A DOE-ILP Assisted Conjugate-Gradient Approach for Power and Stability Optimization in High-κ/ Metal-Gate SRAM

G. Thakral[1], S. P. Mohanty[1], D. Ghai[1], and D. K. Pradhan[2]
Department of Computer Science and Engineering,
University of North Texas, USA.
Department of Computer Science, University of Bristol, UK[2]
Email-ID: saraju.mohanty@unt.edu

Acknowledgments: This research is supported in part by NSF award CNS-0854182.

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

Outline of the Talk

- Introduction
- Novel Contributions of this paper
- Related Prior Research
- Proposed Flow for Optimal Design of High-κ NANO-CMOS SRAM
- Optimization methodologies for 10 Transistor SRAM
- Optimized Results
- Conclusions

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

NSDL
NanoSystem Design Laboratory

University of BRISTOL

Introduction

- Static Random Access Memories arrays are widely used as cache memory in microprocessors and application-specific integrated circuits occupy a significant portion of the die area.

  In January 2010, a leading edge IC contained approximately 2 billion transistors.
- The process technology scaling and push for better performance enabled embedding of millions of SRAM cells into contemporary Integrated Circuits (ICs).
- In an attempt to optimize the power consumption/performance/cost ratio of such chips, designers are faced with a dilemma.

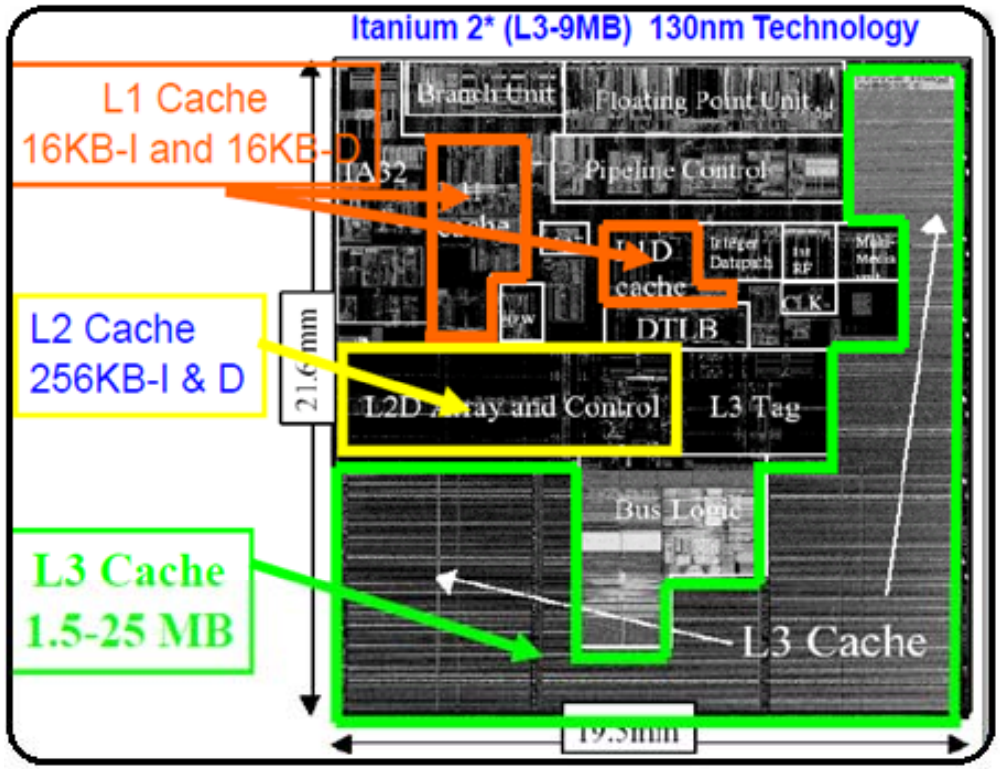
UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

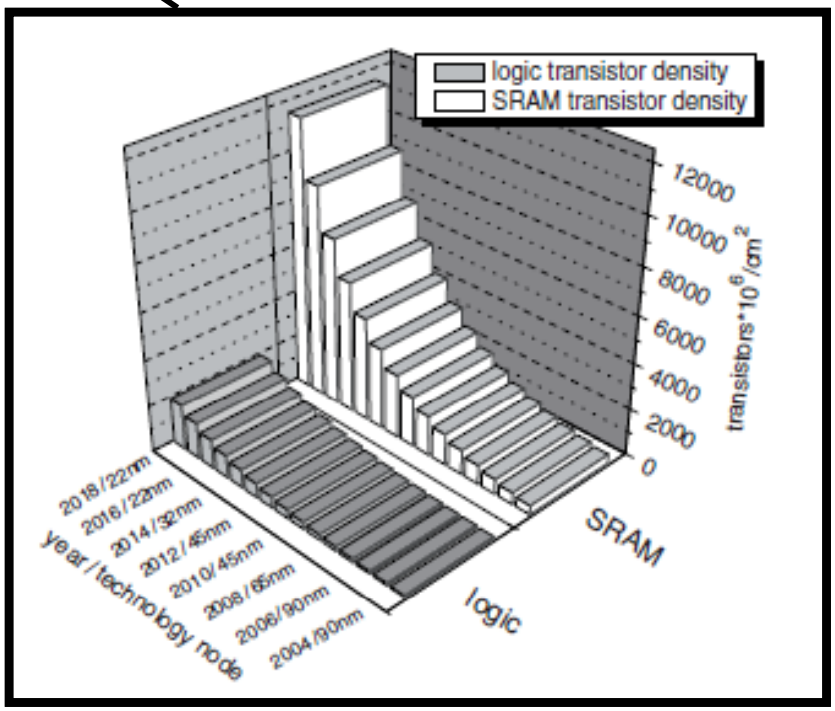
Motivation For SRAM Research

- Millions of minimum-size SRAM cells are tightly packed
- Such areas on the chip can be especially susceptible and sensitive to manufacturing defects and process variations.
- The stability is a growing concern in the design as the process technology continues to scale deeper
- Up 70% of die area is occupied by cache
- To meet performance and throughput requirements

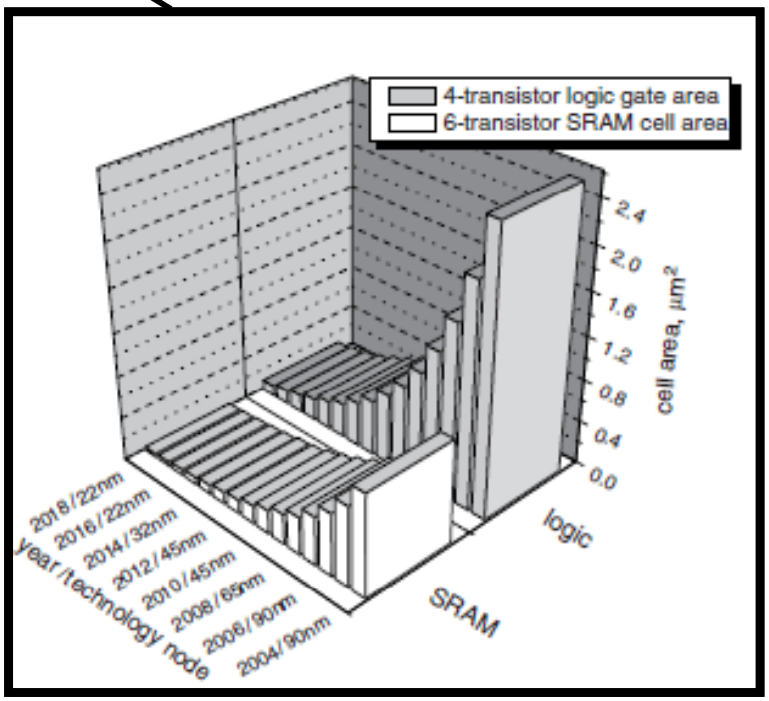


Challenges: A Glimpse

Transistor density trends with scaling: 6T SRAM cell vs. 4T logic gate



Area trends with scaling: 6T SRAM cell area vs. a 4T logic gate



Source: Process-Aware SRAM Design and Test. Authors:Andrei Pavlov & Manoj Sachdev

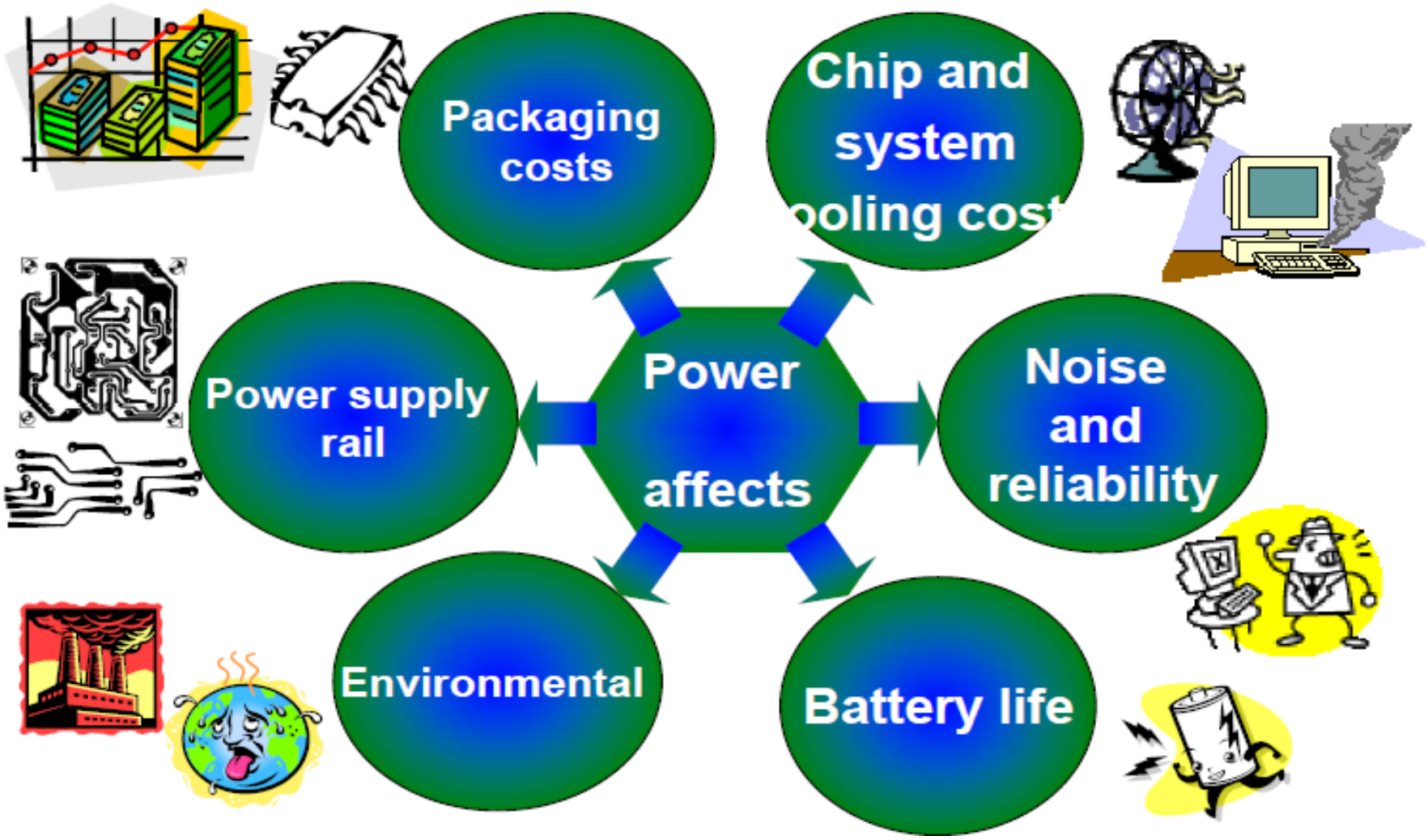
UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

Design Challenges for SRAM

Why Low Power?



UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

NSDL
NanoSystem Design Laboratory

University of BRISTOL

Related Research: SRAM

Reference	Optimization Technique	Power	Performance (SNM)	Number of Transistors
Amelifard et al (2006)	Dual-VTh and dual-Tox	53.5 % decrease	43.8% increase	6
Agrawal et al (2006)	Modeling Based Approach	-	160mV (approx.)	6
Lin et al (2008)	Write bitline balancing circuitry	5 nW (standby)	310 mV	9
Okumura et al (2009)	Column Line Assist Scheme	-	360 mV	10
This research (2010)	**DOE-ILP Assisted Conjugate-Gradient**	**314.5 nW 86% decrease (Total Power)**	**295 mV 8% increase**	**10**

Contributions of This Paper

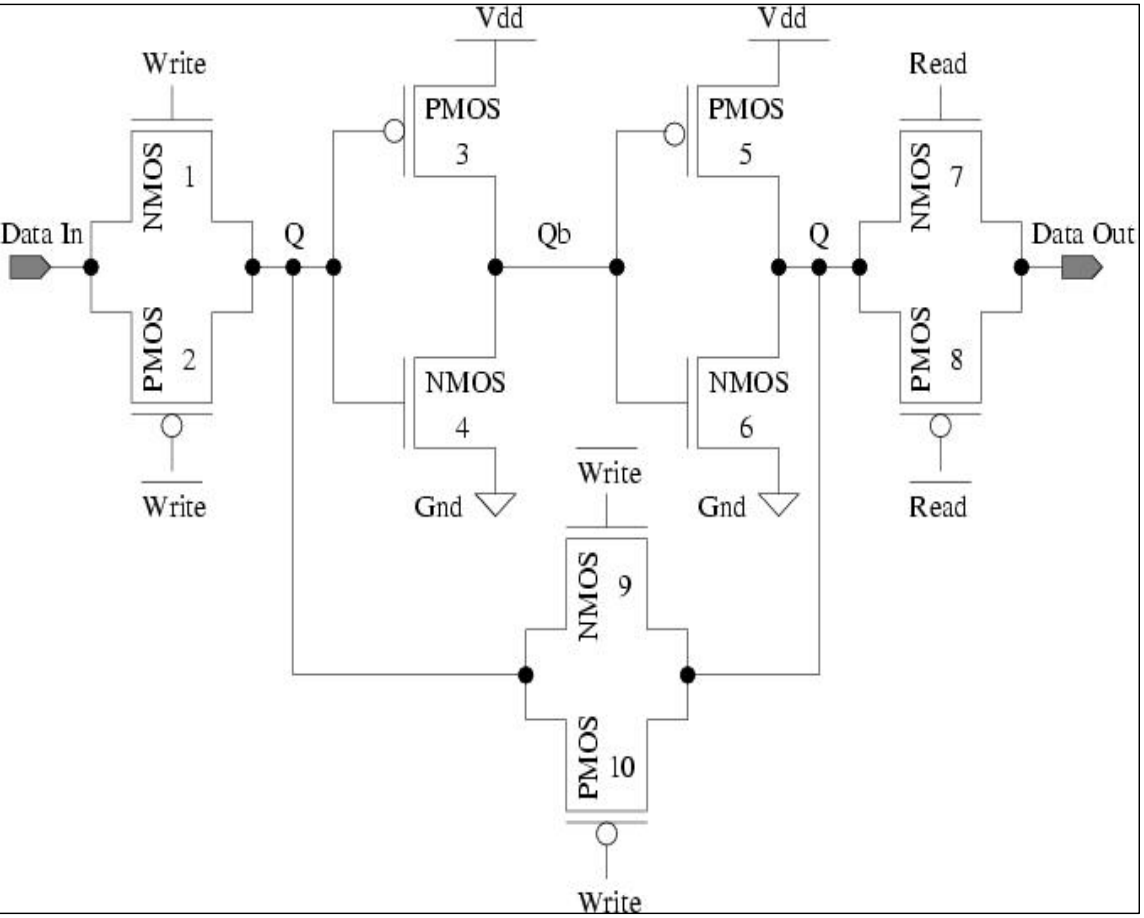
- A novel design flow is proposed for power minimization and stability maximization in nano-CMOS SRAM circuits.
- A high-κ/metal-gate 32nm 10-transistor SRAM is subjected to this methodology to show it's effectiveness.
- A novel DOE-ILP based approach is proposed for power minimization in a SRAM circuit.
- A conjugate-gradient based algorithm is proposed for SNM maximization of the SRAM.
- Process variation analysis for robustness to study the SRAM.
- An 8 × 8 array is constructed using optimal SRAM cells.

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

High-*K* based 10-TRANSISTOR SRAM



Highlights of 10T SRAM

- Two inverters connected back to back in a closed loop fashion in order to store the 1-bit information

- Three transmission gates read, write and hold states, instead of access transistors used in the traditional 6-transistor SRAM

- Transmission gates carefully input and output the data to/ from the cell node Q at full logic level.

- This provide full swing during write and read operation.

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

High-*K* NANO-CMOS SRAM Models

1. For the design and simulation of SRAM presented in this research, a 32nm high-κ/metal-gate CMOS PTM is used.
2. For the PTM based on BSIM4/5, two methods are adopted:

- The model parameter in the model file that denotes relative permittivity (EPSROX) is changed.
- The equivalent oxide thickness (EOT) for the dielectric under consideration is calculated.
- The total power of a nano-CMOS circuit is defined as:

$$P_{total} = P_{dynamic} + P_{subthresho\,ld}$$

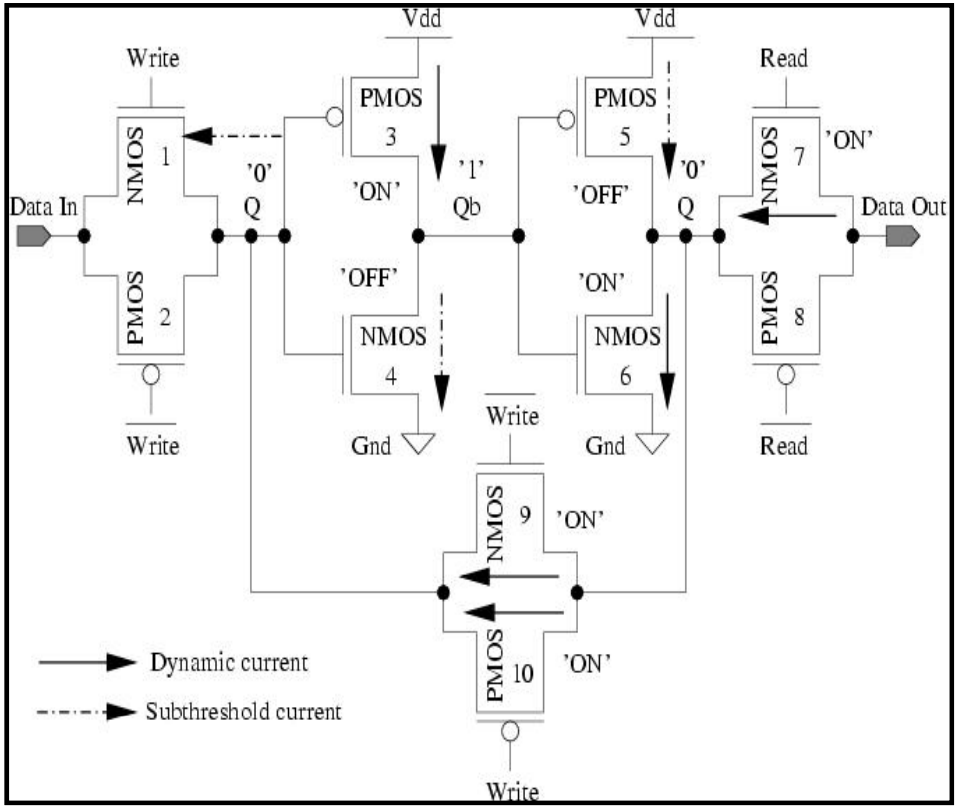
- The use of high-*κ metal-gate technology eliminates the gate leakage in SRAM.*

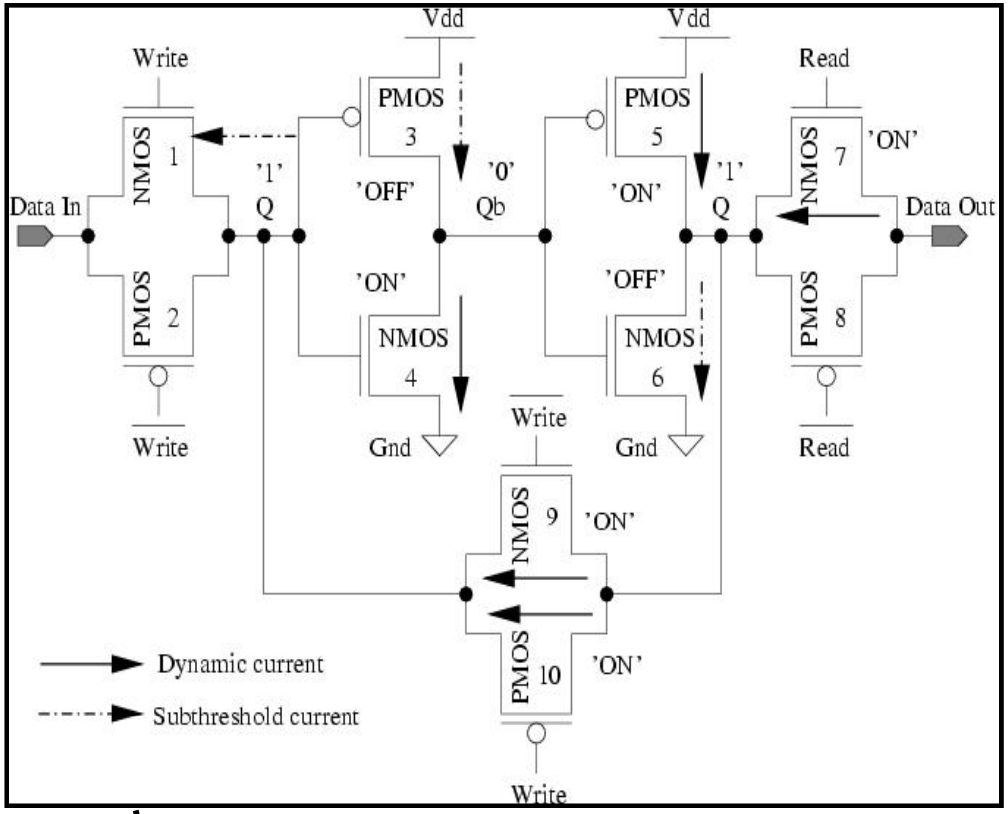
UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

Operations of Proposed SRAM





Current path for Read '1'

Current path for Read '0'

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

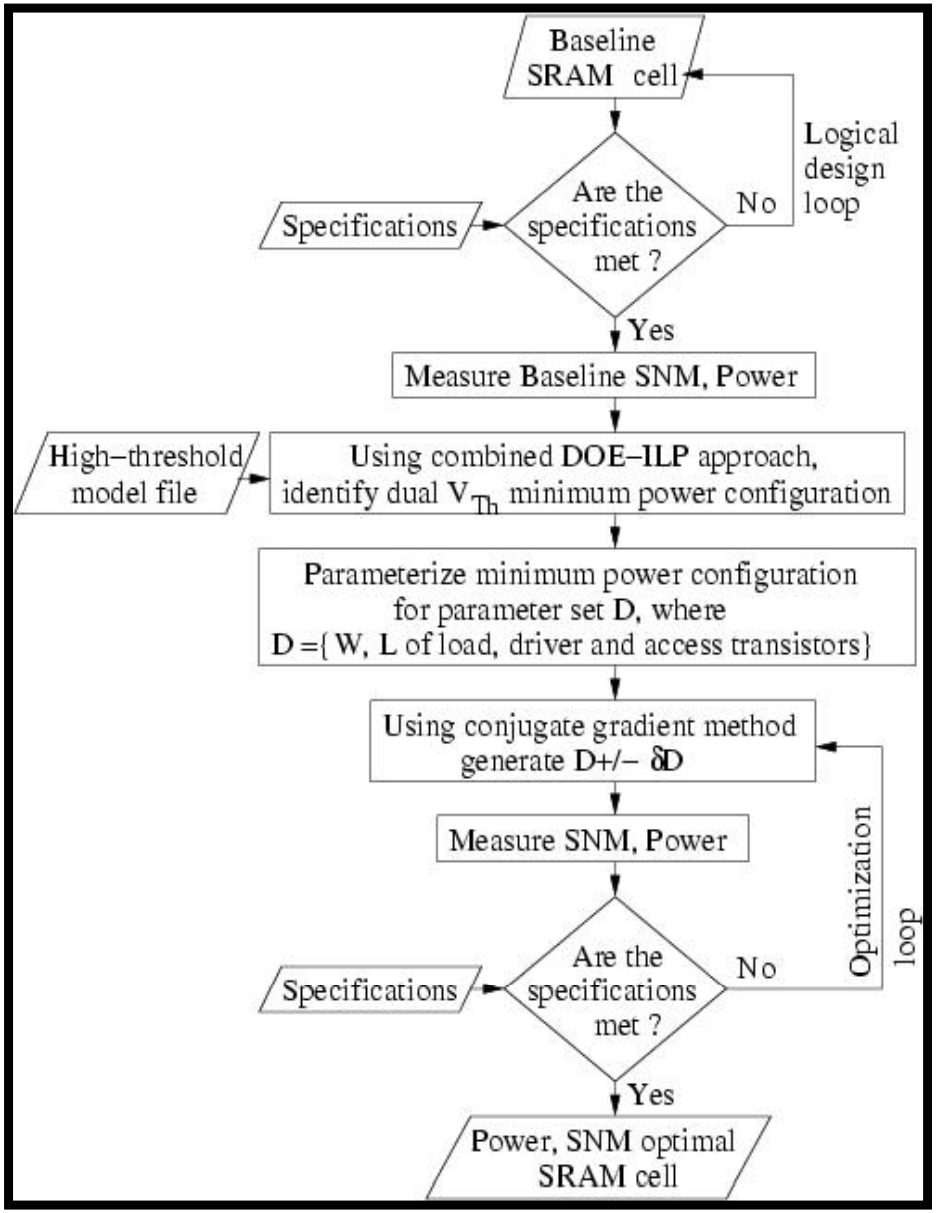
University of BRISTOL

Proposed Flow for Optimal Design of High-*K* NANO-CMOS SRAM

- Dual-V_{Th} voltage technique has strong impact on power dissipation and SNM of the SRAM.

- This is performed using a DOE-ILP based approach

- ILP is used to the linear equations which ensures minimum power SRAM cell configuration.



Proposed Flow for Optimal Design of High-*K* NANO-CMOS SRAM contd…

- However, this results in degradation in the stability (SNM) of the SRAM.

- To improve the stability of the SRAM, the minimum-power configuration SRAM is subjected to the conjugate-gradient based optimization loop for SNM maximization

- The parameter set for optimization includes the widths and lengths of the access, load and driver transistors of the SRAM cell.

- The output of this optimization loop is a highly stable SRAM cell, which consumes minimum power and better performance.

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

Optimization methodologies for 10-Transistor SRAM

DOE-ILP Approach for Minimum Power/Leakage Configuration

- Approach that uses both DOE and ILP is deployed for power minimization of the SRAM.
- Design of Experiments based approach is implemented using a 2-Level Taguchi L-12 array.
- The factors are the V_{Th} *states* of 10 transistors of the SRAM cell, and the response under consideration is the average power consumption of the cell (f_{PSRAM}).

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

DOE-ILP Approach contd…

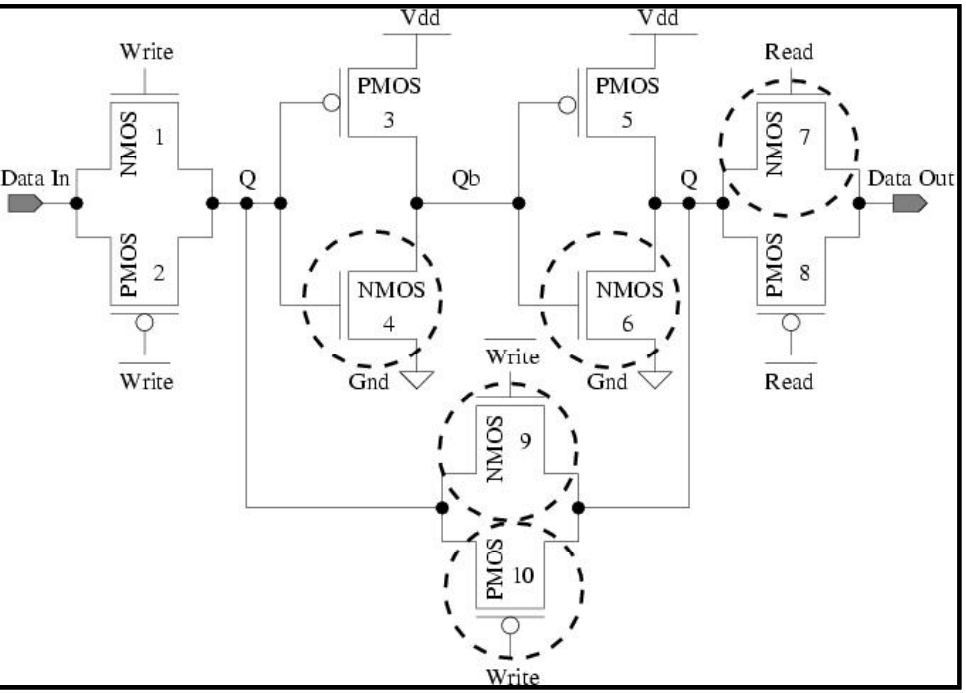
Equation: $\left(\frac{\partial(n)}{2}\right) = \left(\frac{avg(+1) - avg(-1)}{2}\right)$

where: $\left(\frac{\partial(n)}{2}\right)$ is the half effect of nth transistor

avg(+1) avg power when transistor n is in high-V_{th} state.
avg(-1) avg power when transistor n is in low-V_{th} state.

- Using other methods like full factorial would take 2^{10} = 1024 runs, whereas the L-12 Taguchi array requires 12 runs.

DOE-ILP Approach for Minimum Power/Leakage Configuration contd…



Results of baseline SRAM

Parameters	Values
P_{SRAM}	2.27 µW
SNM_{SRAM}	271 mV

Minimum power configuration results

Parameters	Values
P_{SRAM}	314.5 nW
SNM_{SRAM}	230.4 mV

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

Conjugate-gradient approach

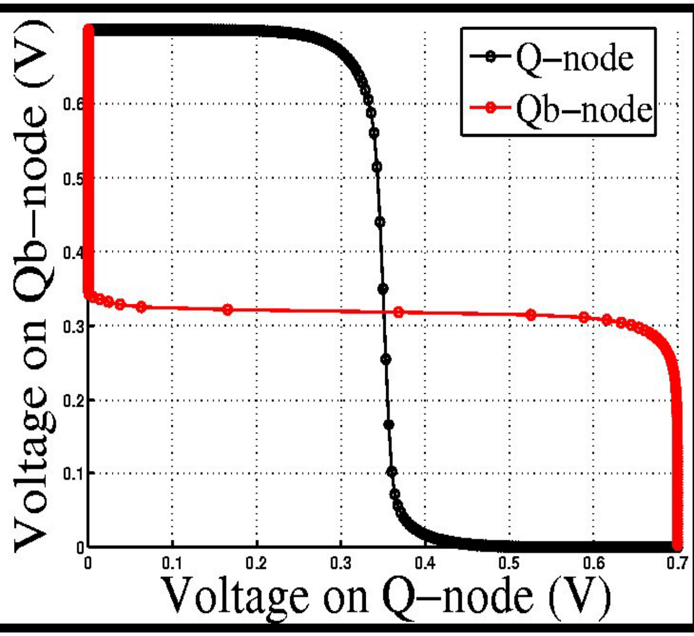
- **Input:** Minimum power configuration SRAM, Baseline model file, High-threshold model file, Objective Set $F = [SNM_{SRAM}, P_{SRAM}]$, Stopping Criteria S, parameter set $D = [W_{pl}, L_{pl}, W_{nd}, L_{nd}, W_{pa}, L_{pa}, W_{na}, L_{na}]$, Lower parameter constraint C_{low}, Upper parameter constraint C_{up}.
- **Output:** Optimized objective set F_{opt}, Optimal parameter set D_{opt} for $S \leq \pm \beta$. $\{1\% \leq \beta \leq 5\%\}$
- Run initial simulation with initial guess of D.
- **while** $(C_{low} < D < C_{up})$ **do**
  Use Conjugate gradient method to generate new set of parameters $D' = D \pm \Delta D$
- Compute $F = [SNM_{SRAM}, P_{SRAM}]$.
- **if** $(S \leq \pm \beta)$ **then**
- **return** $D_{opt} = D'$.
- **end if**
- **end while**
- Using D_{opt}, simulate the optimal SRAM.
- Record F_{opt} for the optimal SRAM.

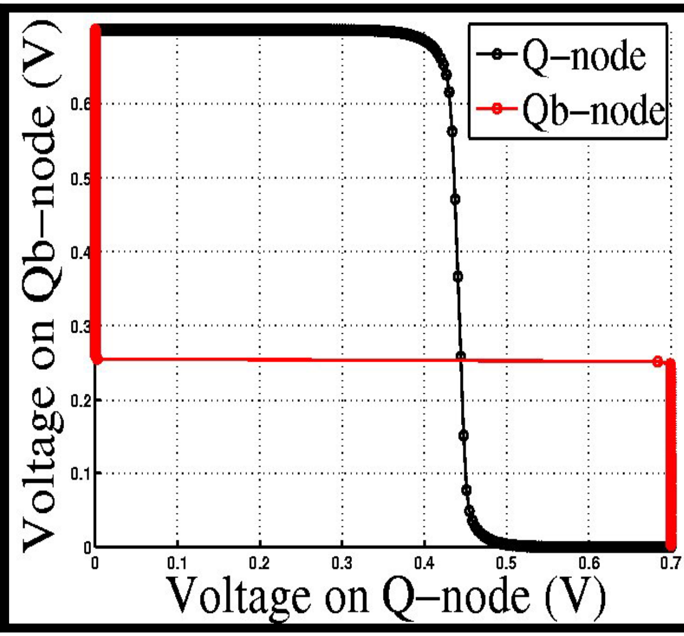
UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

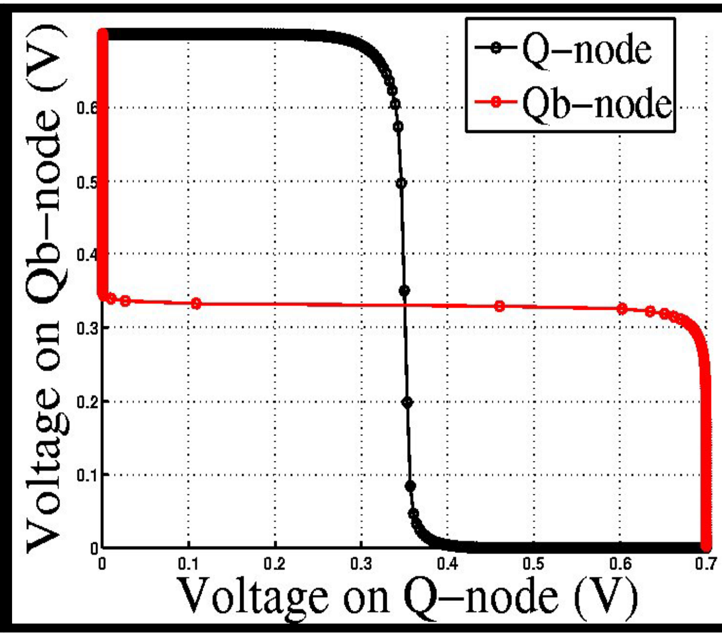
NSDL NanoSystem Design Laboratory

University of BRISTOL

Conjugate-gradient approach…







Baseline

Power Optimal

Power/SNM Optimal

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

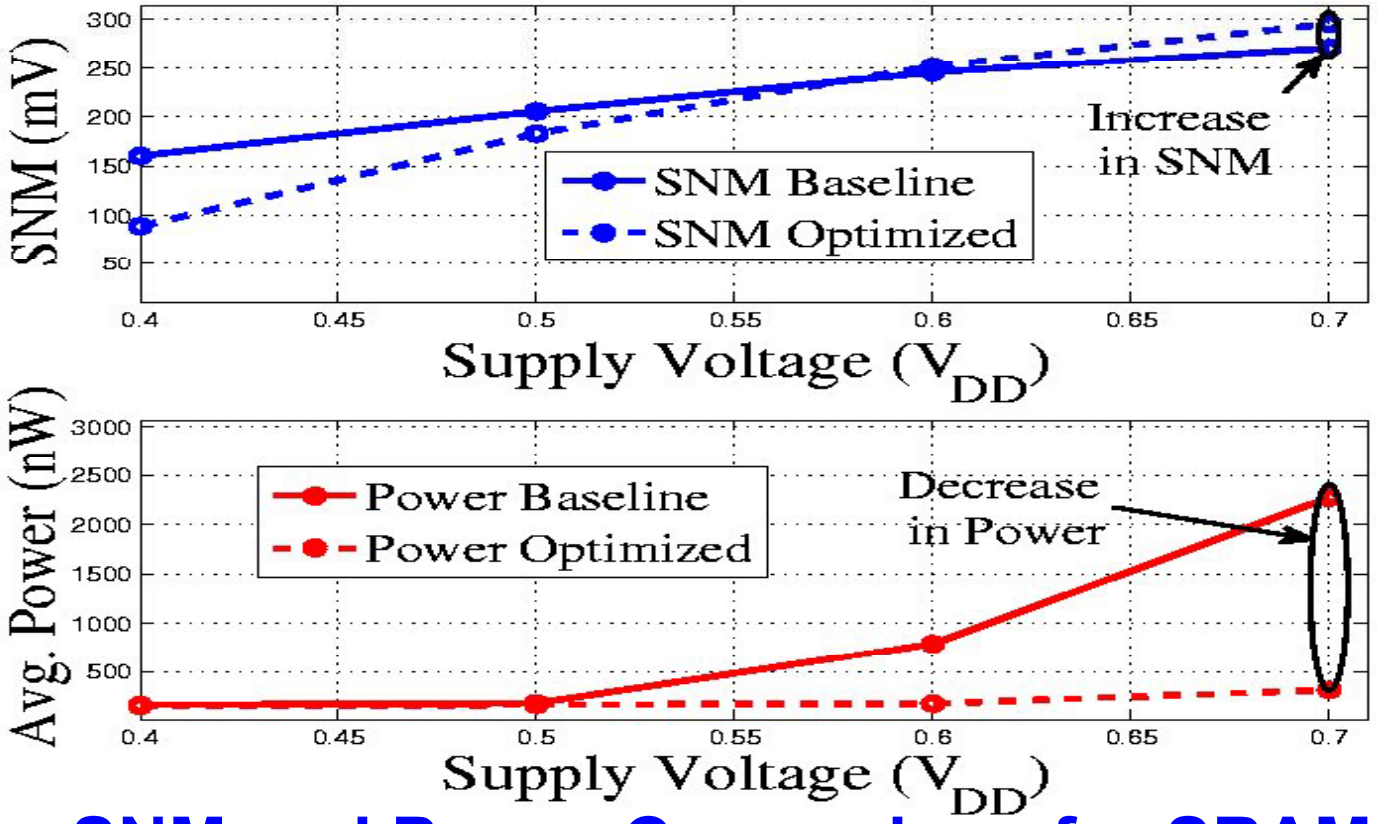
NSDL
NanoSystem Design Laboratory

University of BRISTOL

Optimization Results for Power, Performance and Process Variation

SRAM results after Optimization

Parameters	Values
P_{SRAM}	314.5 nW
SNM_{SRAM}	295 mV



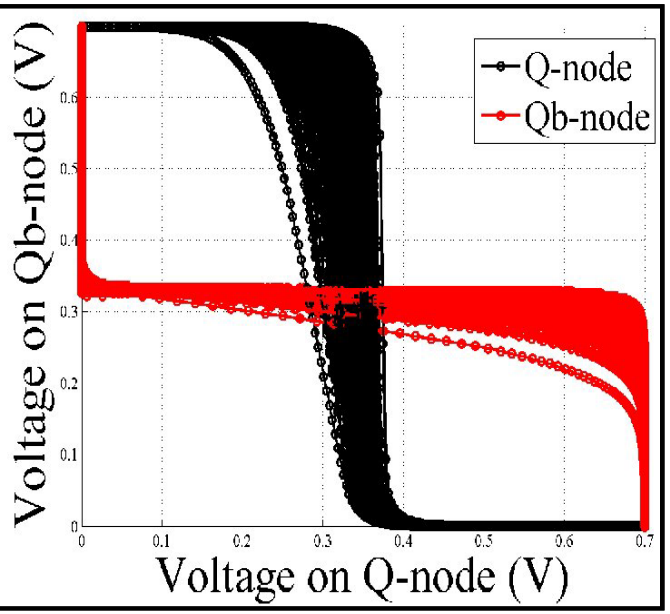
SNM and Power Comparison for SRAM

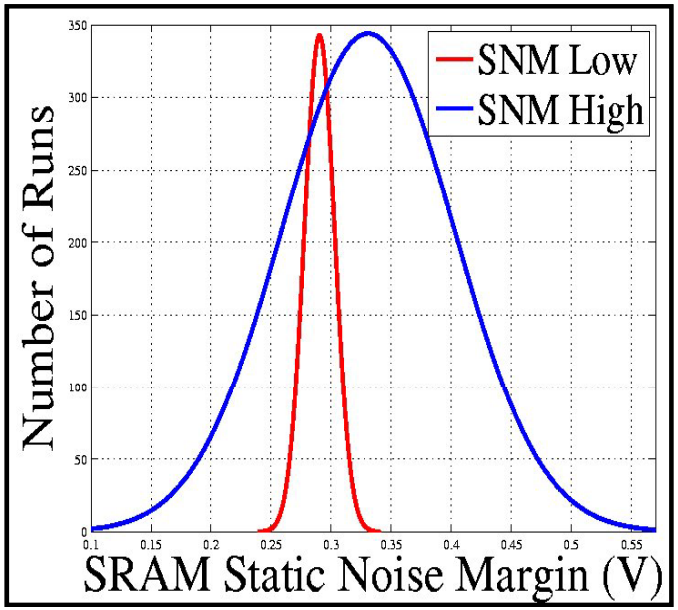
UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

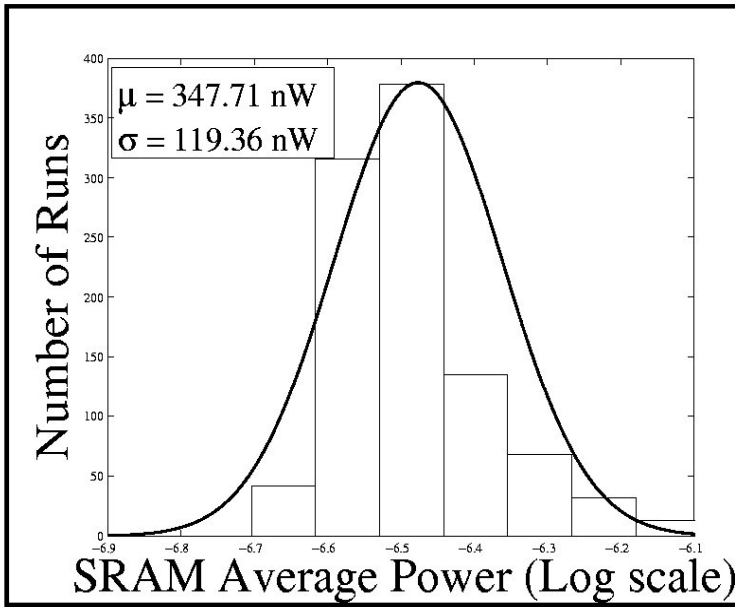
NSDL NanoSystem Design Laboratory

University of BRISTOL

Process Variation for 10T SRAM







Effect of process variation on the butterfly curve of SRAM

Distribution of "High SNM" and "Low SNM"

Distribution of average power of SRAM

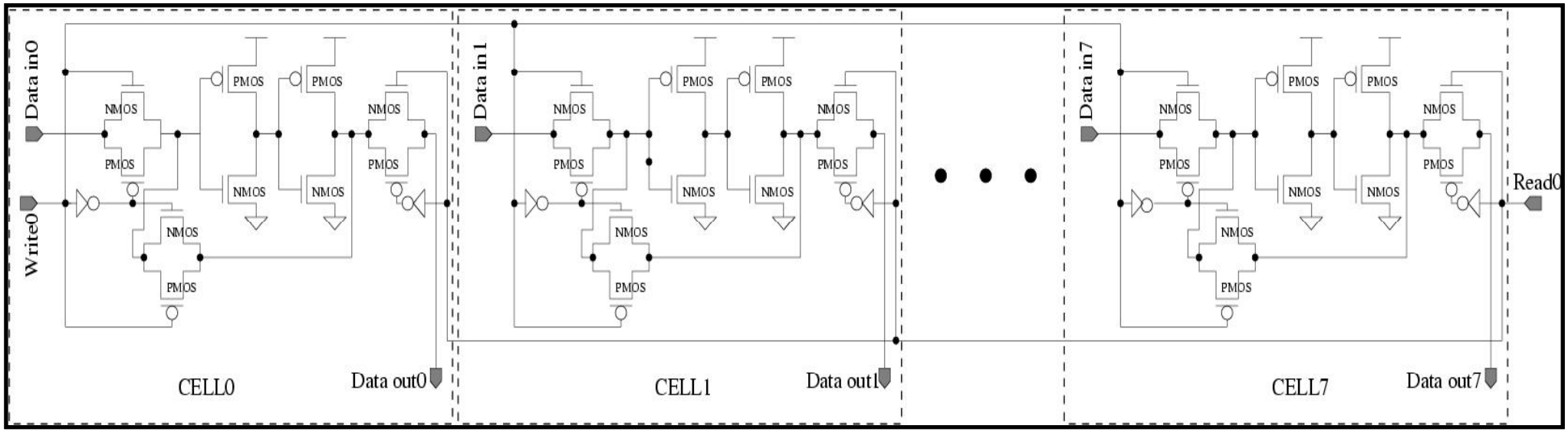
SNM Value	μ (mV)	σ (mV)
SNM High	330.7	71.9
SNM Low	290.3	12.7

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

Array organization for 10T SRAM



- As per the design flow, an 8 × 8 array is constructed using the optimized cell
- The average power consumption of the array is 1.2 μW

Conclusions and Future Work

- A methodology is presented for cell-level optimization of SRAM power and stability.

- A 32nm high-κ metal gate 10-transistor SRAM is subjected to the proposed methodology which has shown 86% reduction in power and 8% increase in SNM.

- A novel DOE-ILP approach has been used for power minimization, and conjugate gradient method is used for SNM maximization.

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

NSDL
NanoSystem Design Laboratory

University of BRISTOL

Conclusions and Future Work …

- The effect of process variation of 12 parameters on the proposed SRAM is evaluated.

- A 8 × 8 array has been constructed using the optimized cell and data for power and read static noise margin is presented.

- The future scope of this research involves array-level optimization of SRAM.

- For array optimization, both mismatch and process variation will be considered as part of the design flow.

UNT UNIVERSITY OF NORTH TEXAS Discover the power of ideas

NSDL NanoSystem Design Laboratory

University of BRISTOL

UNT
UNIVERSITY OF NORTH TEXAS
Discover the power of ideas

NSDL
NanoSystem Design
laboratory

University of
BRISTOL