MESGA: An MPSoC Based Embedded System Solution for Short Read Genome Alignment

Vikkitharan Gnanasambandapillai (Vikki.g@unsw.edu.au)
Arash Bayat
Sri Parameswaran
University of New South Wales, Australia
Agenda

- DNA tests
- Sequencer and current computational platforms
- Embedded system for genomic computation
- MESGA
- Future works & Conclusions
DNA Tests

- Four types bases A, C, G & T
- 3.2 billion base pair
- 99.9% similar
- DNA tests identify genetic markers
Read Genome Sequencing

1. Sample preparation
2. Sequencing
3. Computations

- “Short reads” or “long reads”
Short Read Genome Alignment

- Reference & target
- 30x coverage
- Huge raw data (50 TB)
<table>
<thead>
<tr>
<th>Enormous improvement</th>
<th>Moderate improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Big</td>
</tr>
<tr>
<td>Portable</td>
<td>Remote cloud computers</td>
</tr>
<tr>
<td>Cheap</td>
<td>?</td>
</tr>
<tr>
<td>Fast</td>
<td>?</td>
</tr>
</tbody>
</table>
Remote Sequencing

- Real-time, portable genome sequencing for Ebola surveillance
  - Identify infections agent & monitor spreading rate
  - Sequenced in west Africa and processed in UK
  - High speed satellite internet would cost 2000$

- DNA Sequencing in Space Monitor
  - Astronauts’ health & identify life beyond earth
  - Sequence in ISS and process on earth
  - Would take 15 months to diagnose
# Sequencers vs Computers vs Embedded Processor

<table>
<thead>
<tr>
<th>Exponential improvement</th>
<th>Linear improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Big</td>
</tr>
<tr>
<td>Portable</td>
<td>Remote</td>
</tr>
<tr>
<td>Cheap</td>
<td>Expensive</td>
</tr>
<tr>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Secure</td>
<td>Unsecure</td>
</tr>
<tr>
<td>Slower CPU</td>
<td>Faster CPU</td>
</tr>
<tr>
<td>Smaller memory</td>
<td>Bigger Memory</td>
</tr>
</tbody>
</table>
MESGA

- Reference partition
- MPSoC pipelined architectures
- Processor
- Experiment setup & Results
Reference Partition

- Whole ref in shared memory
  - Memory bottleneck
  - Limited memory size

- Partial ref in local memory
Reference Partition Cont.

- Faster alignment
- Higher failure rate
Linear Pipeline

- Same processor
  - Except ref
- An external interface
- FIFO interfaces
Linear Pipeline

P0
BWA aln
MEM0

P1
BWA aln
MEM1

P2
BWA aln
MEM2

P3
BWA aln
MEM3

P7
BWA aln
MEM7

P6
BWA aln
MEM6

P5
BWA aln
MEM5

P4
BWA aln
MEM4
Linear Pipeline

- Same processor
  - Except ref
- An external interface
- FIFO interfaces
- Pipeline freeze
- Unbalanced load
Linear Pipeline with Optimized Partition

- Different ref size
- Different timeout
- Better balanced load
Ring Pipeline

- Same processor
- Except ref
- Ring FIFO interfaces
- External interfaces
Ring Pipeline

- An animation is removed to reduce the file size to 10MB.
- Please contact the presenter for the original slide
Ring Pipelined

- Same processor
  - Except ref
- Ring FIFO interfaces
- External interfaces
- Balanced load
- Less pipeline freezing
MESGA Processor

- 32-bit Xtensa LX6.0 processors
  - Clock speed 1092 MHz
  - Size 0.08 mm²
- 8kB data cache
- Up to 2GB local memory
Experiments Setup

- Full human genome reference
- One million reads each 125 base pair length
  - 0.09 % SNP mutation error
  - 0.01 % indel mutation error
  - 2 % uniform sequencing error
Results

- Single processor
- Linear pipeline
- Linear pipeline (optimized)
- Ring pipeline (without timeout)
- Ring pipeline (with timeout)
# MESGA vs Other Solutions

<table>
<thead>
<tr>
<th></th>
<th>GPU</th>
<th>FPGA</th>
<th>MESGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Hardware accelerator</td>
<td>Hardware accelerator</td>
<td>Stand alone &amp; Complete</td>
</tr>
<tr>
<td>Feature support</td>
<td>Limited</td>
<td>Limited</td>
<td>Full</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
Future works & Conclusions

- Customised Instructions
- Inter-processor communication optimization
- Long read alignment

- Whole human genome can be processed in embedded system
  - Cheap
  - Small
  - Secure

- Performance improvement
Thank You