Analysis of Read-stability and Write-ability in FinFET SRAM cells

Ashish Kumar Sharma[#1], Nikhil Saxena*[2]

[#1]*M.Tech Scholar, Electronics & Communication Department, ITM Universe Gwalior, Madhya Pradesh (India)*

Abstract*—SRAM cells are designed exclusively to guarantee that the data of the cell is not altered during read access and the cell can quickly change its state during write operation. These conflicting needs for read and write operations are contented by some specific conditions to provide stable read and write operations, SRAM cell read stability and write ability is most important concerns in nanometer scale technologies, due to the progressive increase in intra die variability and Vdd scaling. In conventional six transistors (6T) SRAM cell, read stability is very low due to the voltage division between the access and driver transistors during read operation. Analysis has been done with the factors responsible to improve the read stability and write ability of 6T SRAM cell structures. SRAM cell stability analysis is typically based on Static Noise Margin (SNM) investigation.*

Keywords—*Read stability, write ability, Cell Ratio, Pull up transistor, SRAM cell, Static Noise Margin (SNM)*

I. INTRODUCTION

In present scenario of electronics industry, we generally use the word "memory" interchangeably to RAM. Read/Write memory or random access memory hold instructions temporarily, which is required for CPU (central processing unit) to process the task and permit alteration of data bits which is stored in the memory arrays i.e. volatile memory. In digital logic system design semiconductor memories are most vital microelectronics component such as embedded microcontroller or microprocessor. To store a huge amount of digital information semiconductor memory arrays are designed. The amount of memory is always depends on the application for which it is used [1]. The persistently increasing demands for huge memory capacity have preoccupied the fabrication technology and design rules more compressed and dense, this result in higher storage densities for data. On chip memory arrays are broadly used subsystem in modern VLSI based circuits, commercially used single chip memory has the capacity about 1GB. This trend toward superior storage capacity and higher density will prolong to increase the leading edge of digital system design. According to the type of data storage the semiconductor memory is classified as stable or temporary and times essential to store and recover a data bit in the memory.

For the demand of minimizing power consumption during active operation, supply voltage scaling is often used. However, SRAM reliability is even more deduced at lower voltages. V_{dd} min is the minimum supply voltage for an SRAM array to read and write safely under the required frequency restriction. Therefore, the analysis of SRAM read/write stability is essential for low power SRAMs. SRAM cell design has to cope with a severe constraint on the cell area to achieve high integration density in modern system-on-chips (SoCs). This leads to choosing minimal width-to-length (W/L) ratios for the SRAM cell transistors. As dimensions scale down to nanometer regime, the variations in CMOS transistor parameters, e.g., the threshold voltage(V_{th}) , increase steadily [2] due to random dopant density fluctuations in channel, source and drain. Therefore, two closely placed, supposedly identical transistors can have important differences in their electrical parameters as V_{th} and make the design of the SRAM less predictable and controllable. Moreover, the stability of the SRAM cell is seriously affected by the increase in variability and by decrease in supply voltage (V_{dd}).

II. READ WRITE OPERATION

Like other memories, SRAM cell can be operated in three modes: standby (or hold), read, and write modes. In the standby mode, the word line is set to a low-voltage level and both the internal nodes are isolated from the bit lines. In the read mode, both the bit lines are usually precharged to a high-voltage level before the FinFETs are turned on, the charges in the bit lines will disturb the charges stored in the internal nodes, and if the inverters are not "strong" enough i.e., the static noise margin is too small [3]. The bit lines may not be sufficiently discharged to the expected values, Hold(Stand By) The access transistors(M3 and M4) are disabled by applying word line signal WL to a low voltage level equal to “0” to their gates and both the internal nodes are isolated from the bit lines. The data is held in the latch. The bit lines (b and $\bar{b}$) are charged to the supply voltage. In the large SRAM

array (e.g.>1 MB), most of the cell are in standby state, which dominates the overall power consumption.

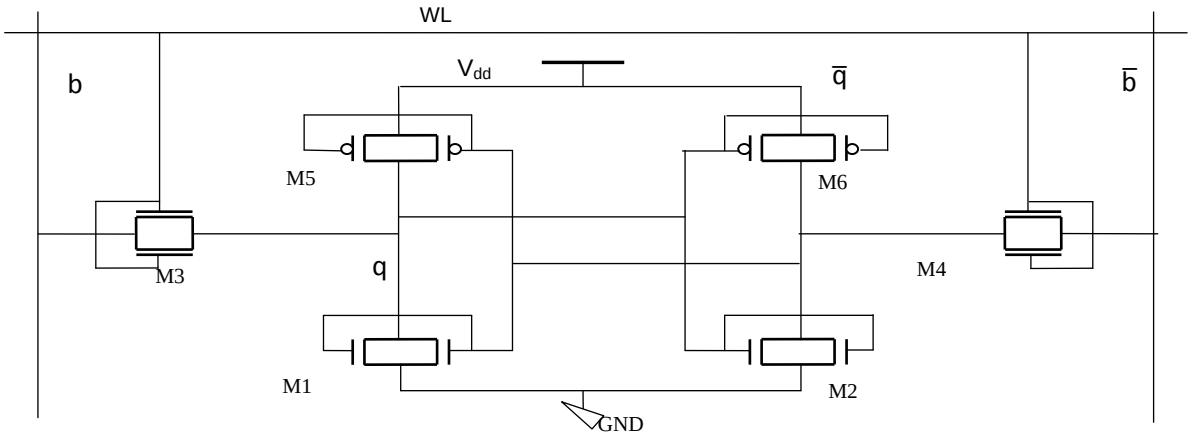

Fig.1 Basic structure of 6T FinFET SRAM cell.

A. Read Operation

In read operation, word line is activated while the external word line driver is disabled. The value can be determined by external logic if the inverter inside the SRAM cell drives the bitlines. The Bit lines (b and $\bar{b}$) of the cell are pre charged as given in the above step if reading is done just after write operation. In the read mode, the charges in the bit lines will hassle the charges stored in the internal nodes, and if the inverters are not strong enough (i.e. the static noise margin significantly smaller), the bit lines may not sufficiently discharged to the expected values.

B. Write Operation

In write operation, the big (external) tri-state drivers need to activate first, to drive the bitlines. The previous state of the cross-couple can easily be over write. It is because the internal driver (small transistor used in the 6T SRAM cell) is much smaller than the external drivers. Now, allow the wordline transistors at rest, when shifting the data, the short circuit will occur for just a few nanoseconds. Data to be written into the cell is applied to the bit lines (b and $\bar{b}$). The access transistors (M3 and M4) are enabled by applying high (1) word line signal WL to the gates. When data "0" is to be written to storage node storing "1", the corresponding bit- line is applied with voltage equal to "0", resulting in a current flow through pull up device (PUP1) to the bit line through the storage node storing high, which is pulled low. When the voltage of the high storage node is below the trip point of the other inverter (M6 pull-up and M2 pull-down) the content of the cell flips due to feed back [4].

The contrary takes place when the storage node voltage is "0" and "1" is to be written. There will be no change in the state of the cell, if the storage node is initially storing "0" and the data on the bit line is also "0". If the storage node is initially storing "1" and the data on the bit line is "1" then also there is no change in the state of the cell. It is very essential that for easy write operation the pull up ratio (PR), given by the ratio of the width to length (W/L) ratio of the pull up device (PUP) to the width to length (W/L) of the access transistor (PG) should be small

Cell Ratio

$$\text{Cell ratio (CR)} = \frac{\text{size of driver transistor}}{\text{size of load transistor}}$$

$$\text{(during read operation)} = \frac{(W1/L1)}{(W3/L3)} = \frac{(W2/L2)}{(W4/L4)}$$

Pull up Ratio

$$\text{Pull up ratio} = \frac{\text{size of load transistor}}{\text{size of access transistor}}$$

$$\text{(during write operation)} = \frac{(W6/L6)}{(W4/L4)}$$

Drive Strength

$$\text{Drive strength } \beta = \mu C_{ox} \frac{W}{L}$$

I. READ AND WRITE STABILITY

In standby mode and during a read access, Data retention and read stability of the SRAM cell is an important functional constraint in advanced technology nodes. The cell becomes less stable with lower supply voltage, it also results in rising

leakage currents and increasing variability from technology scaling. The stability is usually defined by the SNM [5] as the maximum value of DC noise voltage that can be tolerated by the SRAM cell without changing the stored bit. For a successful write, only one cross point should be establish on the butterfly curves, indicating that the cell is mono - stable. WSNM for writing "1" is the width of the smallest square that can be embedded between the lower-right half of the curves. WSNM for writing "0"can be obtained from a similar simulation. The final WSNM for the cell is the bare minimum of the margin for writing "0"and writing "1". A cell with lower WSNM has poorer write ability.

II. SUCCESSFUL READ VS FAILURE READ

SRAM cells are designed to make sure that the contents of the cell should not altered during read access, and the cell can quickly change its state during write operation. These incompatible requirements for read and write operations are satisfied by sizing the bitcell transistors to provide stable read and write operation. In read operation, a static RAM bitcell is most prone to failure. After the WL is enabled, voltage at the internal storage node storing a zero (q) slightly rises due to the voltage divider between the PG transistor (PG1) and the pull-down (PD1), as shown in Fig.2. If the voltage at q rises close to the threshold voltage of the adjacent pull-down PD2, then the cell may flip its state. Therefore, stable read operation requires that PD1 should be stronger than PG1 i.e. Drive strength of pull down device >> Drive strength of access device [6].

$\beta_d \gg \beta_a$

$$\mu\frac{W_d}{L_d}(V_{cell}-V_{th,d}) \gg \mu\frac{W_a}{L_a}(V_{wl}-V_{th,a})^{\alpha}$$

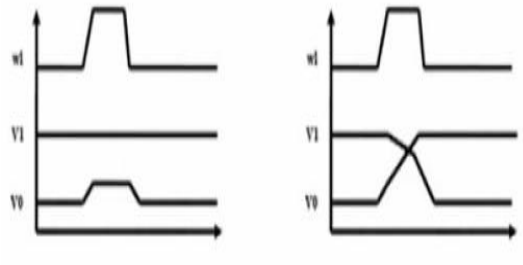

Fig.2 Successful Read Vs failure read

Read stability failure is exacerbated by process variations, which affect all the transistors in the bitcell. To quantify the bitcell's robustness against this type of failure, static noise margin (SNM) is one of the most commonly used metrics. A read stability failure can occur if the bitcell cannot hold the stored data i.e. SNM is zero. Read stability failure can occur any time, if the bitcell is not accessed for either read or write operations and the WL is enabled. For example, in half-selected bitcells, the WL is enabled while the bitlines column is not selected (the bitcells are not actively accessed for read or write). These bitcells experience a dummy read operation because the bitlines are initially precharged to V_{dd}, and the bitlines are discharged after the WL is enabled, hence, the bitcells become prone for read stability failure. Dealing with read stability failures is one of the biggest challenges for SRAM design and has been extensively studied

III. DELAY

To determine delay of signal from input to output circuit during the high to low and vice versa at the output, we use propagation delay times τPHL and τPLH. We can define these terms as, τPHL is referred to the time delay which occurs with 50% of the voltage transition of the rising input voltage and falling output voltage. Similarly, τPLH is referred to the time delay which occurs with the 50% of the voltage transition of the falling input voltage and rising output voltage. For simplification of the analysis and derive the delay expression, we have considered an input voltage waveform having zero rise and fall time acting as an ideal step pulse. Due to this assumption τPHL becomes the time consumed for the output voltage to fall from VOH to 50% of the voltage level and τPLH becomes the time consumed for output voltage to rise from VOL to 50% of the voltage level. The 50% of the voltage point is defined by

$$V_{50\%} = V_{OL} + \frac{1}{2}(V_{OH}-V_{OL}) = \frac{1}{2}(V_{OH}+V_{OL})$$

Now we can define the average propagation delay τ_P of the inverter which gives the average time required for the input signal to propagate through the inverter.

$$\tau_P = \frac{\tau_{PHL}+\tau_{PLH}}{2}$$

Delay of the cell depends upon the time elapsed between the cell from input to output transition of the signal.

IV. LEAKAGE CURRENT & POWER

Power dissipation occurs due to charging and discharging of load capacitances, it refers to as the dynamic power dissipation. Dynamic power is consumed when switching in bits going on either "0" to "1" or "1" to "0". It provides the region of transistors works in active region and cut-off region, the current is:

$$I_D = \mu_n C_{ox}\frac{W}{L}\{(V_{GS}-V_{TH})V_{DS}-\frac{V_{DS}.^2}{2}\}$$

For active region, the current equation is:

$$I_D = \mu_n C_{ox} \frac{W}{L}((V_{GS} - V_{TH})V_{DS})$$

For saturation region, the current equation becomes:

$$I_D = \mu_n C_{ox} \frac{W}{L}(\frac{V_{DS}^2}{2})$$

Scaling down technology to a significant value increases the performance but total power consumption in some portion not decreases due to leakage parameters due to reduced threshold voltage and high packaging density [7]. ITRS in 2014 stated that 13% of feature size shrinks per year. The cost of memory also reduces because of bulk production of units. The gate oxide thickness reduces drastically as feature size reduces then gate tunneling leakage current increases [8].

Power dissipation has four components in digital circuits:

$$P = P_{ds} + P_{sc} + P_{sb} + P_{leakage}$$

In which P is the total power dissipation, $P_{ds}$is the dynamic-switching, P_{sc} is the short circuit, P_{sb} is the static-biasing and $P_{leakage}$ is the leakage power.

The power dissipation in dynamic-switching depends upon frequency, capacitor and supply voltage

$$P_{ds} = CV^2f$$

The short circuit power dissipation is depends on τ rise time or fall time, and clock frequency f. The formula is:

$$P_{sc} = K(V_{dd}-2V_{th})^3\tau f$$

The flow of static current towards ground from supply voltage without input degradation is leakage power. Three leakage mechanisms are: Sub threshold, band to band tunneling (BTBT) and gate oxide.

$$I_s = I_o W_e \frac{V_{gs}-(V_{to}-\eta V_{ds}-\gamma V_{bs})}{\eta V_T}[1-e^{\frac{-V_{ds}}{V_T}}]$$

Where I_s is sub threshold leakage current

V. CONCLUSION

The improved read and write-ability (data stability), reduced dynamic and leakage power dissipation can be achieved by varying the parameters i.e. Pull up ratio, pull down ratio, cell ratio, width to length ratio of transistors, it makes the new approach attractive for nanoscale technology regime, in which process variation is a major design constraint. The conflicting read and write problem is addressed by providing a separate read word line.

REFERENCES

[1] T.Skotnicki , M. Sellier, A. Farcy, and F. Boeuf, "An evaluation of the CMOS technology roadmap from the point of view of variability, interconnects, and power dissipation." IEEE Trans. Electron Devices, vol. 55, no. 6, pp. 1433–1440, Jun. 2008.

[2] B. Cheng *et al.*, "The impact of random doping effects on CMOS SRAM cell," in *Proc. ESSCIRC*, Sep. 2004, pp. 219–222.

[3] Li Ding and P. Mazumder. Dynamic noise margin: definitions and model. In VLSI Design, 2004. Proceedings. 17th International Conference on, pages 1001 – 1006, 2004.

[4] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi and V. De, "Parameter Variations and Impact on Circuits and Microarchitecture", Proc. DAC, 2003, pp. 338-342.

[5] B Calhoun and A Chandrakasan. Static noise margin variation for sub-threshold SRAM in 65-nm CMOS. IEEE Journal of Solid State Circuits, 41:1673, 2006.

[6] M. Khellah, Y. Ye, N. S. Kim, D. Somasekhar, G. Pandya, A. Farhang, K. Zhang, C. Webb and V. De, "Wordline & bitline pulsing schemes for improving SRAM cell stability in low-Vdd 45nm CMOS designs," Symp. on VLSI Circuits, pp. 9-10, June 2006.

[7] D. A. Antoniadis, I. Aberg, C. N. Chléirigh, O. M. Nayfeh,A. Khakifirooz, and J. L. Hoyt, "Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations," IBM J. Research Development, vol.050, no.4, pp. 363–376, Jul. 2006.

[8] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep sub micrometer CMOS circuits," Proceeding of IEEE, vol.91, no.2, pp.305–327, Feb. 2003.