Heterogeneous Simulation of Particulate Flows on GPU Clusters

C. Feichtinger

Chair for System Simulation, University of Erlangen-Nuremberg, Erlangen, Germany

ParCFD 2011, Barcelona, Spain

17. May 2011
Particulate Flows

Fluidization Experiment
Overview: Algorithm

- Time Loop
  - Rigid Body Mapping
  - Fluid Flow Computation
  - Force Calculation
  - Rigid Body Dynamics
  - Communication
Rigid Body Mapping

Implementation Details

- Positions and radii of spheres are transferred from host to device
- 1 CUDA thread block per sphere
- 1 thread per lattice cell
- Atomic operations ensure consistency for overlapping AABB
- In case of overlapping spheres the one with the largest ID is mapped
- New obstacle cells are blocked out
- New fluid cells are reconstructed
LBM Time Step

**Brief Introduction**
- Mesoscopic method for CFD simulations
- Equivalent to a finite difference Navier-Stokes scheme
- Two major steps: Stream step and collision step

\[
f_\alpha(x_i + e_{\alpha,i} \delta t, t + \delta t) - f_\alpha(x_i, t) = -\frac{\delta t}{\tau} \left[ f_\alpha(x_i, t) - f_\alpha^{(eq)}(\rho(x_i, t), u_i(x_i, t)) \right]
\]

\[
\rho u_i = \sum_{\alpha=0}^{18} e_{\alpha,i} \cdot f_\alpha
\]

\[
\rho = \sum_{\alpha=0}^{18} f_\alpha
\]
Side Remark Communication

Time Loop
- Rigid Body Mapping
- Fluid Flow Computation
- Communication
- Force Calculation
- Rigid Body Dynamics

GPU Buffers
- GPU - GPU Copy Operations
- Local Communication
- Swap

CPU Buffers
- GPU - GPU Copy Operations
- PCI Express Transfer

MPI Buffers
- MPI_Irecv
- InfiniBand Transfer

Neighboring Process

C. Feichtinger
ParCFD 2011
Side Remark Heterogeneous Communication

Time Loop
- Rigid Body Mapping
- Fluid Flow Computation
- Force Calculation
- Rigid Body Dynamics

Communication

Process I
- CPU
- OpenMP Parallel
- Socket I

Process II
- OpenMP Parallel
- Socket II

Process III
- GPU

MPI Buffers
Force Calculation

Implementation Details
- Each sphere is treated by one Cuda thread block
- Each thread processes one row in z direction of the AABB
- Temporary force data is stored in shared memory
- Force values have to be reduced and are stored in global memory
- Device-Host transfer of force and torque values

Momentum Exchange

\[ F = \sum_{x_b} \sum_{\alpha=1}^{19} e_\alpha \left[ 2\tilde{f}_\alpha(x_f, t) + 6w_\alpha \rho_w e_\bar{\alpha} \cdot u_w \right] \frac{\Delta x}{\Delta t} \]
Rigid Body Dynamics

Time Loop
- Rigid Body Mapping
- Fluid Flow Computation
- Force Calculation
- Rigid Body Dynamics

Process I
- WaLBerla
- Physics Engine

Process II
- WaLBerla
- Physics Engine
Rigid Body Dynamics

Time Loop
- Rigid Body Mapping
- Fluid Flow Computation
- Communication
- Force Calculation
- Communication

Process I
- WaLBerla
- Physics Engine
- Forces

Process II
- WaLBerla
- Physics Engine
- Forces
Rigid Body Dynamics

Time Loop
- Rigid Body Mapping
- Fluid Flow Computation
- Force Calculation
- Rigid Body Dynamics

Process I
- WaLBerla
- Forces
- Physics Engine

Process II
- WaLBerla
- Forces
- Physics Engine

Sync
Rigid Body Dynamics

The PE Framework

- Massively parallel rigid body C++ software framework
- Largest simulation: 1.1 billion rigid bodies
## Results

### Unit for Performance Measurements
- **MLUPS** - Mega Lattice Updates per Second

### Architecture for Single Node Results

<table>
<thead>
<tr>
<th></th>
<th>GPU</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>TinyGPU @ RRZE</td>
<td>NVIDIA Tesla M2070</td>
<td>Xeon 5650 ”Westmere”</td>
</tr>
<tr>
<td></td>
<td>x 2</td>
<td>2 sockets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12 cores</td>
</tr>
</tbody>
</table>

Theoretical Peak
- Memory Bandwidth [GB/s] 148 64
- Stream Copy 100 (+ECC)/
- Memory Bandwidth [GB/s] 115(-ECC) 43.2
## Results: Single GPU Measurements

### Pure Fluid Flow Measurements

<table>
<thead>
<tr>
<th>$N^3$</th>
<th>SP - ECC [MLUPS]</th>
<th>SP + ECC [MLUPS]</th>
<th>DP - ECC [MLUPS]</th>
<th>DP + ECC [MLUPS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>478</td>
<td>230</td>
<td>257</td>
<td>137</td>
</tr>
<tr>
<td>200</td>
<td>595</td>
<td>285</td>
<td>299</td>
<td>162</td>
</tr>
<tr>
<td>300</td>
<td>625</td>
<td>298</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Fluidization Measurements

<table>
<thead>
<tr>
<th>$N^3$</th>
<th>$\Phi$ [%]</th>
<th>Spheres</th>
<th>SP + ECC [MLUPS]</th>
<th>DP + ECC [MLUPS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2.5</td>
<td>100</td>
<td>176</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>400</td>
<td>116</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>800</td>
<td>72</td>
<td>55</td>
</tr>
<tr>
<td>200</td>
<td>2.5</td>
<td>800</td>
<td>182</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3200</td>
<td>128</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6400</td>
<td>83</td>
<td>61</td>
</tr>
<tr>
<td>300</td>
<td>2.5</td>
<td>2700</td>
<td>225</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10800</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>21600</td>
<td>79</td>
<td></td>
</tr>
</tbody>
</table>

C. Feichtinger  
ParCFD 2011
## Results: Single Node Heterogeneous

### Heterogeneous Hybrid Fluidization Measurements

<table>
<thead>
<tr>
<th>$\Phi$ [%]</th>
<th>$N$</th>
<th>Processes</th>
<th>Blocks</th>
<th>DP + ECC [MLUPS]</th>
<th>Speed Up [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>$600 \times 100 \times 100$</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$450 \times 100 \times 100$</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$450 \times 100 \times 100$</td>
<td>1</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$150 \times 100 \times 100$</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>$600 \times 100 \times 100$</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$450 \times 100 \times 100$</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$450 \times 100 \times 100$</td>
<td>1</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$150 \times 100 \times 100$</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>2</td>
</tr>
</tbody>
</table>
**Future Work**

**Weak Scaling Setup**
- Domain size of each GPU: $825 \times 75 \times 75$
- Domain size of each CPU: $150 \times 75 \times 75$

**GPU Cluster Dirac @ Nersc**
- Each node has 2 x Intel 5530 2.4 GHz quad core Nehalem CPUs
- 44 nodes have an NVIDIA Tesla C2050
Acknowledgment

Thank you for your attention!