RÉPUBLIQUE FRANÇAISE
*Liberté*
*Égalité*
*Fraternité*

METEO FRANCE

# Météo-France RAPS 2024 benchmark

Ryad El Khatib / Météo-France
Norrköping, April 2024

# Overview

**Introduction (RAPS definition and expectations)**

**Source code description**

**Applications and datasets**

**A few results before the RAPS reports**

**Next steps toward the benchmark**

# Introduction

**R.A.P.S. :**

- Stands for « Real Application for Parallel System »
- a simplified benchmark before a comprehensive benchmark for a procurement
- Delivered at the end of February 2024
- Requires vendors to sign a NDA (Non-Disclosure Agreement)

**Vector engines**


?

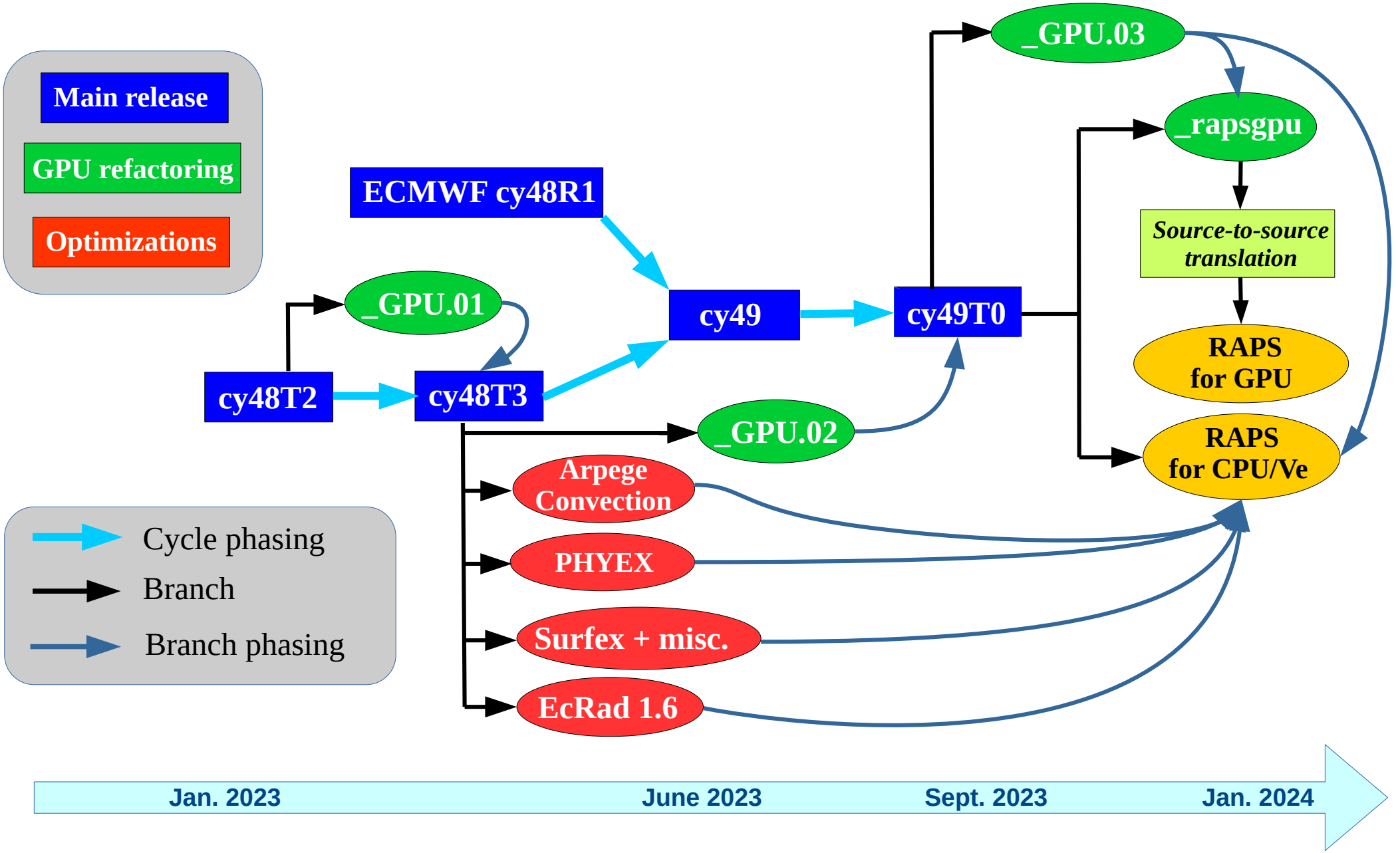
**X86_64, ARM**


**GPU**

# Expected outcomes from a RAPS exercise

- Help the vendors to be ready when the benchmark is delivered (no time spent to get familiar with the build system and the execution scripts, no time lost in fixing portability issues)
- Get feedback about where to improve the benchmark quality (performances issues, scalability, memory cost)
- Get feedback about the feasibility of our expectations (budget, energy cost, etc) against the state of the art of HPC
- Discuss about source code modifications policy for significant optimizations
- Get information about how to formulate the procurement in a way such that it will not disqualify a vendor

# Source code : evolution and status



# Why we provide two flavours of the same source-code

**Recall :**
Porting our applications to GPU accelerators requires deep refactoring of the source code, made in two steps :

- ➔ Refactoring of the top layers of the source-code (in such a way that it will not affect the performance on any architectures)
- ➔ Apply tailored source-to-source transformations to some loops only when targeting a specific GPU accelerator (using fxtran, loki)

- Neither the porting work (in progress) to GPU nor the optimization work for other chips should have slown down the development work of the other teams

- Source-to-source transformers (fxtran, loki) too complex and not mature enough to be let in the hand of benchmarkers (or a huge human support should be provided, too !)

# Status of GPU code

**Recall :**
to be efficient on GPU,
the full time step should be able to run on GPU

- GPU code is provided for ARPEGE only.
  AROME is definitely not ready.

- Only OpenACC port (ie : for Nvidia GPU) is available

- Physical parameterizations except radiation scheme and Surfex can run on GPU

- Nothing more for this benchmark.
  (even not the spectral transforms)

# Applications

ARPEGE and AROME forecasts H24 :

- Configured with cubic grids
- To be run in simple precision
- IO server enabled :
  - ➔ To read AROME coupling files
  - ➔ To write output historical files
- On-line post-processing is disabled :
  - To ease the comparison with the GPU code version
  - => Forecast time-to-solution to be faster by 15 % for ARPEGE and 7 % for AROME)

# Datasets

3 resolutions for each model :

| | ARPEGE | AROME |
|---|---|---|
| **Toy**<br>For porting | T48 linear grid<br>15 levels | E29x29 linear grid (2,5 km), 90 levels |
| **Medium**<br>Operations of today<br>Suitable for optimizations and studies | T1798 linear grid<br>105 levels | E719x767 linear grid (1,3 km), 90 vertical levels |
| **High**<br>Target for the benchmark | T1799 cubic grid<br>120 levels | E624x674, linear grid (750 m), 120 levels |

+ 1 « small » resolution for ARPEGE (T798 linear grid, 90 levels) :

Corresponding to the best we can do
with the few GPU there is on MF supercomputer

# A few results before the RAPS reports

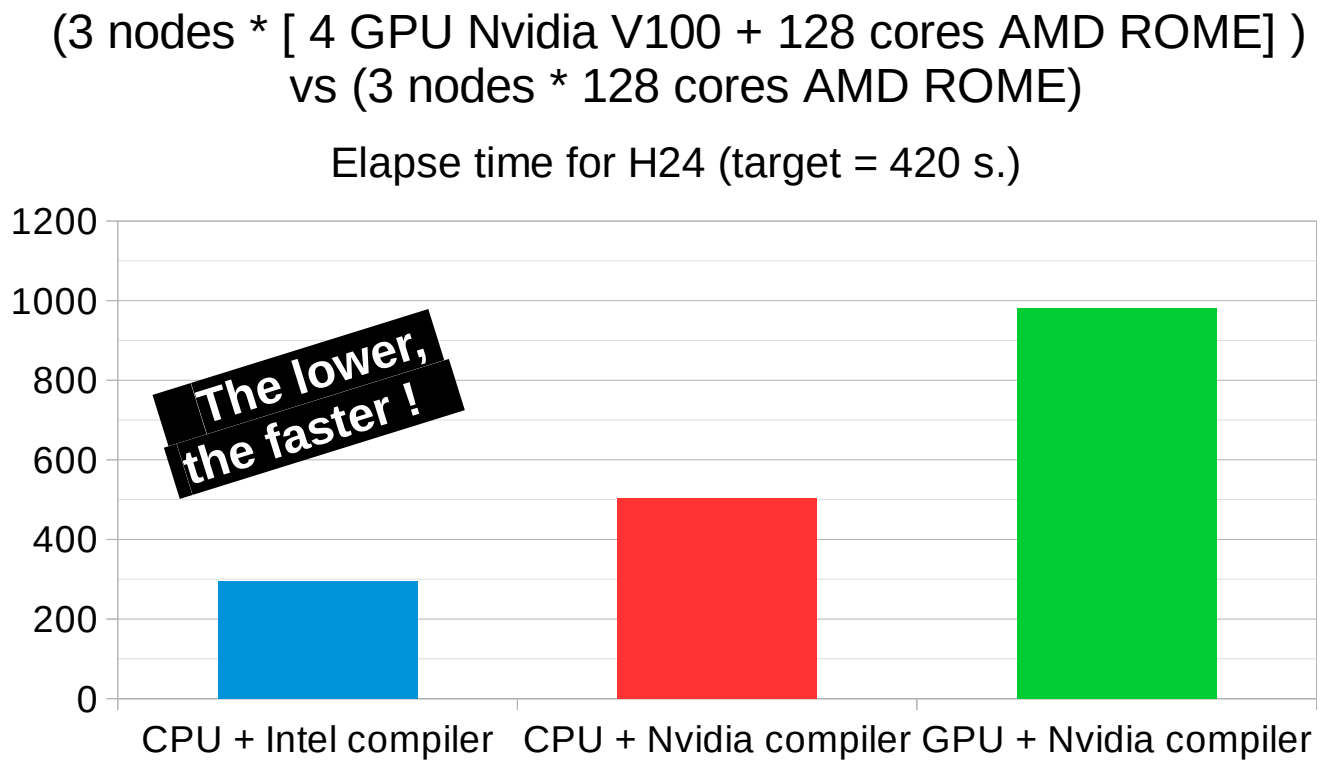
Jobs are running alright at high resolution :-)

Measurements on Atos cluster ‘Belenos ‘ (AMD Rome, 2 sockets x 128 cores per node, 2.2GHz, DDR4 3200 MT/s), 256 GB/node, Intel Compiler 2023.2.0, IntelMpi/2021.10.0:

| Application (High résolution) | Compute nodes | IO nodes | Time to solution |
|---|---|---|---|
| Arpege H24 (target : 7 min) | 190 | 6 | 12 min 22 s |
| Arome H24 (target : 26 min | 109 | 2 | 25 min 49 s. |

*(disclaimer : no fine tuning nor optimizations)*

# Race between CPU and GPU on Arpege T798



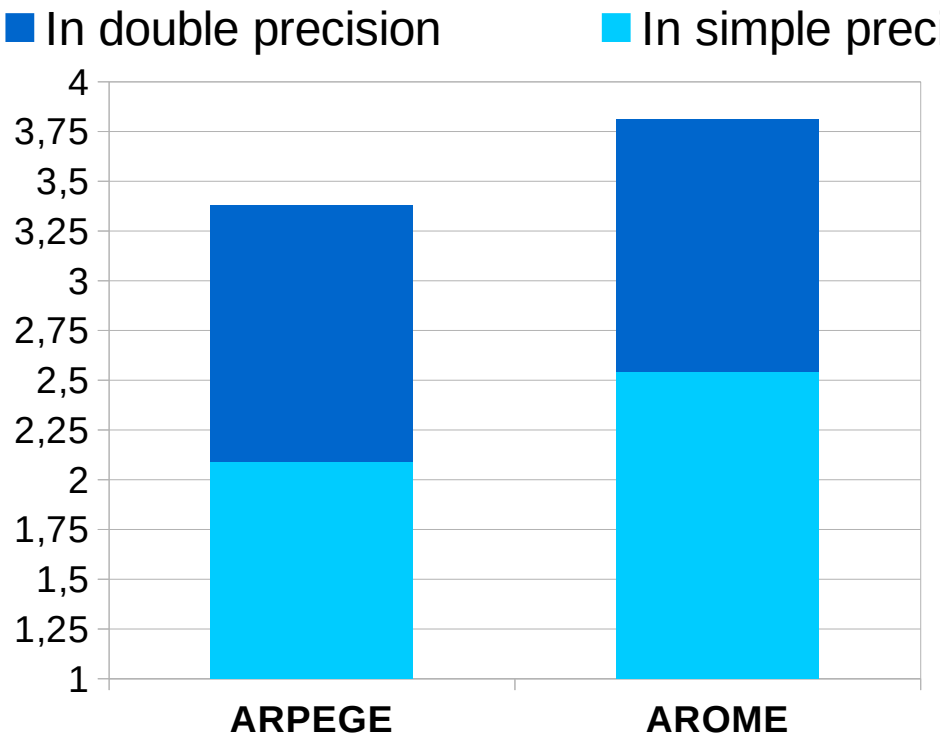

… the compiler matters, too !

# Race between CPU and NEC Vector Engines

Ratio of "System Billing Units"*, or

***how many nodes AMD Rome for 1 node VE 20B to run operations***

≈ 3,5 nodes
AMD Rome
for 1 node
Aurora VE 20B
In double precision




But only
≈ 2,25 nodes
AMD Rome
for 1 node
Aurora VE 20B
In simple precision

* Not considering the energy consumption

# Next steps toward the benchmark

- Select a candidate-application for assimilation
  (Arpege 4DVar or Arome 4DEnvar) and add it to the benchmark suite

- Upgrade the source-code :
  - ➔ Switch to cycle 49T2 (= 49t1 + additional refactoring for GPU)
  - ➔ Backphase optimizations from 49T0_raps not yet in 49T2

- Consider feedback from RAPS reports

- Add more GPU optimizations with OPENACC (spectral transforms, ecRad)

- Work on more memory savings ? (to the benefit of High Bandwidth Memory)

RÉPUBLIQUE FRANÇAISE
*Liberté*
*Égalité*
*Fraternité*

METEO FRANCE

# Thank you for your attention !