

HDS028-16

Room:302

Time:May 24 10:15-10:30

## Large-scale Wave Propagation Simulation using Multi GPU on TSUBAME2.0

Shin Aoi<sup>1\*</sup>, Naoki Nishizawa<sup>2</sup>, Takayuki Aoki<sup>3</sup>

<sup>1</sup>NIED, <sup>2</sup>SGI Japan, Ltd., <sup>3</sup>Tokyo Tech

Ground motion simulation using FDM is one of the key techniques for seismic hazard assessment. Huge computation resources are required to perform large-scale wave propagation simulations using realistic models for high accuracy assessment. To estimate the uncertainty of the assessment caused by uncertainty of the assumptions of the source models for future earthquakes, thousands of various simulations are necessary. To overcome this difficulty, we developed, by using CUDA, the multi GPU version of GMS (Ground Motion Simulator) and performed the large-scale simulation on TSUBAME2.0 which is the Japanese fastest supercomputer operated by Tokyo Tech. The original code (CPU version) of GMS (Aoi et al., 2004) is a total system for seismic wave propagation simulation based on 3-D FDM using discontinuous grids (Aoi&Fujiwara, 1999), which includes the solver as well as the preprocessor tools (parameter generation tool) and postprocessor tools (filter tool, visualization tool, and so on). The computational model is decomposed in two horizontal directions and each decomposed model is allocated to a different GPU. Because the values on the grid at the boundary of the neighbor decomposed models are necessary for the calculations, two grids from the boundary are overlapped each other and the values on these grids are exchanged by MPI. Relative time required for the communication compared to the time for the calculation is longer for GPU than for CPU, because the calculation speed is much faster for GPU. Moreover, the overheads for the communication are larger for GPU because direct communications are not available and values are transmitted to the target GPU via CPU using MPI. Therefore the time for the communication is not negligible and the concealment technique of the communication by overlapping the calculation and the communication is essentially important for achieve high performance parallel computation using GPU. Popular technique for concealing the communication is follows: Values on the overlapped grids are calculated first and then the communication of those values between neighbor decomposed models are performed during the calculation of rest grids. This technique is not efficient because it requires discontinuous memory accesses which are hard for GPU. Considering that our discontinuous grids have two regions having different size of grid spacing, exchanges of the values on the overlapped grid in one region are made during the calculation of another region. Our concealment technique makes it possible to avoid the discontinuous memory accesses. We examined the two kind of performance test for parallel computing; weak and strong scaling tests. For the weak scaling test where the model sizes (number of grids) are increased in proportion to the degree of parallelism (number of GPUs), the result showed almost perfect linearity up to the simulation with 256 GPUs. Here we used the model with about 22 million grids as the unit model and the model size for the 256 GPUs case is about 5.5 billion grids. On the other hand, for the strong scaling test where the model size is independent from the degree of parallelism, the speed-up using the unit-model is 3.2 and 7.3 for 4 and 16 GPUs cases, respectively. The reason of the rapid decrease of the parallel performance is that the communication time increases so that it is no longer possible to conceal by calculation time. Moreover, the number of the threads for each GPU decreases because the model size allocated to each GPU becomes too small. Considering that the time steps for most model we use for simulation are up to hundreds thousand, the turn around times are several minutes to a few hours when the GPU resources appropriate to the size of the model is available. Thus, the performance of the GMS on GPU is practically satisfactory for most cases. The turn around times above do no include the time for outputting the result. Efficient parallel output technique maybe necessary for large files.

Keywords: GPGPU, TSUBAME2.0, parallel computing, ground motion simulation, FDM, seismic hazard assessment