INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Final General Assembly

16-18 January 2023
Toulouse

NEMO: Adaptations for High Performance Computing

Italo Epicoco (CMCC)
NEMO-HPC Working Group

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

NEMO-HPC WG participants

UKRI Science and Technology Facilities Council
Hartree Centre

Met Office

ECMWF

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

MERCATOR OCEAN INTERNATIONAL

Sciences de l'environnement
Institut Pierre Simon Laplace

AtoS
Bull

GENCI
HPC at the service of knowledge

CERFACS
CENTRE EUROPÉEN DE RECHERCHE ET DE FORMATION AVANCÉE EN CALCUL SCIENTIFIQUE

NVIDIA

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

NEMO Development strategy: Adaptations for High Performance Computing

Current HPC limitations in NEMO

- Lack of a per-thread parallelism
- High memory access limits a full exploitation of the computational resources
- Lack of a GPU-based implementation
- HPC optimizations and code transformations often clash with readability/maintainability
- Many optimizations have been introduced in recent years, but still need to be consolidated and improved.

IS-ENES3 has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

Priorities for 2023-2027

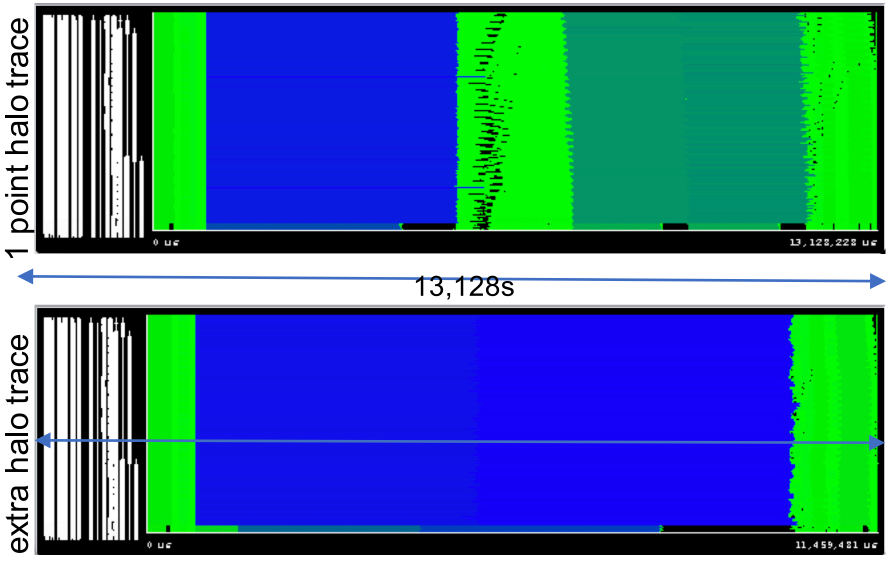
- Consolidating the recent optimizations (Extended halo, Tiling, memory footprint, loop bounds, …)
- Enhance the single node performance through
  - 3D tiling to better exploit cache memory
  - Exploit the mixed precision approach to increase the arithmetic intensity and reduce the memory footprint
  - Make use of an even wider halo (3 or more points) where needed
- Develop a per-thread parallelism
- Develop GPU-oriented parallelism
- Make use of DSL (PSyclone) approach to
  - Automatically apply HPC transformations without changing the developer interface (i.e., loop-fusion)
  - Generate OpenMP/OpenACC version of the code

IS-ENES3 has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

NEMO Performance Analysis: Extended halo

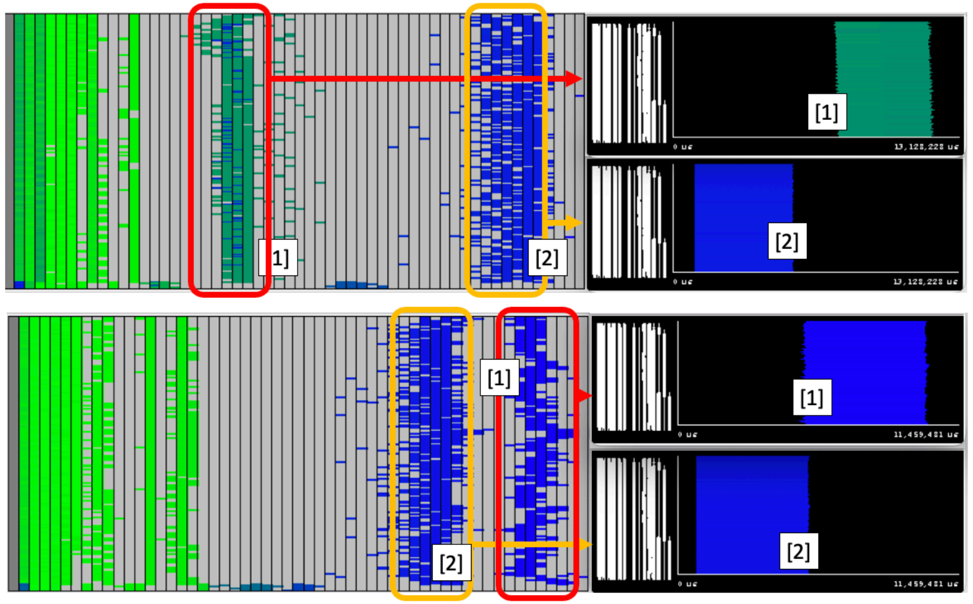
- ✓ ORCA12-like configuration from BENCH TEST case
- ✓ best domain decomposition: 1536 (48 x 32) cores (32 nodes on MN4)
- ✓ MPI subdomains 94 x 103 grid points

Timeline of Useful duration



11,459s

Histogram



- **elapsed time with extra halo is ~13% lower**
- halo 2 have larger computational blocks
- second part of the iteration [1] have smaller blocks of computation in the exp 1
- computational time of the first part of iteration [2] decreases moving from halo 1 to halo 2

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Final General Assembly

16-18 January 2023 Toulouse

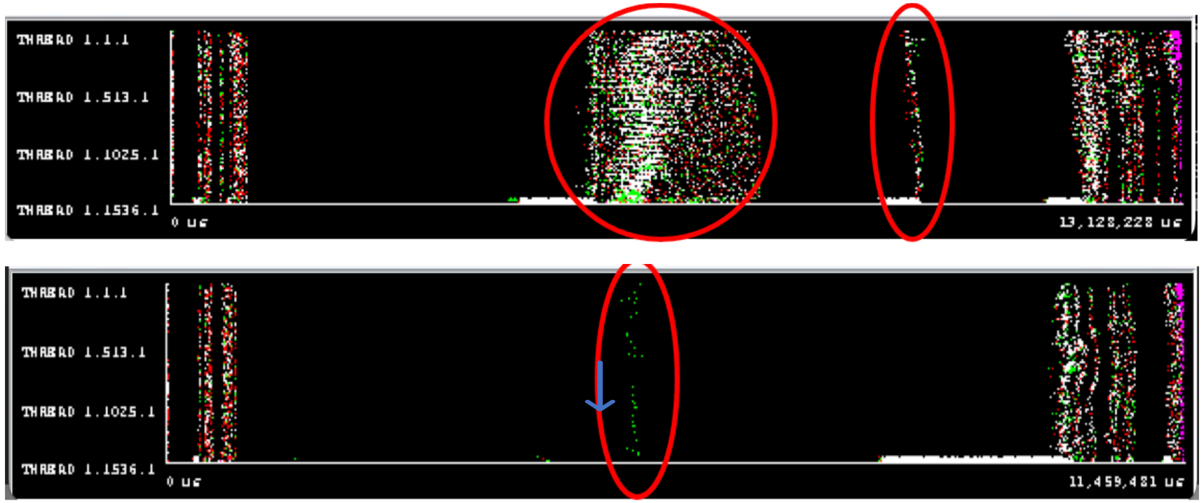
NEMO Performance Analysis: Extended halo

- The MPI calls are drastically reduced with extra halo, especially in the central part of the iteration
- communication time reduced ~3x for lbc_lnk and ~2.7x for lbc_nfd with extra-halo

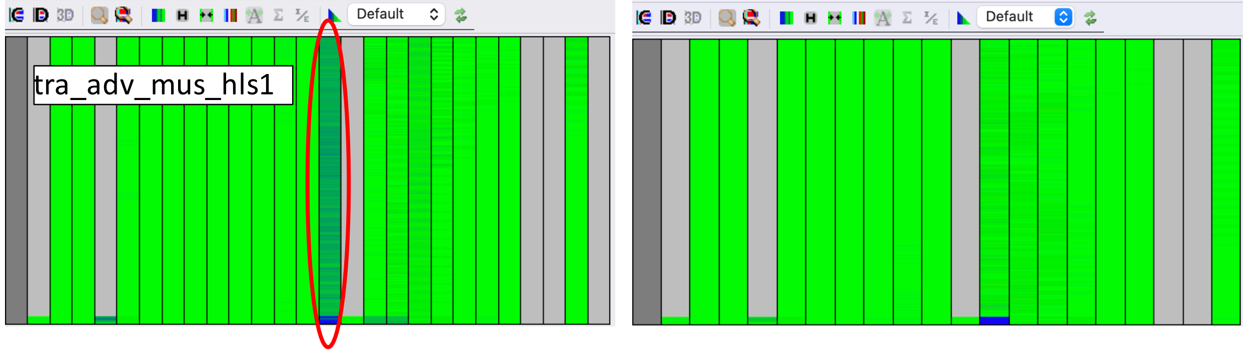
Timeline of MPI call

1 point halo trace

extra halo trace



Histogram of the MPI caller - third level



IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Final General Assembly

16-18 January 2023 Toulouse

NEMO Performance Analysis: Communications; LoopFusion

MPI3: neighbour collectives

Timeline of MPI call

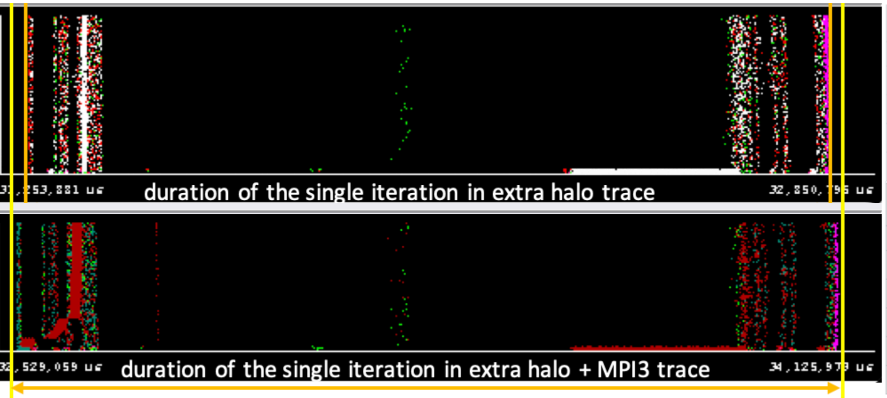


Table of MPI call

	MPI_Recv	MPI_Isend	MPI_Irecv	MPI_Waitall	MPI_Allreduce
Total	665,280	680,960	15,680	182,784	1,536
Average	433.12	443.33	326.67	119	1
Maximum	440	666	336	336	1
Minimum	330	330	224	112	1
StDev	26.63	43.17	30.96	38.97	0

	MPI_Isend	MPI_Irecv	MPI_Wait	MPI_Waitall	MPI_Allreduce	MPI_Ineighbor_alltoallv
Total	15,680	15,680	172,032	91,776	1,536	168,960
Average	326.67	326.67	112	59.75	1	110
Maximum	336	336	112	318	1	110
Minimum	224	224	112	50	1	110
StDev	30.96	30.96	0	47.01	0	0
Avg/Max	0.97	0.97	1	0.19	1	1

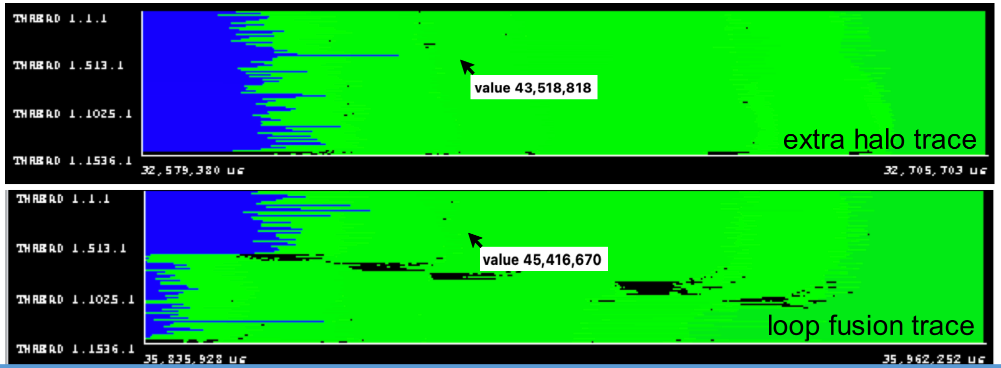
- ✓ The execution with the point-to-point communications was a little bit faster
- ✓ higher MPI message size due to MPI_Ineighbor_alltoallv
- ✓ the number of MPI calls has been reduced globally

no much benefit from the use of MPI3

LoopFusion

- ✓ few routine fused:
  - dyn_ldf_lf: iso-level harmonic operator
  - tra_adv_fct_lf: FCT advection scheme
- ✓ big improve not expected:
  - no difference in terms of duration
  - no difference in MPI calls
  - but…
  - a little increase in instructions
  - a little improve in cache misses

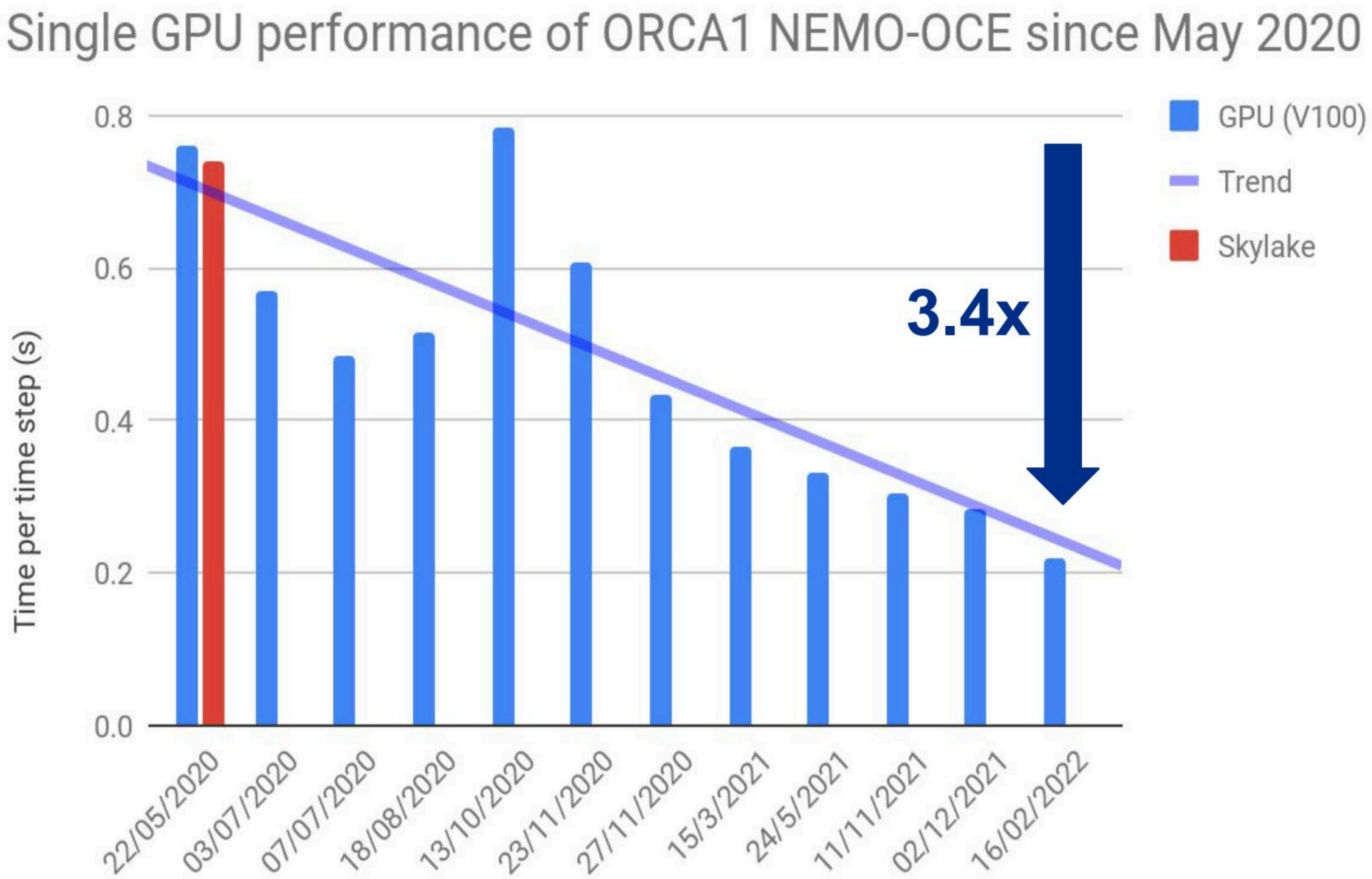
Zoom on the timeline of Useful Instruction



Zoom on L3 Cache misses

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

GPU Performance of PSyclone-processed NEMO



- 250 source files
- 87K lines of code
- Adds 4.3K OpenACC directives
- Some manual tweaks still required

UKRI Science and Technology Facilities Council
Hartree Centre

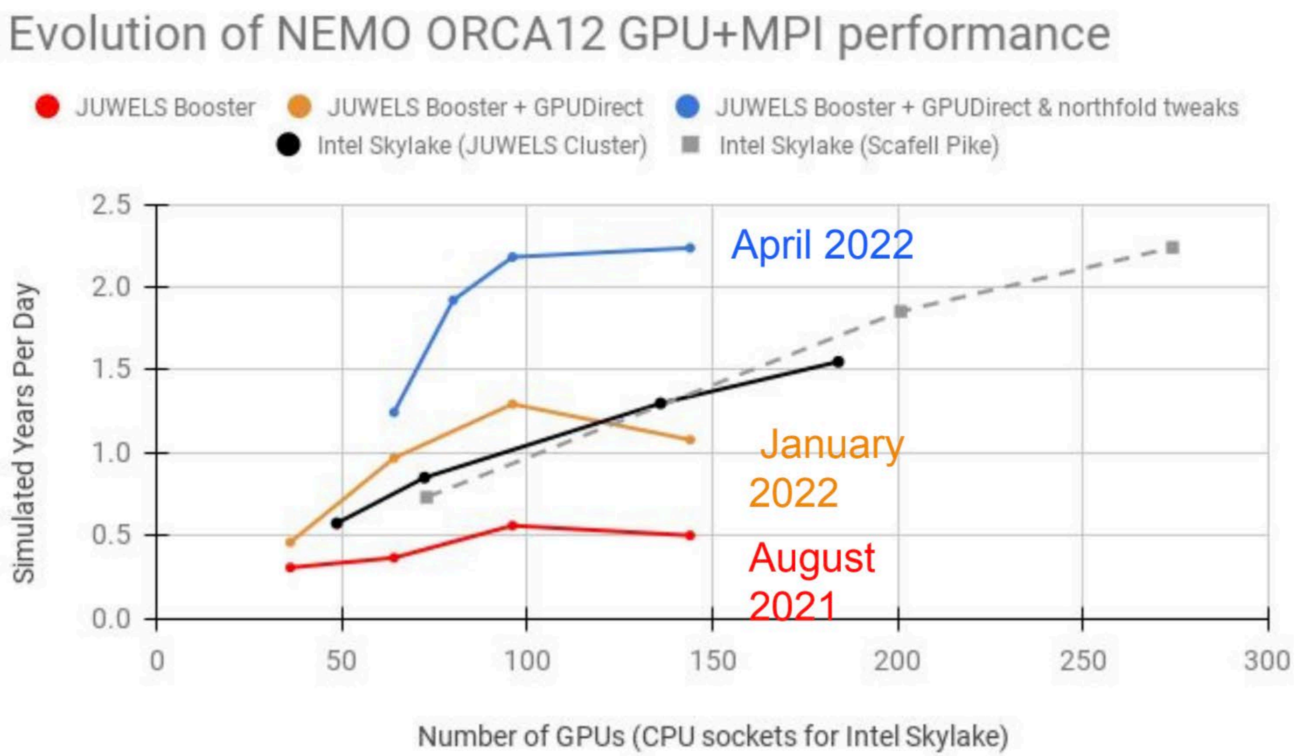
Using 4.0.2 of NEMO & G08 configuration provided by MO.

IS-ENES3 has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

Multi-GPU Performance of PSyclone-processed NEMO

Large-scale resources accessed through **ESiWACE2**

Run on up to **192 GPUs** on **Marconi** (V100) and JUWELS Booster (A100)



Science and Technology Facilities Council
Hartree Centre

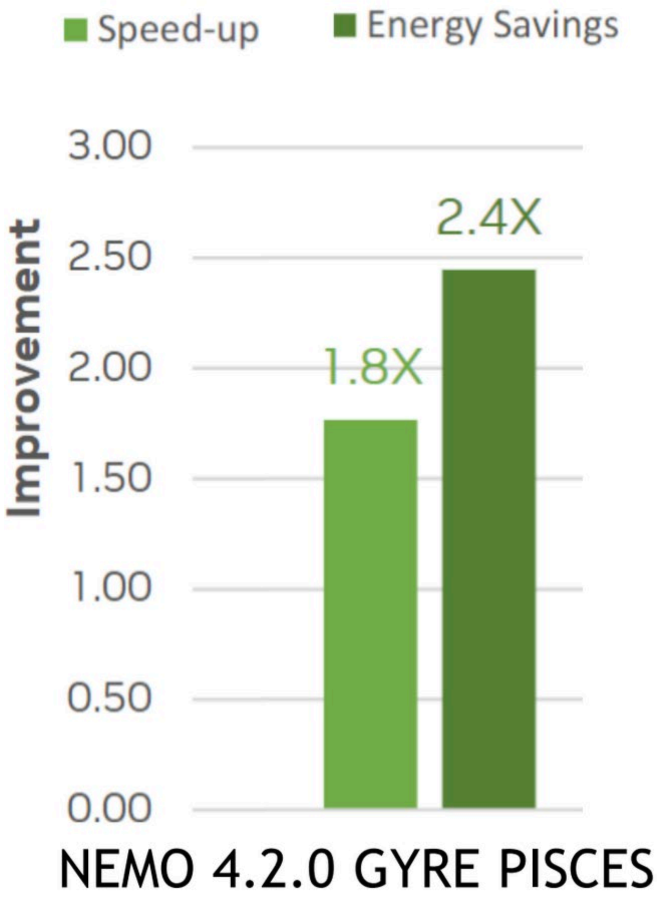
esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHE AND CLIMATE IN EUROPE

IS-ENES3 has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

NEMO Performance on NVIDIA Grace

NEMO Projections of Speed-up and Energy Savings

- Configurations for comparisons:
  - 2 x AMD EPYC 7763, 64 Zen 3 cores each, 128 cores total
  - 2 x NVIDIA Grace Arm, 74 cores each, 144 cores total
- Runtimes projected from actual results
  - 2 x AMD: runtime = 130 sec (actual); BW = 328 GB/s
  - Graviton3: runtime = 215 sec (actual); BW = 260 GB/s
  - 2 x Grace: runtime = 74 sec (projected); BW = 760 GB/s
- Speedup projection based on GV3 runtime:
  - GV3-based = 215 sec / (760 GB/s / 260 GB/s) = 74 sec
  - AMD-based = 130 sec / (760 GB/s / 328 GB/s) = 56 sec
  - Grace speedup vs. AMD = 1 / (74 sec / 130) sec = 1.8x



NVIDIA

unding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824084

Plans for the next future

- Mixed precision
  - Generalization of the AutoRPE tool + automation of parsing tool that can be integrated into the continuous Integration pipeline used in the NEMO GitLab.
  - Integration of mixed-precision version of NEMO4.0 on an ORCA12 grid with IFS-NEMO (Destination Earth framework)
  - Mixed-Precision version of NEMO4.2 implemented in the NEMO (ORCA36) component of the Ocean Twin (EDITO project)
  - Explore half-precision
- GPU porting
  - To explore the porting of some modules in NEMO to GPUs, targeting MN5, LUMI, Leonardo

BSC

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Plans for the next future

Met Office

- Improvements to tiling performance
  - Vertical tiling
  - Removal of halo calculations
  - Implement OpenMP-compatible solution for tiling overlap issue
- Scope OpenMP parallelisation of tiling
- Extend the tiling approach to other NEMO modules TOP/MEDUSA
- Applying PSyclone to NEMO 4.2 & GOSI9
- Apply BSC mixed precision tool to NEMO 4.2 & GOSI9

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Plans for the next future

- Integration of PSyclone DSL into the NEMO compilation chain (strong collaboration with STFC and MetOffice)
  - Enhancement of PSyclone with new transformations tailored for NEMO (e.g. loop fusion)
  - Performance evaluation of the OpenMP OpenACC version of NEMO produced by PSyclone and developing of eventual improvemets for the automatic code parallelization
- Evaluation of AutoRPE tool for obtaining a mixed precision configuration of NEMO at CMCC (strong collaboration with BSC)
  - Preliminary investigation of mixed precision based on Stocastichal Arithmetic by means of CADNA library
- Development of performance monitoring service for NEMO (collaboration with BSC)

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

The Consortium

Coordinated by **CNRS-IPSL**, the IS-ENES3 project gathers **22 partners** in **11 countries**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084

Our website:
https://is.enes.org

Follow us on Twitter!
@ISENES_RI

Contact us at
is-enes@ipsl.fr

Follow our channel!
IS-ENES3 H2020