RANS CFD Solver on Fermi

Peng Wang, NVIDIA
On Behalf of
James Lin
HPC Center, Shanghai Jiao Tong University
GTC @San Jose, May 2012
Outline

- Related Work
- Background
- Design and Implementation
- Results analysis
- Conclusion and Future work
Related Work

Navier-Stokes Stanford University Solver (NSSUS)

15-20x Speedup On NVIDIA 8800GTX

Stanford
E.Elsen etc
2008.12

Table 2
Speed-ups for the hypersonic vehicle computation

<table>
<thead>
<tr>
<th>Mesh size</th>
<th>Multi-grid cycle</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>720 k</td>
<td>Single grid</td>
<td>15.4</td>
</tr>
<tr>
<td>720 k</td>
<td>2 Grids</td>
<td>11.2</td>
</tr>
<tr>
<td>1.5 M</td>
<td>Single grid</td>
<td>20.2</td>
</tr>
<tr>
<td>1.5 M</td>
<td>2 Grids</td>
<td>15.8</td>
</tr>
</tbody>
</table>
Other Work

1. J. C. Thibault et al. implemented a Navier-Stokes (NS) Equation solver for incompressible flow on Tesla C870 boards, 13X on 1 GPU and 21X on 2 GPUs

2. Tobias Brandvik et al. ported Euler Equations onto GPU and a speedup of 29x in 2D and 16x in 3D

3. A. Corrigan et al. implemented 3D unstructured grid Euler solver for the inviscid compressible flow, and get 33x speedup on 1 C1060

4. I. C. Kampolis et al. use of the 2-D Euler solver for inviscid compressible flow to optimize the aerodynamic performance of the airfoil, get 15 X speedup with single precision and 21X with double on GTX 285
RANS Solver in SJTU

- CFD code by SJTU
- Used in COMAC
- Simu. For Wing design
- Based on RANS Equ.
- FVM on Struc. Grid
Flowchart of Sequential code

Begin

Read Grid File and Control Parameters

Geometry Variable Computation (INIT)

Flow Field Initialization (INITFI)

Local Time-Stepping Computation (STEP)

RK Iteration Initialization

Boundary Condition Management (BCOND)

Artificial Viscosity Flux Computation (FILTER)

Viscous Flux Computation (FLUXV)

Convective Flux Computation (FLUX)

Implicit Residual Smoothing (IMSMOO)

RK Iteration

5th step of RK Iteration

Yes

Error Computation

Total Force Computation And Print (FORCT)

Store Computation Results (OUTPUTFLOW)

Solution Converged or Iteration >= Iterationmax

No

Yes

End
Finding hotspots of RANS Solver
Parallel Approach

• Aggregation
  Compute: $R_k$ inter. + accer. Conv.
  IO ops: print and write file

• Mapping
  Compute: $\rightarrow$ GPU
  IO ops: $\rightarrow$ CPU
Optimization Technique

• for better locality, transformed 1D matrices stored intermediate variables to 3D matrices
• to update data in neighbor vertexes, used multiple kernels to synchronize globally
• to eliminate data transfer between host and device memory, paralleled one direction each time in Implicit Residual Smoothing Algorithm
• for better access to global memory after matrices transformation: coalesce access.
Test Environment

Hardware

**CPU:** Intel(R) Core(TM) i7 CPU **920 @ 2.67GHz**, Quad-Core

**Cache:** 32KB x 4 L1 Cache, 256 x 4 L2 Cache, 8 MB L3 Cache;

**Memory:** **6GB** DDR3, 1033MHz;

**GPU:** **Tesla C2050** with 2.78GB global memory, 65.536KB constant memory, 49.152KB shared memory and 32768 32-bit registers per block, 14 multiprocessor, 448 CUDA cores, 1.15GHz.

Software

**OS:** Windows 7 Enterprise Edition;

**Compiler:** Microsoft Visual Studio 2008 C++ compiler;

**CUDA:** CUDA Toolkit **3.1** and SDK;
ORENA M6 Wing 244800 (51x30x160) Mesh

Computation error

<table>
<thead>
<tr>
<th></th>
<th>CL</th>
<th>CD</th>
<th>CZ</th>
<th>CM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serial</td>
<td>0.268945</td>
<td>0.022390</td>
<td>0.005948</td>
<td>-0.126988</td>
</tr>
<tr>
<td>CUDA</td>
<td>0.269663</td>
<td>0.022446</td>
<td>0.005969</td>
<td>-0.127724</td>
</tr>
<tr>
<td>Relative</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.4%</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

Pressure Distribution span section of 44%

<table>
<thead>
<tr>
<th></th>
<th>Serial</th>
<th>Parallel</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>OutPutFlow</td>
<td>15041.5s</td>
<td>740.0s</td>
<td>20.3</td>
</tr>
<tr>
<td>OutPutFlow*</td>
<td>14903.4s</td>
<td>601.9s</td>
<td>24.8</td>
</tr>
</tbody>
</table>
Another Wing 1676480 (208x65x124) Mesh, Ma=0.785, AOA=2.13, Re=2.0E+07

Computation error

<table>
<thead>
<tr>
<th></th>
<th>CL</th>
<th>CD</th>
<th>CZ</th>
<th>CM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serial</td>
<td>0.519426</td>
<td>0.022840</td>
<td>-0.046544</td>
<td>1.349560</td>
</tr>
<tr>
<td>CUDA</td>
<td>0.518257</td>
<td>0.022661</td>
<td>-0.046310</td>
<td>1.346366</td>
</tr>
<tr>
<td>Relative</td>
<td>0.2%</td>
<td>0.8%</td>
<td>0.5%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

Pressure Distribution span section of 55%

<table>
<thead>
<tr>
<th></th>
<th>Serial</th>
<th>Parallel</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>OutPutFlow</td>
<td>139751.5s</td>
<td>4652.4s</td>
<td>30.0</td>
</tr>
<tr>
<td>OutPutFlow*</td>
<td>138926.7s</td>
<td>3769.8s</td>
<td>36.9</td>
</tr>
</tbody>
</table>
Conclusion and Future Work

• RANS solver on GPU get 25X speedup for M6 Wing and 37X for another Wing without sacrifice in accuracy

• Next Step will scale up on GPU cluster
Acknowledgement

This research is financially supported by Academic Partnership Program from NVIDIA
Contact Information

James Lin （林新华） lin-xh@sjtu.edu.cn

SJTU CUDA Center of Excellence
High Performance Computing Center, SJTU

Bring Many-core Thinking to High End Research and Education