MDGRAPE-3 chip: A 165-Gflops application-specific LSI for Molecular Dynamics Simulations

Makoto Taiji
High-Performance Biocomputing Research Team
Genomic Sciences Center, RIKEN

Molecular Dynamics Simulations

-force calculation dominates computational time

Require large computational power
Scientific Backgrounds

- Protein 3000 Project: National project to determine 3,000 protein structures
- Extensive requirements for molecular simulations
  - Drug Design
  - Bio-nanotechnology
- High-performance dedicated computer can solve computational difficulties

What is GRAPE?

- GRAvity PipE
- Special-purpose computers for classical particle simulations
  - Astrophysical $N$-body simulations
  - Molecular Dynamics Simulations
- Accelerate only force calculations
- Univ. Tokyo / RIKEN
- MDGRAPE-3: Petaflops GRAPE for Molecular Dynamics simulations

History of GRAPE computers

Why we build special-purpose computers?

Bottleneck of high-performance computing:

- Parallelization limit / Memory bandwidth
- Power Consumption = Heat Dissipation

These problems will become more serious in future. Special-purpose approach:

- can solve parallelization limit thoroughly
- relax power consumption
- ~100 times better cost-performance
Number of floating-point operations / cycle of microprocessors

- Parallelization within LSI is quite important
- Number of operations / cycle is quite limited in general-purpose computer
- Mainly due to memory bandwidth


Broadcast Parallelization

- Molecular Dynamics Case
- Two-body forces
  \[ F_i = \sum_j f(x_i; x_j) \]
- For parallel calculation of \( F_i \), \( x_j \)
  we can use the same \( x_j \)
- Broadcast Parallelization - relax Bandwidth Problem
Highly-Parallel Operations in LSIs for MD simulations

- For special-purpose computers
  - Broadcast Memory Architecture
  - Efficient: 660 equivalent operations/cycle/chip in MDGRAPE-3 chip
  - Possible to increase according to Moore’s law
- In case of Molecular Dynamics:

<table>
<thead>
<tr>
<th></th>
<th>MDGRAPE</th>
<th>600nm</th>
<th>1 pipeline</th>
<th>1Gflops</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDGRAPE-2</td>
<td>250nm</td>
<td>4 pipelines</td>
<td>16Gflops</td>
<td></td>
</tr>
<tr>
<td>MDGRAPE-3</td>
<td>130nm</td>
<td>20 pipelines</td>
<td>165Gflops</td>
<td></td>
</tr>
</tbody>
</table>

Power Efficiency of Special-Purpose Computers

- General-Purpose Processors
  - Pentium 4 (130nm, 3GHz, FSB800) … 82W
  - 14W/Gflops
- Molecular Dynamics Processors
  - MDGRAPE-2 (250nm, +2.5V, 100MHz) … 1W/Gflops
  - MDGRAPE-3 (130nm, +1.2V, 250MHz) … 0.1W/Gflops
  - Highly-parallel operation at modest frequency
  - Control precision makes power-performance better.
Molecular Dynamics Machine (MDM)

- MDGRAPE-2 chip:
  16 Gflops at 100 MHz
  IBM SA-12E 250nm
- 78 Tflops Performance
- Fastest Computer since 2000
- Small system (MDGRAPE-2) is commercially available


MDGRAPE-3 (aka Protein Explorer)

- Petaflops special-purpose computer for molecular dynamics simulations
- Whole system: FY2006
- MDGRAPE-3 chip
  - Force calculation chip
  - 130 nm technology
    Hitachi HDL4N
  - 165 Gflops/chip at 250 MHz
    Sample: 230Gflops@350MHz

M. Taiji et al, Proc. Supercomputing 2003, on CDROM.
Block Diagram of MDGRAPE-3 chip

- Memory-in-a-chip Architecture
- Memory for 32,768 particles
- The same data is broadcasted to each pipeline

MDGRAPE-3 chip

- Memory Controller
- Particle Type Memories
- Main Memory
- 15.7 mm X 15.7 mm
**Force Pipeline**

- Calculate two-body central forces
  \[ r_{ij} = r_i - r_j \]
  \[ r_{ij}^2 = r_i^2 + r_j^2 \]
  \[ F_i = \sum_{j} r_{ij} \cdot g(\mathbb{R}_{ij}) \]

- 8 multipliers, 9 adders, and 1 function evaluator
- 33 equivalent operations for Coulomb force calculation
- Function Evaluator: approximate arbitrary functions by segmented fourth-order polynomials
- Multipliers: floating-point, single precision
- Adders: floating-point, single precision / fixed-point 40 or 80 bit

---

**Chip Details**

- Hitachi HDL4N 130nm
  Vcore = +1.2V, 7-layer Cu wiring, pitch=360nm
- I/O
  GTL and/or +1.2V CMOS 203 signals (not including test)
- Clock Frequency
  Core: 250 MHz, I/O: 125 MHz in worst case commercial
- Die Size
  15.7mm X 15.7 mm
- Total Gates
  6.1M (2NAND) ~20% for test and clock tree
- Total Memories
  9M bits
- Usage
  53%
- Package
  1444 pin FCBGA
- Power Consumption
  19 W at 350 MHz
Performance

- In total
  - 160 floating-point multipliers
  - 60 floating-point adders
  - 60 40-bit integer adders + floating point converters
  - 60 floating-point to 80-bit integer converters + integer adders
  - 20 function evaluators
    - Table + fourth order polynomial calculation
- All units work simultaneously
- ~660 operations/cycle for Coulomb force
- 165 Gflops at 250 MHz
- Sample LSI worked at 350MHz, 230 Gflops

Design Method

- Synopsis Design Compiler Ultra +
  - In-house multiplier/multiplier-adder generator to use special cells
- VHDL 11,000 lines
  - Except for multipliers, test bench, and comment/blank lines
- Simple: all vhdl, synthesis, simulations has been performed by the presenter alone
  - ~8 man-month
- Test Circuits, Clock Tree, Layout: Hitachi
  - ~18 man-month
- Development period: ~1 year
- 2 scientists work for system design and software
MDGRAPE-3 System

- 12 Chips/Board
- 2 boards/2U subrack
  = 4 Tflops
- Chips are daisy-chained by 81-bit GTL bus (1.27 Gbytes/sec)
- Connected to PCI-X bus via LVDS 10Gbit/s interface on Xilinx Virtex-II Pro
- Small PCI-X system is also available

Petaflops MDGRAPE-3 system

- Host Computer: 512-CPU PC cluster or HPC server like SGI Altix 350
- 6,144 MDGRAPE-3 chips to reach a petaflops
- 32 19” pedestals
  … very small
Price & Power Performance

<table>
<thead>
<tr>
<th></th>
<th>$ / Gflops</th>
<th>W / Gflops</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDGRAPE-3</td>
<td>15</td>
<td>0.2</td>
</tr>
<tr>
<td>BlueGene/L</td>
<td>140</td>
<td>6</td>
</tr>
<tr>
<td>Pentium 4 PC</td>
<td>400</td>
<td>14</td>
</tr>
<tr>
<td>Earth Simulator</td>
<td>8000</td>
<td>128</td>
</tr>
<tr>
<td>MDGRAPE-2</td>
<td>150</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Total development cost … about 15 M$ including our salaries

Applications suitable for broadcast memory architecture

- Computation-intensive (not data-intensive)
- Multiple calculations using the same data
  - Molecular dynamics simulations
  - Astrophysical $N$-body simulations
  - Dynamic programming for genome sequence analysis
  - Boundary value problems
  - Calculation of dense matrices
Quasi-general-purpose machines with broadcast memory architecture

- (F)PU array
  GRAPE-DR Project (2004-2008)
  Prof. Jun Makino, Univ. Tokyo
  1 Tflops/chip

- SIMD vector processor with broadcast memory architecture
  MACE (MAtrix Computing Engine)
  for dense matrix calculation
  3.5Gflops/chip, double precision, 180nm

- Such approach will be effective in future
  when our approach will become more advantageous

Acknowledgements

- Coworkers
  - Hardware developments
    Tetsu Narumi, Ph. D.
    Yousuke Ohno, Ph. D.
  - Applications
    Atsushi Suenaga, Ph.D.
    Noriyuki Futatsugi, Ph.D.
    Noriaki Okimoto, Ph.D.
    Naoki Takada, Ph.D.
  - Bioinformatics Group Director
    Akihiko Konagaya, Dr.Eng.

- GRAPE collaboration
  Univ. Tokyo
  Prof. Junichiro Makino
  Dr. Toshiyuki Fukushige
  RIKEN
  Dr. Toshikazu Ebisuzaki
  Dr. Takahiro Koishi
  Dr. Ryutaro Susukita
  Saitama Inst. Tech.
  Dr. Atsushi Kawai
  Univ. Air
  Prof. Daiichiro Sugimoto

This work is partially supported by `Protein 3000 project',
Ministry of Education, Culture, Sports, Science and Technology