11TH EGU GALILEO CONFERENCE

Solid Earth and Geohazards in the Exascale Era

Towards exascale-prepared codes for tsunami simulation

M. de la Asunción, J. Macías, M. J. Castro

University of Malaga

Contents

1 Introduction to HySEA

1.1 HySEA

- **HySEA** (Hyperbolic Systems and Efficient Algorithms) is a high-performance package developed by the EDANYA group at the University of Malaga, Spain, for the simulation of geophysical flows.

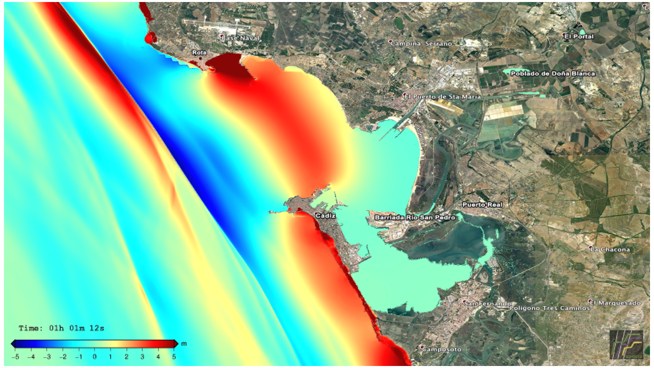
1.2 Main Components of HySEA

- **Tsunami-HySEA**: Simulation of tsunamis generated by earthquakes (multi-GPU).
- **Landslide-HySEA**: Simulation of tsunamis generated by landslides (multi-GPU).
- Other codes to simulate sediment bedload transport, turbidity currents, multilayer flows, dam breaks, landslides with AMR, etc.
- HySEA web page: `https://edanya.uma.es/hysea`

1 Introduction to HySEA

1.3 Tsunami-HySEA

- The deformation of the seafloor (using the Okada model), propagation and inundation are performed in a single code.
- *Features*: multi-GPU (MPI+CUDA), nested meshes, Kajiura filter, multiple Okada faults at any time, resume a stored simulation, compressed NetCDF output files, multiple metrics, time series, asynchronous file writing, variable friction, etc.



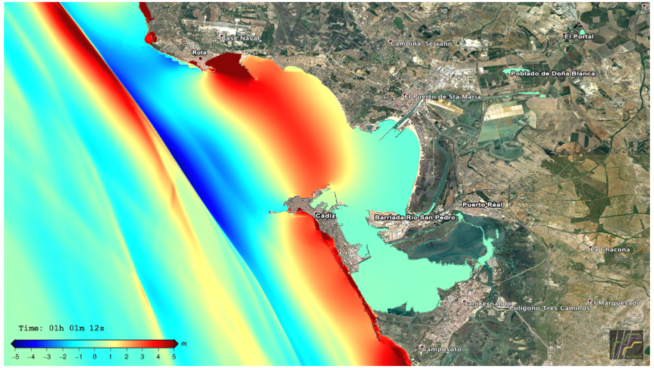
1 Introduction to HySEA

1.3 Tsunami-HySEA

- The deformation of the seafloor (using the Okada model), propagation and inundation are performed in a single code.
- *Features*: multi-GPU (MPI+CUDA), nested meshes, Kajiura filter, multiple Okada faults at any time, resume a stored simulation, compressed NetCDF output files, multiple metrics, time series, asynchronous file writing, variable friction, etc.

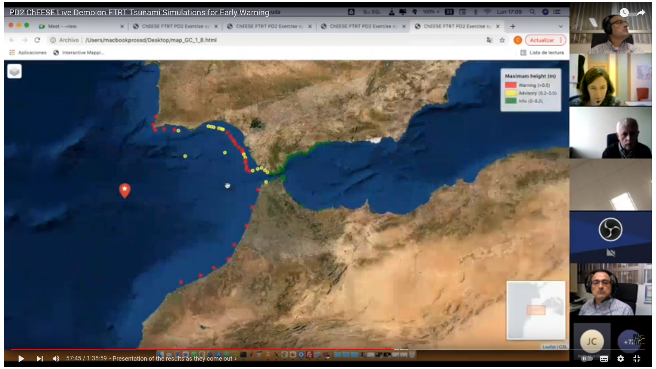
1.4 Tsunami-HySEA Monte Carlo

- A version to perform multiple simulations in the same job, with each simulation running on one GPU. Useful in Probabilistic Tsunami Hazard Assessment (PTHA).
- Used in a live demo on FTRT tsunami simulations with IGN, UMA and CINECA to predict alert levels.



2 ChEESE Live Demo

- Live demo on FTRT tsunami simulations performed by IGN, UMA and CINECA on 11/2021.
- Goal: predict alert levels in coasts of Spain, Portugal and Morocco for a tsunami in the Gulf of Cádiz with variability in magnitude, location, strike and dip angles of the source.
- Schedule: 1) IGN sends UMA data of the seismic sources, 2) UMA launches the simulations on M100, 3) Results are presented as they are obtained.
- 135 scenarios x 4 domains = 540 simulations.
  68 nodes reserved at M100 (272 GPUs).
- A rough estimation at 30 arc-sec in 1 min.
  Refined results at 15 arc-sec in 4 min.
  More refined results at 7.5 arc-sec in 7 min.
- Available at YouTube's grupoedanya channel.



3 Audits in ChEESE

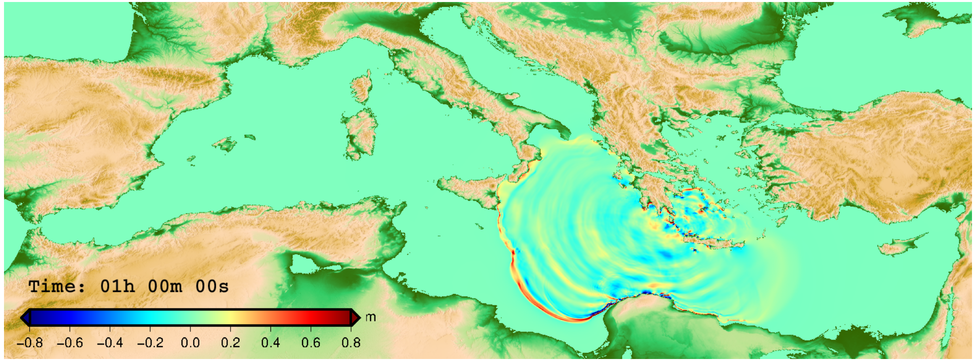
- Performed by the Performance Optimisation and Productivity Centre of Excellence in HPC.
- **First Audit (2019)**
    - **Problems**: 1) Poor load balancing, 2) Synchronous transfers between CPU and GPU memories (cudaMemcpy lasted 24 % of the runtime using 64 GPUs).
    - **Actions**: 1) Recompute the load balancing weights, 2) Implement asynchronous CPU-GPU memory transfers, 3) Implement direct GPU-GPU memory transfers.
    - **Results**: Runtime reduces 33-39 %, and scaling improves 30-40 %, depending on the number of processes.

3 Audits in ChEESE

- Performed by the Performance Optimisation and Productivity Centre of Excellence in HPC.
- **First Audit (2019)**
  - **Problems**: 1) Poor load balancing, 2) Synchronous transfers between CPU and GPU memories (cudaMemcpy lasted 24 % of the runtime using 64 GPUs).
  - **Actions**: 1) Recompute the load balancing weights, 2) Implement asynchronous CPU-GPU memory transfers, 3) Implement direct GPU-GPU memory transfers.
  - **Results**: Runtime reduces 33-39 %, and scaling improves 30-40 %, depending on the number of processes.
- **Second Audit (2020)**
  - **Problems**: Many processes wait too much time during the reduction step.
  - **Actions**: Change some instructions of the reduction kernel.
  - **Results**: Runtime reduces up to 6 %.

4 Runtime Improvement

- Test problem: Tsunami in the Mediterranean Sea.
- Resolution: 30 arc-sec.
- Mesh size: 10.0 Mcells.
- Simulation time: 8 hours.



4

Runtime Improvement

Year	Notes	Runtime (s)	Architecture
2014	Early T-HySEA version	378.1	8 GTX Titan Black
2017		312.1	8 GTX Titan Black
2017		257	2 Tesla P100
2019		284.1	1 NVIDIA V100
2019	Before ChEESE 1st code audit	66.5	8 NVIDIA V100
2019	After ChEESE 1st code audit	48.6	8 NVIDIA V100
2020	After ChEESE 2nd code audit	45.5	8 NVIDIA V100
2021		187.4	1 NVIDIA A100

5 Asynchronous File Writing

- Implemented using C++11 threads.
- Computer: Intel Xeon E5-2620 v4 with 2 Tesla P100.
- Tohoku problem, 1 hour of simulation.
- Storing water height, velocities, max. water height and arrival times, no time series.

	Runtime sync. writing (sec)	**Runtime async. writing (sec)**	**Runtime reduction**
Saving every 10 min.	325.5	256.8	21 %
Saving every 5 min.	385.6	258.1	33 %

6 Saving Disk Space

6.1 Single Precision Storage

- In PTHA applications, we need to store the results of tens of thousands of simulations. Methods to reduce the size of the resulting files are needed.
- All the NetCDF variables are stored in single precision except lon-lat.

6 Saving Disk Space

6.1 Single Precision Storage

- In PTHA applications, we need to store the results of tens of thousands of simulations. Methods to reduce the size of the resulting files are needed.
- All the NetCDF variables are stored in single precision except lon-lat.

6.2 NetCDF Deflate Feature

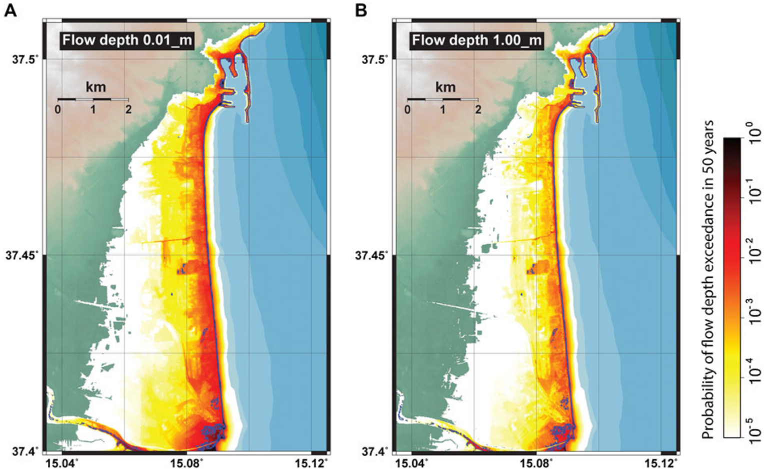
- The *nc_def_var_deflate* C function compresses a particular NetCDF variable.
- Level of compression: 0-9 (0: no compression).
- The more compression, the greater the runtime. Level 5 is a good equilibrium, reaching a compression of 3 – 3.5x in our case.

6.3 NetCDF Scale and Offset Feature

- The scale-offset compression feature of NetCDF scales the values of a variable, so that it is stored using another data type. The range of possible values of the variable should be limited.
- For each variable: real_value = scale_factor x stored_value + add_offset.
- Example: from *float* (4 bytes) to *short int* (2 bytes). 2^{16} values could be stored. You can store the maximum water height from 0 to 150 m. with a resolution of 0.23 cm.
- Added into T-HySEA MC but removed later on due to excessive loss of precision.

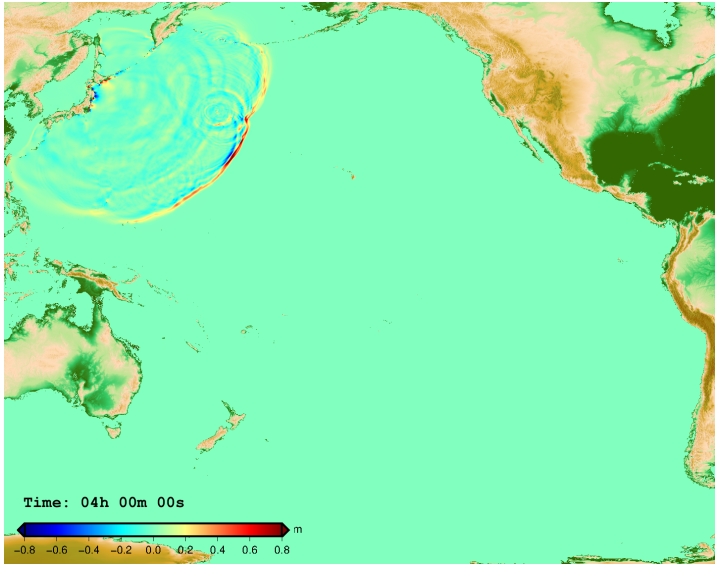
7 T-HySEA Use Case in PTHA

- Probabilistic Tsunami Hazard Assessment in Catania and Siracusa (Sicily, Italy).
- 42,720 scenarios, up to 1,024 simulations in parallel on Marconi-100.
- S. J. Gibbons et al. Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations. Frontiers in Earth Science. 2020.



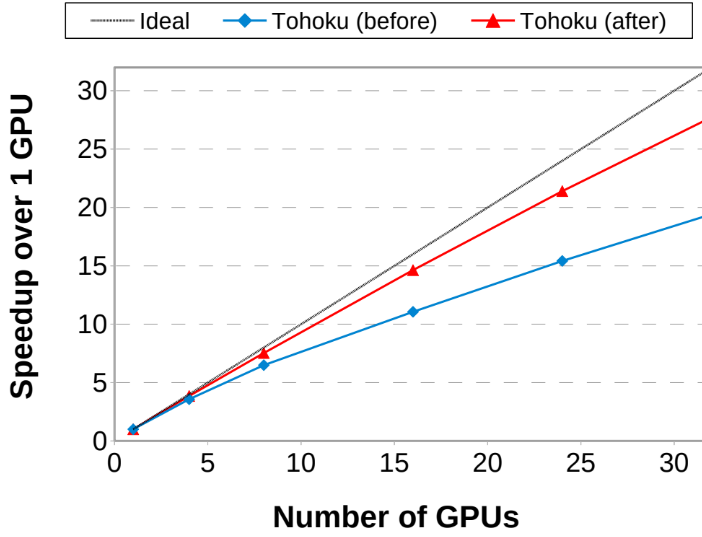
8 Scaling Results

- Test problem: 2011 Tohoku earthquake and tsunami simulation.
- Resolution: 1 arc-min.
- Mesh size: 84.2 Mcells.
- Simulation time: 24 hours.



8 Scaling Results

- Beginning of ChEESE: at CTE-POWER (BSC). End: at Marconi-100 (CINECA).
- Using up to 32 NVIDIA V100.
- Runtime with 1 V100: **7338.7 sec**.



9 Best Practices

- **General Recommendations:**
  - Minimize the size and overhead of communications between processes and between CPU and GPU memories.
  - Obtain good load balancing.
  - Asynchronous file writing.
  - Not use more precision than needed to store the data (usually single precision is enough for many applications).
  - Compress the output files.
  - Use specific instructions for the numerical precision that you are using (e.g. use sqrtf, expf, etc, for single numerical precision operations).

Development and optimization

of HySEA codes

continue in ChEESE-2P

Thank you!