Lecture 12: Efficient SRAM Circuit Design

CSCE 6933/5933 Advanced Topics in VLSI Systems

Instructor: Saraju P. Mohanty, Ph. D.

NOTE: The figures, text etc included in slides are borrowed from various books, websites, authors pages, and other sources for academic purpose only. The instructor does not claim any originality.





Outline

➤ Introduction

- Different SRAM topology
- Different SRAM Figures of Merits
- Proposed Optimal SRAM Design Flows
- Optimal Design of SRAMs





Issues in Nano CMOS







Technology Scaling: Nano-Regime

Process variations affect:

- L: Channel Length
- T_{ox}: Gate Oxide Thickness

100X

Normalized IOFF

150nm, 110°C

10

- V_{th}: Threshold Voltage
- # Dopant Atoms

NMOS

PMOS

2X

0.1

1.4

1.2

1.0

0.8

0.6

0.4

0.01

Normalized Ion





Why Efficient SRAM Design?

Cache (MB

Die

180nm

130nm

130nm

Amount of on-die caches increases

UNIVERSITY OF NORTH TEXAS Discover the power of ideas

- 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

SRAM Challenges ...



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





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



- 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.





Prior Research on SRAM





Related Prior Research in SRAM



Advanced Topics in VLSI Systems

UNIVERSITY OF NORTH TEXAS Discover the power of ideas

Related Prior Research in SRAM ...



UNIVERSITY OF NORTH TEXAS Discover the power of ideas

Advanced Topics in VLSI Systems

SRAM Circuit Topologies





Traditional 6T SRAM







7T SRAM Circuit







Single-Ended 7-Transistor SRAM



Highlights of this SRAM:

•Single-ended I/O latch style 7transistor SRAM.

•Functions in ultra-low voltage regime allowing subthreshold operation.

•Better read stability, better writeability compared to standard SRAM.

•Improved nanoscale process variation tolerance compared to the standard 6-transistor SRAM.

Source: Our publication in SOCC 2008





10T SRAM Circuit







Figures of Merit: Total Power (including leakage) and Static Noise Margin (SNM)





Stability Analysis of SRAM: Performance Metric

• Static Noise Margin (SNM): Minimum DC voltage which is required to flip the state of the SRAM cell during the read/write operation.





Stability Analysis of SRAM ...



Static Noise Margin (SNM): The maximum DC noise voltage $V_{\rm n}$ that can be tolerated by SRAM.





Stability Analysis of SRAM







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.





(a) For Baseline

(b) SPWR-Optimal

(c) SOBJ-Optimal





Power Dissipation in CMOS



$$P_{dynamic} = P_{trans - tunn} + P_{cap - switch} + P_{short - circuit}$$

 $Static = P_{steady - tunn} + P_{subthreshold} + P_{reverse - biased}$

Both dynamic and static power are significant fractions of total power dissipation in a nano-scale CMOS circuit.





Accurate Power Analysis: Problem Statement



 $P_{dynamic} = P_{trans-tunn} + P_{cap-switch} + P_{short-circuit}$ $P_{static} = P_{steady-tunn} + P_{subthreshbd} + P_{reverse-biased}$

 Both dynamic and static power are significant fractions of total power dissipation in a nanoscale CMOS circuit.





Average Power in Operation: Write/Read/Hold Mode:



Where,
$$V_{DD}$$
 = supply voltage
 I_{gate} =is the current associated with tran-tunn or
steady-tunn i.e., gate leakage
 I_{ds} =contributes to the capacitive switching in

some devices and subthreshold leakage in others.





Significant Leakages in NMOS and PMOS devices



(a) Capacitive-Switching Power

(b) Subthreshold Leakage

(c) Gate-oxide Leakage





Gate Leakage Current Analysis







Example Circuits: Six and Seven transistor SRAM Cell







Current paths for the 6T SRAM: Write



(b) Current path for Write "1"





Advanced Topics in VLSI Systems

Current paths for the 6T SRAM: Read



(c) Current path for Read "0"



(d) Current path for Read "1"





Advanced Topics in VLSI Systems

Current paths for the 6T SRAM: Hold



(f) Current path for Hold "1"

Word Line

gate leakage current

-subthreshold leakage current





Advanced Topics in VLSI Systems

Current paths for the 6T SRAM: Hold



Currents in 7-Transistor SRAM: Write







Currents in 7-Transistor SRAM: Read







Current Paths for 10-Transistor SRAM





Advanced Topics in VLSI Systems



Single Ended I/O 8T-SRAM





The Proposed SE-SRAM

- Proposed single ended I/O 8T-SRAM cell design.
- In word oriented design it becomes 6T- SRAM design.
- Minimum size of transistors are used.
- Read Stability: For 0 and V_{dd}.
- No ratio contention.
- 3 signals: W, W0, R; W0 = W.
 Read operation: R, Write operation: W and W0.



Reduction in dynamic power and leakage because of single ended input/output line and stacking of transistors, respectively.





32-Bit Word Organization Using SE-SRAM



- Word oriented design to reduce area and power overhead.
- 6T-SRAM cell with 2T shared among the word cells.
- Read/Write assist transistors are shared by all bits of a word as all 32 bits are accesses simultaneously.
- Wider word will provide better area saving.





Physical Design of a Proposed 32-bit Word



- Bitcell Area: 0.68μm² (0.55μm x 1.22μm).
- 8% higher than standard 6T SRAM.

4-bit array shown for clarity

- Read/write assist transistors half roughly half of a bitcell area per a memory word.
- A 32-bit layout was designed and parasitics were extracted.





Read/Write Assist Transistors Sizing

- The amount of current flowing through the read assist transistor:
- Voltage at the node V_{RA} is $V_{RA} = V_{dd} \exp\left(\frac{-t}{\tau}\right)$ $I_{RA} \approx \mu_n C_{ox} \left(\frac{W}{L}\right)_{RA} \left(V_{dd} V_{th}\right) V_{RA}$

Where, $\tau_d = R_{RA}C_{BL}$ and $\tau_d = \tau$ when voltage at node V_{RA} is 0.36 V_{dd}

Hence, size of the read assist M_{RA}

$$\left(\frac{W}{L}\right)_{RA} = \frac{1}{R_{RA}\mu_n C_{ox}(V_{dd} - V_{th})}$$

• Size of the write assist M_{WA} : A single equivalent minimum size transistor per word for minimum leakage and data retention.









Stability Analysis of SE-SRAM ...

- SNM of traditional SRAM and proposed SE-SRAM.
- Under normal read operation.
- For traditional SRAM cell ratio = 2.
- For proposed SE-SRAM minimum size transistors.



- Proposed cell has 2X higher SNM than the standard cell at beta = 2 and V_{dd} =1.0V.
- For subthreshold operation proposed cell SNM is equal to stdard cell at beta=4 and V_{dd} =0.5V.





Stability Analysis of SE-SRAM

- SNM of standard SRAM and the proposed SE-SRAM.
- Read operation under process variations in V_{th}.
- atio =2.
 For prop. 6T minimum size of transistors



- For the worst case prop. cell has 2.65X higher SNM than the standard cell at beta=2 and V_{dd} =1.0V.
- The worst case standard deviation in the SNM for proposed cell is 11% higher than the standard cell at beta=2 and V_{dd} =1.0V.





Active Power Dissipation of SE-SRAM

- Active power of standard and proposed SRAMs.
- For all possible read and write operations at V_{dd}=1.0V.
- Power pattern is asymmetrical for proposed SE-SRAM, because of asymmetric r/w operation or its structure.



- If the upcoming datum is same either for read of write operation (W1_1 or R1_1) the proposed SRAM has low power consumptions compared standard.
- If the upcoming datum is zero during read operation (R1_0 or R0_0) proposed design has 21% and 29% higher power than the standard SRAM.
- Average active power in the proposed design is 28% lower than the standard.





Significance of the 8T SRAM

- The proposed SE-SRAM design achieves 2.65X better static noise margin compared to a standard 6T-SRAM.
- Improved write-ability of logic '1'.
- Minimum feature size devices.
- No radioed contention or tuning of cell ratio
- Saving of active and leakage power.
- One disadvantage: A marginally high standard deviation in the SNM and active and leakage power due to minimum sized device.





SRAM Optimization Methodology 1: Combined DOE-ILP Approach





Combined DOE-ILP Approach: Solution 1



Discover the power of ideas

- 1: Input: Baseline Psram/SNM sram, Nominal/High VTh 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

Design Flow 1

- 7 Form predictive equations : $\overline{f_{PWR}}$ for power, $\overline{f_{SNM}}$ for SNM.
- 8 : Solve $\overline{f_{PWR}}$ using ILP. Solution set : Spwr.
- 9: Solve $\overline{f_{SNM}}$ using ILP. Solution set : S_{SNM}.
- 10: Form $S_{OBJ} = S_{PWR} \bigcap S_{SNM}$.
- 11: Assign high $V{\sc transistors}$ based on SOBJ.



Combined DOE-ILP Approach: Solution 2



Discover the power of ideas

1: Input: Baseline Psram/SNMsram, Nominal/High - VTh models.

- 2: Output: Objective set SOBJ* = [fPWR*, fSNM*] with transistors identified for high VTh 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: $\overline{f_{PWR}} * and \overline{f_{SNM}} *$.

8 : Form fobj* =
$$\left(\frac{\overline{f_{PWR}} *}{\overline{f_{SNM}} *}\right)$$

9: Solve $\overline{f_{OBJ}}^* = using ILP$. Solution set : SOBJ*.

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

Design Flow 2 Advanced Topics in VLSI Systems



Combined DOE-ILP Approach

Predictive Equation:

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

 χ_n is the V_{Th} -state of transistor n;

- \hat{f} is the response of cell ; (e.g. Power, SNM, etc)
- \overline{f} is the average of response in the cell;

 $\begin{pmatrix} \underline{\Delta(n)} \\ 2 \end{pmatrix}$ is the half effect of the nth transistor ; it is calculated by: $\frac{\underline{\Delta(n)}}{2} = \begin{pmatrix} \frac{avg(1) - avg(0)}{2} \end{pmatrix}$



Selection of Appropriate Transistors



Configuration for Flow 1





Experimental Results: 4 Alternatives

Design Alternative	Parameter	Value	Change
Baseline	Psram	203.6 nW	-
	SNM sram	170mV	-
Spwr	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
Sobj	P sram	113.6 nW	44.2% decrease
Approach 1	SNM sram	303.3 mV	43.9% increase
Sobj *	P sram	100.5 nW	50.6% decrease
Approach 2	SNM sram	303.3 mV	43.9% increase





Experimental Results: SNM



Advanced Topics in VLSI Systems

UNIVERSITY OF NORTH TEXAS Discover the power of ideas

Experimental Results: Power/SNM





Advanced Topics in VLSI Systems



Monte Carlo Distribution Results ...





Advanced Topics in VLSI Systems



Monte Carlo Simulation Results

Optimization	Parameter	Mean	Standard Deviation
S_{PWR}	P _{sram}	28.91 nW	8.26 nW
	SNM _{sram}	180mV	30mV
$\mathbf{S}_{\mathbf{SNM}}$	P _{sram}	147.73nW	101.4nW
	SNM _{sram}	295mV	28mV
S _{OBJ} : Approach 1	P _{sram}	147.73nW	101.4nW
	SNM _{sram}	295mV	28mV
S _{OBJ} : Approach 2	P _{sram}	135.24nW	101.85nW
	SNM _{sram}	295mV	28mV





Array Organization for 7T and 10T SRAM









SRAM Optimization Methodology 2: Statistical DOE-ILP





Statistical DOE-ILP Approach for Nano-CMOS SRAM

- 1: Input : Baseline SRAM.
- 2: **Output**: Optimized P3: power minimization, performance maximization and process variation tolerant SRAM.
- 3: Measure power and SNM of baseline SRAM cell.
- 4: Go To Algorithm 2 for optimizing baseline SRAM
- 5: **Re simulate** SRAM cell to obtain P2 (power minimization and performance maximization) SRAM cell.
- 6: **Perform** process variation characterization of SRAM cell using device paramters(12).
- 7: Obtain P3 optimal SRAM cell.
- 8 : Construct array organization for e.g. 8 × 8 array to observe the feasibility of the optimal SRAM cell.

Design flow for P3 optimal SRAM





Algorithm for P2 optimal SRAM cell

- **Input:** Baseline PWR, SNM of SRAM cell, Baseline model file, High- threshold model file.
- **Output:** Optimized objective set $f_{obj} = [f_{PWR}, f_{SNM}]$ optimal SRAM cell with transistors identified for High V_{Th} assignment.
- Set-up experiment for transistors of SRAM cell using 2-Level Taguchi L-8 array, where the factors are the V_{Th} states of transistors of SRAM cell, the response for average power consumption is $\overline{\mu PWR}$, $\overline{\sigma PWR}$ and the response for read SNM is $\overline{\mu SNM}$, $\overline{\sigma SNM}$.
- For Each 1:8 experiments of 2-Level Taguchi L-8 array do
 - Run 100 Monte Carlo runs
 - Record $\overline{\mu PWR}$, $\overline{\sigma PWR}$ and $\overline{\mu SNM}$, $\overline{\sigma SNM}$
- end for
- Form linear predictive equations

 $\overline{\mu PWR}$, $\overline{\sigma PWR}$ for power

 $\overline{\mu SNM}$, $\overline{\mu SNM}$ for SNM.

- Solve μPWR using ILP: Solution set: $S_{\mu PWR}$
- Solve $\overline{\sigma PWR}$ using ILP: Solution set: $\dot{S}_{\sigma PWR}$
- Solve $\overline{\mu SNM}$ using ILP: Solution set: $S_{\mu SNM}$
- Solve $\overline{\sigma SNM}$ using ILP: Solution set: $S_{\sigma SNM}$
- Form $S_{obj} = S_{\mu PWR} \cap S_{\sigma PWR} \cap S_{\mu SNM} \cap S_{\sigma SNM}$
- Assign high \dot{V}_{Th} transistors based on S_{obj} .
- Re-simulate SRAM cell to obtain optimized objective set.





P3 SRAM Optimal Results



SRAM Optimization Methodology 3: PVT Optimization of SRAM











Ambient Temperature Analysis



Advanced Topics in VLSI Systems

UNIVERSITY OF NORTH TEXAS Discover the power of ideas

Algorithm for PVT-tolerant SRAM

- **Input:** Baseline power and SNM o<u>f the</u> SRAM cell, baseline model file.
- **Output:** Optimized FOM: $\overline{f_{PSR}} = \frac{f_{PWR}}{\overline{f_{SNM}}}$, with transistors identified for optimized W_n and W_p
- Identify worst case ambient temperature (measure at 27°C, 50°C, 75°C, 100°C, 125°C) for defined FOMs (Power, SNM and PSR) of SRAM design.
- Generate power dissipation profile of SRAM design by measuring average (total) power consumption and total leakages.
- for Each range of W_n and W_p of transistors in SRAM do Run simulations, Record power, SNM and PSR.
- end for
- Generate surface plots using Polynomial Regression, for all three FOMs.
- Form polynomial equations: $\overline{f_{PWR}}$ for power, $\overline{f_{SNM}}$ for SNM and $\overline{f_{PSR}}$ for PSR.
- Minimize f_{PWR} using second order differential equation.
- Maximize $\overline{f_{SNM}}$ using second order differential equation.
- Minimize $\frac{1}{f_{PSR}}$ using second order differential equation.
- Optimize: $\overline{f_{PSR}} = \frac{\overline{f_{PWR}}}{\overline{f_{SNM}}}$
- Assign optimized values of W_n and W_p for the NMOS and PMOS transistors.
- Re-simulate SRAM cell to obtain optimized objective $\overline{f_{PSR}}$





Surface Plots and Fit Matrix



Advanced Topics in VLSI Systems

Discover the power of ideas

Optimal Simulation Results

Parameter	Baseline Power	Power optimality	SNM optimality	PSR optimality
Average Power	1.03 μW	1.03 μW	1.23 μW	1.03 μW
SNM	150.1 mV	150.1 mV	154 mV	154 mV
PSR	18.94	18.94	20.84	18.94





PVT Tolerant SRAM Optimal Results



Significance of the Methodology

- Design of Experiments-Integer Linear Programming (DOE-ILP) approach.
- Design of Experiments (DOE) assisted conjugate gradient approach.
- Statistical Design of Experiments-Integer Linear Programming (DOE-ILP) approach.
- Polynomial regression based technique.
- The following circuits have been subjected to these optimization methodologies:
- ➢ 45 nm 6-Transistor SRAM
- 45 nm 7-Transistor SRAM
- > High-κ/Metal-Gate 32 nm 10-Transistor SRAM





Comparative Perspective

Approach	Power (nW/ μW)	Performance (SNM) (mV)	Temp. (°C)	No. of Transistors	Technology
Combined DOE-ILP	100.5 nW	303.3 mV	27	7T	45nm nano CMOS node
DOE-ILP Assisted Conjugate Gradient	314.5 nW	295 mV	27	10T	High-K/Metal- Gate 32nm node
Statistical DOE-ILP	113.6 nW	303.3 mV	27	7T	45nm nano CMOS node
Polynomial Regression	1.03 μW	154 mV	125	7T	45nm nano CMOS node



Advanced Topics in VLSI Systems

