The Potential of a Temperature-Aware Configurable Cache on Energy Saving

Hamid Noori, Maziar Goudarzi, Koji Inoue, and Kazuaki Murakami

Kyushu University
Outline

- Motivations and Observations
- Energy Evaluation
- Problem Definition
- Experimental Results
- Conclusion
Outline

- Motivations and Observations
Motivations and Observations (1/4)

- Energy efficiency is very important in embedded systems!
- Leakage power is increasing in new nanometer-scale technologies
Motivations and Observations (2/4)

- 4-way set-associative cache with 16-byte block size
- Dynamic: 180nm ~ 4x 100nm & 9x 70nm (CACTI 4.1)
- Static: 70nm ~ 400x 180nm & 5x 100nm (CACTI 4.1)
Motivations and Observations (4/4)

![Leakage Power for Cache 32K (mW) vs Temperature]

- 180nm
- 100nm
- 70nm

Leakage Power for Cache 32K (mW)

Temperature: 0°C, 20°C, 40°C, 60°C, 80°C, 100°C
Goal

- The effect of temperature on cache configuration selection in low-energy embedded systems
Outline

- Energy Evaluation
Energy Evaluation (1/3)

- Static
- Dynamic

\[
\text{energy}_{\text{memory}}(\text{Config, Temp, Tech}) = \\
\text{energy}_{\text{dynamic}}(\text{Config, Tech}) + \\
\text{energy}_{\text{static}}(\text{Config, Temp, Tech})
\]
Energy Evaluation (2/3)

\[
\begin{align*}
\text{energy\_dynamic}(\text{Config}, \text{Tech}) &= \text{cache\_accesses}(\text{Config}) \times \text{energy\_cache\_access}(\text{Config}, \text{Tech}) + \\
&\quad \text{cache\_misses}(\text{Config}) \times \text{energy\_miss}(\text{Config}, \text{Tech})
\end{align*}
\]

\[
\begin{align*}
\text{energy\_miss}(\text{Config}, \text{Tech}) &= \text{energy\_off\_chip\_access} + \\
&\quad \text{energy\_cache\_block\_refill}(\text{Config}, \text{Tech})
\end{align*}
\]

\[
\begin{align*}
\text{energy\_static}(\text{Config}, \text{Temp}, \text{Tech}) &= \text{executed\_clock\_cycles}(\text{Config}) \times \text{clock\_period} \times \\
&\quad \text{leakage\_power}(\text{Config}, \text{Temp}, \text{Tech})
\end{align*}
\]
Energy Evaluation (3/3)

- **Simplescalar**
  - \textit{cache\_accesses}
  - \textit{cache\_misses}
  - \textit{executed\_clock\_cycles}

- **CACTI 4.1**
  - \textit{energy\_cache\_access}
  - \textit{energy\_cache\_block\_refill}
  - \textit{leakage\_power}
  - \textit{energy\_off\_chip\_access} = 20 nJ
Outline

- Problem Definition
Problem Definition

“For a given application, processor architecture, and technology, find the cache configuration that results in minimum energy consumption in different temperatures (i.e. minimizes Energy Equation for a given temperature) over the entire application run.”
Outline

- Experimental Results
Experimental Results (1/2)

- Applications from Mibench
- SimpleScalar
- CACTI 4.1
  - Three technologies: 180nm, 100nm, and 70nm
  - Six Temperatures: 0° C, 20° C, 40° C, 60° C, 80° C, 100° C
- Configurable Cache
  - Size: 64KB~1KB
  - Ways: 4, 2, and 1
  - Line size: 32-, 16-, and 8-byte
Experimental Results (2/2)

- For each application
  - 7 (size) x 3 (ways) x 3 (line size) x 2 (I$, D$) = 126 simulations per application

- Dynamic energy
  - 7 (size) x 3 (ways) x 3 (line size) x 3 (Tech) = 189

- Leakage power
  - 7 (size) x 3 (ways) x 3 (line size) x 3 (Tech) x 6 (Temp) = 1134
<table>
<thead>
<tr>
<th>Application</th>
<th>basimath</th>
<th>bitcounts</th>
<th>qsort</th>
<th>djpeg</th>
<th>lame</th>
<th>dijkstra</th>
<th>patricia</th>
<th>blowfish</th>
</tr>
</thead>
<tbody>
<tr>
<td>180nm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0°C</td>
<td>32K, 16, 1</td>
<td>2K, 16, 1</td>
<td>32K, 8, 2</td>
<td>8K, 16, 2</td>
<td>16K, 16, 1</td>
<td>16K, 32, 2</td>
<td>16K, 16, 1</td>
<td>32K, 8, 4</td>
</tr>
<tr>
<td>20°C</td>
<td>32K, 16, 1</td>
<td>2K, 16, 1</td>
<td>32K, 8, 2</td>
<td>8K, 16, 2</td>
<td>16K, 16, 1</td>
<td>16K, 32, 2</td>
<td>16K, 16, 1</td>
<td>32K, 8, 4</td>
</tr>
<tr>
<td>40°C</td>
<td>32K, 16, 1</td>
<td>2K, 16, 1</td>
<td>32K, 8, 2</td>
<td>8K, 16, 2</td>
<td>16K, 16, 1</td>
<td>16K, 32, 2</td>
<td>16K, 16, 1</td>
<td>32K, 8, 4</td>
</tr>
<tr>
<td>60°C</td>
<td>32K, 16, 1</td>
<td>2K, 16, 1</td>
<td>32K, 8, 2</td>
<td>8K, 16, 2</td>
<td>16K, 16, 1</td>
<td>16K, 32, 2</td>
<td>16K, 16, 1</td>
<td>32K, 8, 4</td>
</tr>
<tr>
<td>80°C</td>
<td>32K, 16, 1</td>
<td>2K, 16, 1</td>
<td>32K, 8, 2</td>
<td>8K, 16, 2</td>
<td>16K, 16, 1</td>
<td>16K, 32, 2</td>
<td>16K, 16, 1</td>
<td>32K, 8, 4</td>
</tr>
<tr>
<td>100°C</td>
<td>32K, 16, 1</td>
<td>2K, 16, 1</td>
<td>32K, 8, 2</td>
<td>8K, 16, 2</td>
<td>16K, 16, 1</td>
<td>16K, 32, 2</td>
<td>16K, 16, 1</td>
<td>32K, 8, 4</td>
</tr>
</tbody>
</table>

| 100nm       |         |          |       |       |      |          |          |         |
| 0°C         | 32K, 16, 1 | 2K, 16, 1 | 32K, 16, 2 | 8K, 16, 2 | 32K, 32, 1 | 32K, 32, 4 | 16K, 16, 1 | 32K, 32, 4 | 32K, 16, 1 |
| 20°C        | 8K, 8, 4 | 2K, 16, 1 | 16K, 8, 4 | 8K, 16, 2 | 16K, 16, 1 | 16K, 32, 2 | 8K, 8, 4 | 32K, 32, 4 | 8K, 8, 4 |
| 40°C        | 8K, 8, 4 | 2K, 32, 1 | 16K, 8, 4 | 8K, 16, 2 | 16K, 32, 1 | 8K, 32, 4 | 8K, 8, 4 | 32K, 32, 4 | 8K, 8, 4 |
| 60°C        | 8K, 8, 4 | 1K, 8, 2 | 16K, 32, 4 | 8K, 32, 2 | 16K, 32, 2 | 8K, 32, 4 | 8K, 8, 4 | 32K, 32, 4 | 8K, 8, 4 |
| 80°C        | 8K, 16, 4 | 1K, 16, 2 | 16K, 32, 4 | 8K, 32, 2 | 8K, 32, 2 | 8K, 16, 4 | 8K, 32, 4 | 32K, 32, 4 | 8K, 16, 4 |
| 100°C       | 8K, 16, 4 | 1K, 16, 2 | 16K, 32, 4 | 8K, 32, 2 | 8K, 32, 4 | 8K, 16, 4 | 8K, 32, 4 | 32K, 32, 4 | 8K, 16, 4 |

| 70nm        |         |          |       |       |      |          |          |         |
| 0°C         | 8K, 16, 4 | 1K, 16, 2 | 16K, 16, 4 | 8K, 32, 2 | 16K, 32, 1 | 8K, 32, 4 | 8K, 16, 4 | 32K, 32, 4 | 8K, 16, 4 |
| 20°C        | 8K, 16, 4 | 1K, 16, 2 | 16K, 16, 4 | 8K, 32, 2 | 8K, 32, 4 | 8K, 16, 4 | 32K, 32, 4 | 8K, 16, 4 |
| 40°C        | 8K, 16, 4 | 1K, 16, 2 | 16K, 16, 4 | 8K, 32, 2 | 8K, 32, 4 | 8K, 16, 4 | 32K, 32, 4 | 8K, 16, 4 |
| 60°C        | 8K, 32, 4 | 1K, 32, 2 | 16K, 32, 4 | 8K, 32, 4 | 8K, 32, 4 | 8K, 32, 4 | 32K, 32, 4 | 8K, 32, 4 |
| 80°C        | 8K, 32, 4 | 1K, 32, 2 | 16K, 32, 4 | 8K, 32, 4 | 8K, 32, 4 | 8K, 32, 4 | 32K, 32, 4 | 8K, 32, 4 |
| 100°C       | 8K, 32, 4 | 1K, 32, 2 | 16K, 32, 4 | 8K, 32, 4 | 8K, 32, 4 | 8K, 32, 4 | 32K, 32, 4 | 8K, 32, 4 |
## Data Cache

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Application</th>
<th>basismath</th>
<th>susan</th>
<th>qsort</th>
<th>cjpeg</th>
<th>djpeg</th>
<th>lame</th>
<th>dijkstra</th>
<th>patricia</th>
<th>blowfish</th>
</tr>
</thead>
<tbody>
<tr>
<td>180nm</td>
<td>0°C</td>
<td>8K, 8, 1</td>
<td>2K, 16, 4</td>
<td>64K, 32, 1</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>16K, 32, 2</td>
<td>32K, 8, 2</td>
</tr>
<tr>
<td>0°C</td>
<td>20°C</td>
<td>8K, 8, 1</td>
<td>2K, 16, 4</td>
<td>64K, 32, 1</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>16K, 32, 2</td>
<td>32K, 8, 2</td>
</tr>
<tr>
<td>20°C</td>
<td>40°C</td>
<td>8K, 8, 1</td>
<td>2K, 16, 4</td>
<td>64K, 32, 1</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>16K, 32, 2</td>
<td>32K, 8, 2</td>
</tr>
<tr>
<td>40°C</td>
<td>60°C</td>
<td>8K, 8, 1</td>
<td>2K, 16, 4</td>
<td>64K, 32, 1</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>16K, 32, 2</td>
<td>32K, 8, 2</td>
</tr>
<tr>
<td>60°C</td>
<td>80°C</td>
<td>8K, 8, 1</td>
<td>2K, 16, 4</td>
<td>64K, 32, 1</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>16K, 32, 2</td>
<td>32K, 8, 2</td>
</tr>
<tr>
<td>80°C</td>
<td>100°C</td>
<td>4K, 32, 1</td>
<td>2K, 16, 4</td>
<td>32K, 32, 1</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>32K, 32, 1</td>
<td>32K, 16, 2</td>
<td>16K, 32, 2</td>
<td>32K, 8, 2</td>
</tr>
<tr>
<td>100°C</td>
<td>0°C</td>
<td>4K, 16, 2</td>
<td>4K, 32, 4</td>
<td>32K, 32, 2</td>
<td>32K, 32, 2</td>
<td>32K, 32, 4</td>
<td>32K, 32, 4</td>
<td>32K, 32, 2</td>
<td>32K, 32, 4</td>
<td>32K, 16, 2</td>
</tr>
<tr>
<td>0°C</td>
<td>20°C</td>
<td>4K, 16, 2</td>
<td>4K, 32, 4</td>
<td>16K, 32, 2</td>
<td>16K, 32, 2</td>
<td>32K, 32, 4</td>
<td>32K, 32, 4</td>
<td>32K, 32, 2</td>
<td>16K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>20°C</td>
<td>40°C</td>
<td>4K, 16, 2</td>
<td>2K, 32, 4</td>
<td>4K, 32, 4</td>
<td>16K, 32, 2</td>
<td>32K, 32, 4</td>
<td>32K, 32, 4</td>
<td>32K, 32, 2</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>40°C</td>
<td>60°C</td>
<td>2K, 8, 2</td>
<td>2K, 32, 4</td>
<td>2K, 32, 4</td>
<td>16K, 32, 2</td>
<td>16K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>60°C</td>
<td>80°C</td>
<td>8K, 8, 2</td>
<td>2K, 32, 4</td>
<td>2K, 32, 4</td>
<td>16K, 32, 2</td>
<td>16K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>80°C</td>
<td>100°C</td>
<td>2K, 16, 1</td>
<td>2K, 32, 4</td>
<td>2K, 32, 4</td>
<td>16K, 32, 2</td>
<td>16K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>100°C</td>
<td>0°C</td>
<td>4K, 32, 2</td>
<td>2K, 32, 4</td>
<td>4K, 32, 4</td>
<td>16K, 32, 2</td>
<td>32K, 32, 4</td>
<td>32K, 32, 4</td>
<td>16K, 32, 2</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>0°C</td>
<td>20°C</td>
<td>4K, 16, 4</td>
<td>2K, 32, 4</td>
<td>2K, 32, 4</td>
<td>8K, 32, 2</td>
<td>16K, 32, 4</td>
<td>16K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>20°C</td>
<td>40°C</td>
<td>2K, 16, 4</td>
<td>2K, 32, 4</td>
<td>2K, 32, 4</td>
<td>8K, 32, 2</td>
<td>8K, 32, 4</td>
<td>16K, 32, 2</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>40°C</td>
<td>60°C</td>
<td>2K, 32, 1</td>
<td>2K, 32, 4</td>
<td>2K, 32, 4</td>
<td>8K, 32, 8</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>4K, 16, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>60°C</td>
<td>80°C</td>
<td>2K, 32, 1</td>
<td>2K, 32, 4</td>
<td>8K, 32, 2</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>8K, 32, 4</td>
<td>4K, 16, 4</td>
<td>8K, 32, 4</td>
</tr>
<tr>
<td>80°C</td>
<td>100°C</td>
<td>2K, 32, 1</td>
<td>1K, 8, 2</td>
<td>1K, 32, 4</td>
<td>8K, 32, 2</td>
<td>4K, 32, 4</td>
<td>8K, 32, 4</td>
<td>4K, 32, 2</td>
<td>4K, 16, 4</td>
<td>8K, 32, 4</td>
</tr>
</tbody>
</table>
Energy Saving & Performance Penalty

Energy Saving =
\[
\frac{(energy\_cache0\_tempN - energy\_cacheN) \times energy\_cache0\_tempN}{exec\_time\_cacheN - exec\_time\_cache0} / exec\_time\_cache0
\]
Instruction Cache – Energy Saving

100nm: 8%, 27%, and 41% for 20°C, 60°C, 100°C (max: 65%)
70nm: 1%, 6%, and 16% for 20°C, 60°C, 100°C (max: 45%)
Instruction Cache – Performance Penalty

100nm: 1%, 1.2%, and 2.2% for 20°C, 60°C, 100°C
70nm: 0.6%, 2.3%, and 16% for 20°C, 60°C, 100°C
Data Cache – Energy Saving

100nm: 3.3%, 25%, and 47% for 20°C, 60°C, 100°C (max: 75%)
70nm: 7%, 22%, and 33% for 20°C, 60°C, 100°C (max: 65%)
Data Cache – Performance Penalty

100nm: 0.8%, 5.3%, and 8% for 20°C, 60°C, 100°C
70nm: 3.6%, 10%, and 20% for 20°C, 60°C, 100°C
Architecture and Reconfiguration Flow for a Temperature-Aware Configurable Cache

- Configurable Cache +
  - Hardware
    - Thermal sensor
    - Accessible read port
  - Software
    - A table in Operating System (OS) for recoding temperature ranges and their suitable cache configuration
Flow of configuring Temperature-Aware Configurable Cache

1. **Evaluation phase (offline)**
   - Detect the current temperature
   - Use the lookup table and load the proper configuration for the current temperature
   - Execute the application

2. **Determine the lowest energy cache configuration for different target temperatures**
   - Execution time, number of hits and misses for different cache configurations obtained through running the application on an ISS

3. **Fill the lookup table of the configurable cache with proper configuration for each temperature**

4. **Static and dynamic power for different cache configuration and temperatures for the target technology**
Temperature measurement accuracy (1/2)

\[ T_j = T_a + \theta_{JA} \cdot P \]

- \( T_j \): Junction Temperature
- \( T_a \): Ambient Temperature
- \( P \): Power
- \( \theta_{JA} \): Junction-to-Ambient Thermal Resistance
Temperature measurement accuracy (2/2)

\[ \theta_{JA} : 7 \, ^\circ C/W \sim 35 \, ^\circ C/W \]

\[ \Delta T = (T_j - T_a) \sim 5 \, ^\circ C \]

<table>
<thead>
<tr>
<th></th>
<th>ARM7TDMI</th>
<th>ARM966E-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>180nm</td>
<td>Power consumption</td>
<td>24.15 mW</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>115 MHz</td>
</tr>
<tr>
<td>130nm</td>
<td>Power consumption</td>
<td>7.98 mW</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>133 MHz</td>
</tr>
<tr>
<td>90nm</td>
<td>Power consumption</td>
<td>7.08 mW</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>236 MHz</td>
</tr>
</tbody>
</table>
Conclusions

- Our results show that up to 66% and 45% energy consumption can be saved for 100nm and 70nm for instruction cache when the temperature changes from 0°C to 100°C.
- Due to the increase of leakage effect in finer technologies and higher temperatures, the smaller caches will be more energy efficient for future low energy systems.
- Since the smaller caches are more suitable for low energy systems in finer technologies and higher temperatures, finding an optimal cache configuration that simultaneously optimizes performance and energy is increasingly more difficult in future, specially at high temperatures.
- Since the accesses to data cache are less than the accesses to instruction cache, the data cache is more easily affected by temperature and technology than instruction cache. By using a configurable data cache, up to 74% and 64% energy can be saved for 100nm and 70nm respectively.