-clusters.
  - Small clusters including ARMv8 (64 bit) platforms.

- 4 nodes devoted to computation, each equipped with:
  - 2x sockets Cavium ThunderX
    - 48x ARMv8-A cores each (i.e. 96 cores with shared mem each node) @ 1.8 GHz
  - 128 GB memory
  - 2x 10GbE (data network)
  - 1x 40GbE (not connected)
  - 1x 128GB SSD





EC-Earth in ThunderX

- EC-Earth 3.2 built with:
  - GCC 4.8.4, SZIP 2.1, OPENMPI 1.10.2, NETCDF 4.4.0, HDF5 1.8.17, LAPACK 3.6.0
  - T255L91-ORCA1 configuration
- Successful run of one month of simulation (output included)
- First execution times needs to be improved
  - 10.508 seconds using 258 cores (128 IFS, 128 NEMO, 1 XIOS, 1 runoff)
  - 923 seconds using the same number of cores in MN3
  - Work required to improve these numbers

OmpSs programming model

- Extend OpenMP with new directives to support asynchronous parallelism and heterogeneity
  - Overlap communication and computation
  - Apply Dynamic Load Balancing library

- Two stages compilation
  - Mercurium compiler which introduces OmpSs runtime
  - Usual compiler (gcc, intel, ibm) to generate the executable

- Previous work done
  - EC-Earth 3.1 compiled and executed with OmpSs
    - NEMO is not OpenMP ready so not really useful
  - Moved to OpenIFS 40r1
    - Bugs with mercurium compiler not yet solved
    - Some corrections reported to OpenIFS developers to make the code ready for Mercurium

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Thank you!

For further information please contact
kim.serradell@bsc.es