NEMO Benchmark configuration

Eric Maisonnave (CERFACS)

Sébastien Masson (LOCEAN)

April-June 2018

Questions

- Is scalability bounded by extra computations, MPI communications or load imbalance ?
    - Model needs to be simplified to remove ice model, IO and MPI collective calls (~light e-ORCA)
    - NEMO is instrumented to measure separately the 3 effects
- Is the scalability limit the same at 1, 0.25 and 1/12 degree ?
    - 3 namelists to reproduce realistic physics of 3 resolutions
- Can the result be reproduced on various machines ?

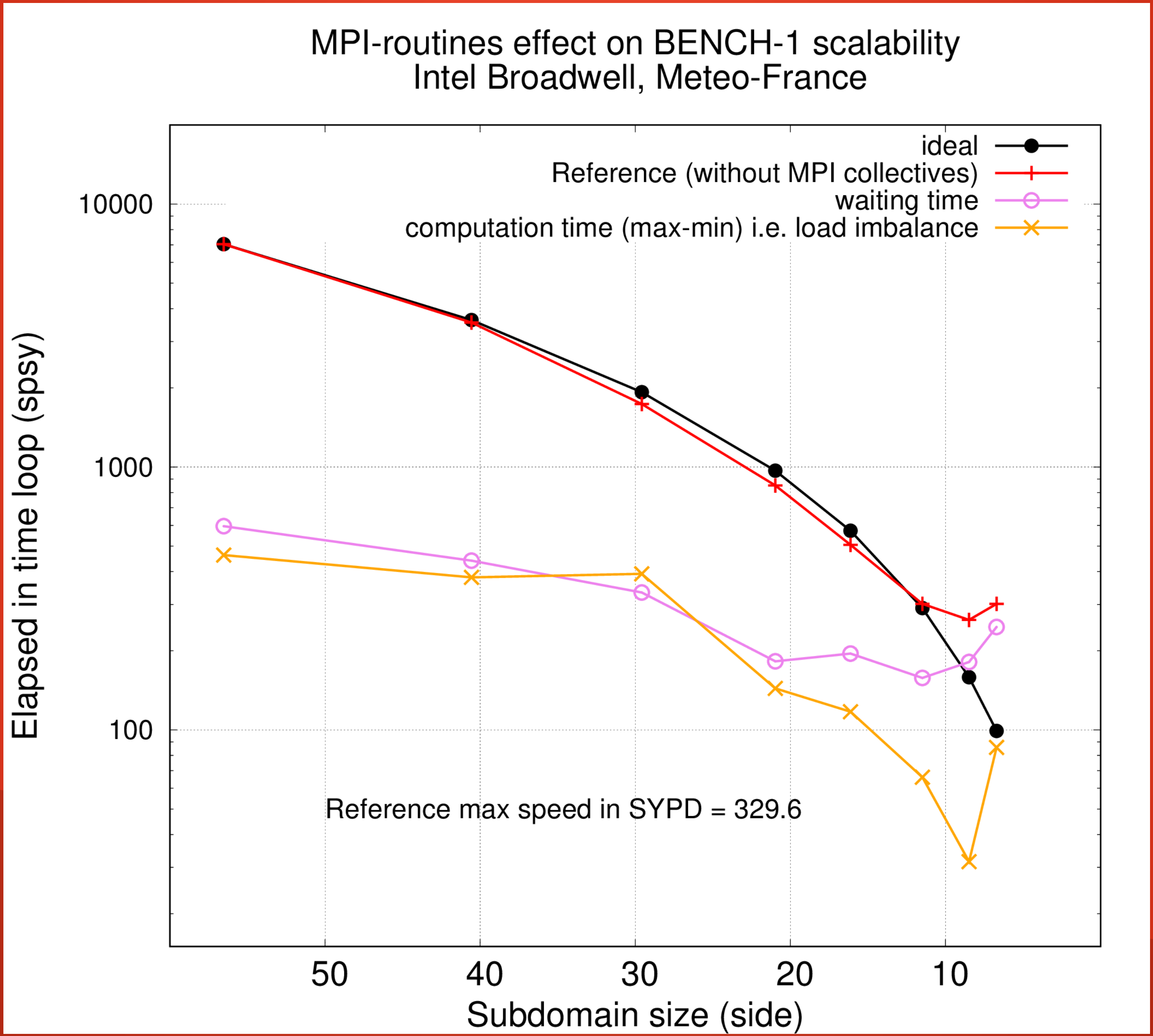
„BENCH“ implementation

- Starting from branch dev_r9759_HPC09_ESIWACE

```
svn co http://forge.ipsl.jussieu.fr/nemo/svn/NEMO/branches/2018/dev_r9759_HPC09_ESIWACE
```

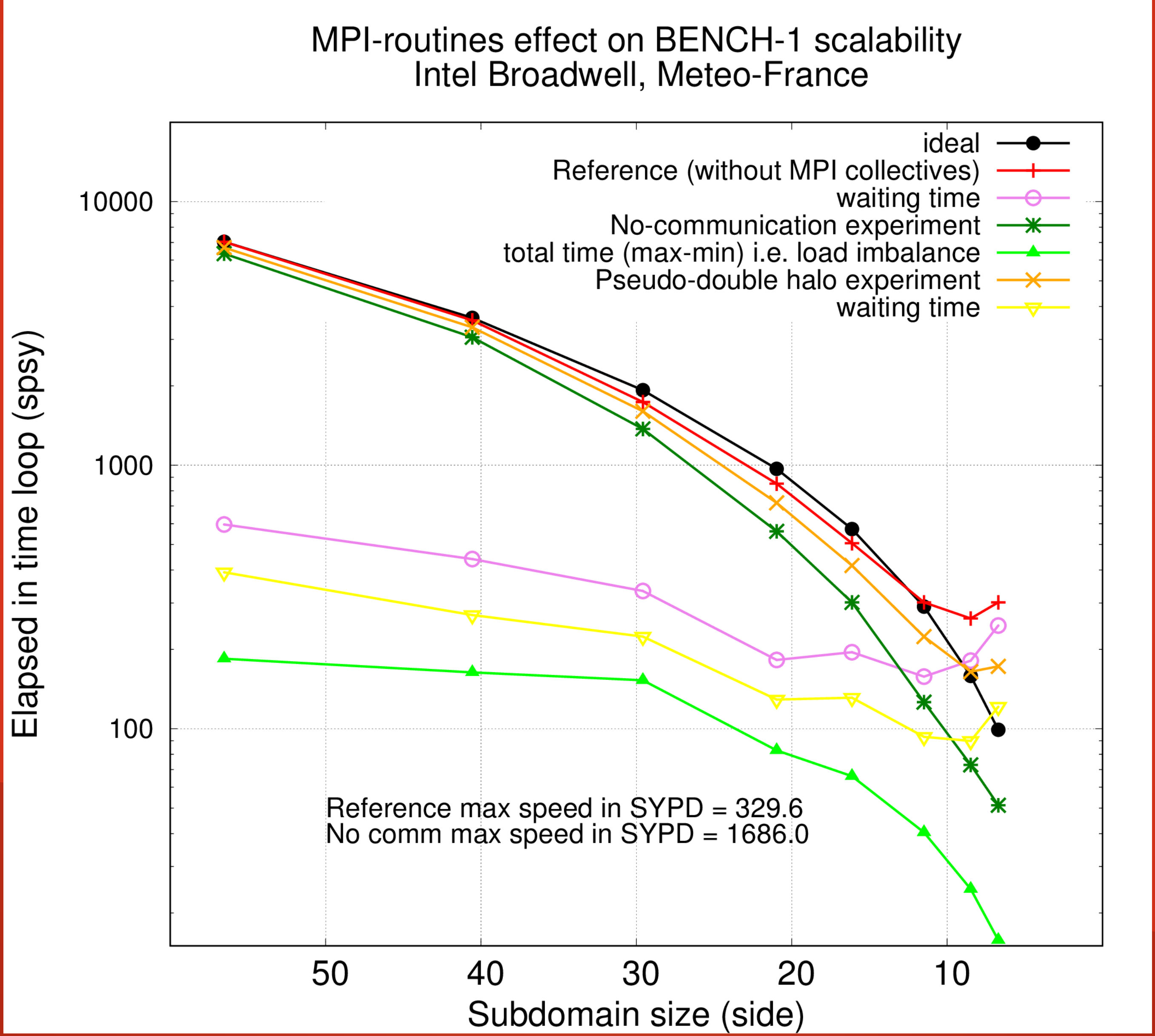
- „quiet“ e-ORCA (no risk of numerical explosion)
- 1 timer (MPI_Wtime) between 2nd and n-1 time step
  - Removes init/end (contributes to save CPU during benchmarking)
- Add MPI_Wtime before and after halo exchanges
  - Inner timing: „waiting“ time (MPI comm + load imbalance)
  - Outer timing: „computation“ time (spread: ~ computational load imbalance)
- Modular frequency of halo exchange call (possibility to mimic double sized halos, to avoid any communication ...)
- Identify # and size of halo exchanges to possibly replace time steping by halo exchange only

First Results



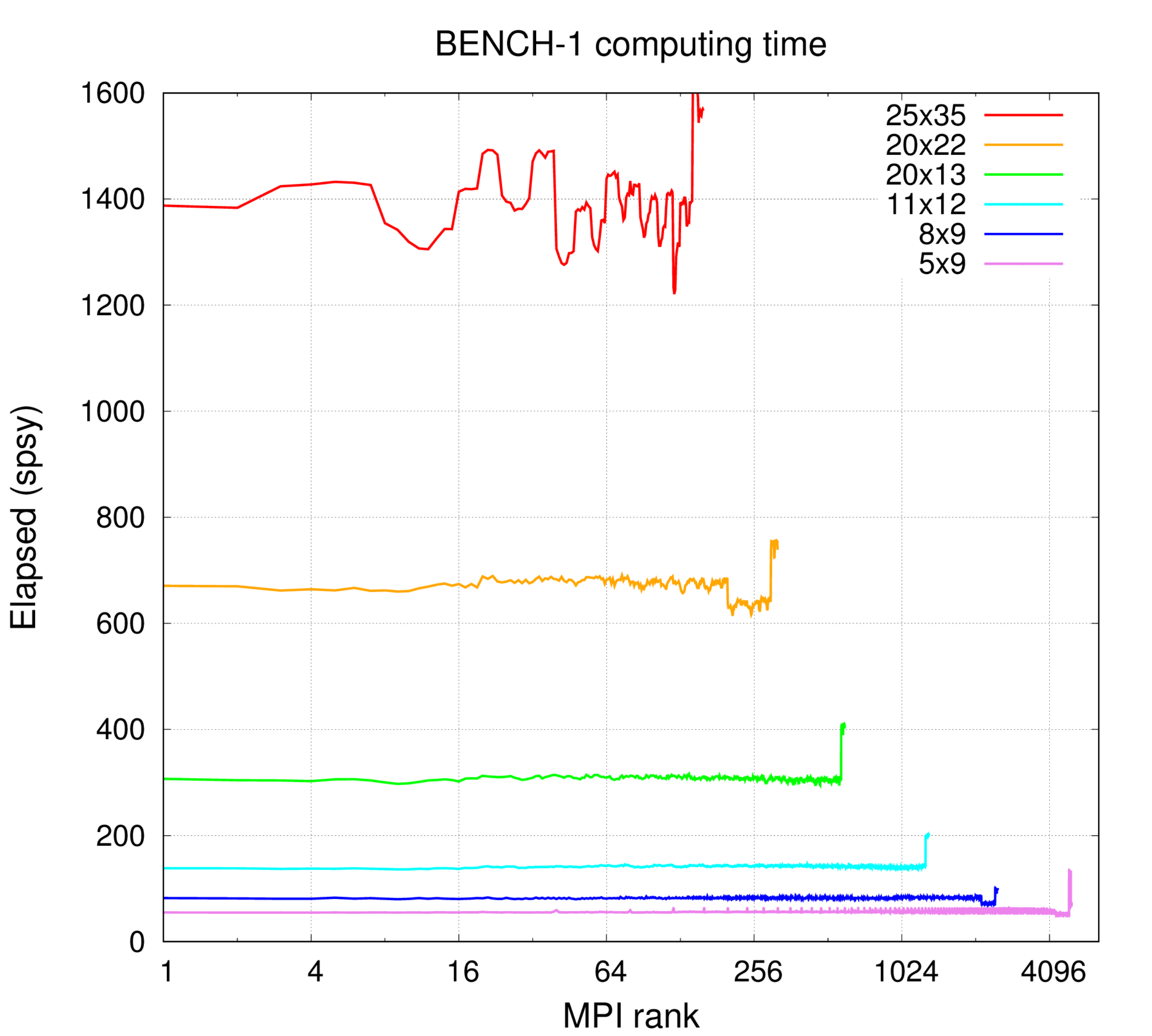
BENCH-1 (ORCA-1 like) scalability and limit „attribution“

First Results



Same experiment with halved or zero halo exchanges

First Results



Subdomain size

Visualisation of computation time spread, fn(parallelism)

Discussion

- Much (!) more results (for 1, 1/12 and 12° configs)
- Is instrumentation able to guide future improvements ?
- e.g. can we evaluate potential impact of double halo ?
- Where should/could we reduce lbc_lnk calls ?
- Can we reproduce the results on several machines ?
- Can we extend the exercise to build a strategy for IO/ice model ?

- instrumented code + namelists + launching scripts + gnuplot scripts available (branch dev_r9759_HPC09_ESIWACE)