Status Olivia pilot

B. Arntsen

NTNU
Department of Geoscience and petroleum
borge.arntsen@ntnu.no

August 17, 2025

Olivia system

1. Used only the Cray default and nvidia environment
2. Compilers for AMD/ARM/GH200 (gcc and nvcc) works without problems
3. Batch system works without problems for both cpu and gpu jobs
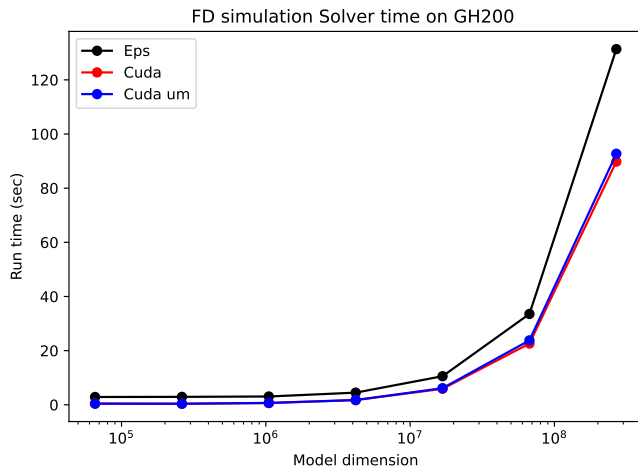4. No change or tweaking of code was necessary

Performance tests

1. Measured runtimes for finite-difference wave equation solver on single GH200 GPU
2. Measured runtimes for finite-difference wave equation solver on 128 core AMD EPYC Turin (omp)
3. Comparison with A100 (Idun) and RTX4070 (Dell Allienware gaming laptop)
4. No comparison with LUMI/AMD (No access to Lumi)

Codes used for performance test

1. Finite-difference wave equation solver in handwritten CUDA with traditional memory managment
2. Finite-difference wave equation solver in handwritten CUDA with unified memory managment
3. Experimental machine generated finite-difference wave equation solver (C,Multicore C (omp),Cuda,Hip)

Runtime GH200

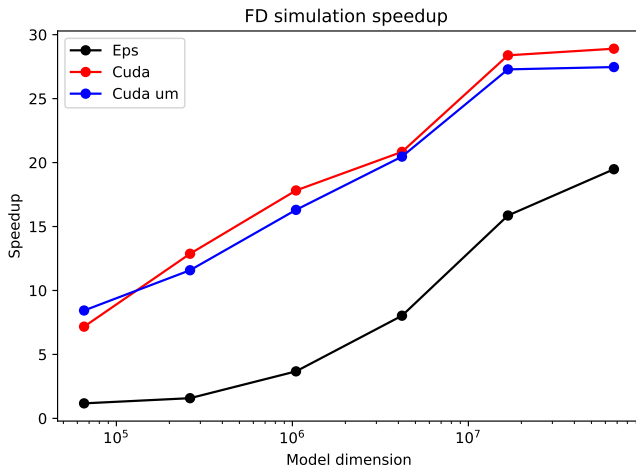FD simulation Solver time on GH200

Eps: Machine generated cuda
Cuda: Handwritten cuda with trad memory managment
Cuda um: Handwritten cuda with unified memory
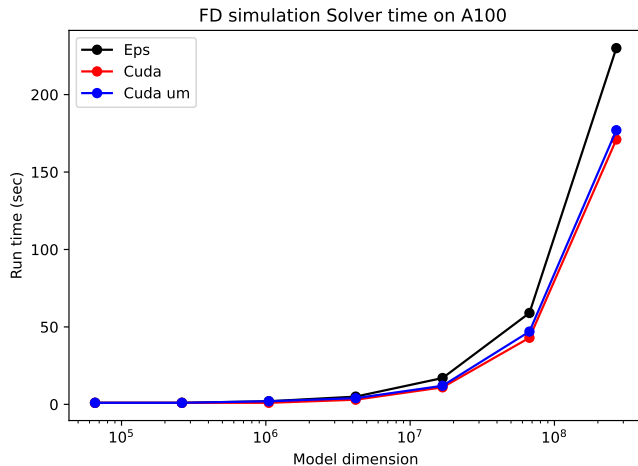
Speedup GH200/AMD Epyc Turin

Relative to AMD EPYC Turin 128 cores (OpenMP, not optimal)
Eps: Machine generated cuda
Cuda: Handwritten cuda with trad memory managment
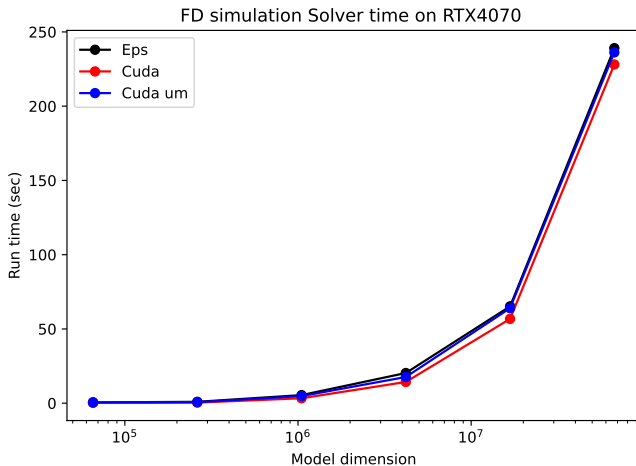Cuda um: Handwritten cuda with unified memory

Runtime A100

FD simulation Solver time on A100

Eps: Machine generated cuda
Cuda: Handwritten cuda with trad memory managment
Cuda um: Handwritten cuda with unified memory

Runtime RTX4070

FD simulation Solver time on RTX4070

Eps: Machine generated cuda
Cuda: Handwritten cuda with trad memory managment
Cuda um: Handwritten cuda with unified memory

Performance

- Single GH200 GPU on Olivia runs twice as fast as single A100 GPU on Idun
- Speedup relative to AMD EPYC Turin node on Olivia approx 30 for handwritten cuda and approx 20 for machine generated cuda
- Results are preliminary, no tweaking of code
- Machine generated code suboptimal, but fix is known.

Conclusions

- Olivia basic system works without any problems
- Olivia Documentation sufficient for pilot (No need for additional support)
- Finite-difference wave simulation codes runs twice as fast relative to Idun (A100)
- No penalty for using unified memory
- Implications for research are faster turn-around due to speed and larger number of gpu's