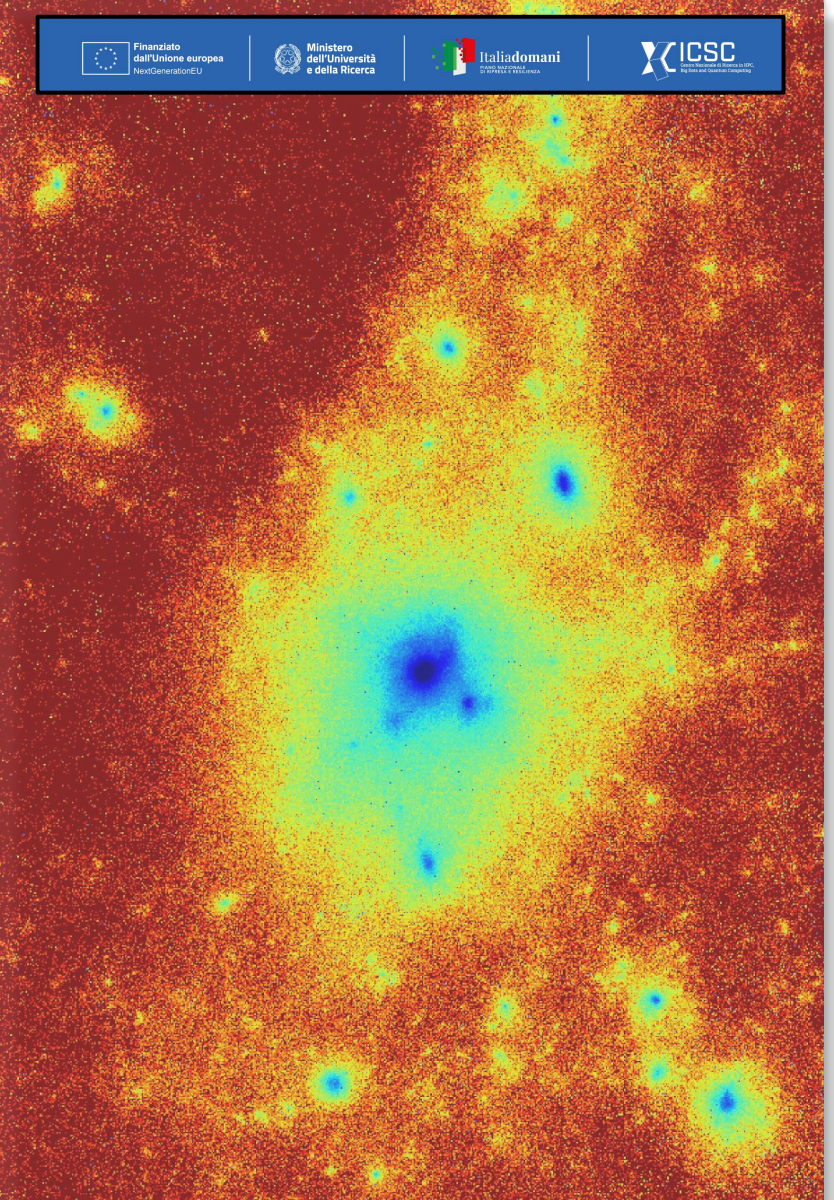


Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

RAMSES GPU

Presented by:
Raffaele Pascale

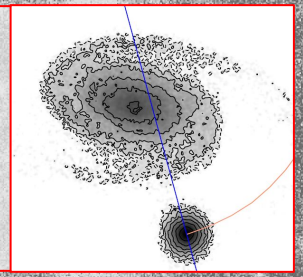
Collaborators:
Francesco Calura, Claudio Gheller, Emanuele De Rubeis, Donatella Romano, Valentina Cesare

INAF OAS

WPs meeting - 20/12/23

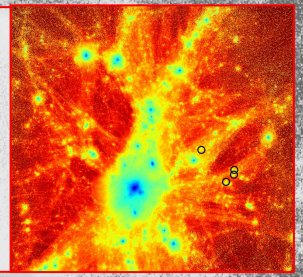
Context:

- In modern astrophysics, hydrodynamical *N*-body simulations play a pivotal role in testing theories of galaxy formation and evolution.
- The need for a deep understanding of the involved physical processes drives the demand for simulations with higher spatial resolution.



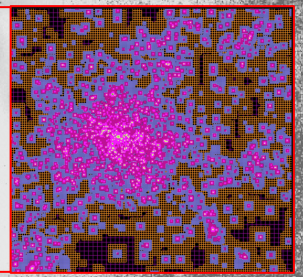
Computational Challenges:

- As spatial resolution increases, computational demands escalate dramatically, posing a significant challenge.
- Addressing this challenge requires innovative solutions to optimize and accelerate computations.



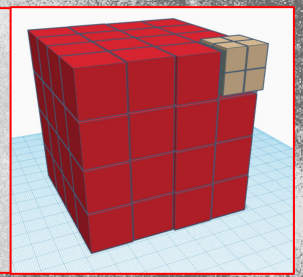
GPU Optimization Strategy:

- An effective strategy involves porting and optimizing specific, time-consuming modules of N-body hydrodynamical codes onto GPU architecture.
- This optimization aims to significantly reduce computational time and enhance the efficiency of simulations.



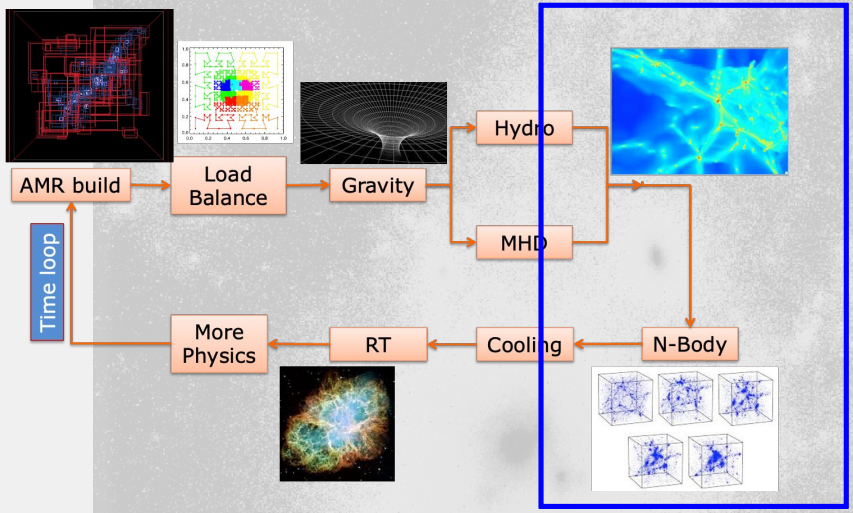
Application to (MINI)RAMSES:

- Focusing on the (MINI)RAMSES code, known for its Eulerian nature and specialization in cosmological simulations.
- (MINI)RAMSES, written in Fortran90, uses Adaptive Mesh Refinement (AMR), dynamically increasing spatial resolution in regions meeting specific criteria (e.g., mass density).



Main goals:
Improving Performance and **Reducing** Computational Time.
Porting hydrodynamical and *N*-body modules to **GPU architecture**, yielding a 'substantial' speedup factor

how RAMSES works

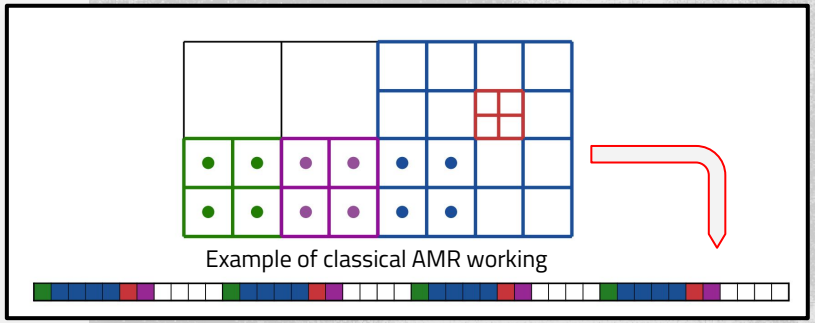


Main goals:
Improving Performance and **Reducing** Computational Time.
Porting hydrodynamical and *N*-body modules to **GPU architecture**, yielding a 'substantial' speedup factor

RAMSES exploits **AMR** (adaptive mesh refinement) **to increase resolution** on the mesh only where it is needed

Not suited for GPU porting

RAMSES



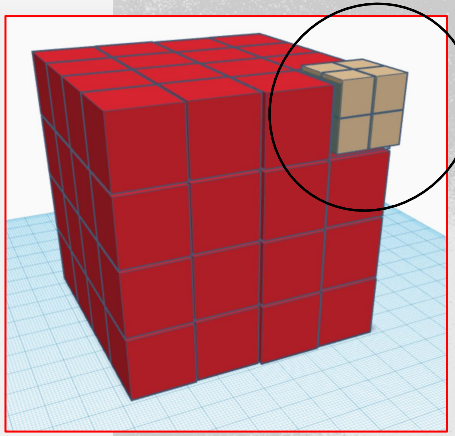
- Identification of cells suitable for refinement
- creation of grids with higher resolution
- **Refined cells saved in non-contiguous memory**

Main goals:
Improving Performance and **Reducing** Computational Time.
Porting hydrodynamical and *N*-body modules to **GPU architecture**, yielding a 'substantial' speedup factor

RAMSES exploits **AMR** (adaptive mesh refinement) **to increase resolution** on the mesh only where it is needed

Moved to **MINIRAMSES** with a better optimization of AMR, **designed for GPU parallelization**

MINIRAMSES



- Introduces the concept of super-oct in cell refinement.
- groups of adiaject ocs saved in contiguous memory locations
- **cells in super-oct saved in contiguous memory**

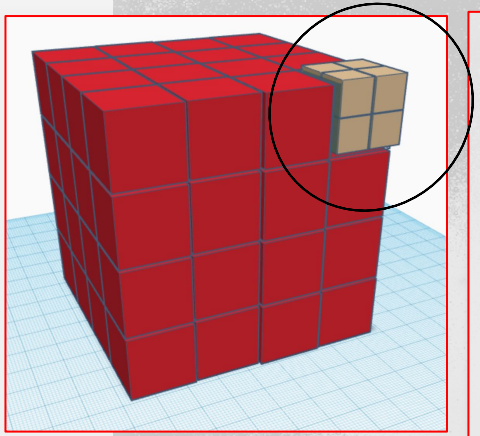
Main goals:
Improving Performance and **Reducing** Computational Time.
Porting hydrodynamical and *N*-body modules to **GPU architecture**, yielding a 'substantial' speedup factor

Identification of two main parts of the code suitable for GPU porting:
***N*-body + Hydro**

The **porting strategy** is based on:

- OpenACC directives to parallelize time-consuming loops and critical code regions;
- applying optimization techniques such as memory management, kernel optimization, and reduction of communication between CPU and GPU;
- profiling methods.

MINIRAMSES



- Introduces the concept of super-oct in cell refinement.
- groups of adiajject ocs saved in contiguous memory locations
- **cells in super-oct saved in contiguous memory**

Timescale and milestones

Analysis and Preparation:

Investigation of MINIRAMSES to identify sections suitable for GPU parallelization - *partially done (hydro and Nbody)*

GPU porting and parallelization of hydrodynamical modules

Identification of suitable modules to port on GPU (*partially done - Profiling*)
Implementation and run of suitable tests (hydrodynamical sedevo3d test).
Evaluation of initial performance and identification of any issues or bugs
Optimization of the code on GPU to maximize performance

Testing phase (Hydro)

GPU porting and parallelization of Nbody modules

Identification of suitable modules to port on GPU
Implementation and run of suitable tests (cosmological simulation)

Testing phase (Nbody)

Integration (Nbody and Hydro):

Integration in principal version of the code
Execution of tests to evaluate scalability

Accomplished work - KPIs
(presentation during hackathon)
May 2023 - August 2023

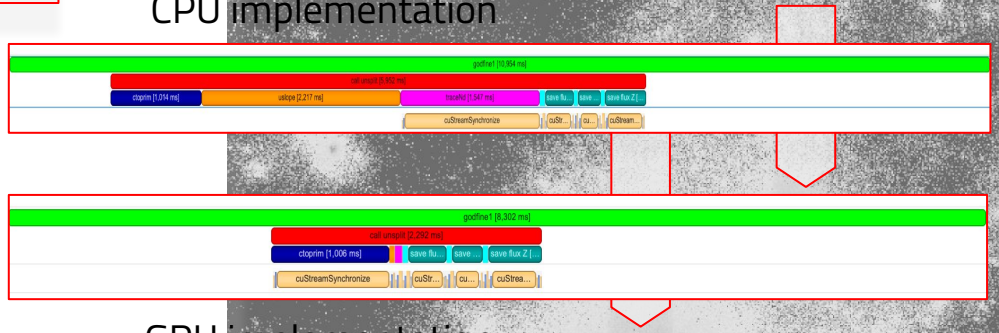
Project started during June 2023, taking advantage of the **hackathon event @ Cineca**

Focus on MINIRAMSES, abridged version of RAMSES designed for GPU porting

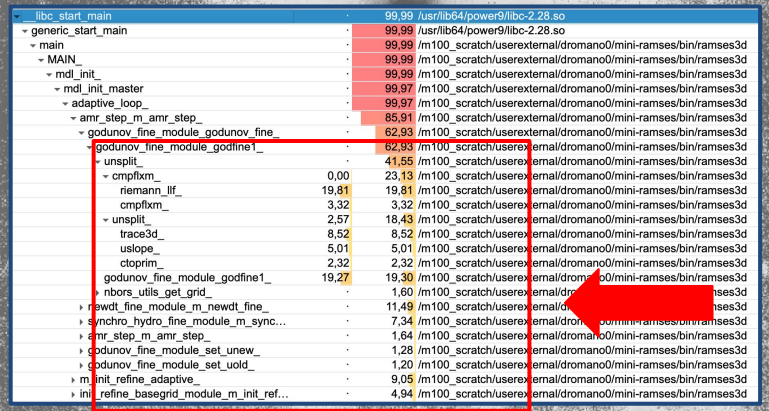
Accomplished work

- Identification of modules hydrodynamical modules to port on GPU.
- Partial GPU porting using openACC directives.
- Generation of profiling data to assess code performance before and after the porting.
- Analysis of code performance on the GPU, especially memory transfers.

CPU implementation



GPU implementation



Ongoing work - KPIs
September 2023 - February 2024

Submission of proposal for GPU hours @ Cineca (ISCRA C) - **done and accepted**

CINECA ISCRA

ISCRA Application form

Class C Projects

code: HP10CLVXSG

Section 1: You and Your Group

Accepted

Principal Investigator

Title	*Dr*
Name	*Raffaele*
Surname	*PASCALE*
Position	*Researcher*
Institution	*Istituto Nazione di Astrofisica (INAF)*
Department	*INAF OAS Osservatorio di Astrofisica e Scienza dello spazio di Bologna*
Address	*via piero gobetti 93/3, 40129, BO, Bologna*
E-mail	*raffaele.pascale@inaf.it*
Phone Number	*+393349265305*

Ongoing work - KPIs
September 2023 - February 2024

Submission of proposal for GPU hours @ Cineca (ISCRA C) - **done and accepted**

Get the code working on Leonardo

Attempt to replicate the same results: profiling data to assess code performance

CINECA ISCRA

ISCRA Application form

Class C Projects

code: HP10CLVXSG

Section 1: You and Your Group

Accepted

Principal Investigator

Title	*Dr*
Name	*Raffaele*
Surname	*PASCALE*
Position	*Researcher*
Institution	*Istituto Nazione di Astrofisica (INAF)*
Department	*INAF OAS Osservatorio di Astrofisica e Scienza dello spazio di Bologna*
Address	*via piero gobetti 93/3, 40129, BO, Bologna*
E-mail	*raffaele.pascale@inaf.it*
Phone Number	*+393349265305*

Ongoing work - KPIs
September 2023 - February 2024

Submission of proposal for GPU hours @ Cineca (ISCRA C) - **done and accepted**

Get the code working on Leonardo

Attempt to replicate the same results: profiling data to assess code performance

Problems:

- unable to use nvidia compilers (now solved?)
- unable to make nvtx work

CINECA ISCRA

ISCRA Application form

Class C Projects

code: HP10CLVXSG

Section 1: You and Your Group

Accepted

Principal Investigator

Title	*Dr*
Name	*Raffaele*
Surname	*PASCALE*
Position	*Researcher*
Institution	*Istituto Nazione di Astrofisica (INAF)*
Department	*INAF OAS Osservatorio di Astrofisica e Scienza dello spazio di Bologna*
Address	*via piero gobetti 93/3, 40129, BO, Bologna*
E-mail	*raffaele.pascale@inaf.it*
Phone Number	*+393349265305*

The End