MicroHeterogeneous Computing
for Medical Processing Applications

Sam Skalicky
Ganesh Khedkar
+ Curious George

let's get curious!
Outline

- Overview Heterogeneous Computing
- Overview Medical Processing Application
- Different hardware architectures/performance
- Simulator
- Conclusions
Heterogeneous Computing

- Combination of multiple different types of processors in a single system
  - CPU, GPU, FPGA
  - Good mix (covering highly seq., massively parallel, & in-between)
- Unlike Clusters which tend to be homogenous
- Better performance by tailoring hardware to different parts of the program (phases)
Medical Processing Applications

- Image Processing
  - Registration, Segmentation, Analysis, Visualization
- Data Manipulation
  - ECG, Ultrasound, other signal processing
- Combine all into a single system
  - Heart ECG Visualization
Heart ECG Visualization

- Model all Electrical and muscle movements
- Monitor patient for irregular activities
- Reduce need for radiation exposure due to harmful imaging modalities
ECG Sampling
NTEPI Algorithm

Noninvasive

Transmural

Electro-

Physiological

Imaging

---

Generate sample vectors to form $U_{nX(2n+1)}$

Monte-Carlo sim of Aliev-Panfilov models on set of samples $U_{nX(2n+1)}$

Calculate mean $u_{nX1}$ & covariance matrix $P_{nXn}$ from $U_{nX(2n+1)}$

Update mean $u_{nX1}$ & covariance $P_{nXn}$ by Kalman updated

Maintain the extra variable $v_{nX1}$ and its sample set $V_{nX(2n+1)}$

$n \sim 2000-3000$ iterations
Computations

- Kalman Filter

\[
\begin{align*}
\hat{x}_{k|k-1} &= F_k \hat{x}_{k-1|k-1} + B_k u_k \\
P_{k|k-1} &= F_k P_{k-1|k-1} F_k^T + Q_k \\
S_k &= H_k P_{k|k-1} H_k^T + R_k \\
\hat{y}_k &= z_k - H_k \hat{x}_{k|k-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k \hat{y}_k \\
K_k &= P_{k|k-1} H_k^T S_k^{-1} \\
P_{k|k} &= (I - K_k H_k) P_{k|k-1}
\end{align*}
\]

- Matrix Computations
  - Add, Mult, Inv, Decomp.
  - Very Large sizes from 300^2 (90k) to 1000^2 (1M)
High Performance Solution

**CPU**
- High Speed 2-5 GHz
- Single Ops (Variable level)

**GPU**
- Med Speed 5-800MHz
- Massively Parallel Ops (Matrix Add…)

**FPGA**
- Low Speed 2-500MHz
- Custom H/W pipelines (image proc.)
High Performance Solution

**GPU**
- Massively Parallel Ops

**FPGA**
- Opportunities for custom pipeline/accel.

**CPU**
- Remaining Ops are intermediary or excessively sequential
Heterogeneous System

Workstation System

Control CPU

System Scheduler

Computations

CPU

GPU

FPGA

Heterogeneous Computing for Medical Processing Applications
Heterogeneous System Simulator

Heterogeneous Computing for Medical Processing Applications
# NTEPI Algorithm Computations

<table>
<thead>
<tr>
<th>Computation</th>
<th>Quantity</th>
<th>CPU</th>
<th>GPU</th>
<th>FPGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Add</td>
<td>77220</td>
<td>3087us</td>
<td>166us</td>
<td>11us</td>
</tr>
<tr>
<td>M Sub</td>
<td>65520</td>
<td>3157us</td>
<td>167us</td>
<td>11us</td>
</tr>
<tr>
<td>M Mult</td>
<td>70200</td>
<td>42362us</td>
<td>4141us</td>
<td>12000us</td>
</tr>
<tr>
<td>M Scale</td>
<td>79560</td>
<td>1984us</td>
<td>107us</td>
<td>15us</td>
</tr>
<tr>
<td>M Norm</td>
<td>14040</td>
<td>210us</td>
<td>143us</td>
<td>11us</td>
</tr>
<tr>
<td>M Transpose</td>
<td>11700</td>
<td>4019us</td>
<td>296us</td>
<td>11us</td>
</tr>
<tr>
<td>M Inverse</td>
<td>2340</td>
<td>79152us</td>
<td>30159us</td>
<td>8000us</td>
</tr>
<tr>
<td>M Col Sum</td>
<td>4680</td>
<td>210us</td>
<td>527us</td>
<td>11us</td>
</tr>
<tr>
<td>M Cholesky</td>
<td>2340</td>
<td>10367us</td>
<td>16769us</td>
<td>3000us</td>
</tr>
<tr>
<td>FP Add</td>
<td>2340</td>
<td>0.12ns</td>
<td>45us</td>
<td>2.22ns</td>
</tr>
<tr>
<td>FP Sub</td>
<td>4680</td>
<td>0.12ns</td>
<td>45us</td>
<td>2.22ns</td>
</tr>
<tr>
<td>FP Div</td>
<td>23400</td>
<td>0.20ns</td>
<td>260us</td>
<td>100ns</td>
</tr>
</tbody>
</table>
## NTEPI Algorithm Computations

<table>
<thead>
<tr>
<th>Computation</th>
<th>Quantity</th>
<th>CPU</th>
<th>GPU</th>
<th>FPGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Add</td>
<td>77220</td>
<td>3087us</td>
<td>166us</td>
<td>11us</td>
</tr>
<tr>
<td>M Sub</td>
<td>65520</td>
<td>3157us</td>
<td>167us</td>
<td>11us</td>
</tr>
<tr>
<td>M Mult</td>
<td>70200</td>
<td>42362us</td>
<td>4141us</td>
<td>12000us</td>
</tr>
<tr>
<td>M Scale</td>
<td>79560</td>
<td>1984us</td>
<td>107us</td>
<td>15us</td>
</tr>
<tr>
<td>M Norm</td>
<td>14040</td>
<td>210us</td>
<td>143us</td>
<td>11us</td>
</tr>
<tr>
<td>M Transpose</td>
<td>11700</td>
<td>4019us</td>
<td>296us</td>
<td>11us</td>
</tr>
<tr>
<td>M Inverse</td>
<td>2340</td>
<td>79152us</td>
<td>30159us</td>
<td>8000us</td>
</tr>
<tr>
<td>M Col Sum</td>
<td>4680</td>
<td>210us</td>
<td>527us</td>
<td>11us</td>
</tr>
<tr>
<td>M Cholesky</td>
<td>2340</td>
<td>10367us</td>
<td>16769us</td>
<td>3000us</td>
</tr>
<tr>
<td>FP Add</td>
<td>2340</td>
<td>0.12ns</td>
<td>45us</td>
<td>2.22ns</td>
</tr>
<tr>
<td>FP Sub</td>
<td>4680</td>
<td>0.12ns</td>
<td>45us</td>
<td>2.22ns</td>
</tr>
<tr>
<td>FP Div</td>
<td>23400</td>
<td>0.20ns</td>
<td>260us</td>
<td>100ns</td>
</tr>
</tbody>
</table>
## Initial System Simulations

<table>
<thead>
<tr>
<th>Scheduling Algorithm</th>
<th>Total Proc.</th>
<th>Processor Count</th>
<th>% Optimal Assign</th>
<th>Exec. Time mm:ss.s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best/Avail.</td>
<td>3</td>
<td>1 1 1</td>
<td>34%</td>
<td>01:23.64</td>
</tr>
<tr>
<td>Best/Only</td>
<td>1</td>
<td>1 1 1</td>
<td>100%</td>
<td>01:35.94</td>
</tr>
<tr>
<td>Best/Support</td>
<td>1</td>
<td>1 1 1</td>
<td>86%</td>
<td>00:59.38</td>
</tr>
<tr>
<td>Best/Support</td>
<td>4</td>
<td>2 1 1</td>
<td>80%</td>
<td>00:46.47</td>
</tr>
<tr>
<td>Best/Support</td>
<td>4</td>
<td>1 2 1</td>
<td>89%</td>
<td>00:37.52</td>
</tr>
<tr>
<td>Best/Support</td>
<td>4</td>
<td>1 1 2</td>
<td>87%</td>
<td>00:52.41</td>
</tr>
<tr>
<td>Best/Support</td>
<td>5</td>
<td>2 2 1</td>
<td>83%</td>
<td>00:31.81</td>
</tr>
<tr>
<td>Best/Support</td>
<td>5</td>
<td>1 3 1</td>
<td>91%</td>
<td>00:27.39</td>
</tr>
<tr>
<td>Best/Support</td>
<td>8</td>
<td>2 5 1</td>
<td>87%</td>
<td>00:16.35</td>
</tr>
</tbody>
</table>

Heterogeneous Computing for Medical Processing Applications
Conclusion

- Heterogeneous Systems offer faster performance

- System can contain any combination of many processors

- Simulator provides easy design space exploration

- Estimated execution for any program

- Configure the system for optimal program performance
Show Me The Monkey!

Questions or Comments