A 90nm Data Flow Processor Demonstrating Fine Grained DVS for Energy Efficient Operation from 0.25V to 1.2V

University of Virginia, Charlottesville, VA
Custom Integrated Circuits Conference
September 21, 2011
Motivation

Portability applications require extended battery life & small form factor.

High performance struggles with power density.

- Portable applications, microsensors: Energy constrained
- Medical devices, microsensors: Energy constrained
- Portable electronics, ambient intelligence: Workloads vary; Maximize lifetime

- High Performance Apps: Power density

Electronics need high performance for a fraction of their life.
Classic Dynamic Voltage Scaling (DVS)

- Adjust $V_{\text{DD}}$ to adjust speed to match workload
- Usually chip or core level
- Use DC-DC converters to adjust $V_{\text{DD}} = \text{Slow, infrequent } V_{\text{DD}}$ transitions
- Constrained by block with highest workload

→ Need for finer granularity in space and time
Panoptic DVS (PDVS) Structure

- **Single-\(V_{DD}\)**
  - All components share one \(V_{DD}\)

- **Multi-\(V_{DD}\)**
  - Each component statically tied to a \(V_{DD}\)

- **PDVS**
  - PMOS header switches used to select a specific \(V_{DD}\)
  - Small set of voltage rails (2-4)
  - Uses common components
  - Fine temporal granularity
  - Fine spatial granularity

Contributions

This work:

- Demonstrates DSP Processor using PDVS
- Demonstrates single clock cycle $V_{DD}$-switching & $V_{DD}$ dithering for near optimal energy scalability
- Demonstrates switch efficiently between high performance DVS and subthreshold modes.
- Demonstrates energy savings compared to single-$V_{DD}$ and multi-$V_{DD}$ alternatives.
Outline

- Chip Architecture
- Measured Results
- Subthreshold Mode
- Conclusion
Chip Architecture

- **32b Data Flow Processor**
  - 4 Kogge Stone Adders
  - 4 Baugh Wooley Multipliers
  - PMOS Header
  - Level Converters

- Execute arbitrary data flow graphs (DFGs)

- **32 Kb Data Memory**

- **40 Kb Instruction Memory**

- **Register File**
2-stage Pipelined SRAM

- During the first phase of a conventional SRAM read access when row decode occurs, the SA lies idle.
- In our scheme, in the first cycle, row decode and BL droop development is completed.
- At the beginning of the second cycle, SA enable signal occurs.
- Col Mux acts as pipeline register.
- Helps lower the cycle time by $T_{SA}$.
- 23% reduction of the cycle time.
## Chip Vitals

<table>
<thead>
<tr>
<th>Feature</th>
<th>This Chip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>90nm CMOS Bulk w/ Dual $V_T$</td>
</tr>
<tr>
<td>Area</td>
<td>4.3mm x 3.3mm</td>
</tr>
<tr>
<td>Transistor Count</td>
<td>~2 million</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>250mV – 1.2V</td>
</tr>
<tr>
<td>SRAMs</td>
<td>40kb &amp; 32kb</td>
</tr>
</tbody>
</table>
Outline

- Chip Architecture
- **Measured Results**
- Subthreshold Mode
- Conclusion
Dithering

Time Constraint

Dithering

Required rates

Quantized rates

[Gutnik, SympVLSI'96]
System Results

Measured Energy Savings

Normalized Energy Workload

SV

MV

PDVS
System Results Cont’d

- Measured Energy Savings
- Normalized Energy Workload
- SV\textsubscript{DD}, MV\textsubscript{DD}, PDVS

Graphs showing normalized energy consumption and measured energy savings for different workloads and compared with SV\textsubscript{DD}, MV\textsubscript{DD}, and PDVS.
System Results Cont’d

Introducing slack increases savings

Measured Energy Savings

Area Savings

Normalized Energy

Workload
PDVS area savings are a result of reducing the number of copies
Measured Dithering
# Overheads

<table>
<thead>
<tr>
<th></th>
<th>32b Adder</th>
<th>32b Mult.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Header Area</td>
<td>2.4%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Level Conv. Delay</td>
<td>32.0%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Level Conv. Energy</td>
<td>8.0%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Level Conv. Area</td>
<td>11.4%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Sw. Delay</td>
<td>10.4%</td>
<td>12.0%</td>
</tr>
<tr>
<td>Sw. Energy</td>
<td>215.3%</td>
<td>35.0%</td>
</tr>
<tr>
<td>Breakeven Cycles (N_{BE})</td>
<td>&lt; 4</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

\[
N_{BE} = \frac{(E_{High} - E_{Low})}{E_{switch}}
\]
Outline

- Chip Architecture
- Measured Results
- Subthreshold Mode
- Conclusion
Subthreshold Operation

- Chip provides subthreshold mode
  - E.g. $V_{\text{DDH/M/L}}$ @ 1V, 0.5V 0.25V
- Minimum energy operating point
Subthreshold Circuit Optimizations

- Added PMOS headers to:
  - Register bank
  - Crossbar
- Tied circuit & Sub-\(V_T\) header body connection to Virtual \(V_{DD}\)
- Optimized level converter for subthreshold
  - Converts from 0.25 up to 1.0V
- Added bypass for these level converters
Outline

- Chip Architecture
- Measured Results
- Subthreshold Mode
- Conclusion
Conclusion

- First processor using PDVS
- Demonstrate single clock cycle $V_{DD}$-switching & $V_{DD}$ dithering for near optimal energy scalability
- Switch efficiently between high performance DVS and subthreshold modes.
- Energy savings up to 50% and 46% compared to single-$V_{DD}$ and multi-$V_{DD}$ alternatives.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Howard ISSCC10</th>
<th>Truong VLSI08</th>
<th>Nam ISSCC07</th>
<th>This work</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{DD}$ Granularity</td>
<td>6 cores</td>
<td>1 core</td>
<td>1 core</td>
<td>Add, Mult</td>
</tr>
<tr>
<td>Speed of $V_{DD}$ change</td>
<td>&gt;10µs (e.g.[4])</td>
<td>2-5ns</td>
<td>&gt;10µs (e.g. [4])</td>
<td>&lt;2ns</td>
</tr>
<tr>
<td>$V_{DD}$ dithering</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sub-threshold</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Thank you!

Any Questions?