A Software Technique to Improve Yield of Processor Chips in Presence of Ultra-Leaky SRAM Cells Caused by Process Variation

Maziar Goudarzi, Tohru Ishihara, Hiroto Yasuura

System LSI Research Center
Kyushu University, Fukuoka, Japan

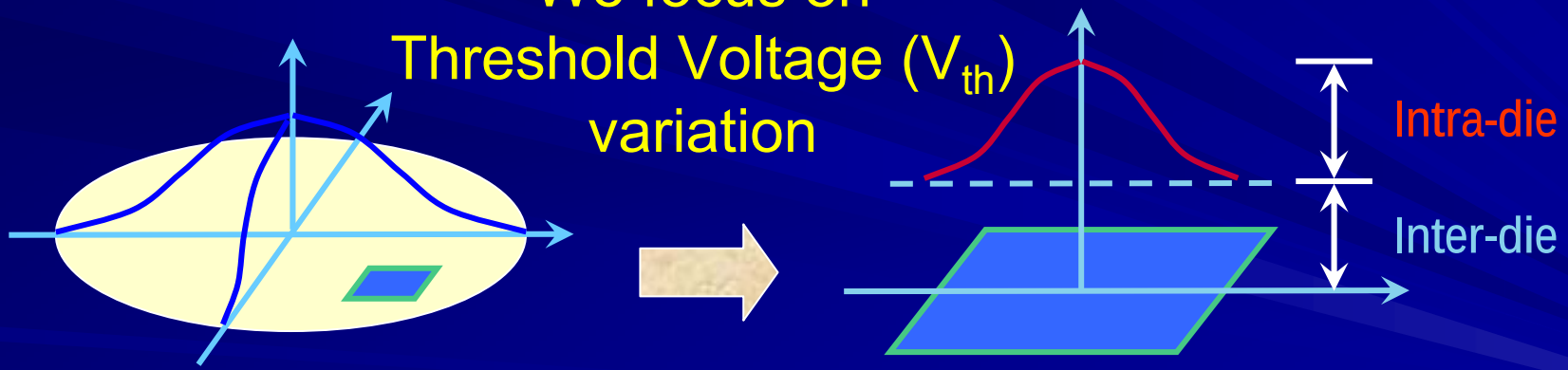
Outline

- Background
  - Process variation
  - Power in nanometer embedded processor-based systems
- Our work[1]
  - Motivation
  - Approach
  - Experiments
- Summary and Future work

[1] This is part of the CREST "Ultra Low Power Design Projects" sponsored by Japan Science and Technology Corporation (JST), http://www.slrc.kyushu-u.ac.jp/~ishihara/CREST/e_kenkyu.html

Background: Process Variation

We focus on Threshold Voltage (V_{th}) variation



Both inter-die and intra-die variations become increasingly important!

* Source: X. Li, J. Le, L. Pileggi, "Projection-Based Statistical Analysis of Full-Chip Leakage Power with Non-Log-Normal Distributions," DAC, 2006.

Our Focus: Intra-die (Within-die) V_{th} Variation

Large Intra-Die Variation

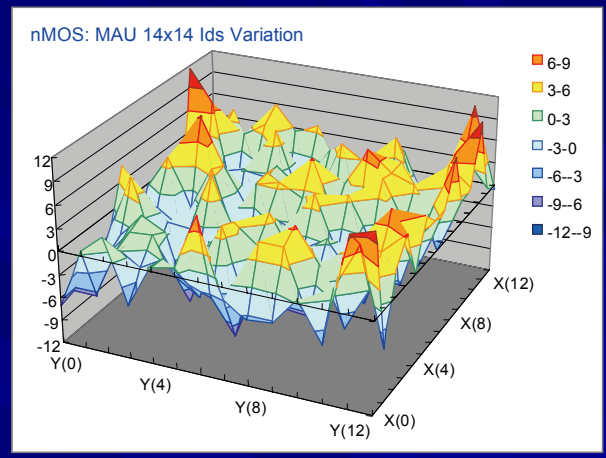
Current 3-sigma = 13%

V_{th} 3-sigma = 67mV

Variation is huge in small transistors

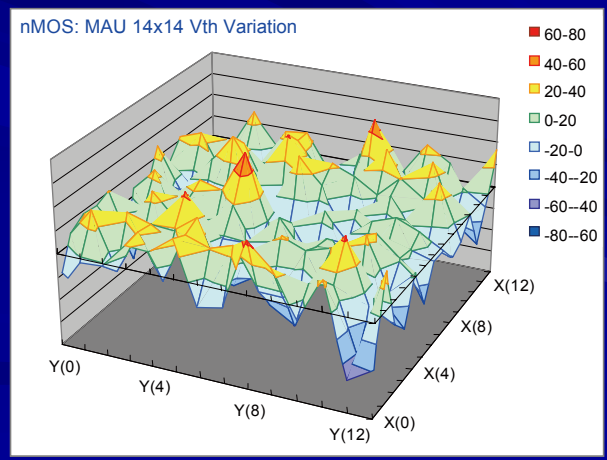
$$\sigma_{Vth} = \frac{q}{C_{ox}} \sqrt{\frac{N_a \cdot W_{dm}}{3 \cdot L \cdot W}}$$

L, W: Effective channel length and width
q: electron charge
C_{ox}: oxide capacitance
N_a: substrate doping concentration
W_{dm}: maximum depletion width



L = 0.1um
W= 0.4um

Av. = 203.7uA
Sigma = 4.4%
min. = -11.4%
max. = 11.4%

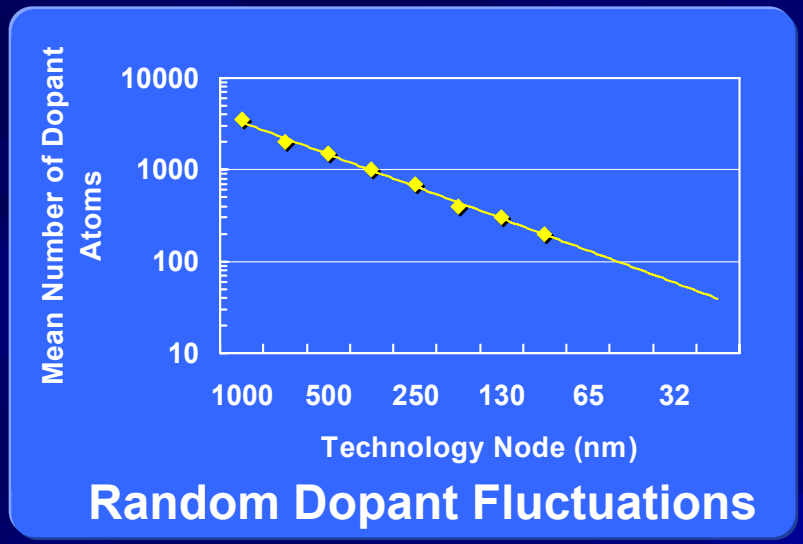


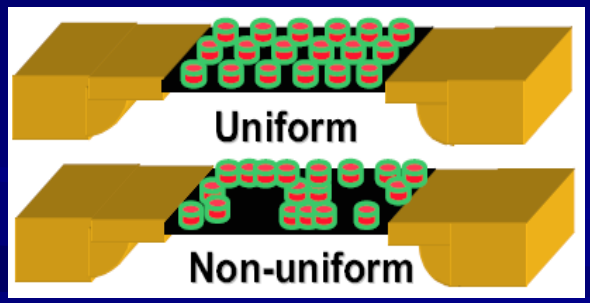
L = 0.1um
W= 0.4um

Av. = 308.3uA
Sigma = 22.1mV
min. = -66.6mV
max. = 57.0mV

Eijiro Toyoda, "DFM: Device & Circuit Design Challenges", Int'l Forum on Semiconductor Technology, 2004

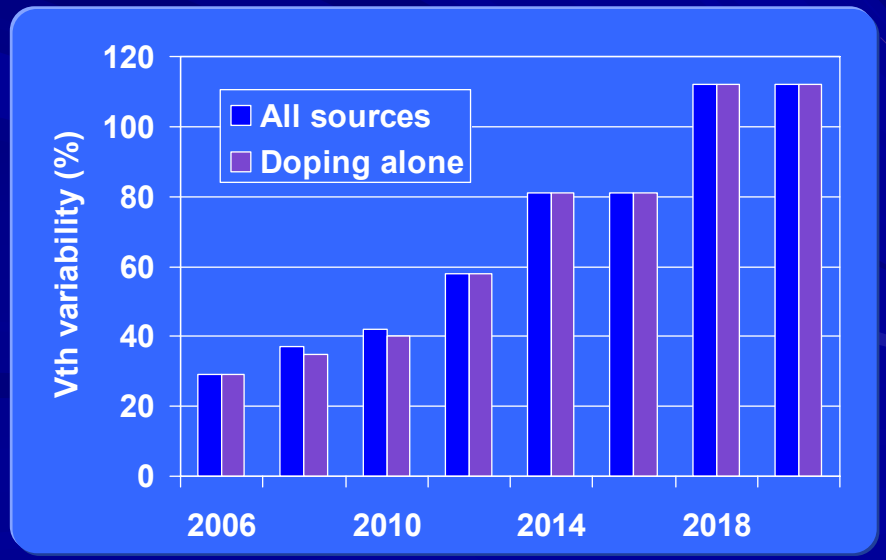
Unavoidable Cause of V_{th} Variation: Random Dopant Fluctuation (RDF)





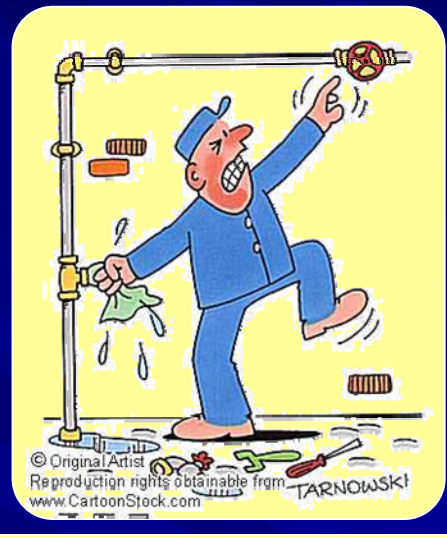
Source: S. Borkar

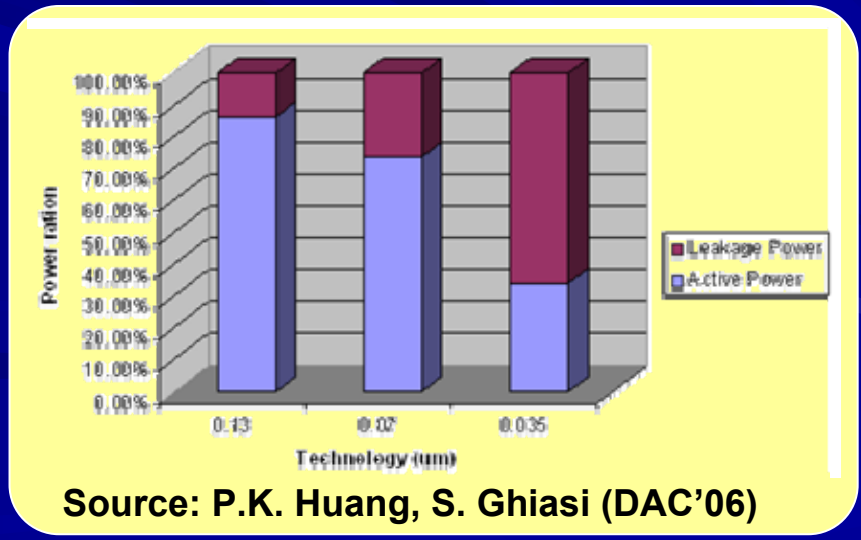
- Nature of variations
  - Systematic
  - Random
- ITRS-2005 roadmap forecast



Our Focus: Leakage Power

- Power consumption
  - Dynamic
    - activity-based
  - Static (leakage)
    - activity-independent
- Trend
  - Traditionally:
    - Dynamic >> Static
  - Nanometer technologies
    - Static >> Dynamic





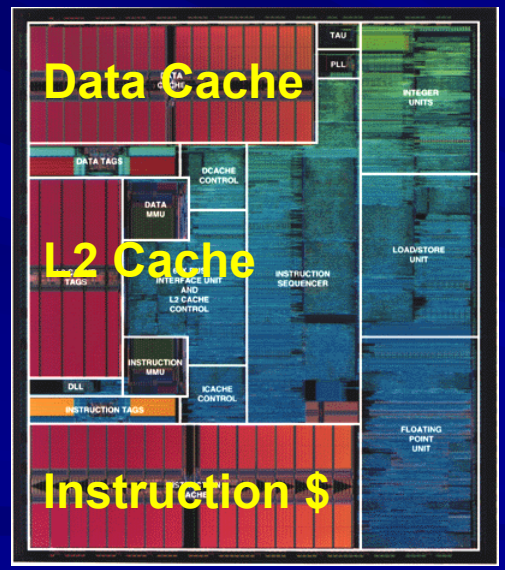
Source: P.K. Huang, S. Ghiasi (DAC'06)

Our Focus: Caches Memories

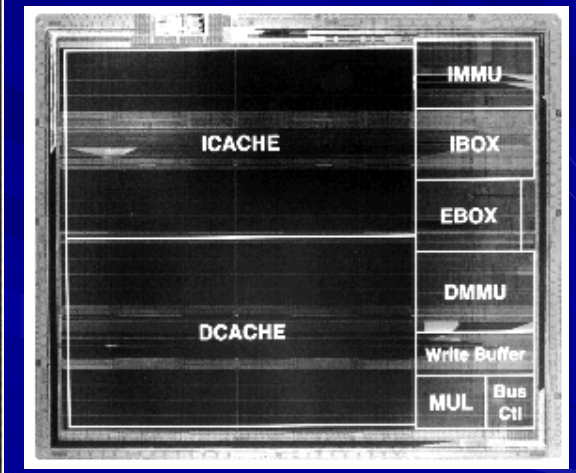
- Largest portion of chips
  =>
  biggest leakage
- Minimum-area transistors
  =>
  most susceptible to process variation

$$\sigma_{Vth} = \frac{q}{C_{ox}} \sqrt{\frac{N_a \cdot W_{dm}}{3 \cdot L \cdot W}}$$

PowerPC™
40% of core area



StrongARM-110™
75% of core area



Process Variation at 90nm

$$I_{Subthreshold} \propto \frac{W \cdot V_T^2}{T_{ox} \cdot L} \cdot \exp\left(\frac{-V_{th}}{\alpha \cdot V_T}\right)$$

V_T: Thermal voltage (25mV@room temperature)

α: Sub-threshold factor (1.40~1.65)

T_{ox}: Oxide thickness

Year	min. L [nm]	$^1V_{TH}$ [V]	$^2V_{TH}$ [V]
2004	37 (90)	0.32	0.12
2005	32 (80)	0.33	0.09
2006	28 (70)	0.34	0.06

1: Low Operating Power Process 2: MPU process

Ultra-Leaky Transistor (ULT): Transistors that leak beyond a given constraint

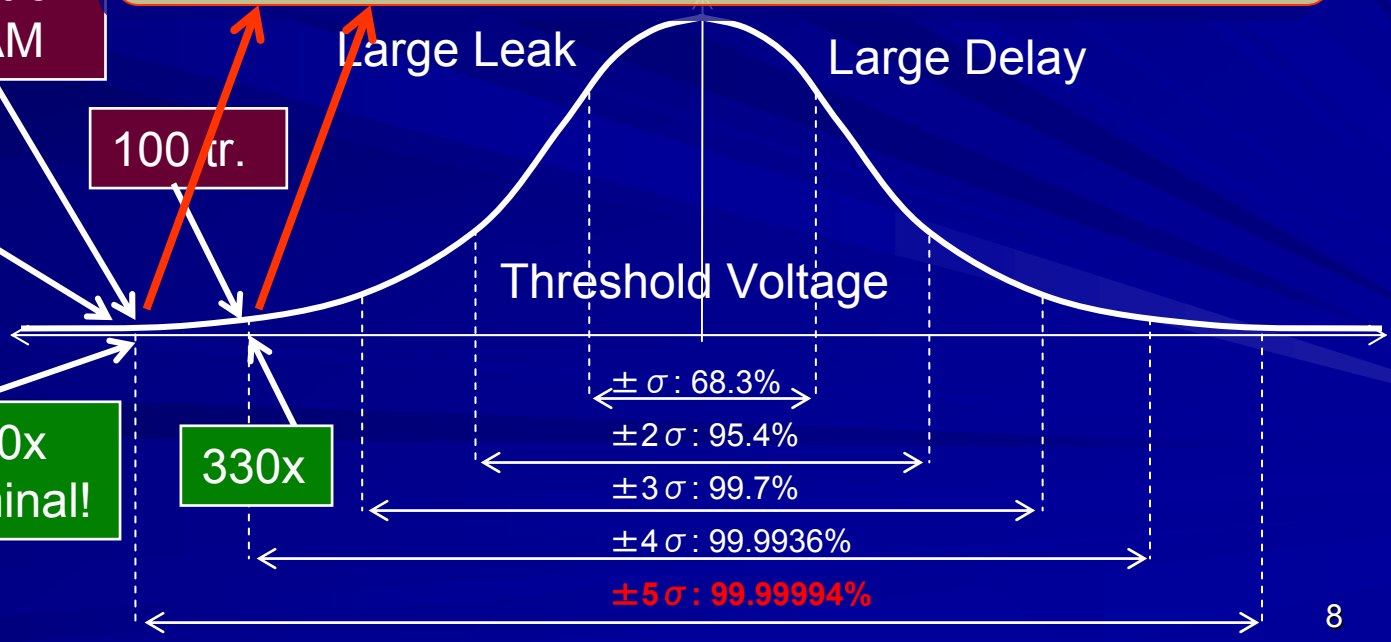
1 transistor out of 64K-Byte SRAM

100 tr.

$5\sigma_{Vth}$=0.3V

Leakage is 1,400x higher than nominal!

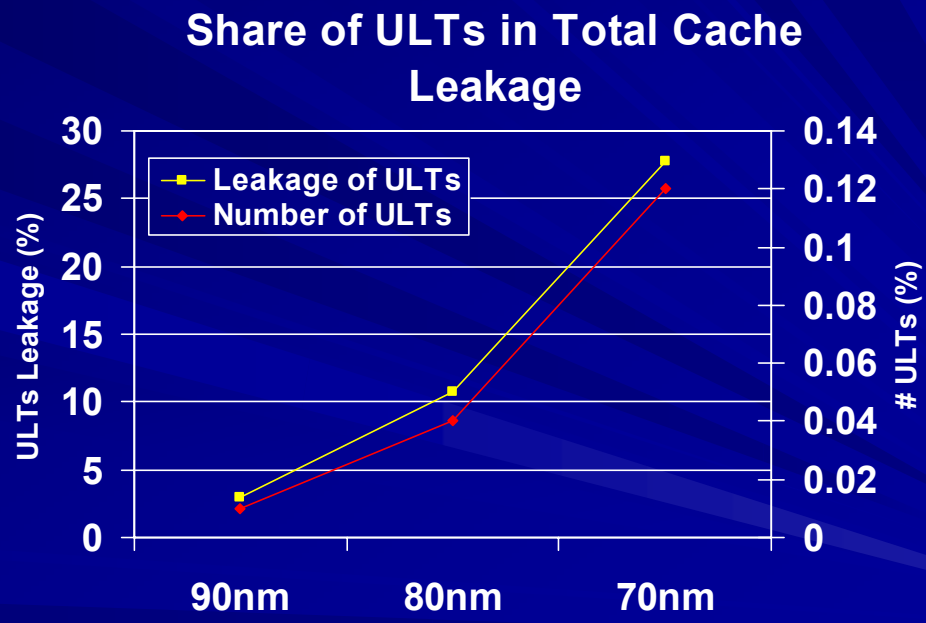
330x



Ultra-Leaky SRAM Cells Problem

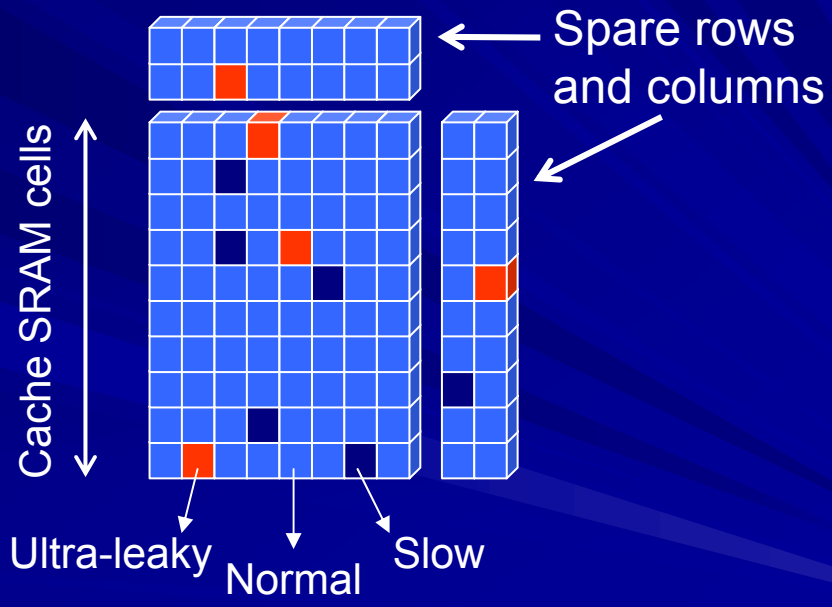
Ultra-Leaky Cache Cells and Ultra-leaky Cache Lines:
Those containing one or more ULT

- Problem
  - Ultra-leaky cache cells dissipate lots of power
  - Especially for long-standby applications, cause rapid discharge of battery

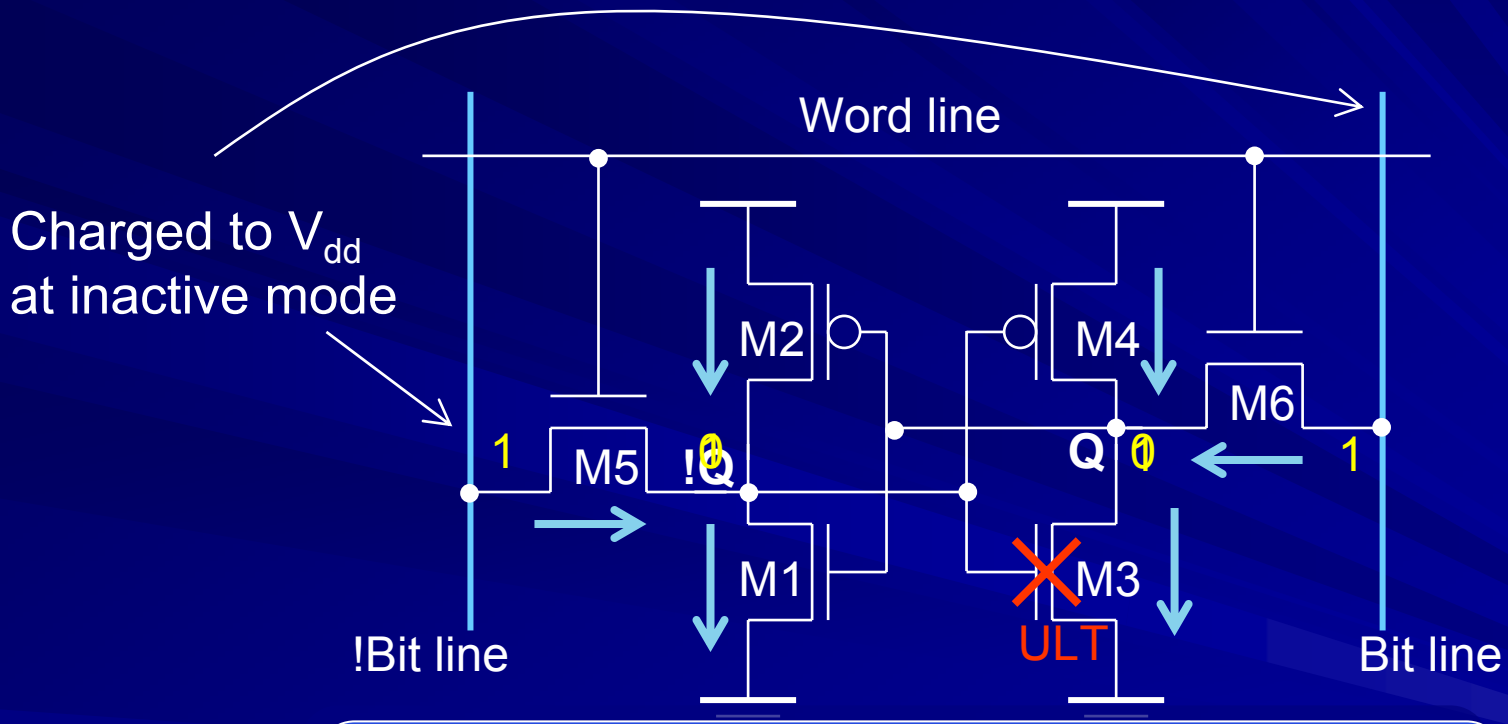


Ultra-Leaky SRAM Cells Problem (cont'd)

- Naïve solution
  - Mark as faulty, replace with spare row/column
  - Disadvantages
    - Spares may be leaky themselves
    - Spares should replace slow/faulty cells as well
    - Fuse-blowing expensive and slow
    - Aging may introduce ULTs over time
    - Temperature may also introduce ULTs



Our Fundamental Observation: Cell Leakage is Value-Dependant

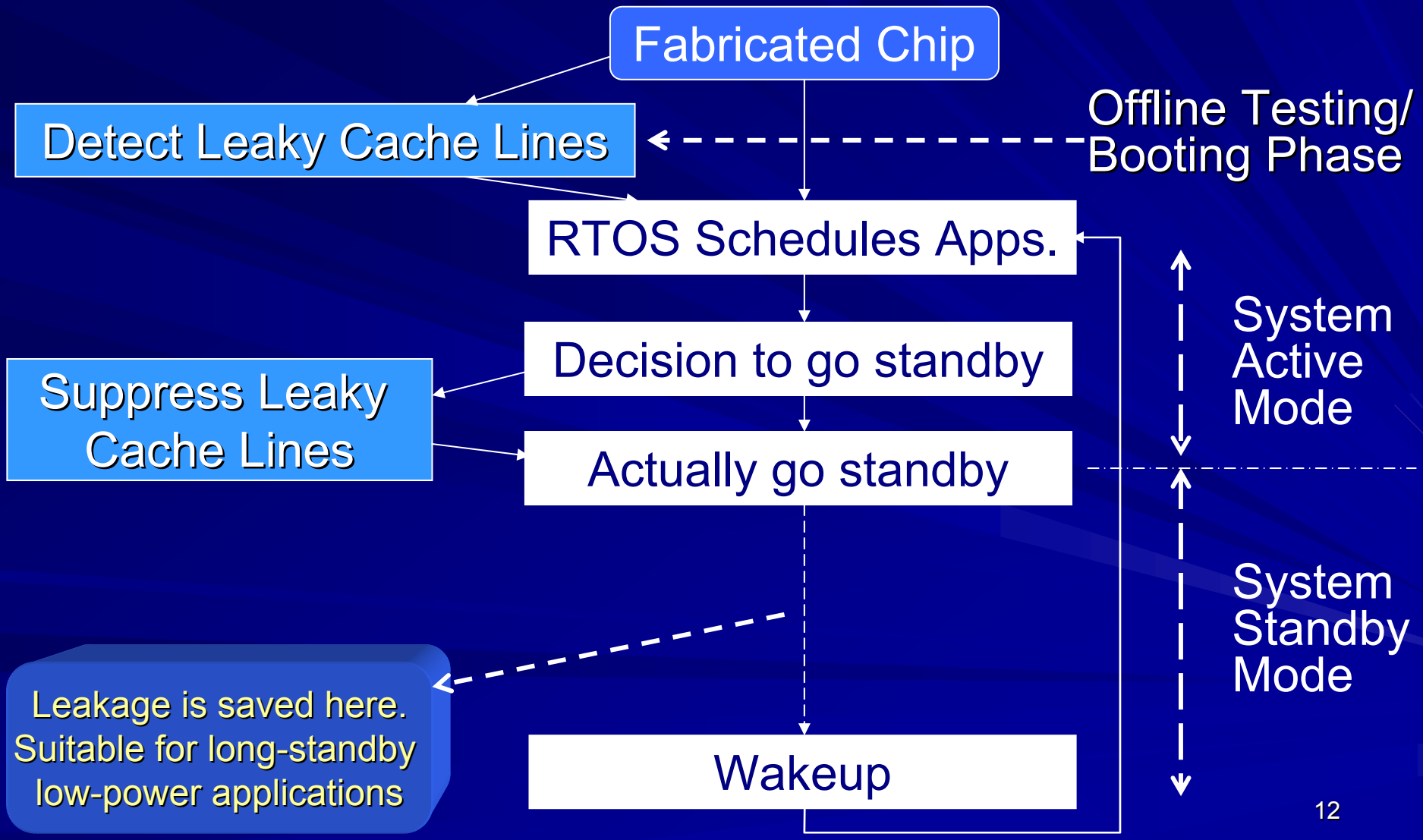


If M2, M3, or M5 is leaky, the SRAM cell is 1-leaky

If M1, M4, or M6 is leaky, the SRAM cell is 0-leaky

Our Approach:
Store the Leakage-Safe Value
when entering standby mode

Flow of Operations

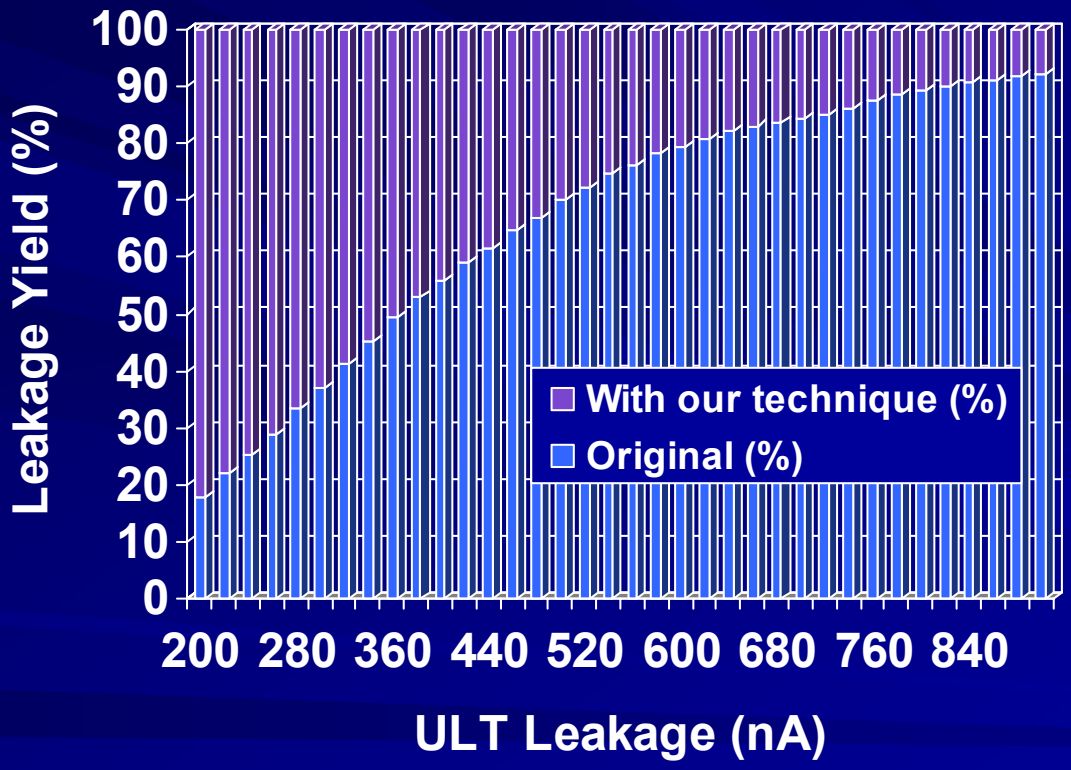


Offline Testing Phase

- Goal:
  - Detect location of ULTs
  - Location accuracy: cache line or cache cell
- Idea
  - ΔI_{DDQ} Testing:
    - If the leaky cell is sensitized, the quiescent current reflects an abnormal change.
- General outline
  - Write all 0's, then all 1's to every cache line and measure the leakage current

Improvement in Leakage Yield

Leakage Yield = % of chips meeting a given leakage constraint

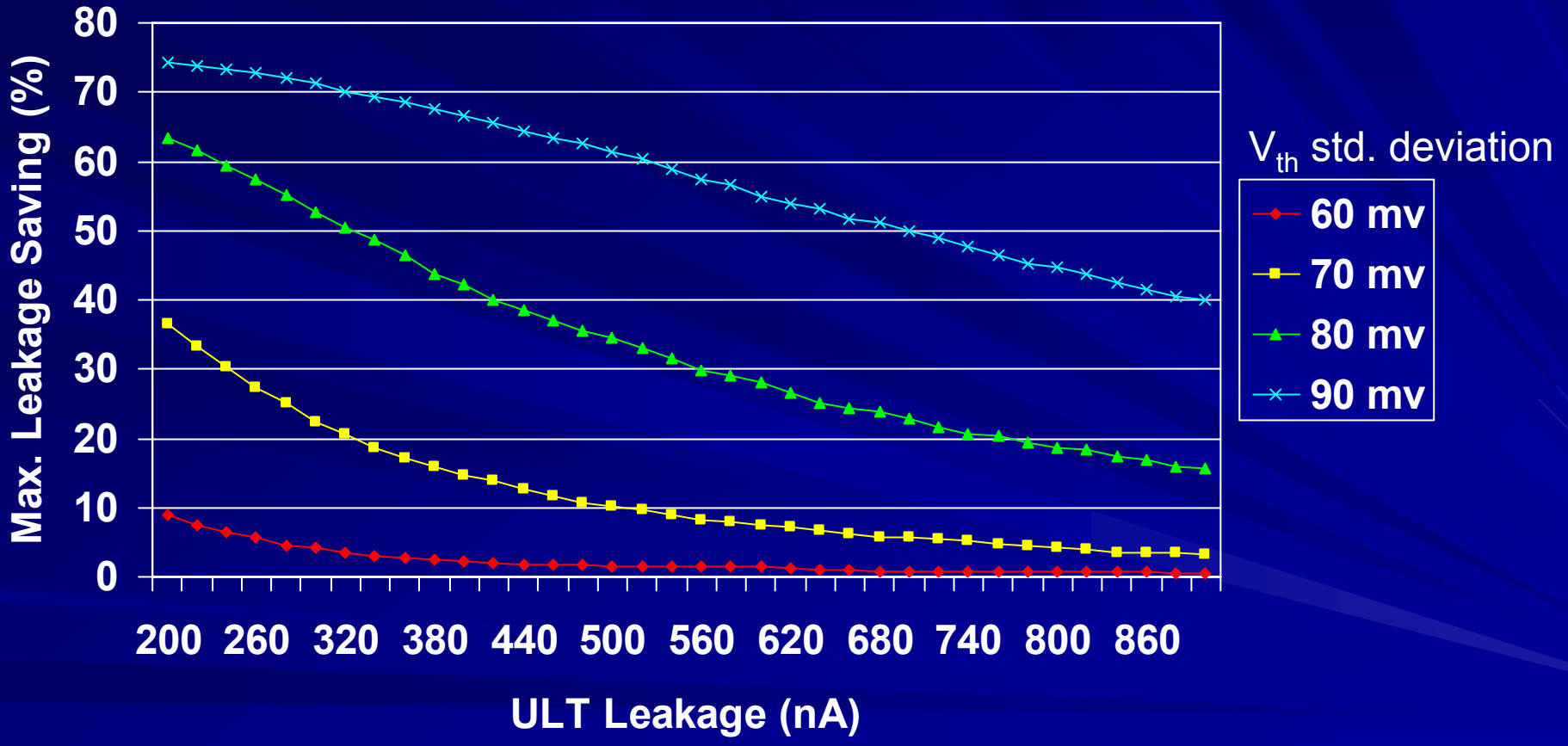


Experiments:
- Monte Carlo simulation
- 1000 chips
- 32 Kb data + 22 Kb tag
- 60mv within-die V_{th} variation
- Nominal values from a 90nm process

V_{th}=320mv

Nominal transistor leakage =0.345 nA

Maximum Leakage Power Saving vs. Within-die Variation



Nominal transistor leakage =0.345 nA

Associated Costs

Costs	Why to pay	When to pay
Power	Run instructions to store leakage-safe values in leaky cache lines	When going to standby mode
Performance	Invalidated, but later-referenced, cache contents	After returning from standby mode
Area	Leakage-measurement on-chip circuitry	Chip design & manufacturing

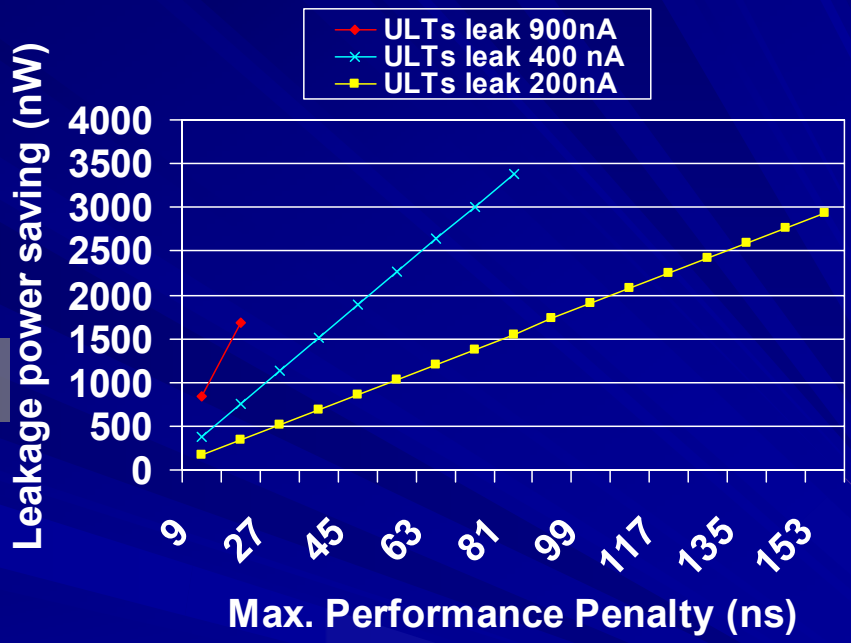
Analysis of Costs

- Energy benefit & Performance cost linearly depend on the number of leaky cells cured (*N*)

$$EnergySaving(t) = N \times (P_{leak} \times t - E_{lock} - E_{fetch})$$

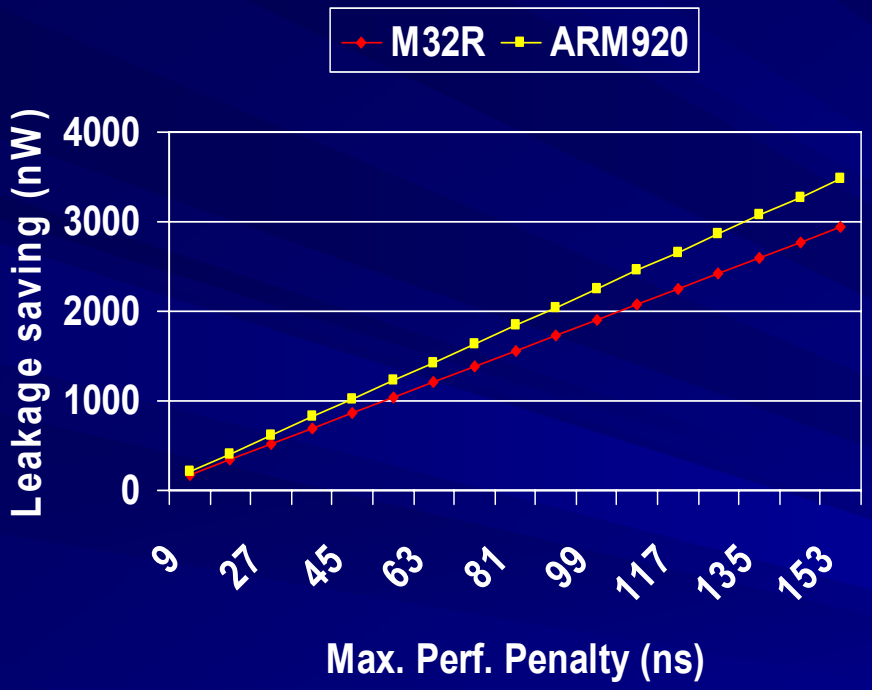
$$Perf.Penalty \leq N \times (T_M - T_c)$$

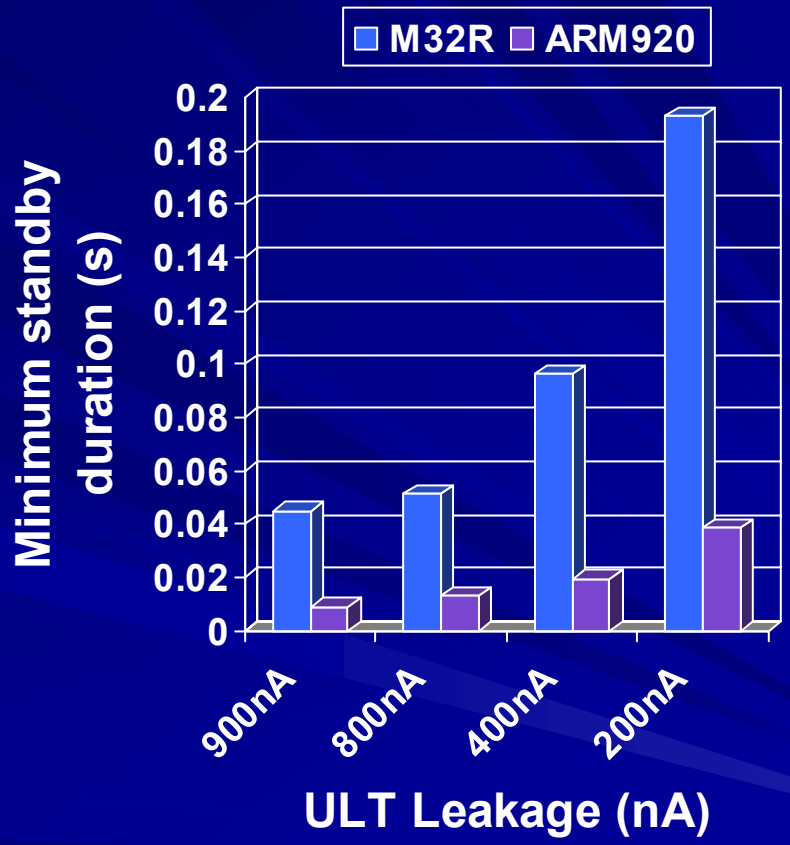
N: Number of leaky cells cured
t: Time duration spent in standby
P_{leak}: Avg. power saved per cured cache line
E_{lock}: Energy for locking leakage-safe value in the cache
E_{fetch}: Energy for fetching invalidated data if needed
T_M: Memory access time
T_c: Cache access time



Results for M32R processor:
0.18u process, 200mW @ 50MHz
Memory latency: 10 ns
Cache latency: 1 ns

Effect of the Processor Used





M32R: 0.18u, 200mW @ 50 MHz

ARM920: 0.18u, 0.8mW / MHz [1]

[1] http://www.arm.com

Summary & Future Work

Thanks! + Q&A

- Presented a *software* technique to suppress, during standby mode, leakage of ultra-leaky transistors
  - No major hardware/circuit change required
  - Only uses already-popular cache-control instructions
  - Useful even for dynamic effects such as aging and temperature
- Results
  - Reduced leakage power in standby mode
  - Salvage chips containing ULTs => higher yield for long-standby low-power applications
- Future work
  - Reduce leakage power, even in active mode, by matching cache contents with the less-leaky state of cache cells