A Combined DOE-ILP Based Power and Read Stability Optimization in Nano-CMOS SRAM

G. Thakral, S. P. Mohanty and D. Ghai
Dept. of Comp. Science & Engineering
University of North Texas, USA.
Email: saraju.mohanty@unt.edu

Dhiraj K. Pradhan
Dept of Computer Science
University of Bristol, UK.
Email: pradhan@compsci.bristol.ac.uk

Acknowledgment: This research is supported in part by NSF award numbers CCF- 0702361 and CNS-0854182.

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

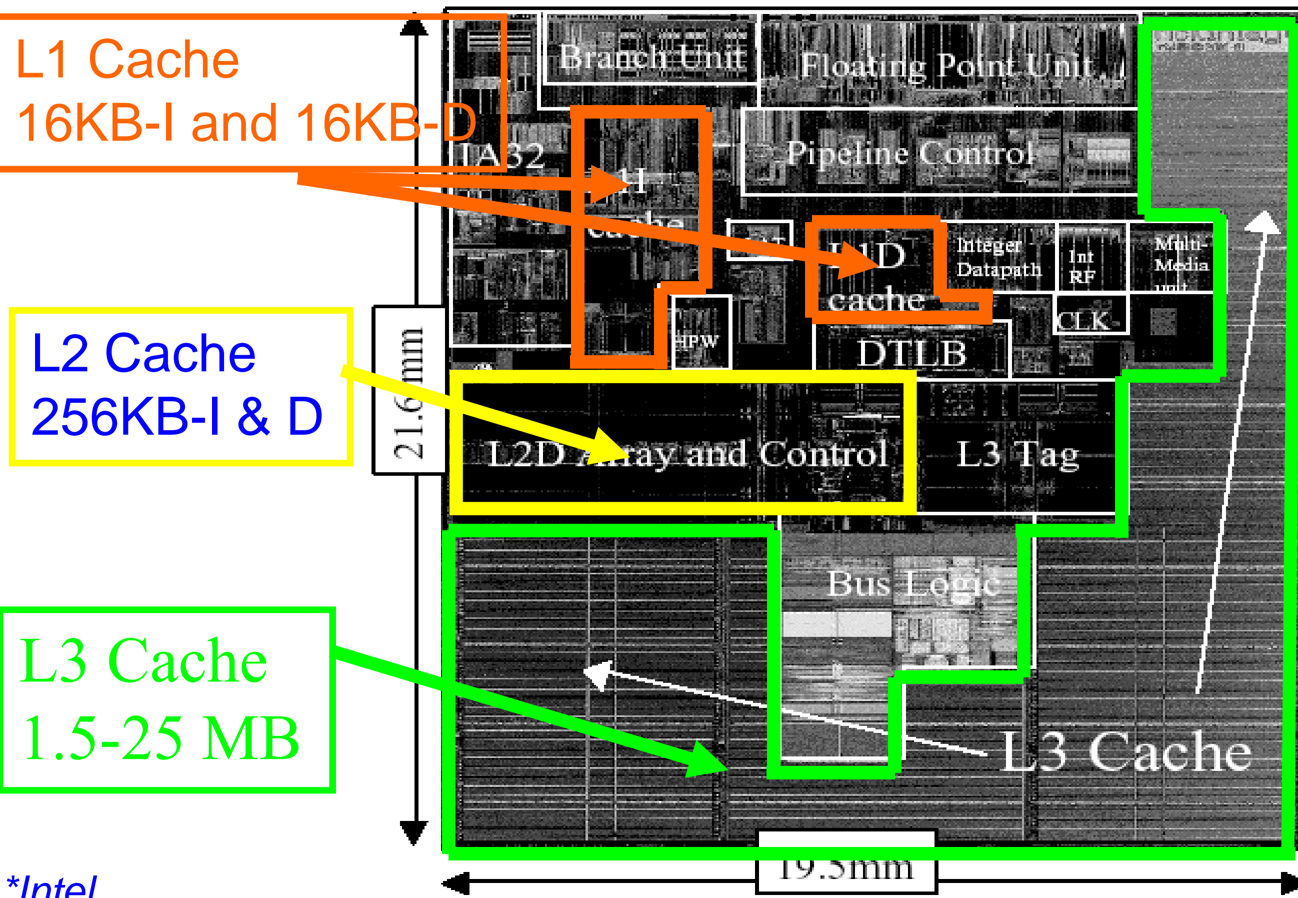
Outline of the talk

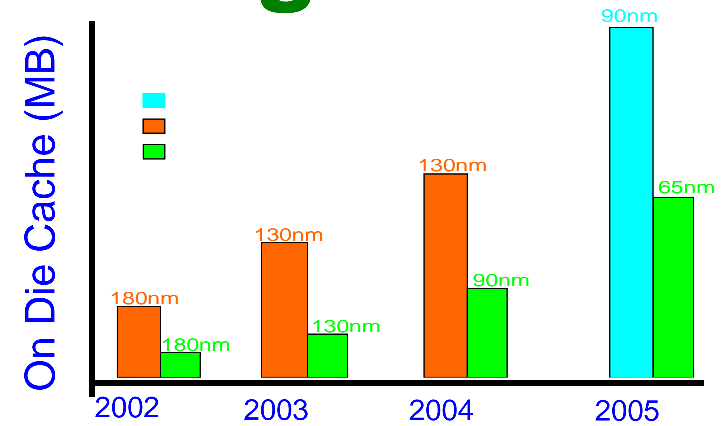
- Introduction
- Problem Statement: How to decrease power while maintaining performance of SRAM?
- Solutions: Assigning high/low V_{th} to transistors
- Proposed Optimal SRAM Design Flows
- Experimental Results: Nominal an Monte Carlo
- Related Prior Research
- Conclusions and Future Research

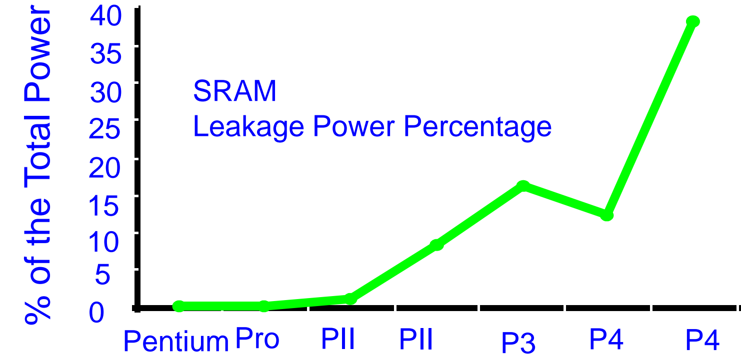
Why Efficient SRAM Design?

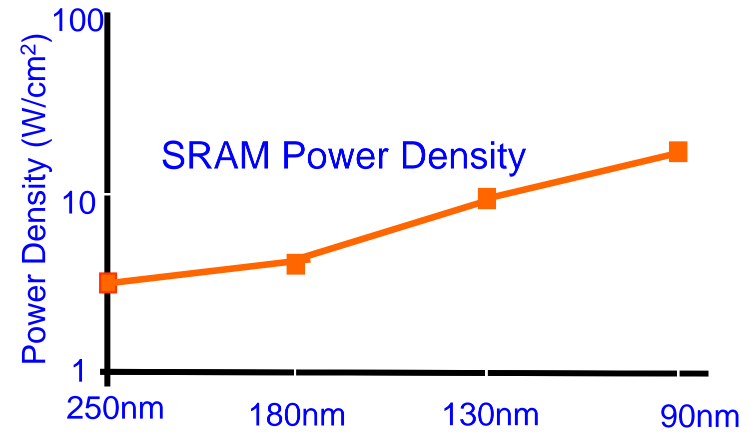
- Amount of on-die caches increases
- Up to 60% of the die area is devoted for caches in typical processor and embedded application.
- Largely contributes for leakage and power density.

Itanium 2* (L3-9MB) 130nm Technology





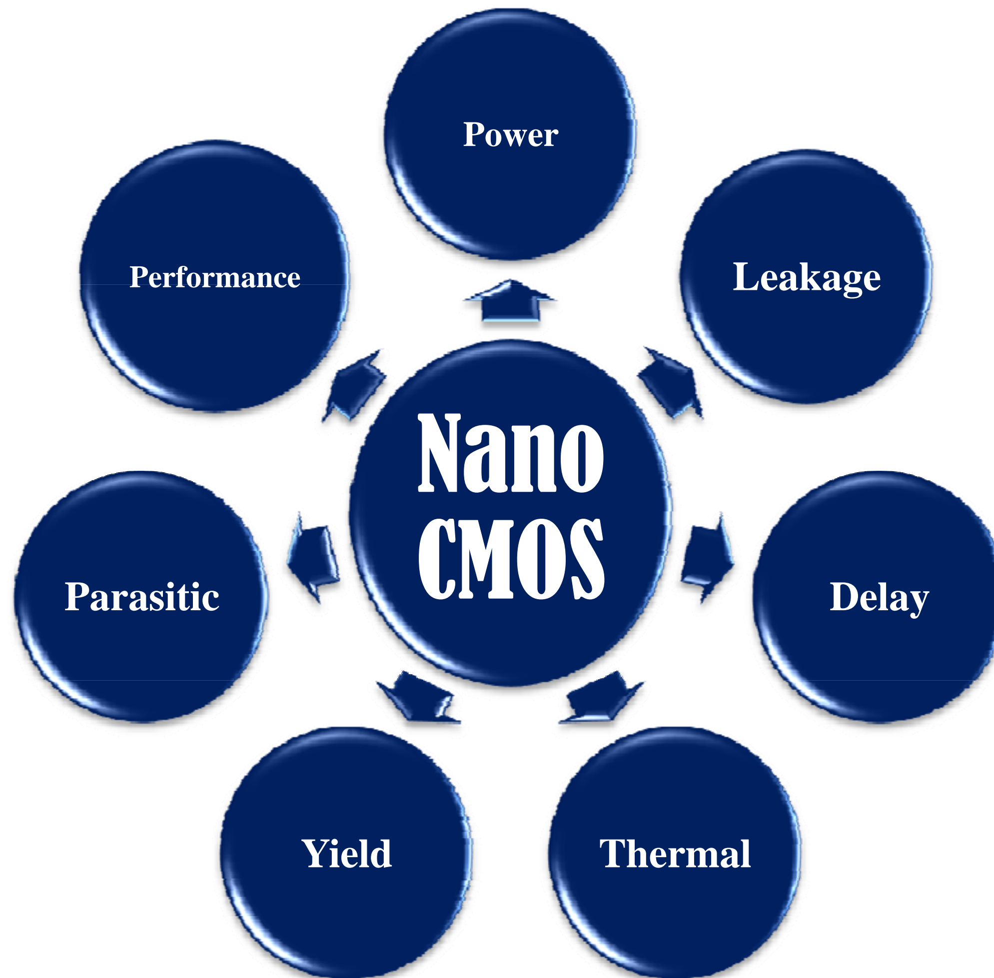




*Intel

UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

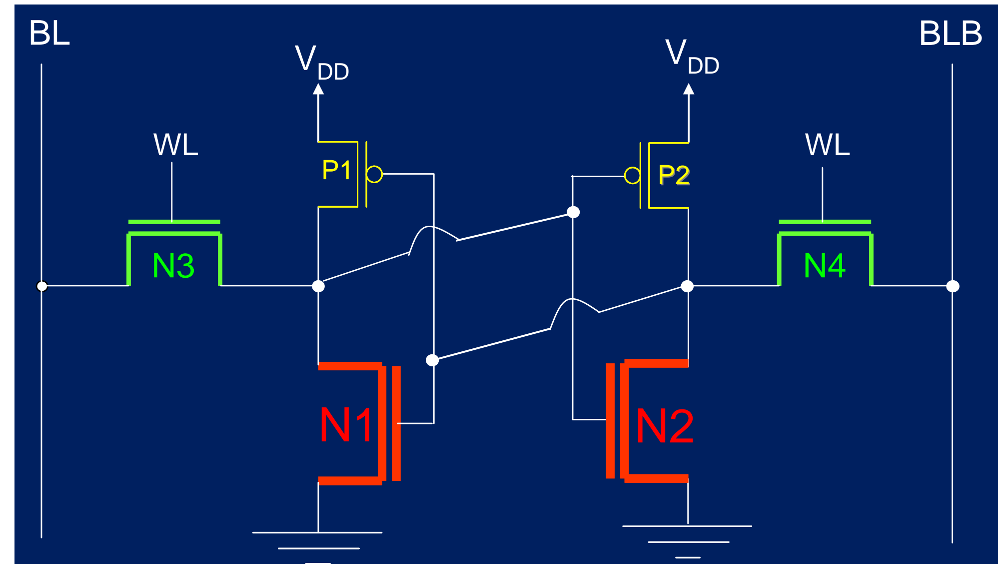
Issues in Nano CMOS



Nano-CMOS SRAM Design Challenges ...

In nano-CMOS regime following are the major issues:

- Data stability and functionality
  - Non-destructive read
  - Successful write
  - Noise sensitivity
- Proper sizing of the transistors
  - To improve the write ability
  - To improve the read stability
  - To improve the data retention
- Minimum size of transistors to maximize the memory density.
- Minimum leakage for low-power design.
- Minimum read access time to improve the performance.



6transistor-SRAM

Nano-CMOS SRAM Design Challenges

BL

BLB

P1 << N3

P2 << N4

V_{DD} WL

Load

Load

WL V_{DD}

P1

P2

N3

0

V_{DD}

N4

Access

Access

N1 >> N3

N1

N2

N2 >> N4

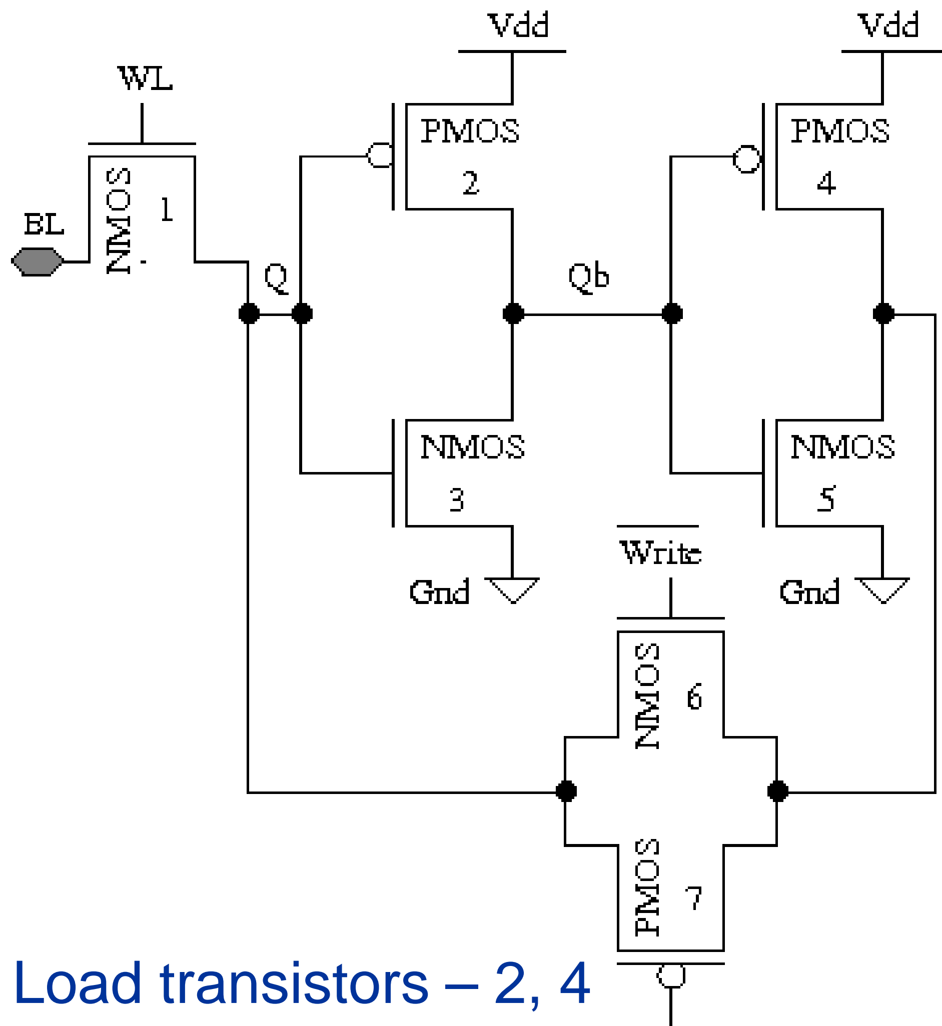
Driver

Driver

- For proper read stability: N1 and N2 are sized wider than N3 and N4.
- For successful write: N3 and N4 are sized wider than P1 and P2.
- Minimum sized transistors do not provide good stability and functionality.
- SRAM cell ratio (β): ratio of driver transistor's W/L to access transistor's W/L.

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Single-Ended 7-Transistor SRAM



Load transistors – 2, 4
Driver transistors – 3, 5
Access transistors – 1, 6, 7

Highlights of this SRAM:

- Single-ended I/O latch style 7-transistor SRAM.
- Functions in ultra-low voltage regime allowing subthreshold operation.
- Better read stability, better write-ability compared to standard SRAM.
- Improved nanoscale process variation tolerance compared to the standard 6-transistor SRAM.

Source: Our publication in SOCC 2008

Research Question

How to reduce power dissipation while maintaining/enhancing stability of SRAM.

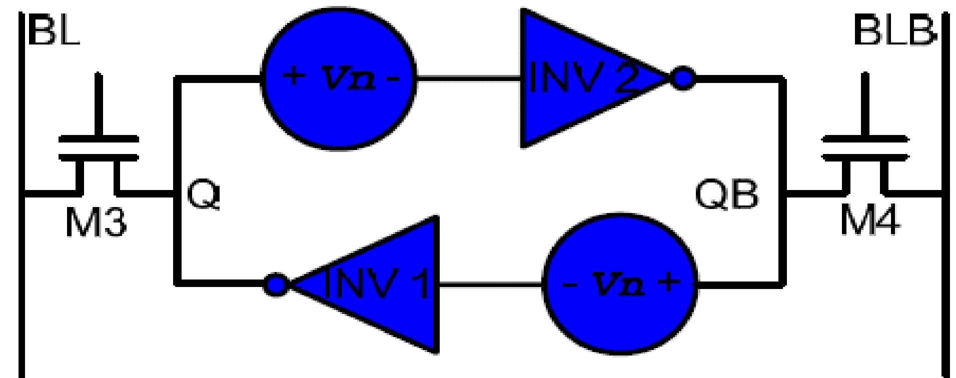
The Solution Explored in This Paper

- To reduce the power consumption this research investigates the process level technique, called dual-V_{th}.
- Important is the selection of appropriate transistors for high-Vth assignment so that performance of SRAM is not degraded.
- SRAM is subjected to the dual-V_{th} assignment using a novel combines Design of Experiments-Integer Linear Programming (DOE-ILP) algorithms.

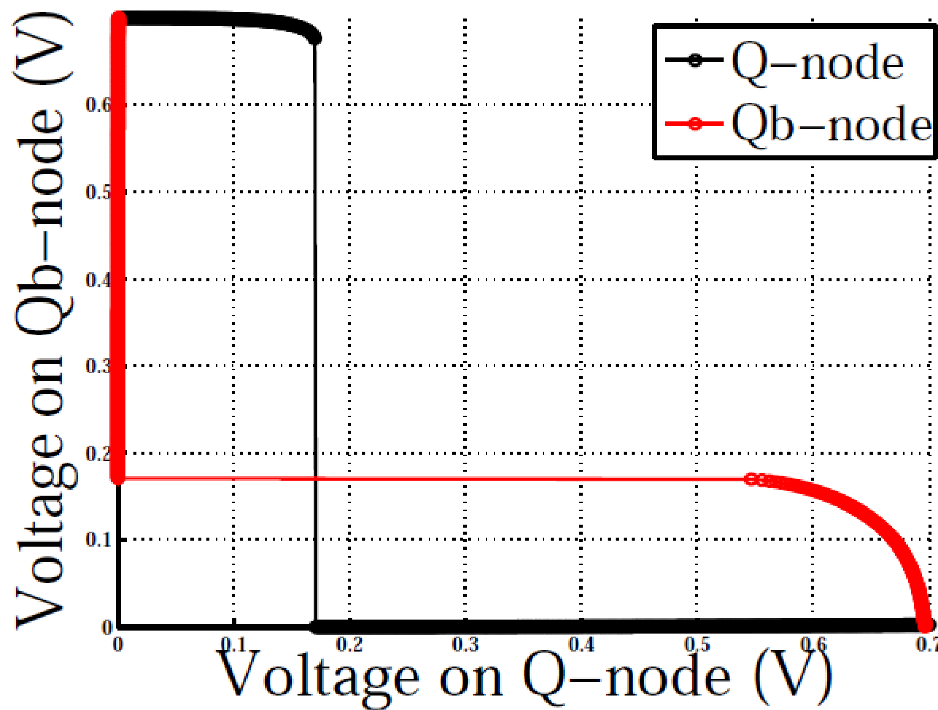
UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Stability Analysis of SRAM: SNM

- Static Noise Margin (SNM): It is the amount of maximum DC voltage (Vn) in this case, that SRAM can tolerate.



Noise model for stability analysis



Butterfly curve for baseline SRAM.

Currents in 7-Transistor SRAM: Write

Current Path for Write '1'

WL '1'
'1' BL
NMOS 1
'1' Q
PMOS 2 'OFF'
'0' Qb
PMOS 4 'ON'
NMOS 3 'ON' Gnd
NMOS 5 'OFF' Gnd
Write
NMOS 6 'OFF'
PMOS 7 'OFF'
Write

dynamic current
gate leakage current
subthreshold leakage current

Current Path for Write '0'

WL '1'
'0' BL
NMOS 1
'0' Q
PMOS 2 'ON'
'1' Qb
PMOS 4 'OFF'
NMOS 3 'OFF' Gnd
NMOS 5 'ON' Gnd
Write
NMOS 6 'OFF'
PMOS 7 'OFF'
Write

dynamic current
gate leakage current
subthreshold leakage current

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Currents in 7-Transistor SRAM: Read

Current Path for Read '1'

Current Path for Read '0'

Parameters	Value
P_{sram}	203.6 nW
SNM_{sram}	170 mV

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

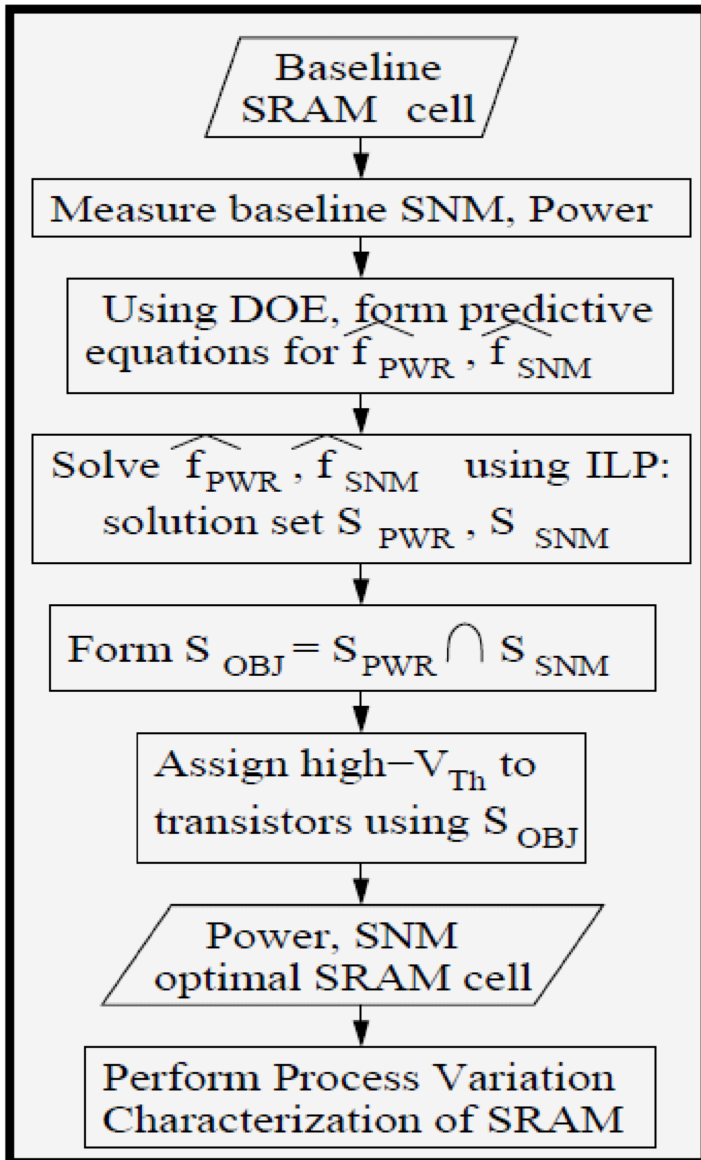
Combined DOE-ILP Approach

- **Design of Experiments (DOE)** consists of purposeful changes of inputs (factors) to a process in order to observe the corresponding changes in the outputs (responses).

- **Integer linear programming (ILP)** is a technique for optimization of a linear objective function, subject to linear equality and linear inequality constraints. ILP determines the way to achieve the best outcome (such as maximum profit or lowest cost) in a given mathematical model and given some list of requirements represented as linear equations.

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Combined DOE-ILP Approach: Solution 1



Design Flow-1

1: Input: Baseline circuit, Nominal/High-V_{Th} models.
2: Output: Objective set $S_{OBJ} = [f_{PWR}, f_{SNM}]$ with transistors identified for high V_{Th} assignment
3: Setup experiment for transistors of SRAM cell using 2-Level Taguchi L-8 array, where the factors are the transistors and the responses are average P_{sram} and read SNM_{sram}.
4: for Each 1:8 experiments of 2-Level Taguchi L-8 array do
5: Perform simulations and record P_{sram} and SNM_{sram}.
6: end for
7: Form predictive equations $\hat{f}_{PWR}$ for power, $\hat{f}_{SNM}$ for SNM.
8: Solve $\hat{f}_{PWR}$ using ILP. Solution set: S_{PWR}.
9: Solve $\hat{f}_{SNM}$ using ILP. Solution set: S_{SNM}.
10: Form $S_{OBJ} = S_{PWR} \cap S_{SNM}$.
11: Assign high V_{Th} to transistors based on S_{OBJ}.

Algorithm -1

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

DOE Predictive Equations

$$\hat{f} = \overline{f} + \sum_{n=1}^{7} \left(\frac{\Delta(n)}{2} \times x_n \right),$$

Where:

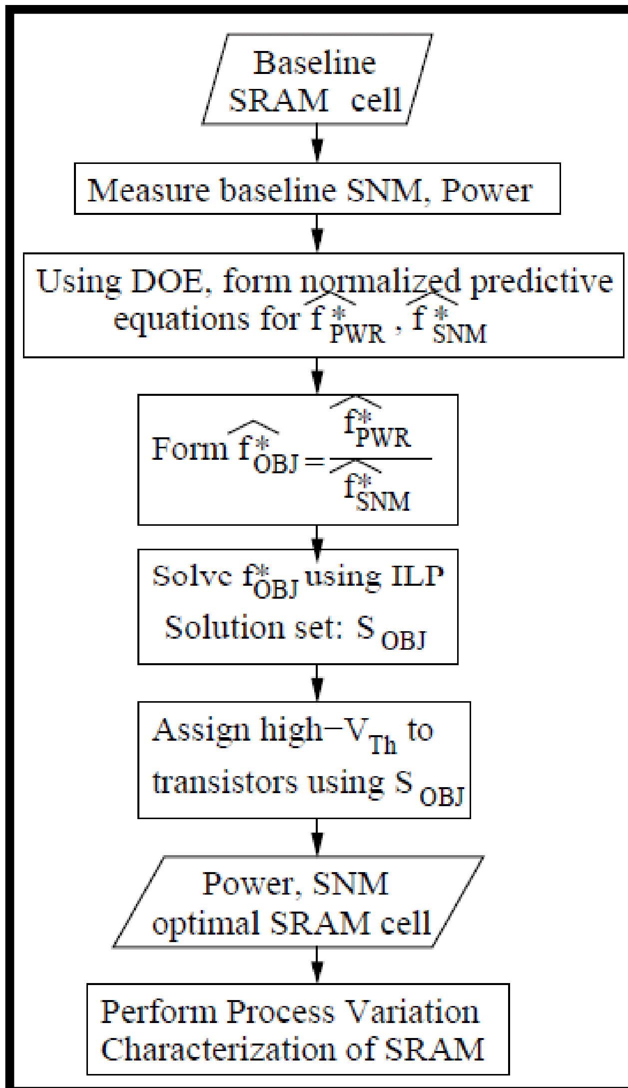
x_n is the V_{Th} -state of transistor of nth transistor ;

$\hat{f}$ is the response of the transistor; (e.g. Power, SNM)

$\left(\frac{\Delta(n)}{2} \right)$ is the half-effect of the nth transistor ;

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Combined DOE-ILP Approach: Solution 2



Design Flow-2

1: Input: Baseline circuit, Nominal/High-V_{Th} models.

2: Output: Objective set $S_{OBJ}* = [f_{PWR}*, f_{SNM}*]$ with transistors identified for high V_{Th} assignment

3: Setup experiment for transistors of SRAM cell using 2-Level Taguchi L-8 array, where the factors are the transistors and the responses are average P_{sram} and read SNM_{sram}.

4: for Each 1:8 experiments of 2-Level Taguchi L-8 array do

5: Perform simulations and record P_{sram} and SNM_{sram}.

6: end for

7: Form normalized predictive equations: $\hat{f}_{PWR}*$ and $\hat{f}_{SNM}*$.

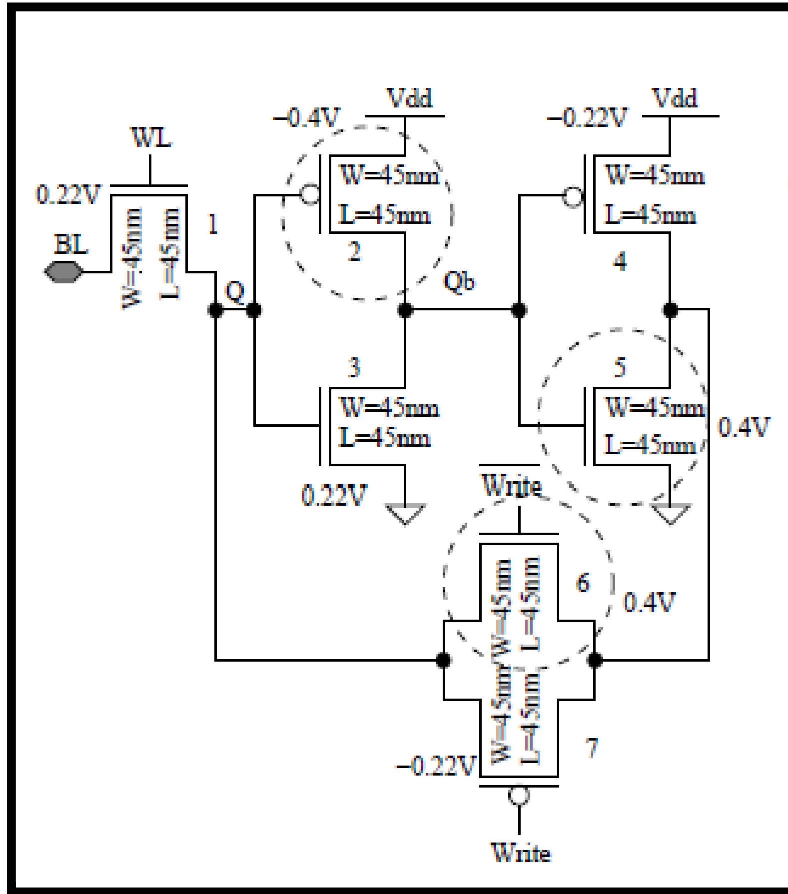
8: Form $\hat{f}_{OBJ}* = \left(\frac{\hat{f}_{PWR}*}{\hat{f}_{SNM}*}\right)$.

9: Solve $\hat{f}_{OBJ}* =$ using ILP. Solution set: $S_{OBJ}*$.

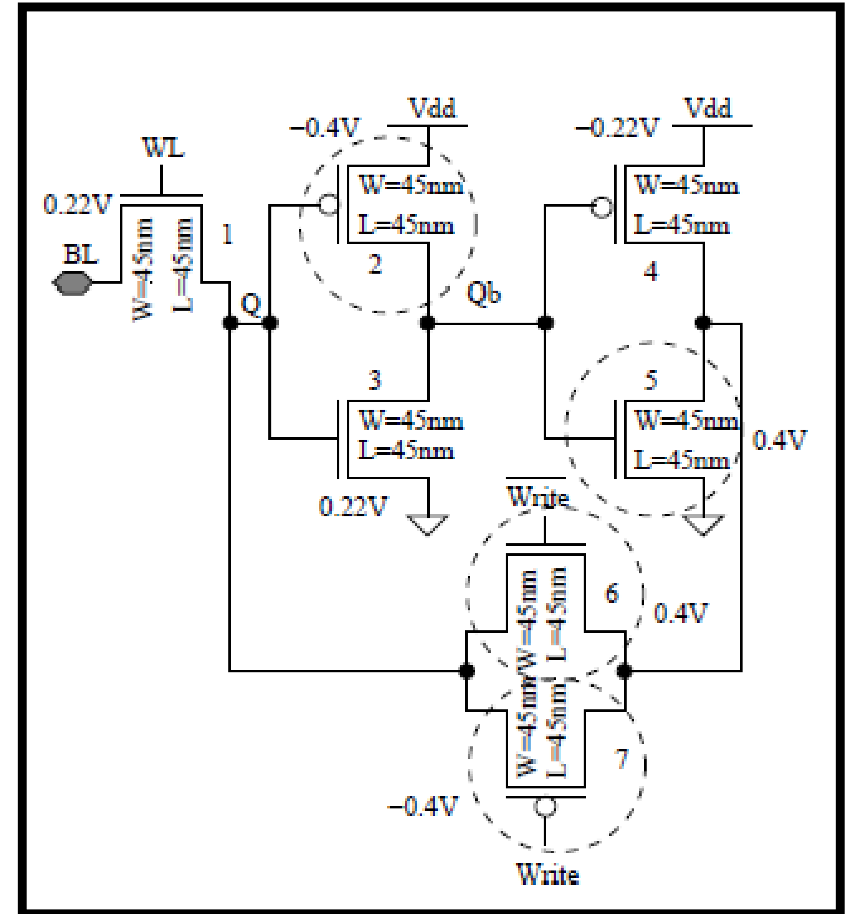
10: Assign high V_{Th} to transistors based on $S_{OBJ}*$.

Algorithm - 2

Selection of Appropriate Transistors



Configuration for flow 1

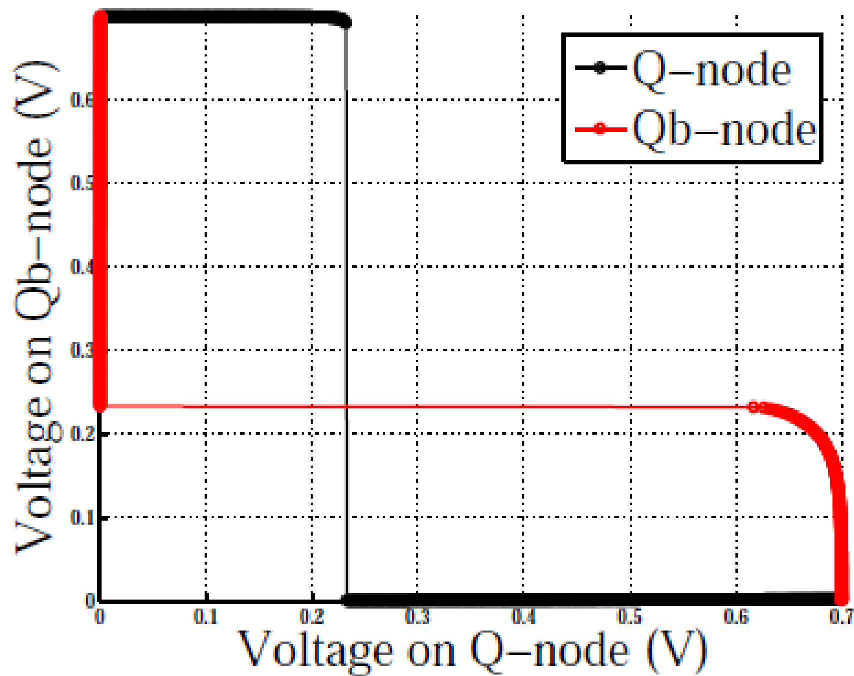


Configuration for flow 2

Experimental Results: 4 Alternatives

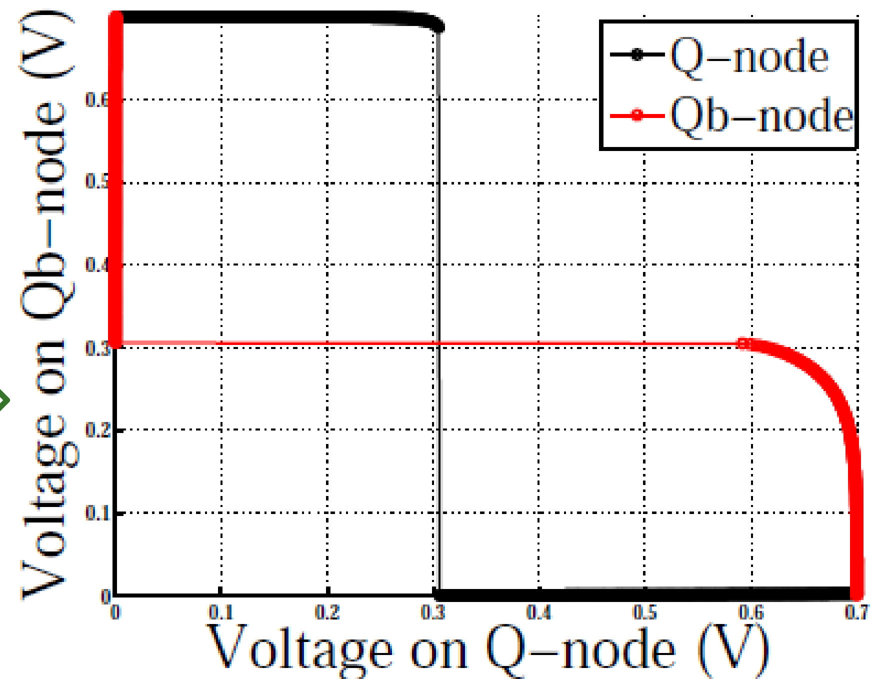
Design Alternative	Parameter	Value	Change
Baseline	P_{sram}	203.6 nW	-
	SNM_{sram}	170mV	-
S_{PWR}	P_{sram}	26.34 nW	87.1%decrease
	SNM_{sram}	231.9 mV	26.7%increase
S_{SNM}	P_{sram}	113.6 nW	44.2%decrease
	SNM_{sram}	303.3 mV	43.9%increase
S_{OBJ} Approach 1	P_{sram}	113.6 nW	44.2%decrease
	SNM_{sram}	303.3 mV	43.9%increase
S_{OBJ} * Approach 2	P_{sram}	100.5 nW	50.6%decrease
	SNM_{sram}	303.3 mV	43.9%increase

Experimental Results: SNM



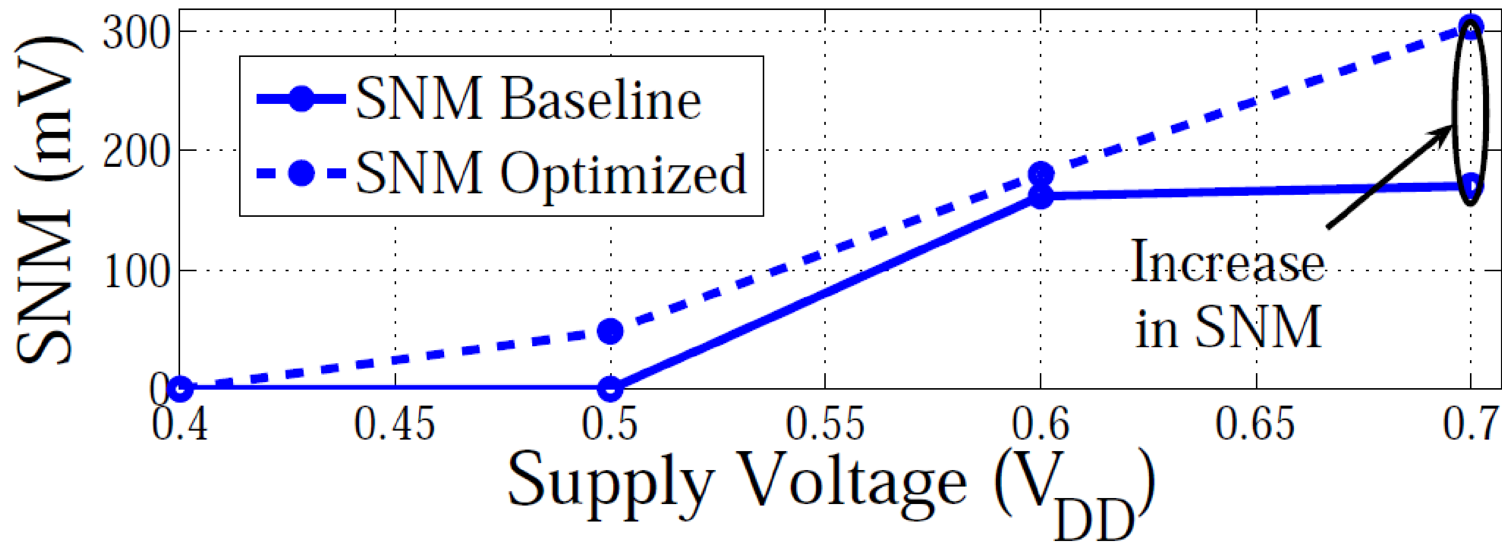
Butterfly curve for reduced power SRAM.

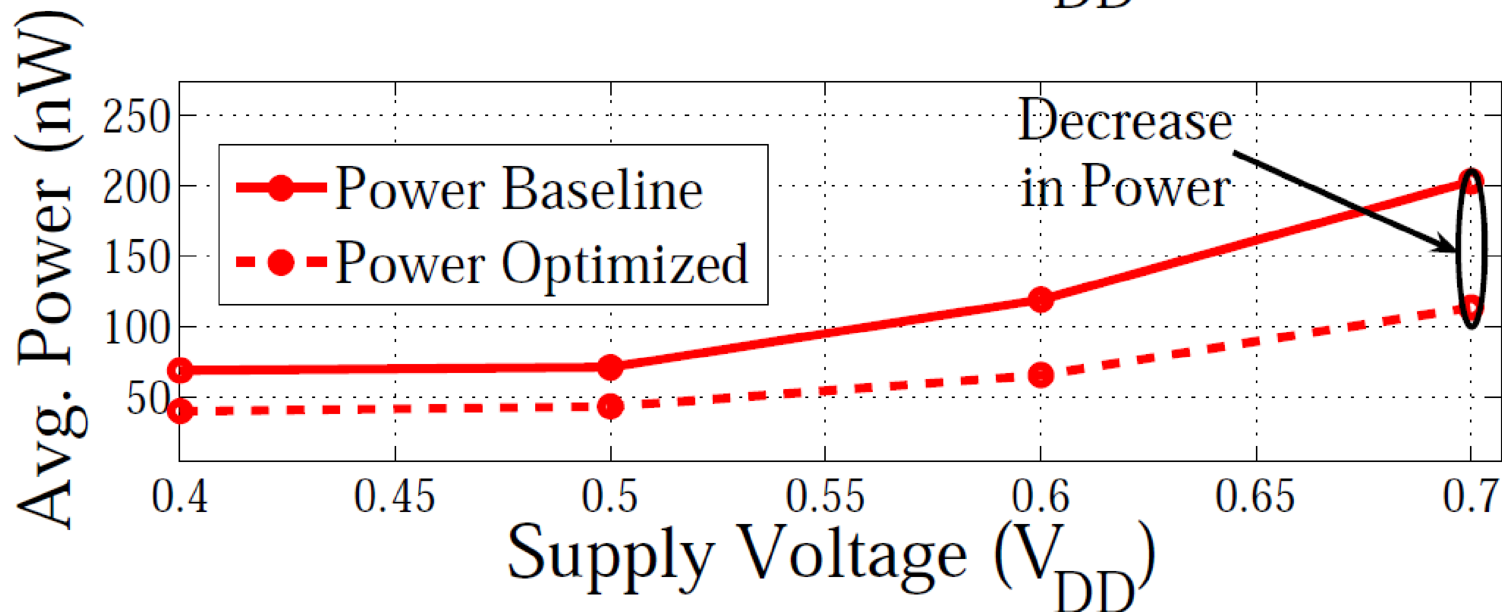
Butterfly curve for the optimal SRAM.



UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Experimental Results: Power/SNM

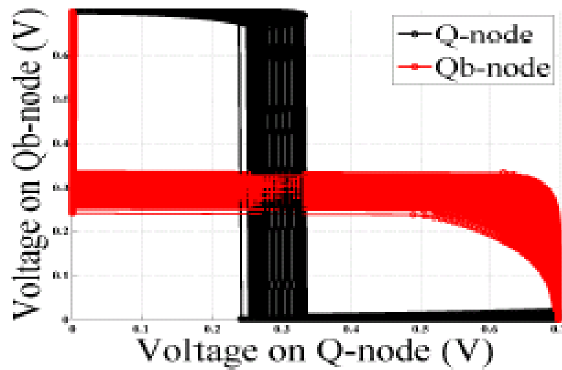


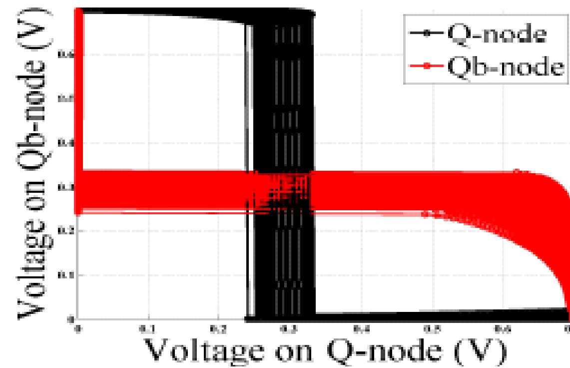


UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Monte-Carlo Distribution Results …

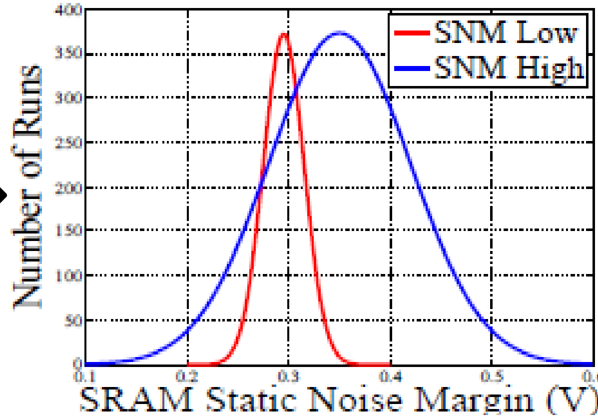
Butterfly curve for Flow 1

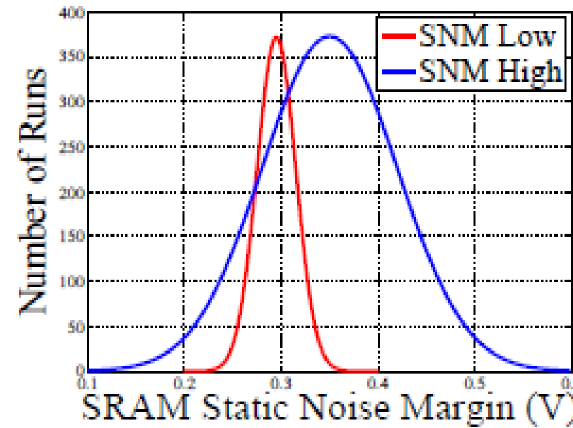




Butterfly curve for Flow 2

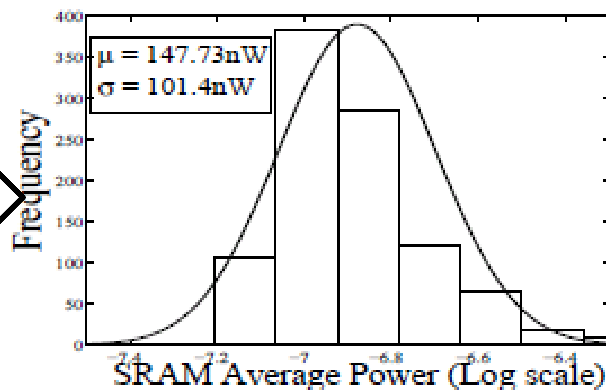
SNM Distribution for Flow 1

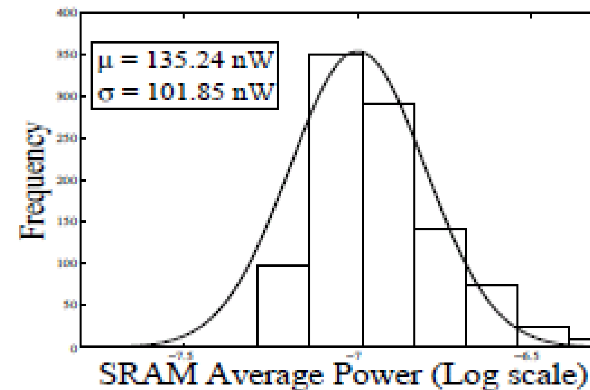




SNM Distribution for Flow 2

Power Distribution for Flow 1





Power Distribution for Flow 2

Monte Carlo Simulation Results

Optimization	Parameter	Mean	Standard deviation
S_{PWR}	P_{sram}	28.91 nW	8.26 nW
	SNM_{sram}	180 mV	30 mV
S_{SNM}	P_{sram}	147.73 nW	101.4 nW
	SNM_{sram}	295 mV	28 mV
S_{OBJ} : $Approach1$	P_{sram}	147.73 nW	101.4 nW
	SNM_{sram}	295 mV	28 mV
S_{OBJ} : $Approach2$	P_{sram}	135.24 nW	101.85 nW
	SNM_{sram}	295 mV	28 mV

Conclusions

- A methodology for simultaneous optimization of SRAM power and read stability is presented.
- A 45nm single ended seven transistor SRAM was subjected to the proposed methodology (novel DOE-ILP algorithms) leading to 50.6% power reduction and 43.9% increase in read stability (read SNM).
- The effect of process variation of twelve process parameters on the SRAM is evaluated, and it is found to be process variation tolerant.
- A 8 × 8 array has been constructed using the optimized cells whose average power consumption is 4.5μW.

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

Future Research

- Future research will involve SRAM-array optimization where variability will be accounted in flow.
- Along with the states of transistors, the sizes will also be considered which will increase the solution space of the algorithms.
- In addition to the power, performance and process variation, thermal effects will also be taken into account.

Thank you !!!