Sub-Threshold Design: The Challenges of Minimizing Circuit Energy

by

Ben Calhoun[1], Alice Wang[2], Naveen Verma[3], and Anantha Chandrakasan[3]

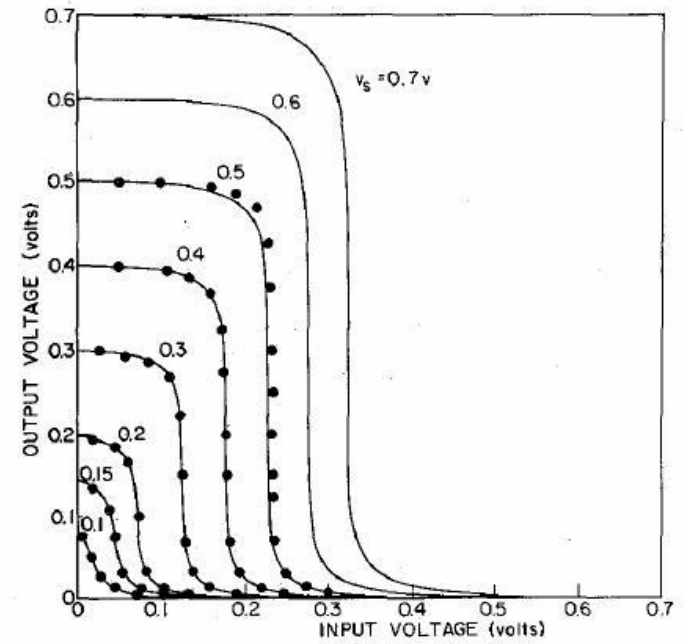
[1]University of Virginia
[2]Texas Instruments
[3]Massachusetts Institute of Technology

Outline

- Background
- Sub-threshold FFT Processor
- Ultra-Dynamic Voltage Scaling (UDVS)
- Sub-threshold SRAM
- Conclusions

History of Sub-Threshold Digital Circuits

- 1972: Sub-threshold theorized for minimum V_{DD} operation (*Swanson & Meindl, JSSC*)
- 1972: Low voltage digital circuits for wristwatch (*Vittoz et al, JSSC*) operated into sub-V_T (*Vittoz, pers. comm.*)
- 1977: Analog sub-V_T (*Vittoz & Fellrath, JSSC*)

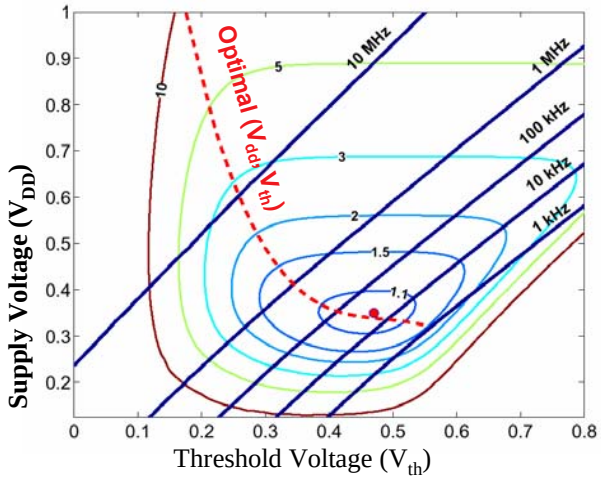


Swanson & Meindl, 1972

- 1970s-80s: Analog sub-V_T popular
- 1999: Sub-threshold digital again analyzed for low power (*Soeleman & Roy, ISLPED*)

History of Sub-Threshold Digital Circuits (2)

- 2001: Sub-V_T adders (*Paul, Soeleman, Roy, ESSCIRC*)
- 2002: Sub-V_T ring osc. (*Deen et al., ICCDCS*)
- 2002: Optimal energy in sub-V_T (*Wang, Chandrakasan, Kosonocky, SVLSI*)



- 2003: Sub-V_T DLMS filter (*Kim, Soeleman, Roy, TVLSIS)*
- 2004: Sub-V_T FFT Processor (*Wang & Chandrakasan, ISSCC)*
- 2004: Min. energy numerical model (*Zhai, Blaauw, Sylvester, Flautner, DAC)*
- 2004: Min. energy analytical model (*Calhoun & Chandrakasan, ISLPED*)
- …

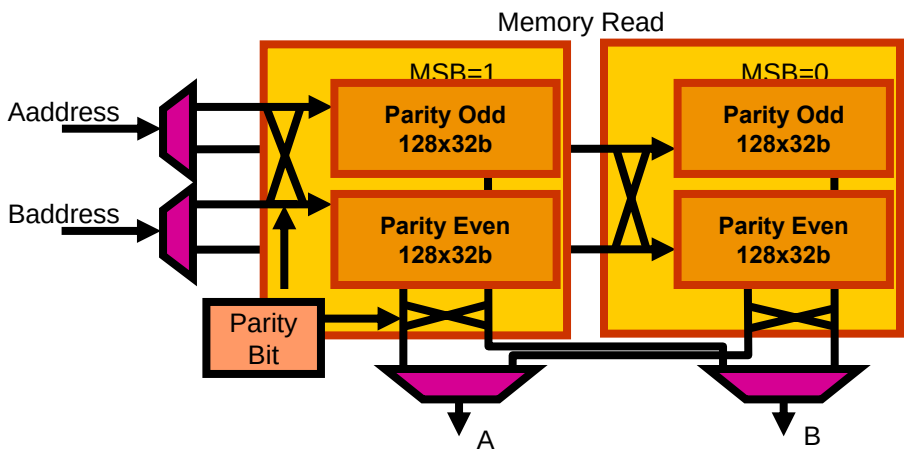
Applications for Sub-threshold

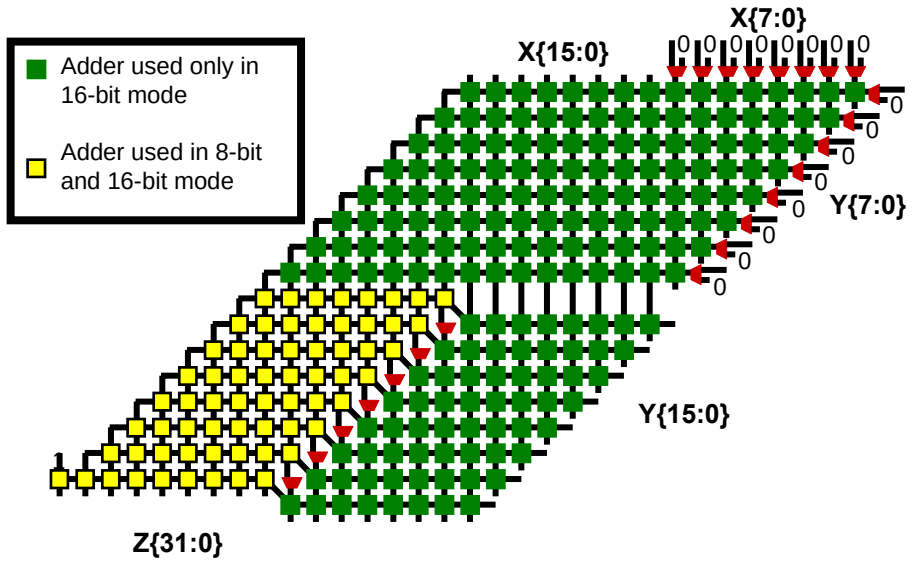
- Energy Constrained Applications
  - Wireless microsensors
  - RFID tags
  - Medical implants
- Burst-mode Applications
  - Used for low performance operations
  - Power management circuits
- Throughput-centric Applications
  - Large amounts of parallelism

Outline

- Background
- **Sub-threshold FFT Processor**
- Ultra-Dynamic Voltage Scaling (UDVS)
- Sub-threshold SRAM
- Conclusions

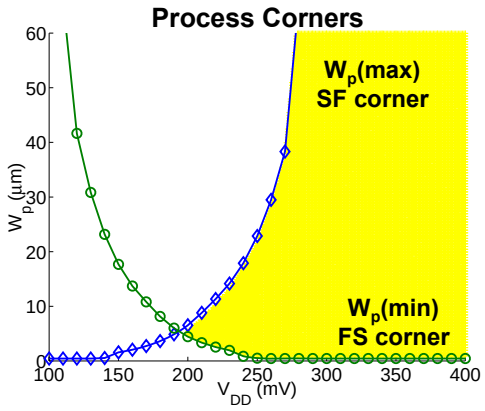
Energy-Scalable Architecture for FFT



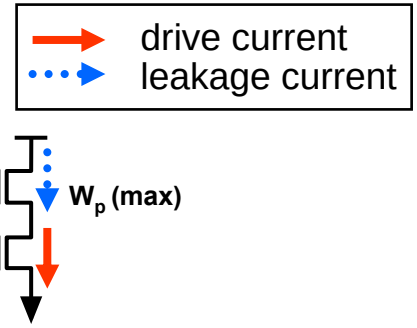


- Bit-precision scaling architectures are used in the butterfly datapath, data memory and Twiddle ROMs.
- Fine-grained gating reduces activity factor and achieves energy savings with minimal area overhead.

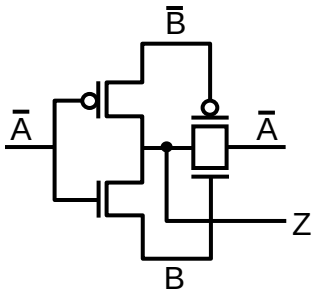
Circuit Practices for Sub-threshold



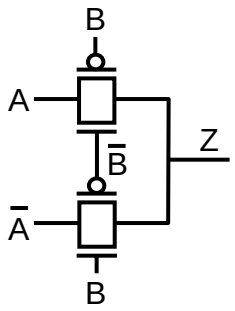
Inverter with a minimum sized W_n

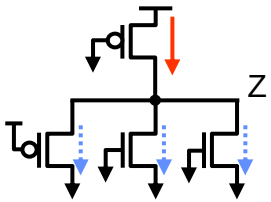


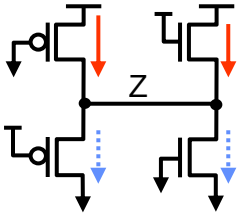
Tiny XOR



TX-gate XOR

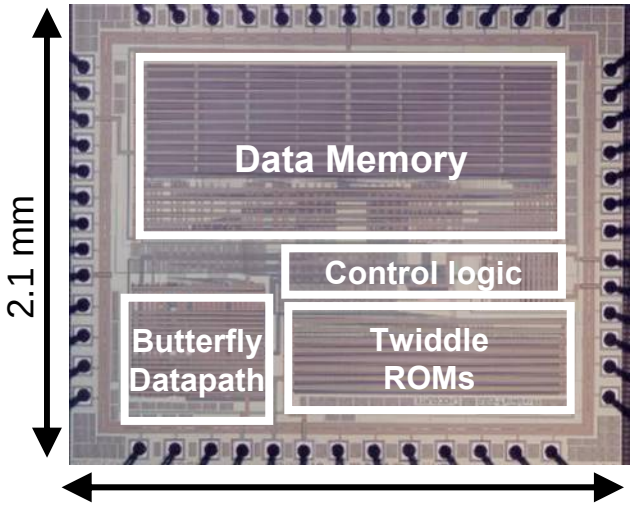


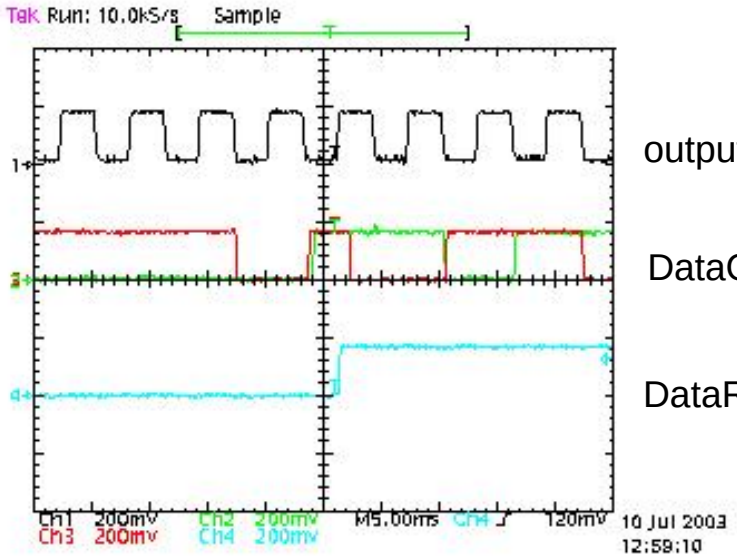




- Process variations affect P/N ratio → Sizing can compensate somewhat
- Parallel leakage harmful due to lower Ion/Ioff → Alternative circuits improve w-c Ion/Ioff

Custom Sub-threshold FFT





output clock

DataOutput[1-0]

DataReady

Design Flow

- Custom sub-threshold logic cells
- Custom Skill-based memory generators and multipliers
- Skill code place-and-route

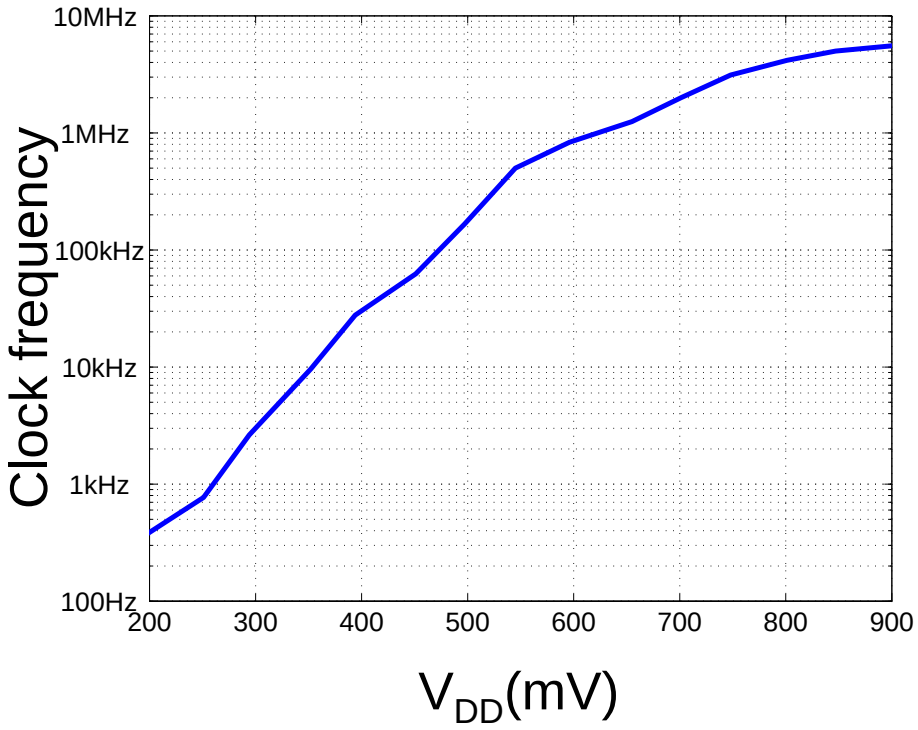
Process Details

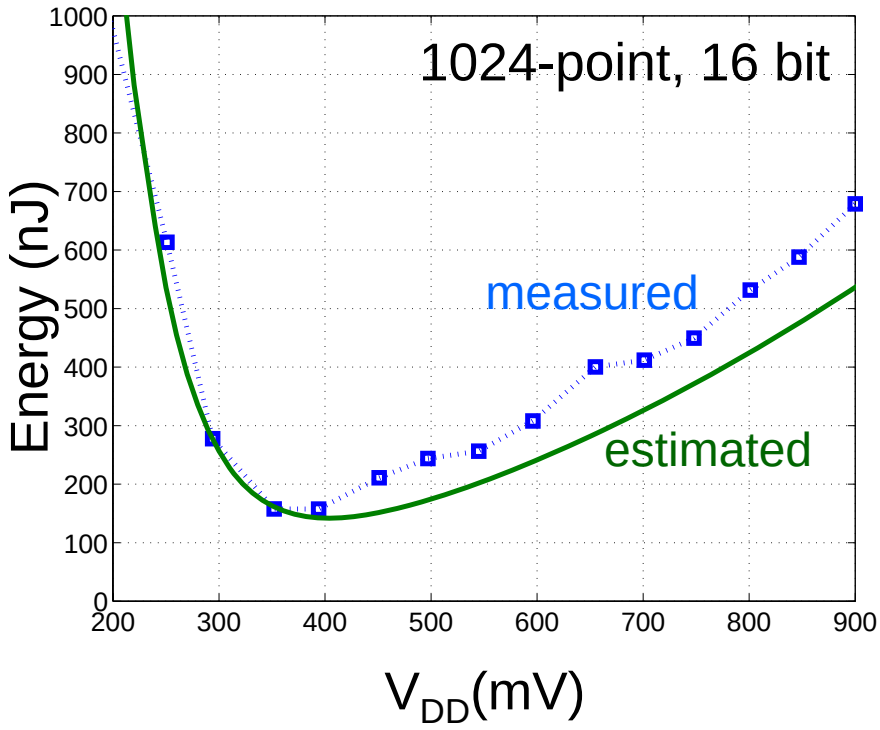
- 0.18μm CMOS process
- 6 layer metal
- 628k transistors

Results

- 180mV V_{DD} for 16-bit, 1024-point operation (164 Hz).

Minimum Energy Operation





- The FFT operates between V_{DD}=180mV-900mV and clock frequency of 164Hz-6MHz.
- The minimum energy dissipated is 155nJ/FFT at 350 mV for a 1024-point 16b FFT. The clock frequency is 10kHz and the FFT processor dissipates 0.6μW.

Outline

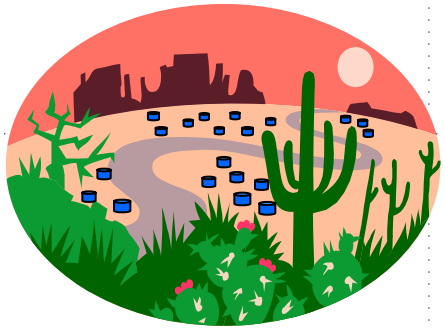
- Background
- Sub-threshold FFT Processor
- **Ultra-Dynamic Voltage Scaling (UDVS)**
- Sub-threshold SRAM
- Conclusions

Ultra-DVS Motivation

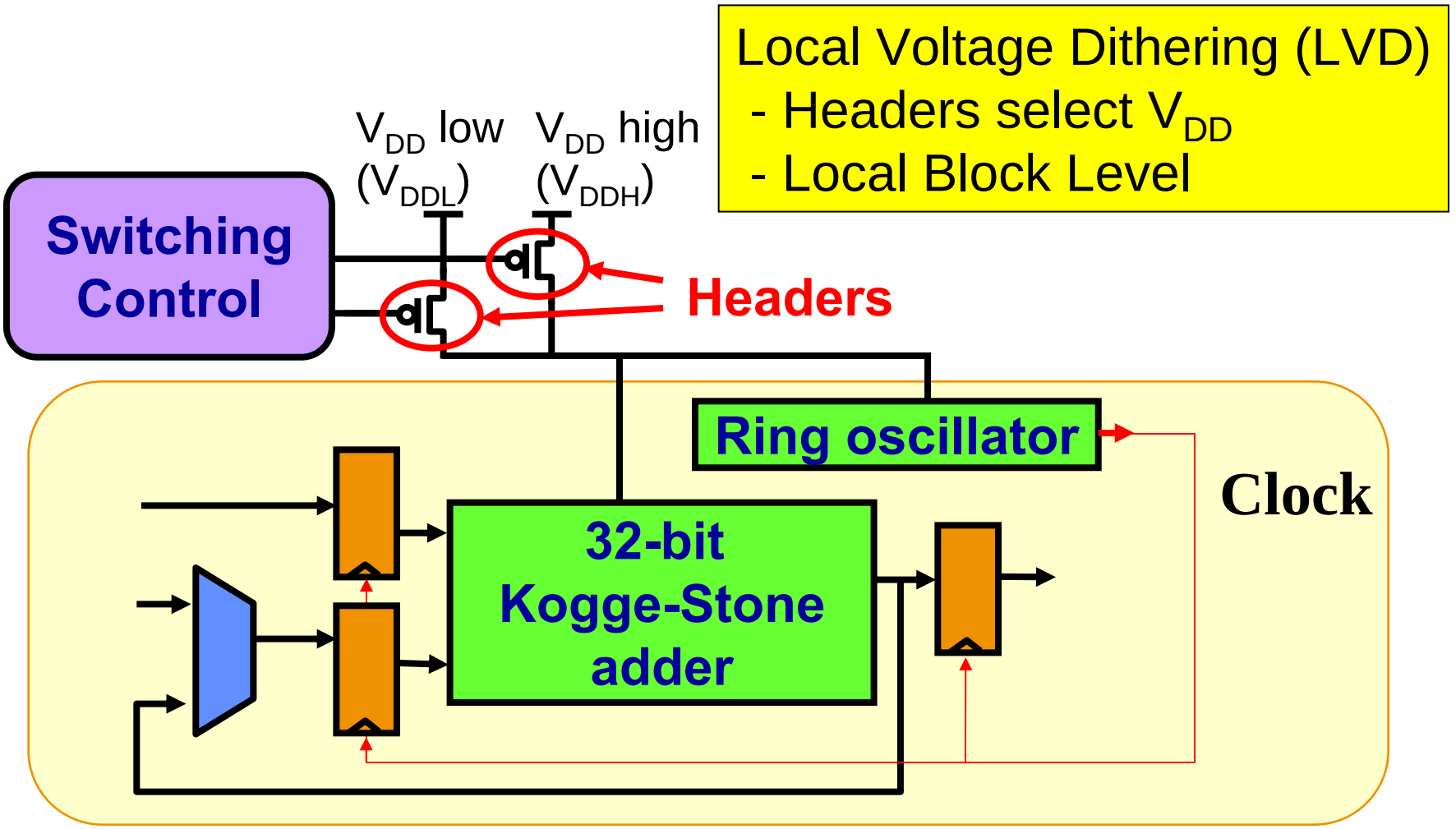
Strong Inversion Operation: fast, power-hungry

Unite the regions

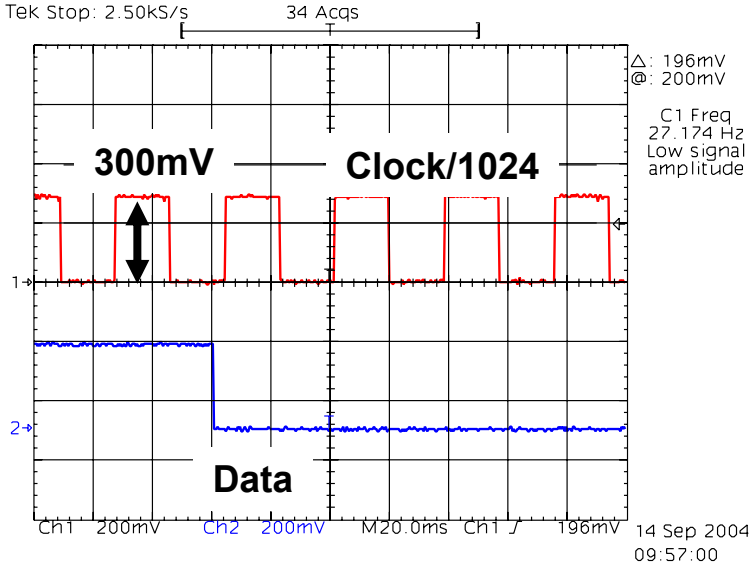
Sub-threshold operation: Slow, lower power, minimum energy operation becomes possible

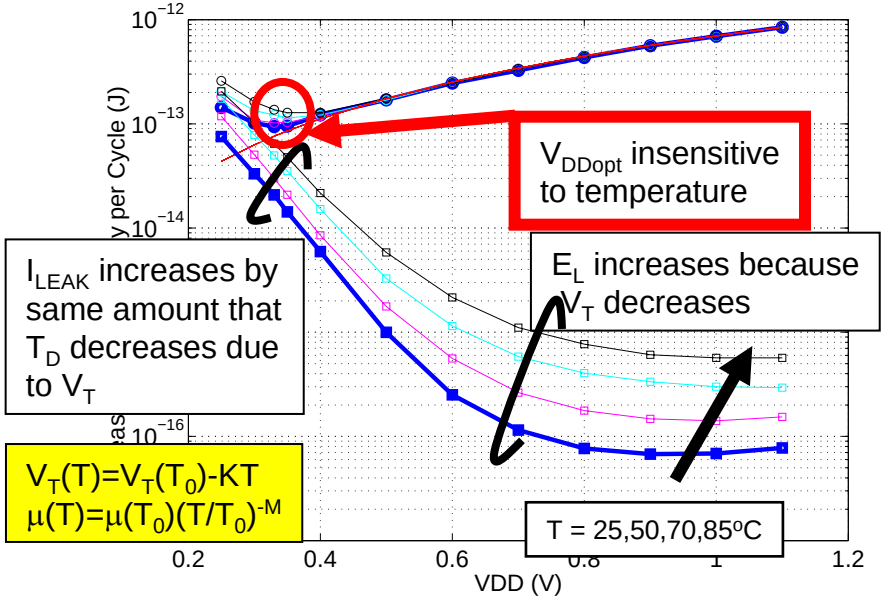


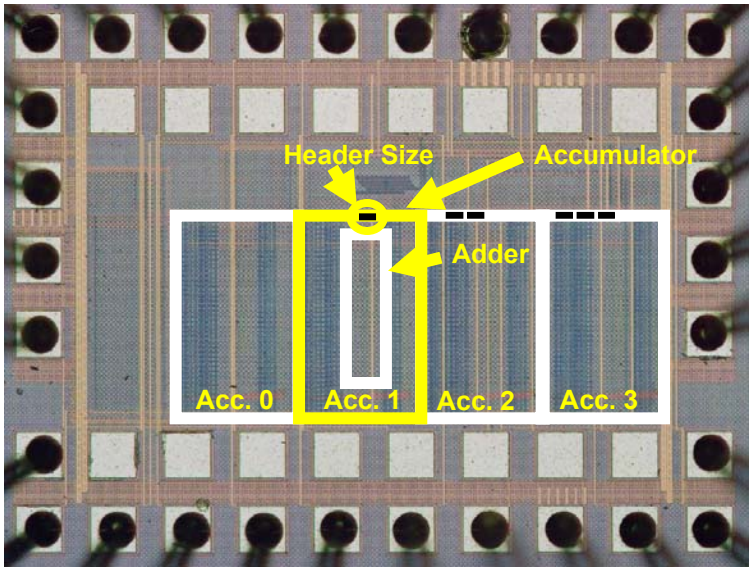
90nm Test Chip Circuit



Test Chip Results



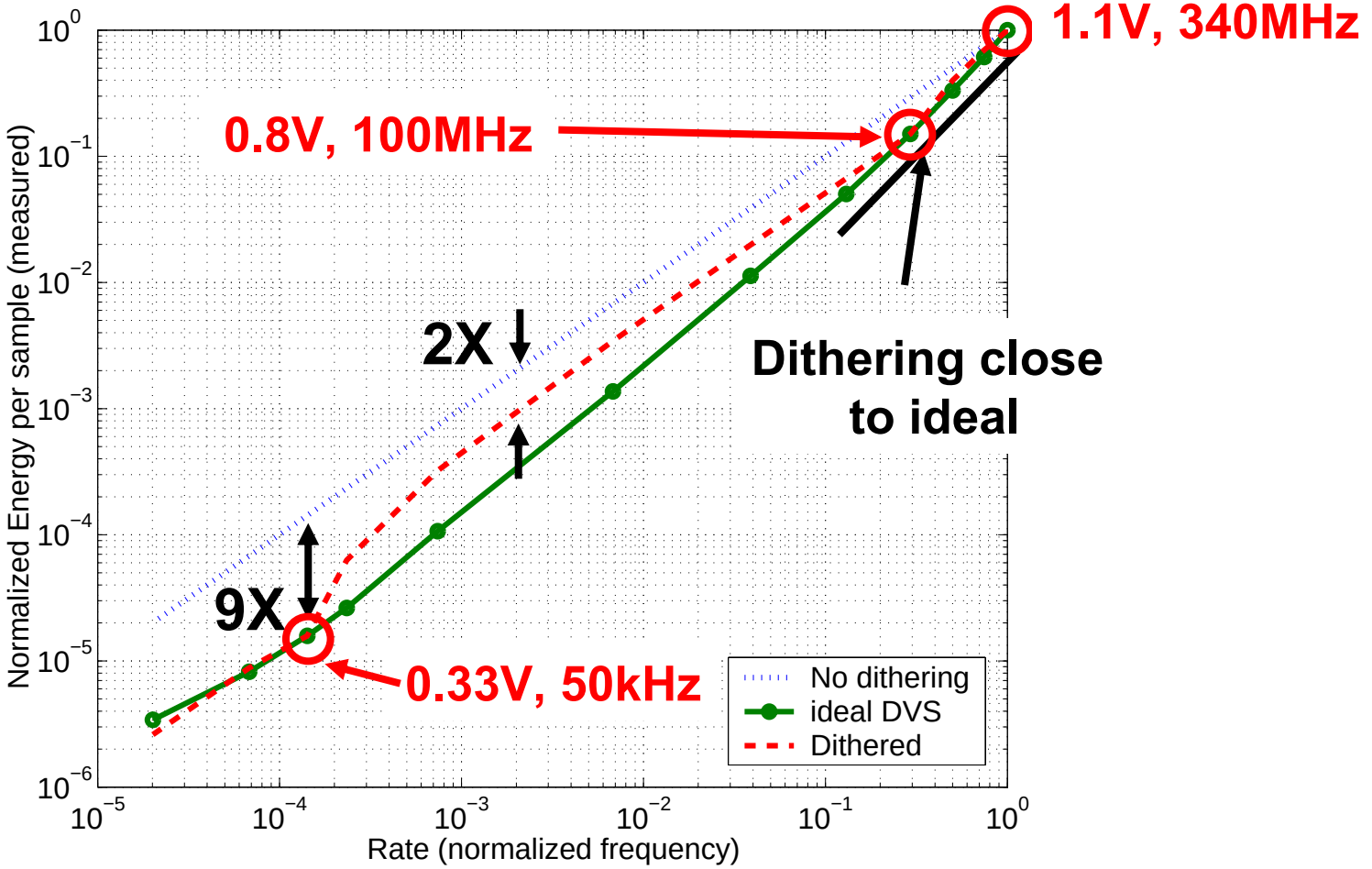




90nm, 6LM

- **Adders work to <200mV**
- **Min E ~330mV, relatively insensitive to T**
- **Voltage dithering reduces energy for varying workload at higher rates**

UDVS Results



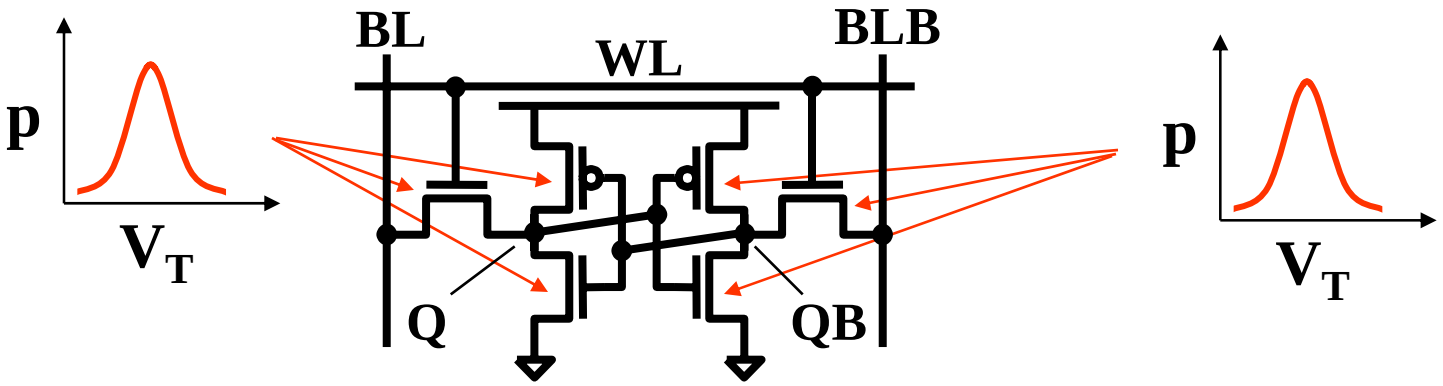
Dither during high performance operation and switch to sub-threshold minimum energy operation when speed is not important

Outline

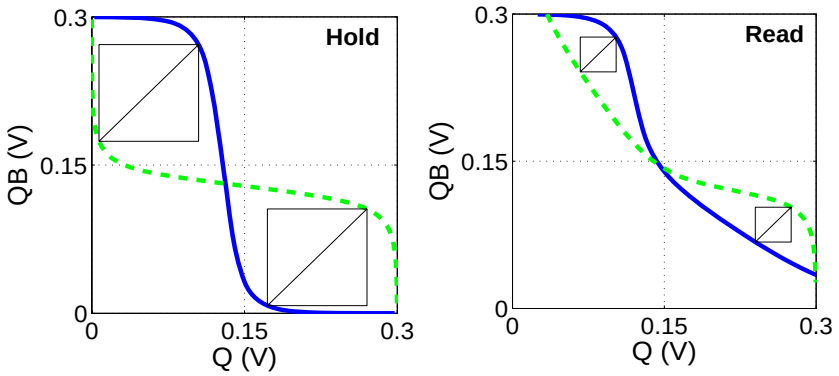
- Background
- Sub-threshold FFT Processor
- Ultra-Dynamic Voltage Scaling (UDVS)
- Sub-threshold SRAM
  - Challenges
  - Potential Bitcells
  - Sub-threshold SRAM Example
- Conclusions

6T Read – Static Noise Margin

- **Variation reduces yield because of SNM**

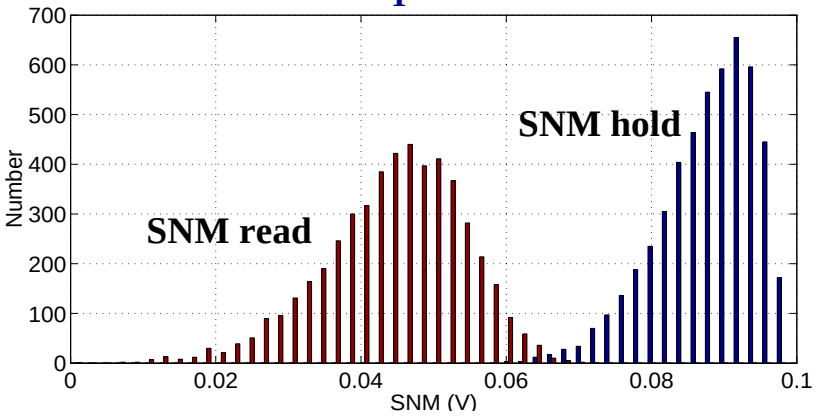


Butterfly Plots: Read SNM worst

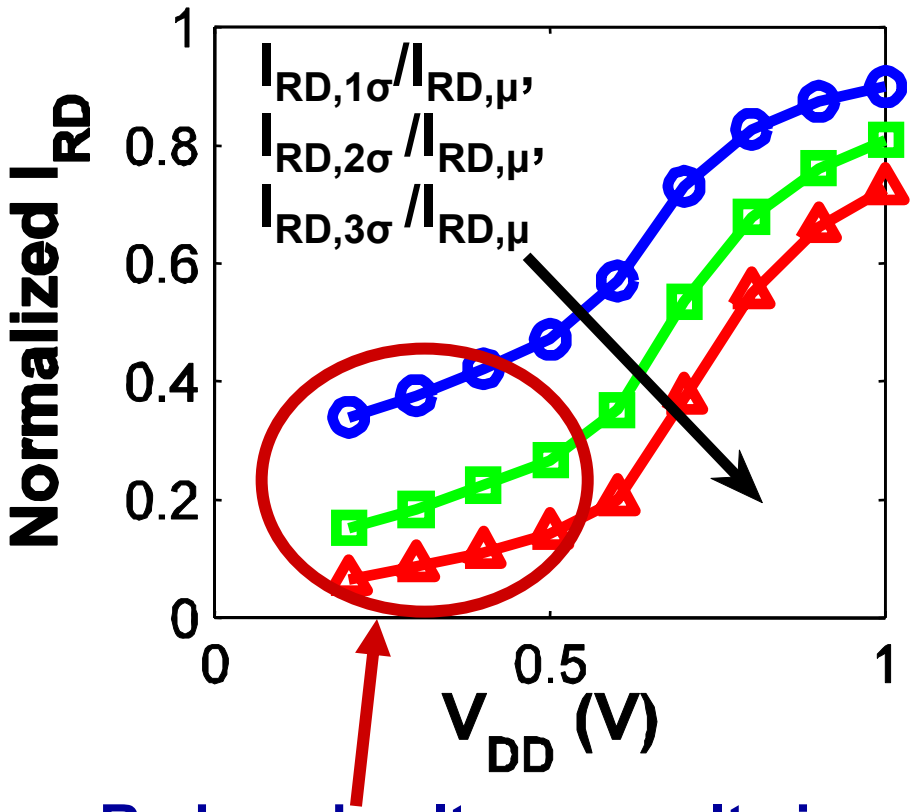


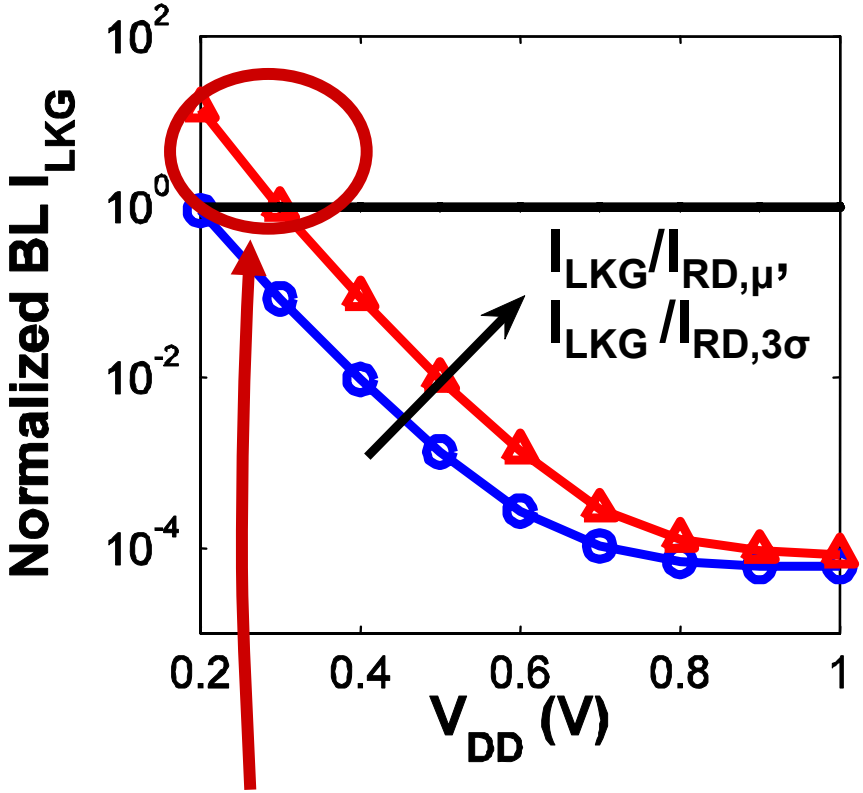
Variation aggravates situation

SNM for sub-V_T, 6T cell at 300mV



Read Current Variation





- Reduced voltage results in larger deviation between worst case I_{RD} and mean I_{RD}

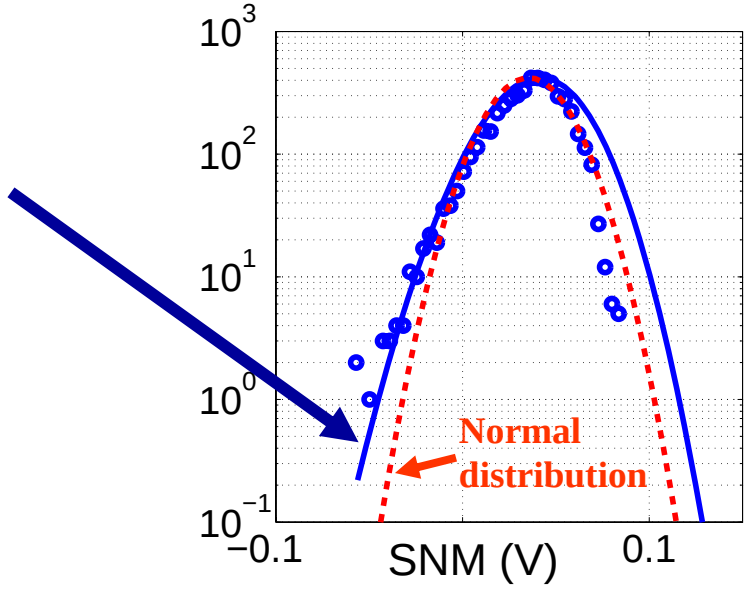
overall performance is severely degraded

- Leakage current from 127 unaccessed cells exceeds worst case I_{RD}

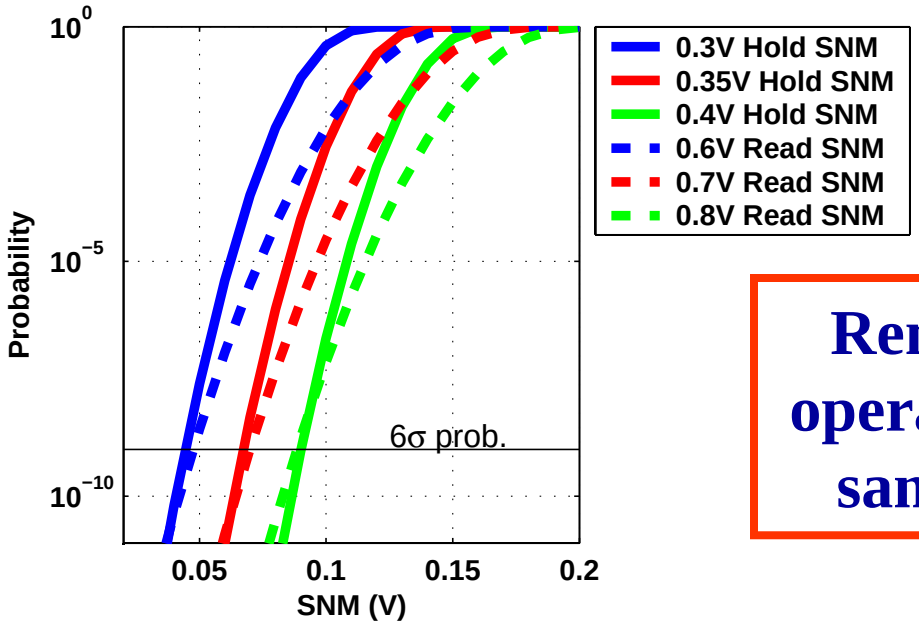
data is indistinguishable!

Simulations Compared to Model

Model* gives good estimate for the distribution of SNM at the worst-case tail



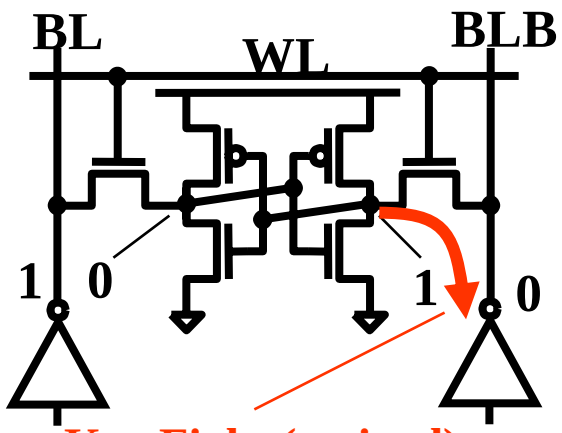
log plot of CDFs



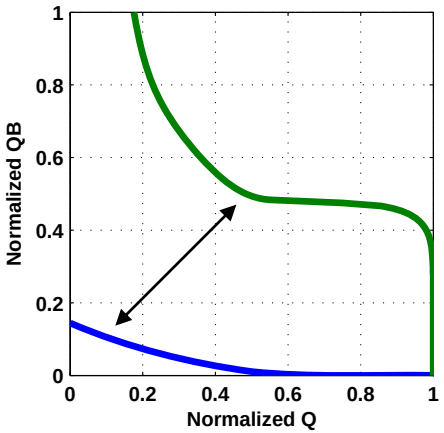
Removing Read SNM allows operation at lower V_{DD} with the same stability at the corners

*Calhoun and Chandrakasan, ESSCIRC 2005.

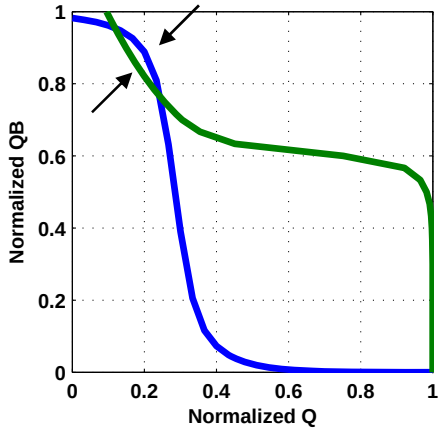
Inability to Write

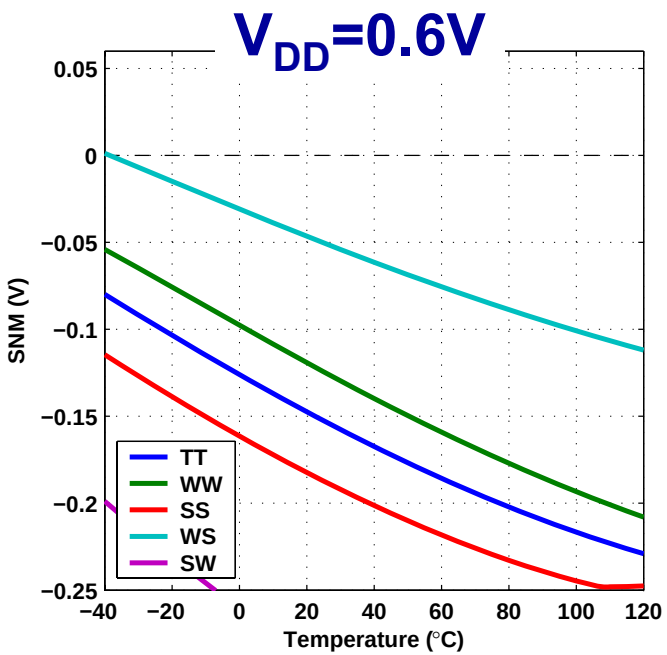


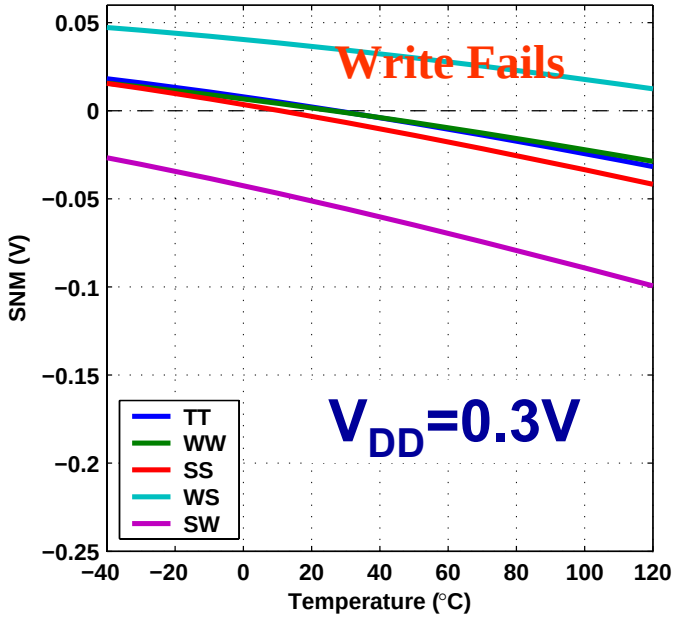
Successful write: Negative "SNM"



Write failure: Positive SNM



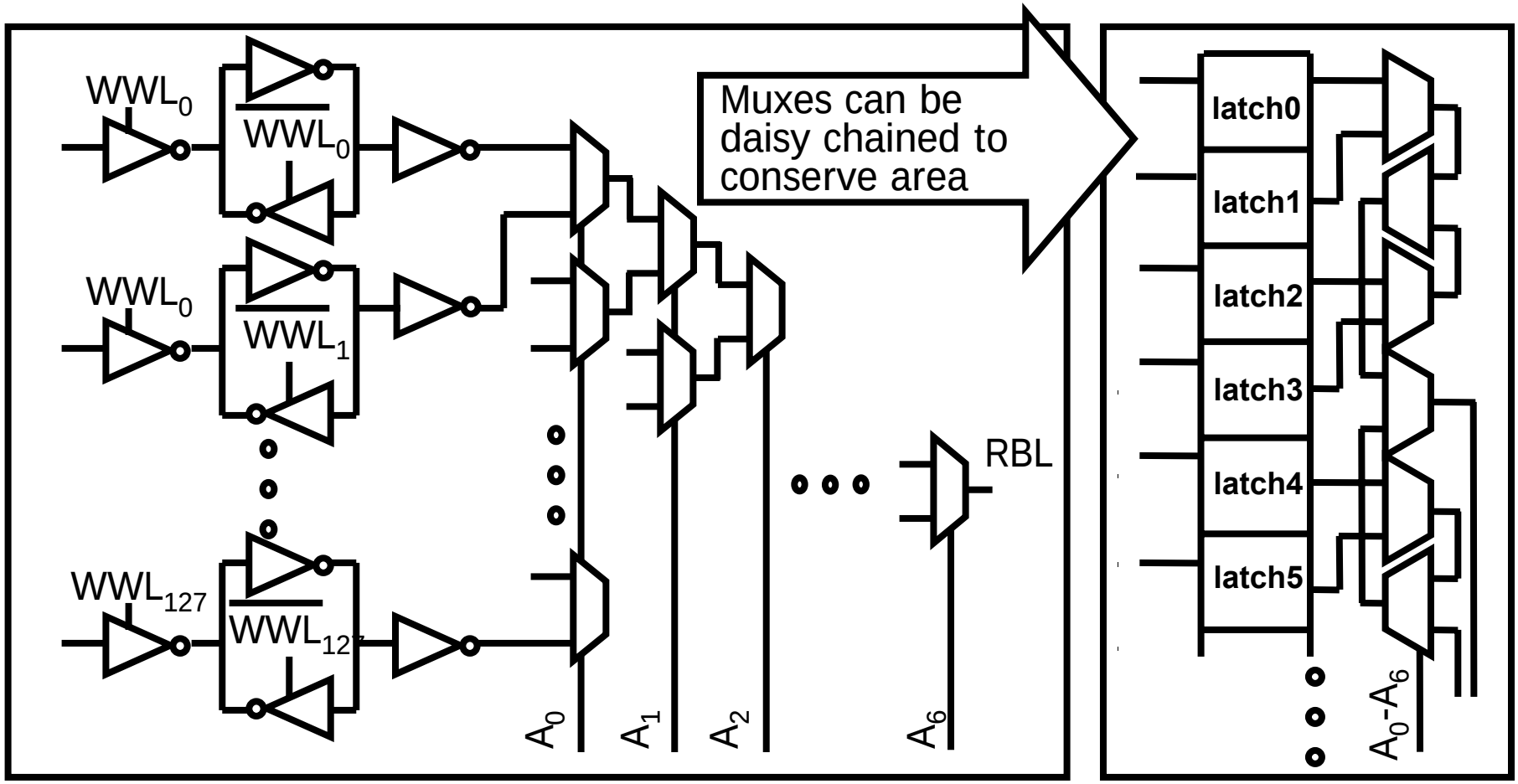




Conventional Write: 0.6V is about the best we can hope for

Mismatch will make this even worse

Previous Work



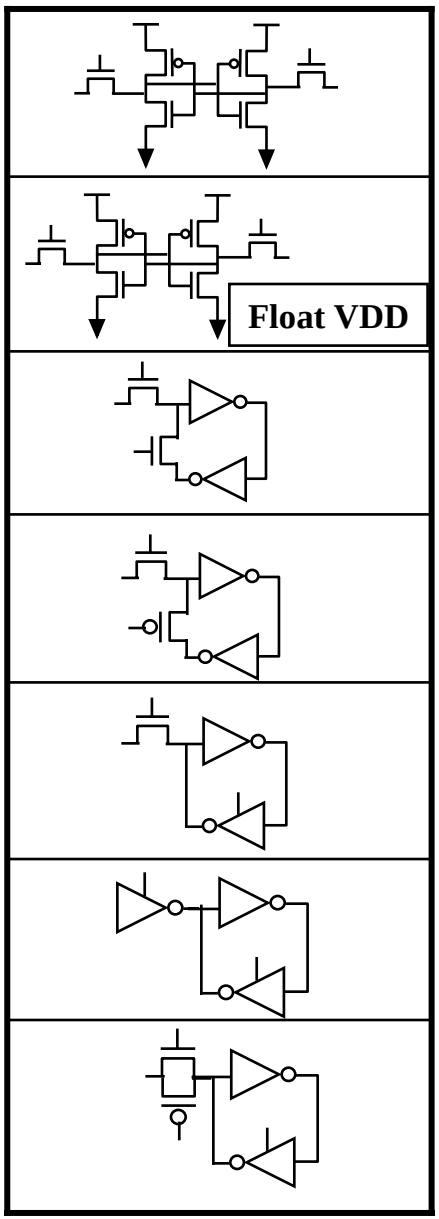
- **18T cell equivalent (when you include the mux)**
- **20-50X slower than 6T at 1.2V**

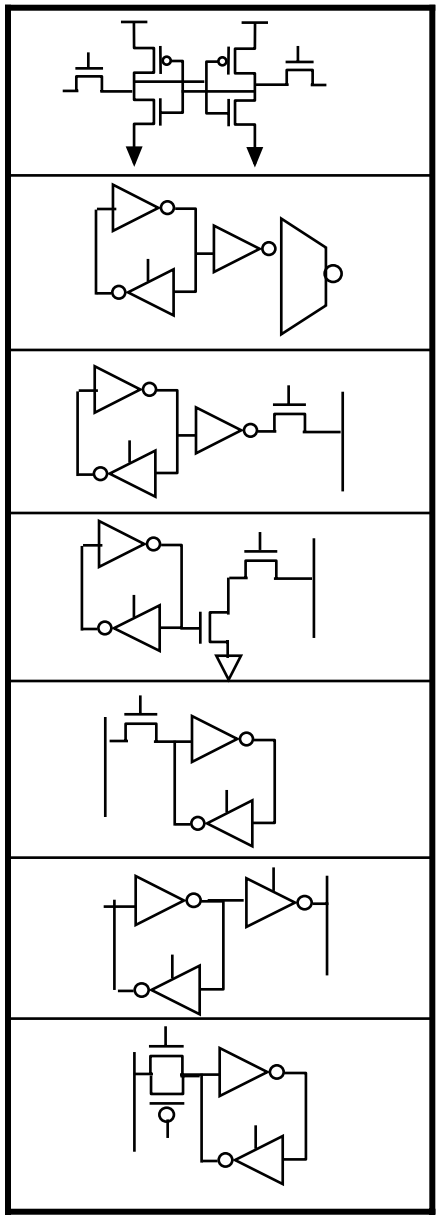
Wang and Chandrakasan *ISSCC* 2004

SubV$_T$ bitcell options (65nm)

Write approaches

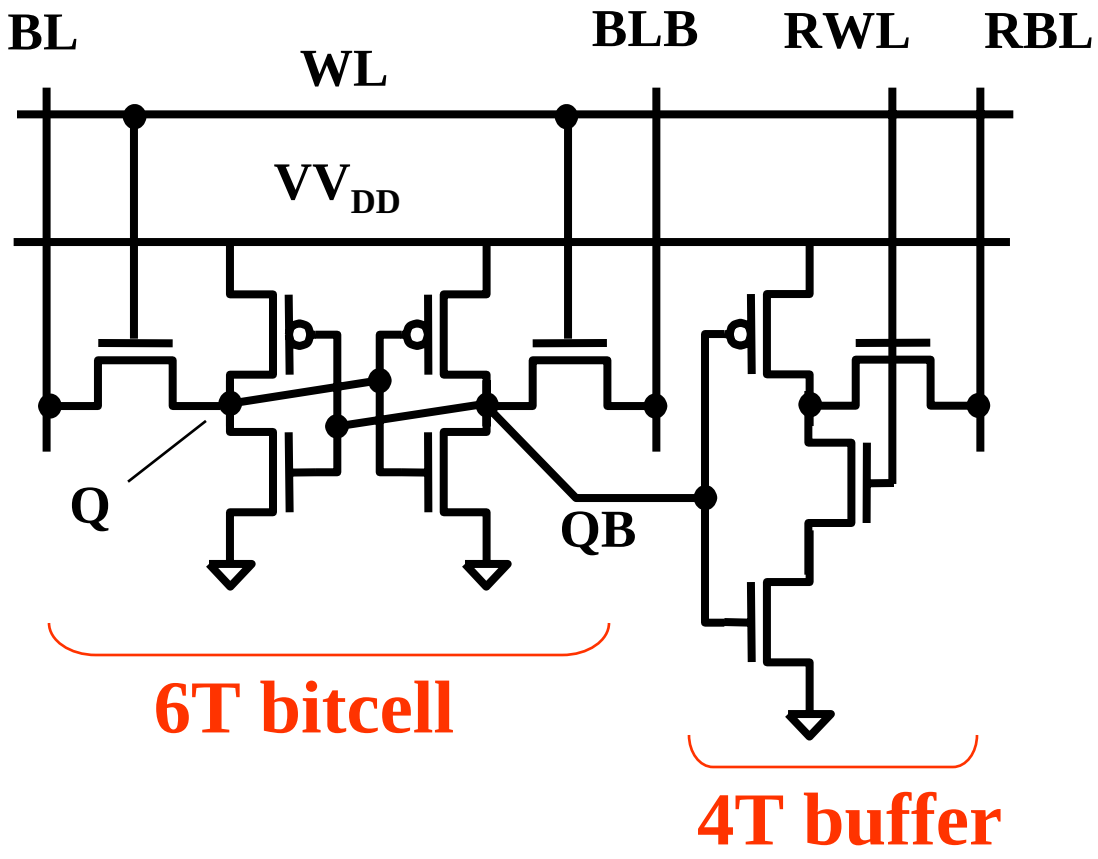
- Combinations tried for sub-threshold
- All failed at some point or required very short BLs for this 65nm process



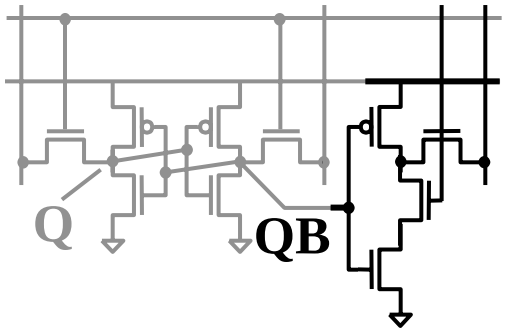


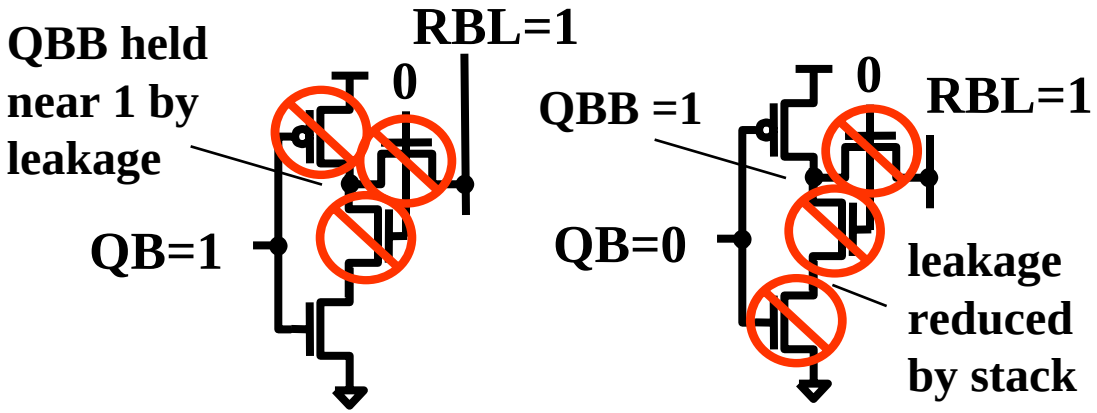
Read approaches

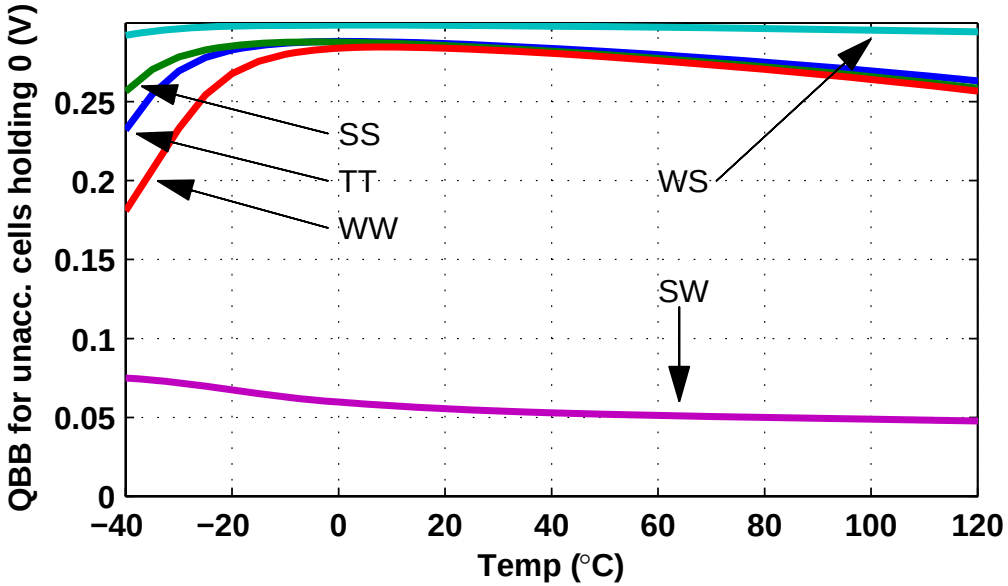
Proposed 10-T bitcell for Sub-V_T



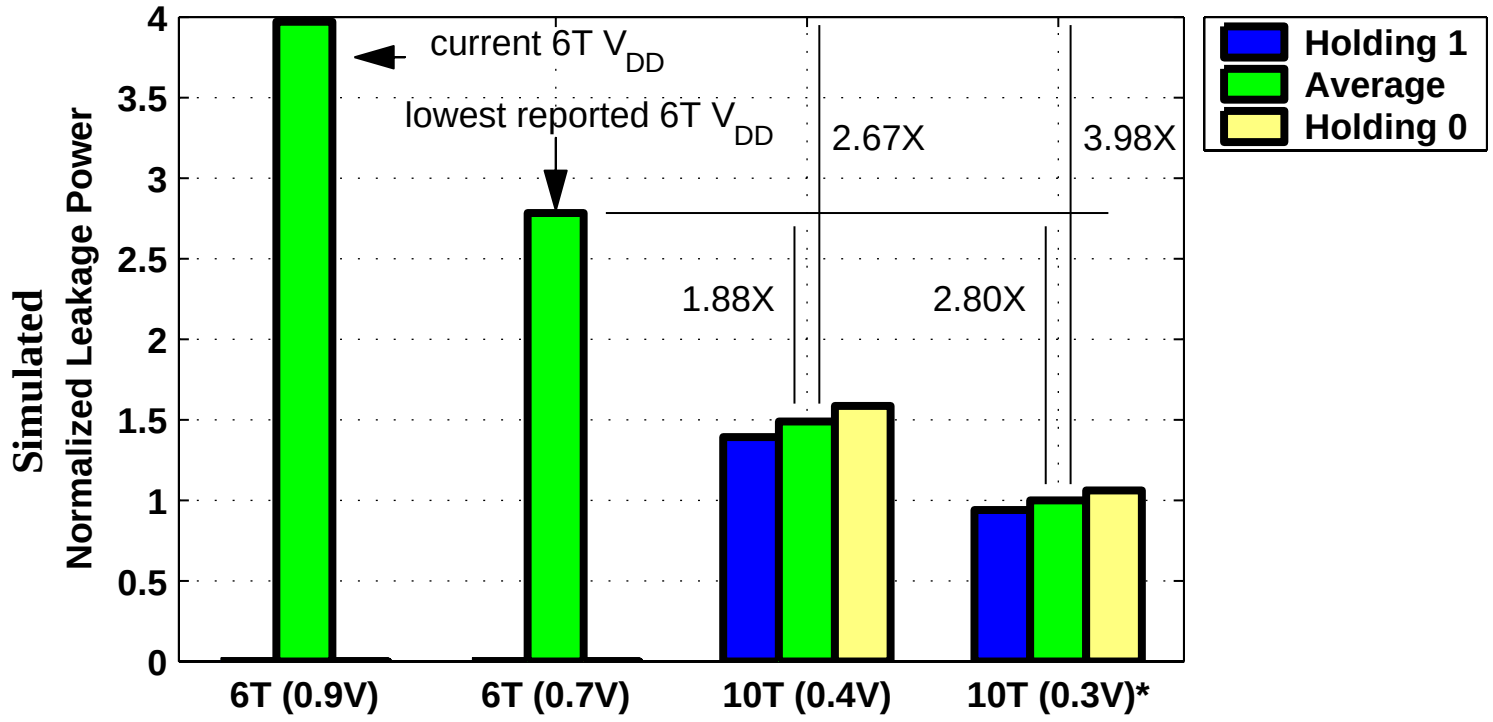
10T Bitcell Reduces Bitline Leakage







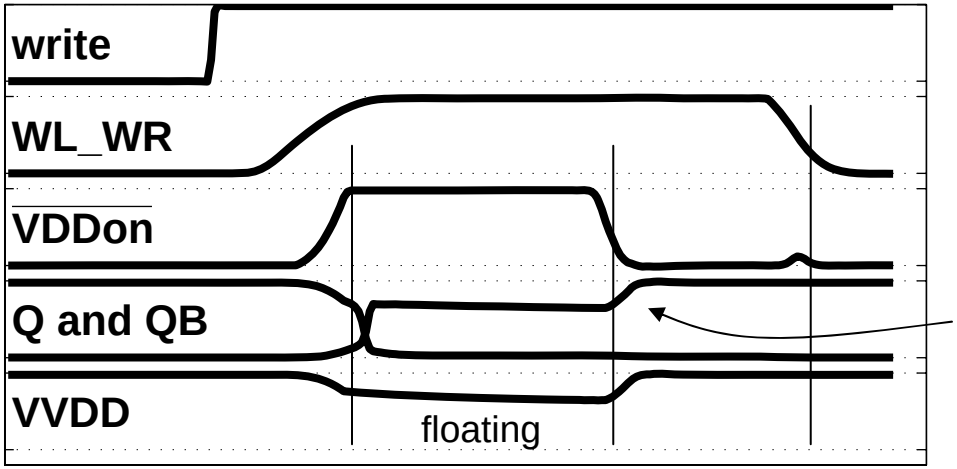
Leakage Power Savings with 10T Bitcell



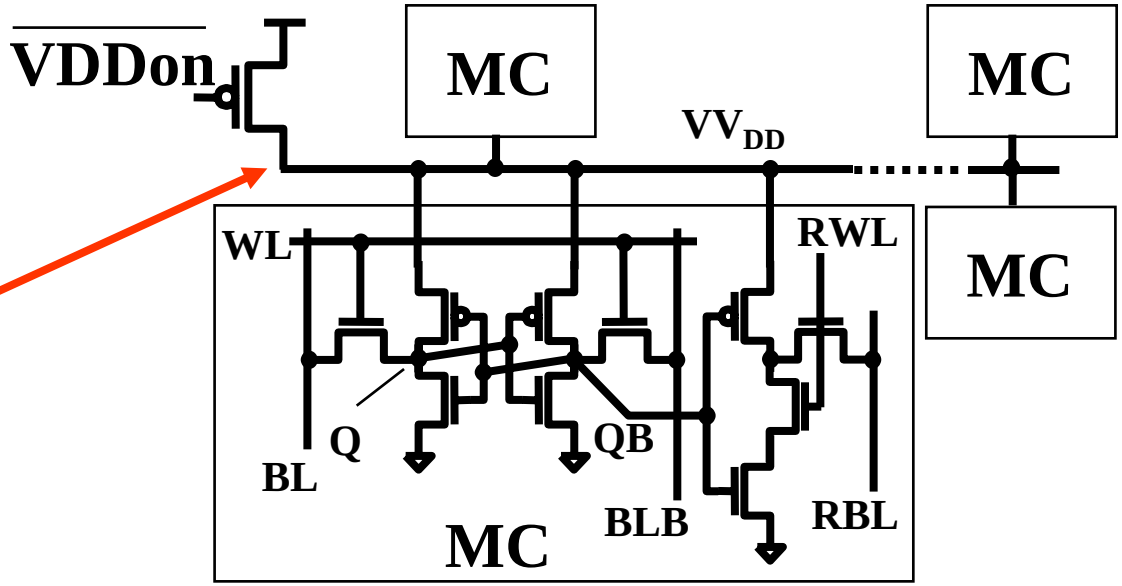
- 6T memories in 65nm usually at 0.9V or greater (lowest reported is 0.7V)
- 10T bitcell allows scaling to lower voltages
- Lower voltage operation reduces leakage power dramatically for unaccessed cells

10T Bitcell Allows Sub-V_T Write

Floating V_{DD} weakens feedback and allows write

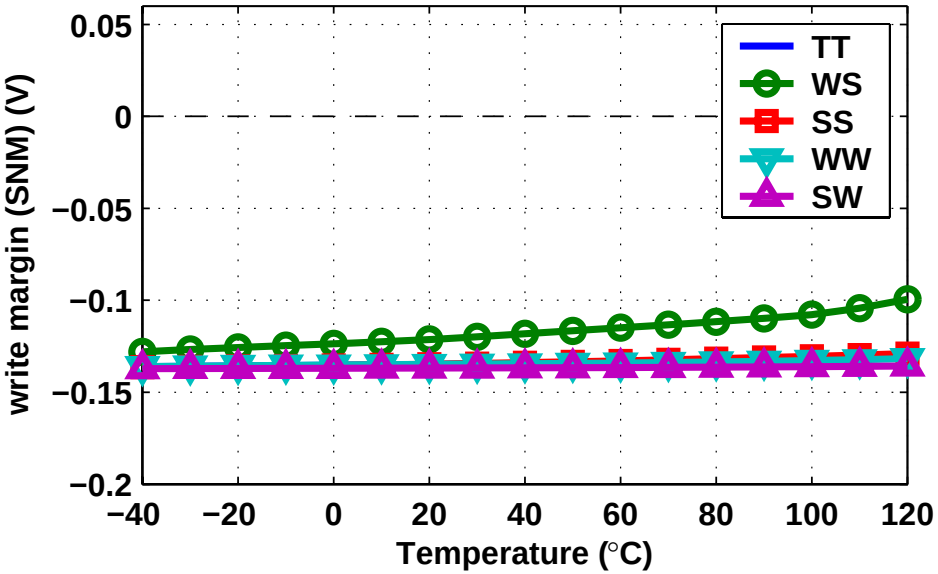


feedback restores '1' to V_{DD}

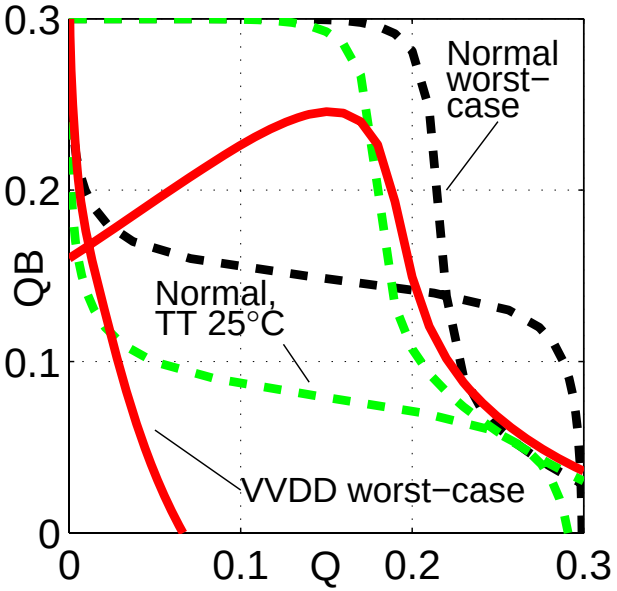


Folded WL shares VV_{DD}

10T Bitcell Allows Sub-V_T Write

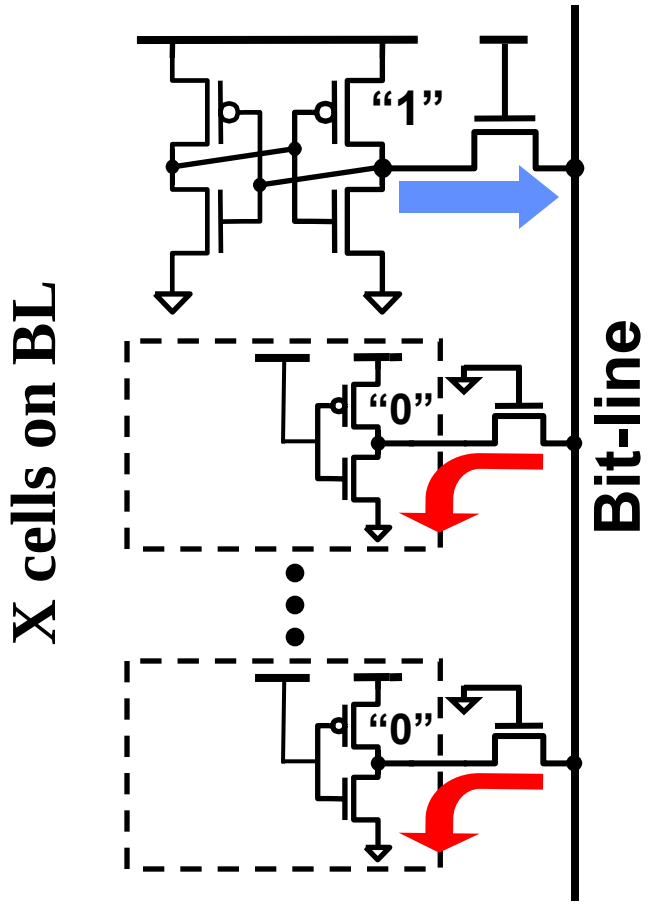


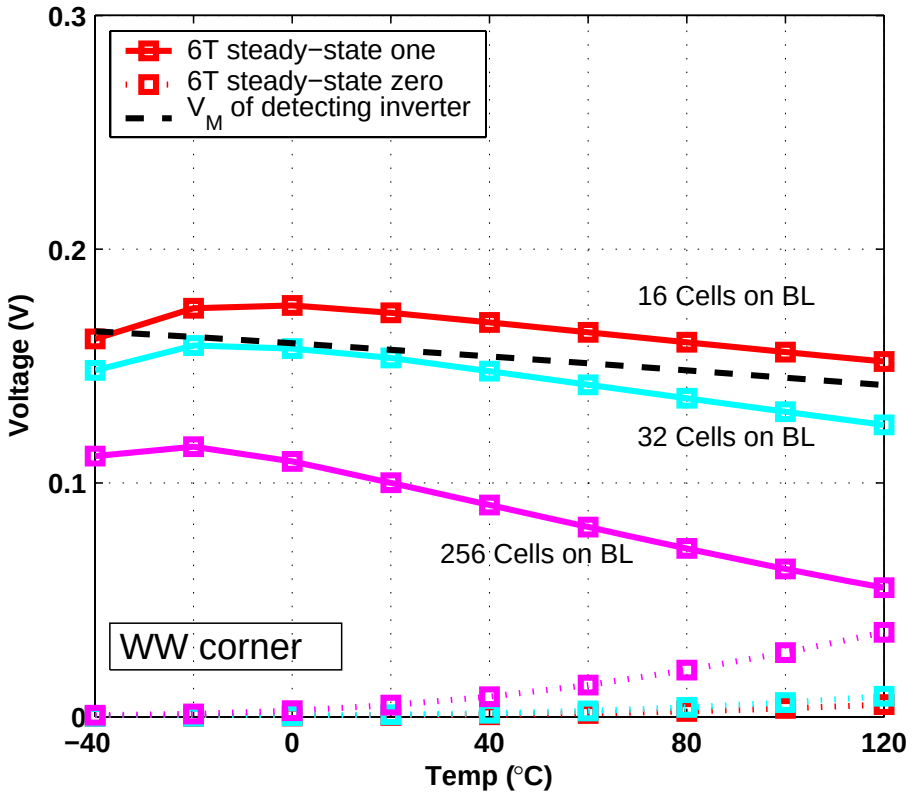
Write successful across process corner and temperature



Write margin correctly negative at worst-case corner

Bitline Leakage Limits Integration Level

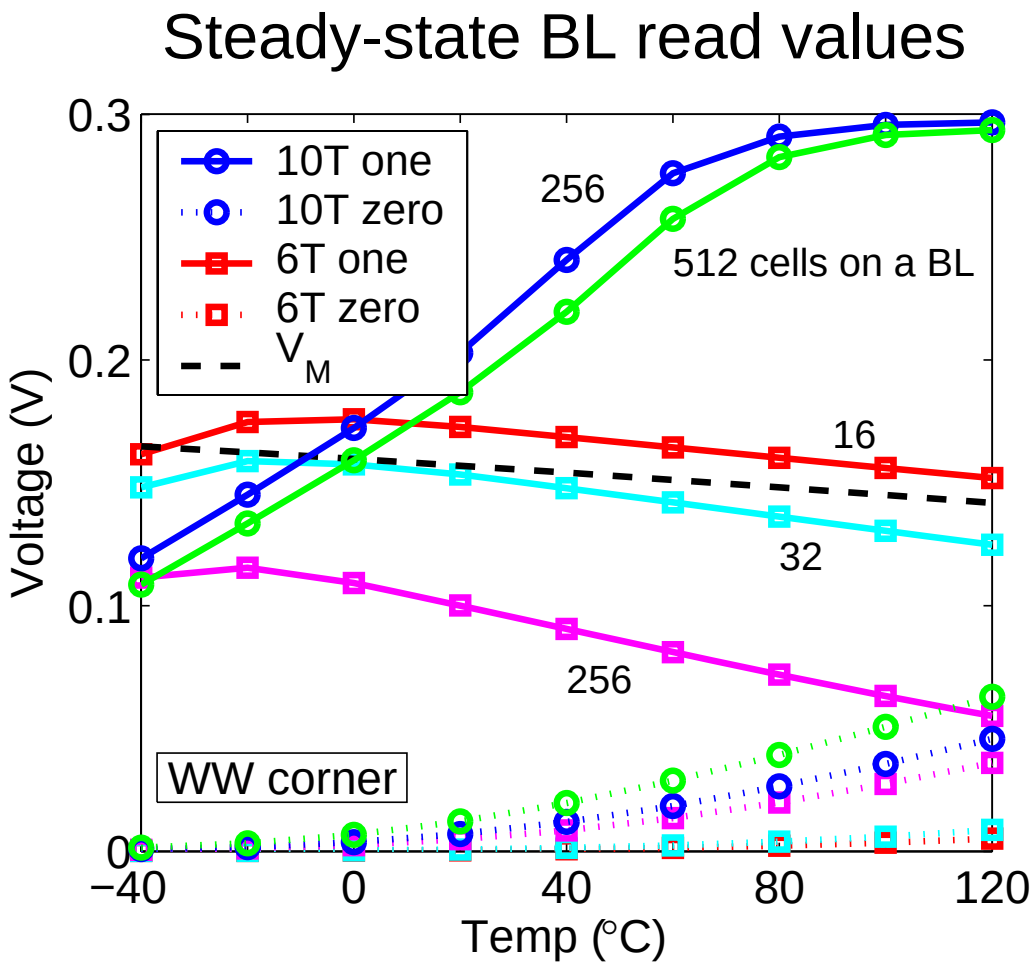




- BL leakage degrades read values

16 bitcells on bitline is best can hope for standard 6T

10T Bitcell Lowers Bitline Leakage

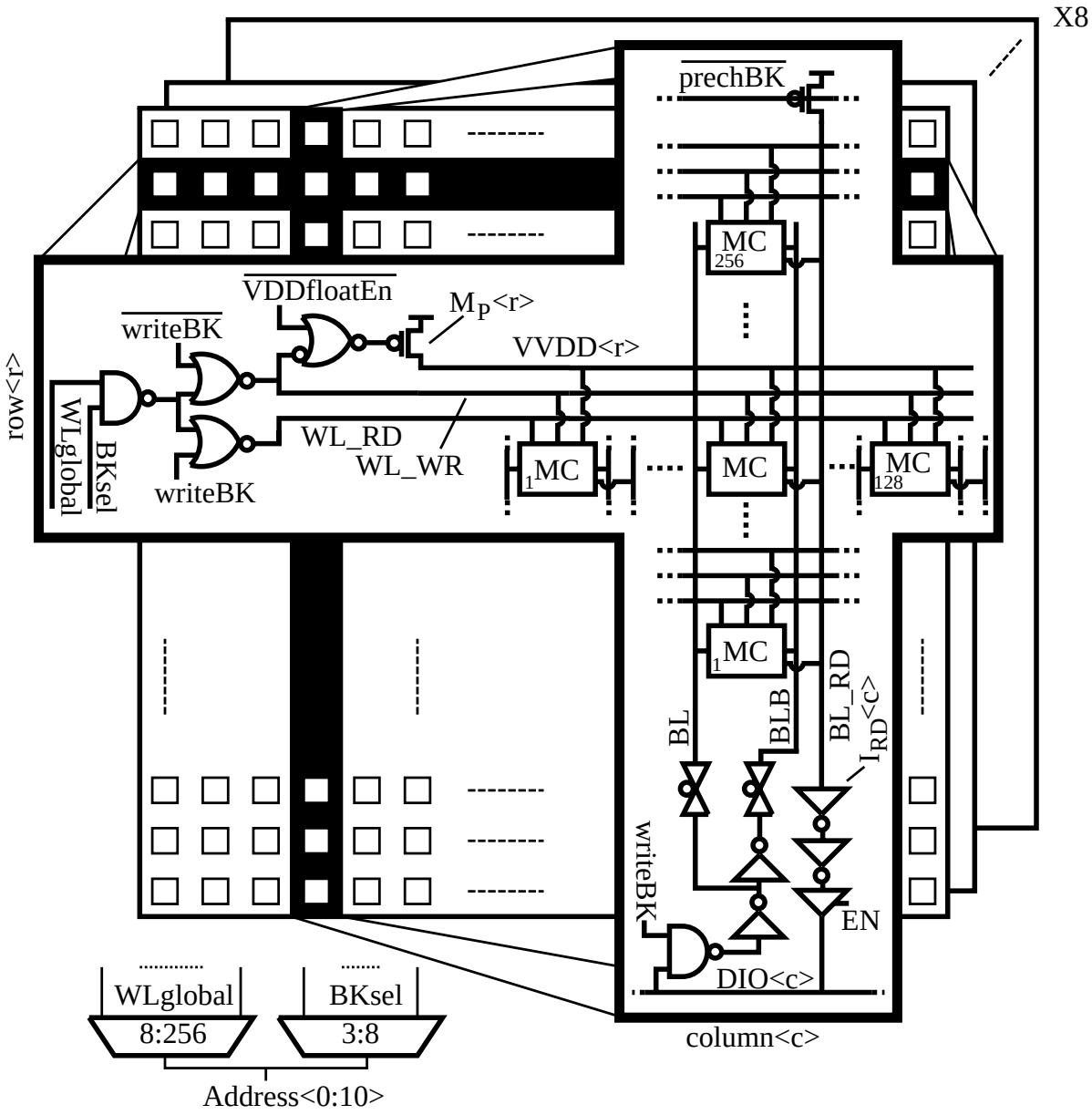


10T bitcell enables higher level of integration on BL

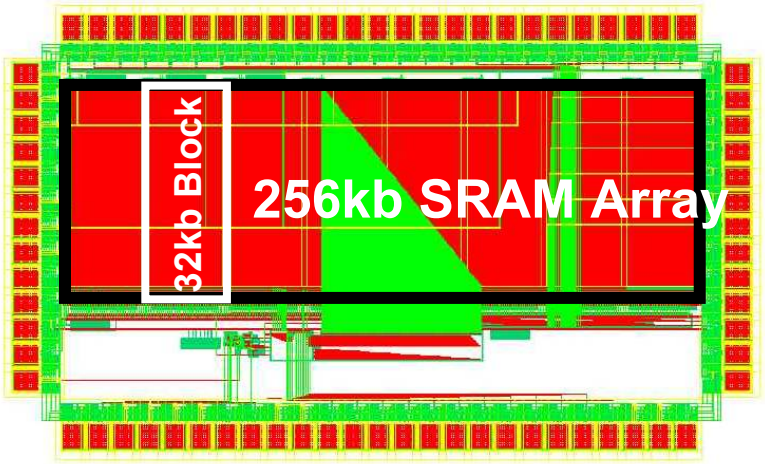
Test Chip Architecture

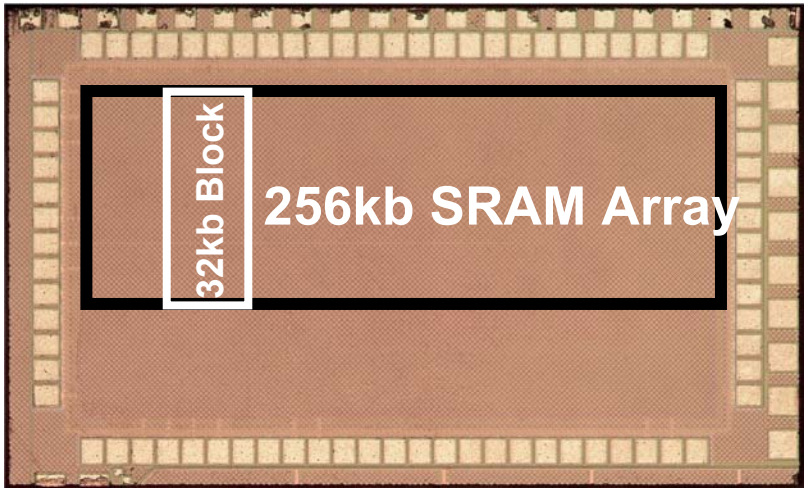
- 256 rows and 128 columns per block
- Static CMOS peripherals
- Separate WL V_{DD} for boosting
- Assumed 1x1 redundancy
- Simulation:

Operates at 300mV across all process corners from 0 to 100°C



256Kb 65nm Sub-V_T memory



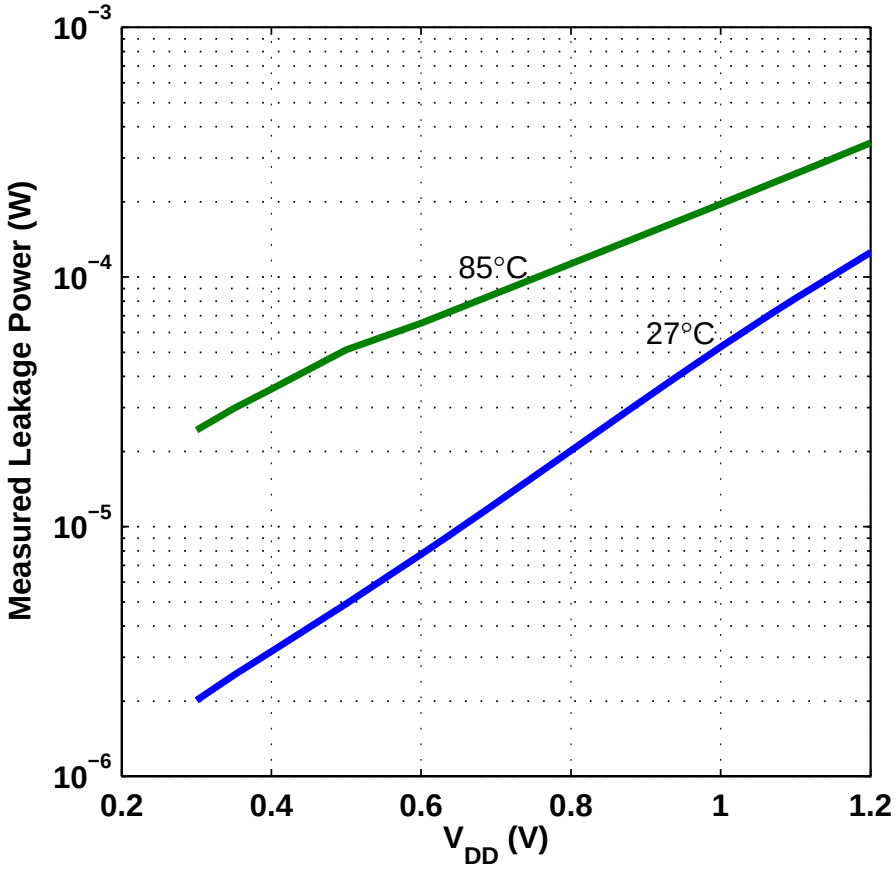


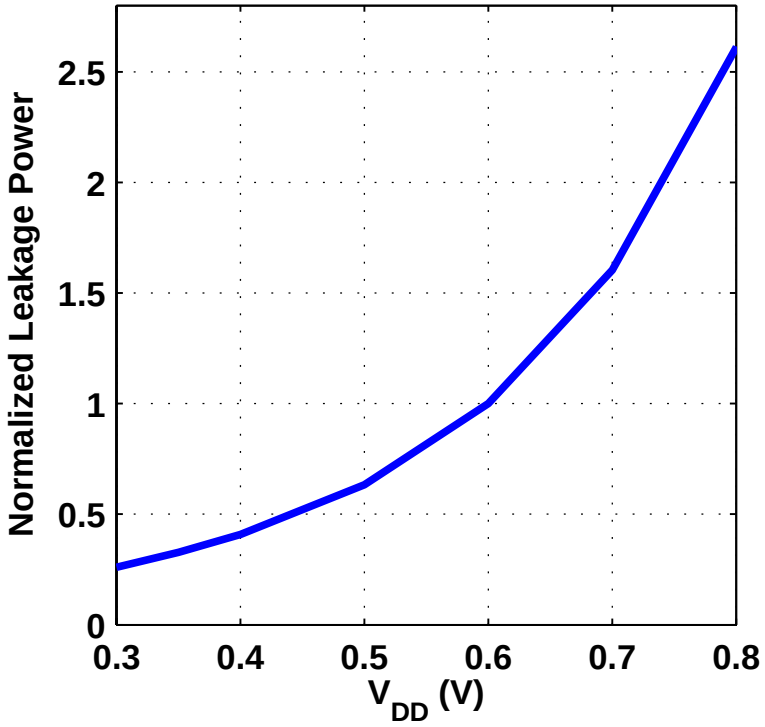
Test chip addressing the sub-V_T problems using 10T bitcell: 1.89mm by 1.12mm

Chip functions to below 400mV, holds without error to <250mV:
At 400mV, 3.28μW and 475kHz at 27°C

Reads without error to 320mV (27°C) and 360mV (85°C)
Write without error to 380mV (27°C) and 350mV (85°C)

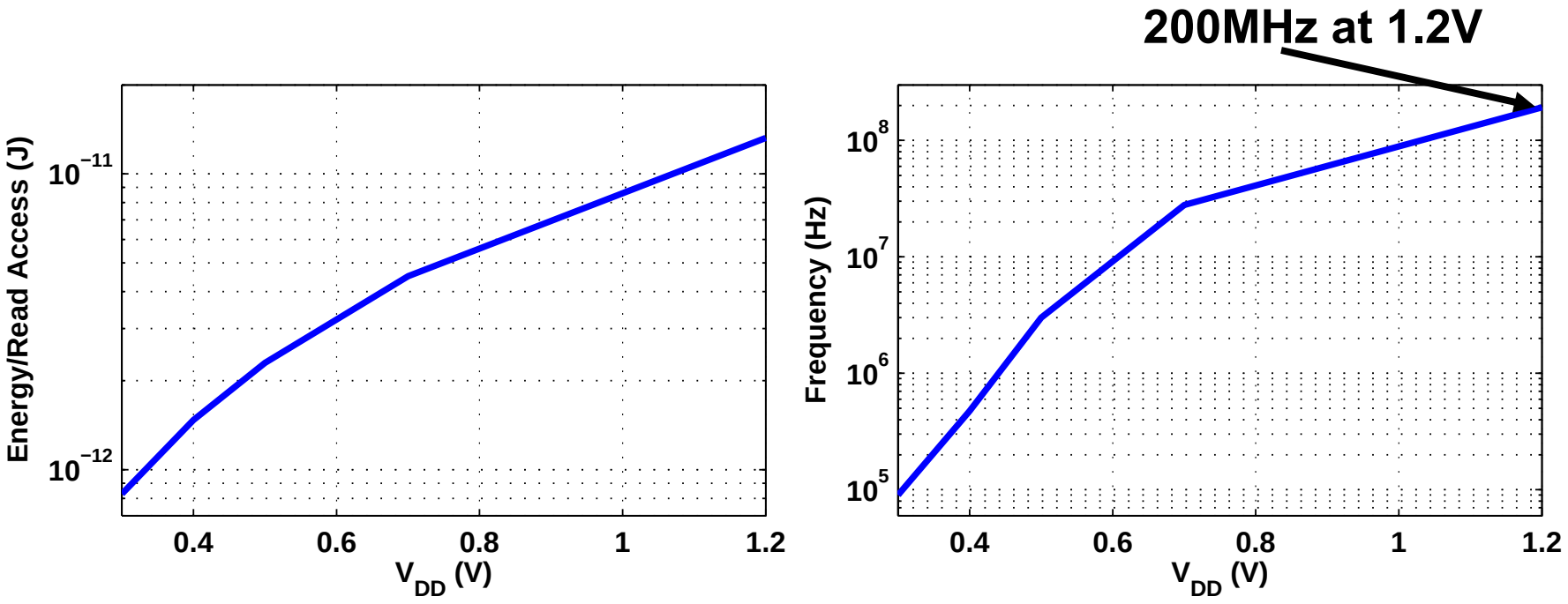
Power Measurements





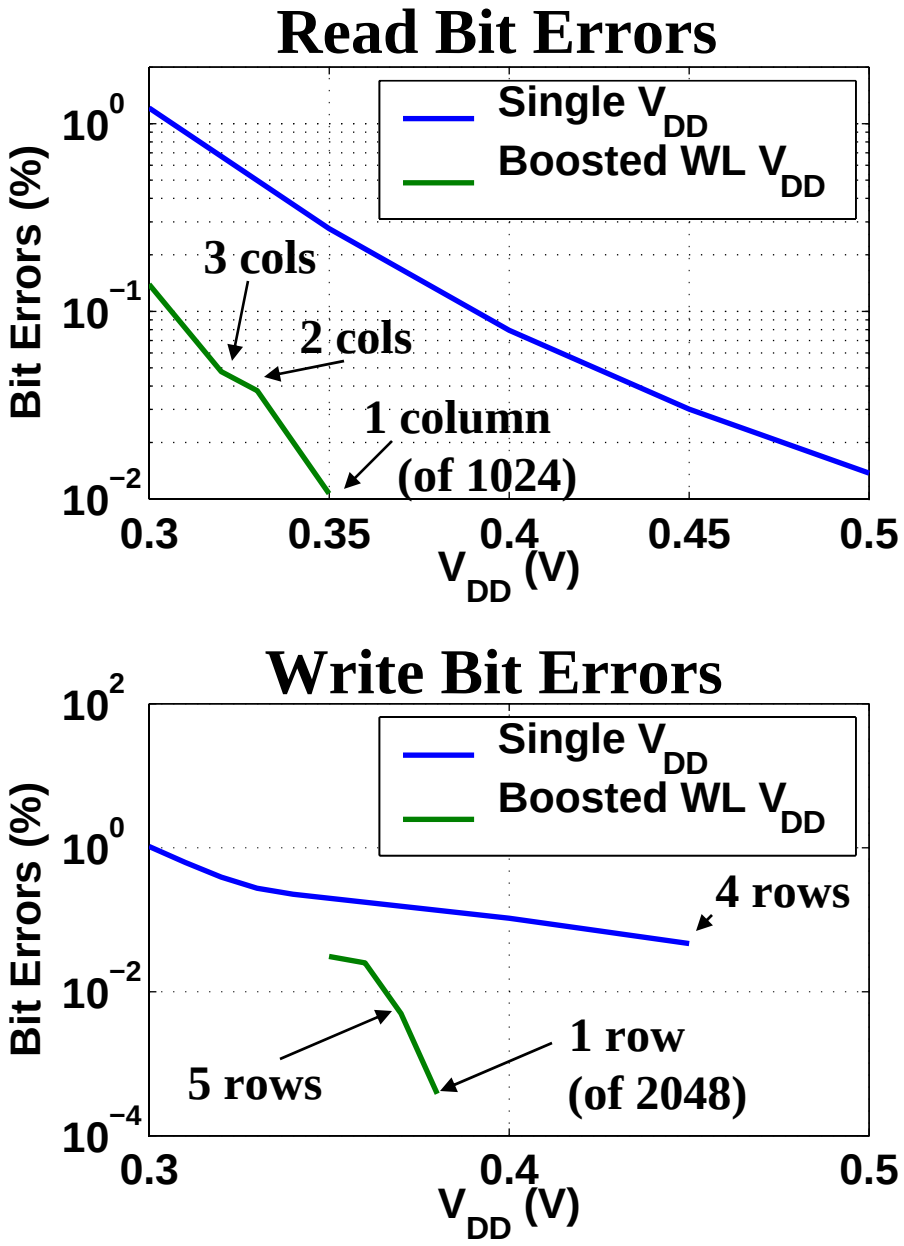
Relative to 0.6V 6T SRAM, 2.2X less leakage power at 0.4V and
3.3X less leakage power at 0.3V
>60X less leakage power than 1.2V

Active Energy Savings with 10T Bitcell



- 6T memories in 65nm usually at 0.9V or greater (lowest reported is 0.7V)
- Operating 10T bitcell at lower voltages saves energy
- 10T memory can provide high frequency operation at higher voltages when necessary

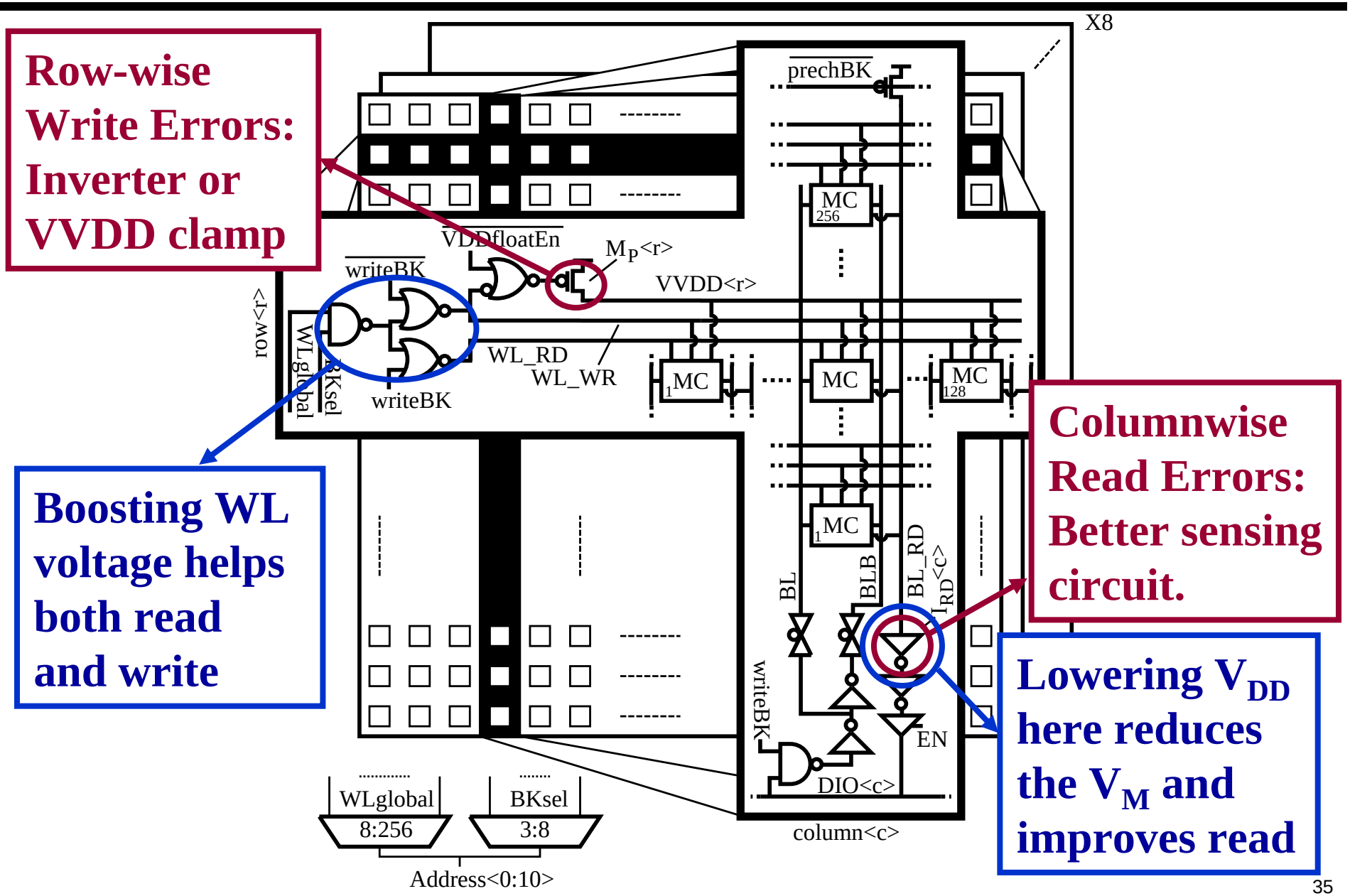
V_{DD} Scaling Limits



Redundancy and/or boosted WL account for mismatch

1x1 redundancy and WL boosting:
Read works to 320mV
Write works to 380mV

Voltage Limits and Proposed Improvements



Conclusions

- Sub-threshold allows minimum energy operation
- Process variations limit sub-threshold operation of CMOS circuits
- Standard 6T SRAM limited to ~0.6-0.7V and 16 cells per bitline
- Proposed 10T bitcell shows sub-threshold operation with overall power and energy savings
- Sub-V_T memory requires circuits and architectures to manage variability and low I_{on}/I_{off}

We acknowledge Texas Instruments and DARPA for funding the research described in this talk, and we thank Texas Instruments for chip fabrication and helpful collaboration.