System Level Power-Performance Trade-Offs in Embedded Systems Using Voltage and Frequency Scaling of Off-Chip Buses and Memory

Kiran Puttaswamy\textsuperscript{1}, Kyu-Won Choi\textsuperscript{1}, Jun Cheol Park\textsuperscript{1}, Vincent J. Mooney III\textsuperscript{1,2}, Abhijit Chatterjee\textsuperscript{1,3} and Peeter Ellervee\textsuperscript{4}

\{kiranp, kwchoi, jcpark, chat, mooney\}@ece.gatech.edu

lrv@cc.ttu.ee

\textsuperscript{1}Center for Research on Embedded Systems and Technology (CREST), http://crest.ece.gatech.edu

\textsuperscript{2}Assistant Professor, \textsuperscript{1}Electrical and Computer Engineering

\textsuperscript{3}Professor, \textsuperscript{1}Electrical and Computer Engineering

\textsuperscript{2}Adjunct Assistant Professor, College of Computing

\textsuperscript{4}Tallin Technical University, Tallin, Estonia

\textsuperscript{2}Hardware/Software Codesign Group, http://codesign.ece.gatech.edu
Overview

- Introduction
- Motivation
- Contribution
- Framework
- Methodology
- Results
Introduction

• Embedded Systems – essential components of living
• Constraining Factor: Power
Motivation

• Limited Battery Capacity

• Battery Energy Supplying Characteristic
  10 mA, 1.5 volts = 1000 hours
  100 mA, 1.5 volts = 80 hours
Previous Work

• Three broad approaches to memory optimization for power/energy reduction
  – Cache optimizations
  – Memory access reduction (especially of off-chip memory)
  – Memory sizing/structuring and memory intensive voltage scaling
Our Contribution

• Combination of an architectural technique (store buffer) and a circuit level technique (voltage and frequency scaling) to realize savings in both power and energy in an embedded system composed of an ARM-like processor chip plus a separate memory chip
• System savings in power from 28% to 36%
• System savings in energy from 13% to 35%
Computation Part of an Embedded System

CPU

Data cache

Instruction cache

Off-Chip Memory

64

96

32

32

October 2002

ISSS
Power Models

- Verilog RTL model for processor (excluding caches)
- Compaq Personal Server PCB Board called “Skiff”
- Analytical memory model for caches and off-chip memory
Framework

Benchmark Programs (c)

VHX Translation

MARS Simulator

Toggle Rate Generation

Processor Core Power Model

Off-Chip Memory Power Model

Off-chip Bus Power Model

System Level Power

October 2002

ISSS
Wither the power?

• Computation in system
  – MARS processor (U. Michigan, www.eecs.umich.edu/~jringenb/power)
    • ~30K lines Verilog
      – synthesized using TSMC .25u std. cell lib. from LEDA Systems
    • 4KB Icache, 4KB Dcache
  – 0.5MB SRAM memory chip (L2)

• Approximately 50% of the power consumed by processor chip (excluding I/O pads and drivers)

• 50% of the power consumed to drive L2 memory: the 0.5MB memory chip + PCB bus + I/O pads/drivers

• => reduce power to drive L2 memory by 60%, system power reduced 30%
3.3 V -> 2V, SRAM delay doubles, power reduces up to 66%

Use TSMC 0.25 μ tech. param. from MOSIS
Embedded System

CPU

Data cache

Instruction cache

Off-Chip Memory

64
96
32
32
Embedded System (with Store Buffer)
Methodology

- Voltage/frequency scaling of L2 memory accesses
- Store buffer technique
Voltage/Frequency Scaling

Processor

Off-chip Buses

2 Volts, 50 Mhz

2.75 Volts, 100 Mhz

Off-chip Memory

2 Volts, 50 Mhz
<table>
<thead>
<tr>
<th>benchmark</th>
<th>Executable size (kB)</th>
<th>Dynamic instruction count</th>
<th>Input data size</th>
<th>Data cache accesses</th>
<th>Data cache misses</th>
<th>Data cache miss %</th>
</tr>
</thead>
<tbody>
<tr>
<td>bubble</td>
<td>34.852</td>
<td>7503</td>
<td>50 integers array</td>
<td>1675</td>
<td>107</td>
<td>6.39</td>
</tr>
<tr>
<td>factorial</td>
<td>34.634</td>
<td>6033</td>
<td>1 integer</td>
<td>2006</td>
<td>250</td>
<td>12.46</td>
</tr>
<tr>
<td>fib</td>
<td>34.651</td>
<td>30602</td>
<td>1 integer</td>
<td>11840</td>
<td>262</td>
<td>2.21</td>
</tr>
<tr>
<td>matmul</td>
<td>34.857</td>
<td>21642</td>
<td>0.5 kB</td>
<td>7358</td>
<td>4916</td>
<td>66.81</td>
</tr>
<tr>
<td>sort_int</td>
<td>34.763</td>
<td>23171</td>
<td>0.5 kB</td>
<td>7808</td>
<td>104</td>
<td>1.33</td>
</tr>
</tbody>
</table>

**Table 2: Execution Statistics for Various Benchmarks**

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Off-chip Bus, Memory at 100 MHz, 3.3 V</th>
<th>Off-chip Bus, Memory at 50 MHz, 2 V</th>
<th>% Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cpu+cache (W)</td>
<td>bus (mW)</td>
<td>L2 memory(mW)</td>
</tr>
<tr>
<td>bubble</td>
<td>1.24</td>
<td>301.64</td>
<td>1276.49</td>
</tr>
<tr>
<td>factorial</td>
<td>1.18</td>
<td>444.35</td>
<td>1236.96</td>
</tr>
<tr>
<td>fib</td>
<td>1.25</td>
<td>287.68</td>
<td>1228.23</td>
</tr>
<tr>
<td>matmul</td>
<td>1.07</td>
<td>637.48</td>
<td>1713.34</td>
</tr>
<tr>
<td>sort_int</td>
<td>1.27</td>
<td>336.78</td>
<td>1485.92</td>
</tr>
</tbody>
</table>

**Table 3: System Level Power Estimates**
<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Execn Time (μs)</th>
<th>Power (W)</th>
<th>Energy (mJ)</th>
<th>Execn Time (μs)</th>
<th>Power (W)</th>
<th>Energy (mJ)</th>
<th>Execn Time increase (%)</th>
<th>Energy decrease (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bubble</td>
<td>113.945</td>
<td>2.817</td>
<td>0.321</td>
<td>122.265</td>
<td>1.857</td>
<td>0.227</td>
<td>7.3</td>
<td>29.3</td>
</tr>
<tr>
<td>factorial</td>
<td>116.115</td>
<td>2.861</td>
<td>0.332</td>
<td>129.325</td>
<td>2.040</td>
<td>0.264</td>
<td>11.37</td>
<td>20.48</td>
</tr>
<tr>
<td>fib</td>
<td>456.795</td>
<td>2.766</td>
<td>1.263</td>
<td>463.245</td>
<td>1.859</td>
<td>0.861</td>
<td>1.4</td>
<td>31.83</td>
</tr>
<tr>
<td>matmul</td>
<td>924.735</td>
<td>3.421</td>
<td>3.164</td>
<td>1192.98</td>
<td>2.313</td>
<td>2.759</td>
<td>29.0</td>
<td>12.8</td>
</tr>
<tr>
<td>sort_int</td>
<td>296.425</td>
<td>3.093</td>
<td>0.917</td>
<td>300.265</td>
<td>1.986</td>
<td>0.596</td>
<td>1.29</td>
<td>35.0</td>
</tr>
</tbody>
</table>

Table 4: System Level Design Space Exploration
Conclusion

• Reduction in both power and energy
  – For an ARM-like processor chip plus a separate memory chip:
  – System savings in power from 28% to 36%
  – System savings in energy from 13% to 35%
  – Increase in execution time from 1% to 29%

• Possible technique for power modulation by user/application