A 256kb 6T Self-Tuning SRAM with Extended 0.38V-1.2V Operating Range using Multiple Read/Write Assists and $V_{MIN}$ Tracking Canary Sensors

*Electrical and Computer Engineering
University of Virginia, Charlottesville

** John Poulton, **C. Thomas Gray
** Nvidia, Durham, North Carolina
IoE market rapidly growing
Motivation

- IoE market rapidly growing
- Battery recharge and replacement problems

IoE = Internet of Everything
Motivation

- IoE market rapidly growing
- Battery recharge and replacement problems
- Soln: ULP wide-range DVS

IoE = Internet of Everything, ULP = Ultra-low Power, DVS = Dynamic Voltage Scaling
Bottleneck in ULP wide-$V_{DD}$ Range

- $V_{\text{MIN}}$ guard-banding in 6Ts
Bottleneck in ULP wide-$V_{\text{DD}}$ Range

- $V_{\text{MIN}}$ guard-banding in 6Ts
- SRAM $V_{\text{MIN}}$ DVS bottleneck
Bottleneck in ULP wide-$V_{DD}$ Range

- $V_{MIN}$ guard-banding in 6Ts
- SRAM $V_{MIN}$ DVS bottleneck
- Dual rail SRAMs SoC level tradeoffs
Bottleneck in ULP wide-$V_{DD}$ Range

- $V_{MIN}$ guard-banding in 6Ts
- SRAM $V_{MIN}$ DVS bottleneck
- Dual rail SRAMs SoC level tradeoffs
- Different solutions across applications
SRAM Solutions for ULP Applications

Trading off performance and area

[Source: Cliff Hou, TSMC, ISSCC 2017]
Scope of 6T SRAM Improvements

- Peripheral assist for $V_{MIN}$ improvement

28nm TT_80C at $V_{MIN}$ (without assist)

[Source: A. Banerjee. et al. ISQED 2014]
Scope of 6T SRAM Improvements

- Peripheral assist for $V_{\text{MIN}}$ improvement
- Lower $V_{\text{MIN}}$ guard-band

Tracking $V_{\text{MIN}}$ could save energy

Normalized SRAM write energy per cycle at $V_{\text{MIN}}$

[Source: A. Banerjee. et al. ISQED 2014]
Scope of 6T SRAM Improvements

- Peripheral assist for $V_{\text{MIN}}$ improvement
- Lower $V_{\text{MIN}}$ guard-band
- Proposed Solution
  - Combined assist$^1$ and Canary based $V_{\text{MIN}}$ tracking$^2$ reducing guardbanding

---

[Source: A. Banerjee. et al. ISQED 2014]

---

$^{[1]}$E. Karl et al., 2012; $^{[2]}$A. Banerjee et al. 2015]
Agenda

- Canary SRAM Sensors
- Peripheral Assists and Reverse Assists
- 256kb Self-tuning SRAM Architecture
- Experiments & Results
- Comparison
- Conclusion
Canary SRAM Sensors

- Canary SRAM a sensor or detector

[Source: http://animalphotos.info/a/topics/animals/birds/canaries/]
Canary SRAM Sensors

- Canary SRAM a sensor or detector
- Fails earlier than the population of SRAM bits

[Source: http://animalphotos.info/a/topics/animals/birds/canaries/]
Canary SRAM Sensors

- Canary SRAM a sensor or detector
- Fails earlier than the population of SRAM bits
- Prior work was in SRAM DRV* tracking¹

¹ [J. Wang, and B. H. Calhoun, CICC, 2007], *DRV=Data retention voltage;
Agenda

- Canary SRAM Sensors
- Peripheral Assists and Reverse Assists
- 256kb Self-tuning SRAM Architecture
- Experiments & Results
- Comparison
- Conclusion
Peripheral Assists and Reverse Assist

- What is a peripheral assist (PA) in SRAM context?
  - An auxiliary circuit that improve read/write-ability
Peripheral Assists and Reverse Assist

What is a peripheral assist (PA) in SRAM context?
- An auxiliary circuit that **improve** read/write-ability

What is reverse assist?
- An auxiliary circuit that **degrades** read/write-ability

SRAM bitcell + Reverse Assist = SRAM bitcell
Example: SRAM write $V_{\text{MIN}}$ Distribution with Reverse Assist Settings (RAS)

Canary write $V_{\text{MIN}}$ distribution shifts right with increasing RAS

[Source: A. Banerjee. et al. ISQED 2014]
## Input and Output Design Metrics

<table>
<thead>
<tr>
<th><strong>Input Metrics</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>Number of SRAM bits on a chip</td>
</tr>
<tr>
<td><strong>$Y_{SRAM}$</strong></td>
</tr>
<tr>
<td>Core SRAM target yield</td>
</tr>
<tr>
<td><strong>C</strong></td>
</tr>
<tr>
<td>Number of canary SRAM bits</td>
</tr>
<tr>
<td><strong>$F_{th}$</strong></td>
</tr>
<tr>
<td>Canary failure threshold condition</td>
</tr>
<tr>
<td><strong>$V_{RA}$</strong> ($RAS^1$)</td>
</tr>
<tr>
<td>Canary BL type reverse assist voltage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Output Metrics</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$P_{fc}$</strong></td>
</tr>
<tr>
<td>Canary SRAM chip failure probability</td>
</tr>
</tbody>
</table>

[Source: A. Banerjee. et al. ISQED 2014]

$^1$RAS = Reverse assist settings; $F_{th}$ = Failure threshold condition
Agenda

- Canary SRAM Sensors
- Peripheral Assists and Reverse Assists
- 256kb Self-tuning SRAM Architecture
- Experiments & Results
- Comparison
- Conclusion
256kb Self-tuning SRAM Architecture

WLB = Wordline Boost, NBL = Negative Bitline, VDD = Vdd Boost.
$V_{\text{MIN}}$ Self-tuning Operation

1. Start
2. TRACK=1?
   - No: Keep the prev. $V_{DD}$ and Assist Settings
   - Yes: Initialize LDO $V_{DD}$ and Assists
3. Run CBIST
4. Canary Pass?
   - No: Increase LDO $V_{DD}$
   - Yes: Decrease LDO $V_{DD}$
5. Select Assists @ $V_{DD}$
6. End
Agenda

- Canary SRAM Sensors
- Peripheral Assists and Reverse Assists
- 256kb Self-tuning SRAM Architecture
- Experiments & Results
- Comparison
- Conclusion
Experiments and Results

- SRAM + PAs = Max 240mV $V_{\text{MIN}}$ improvements
- Does not eliminate $V_{\text{MIN}}$ guard-bands

$PA =$ Peripheral assists
$V_{\text{MIN}}$ Lowering using Combined Read/Write Assists

Measured CDF showing $V_{\text{MIN}}$ improvement w/ combined assist
Experiments and Results

- SRAM + PAs + Canaries = Arbitrary guard-band lowering can save 1444X active power

- SRAM + PAs + Canaries = 12.4X leakage savings

*PA=Peripheral assists*
Active Power Reduction using Combined Assist and Guard-band Lowering Canary Tracking

- Actual chip $V_{\text{MIN}}$
- 90 percentile worst case $V_{\text{MIN}}$
- Combined Assists
- 337X reduction
- 1444X reduction
- 0.38V, 54.3µW
- 0.47V, 12.6µW
- 4.3X reduction ($V_{\text{MIN}}$ tracking)
- 0.6V, 0.65V
- No assist
- $V_{\text{DDDB} + \text{WLB} + \text{NBLA}}$

Measured active power improvements
Active Power Reduction using Combined Assist and Guard-band Lowering Canary Tracking

Measured active power improvements

- Actual chip $V_{\text{MIN}}$
- 90 percentile worst case $V_{\text{MIN}}$
- Combined Assists
- 1444X reduction
- 0.38V, 12.6µW
- 0.47V, 54.3µW
- 4.3X reduction ($V_{\text{MIN}}$ tracking)
- 0.6V
- 0.65V
- 337X reduction
- 1.2V, 18.3mW
Active Power Reduction using Combined Assist and Guard-band Lowering Canary Tracking

Measured active power improvements

- Actual chip $V_{\text{MIN}}$
- 90 percentile worst case $V_{\text{MIN}}$
- Combined Assists

- 1444X reduction
- 0.38V, 12.6µW
- 0.47V, 54.3µW
- 4.3X reduction ($V_{\text{MIN}}$ tracking)
- 0.6V, 0.65V

- 337X reduction
- 0.6V, 0.65V

- 1.2V, 18.3mW
Leakage Power Reduction using Combined Assist and Canary Tracking

Actual chip $V_{\text{MIN}}$

12.4X leakage reduction from $V_{\text{DD}}$ scaling at 0.38V

Measured leakage power improvements
Canary $V_{\text{MIN}}$ Tracking @ 130nm Bulk

Measured canary $V_{\text{MIN}}$ tracking across frequencies

Fth=1500, RAS=011
Canary $V_{\text{MIN}}$ Tracking @ 130nm Bulk

Fth=1500, RAS=011

Settled system $V_{\text{MIN}} > $ SRAM $V_{\text{MIN}}$

Measured canary $V_{\text{MIN}}$ tracking across frequencies
Canary $V_{\text{MIN}}$ Tracking @ 130nm Bulk

Measured canary $V_{\text{MIN}}$ tracking across frequencies

Fth=1500, RAS=011
Scalability of Canary Tracking @ 32nm FDSOI

Simulation results of canary based $V_{\text{MIN}}$ Tracking at TT corner

Operating frequency (MHz)

Canary/SRAM $V_{\text{MIN}}$ (V)

Settled system $V_{\text{MIN}} >$ SRAM $V_{\text{MIN}}$

Canary tuning range

32nm 27C
Overhead

- Canary area overhead only 0.77% (array)

- Combined assist area overhead 2.8% in SRAM

- Total system components without BISTs 1.8%

- Onetime canary tuning (matching the worst case SRAM bitcell) overhead

- Running ~ 90/282 cycles/$V_{DD}$ granularity per frequency/temp change for full 512b/2kb canaries
Agenda

- Canary SRAM Sensors
- Peripheral Assists and Reverse Assists
- 256kb Self-tuning SRAM Architecture
- Experiments & Results
- Comparison
- Conclusion
### Comparison

<table>
<thead>
<tr>
<th>Memory Features</th>
<th>VLSI’15 Technology</th>
<th>VLSI’15 Cell type</th>
<th>VLSI’15 Capacity</th>
<th>This work Technology</th>
<th>This work Cell type</th>
<th>This work Capacity</th>
<th>ISSCC’15 Technology</th>
<th>ISSCC’15 Cell type</th>
<th>ISSCC’15 Capacity</th>
<th>VLSI’14 Technology</th>
<th>VLSI’14 Cell type</th>
<th>ISSCC’12 Technology</th>
<th>ISSCC’12 Cell type</th>
<th>ISSCC’12 Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>14nm</td>
<td>8T</td>
<td>288kb</td>
<td>130nm</td>
<td>6T</td>
<td>256kb</td>
<td>28nm</td>
<td>6T</td>
<td>256kb</td>
<td>180nm</td>
<td>8T</td>
<td>22nm</td>
<td>6T</td>
<td>576KB</td>
</tr>
<tr>
<td>Cell type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capacity</td>
<td>288kb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVS/VMIN Features</td>
<td>DVS range</td>
<td>1-0.3V (700mV)</td>
<td>288kb</td>
<td>1.2-0.38V (850mV)</td>
<td>256kb</td>
<td>0.9-0.58V (320mV)</td>
<td>1.8-0.6V (1200mV)</td>
<td>1-0.625V (375mV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V&lt;sub&gt;MIN&lt;/sub&gt; Tracking</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V&lt;sub&gt;MIN&lt;/sub&gt;</td>
<td>0.3V</td>
<td>0.38V</td>
<td>0.58V</td>
<td>0.6V</td>
<td>0.7V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply / Power</td>
<td>Sub-VT Operation</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max power Reduction</td>
<td>-</td>
<td>1444X</td>
<td>-</td>
<td>16.4X</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- DVS: Dynamic Voltage Scaling
- VMIN: Minimum Voltage
- Sub-VT: Sub-threshold Voltage Technique
- Power Reduction

---

**Source:**
- VLSI’15
- ISSCC’15
- VLSI’14
- ISSCC’12
Conclusion

- A wide DVS range (1.2V-0.38V) with lower SRAM $V_{\text{MIN}}$ (0.38V) achieved using multiple assists (write/read) across supplies

- Canary sensors track SRAM $V_{\text{MIN}}$ for margin guard-band minimization

- Demonstrated a reliable and an adaptive SRAM system selecting optimal $V_{\text{DD}}$ and assist techniques for ULP IoE enablement
Acknowledgements

- Advisor: Professor Ben Calhoun

- UVa Colleagues: Harsh Patel, Ningxi Liu, Farah Yahya, Divya A. K., Kevin Leach, Dilip Vasudevan, Terry Tigner

- Nvidia Colleagues: Tom Gray and John Poulton

- These projects was supported in part by NVIDIA through the DARPA PERFECT program
Thank You