A Power-Efficient 32b ARM ISA Processor Using Timing-Error Detection and Correction for Transient-Error Tolerance and Adaptation to PVT Variation

David Bull¹, Shidhartha Das¹, Karthik Shivashankar¹, Ganesh Dasika², Krisztian Flautner¹, David Blaauw²

¹ARM Ltd., U.K.
²University of Michigan
Design Margins

CLK

- process
- voltage
- temp
- coupling
- jitter
- ageing

Safety

<table>
<thead>
<tr>
<th>STATIC</th>
<th>SLOW-CHANGING</th>
<th>FAST-CHANGING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-die process variation (BTI, TDDB, EM)</td>
<td>Regulator Ripple Ambient temperature variation</td>
<td>PLL jitter IR drop Ldi/dt</td>
</tr>
<tr>
<td>Wear-out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOCAL GLOBAL</td>
<td>Intra-die process variation</td>
<td>Hot-spots</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Razor principles

**Key idea:** Exploit the dynamic nature of variations

- Speculatively operate without full setup margin
- Explicitly check for late-arriving signals
- In the event of a timing error, invoke system recovery mechanism
- Adapt VDD/CLK to target near-zero error-rate operation

**Survive fast moving and transient changes**

- Ldi/dt
- Localized IR drop
- PLL Jitter
- Capacitive coupling
- Critical-path sensitization

**Adapt to slower moving or static conditions**

- Global or long-term IR drop
- Low-frequency supply ripple
- Temperature
- Ageing
- Process variation
Razor-enabled energy-efficient ARM processor

UMC 65SP (High Performance) Process
- 1V nominal VDD and 1.1V Overdrive

Implements a sub-set of ARM ISA
- Critical-paths representative of ARM industrial processor designs

87 die from split lots
- 30FF/37TT/20SS

724MHz sign-off frequency
- 0.9V/SS/125C

Adaptive Control Experiments
- Adaptive Frequency Control - DFS
- Adaptive Voltage Control - DVS
Outline

- Motivation and Razor background

  ➤ Transition-Detector circuit design

- Micro-architecture design

- Adaptive voltage and frequency scaling

- Parametric yield improvement with Razor

- Conclusion
Transition-Detector Circuit Design

D → Main Flip-Flop → Q

Delay on CK defines CK pulse width

D → Pulse-generators generate pulses out of transitions on D.

ERN → Sticky Error history bit identifies failing FF for off-line diagnostics

HRN
Transition-Detector Circuit Design

Earliest Detection

CK

nCK

D

DP

TD

ERROR

TCK

D

ERROR

Earliest Detection
Transition-Detector Circuit Design

Error Detection Window

= TD + TCK – 2TOV

Latest Detection
Transition-Detector Circuit Design

Min Delay Constraint

= TCK – TOV

Minimum Delay
Transition-Detector Comparison

Advantages

- Reduced min-delay constraint
- Operates with conventional 50% clocking
- Simplifies integration with a conventional ASIC flow

Disadvantages

- Flagging errors before actual failure occurs incurs performance penalty
- Additional transistors on the clock network

Trade-off setup pessimism for reduced min-delay
Micro-architecture Design

Balanced pipe-stages with critical-endpoints at clock-gating, IRAM and DRAM inputs protected by Transition-Detectors
Stabilization stages allow sufficient time for Razor validation of critical signals and synchronization overhead of ERROR
Recovery occurs by replaying the pipeline from the last un-committed instruction at half-frequency
## Implementation Details

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flip-flops</td>
<td>2976</td>
</tr>
<tr>
<td>Flip-flops with TD</td>
<td>503 (17%)</td>
</tr>
<tr>
<td>ICGs</td>
<td>149</td>
</tr>
<tr>
<td>ICGs with TD</td>
<td>27</td>
</tr>
<tr>
<td>TD for RAMs</td>
<td>20</td>
</tr>
<tr>
<td>TD Power Overhead</td>
<td>5.7%</td>
</tr>
<tr>
<td>Power Overhead of Min-delay Buffers</td>
<td>1.3%</td>
</tr>
<tr>
<td>Stabilization Stages Power Overhead</td>
<td>2.4%</td>
</tr>
<tr>
<td>Total Power Overhead</td>
<td>8.4%</td>
</tr>
<tr>
<td>Total Area Overhead @ 70% utilization</td>
<td>6.9%</td>
</tr>
<tr>
<td>Measured Setup Pessimism of TD</td>
<td>5% @ 1GHz/1V</td>
</tr>
<tr>
<td>IRAM and DRAM size</td>
<td>2KB</td>
</tr>
</tbody>
</table>
• 4 TDs fail at 1.1GHz compared to 122 at 1.2GHz
Comparing Different Workloads - #TT9 1V VDD

- Significant variation in PoFF across workloads
Frequency Tuning – Fixed 1V VDD #TT9

NOP  Power Virus  Typical

Slow-down on every error
Speedup on 1024 cycles without error
Voltage Tuning – Fixed 1GHz Frequency #TT9

- NOP
- Power Virus
- Typical

Voltage (V)

- 1.07V
- 0.97V
- TT9

Time

TT9 Errors

- 2001
- 1501
- 1001
- 501
- 1
Voltage Tuning – Fixed 1GHz Frequency #TT9

1.1V (3% margin)

1.07V

0.97V

30% power saving

Voltage Tuning

Time

TT9 Errors
SS/TT/FF Comparison – 1GHz Frequency

<table>
<thead>
<tr>
<th>Condition</th>
<th>NOP</th>
<th>Power Virus</th>
<th>Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage (V)</td>
<td>1.17V</td>
<td>1.17V</td>
<td>1.07V</td>
</tr>
<tr>
<td>Voltage (V)</td>
<td>1.08V</td>
<td>1.08V</td>
<td>1.07V</td>
</tr>
<tr>
<td>Voltage (V)</td>
<td>1.03V</td>
<td>1.03V</td>
<td>1.07V</td>
</tr>
<tr>
<td>Voltage (V)</td>
<td>0.97V</td>
<td>0.97V</td>
<td>0.92V</td>
</tr>
<tr>
<td>Voltage (V)</td>
<td>0.92V</td>
<td>0.92V</td>
<td>0.92V</td>
</tr>
</tbody>
</table>

Time
Minimum Voltage – 1GHz Operation

- **NOP**
  - 1.17V
  - 1.08V
  - 1.03V

- **Power Virus**
  - 1.2V
  - 3% margin

- **Typical**
  - 1.07V SS6
  - 0.97V TT9
  - 0.92V FF5
1.2V vs Razor – Typical Workload

- Tune voltage to zero margin point using Razor
- SS6 part now consumes maximum power
- Power outlier for distribution reduces from 100mW to 48mW with Razor for typical code

Diagram showing power savings with Razor tuned VDD, with a 52% power saving.
Power distribution at 1.2V vs Razor

- Power distribution without Razor is wide
- Razor improves both the $\mu$ and the $\sigma$ of the distribution

Power Distribution
- OD (1.2V): 30mW (40%)
Parametric Yield

With Razor 1GHz operation is possible at 1.1V
- All code except pathological power virus runs below 1.1V

Without Razor 1GHz operation is only possible at 1.2V
- Power virus code requires 1.2V for SS6
- 1.2V exceeds 1.1V overdrive limit of the process
- Excessive leakage and wear-out implications

Discarding fast/leaky parts and slow parts might be correct trade-off without Razor
- Limit overdrive to 1.1V with parametric screening
Parametric Yield – Native Distribution

87 devices at 1.1V

<table>
<thead>
<tr>
<th>Power at 1GHz (mW)</th>
<th>Number of Chips</th>
</tr>
</thead>
<tbody>
<tr>
<td>87 devices at 1.1V</td>
<td>30</td>
</tr>
<tr>
<td>FF (30)</td>
<td></td>
</tr>
<tr>
<td>TT (37)</td>
<td></td>
</tr>
<tr>
<td>SS (20)</td>
<td></td>
</tr>
</tbody>
</table>

Maximum Frequency at 1.1V (MHz)

- FF5
- SS6
Parametric Yield – Power vs Frequency

Power at 1GHz (mW) vs Maximum Frequency at 1.1V (MHz)

FF (30) 1.1V OD
TT (37)
SS (20)

87 devices at 1.1V
Parametric Yield – Power vs Frequency

87 devices at 1.1V

Power at 1GHz (mW)

Maximum Frequency at 1.1V (MHz)

Power Limit

1.1V OD

Frequency Limit

FF (30)

TT (37)

SS (20)
Parametric Yield – Prune Distribution

Power at 1GHz (mW)

Maximum Frequency at 1.1V (MHz)

1.1V OD
Power Limit
Yielding Parts

Frequency Limit

Yielding Parts

28
Parametric Yield – Prune Distribution

Power at 1GHz (mW) >60mW (21)

Yielding Parts = 28 out of 87

1.1V OD

<1GHz (38)

Maximum Frequency at 1.1V (MHz)
Parametric Yield – Razor

1.1V OD

Power at 1GHz (mW)>

60mW (0)

Yielding Parts

= 87 out of 87

<1GHz

(0)

<1GHz

(0)

Razor

Maximum Frequency at 1.1V (MHz)
Parametric Yield – Razor

- Power at 1GHz (mW) >60mW (0)
- Yielding Parts = 87 out of 87
- 1.1V OD
- Maximum Frequency at 1.1V (MHz) 20% power saving
Parametric Yield – 100% yield at 1.1V vs Razor

Yielding Parts = 87 out of 87

Razor

Power at 1GHz (mW)

Maximum Frequency at 1.1V (MHz)

- 1.1V OD
- 78mW
- 38% power saving
- 890MHz
- 14% Fmax gain
Summary and Conclusion

- Reclaim margins for gains in energy-efficiency and parametric yield
- Obtained 52% power saving at 1GHz operation on an ARM prototype through Razor
- Developed a new Transition-Detector design with reduced min-delay impact
- Demonstrated run-time adaptation to PVT variations and tolerance to fast transients
- Demonstrated potential for parametric yield improvements using Razor
Backup Slides
Tracking Circuits

Multiple worst-case paths converge to the same end-point

- 100 paths within 70ps (3%) of the critical-path to same endpoint
- 377 unique instances and 119 unique cell masters covered by the paths
- Extracted critical-path spice netlist has 9120 resistors, 2413 coupling and ground capacitors and 1442 instances including aggressors

Requires multiple tracking circuits for reasonable approximation

*Alternatively, just 1 Razor flop at the end-point is sufficient*
**Transition-Detector Timing Diagram**

Min Delay Constraint = TCK - TOV

Error Detection Window = TCK + TD - 2TOV

**Advantages**
- Reduced min-delay constraint
- 50% duty-cycle clocking

**Disadvantages**
- Setup pessimism
- Extra clock transistors
## Parametric Yield – Yield Loss TT Lot

<table>
<thead>
<tr>
<th>PASS</th>
<th>FAIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT Chip</td>
<td>PV PoFF</td>
</tr>
<tr>
<td>TT51</td>
<td>1.026</td>
</tr>
<tr>
<td>TT56</td>
<td>1.035</td>
</tr>
<tr>
<td>TT52</td>
<td>1.054</td>
</tr>
<tr>
<td>TT54</td>
<td>1.054</td>
</tr>
<tr>
<td>TT55</td>
<td>1.060</td>
</tr>
<tr>
<td>TT5</td>
<td>1.061</td>
</tr>
<tr>
<td>TT7</td>
<td>1.062</td>
</tr>
<tr>
<td>TT14</td>
<td>1.063</td>
</tr>
<tr>
<td>TT19</td>
<td>1.065</td>
</tr>
<tr>
<td>TT57</td>
<td>1.066</td>
</tr>
<tr>
<td>TT58</td>
<td>1.066</td>
</tr>
<tr>
<td>TT17</td>
<td>1.068</td>
</tr>
<tr>
<td>TT8</td>
<td>1.068</td>
</tr>
<tr>
<td>TT60</td>
<td>1.069</td>
</tr>
<tr>
<td>TT9</td>
<td>1.071</td>
</tr>
<tr>
<td>TT31</td>
<td>1.071</td>
</tr>
<tr>
<td>TT47</td>
<td>1.072</td>
</tr>
<tr>
<td>TT53</td>
<td>1.075</td>
</tr>
<tr>
<td>TT34</td>
<td>1.079</td>
</tr>
<tr>
<td>TT3</td>
<td>1.08</td>
</tr>
<tr>
<td>TT18</td>
<td>1.08</td>
</tr>
<tr>
<td>TT32</td>
<td>1.084</td>
</tr>
<tr>
<td>TT16</td>
<td>1.084</td>
</tr>
<tr>
<td>TT10</td>
<td>1.087</td>
</tr>
<tr>
<td>TT33</td>
<td>1.09</td>
</tr>
<tr>
<td>TT45</td>
<td>1.09</td>
</tr>
<tr>
<td>TT11</td>
<td>1.094</td>
</tr>
<tr>
<td>TT12</td>
<td>1.097</td>
</tr>
</tbody>
</table>
Voltage Controller Transition Response

NOP to Power Virus Transition

Power Virus to Typical Transition
Transition-Detector Timing Diagram

- Cover setup time with sufficient margin

Min Delay Constraint = TCK - TOV

TD + TCK - 2TOV  Earliest D Detection
Throughput versus Frequency

Typical Workload

Normalized Throughput

Frequency (MHz)

Signoff Frequency

PoFF

Number of Failing TDs
Adaptive Frequency Control

- **31-tap Ring Oscillator** used as the clock source in adaptive mode
- Course selection through changing tap setting
- Switched cap network for fine-grained frequency setting
- Programmable control algorithm