

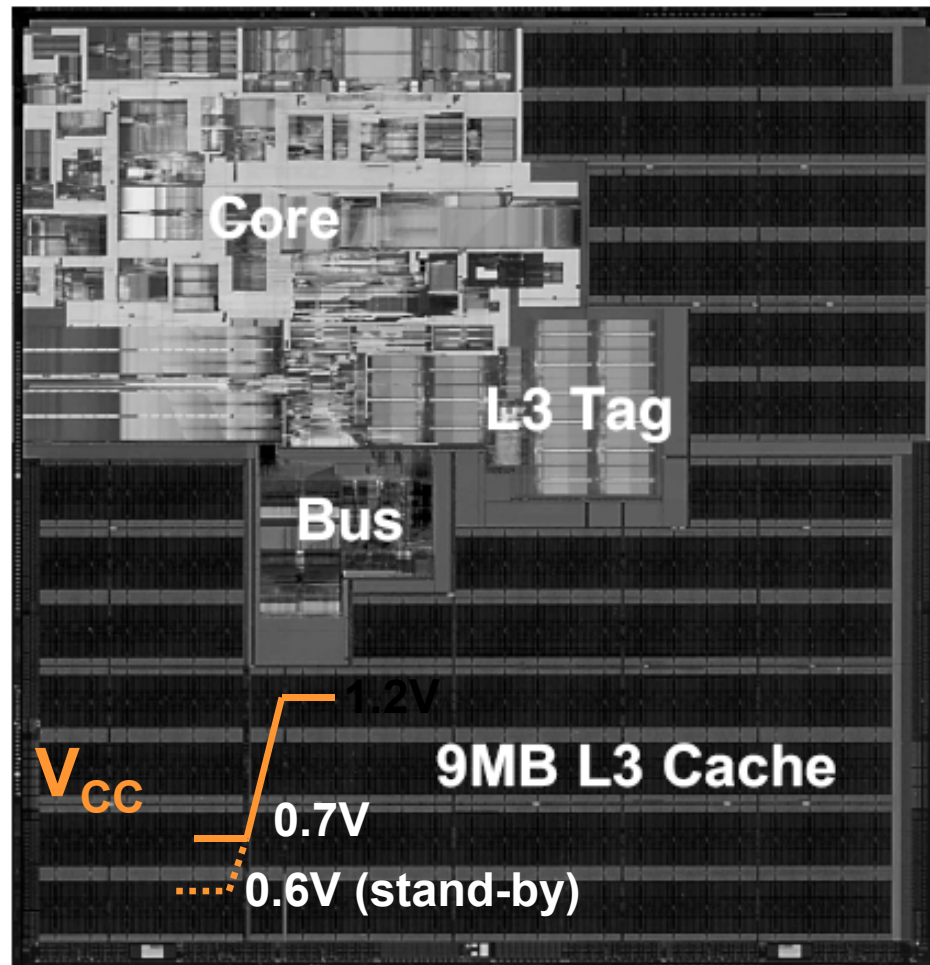
SRAM Scaling Limit: Its Circuit & Architecture Solutions

Nam Sung Kim, Ph.D.
Assistant Professor

Department of Electrical and Computer Engineering
University of Wisconsin - Madison

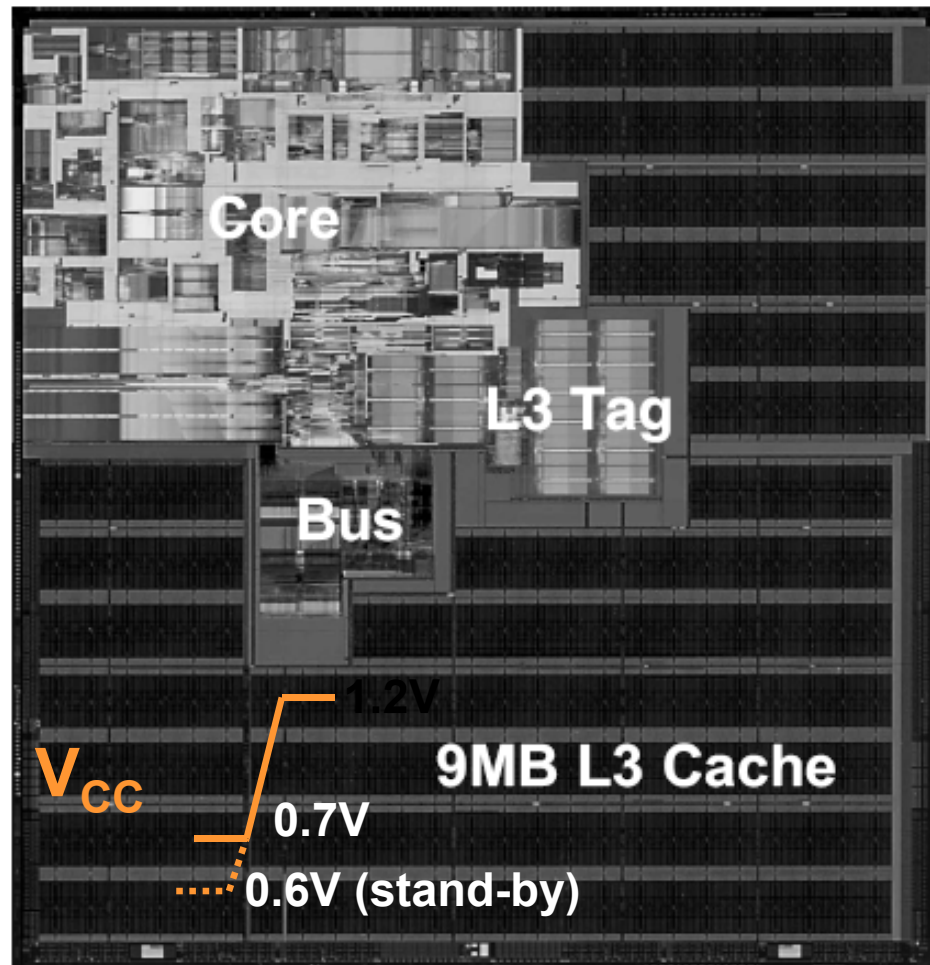


SRAM $V_{CC_{min}}$ Challenges



- VCC/Freq scaling for power efficient computing requires SRAMs to operate at low VCC
- Increasing process variations exacerbate SRAM failures limiting the lowest core operating VCC – $V_{CC_{min}}$

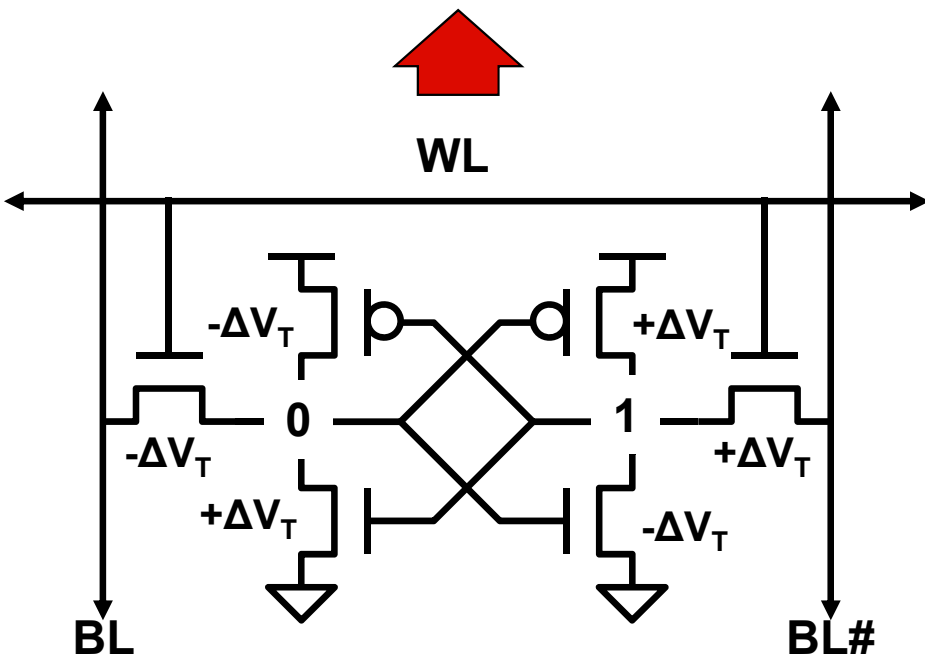
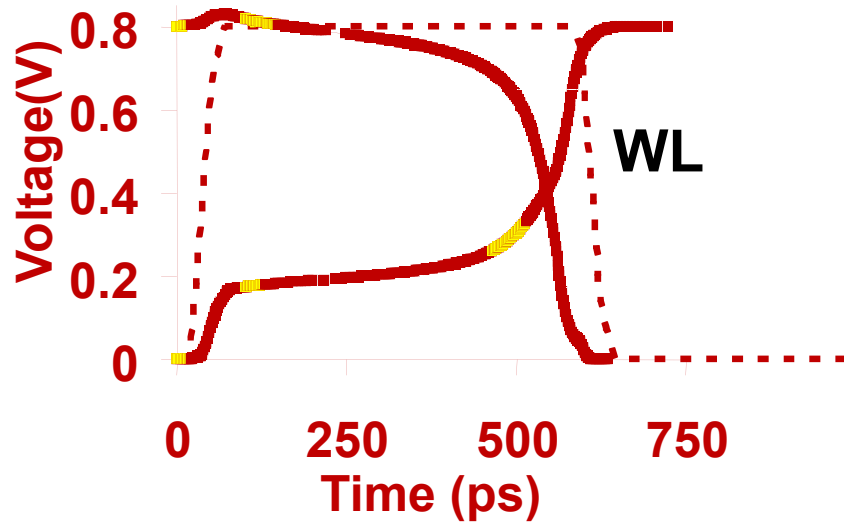
SRAM $V_{CC_{min}}$ Challenges



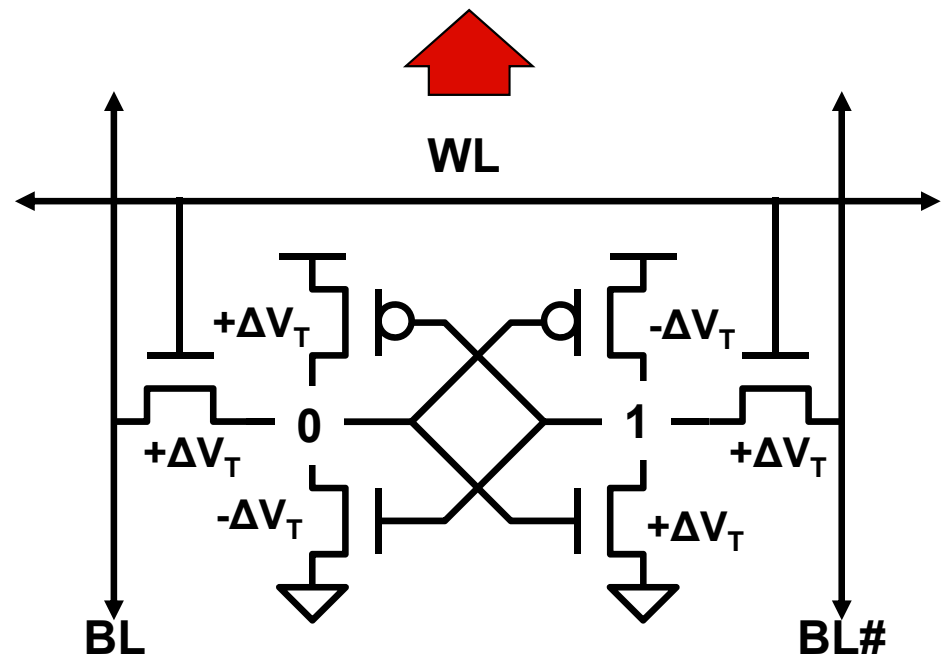
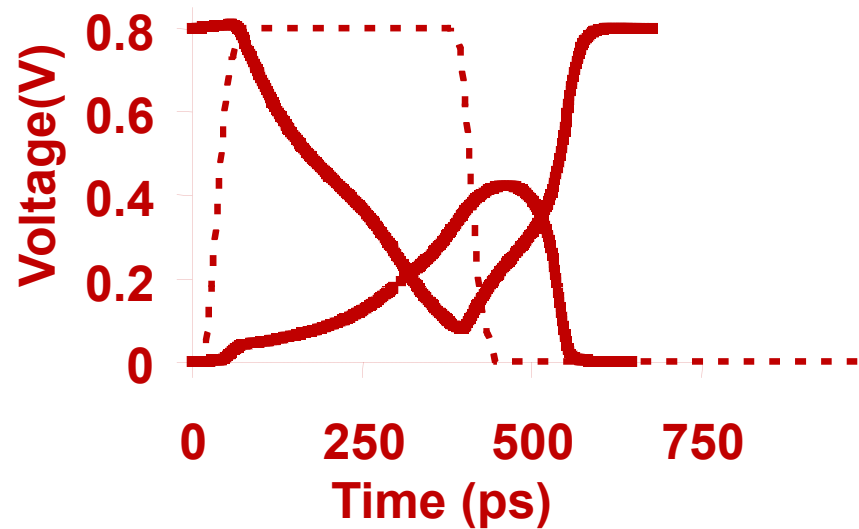
SRAM area scaling is getting harder because of process variations and voltage scaling!

SRAM Failure Mechanisms

Read Failure



Write Failure



Process Variation Trend

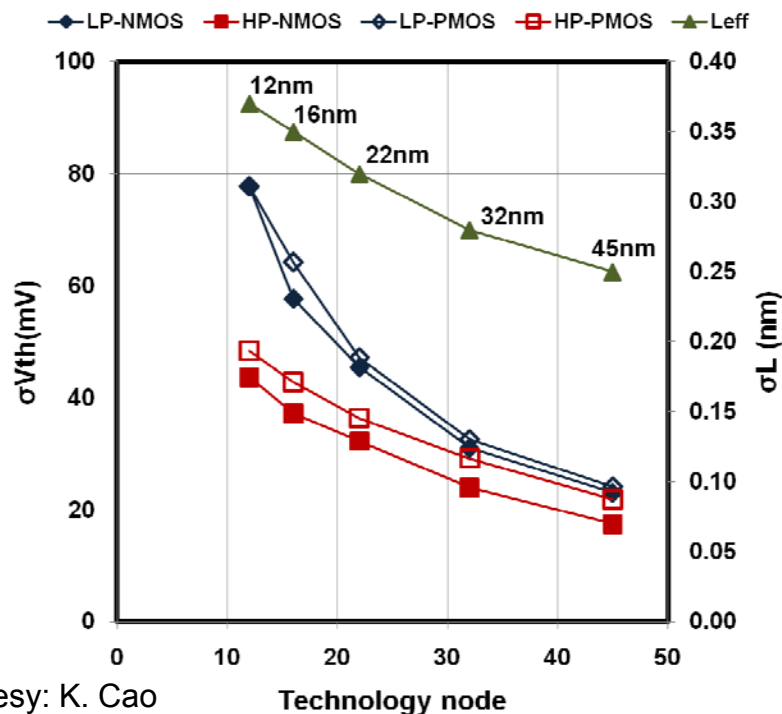
- Threshold variation:

$$\sigma V_{th0} = \frac{C_1}{\sqrt{WL}} \quad \text{where } C_1 = \sqrt{\frac{q}{3\epsilon_{ox}}} \cdot \sqrt{T_{oxe}(V_{TH} - V_{FB} - 2\phi_B)}$$

- Gate length variation:

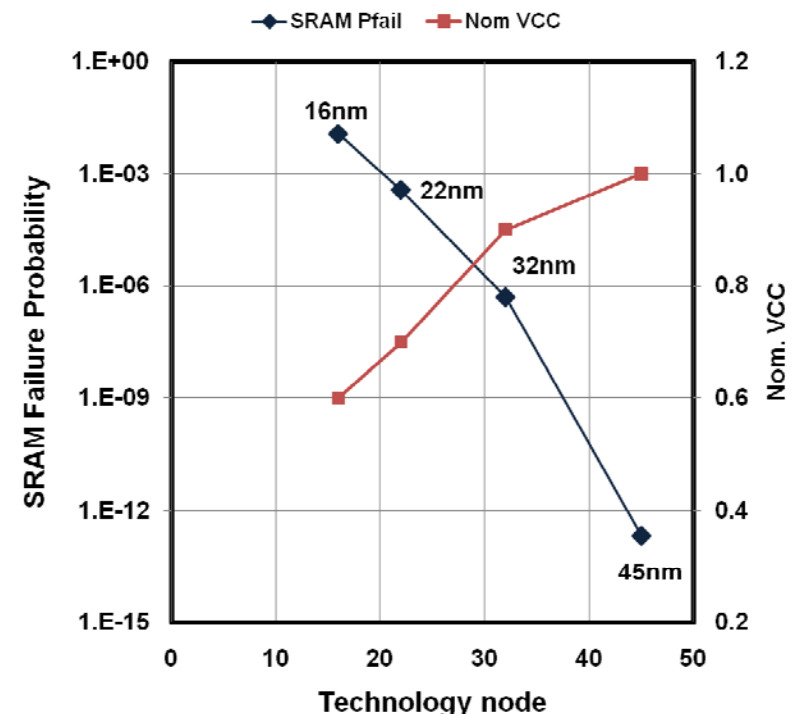
$$\sigma L = \frac{\sigma L_{LER}}{\sqrt{1 + \frac{W}{W_c}}} \quad \text{where } \sigma L_{LER}=0.5 \text{ and } W_c=15\text{nm}$$

ITRS Projection for Vth and Leff Variations



Courtesy: K. Cao

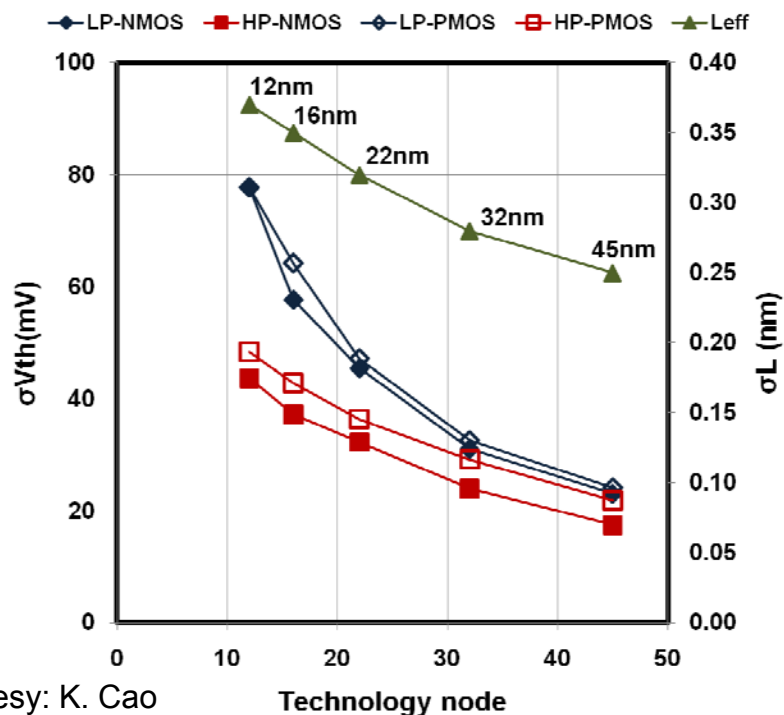
Corresponding SRAM Failure Probability vs VCC



Process Variation Trend

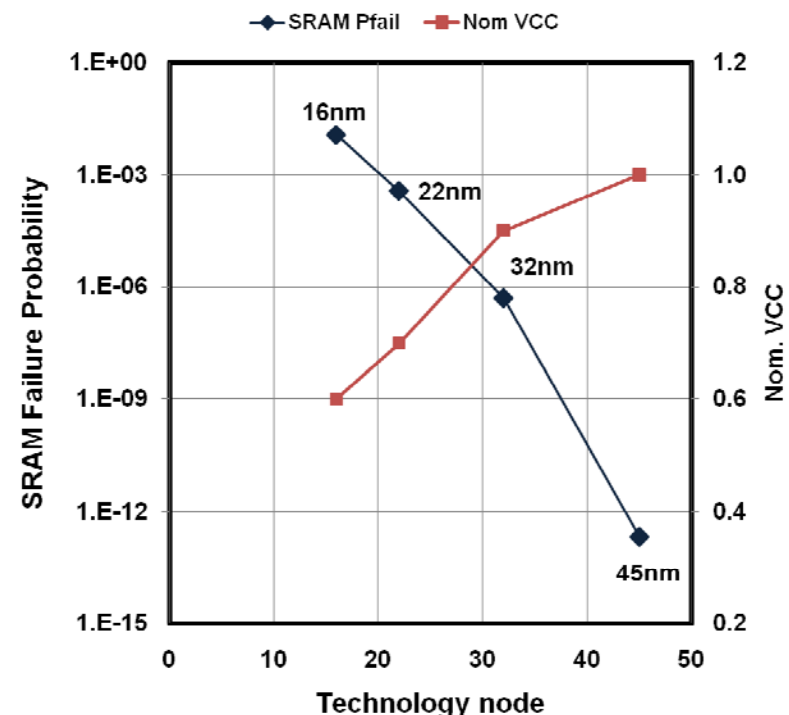
Increasing random variations & decreasing VCC w/ technology scaling begin to limit SRAM size & VCC scaling!

ITRS Projection for Vth and Leff Variations



Courtesy: K. Cao

Corresponding SRAM Failure Probability vs VCC



Circuit Solution

- Dynamic/adaptive techniques 6T SRAM

- ⊙ Dual supply column-based technique¹
- ⊙ Assisted read/write techniques^{2,3,4}

1. K. Zhang et al. A 3-GHz 70-Mb SRAM in 65-nm CMOS Technology With Integrated Column-Based Dynamic Power Supply. IEEE J. Solid-State Circuits vol 41 no 1, pp 146–151, 2006.
2. M. Khellah, N. Kim, et al. PVT-Variations and Supply-Noise Tolerant 45nm Dense Cache Arrays with Diffusion-Notch-Free (DNF) 6T SRAM Cells and Dynamic Multi-Vcc Circuits. In Proc. IEEE VLSI Circuit Symposium, Jun 2008.
3. F. Hamzaoglu, K. Zhang, et al. A 153Mb-SRAM Design with Dynamic Stability Enhancement and Leakage Reduction in 45nm High-κ Metal-Gate CMOS Technology. ISSCC 2008.
4. S. Ohbayashi. A 65-nm SoC Embedded 6T-SRAM Designed for manufacturability With Read and Write Operation Stabilizing Circuits. IEEE J. Solid-State Circuits vol 42 no 4, pp 820–829, 2007.

- SRAM cell sizing +ECCs

- ⊙ 6T SRAM cell area vs. failure rate trade-off
 - Carefully sized 6T SRAM cells for large caches have been more area efficient than 8T¹ and 10T^{2,3} SRAMs at the same VCC_{min}
- ⊙ Stronger ECCs allow us to continue VCC_{min} scaling (for now)

1. N. Verma, A. Chandrakasan. A 65nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy. ISSCC 2007.
2. B. Calhoun, A. Chandrakasan. A 256kb Sub-threshold SRAM in 65nm CMOS. ISSCC 2006.
3. I. Chang, J. Kim, K. Roy. A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS. ISSCC 2008.
4. Z. Chishti, et al. Improving Cache Lifetime Reliability at Ultra-low Voltages. MICRO 2009.

Circuit Solution

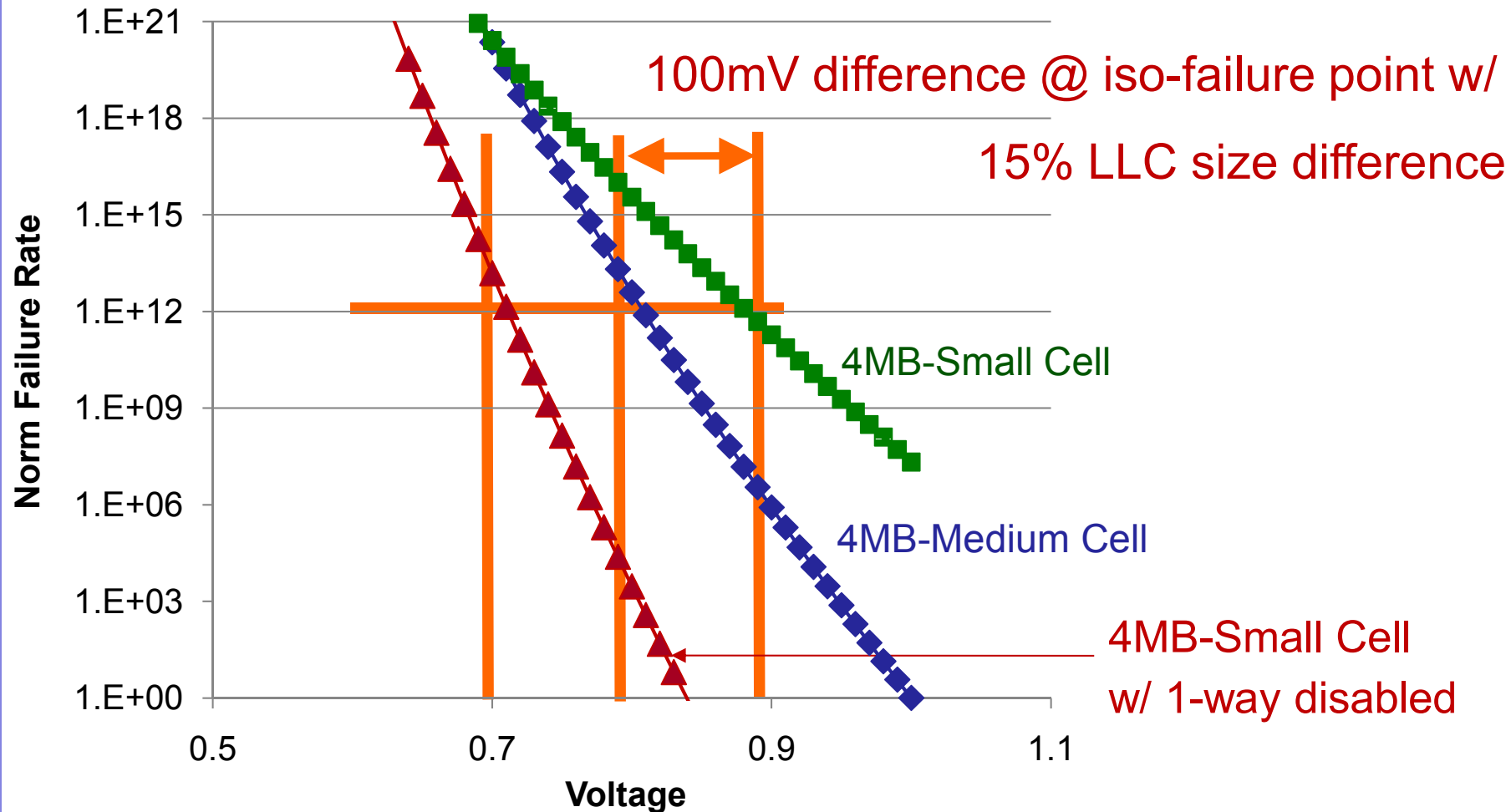
- Dynamic/adaptive techniques 6T SRAM
 - ⊙ Dual supply column-based technique¹
 - ⊙ Assisted read/write techniques^{2,3,4}

**Order-of-magnitude failure rate reduction w/
conventional 6T SRAM + small overhead!**

- SRAM cell sizing +ECCs
 - ⊙ 6T SRAM cell area vs. failure rate trade-off
 - Carefully sized 6T SRAM cells for large caches have been more area efficient than 8T¹ and 10T^{2,3} SRAMs at the same VCC_{min}
 - ⊙ Stronger ECCs allow us to continue VCC_{min} scaling (for now)

**Can we continue the current trend w/ 6T SRAM?
Probably not.**

Architecture Solution



- Small cell is 15% smaller, but 100mV higher VCC_{min} than medium one
- Allowing failure in any one LLC way in each set w/ small cell give 100mV lower VCC_{min} while 15% smaller overall cache area.

Dynamic Cache Resizing

- Designing a large cache operating at both high and low voltages is very challenging
 - ⊙ Lower operating voltage requires a larger area per bit
- Can we design a configurable cache?
 - ⊙ Allow as **big cache size** as possible **when performance is important**
 - ⊙ Allow as **low voltage** as possible at the expense of cache capacity **when power is important**
- At lower voltages and frequencies
 - ⊙ Processor performance is less sensitive to on-chip cache size due to reduced frequency gap b/w main memory and on-chip cache

Reduce cache size to lower VCC_{min} at lower freq since performance impact is very small!

Conclusion

- VCC/Freq scaling for power efficient computing
 - ⊙ Require SRAMs to operate at low VCC
- Increasing random variations & decreasing VCC w/ technology scaling
 - ⊙ Begin to limit SRAM size scaling!
- Various adaptive/dynamic + sizing + ECC techniques
 - ⊙ Have reduced the SRAM failure rate by order-of-magnitude failure rate w/ conventional 6T SRAM + small overhead.
 - So far, 6T SRAM has been more area efficient than 8T and 10T SRAM for large cache structures
- Incorporating architecture techniques
 - ⊙ Lower VCC_{min} by trading cache capacity w/ lower VCC_{min}
 - The performance impact is very small due to reduced frequency gap b/w main memory and on-chip caches