学際大規模情報基盤共同利用・共同研究拠点 公募型共同研究 平成26年度採択課題

6th Symposium

Joint Usage / Research Center for Interdisciplinary Large-scale Information Infrastructures

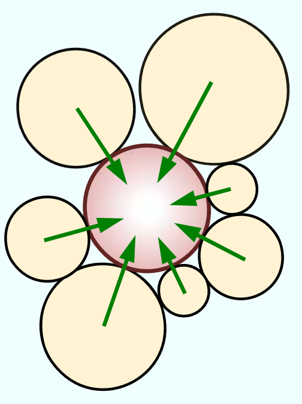
jh140036-NA20

青木尊之（東京工業大学 学術国際情報センター）

動的負荷分散によるGPUスパコンを用いた粒子法の大規模シミュレーション手法の開発

Background

Distinct Element Method (DEM) is used for numerical simulations of granular mechanics. Each particle collides with the contacting particles.



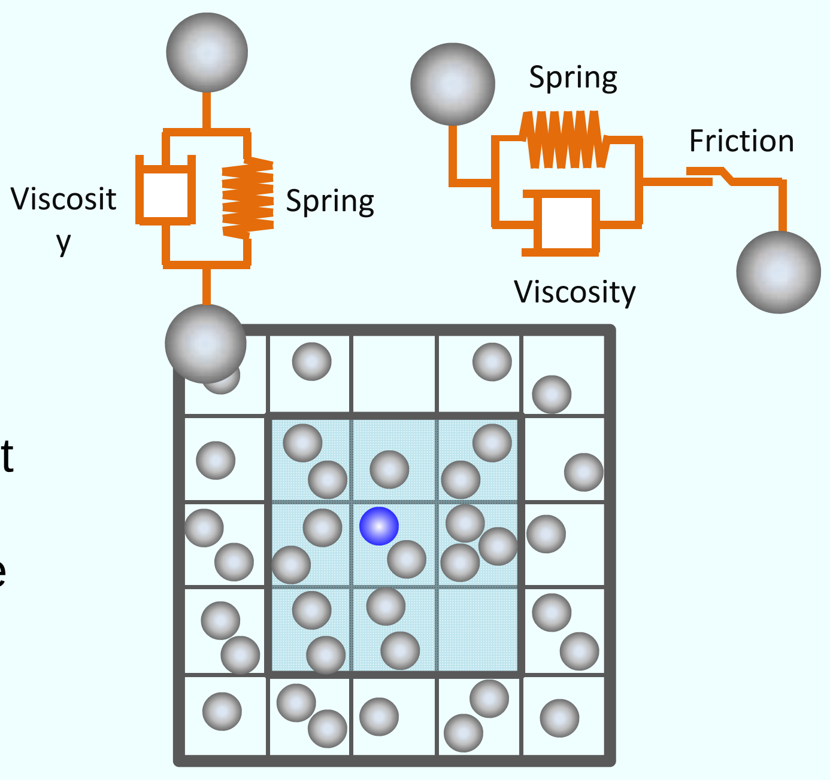
Most of real granular phenomena consist of more than billions of stones or particles. Due to computational resources, a single particle often represents more than 10 thousands particles in the previous simulations. In order to bring the simulation closer to the real phenomena for the purpose of quantitative studies, it is necessary to execute large-scale DEM simulations on modern high-performance supercomputers. However, in the static domain decomposition, the spatial particle distribution changes in time and the computational load for each domain becomes quit non-equal.

Distinct Element Method

✓ The particle interaction is a spring force of the contacting two particles. The friction in the tangential direction is also taken into account. The force among *i*- and *j*-particles is described, as follows:

$$F_{ij} = -kx_{ij} - \gamma \dot{x}_{ij} \qquad (2)$$

✓ The neighbor-particle list is commonly used to reduce the amount of particle interactions. we reduce the computational cost from $O(N^2)$ to $O(N)$, however the memory use often becomes a severe problem in large-scale simulations



Objectives

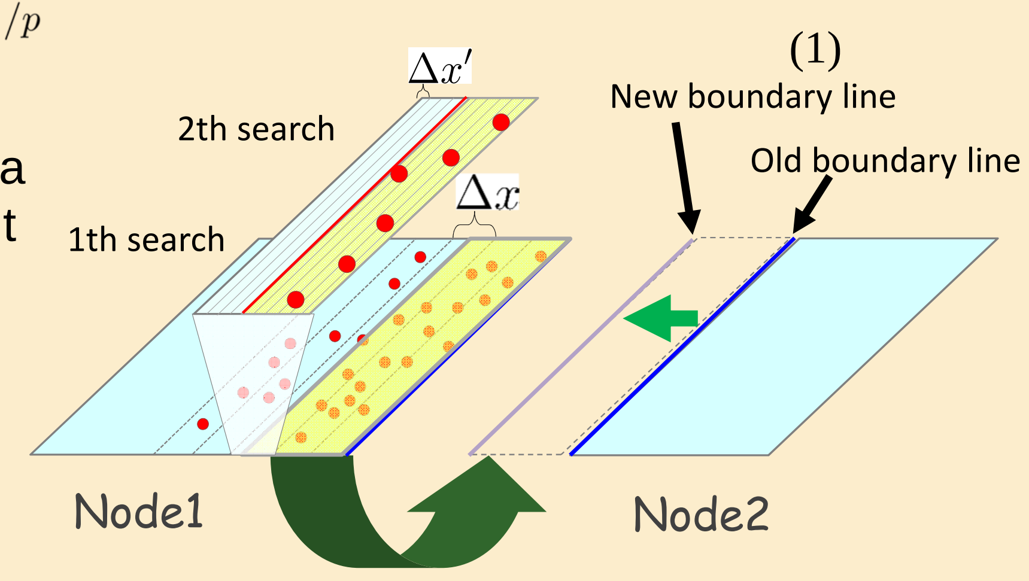
- **Dynamic load balance among GPUs**
  By applying the slice-grid method to our particle simulation, we maintain the same number of particles in each domain.
- **De-fragmentation of GPU memory**
  Frequency of the memory de-fragmentation process is determined by considering the trade-off of the data communication between CPU and GPU.
- **Linked-list method**
  We introduce the linked-list method for the neighbor particle list to save the memory drastically.

Dynamic Load Balance

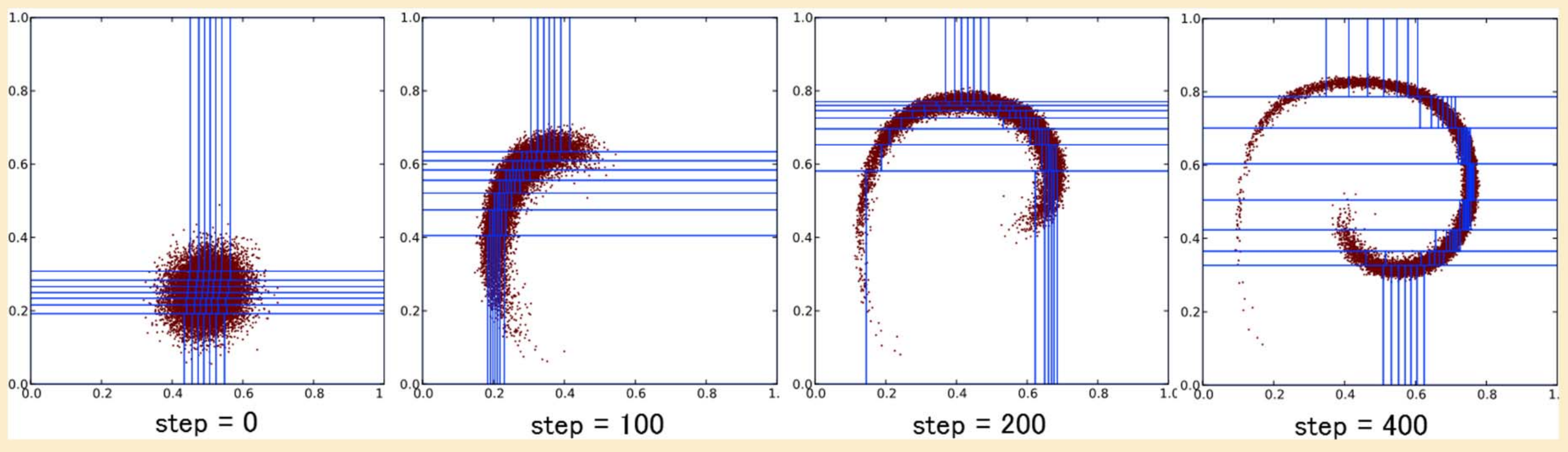
- The whole computational domain is decomposed into several subdomains (local domain) and a GPU is assigned to compute each subdomain.
- The 2-dimensional slice-grid method is introduced to realize equal load balance among GPUs. The horizontal and vertical boundary lines are shifted in turn to keep the same number of particles in each subdomain.

$$\begin{aligned}\Delta N_i &= \Delta N_{i-1} + N_i - N_{total}/p \\ \Delta N_0 &= N_0 - N_{total}/p\end{aligned} \qquad (1)$$

- We divide subdomain into a proper space $\mathit{\Delta x}$ and count the number of particles within $\mathit{\Delta x}$.
- The data of the particles moving to the neighbor subdomains are copied through PCI-Express bus and CPUs memory.

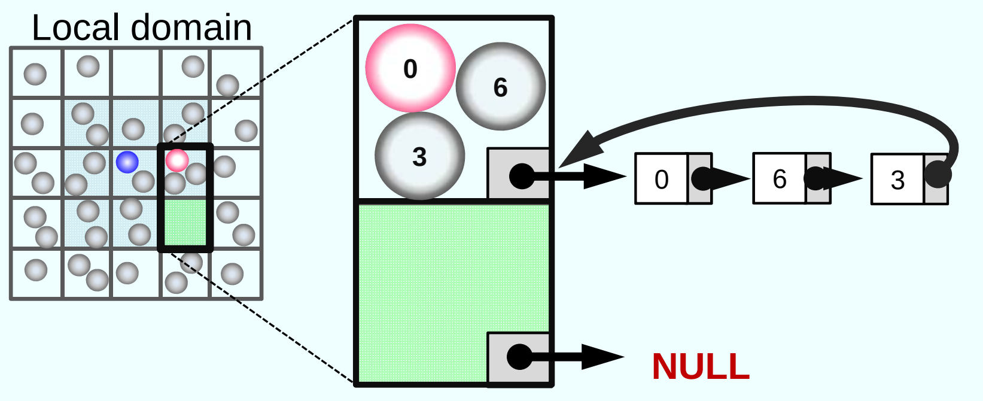


- A benchmark of passive scalar particles for the given vortex velocity.



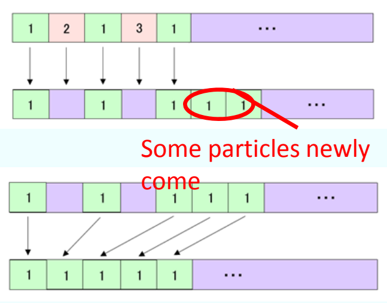
Linked-list Method

✓ A memory pointer is added to each particle data to have a reference to the next particle in sequence in a subdomain.



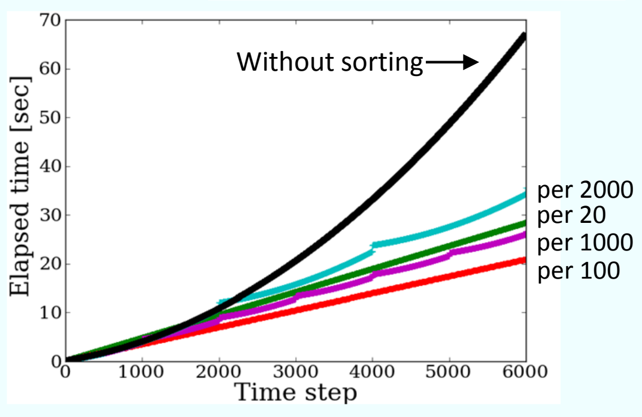
De-fragmentation

✓ Fragmentation of GPU memory degrades the computational performance.



✓ De-fragmentation keeps the performance with linear increase.

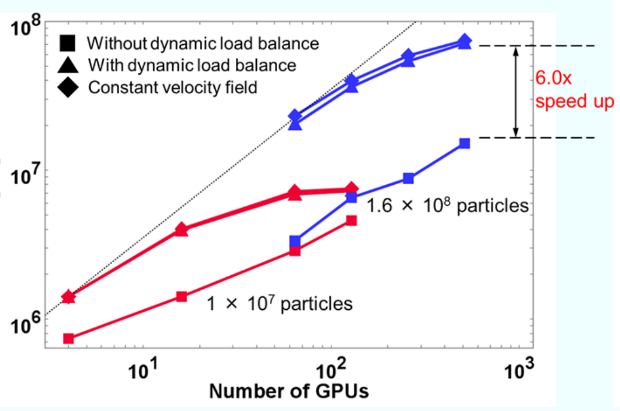
✓ Frequency of de-fragmentation is optimized.

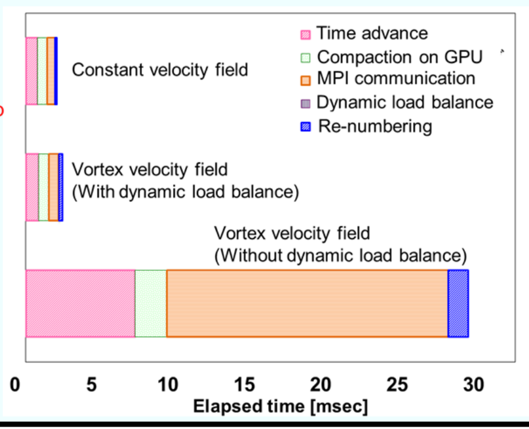


Performance scalability on TSUBAME 2.0

The strong and weak scaling for the benchmark computation of passive scalar particles are measured on TSUBAME 2.0.

Number of particles	:	1.6×10^8
Decomposition	:	2 dimensional slice-grid method
Velocity field	:	Rotational velocity field
Time integration	:	4th-Runnge Kutta
Initial condition	:	Gaussian distribution

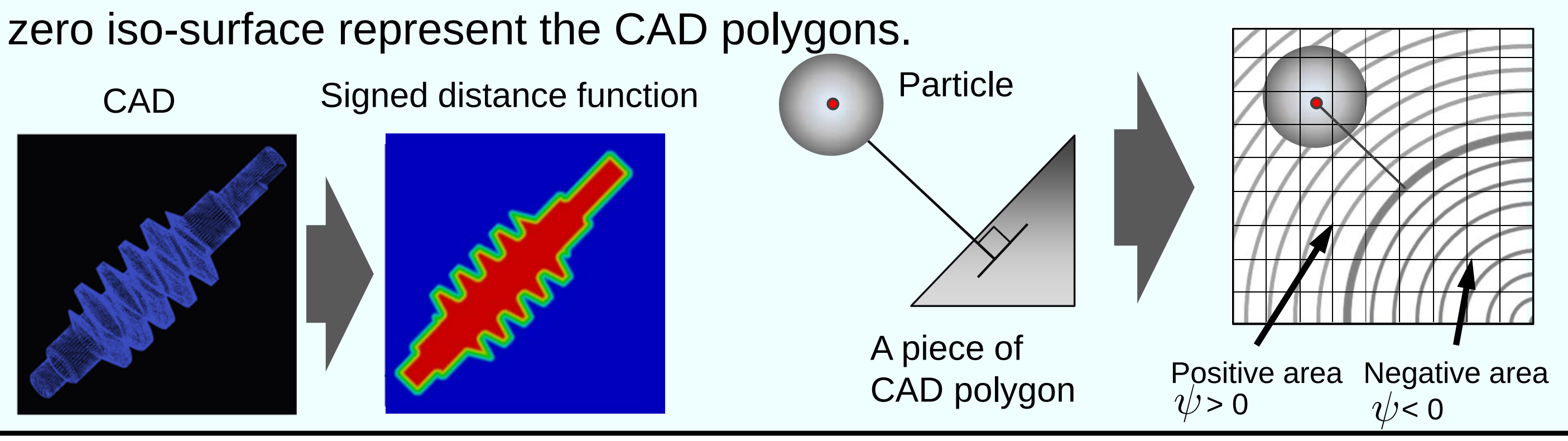




The breakdown of the computational time for the case of 512 GPUs.

Level-Set method

Signed distance field is generated from CAD data and the zero iso-surface represent the CAD polygons.



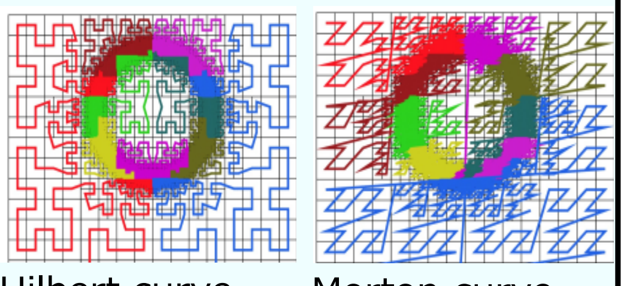
Research Plan

DEM

- Measurement of the weak and strong scalability from 0.1 to 1.0 billon particles on TSUBAME2.5.
- Comparison of the performance depending on memory stores between Array of Structure (AOS) and the Structure of Array (SOA) from the view point of the coalesced memory access to the GPU device memory.
- Optimization by using the On-Chip shared memory of GPU to store the particles within the same CUDA blocks to reduce the device memory access.
- Executions of several practical problems with using 100 million particles such as a golf bunker shot on TSUBAME2.5.

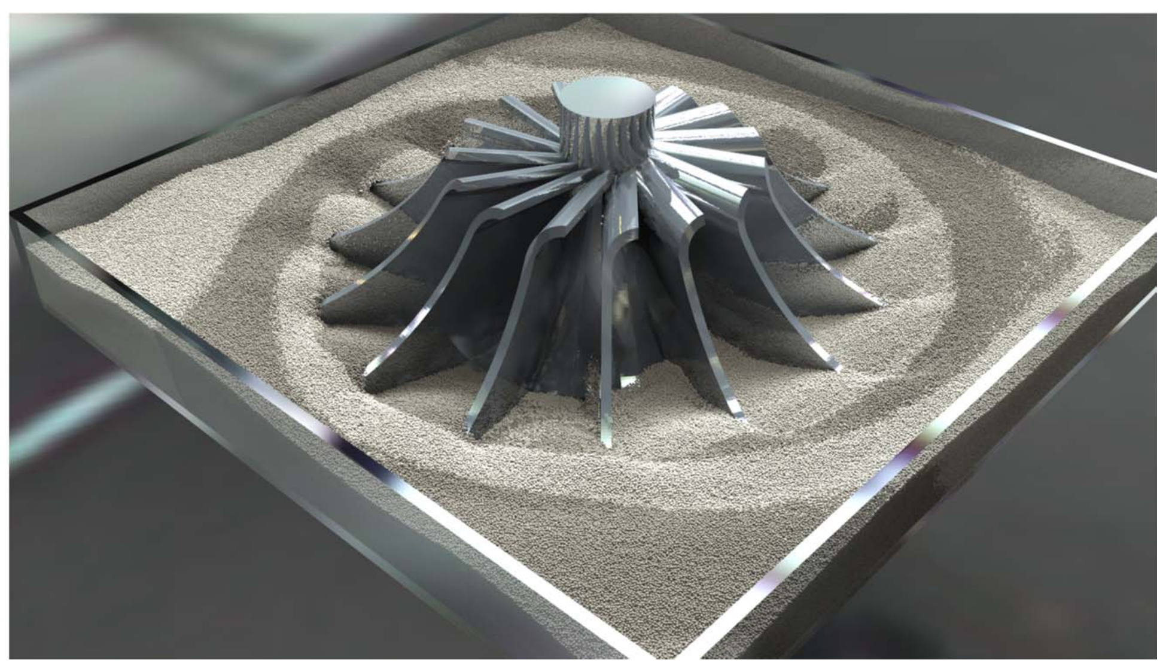
SPH (Smoothed Particle Hydrodynamics)

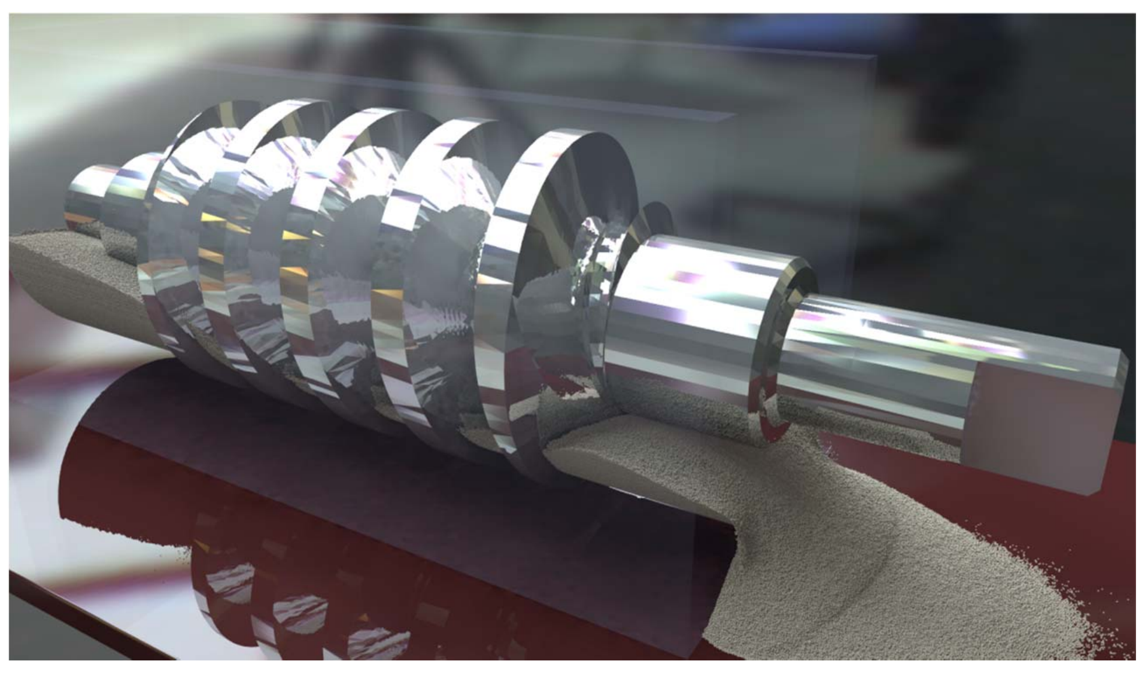
- Development of a SPH simulation code with dynamic load balance based on the slice-grid method. Summation algorithm of forces on a particle from particles within the kernel radius.
- Overhead estimation of a node-to-node communication among GPUs for the transferred data size in SPH including more than 100 particles within a kernel radius (c.f. 10 particles in DEM).
- Study on the advantage and disadvantage of the other dynamic load balance algorithm: Hilbert and Morton curve algorithm and compare the computational performance of the three.



Agitation/Convey analysis on TSUBAME 2.5

✓ Simulations with 4M particles run on 64 GPUs for 200,000 time steps.



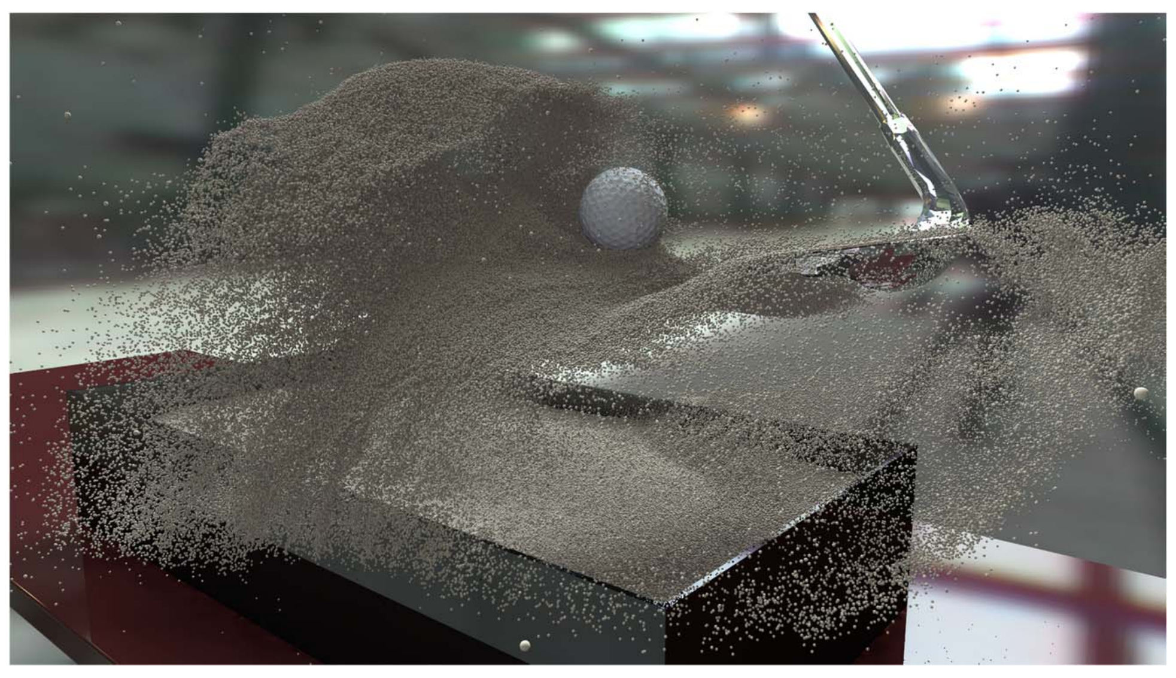


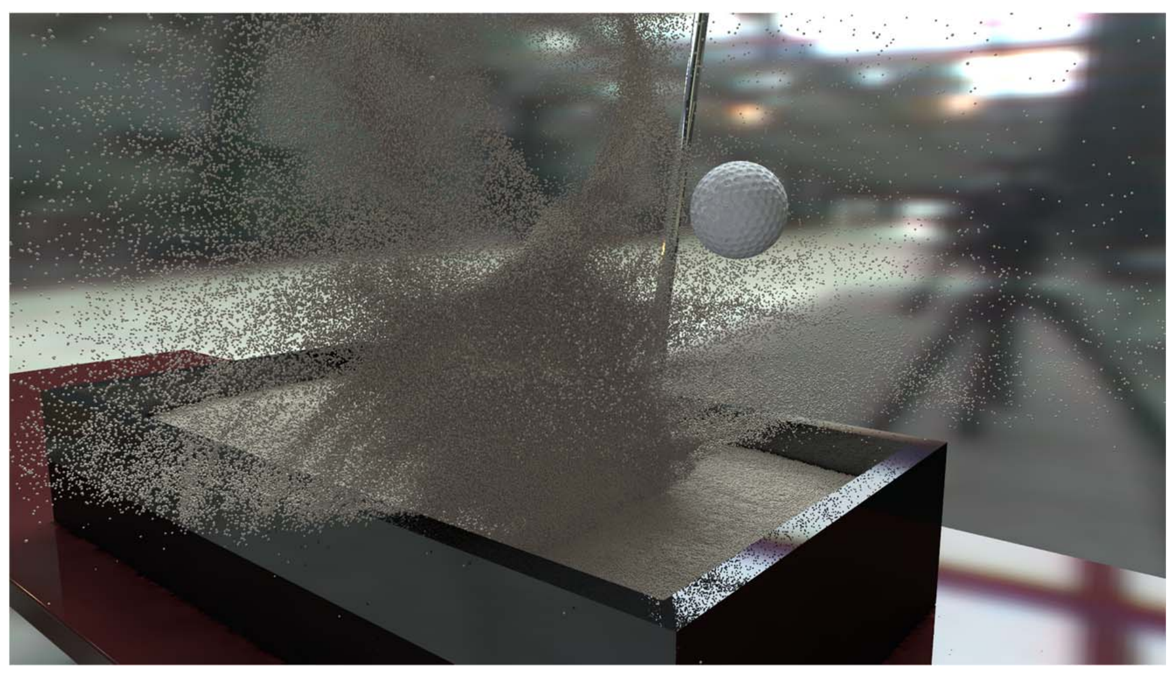
Spiral Slider simulation

✓ 4 million particles with 32 GPUs are used for the 201000 computational steps.

Demonstration of golf bunker shot

✓ It takes 144 hours to run the computation of 130M particles for 47,200 time steps on 256 GPUs.
✓ Simulations using 16.7M particles for various swing trajectories are carried out on 64 GPUs.





JHPCN 学際大規模情報基盤共同利用・共同研究拠点 第6回シンポジウム

Japan High Performance Computing and Networking plus Large-scale Data Analyzing and Information Systems

2014年7月10日, 11日

THE GRAND HALL（品川）