JCAD 2020
Journées calcul et données

Du mésocentre HPC régional au centre de calcul européen Tiers-0 Prace : Analyse des performances HPC de NEPTUNE_CFD pour la simulation numérique d'un réacteur à lit fluidisé gaz-particule réactif à l'échelle industrielle (de 1 à 64 milliards de mailles)

Hervé NEAU[1,2,7] - Maxime PIGOU[1,2,7] (Engineer in scientific computing - HPC expert)

Nicolas RENON[5,7] - Cyril BAUDRY[6] - Yvan FOURNIER[6] - Nicolas MERIGOUX[6]

Renaud ANSART[3,4,7] - Enrica MASI[1,7] - Pascal FEDE[1,7] - Olivier SIMONIN[1,3,7]

❶ IMFT INSTITUT DE MECANIQUE DES FLUIDES DE TOULOUSE

❷ CNRS dépasser les frontières

❸ INP TOULOUSE

❹ LABORATOIRE DE GÉNIE CHIMIQUE TOULOUSE • UMR 5503

❺ calmip

❻ EDF R&D

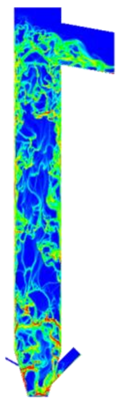
❼ Université de Toulouse

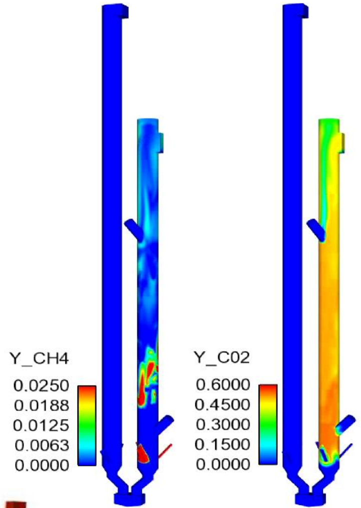
IMFT/LGC research context : Modeling and Simulation of industrial fluid-particle reactive flows

Modeling, Computational Fluid Dynamic, HPC, Experimentations

- **Simulation from laboratory scale up to industrial scale**
- **Industrial applications**

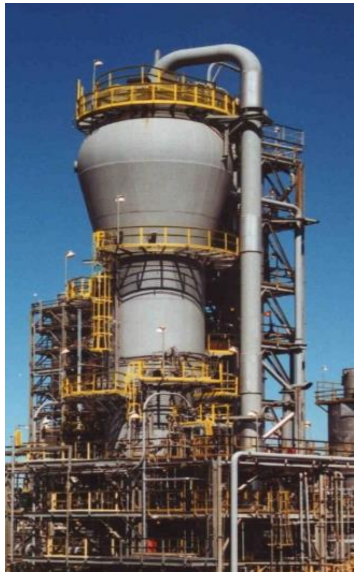
  Polymerization reactor, Chemical looping combustion, coal-fired furnaces, Transport of solid, Concentrated solar power system, Biomass gasifier, FCC riser, Deposition of droplets or particles, Zircon chlorination reactor, Uranium oxide fluorination reactor, …

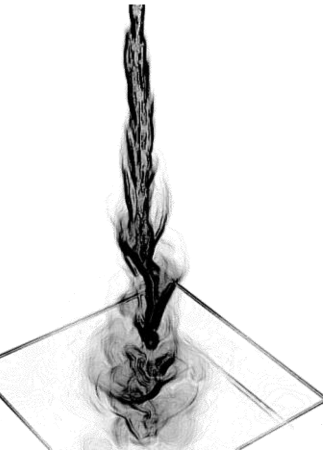




25 years of research on Polyethylene fluidized bed reactor at IMFT/LGC

- **Scale-up studies**
- **Hydrodynamic studies**
- **Reactive studies:** heat and mass transfers
- **HPC studies ⇨ setting an industrial polydispersed reactive case to evaluate HPC capabilities of solvers since 1995**





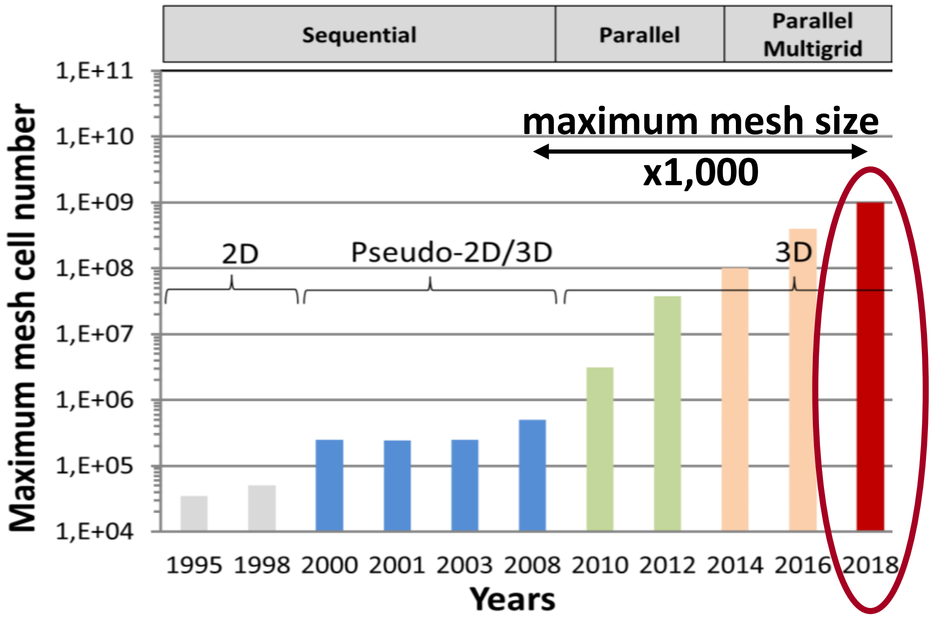
State of art for NEPTUNE_CFD HPC capabilities

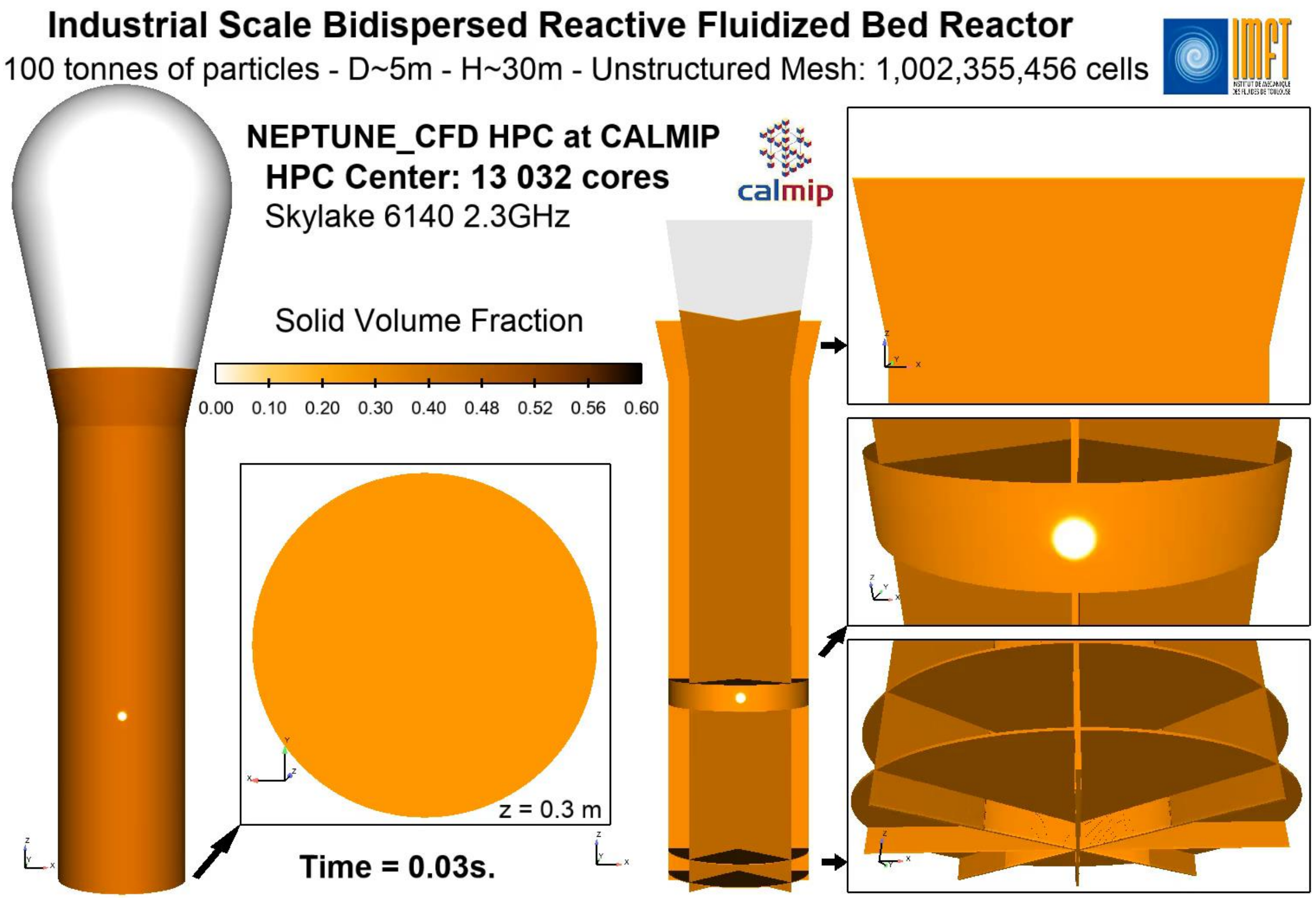
Euler/Euler modeling approach for industrial-scale geometries ⇨ strong sensitivity with respect to mesh size

Maximum mesh size for
- **20 s of simulation of** a reference industrial Fluidized Bed
- max simulation duration: **15 days**

Meso-Challenge CALMIP 2018 ⇨ Presentation JCAD 2018

1 002 355 456 hexahedrons (mesh file: 195 GB)



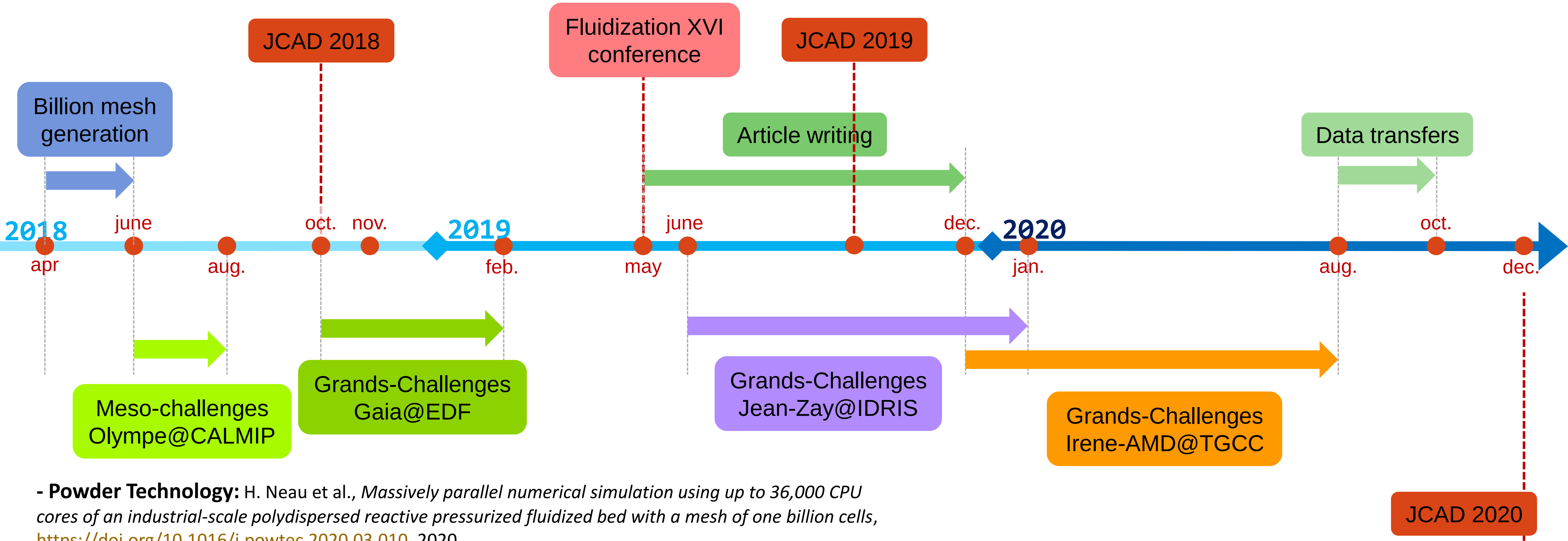

Industrial Scale Bidispersed Reactive Fluidized Bed Reactor

100 tonnes of particles - D~5m - H~30m - Unstructured Mesh: 1,002,355,456 cells

NEPTUNE_CFD HPC at CALMIP
HPC Center: 13 032 cores
Skylake 6140 2.3GHz

Time = 0.03s.

2018-2020: A continuation of NEPTUNE_CFD Meso- and Grands-Challenges at CALMIP, EDF, IDRIS and TGCC


JCAD 2018
Fluidization XVI conference
JCAD 2019
Billion mesh generation
Article writing
Data transfers
2018
apr
june
aug.
oct.
nov.
2019
feb.
may
june
dec.
2020
jan.
aug.
oct.
dec.
Meso-challenges Olympe@CALMIP
Grands-Challenges Gaia@EDF
Grands-Challenges Jean-Zay@IDRIS
Grands-Challenges Irene-AMD@TGCC
JCAD 2020

- Powder Technology: H. Neau et al., *Massively parallel numerical simulation using up to 36,000 CPU cores of an industrial-scale polydispersed reactive pressurized fluidized bed with a mesh of one billion cells*, https://doi.org/10.1016/j.powtec.2020.03.010, 2020

- Fluidization XVI: H. Neau et al., *Massively Parallel Numerical Simulation of Hydrodynamics and Transfers in a Polydispersed Reactive Gas-Particle Fluidized Bed at Industrial Scale with a Very Fine Mesh, over One Billion of Cells*, Guilin, Chine, 2019

- JCAD'18: H. NEAU et al., *Massively Parallel Numerical Simulation of Hydrodynamics and Transfers in a Polydispersed Reactive Gas-Particle Fluidized Bed at Industrial Scale, Lyon, France*, 2018

- **Feedback of these 3 years**
- **1st results and analysis**

Tiers-2 ⇨ Tiers-1 ⇨ Tiers-0: from regional academic computing center up to European supercomputer

Olympe at CALMIP (2018)	
Perf. Peak: 1.37 Pflop/s	Atos Bull SEQUANA X1000 cluster
13,392 cores (2.3 GHz)	RAM: 192 GB/n
360 CPU nodes	2x18c/n - Intel® Xeon® Gold Skylake 6140
Lustre	Infiniband EDR (100 Gb/s)

Gaïa at EDF R&D (2018)	
Perf. Peak: 3.05 Pflop/s 244th Top 500 11/2020	Atos Bull Cluster
42,912 cores (2.3 GHz)	RAM: 192 GB/n
1,192 CPU nodes	2x18c/n - Intel® Xeon® Gold Skylake 6140
GPFS	Intel OPA v1

Jean-Zay at IDRIS (2019)	
Perf. Peak: 16 Pflop/s 108th Top 500 11/2020	HPE SGI 8600
61,120 cores (2.5 GHz)	RAM: 192 GB/n
1,528 CPU nodes	2x20c/n - Intel® Cascade Lake 6248
IBM spectrum scale (ex-GPFS)	Intel OPA (100 Gb/s)

Joliot-Curie Rome Irene-AMD at TGCC (2020)	
Perf. Peak: 11.75 Pflop/s 38th Top 500 11/2020	Bull Sequana XH2000
293,376 (2.6 GHz)	RAM: 256 GB/n
2,292 CPU nodes	2x64c/n – AMD Rome (Epyc) 7H12
Lustre	Infiniband HDR100

2018 CALMIP Meso-Challenge / EDF Grands-Challenges: A First Worldwide numerical simulation

Finite Volume Solver: NEPTUNE_CFD*: 3D reactive turbulent unsteady multiphase flows (C/C++, MPI, QT-Python GUI)

- **Massively Parallel** code: **MPI**, parallel mesh reading, parallel partitioning, pressure **parallel multigrid** solver, MPI I/O
- **Unsteady Multi-Fluid Modeling approach (N-Euler)**
- **NEPTUNE_CFD is proprietary** and powered by the open-source software **Code_Saturne** very high HPC capabilities ⇨ http://code-saturne.org

* NEPTUNE_CFD is developed jointly by EDF and CEA with financial support of IRSN and FRAMATOME in the framework of the NEPTUNE project for nuclear applications

Results: Olympe CALMIP: 16 s of physical time, Gaïa EDF: 25 s of physical time with the same 1 billion cells mesh

- Runs from 35 nodes (1,260 cores) up to 1,000 nodes (36,000 cores)
- 15 millions CPU hours
- **NEPTUNE_CFD HPC capabilities assessed up to 36,000 cores** on a fluidized bed at industrial scale
- **Proof of feasibility of computations with more than one billion cells using the whole supercomputer capabilities!**

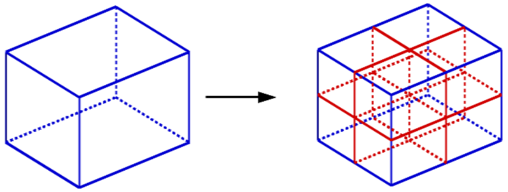
⇨ **First hand experience of massively computation on the full supercomputer scale**

⇨ **Key experience to attempt a more refined case on Tiers-1 facilities**

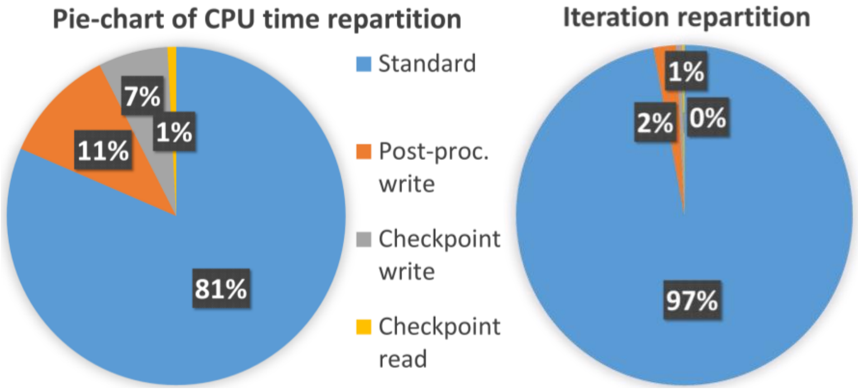
Grands-Challenges IDRIS: Jean-Zay – Refining the industrial case to 8 billion-cells mesh (2019)

Goals of this challenge:

- **Implement splitting the 1 billion-cells mesh into a 8 billion-cells mesh: from 215 GB to 1.6 TB**
  - Native **Code_Saturne functionality:** divide each cell by 2 in each direction: 1 hexa ⇨ 8 hexa
  - Splitting job uses at least **250 nodes, *i.e.* 10,000 cores**
    ⇨ **new mesh file: 8 018 843 648 hexahedrons**
    $\Delta_x \sim \Delta_y \approx 3\ \text{mm} - \Delta_z \approx 5\ \text{mm}\ -\ V_{cell} \approx 45\ \text{mm}^3 - \phi \sim 1\ 500\ \text{cells}$



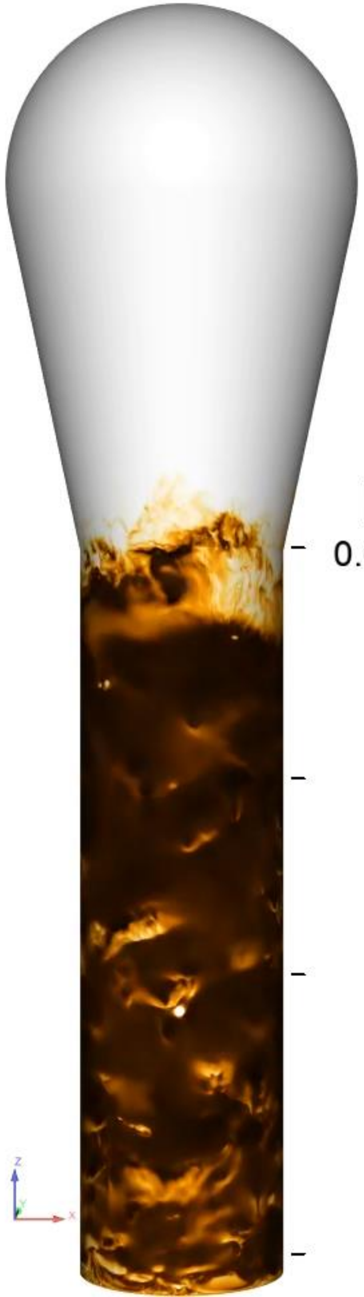
- **Implement interpolation** of the 1 billion-cells restart file at 25 s onto the 8 billion-cells mesh using native **NEPTUNE_CFD functionality:** at least **300 nodes, *i.e.* 12,000 cores**
  ⇨ **new restart file: 9.8 TB**



- **Computations with the 8 billion-cells mesh:**
  - Continue computation from 25 s up to 26.7 s: **1.7 s of physical time ⇔ 31 M CPU h**
  - **Extremely long computations with "short" walltime (20h)** ⇨ 53 parts and restart files of 9.8 TB each
  - **53 post-processing data sets** (binary ensight gold)**: 57 TB**
  - **HPC assessment: speed-up, efficiency, sensibility studies** (IO, process CPU binding)
  - Runs using **240 nodes up to 1,500 nodes (98.2% Jean-Zay), *i.e.* 9,600 up to 61,120 cores**

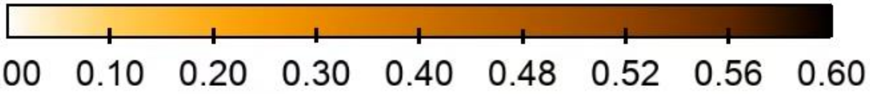
Industrial-scale Polydispersed Reactive Fluidized Bed 3D simulation with NEPTUNE_CFD on Jean-Zay (IDRIS)

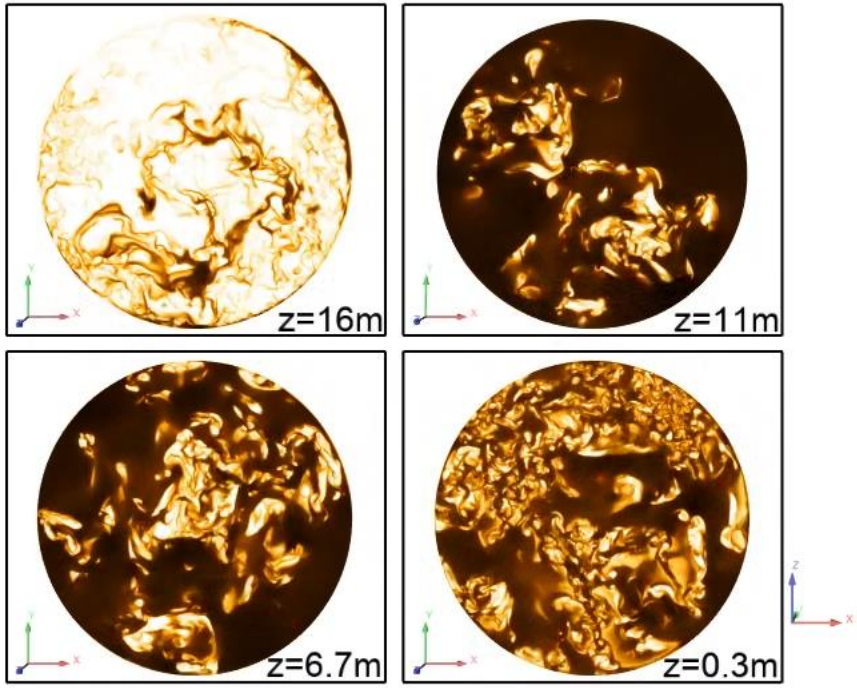
Unstructured mesh of 8,018,843,648 cells
8,560 to 51,840 MPI processes

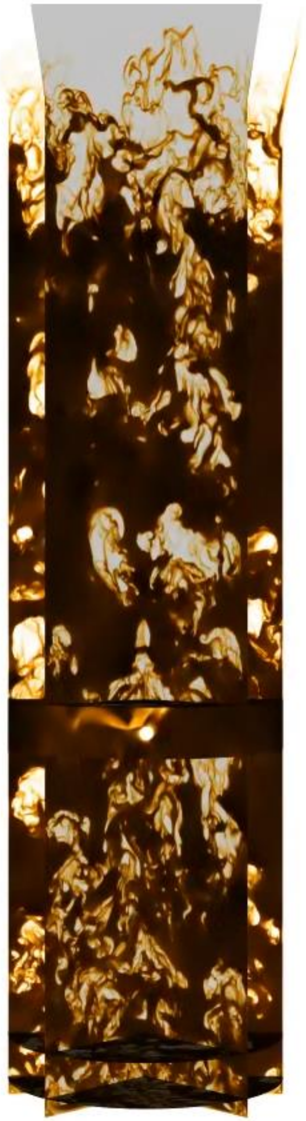


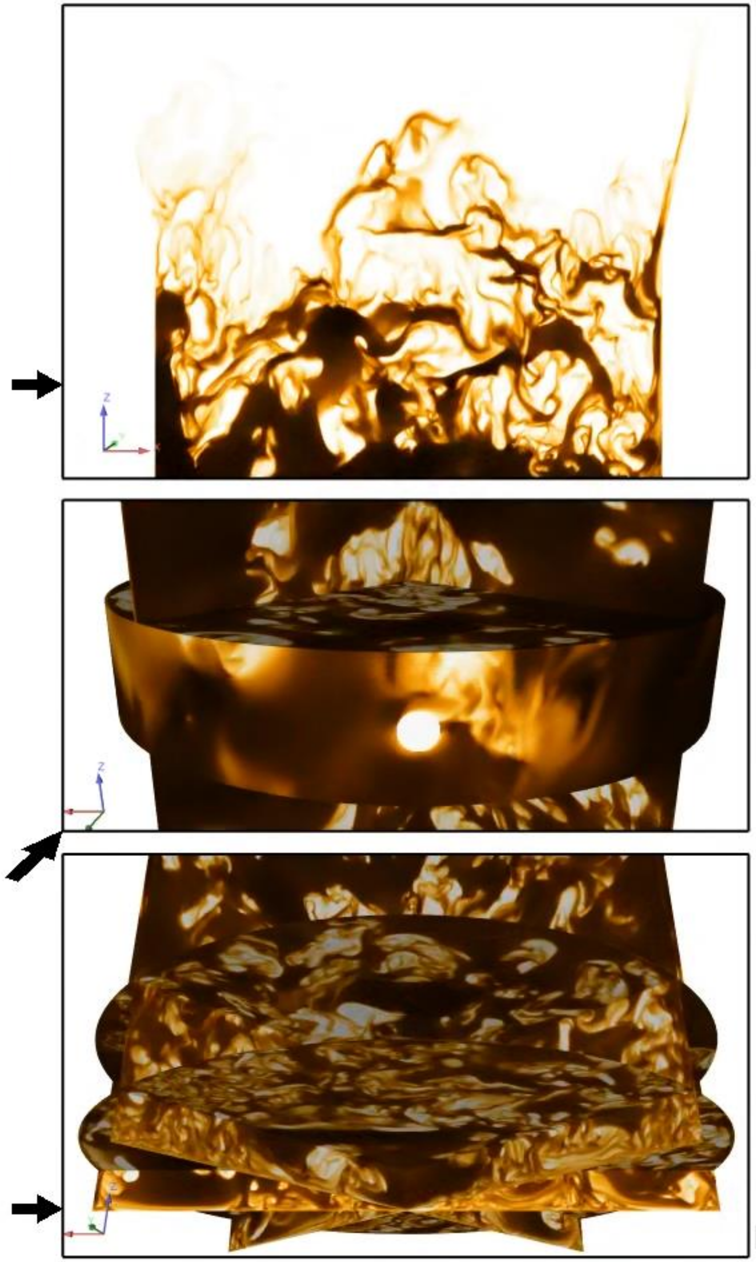
Time = 26.2168s.

Solid Volume Fraction





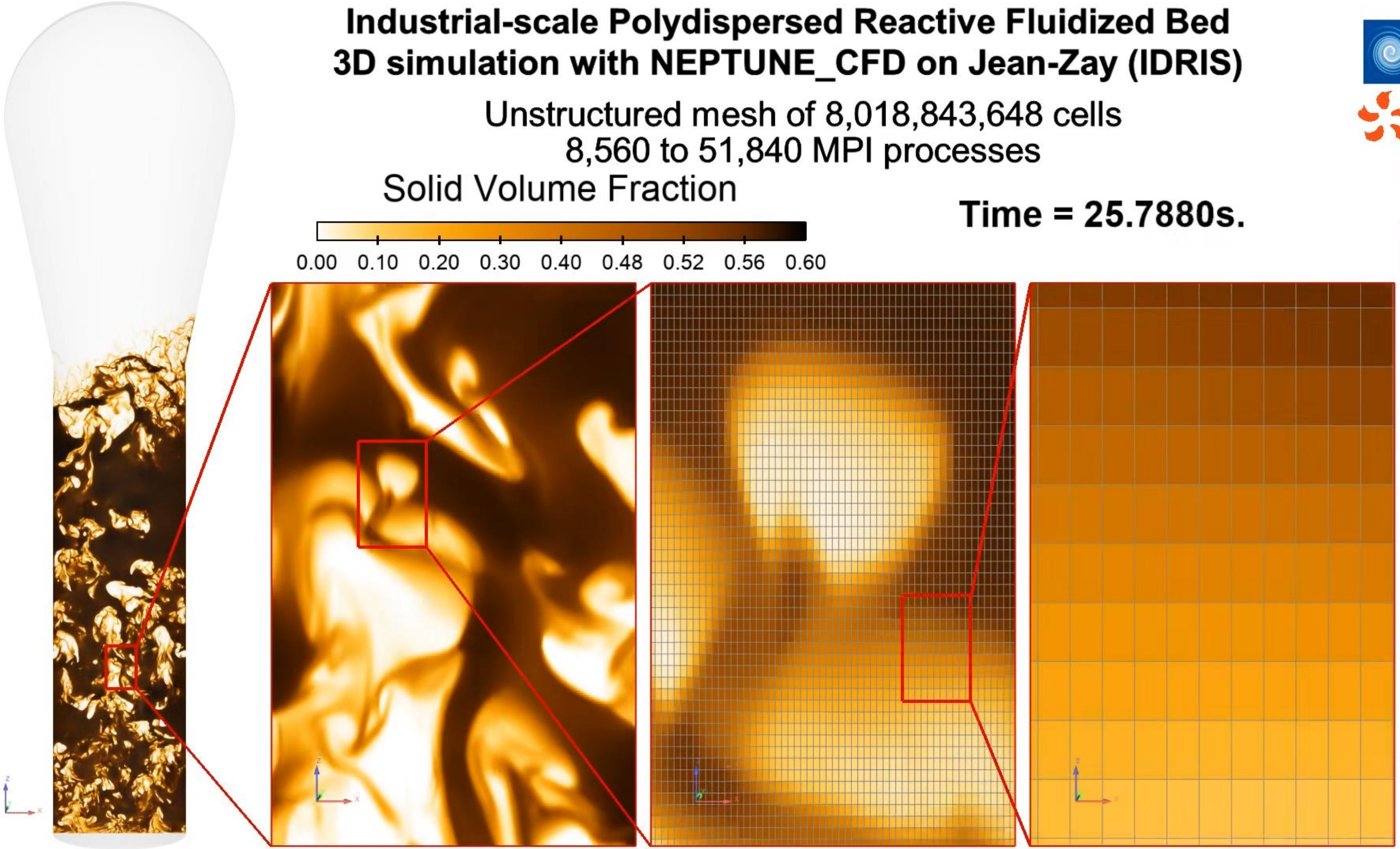




IMFT INSTITUT DE MECANIQUE DES FLUIDES DE TOULOUSE
eDF
cnrs
INP TOULOUSE
calmip
cnrs IDRIS
TGCC Très Grand Centre de calcul du CEA
PRACE
Herve Neau
Maxime Pigou

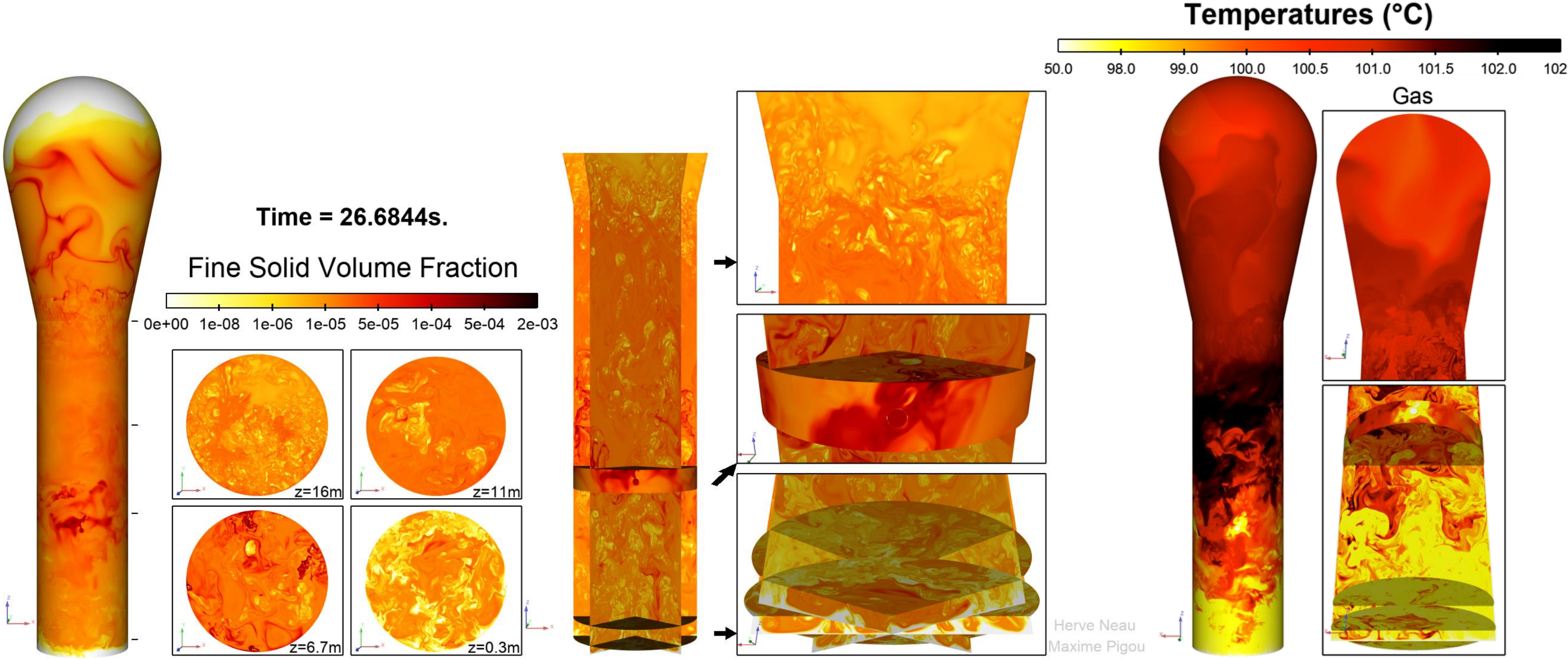
Industrial-scale Polydispersed Reactive Fluidized Bed 3D simulation with NEPTUNE_CFD on Jean-Zay (IDRIS)

Unstructured mesh of 8,018,843,648 cells
8,560 to 51,840 MPI processes



H. Neau
M. Pigou

Polydispersed and reactive flows: α_i, T_i, U_i, …



Herve Neau
Maxime Pigou

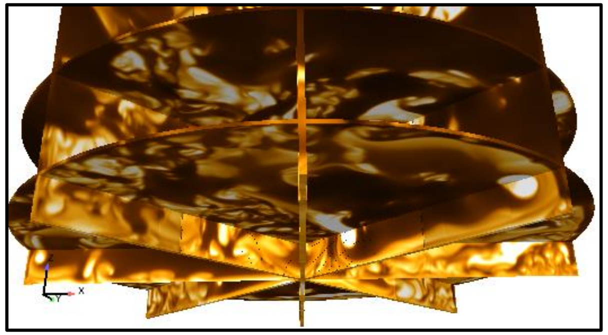
TGCC Grands-Challenges on Joliot-Curie AMD Irene ROME: scaling further up from 8 to 64 billion cells (2020)

Initial goals (5 million CPU-h requested):

- **Test NEPTUNE_CFD on AMD hardware architecture** while still using **Intel compiler and MPI library: v18.2**
- **Complete HPC scaling curve for 1 and 8 billion-cells** at higher core count: 30 up to **2 250 nodes**, *i.e.* **288 000 cores**
- Attempt generating a 64 billion-cells mesh and restart binary files
- **Mesh: single file of 12 TB - 64 150 749 184 hexahedrons** $\Delta_x \sim \Delta_y \approx 1.5 \text{ mm} - \Delta_z \approx 2.5 \text{ mm} - V_{cell} \approx 5.6 \text{ mm}^3$
- **Restart: single file of 80 TB** with double precision scalar and vector fields at t = 26.6 s
- **Only performance measurements and profiling on few iterations**, no significant time advancement

Additional goals (31 million CPU-h total):

- **Many sensibility studies**: I/O tuning on Lustre filesystem (striping, MPI IO collective vs non-collective, …), CPU binding of MPI processes, network topology of selected computation nodes, node depopulation, …
- Attempt generating a **512 billion cells mesh**: 100 TB of expected size
- **Advanced ARM MAP profiling**
- **Comparison of NEPTUNE_CFD v4.1 and v6.0 performances**



Generated data: ⇨ **a few large binary files: mesh + restart in v4 and v6 formats: 400 TB**
⇨ **valuable data sets: 300 GB**
⇨ **many small files of high value with profiling results: 2 GB**

A focus on some challenges faced during this series of Grands-Challenges

Many steps and parameters to control – A human-error prone environment:

- **Job configuration errors:** wrong cpu binding (64 process on 16 cores), option mix between different studies, symlink (to prevent file duplication) sometime broken, job walltime too short, …
- `rm -Rf ./*` **performed wrongfully** by support staff in root of our SCRATCH storage instead of theirs
- **Jobs cancelled by error before or during run**

Data management and transfer in a multi-site project:

- **High volume of data to manage: CALMIP ⇨ 15 TB, EDF ⇨ 20 TB, IDRIS ⇨ 150 TB, TGCC ⇨ 350 TB**
- **Slow transfers:** often limited to **at best 1 Gbps** and sometime at most 1Mbps
- **No robust transfer method** for files of such a large size: **manual file splitting + checksum** proved to be the most efficient method (but induced temporary data duplication)
- **Data Management Plan** required for long-term storage of high-value files
- Two examples of data transfers:
  - EDF ⇨ IMFT: 16 TB transferred through… **hard-drive mailing ⇨ 1 week**
  - Post TGCC Grands-Challenges: 300 TB transferred toward other project (200 TB), IDRIS (50 TB), CALMIP (50 TB) and IMFT (30 TB) **over 2 months**

A focus on some challenges faced during this series of Grands-Challenges

Implication of running jobs on thousand of nodes:

- **Waiting for jobs to start:** resources not always available, jobs often ran during night, week-ends and vacations, job failure if stayed in queue for too long, overlap Grands-Challenges/Production, …
- **Once started, issues with MPI library:**
  - "Having large MPI runs with multiple thousands of ranks, these MPI_INIT and MPI_FINALIZE operations can consume a huge part of the MPI initialization phase time" **(IntelMPI doc.)**
    ⇨ **Up to 60 minutes just to start the preprocessing steps on biggest jobs**
  - **MPI communication buffers could exceed available RAM thus crashing the simulation**
- **During run: hardware failure highly likely to occur when more than 1 000 nodes are used**

From 1 to 512 billion-cells meshes:

- **Code Saturne interpolation** works for any two meshes of the same geometry with a robust but slow algorithm
  ⇨ A tailor-made algorithm has been implemented **by EDF for split-hexahedrons meshes (x40 speed-up)**
- **More than 2^{32} elements** on some processes ⇨ **overflow on uint32 typed index**
- **By default, max file-size of 4 TB on Irene AMD:** `ulimit –f ulimited` to lift this constraint
- Due to RAM requirements, **2 000 computes nodes** required to attempt generating the **512 billion-cells mesh**

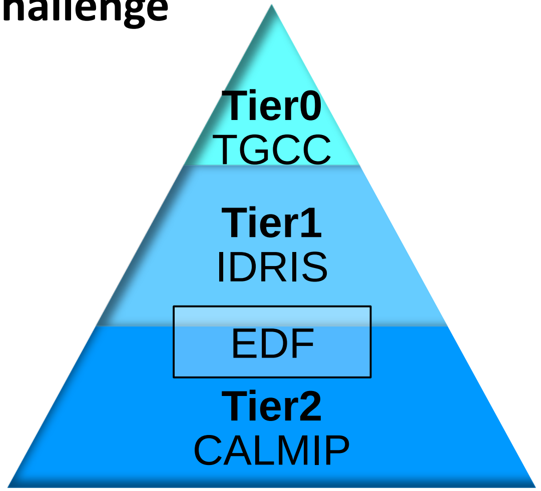
A focus on some challenges faced during this series of Grands-Challenges

Many more challenges:

- **COVID impact** (part of supercomputers reserved: 2 000 nodes max, remote working)
- **Different software environments:** recompilation of NEPTUNE_CFD dependencies ➪ homogeneous software stack
- **Need to pre-define the post-processing:** EnSight Gold binary data files for 717 time steps on 12 selected thick planes, 1 cylinder and 1 external surface saving 20 variables on Jean-Zay
- **Post-processing and visualization of 53 TB of data: data transfers, use of HPC resources, ParaView in client/server**
- **Human challenge**: constant work since 2018 - need of expert at each step: mesh, partitioning, splitting, interpolate, run

To tackle all these issues along the way:

- **Climbing the HPC Pyramid** was only possible with the experience from **initial CALMIP Meso-Challenge**
- **Really tight collaboration with EDF Software Engineers** to adapt some tools to this scale
- **Strong support from supercomputing centers teams** whose help was required on very specific questions (hardware architecture, profiler usage, MPI environment tuning, …)
- **Direct help and strong interest from Heads of computing centers** who offered:
  - Supplementary allocations of computation time
  - Help to increase jobs priority when possible



First Irene-AMD study: tuning I/O for file reading

- **Out-of-the-box input data read times:**

	Jean-Zay	Irene ROME
1 billion cells case: 200 GB mesh + 1.2 TB checkpoint	2 min	6 min
8 billion cells case: 1.6 TB mesh + 9.8 TB checkpoint	6 min	55 min

- **Jean-Zay SCRATCH** uses a **General Parallel File System** (GPFS)

  ⇨ Data is **automatically distributed** and accessed in parallel among multiple storage nodes

  ⇨ **Transparent and turn-key solution** for users, though maybe not with optimum tuning for all users

- **Irene SCRATCH** relies on a **Lustre Filesystem with 42 storage nodes** (OST)
  - **By default**, data is **not "stripped"** (distributed) among those 42 OSTs
  - User may **manually** define the **number of stripes**, and the **stripe block size** (support recommended 4 MB)

  ⇨ Fine-tuning required to reduce simulation preprocessing time

First Joliot-Curie Irene-AMD Rome study: tuning I/O for file reading

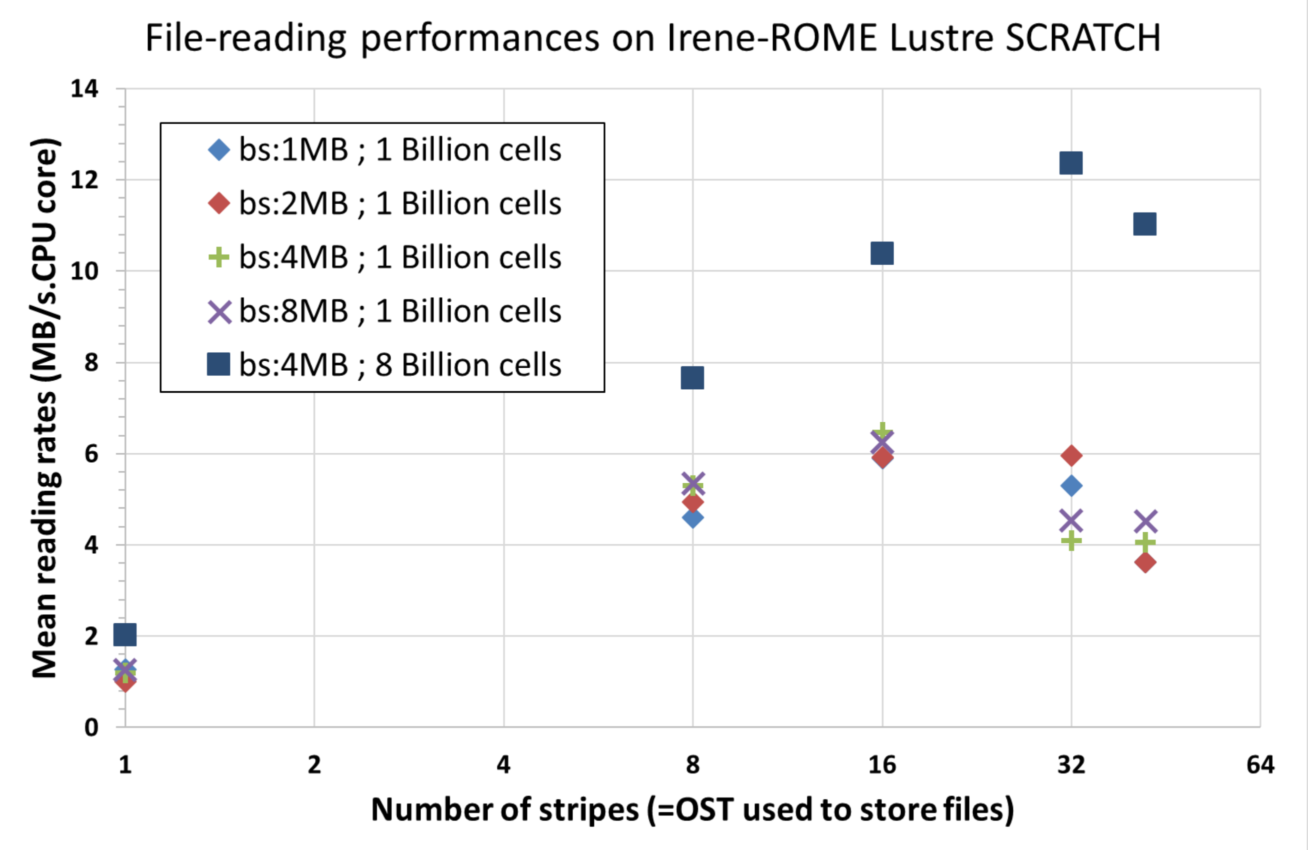
- **Multi-variable parametric study: mesh and checkpoint read times vs**
  - **Lustre number of stripes**
  - **Lustre block-size**
  - **Number of nodes**
  - **Proximity of nodes** on the network topology
  - NEPTUNE_CFD reading method (standard sequential/parallel, MPI IO collective/non-collective)

⇨ **Extensive results database to analyze**. Quick summary:

- Low impact of Lustre block-size
- Good scalability when increasing number of stripes … but loss of performances when using all 42 OSTs

⇨ **With correct tuning, speedup of**
x22 for mesh (1.6 TB)
x15 for checkpoint (9.8TB)

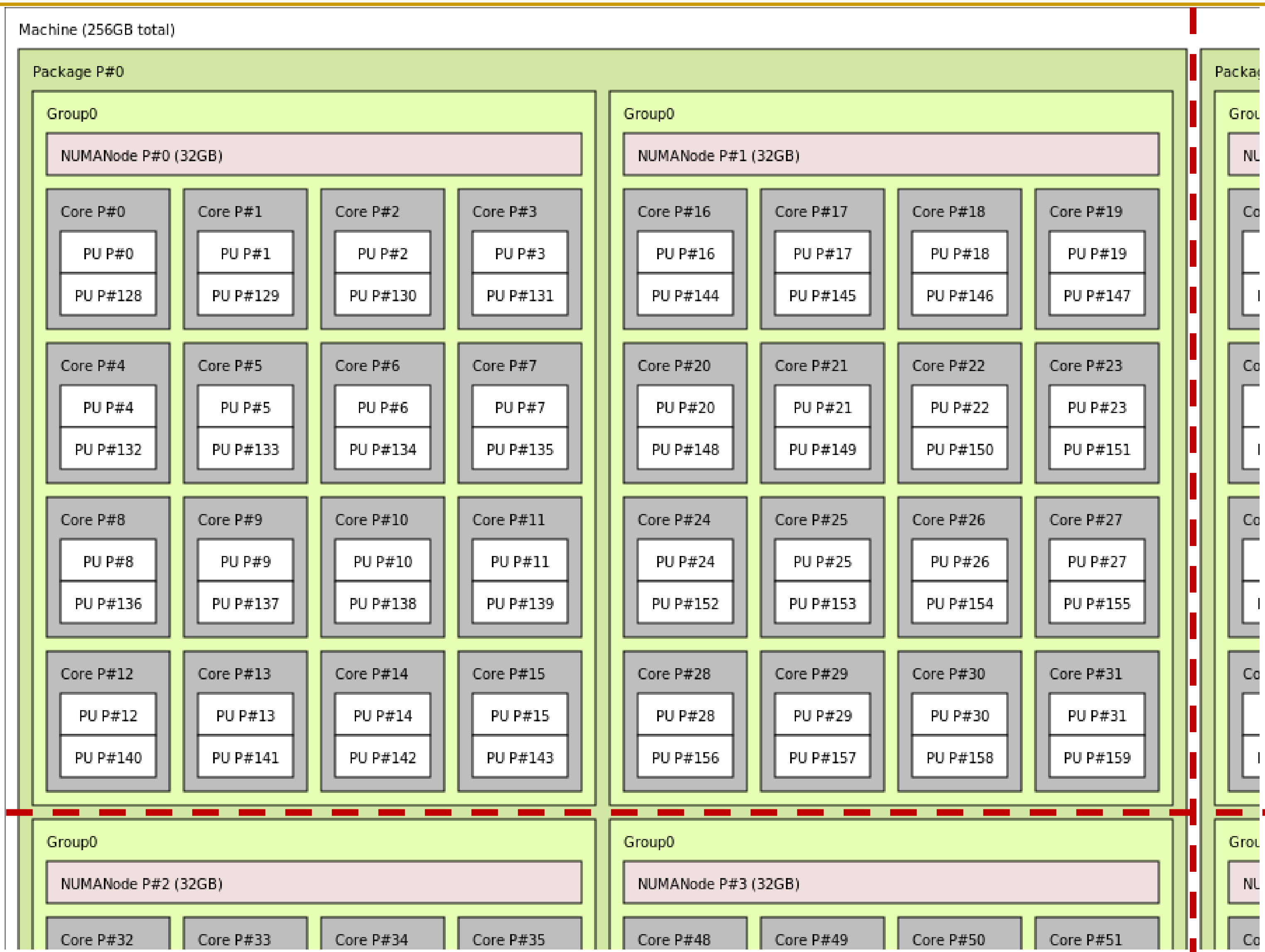
⇨ **From 55 min of read-time down to 4min30!**



Joliot-Curie Irene-AMD: How many cores to use?

Irene Rome: AMD Rome (Epyc) 7H12

- 128 cores per node: 2x64
- 4 cores
  x 4 (L3 cache)
  x 4 groups
  x 2 sockets
- **Which MPI placement apply?**



Machine (256GB total)

Package P#0

Group0

NUMANode P#0 (32GB)

Core P#0 PU P#0 PU P#128	Core P#1 PU P#1 PU P#129	Core P#2 PU P#2 PU P#130	Core P#3 PU P#3 PU P#131
Core P#4 PU P#4 PU P#132	Core P#5 PU P#5 PU P#133	Core P#6 PU P#6 PU P#134	Core P#7 PU P#7 PU P#135
Core P#8 PU P#8 PU P#136	Core P#9 PU P#9 PU P#137	Core P#10 PU P#10 PU P#138	Core P#11 PU P#11 PU P#139
Core P#12 PU P#12 PU P#140	Core P#13 PU P#13 PU P#141	Core P#14 PU P#14 PU P#142	Core P#15 PU P#15 PU P#143

Group0

NUMANode P#1 (32GB)

Core P#16 PU P#16 PU P#144	Core P#17 PU P#17 PU P#145	Core P#18 PU P#18 PU P#146	Core P#19 PU P#19 PU P#147
Core P#20 PU P#20 PU P#148	Core P#21 PU P#21 PU P#149	Core P#22 PU P#22 PU P#150	Core P#23 PU P#23 PU P#151
Core P#24 PU P#24 PU P#152	Core P#25 PU P#25 PU P#153	Core P#26 PU P#26 PU P#154	Core P#27 PU P#27 PU P#155
Core P#28 PU P#28 PU P#156	Core P#29 PU P#29 PU P#157	Core P#30 PU P#30 PU P#158	Core P#31 PU P#31 PU P#159

Group0

NUMANode P#2 (32GB)

Core P#32 Core P#33 Core P#34 Core P#35

Group0

NUMANode P#3 (32GB)

Core P#48 Core P#49 Core P#50 Core P#51

Packag

Grou

NU

Core P#64 - P#79

Core P#80 - P#95

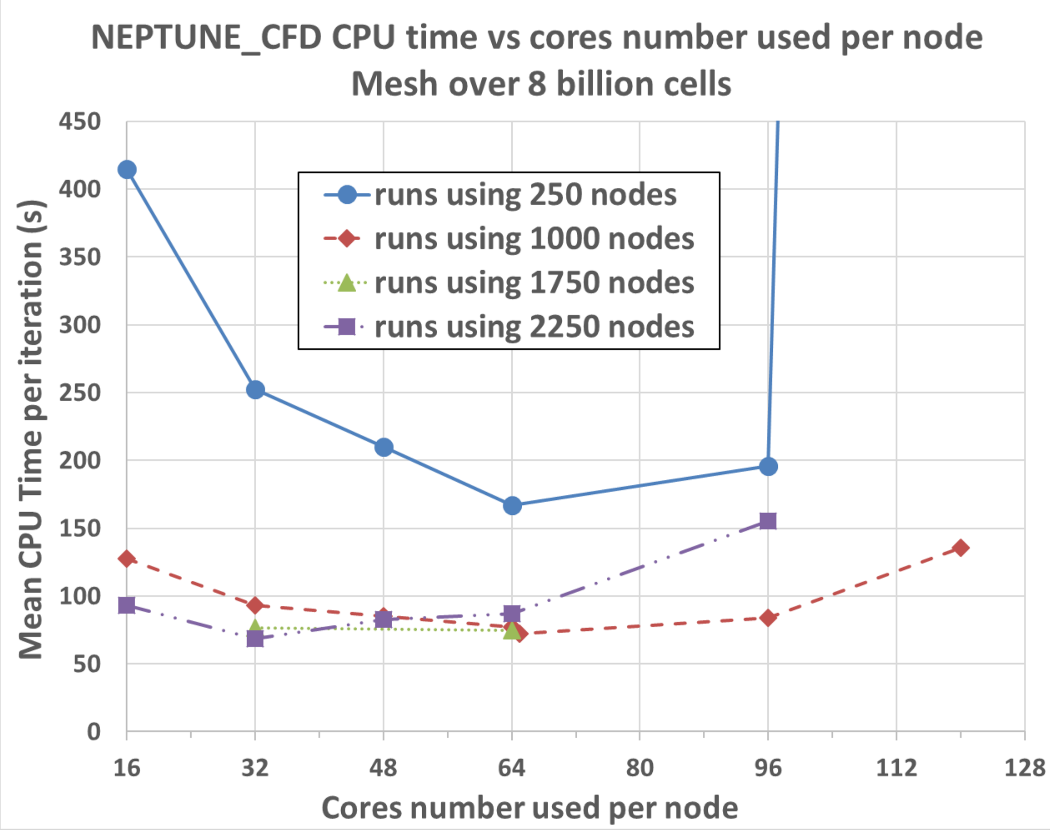
Core P#96 - P#111

Core P#32 - P#47

Core P#48 - P#63

Core P#112 - P#127

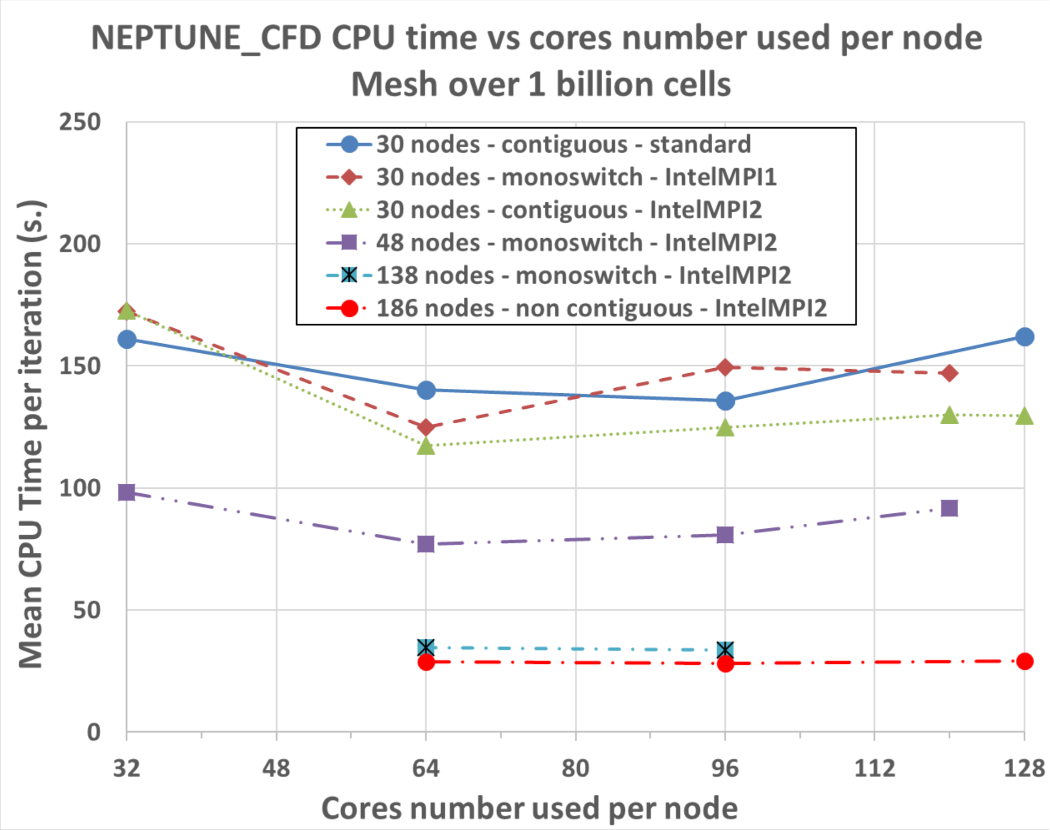
Sensibility to the number of cores used per node



Strong sensitivity to node depopulation

120 and 128 cores ⇨ lower performances when using all cores

⇨ computation often hanged at startup

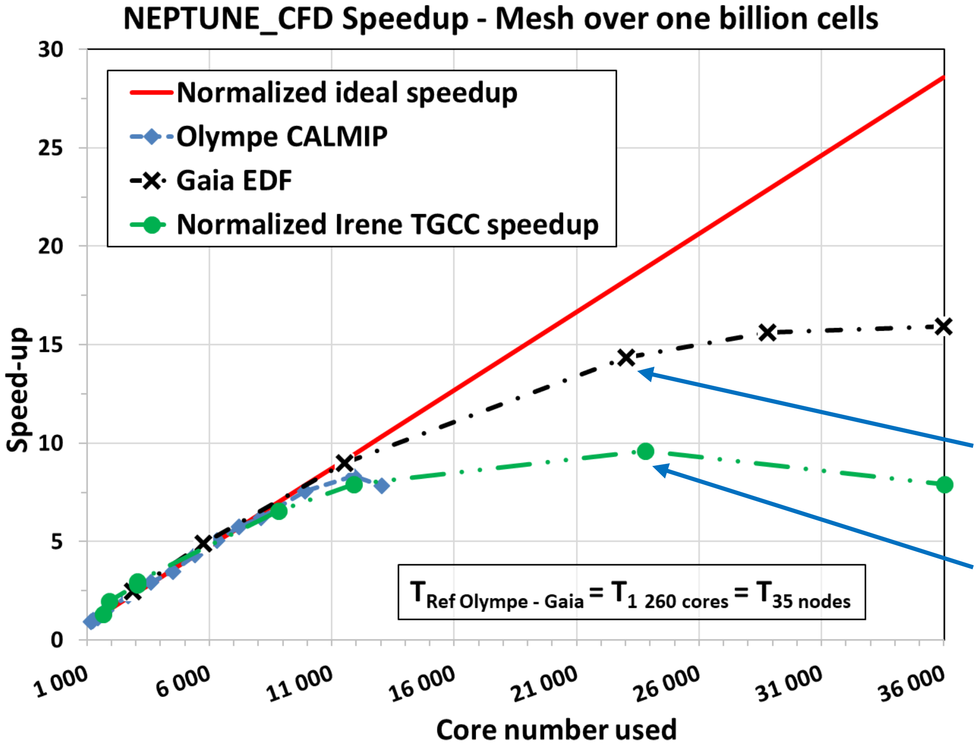


The bigger the simulation the stronger the depopulation

- 96 cores ⇨ 1 billion cells - 30 nodes
- 64-96 cores ⇨ 1 billion cells - 48-186 n
- 64 cores ⇨ 8 billion cells - 250-1500 n
- 32 cores ⇨ 8 billion cells - 1500-2250 n

Manual CPU pining for each core count to maximize L3 cache per core

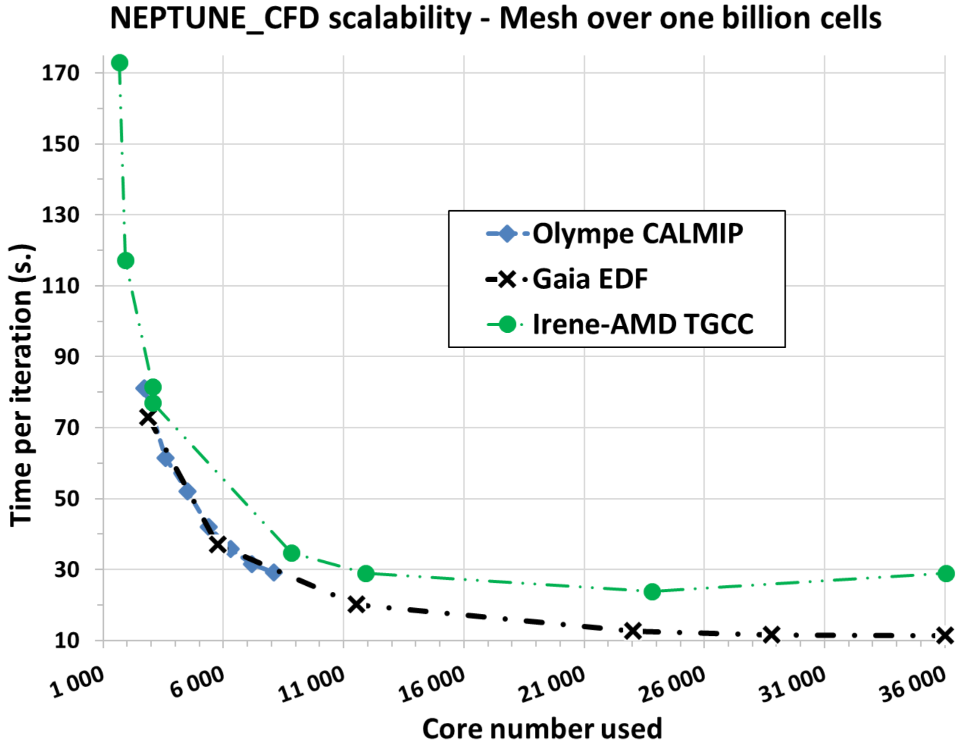
NEPTUNE_CFD HPC capabilities demonstrated up to 36,000 cores on a fluidized bed at industrial scale (1,002,355,456 cells)



Olympe: 36/36 cores
Gaïa: 35/36 cores
Jean-Zay: 40/40 cores
Irene-AMD: 64/128 cores

Efficiency 80% on Gaïa using 22,000 cores

Efficiency 50% on Irene-AMD

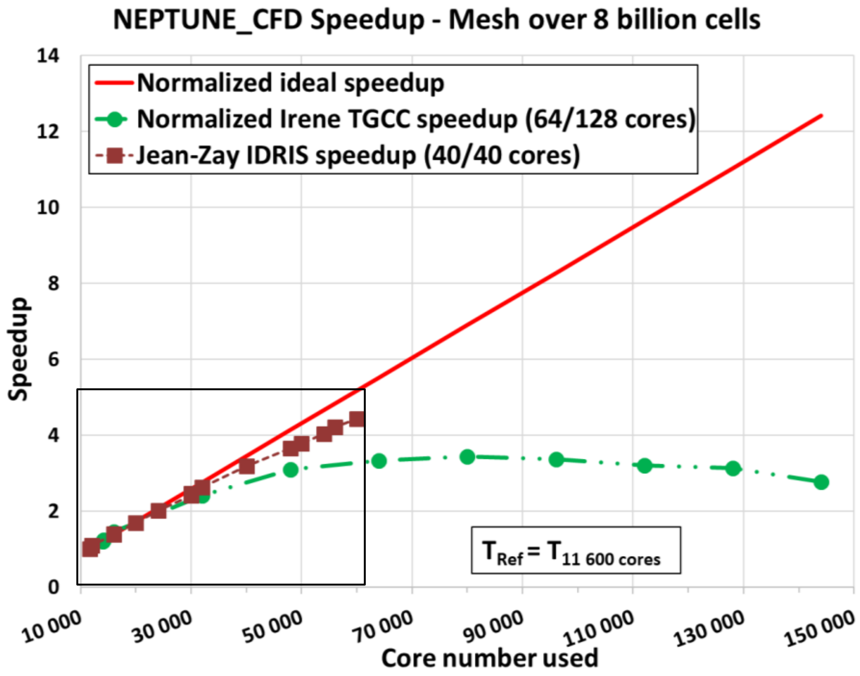


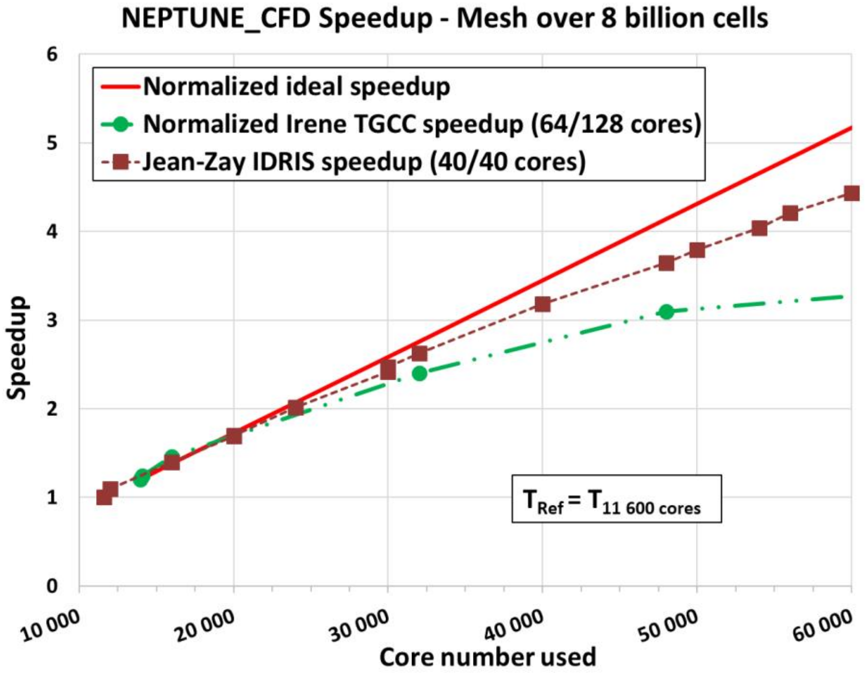
⇨ **Ideal speedup up to 12,000 cores** (> 70,000 cells/core)

⇨ **Excellent speedup up to 22,000 cores on Olympe and Gaïa** (> 45,000 cells/core)

⇨ **For NEPTUNE_CFD with current configuration**, for any number of cores: **Irene-AMD TGCC < Olympe CALMIP ≤ Gaia EDF**

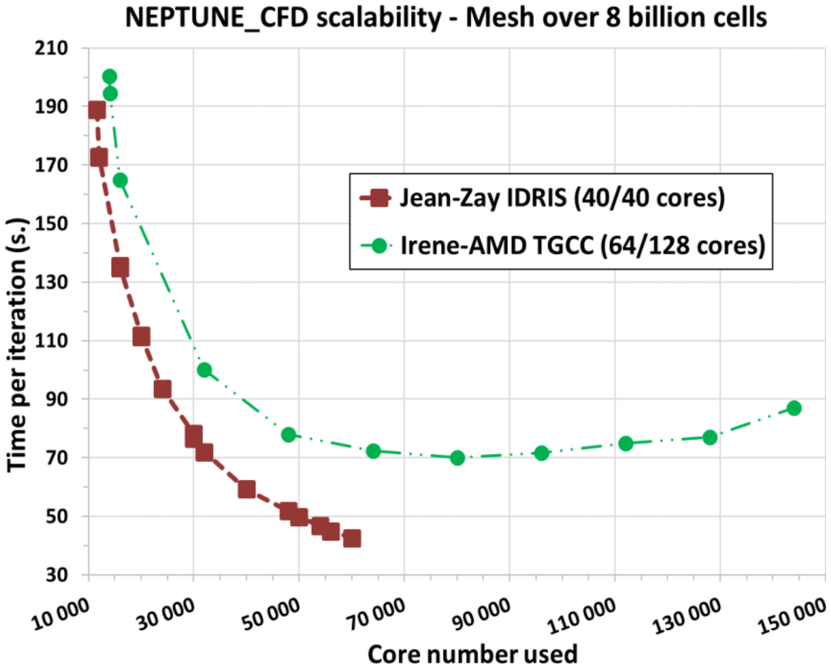
NEPTUNE_CFD HPC capabilities demonstrated up to 60,000 cores on a fluidized bed at industrial scale with 8×10^9 cells mesh





Efficiency 85% on Jean-Zay using 60,000 cores

Efficiency 60% on Irene-AMD



⇨ **Excellent scaling capabilities and efficiency up to 60,000 cores on entirety of Jean-Zay** (> 130,000 cells/core)

⇨ **Excellent speedup up to 32,000 cores on Irene, correct up to 60,000 cores, but no gain at higher core counts**

⇨ **For NEPTUNE_CFD with current configuration**, for any number of cores:
Irene-AMD TGCC < Jean-Zay

Journey to the Center of the Code:
an in-depth profiling of NEPTCFD v4/v6

Two profiling methods:

- **NEPTUNE_CFD internal API**: per-it. timers

⇨ 200+ CSV tables to analyze

- **Plug-and-play profiling tool: Arm MAP**

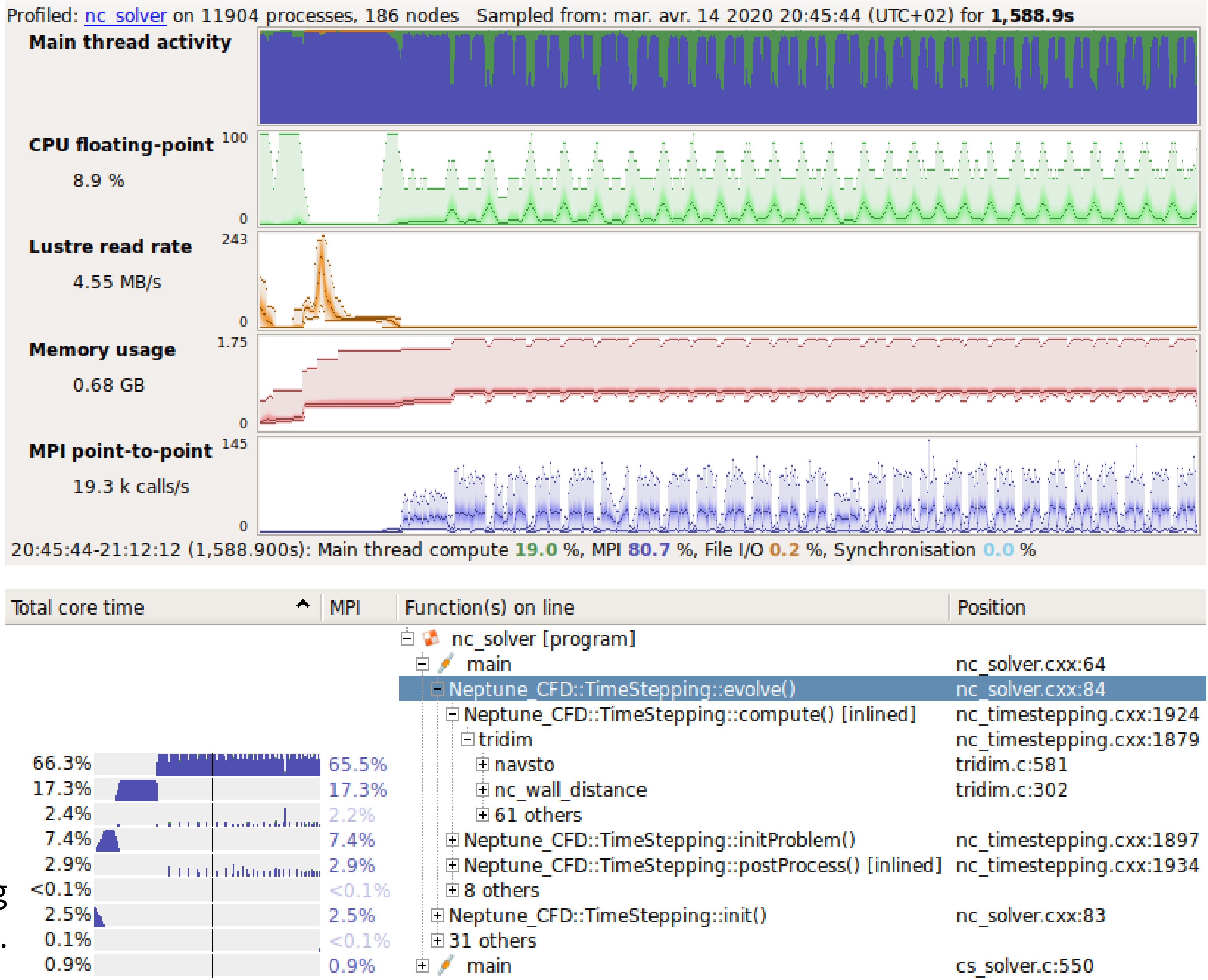
```
$ map --profile srun ./solver.exe
```

(requires compilation **-g** debug flag)

⇨ Min/Max/Mean of **macroscopic metrics** (CPU usage, MPI communications, Lustre usage, memory consumption per code/node, …)

⇨ **Total time spent in subroutines** over time slices

- Used from 960 up to **72 064 MPI processes**
- **5 to 50% performance penalty** with MAP binding not due to -g flag which actually improved perfs.


Profiled: nc_solver on 11904 processes, 186 nodes Sampled from: mar. avr. 14 2020 20:45:44 (UTC+02) for **1,588.9s**

Main thread activity

CPU floating-point 8.9 %

Lustre read rate 4.55 MB/s

Memory usage 0.68 GB

MPI point-to-point 19.3 k calls/s

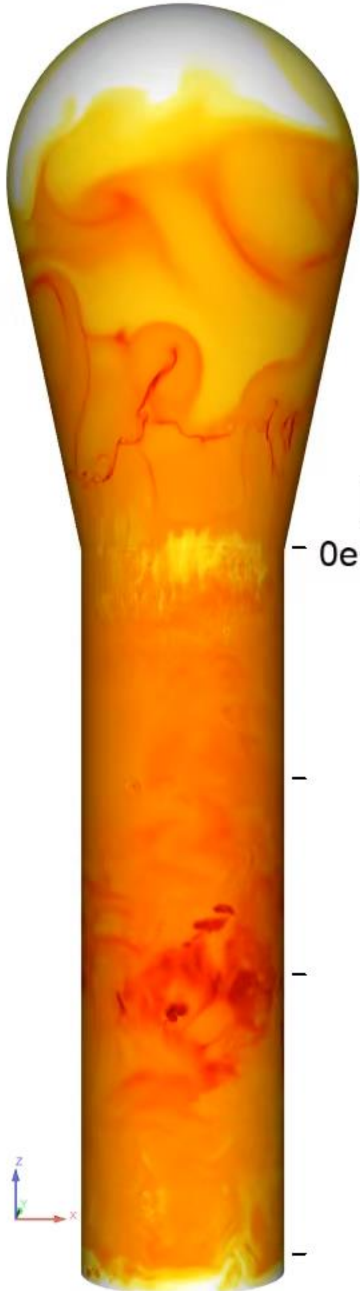
20:45:44-21:12:12 (1,588.900s): Main thread compute 19.0 %, MPI 80.7 %, File I/O 0.2 %, Synchronisation 0.0 %

Total core time	MPI	Function(s) on line	Position
		⊟ nc_solver [program]	
		⊟ main	nc_solver.cxx:64
		⊟ Neptune_CFD::TimeStepping::evolve()	nc_solver.cxx:84
		⊟ Neptune_CFD::TimeStepping::compute() [inlined]	nc_timestepping.cxx:1924
		⊟ tridim	nc_timestepping.cxx:1879
66.3%	65.5%	⊞ navsto	tridim.c:581
17.3%	17.3%	⊞ nc_wall_distance	tridim.c:302
2.4%	2.2%	⊞ 61 others	
7.4%	7.4%	⊞ Neptune_CFD::TimeStepping::initProblem()	nc_timestepping.cxx:1897
2.9%	2.9%	⊞ Neptune_CFD::TimeStepping::postProcess() [inlined]	nc_timestepping.cxx:1934
<0.1%	<0.1%	⊞ 8 others	
2.5%	2.5%	⊞ Neptune_CFD::TimeStepping::init()	nc_solver.cxx:83
0.1%	<0.1%	⊞ 31 others	
0.9%	0.9%	⊞ main	cs_solver.c:550

Industrial-scale Polydispersed Reactive Fluidized Bed
3D simulation with NEPTUNE_CFD on Jean-Zay (IDRIS)

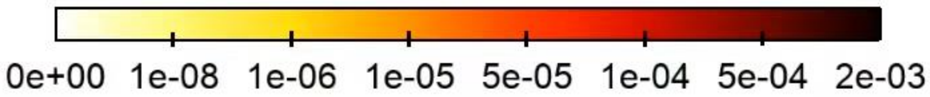
IMFT INSTITUT DE MECANIQUE DES FLUIDES DE TOULOUSE
eDF

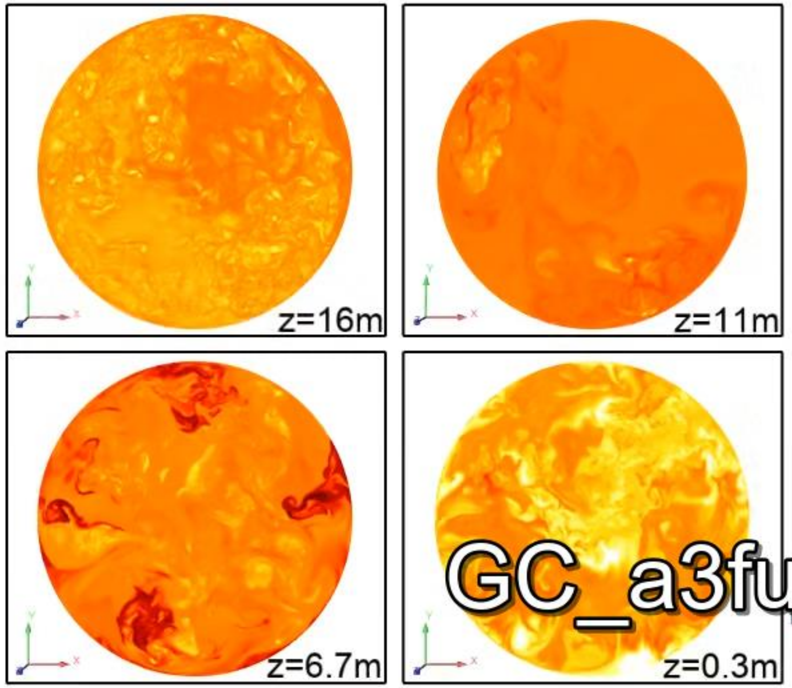
Unstructured mesh of 8,018,843,648 cells
8,560 to 51,840 MPI processes

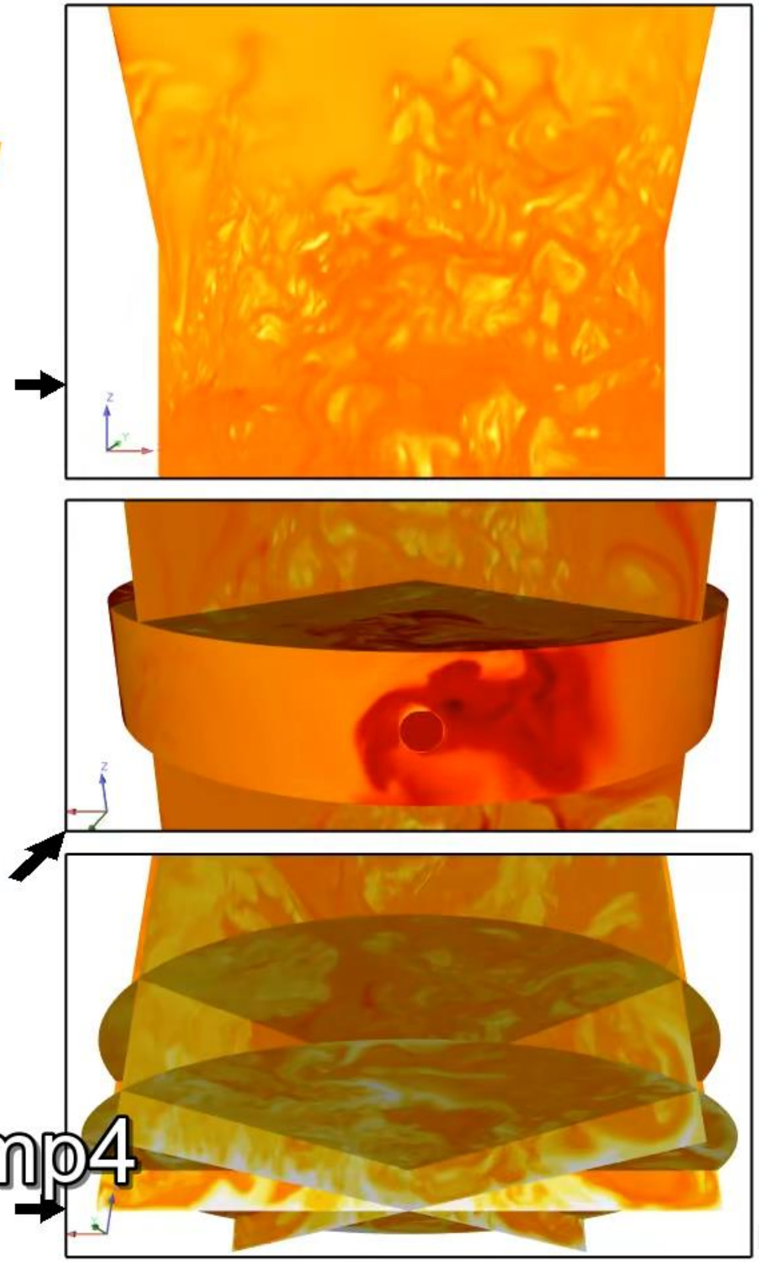


Time = 25.0137s.

Fine Solid Volume Fraction



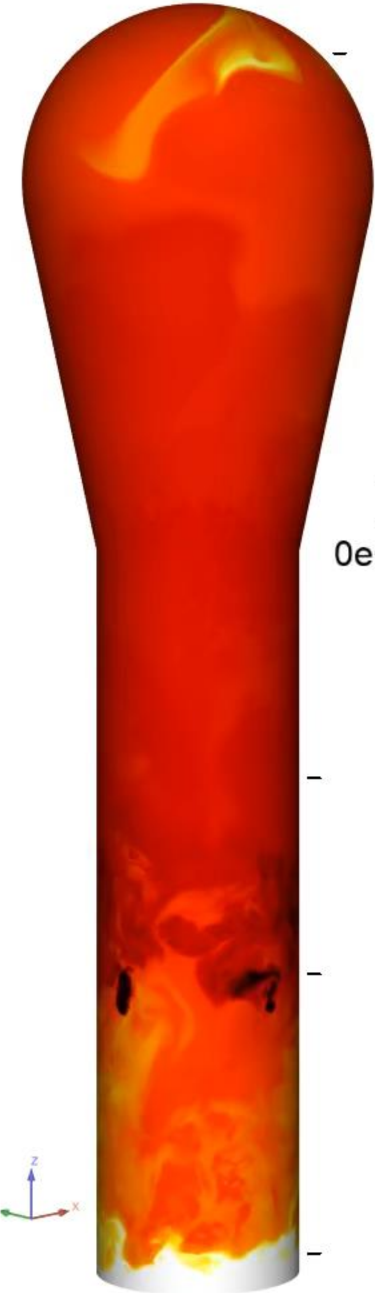




GC_a3full_test04.mp4

cnrs
INP TOULOUSE
calmip
cnrs IDRIS
TGCC
PRACE
Herve Neau
Maxime Pigou

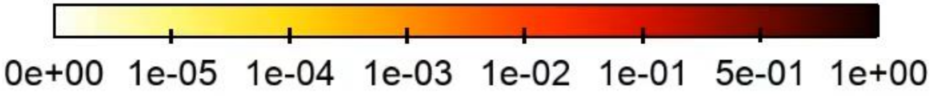
Industrial-scale Polydispersed Reactive Fluidized Bed 3D simulation with NEPTUNE_CFD on Jean-Zay (IDRIS)

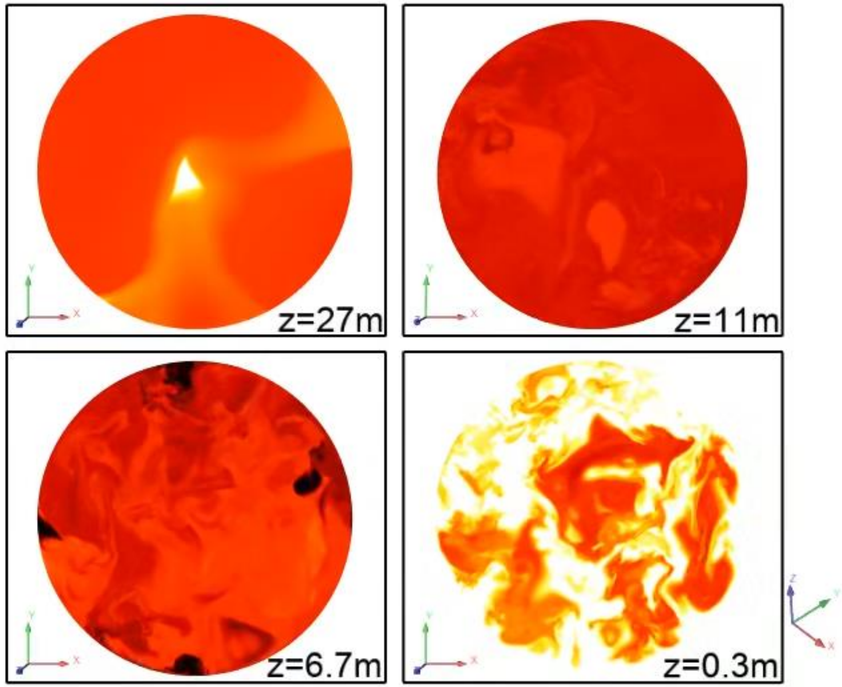


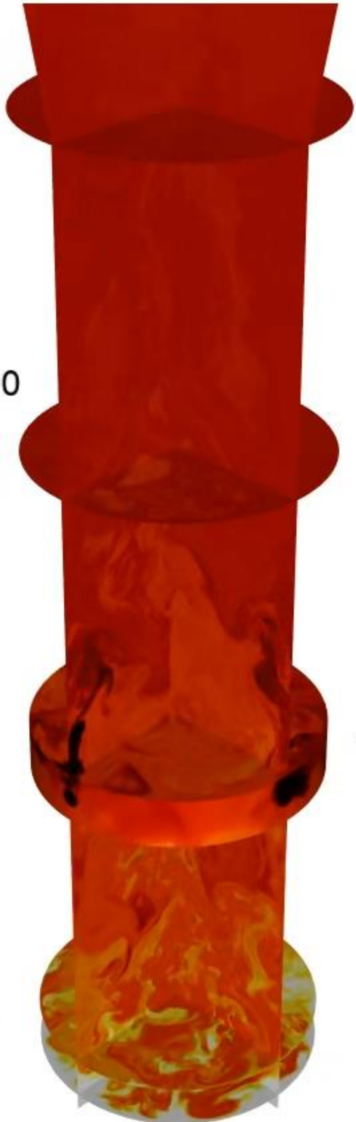
Unstructured mesh of 8,018,843,648 cells
8,560 to 51,840 MPI processes

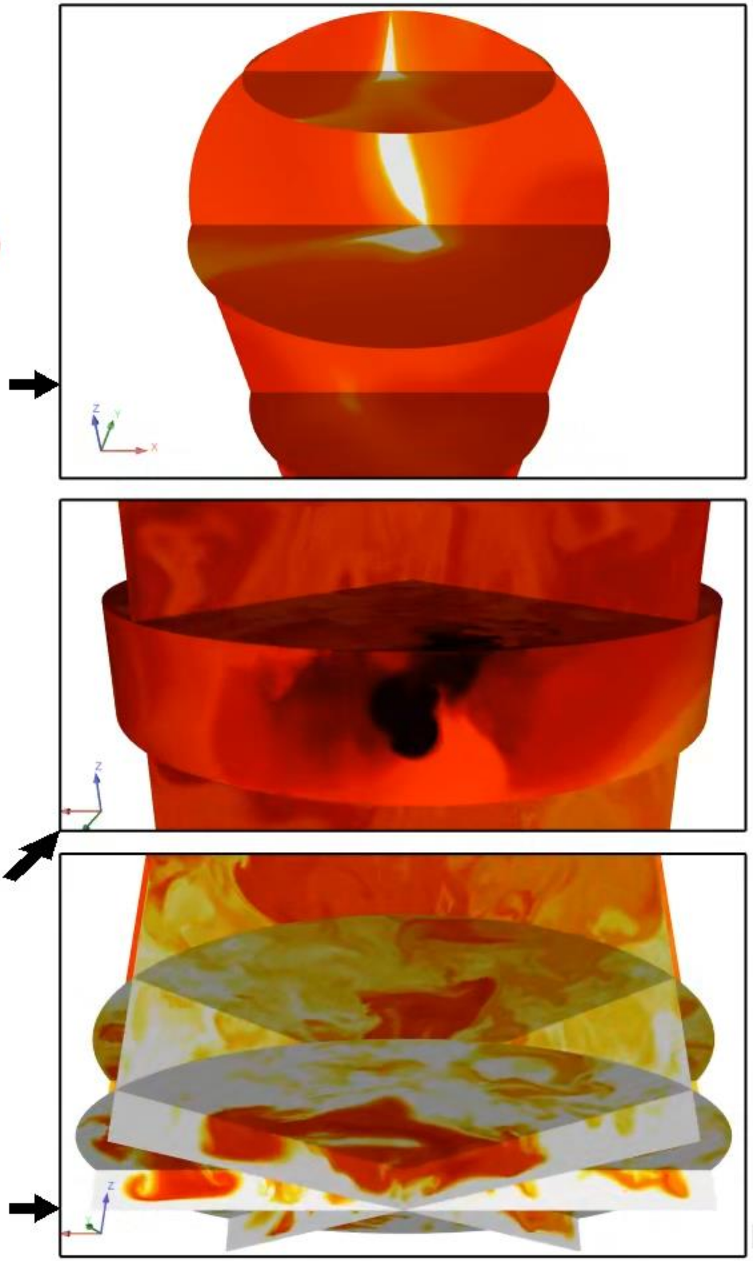
Time = 26.5543s.

Catalyst carrier gas tracer









Herve Neau
Maxime Pigou

Conclusion and prospects: NEPTUNE_CFD HPC capabilities

An era of Worldwide Premiere and of frontier simulations

- **CALMIP/EDF: Worldwide Premiere with 10^9 cells unstructured mesh**

⇨ possibility of physical and statistical analysis (25 s simulated)

- **IDRIS: New Worldwide Premiere with 8 times bigger mesh**

⇨ Limited physical analysis: 1.7 s simulated

⇨ Reaching post-processing limits: storage of 53 TB of data, data transfer limitations, limited toolset for visualization, …

- **TGCC: Yet again a Worldwide Premiere with 64.10^9 cells mesh**

⇨ No physical analysis possible, only few iterations

⇨ Reaching limits of both solver, MPI libraries and supercomputers

⇨ *e.g.* failure when attempting to generate 512 billion cells mesh due to these limitations

Conclusion from JCAD'18

A Worldwide Premiere highly-detailed numerical simulation of industrial reactive fluidized-bed

We have demonstrated we are now able to compute with more than one billion cells using the whole supercomputer capabilities! (up to 36,000 cores ⇔ 1,000 nodes)

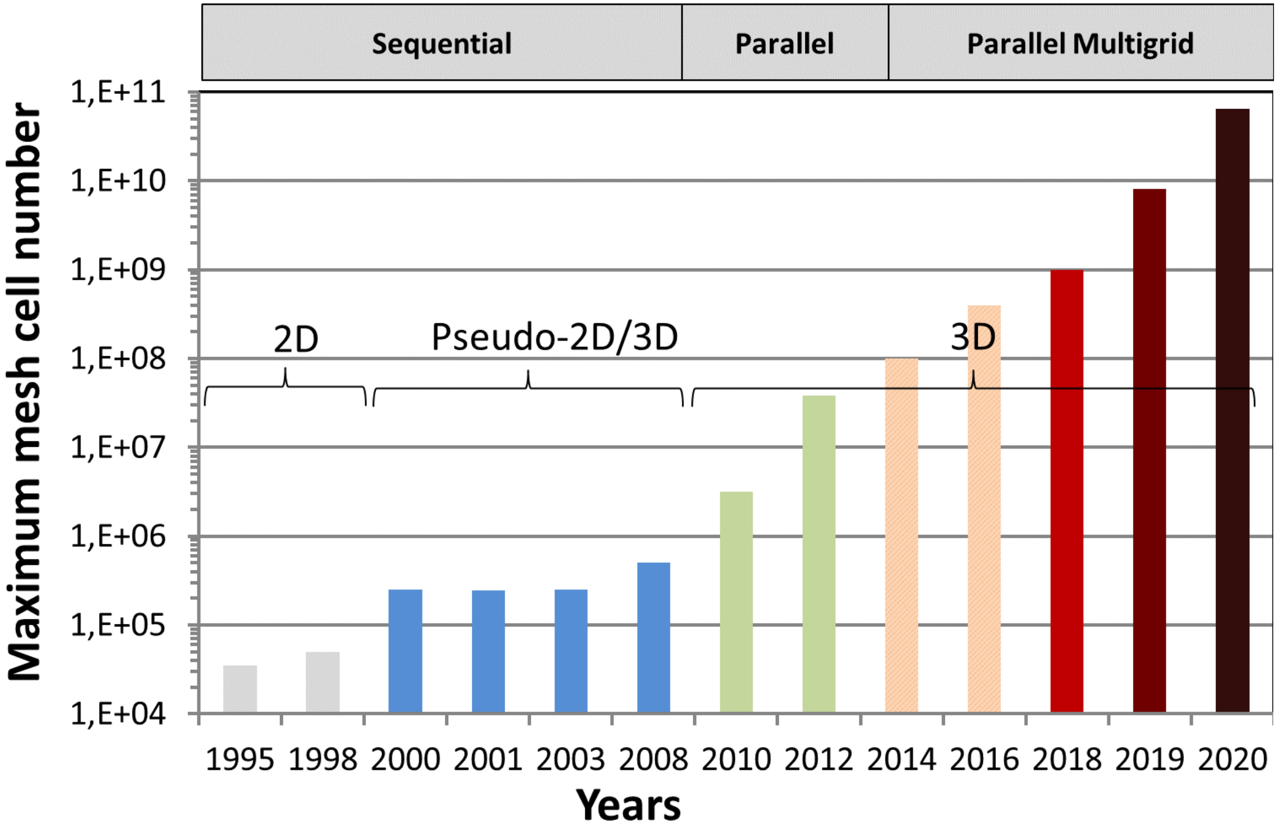
Challenges tackled thanks to **close cooperation between IT, computing and modeling experts**

Each challenge generated an unprecedented database for multiple kinds of analysis:

- Physical analysis to improve fluidized bed modeling
- Advanced solver profiling to further improve NEPTUNE_CFD scaling and prepare for upcoming exascale supercomputer
- Detect and overcome limits in each part of supercomputers hardware

Conclusion and prospects

Exponential growth of simulated case size since 2003



Olympe CALMIP, Gaïa EDF, Jean-Zay IDRIS

3 supercomputers with similar CPU architecture

⇨ Straightforward experience with good performances

⇨ No major impact of interconnects and file systems

Irene-AMD ROME: new CPU architecture thrice the core count but same bandwidth and RAM/node

⇨ Disputable choice of keeping Intel compiler and MPI

⇨ Evaluate OpenMPI library

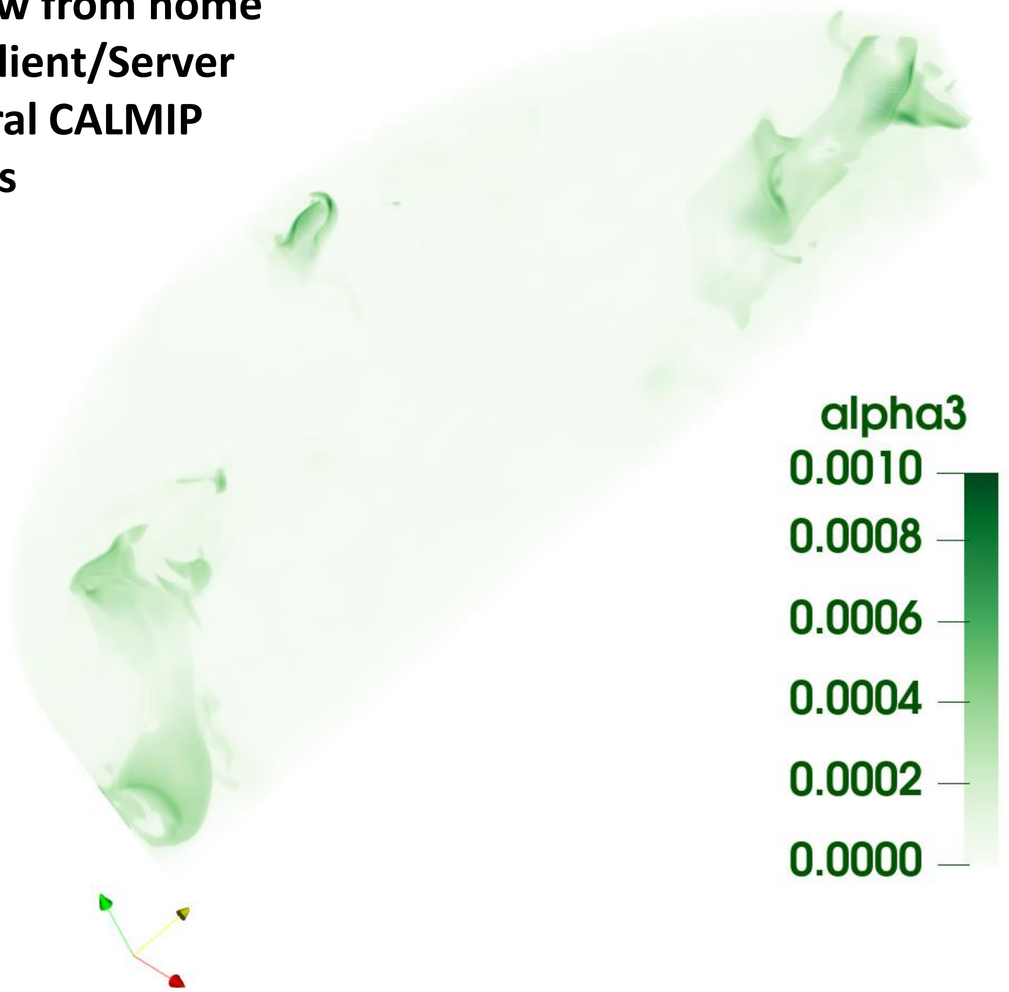
⇨ Necessity to use only 64/128 cores per node

⇨ **RAM/core and bandwidth/core to be increased to match Intel-based supercomputers**

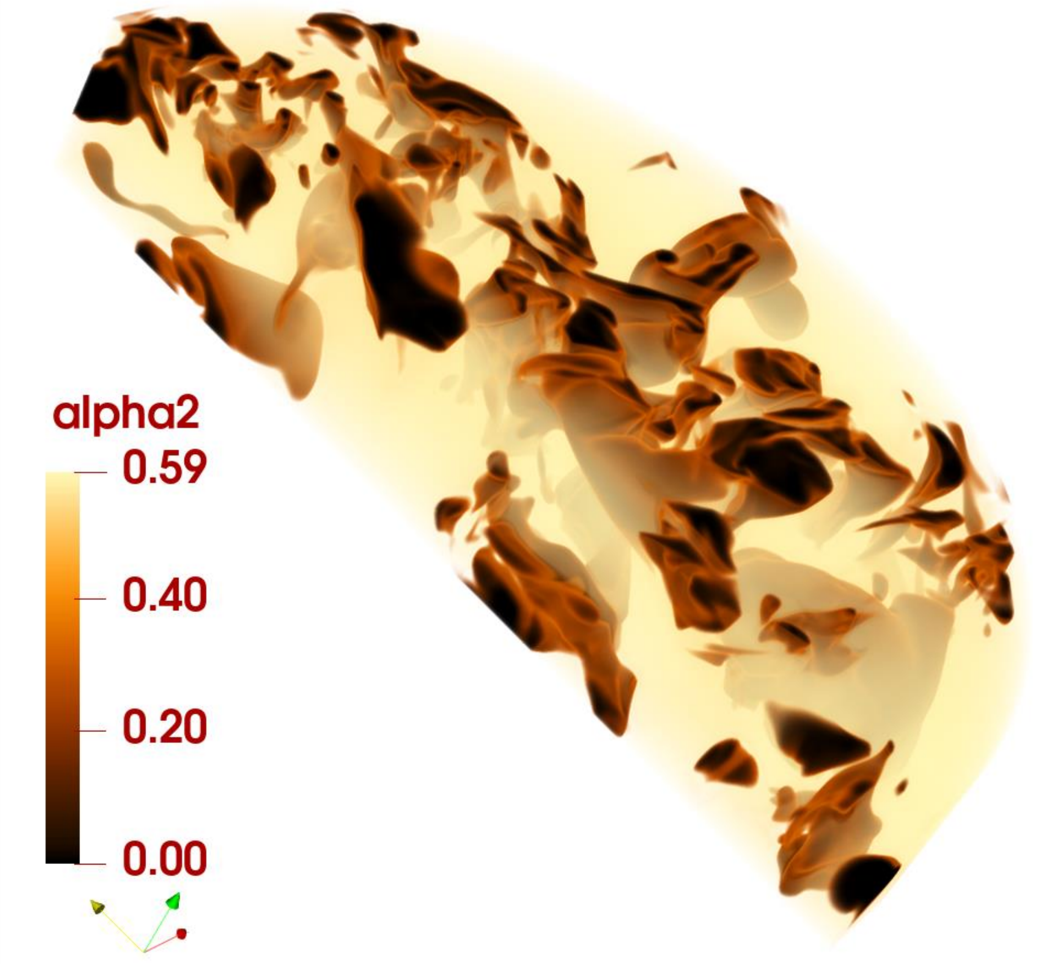
If AMD manycore architecture becomes the new standard: heavy code adaptation required to benefit from full computing potential, *e.g.* hybrid MPI/OpenMP parallelization

From HPC computations to HPC post-processing: ParaView client/server at CALMIP from home

Full packaged and secured script to use ParaView from home (Linux) using Client/Server mode on several CALMIP compute nodes



alpha3
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000



alpha2
0.59
0.40
0.20
0.00