



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



EXCELENCIA
SEVERO
OCHOA

Scaling NEMO4 I/O using the new ORCA36 configuration

Miguel Castrillo

BSC-ES Performance Team, Computational Earth Sciences

NEMO HPC Group

28/07/2020

ORCA36 scaling



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

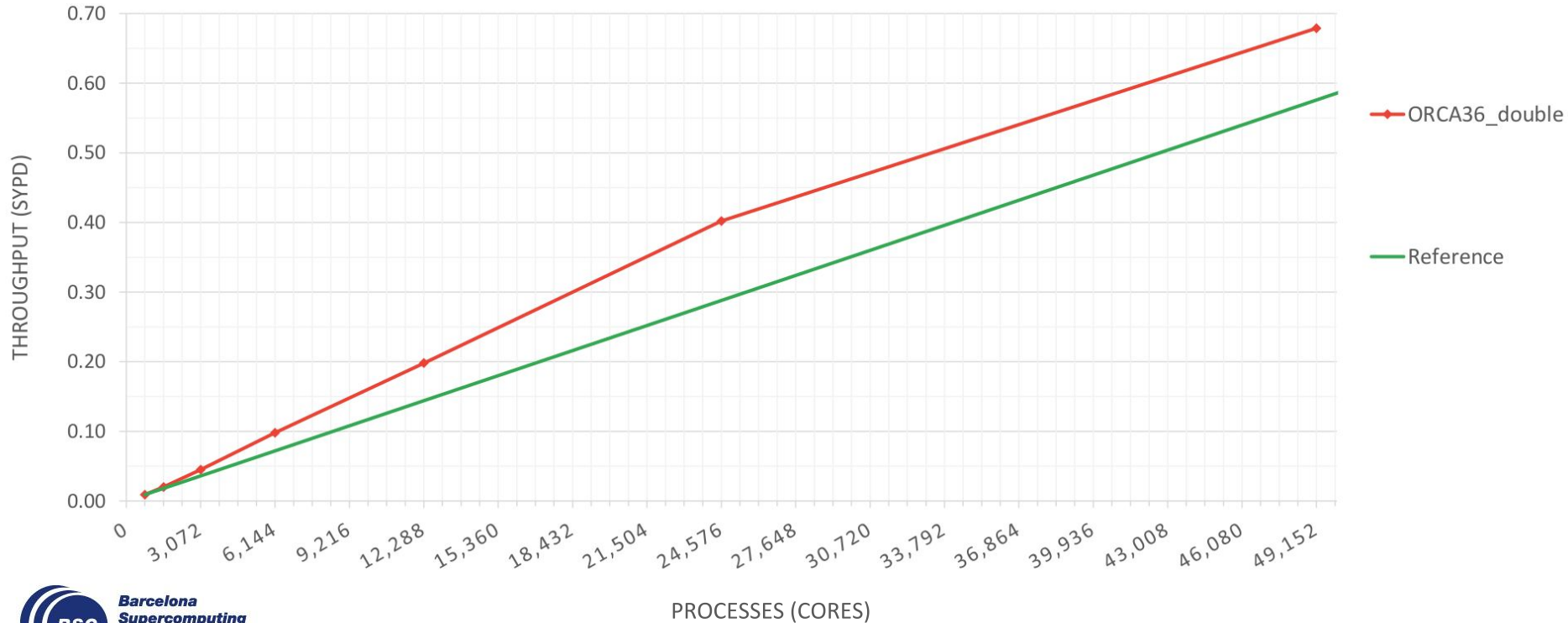
ORCA025 scalability (MN4)

ORCA025 scalability



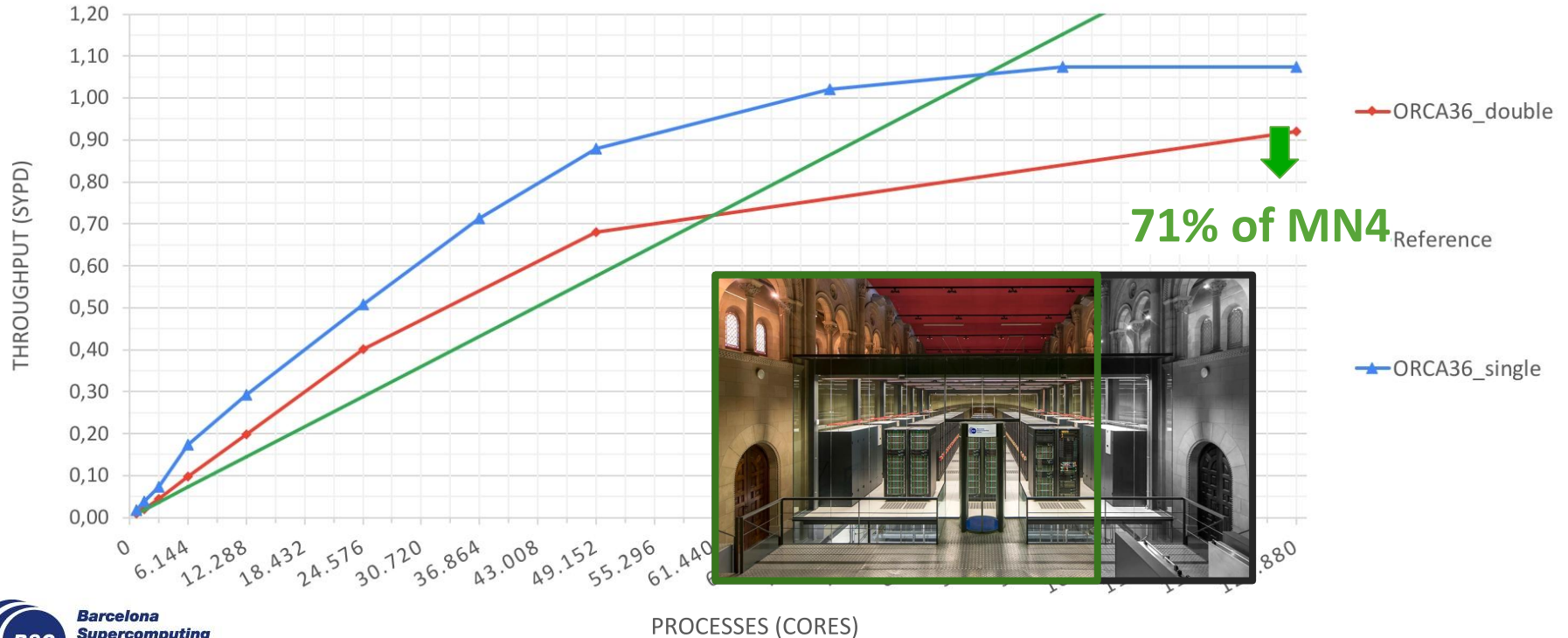
ORCA36 scalability (MN4)

ORCA36 scalability



ORCA36 scalability (MN4)

ORCA36 scalability – Double precision vs Single precision – Grand challenge (2019)



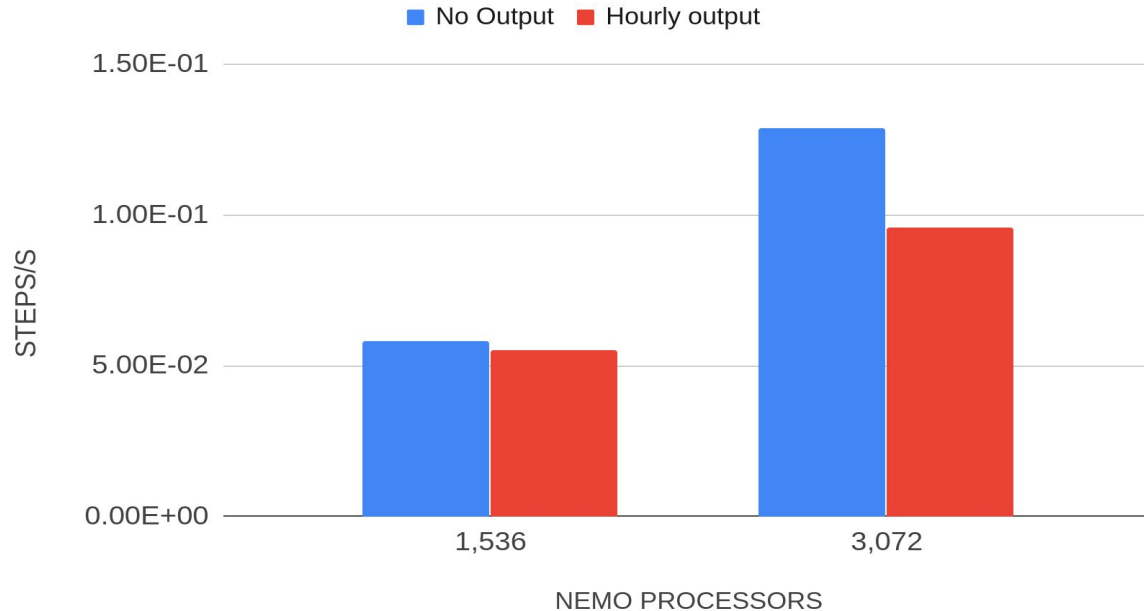
ORCA36 scalability with I/O

Test description

- NEMO 4.0 running with XIOS 2.5.
- **OCE** and **ICE** modules.
- MareNostrum4 supercomputer, Intel 2017.4 compiler and Intel MPI 2018.4.
- ORCA36 configuration provided by Mercator International, CMEMS project.
- **30 seconds** timestep for NEMO. (Clement B. using 120s in “production mode”).
- **Memory mode** used for XIOS.
- XIOS and NEMO running on independent resources.

ORCA36 scalability with I/O

NEMO-XIOS ORCA36 scalability. Early results



ORCA36 scalability with I/O

No output / 2D output

No output

NEMO proc.	XIOS proc.	NEMO step time	XIOS step time	Steps/second
1536	1536	~17s	-	0.058
3072	1536	~8s	-	0.129

2D variables

NEMO proc.	XIOS proc.	NEMO step time	XIOS step time	Steps/second
1536	1536	~17s	~43s	0.058
3072	1536	~8s	~34s	0.126

ORCA36 scalability with I/O

3D hourly output

One file mode

NEMO proc.	XIOS proc.	NEMO step time	XIOS step time	Steps/second
1536	1536	~18s	~366s	0.05
3072	1536	~8s	~348s	0.097
3072	1920	~8s	~376s	0.095

Multiple file mode

NEMO proc.	XIOS proc.	NEMO step time	XIOS step time	Steps/second
1536	1536	~18s	~17s	0.056
3072	1536	~8s	~17s	0.122

ORCA36 scalability with I/O

Some questions to answer

- Can we reduce the wait (XIOS step) by using **performance** mode?
- Multiple file mode reduces the overhead significantly, but may we scale by adding **more processing elements** (servers)?
- Can we run NEMO and XIOS in the **same nodes** and reduce the overhead?
Memory may be an issue.
- Can we speed up the executions by writing in the **local disk** instead of using GPFS?
- Can we benefit from using **Lever-2** servers?

ORCA36 scalability with I/O

Grand challenge executions (2020)

- NEMO 4.0.2 and XIOS 2.5 r1903.
- From 3,072 to 50K (or 100K) cores.
- Intel MPI and Open MPI environment.
- Multiple-file mode.
- Test if the I/O overhead can be reduced by adding more servers and/or using performance mode.



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



EXCELENCIA
SEVERO
OCHOA

Thank you

miguel.castrillo@bsc.es