

New SRAM Cell Design for Low Power and High Reliability using 32nm Independent Gate FinFET Technology

Young Bok Kim
Dept. of Electrical and
Computer Engineering
Northeastern University
youngbok@ece.neu.edu

Yong-Bin Kim
Dept. of Electrical and
Computer Engineering
Northeastern University
ybk@ece.neu.edu

Fabrizio Lombardi
Dept. of Electrical and
Computer Engineering
Northeastern University
lombardi@ece.neu.edu

Abstract

This paper proposes new methods for SRAM cell design in FinFET technology. One of the most important features of FinFET is that the independent front and back gates can be biased differently to control the current and the device threshold voltage. By controlling the back gate voltage of a FinFET, a SRAM cell can be designed for low power consumption. This paper proposes a new 8T (8 transistors) SRAM structure that reduces dynamic power for the write operation and achieves a wider SNM (Static Noise Margin). Using the new FinFET based 8T SRAM cell, dynamic power consumption is reduced by nearly 48% and the SNM is widened up to 56% compared to the conventional 6T SRAM at the expense of 2% leakage power and 3% write delay increase.

1. Introduction

Recently, the double-gate FET has become a popular device candidate for future generations of CMOS technology [1][2]. The FinFET, a vertically oriented embodiment, has received a great deal of attention lately as a double-gate FET that is more amenable to conventional CMOS processing

Unlike planar single- and double-gate devices, the FinFET effective channel width is perpendicular to the semiconductor plane. Therefore, it is possible to increase the effective channel width and drive current per unit planar area by increasing the fin-height. Interconnect dominated circuits such as memory arrays are likely to get benefited from the increased driving current. An estimated 70% of the transistors in a billion-transistor superscalar microprocessor are expected to be used for memory arrays, especially for large L2 and L3 SRAM data caches. Therefore, it is

imperative to develop a low power SRAM design technique for new device such as FinFETs.

In this paper, we proposed new SRAM cell structure with FinFET technology and examine regarding performance, power, stability and process variation.

2. FinFET technology

The FinFET transistor is a vertical double-gate device and is regarded as a promising alternative for sub-45 nm bulk devices [3]

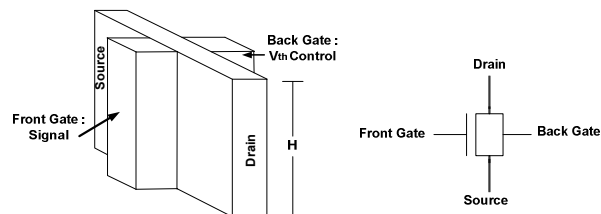


Figure 1. FinFET Structure and Symbol

Figure 1 shows the structure of a multi-fin double-gate FinFET device. Current flow is parallel to the wafer plane. The thickness t_{si} of the single fin is equal to the silicon channel thickness. Each fin contributes to the width of the device, and H is the height of each fin. The FinFET circuit behavior can be studied using PTM (Predictive Technology Model) of 32 nm CMOS FinFET technologies [4]. A unique property of the FinFET is the electrical coupling between the front and back gates. The implication of this coupling is that the threshold voltage of the front gate (V_{thf}) is not only established by the process, but also it can be controlled by the back gate voltage (V_{Gb}). This is similar to the body effect in a bulk transistor.

An independent-gate FinFET operates in the dual-gate mode (DGM) when both gates are biased to induce channel inversion. Alternatively, an

independent-gate n-FinFET (p-FinFET) operates in the single-gate mode when one of the gates is deactivated by connecting the gate to ground (V_{DD}). Disabling one of the gates in the single-gate mode (SGM) increases the absolute value of the threshold voltage compared to DGM. Therefore, it is possible to modulate the threshold voltage of the FinFET by biasing the two gates independently [5].

3. Proposed 8T FinFET SRAM Cell

3.1 Write/Read Operations

In the proposed 8T SRAM, the write and read bits are separated. While the bit and bit-bar lines are used for writing data in a traditional 6T SRAM, only the WRITE_BIT in Figure 2 is used in the proposed SRAM cell to write both “0” and “1” data. The write operation starts by disconnecting the feedback loop of the two inverters. By setting the ‘W_bar’ signal to “0”, the feedback loop is disconnected. The data to be written is determined by the WRITE_BIT voltage. If the feedback connection is disconnected, the SRAM cell has just two cascaded inverters. WRITE_BIT transfers the complementary of the input data to Q2, which drives the other inverter (P2 and N2) and generate Q_bar. WRITE_BIT has to be pre-charged "high" before and right after each write operation. When writing a "0", a negligible writing power is consumed because there is no discharging activity at WRITE_BIT. To write a '1' at Q2, the WRITE_BIT has to be discharged to ground, similar to the 6T SRAM cell. In this case, the dynamic power consumed by the discharging process is the same as the 6T SRAM. The write circuit does not discharge for every write operation but it needs to discharge only when the cell writes a “1”, and the activity factor of the discharging WRITE_BIT is less than "1", which makes the proposed SRAM cell more efficient in power consumption during write operation compared with conventional cells.

All Read_Bit lines are pre-charged before READ operation. During read operation, the transistor N5 is turned on by setting the W_bar signal high and the READ_ROW (RD) is “high” to turn on N6. When Q2="0", N4 is off thus the READ_BIT voltage does not change from the pre-charged value, i.e. Q2 holds a “0”. If Q2= “1”, the transistors N4 and N6 are turned on. In this case, due to charge sharing, the READ_BIT voltage will drop about 100~200mV, this is sufficient to be detected by the sense amplifier.

3.2 Front and back gate configuration of the FinFET

The operation of writing “1” is stable because the NMOS transistor N3 can pass a “0” unaffected.

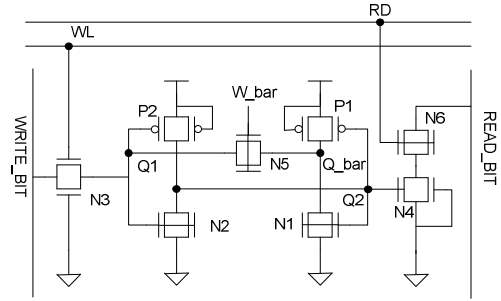


Figure 2. The proposed 8T SRAM structure with independent gate control

However, when writing a “0”, WRITE_BIT is pre-charged high (V_{DD}) and N5 is turned off. The node voltage at Q1 is less than V_{DD} due to the threshold voltage drop between the gate and source of the transistor N3. To compensate for this voltage drop, the transistors N2 and P2 must be designed as a low-skew inverter to ensure Q2 to be at solid ground level to represent the “0” state. The low-skewed inverter has a weaker PMOS transistor. If the PMOS FinFET back gate is connected to V_{DD} , the current is reduced by making the V_{th} (threshold voltage) of the transistor P2 higher than the case in which the front and back gates are tied together. By connecting the back gate to V_{DD} in PMOS, the leakage current can also be reduced.

Assume that the cell stores “0” at Q2 and “1” at Q_bar after WL (Word Line) is deactivated and W_bar is activated. In this case, the voltage at Q1 is less than V_{DD} due to the threshold voltage drop across the gate and the source of the transistor N5. The degraded voltage at Q1 may slightly turn on the transistor P2 causing a short circuit current through transistors P2 and N2. To overcome this problem, the low skewed inverter (N2 and P2, as required for writing in the “0” case) is justified again and the V_{th} of the transistor N5 needs to be controlled low to reduce the voltage difference between Q_bar and Q1. To control the V_{th} of the transistor N5 low, the front and back gates must be connected together as shown in Figure 2.

To implement a low skewed inverter with transistors N2 and P2, the transistor ratio of N2 to P2 should be at least 2 to ensure a solid ground level for Q2. However, by correctly connecting the back gates of the FinFETs, the sizes of P2 and N2 can be same, i.e. if the back gate of P2 is connected to V_{DD} , and the front and back gates of N2 are tied together, then the driving current ratio N2/P2 can be more than 2. Therefore, the inverter transistor ratio N2/P2 can be less than 2 if the FinFET’s threshold voltages are

controlled by correctly connecting the back gates. The transistor ratio N3/P2 of 1.3, N1/P2 of 3, and a low V_{th} for the transistor N5 guarantee a stable READ operation when Q_{bar} stores a "0". However, if the sizing approach similar to N2/P2 is used to optimize the transistor ratios among N1, N3 and P2, the transistor sizes can be further reduced. If the back gate of N3 is connected to GND, the front and back gates of N1 are tied together, and the back gate of P2 is connected to V_{DD} , the transistor N1 needs to be only 1.5 times larger than the transistor P2 to satisfy the relationships among N1, N3, and P2.

By combining the threshold voltage controllability of the FinFET using appropriate back gate connection and transistor sizing, the proposed 8T SRAM accomplish low power consumption by reducing node capacitance and tuning V_{th} at the cost of minimal area overhead.

4. Simulation results

In this paper, four different SRAM cells are designed; Each 6T and 8T SRAM cells are designed using tied FinFETS (front and back gates of the FinFETS are tied together) and independent double gates FinFETS (front and back gates are independently controlled). SRAM (6T-Ind) is implemented by using an independent gate control that connects the back gates of the NMOS (PMOS) transistors to GND(V_{DD}) to reduce the leakage current. The proposed 8T SRAM (8T-Ind) configuration is shown in Figure 2.

The technology parameters for the FinFETs are; Channel length (L) = 32nm, Fin Height (H_{fin}) = 32nm, Fin thickness (t_{si}) = 8nm, Oxide thickness(t_{ox}) = 1.6nm, Channel doping = $2 \times 10^{20} \text{ cm}^{-3}$, Source/ Drain doping = $2 \times 10^{20} \text{ cm}^{-3}$, Work functions (N-FinFET) = 4.5eV, Work functions (P-FinFET) = 4.9eV, and V_{DD} = 0.8 V.

HSPICE using the PTM model is used to simulate the proposed 8T SRAM and conventional 6T SRAM. Table 1 shows the summary of the results to compare the proposed 8T SRAM characteristics to the conventional 6T SRAM's.

4.1 Dynamic power consumption

The proposed 8T SRAM achieves 48% writing power saving while maintaining cell performance, read/write delay, and stability of the conventional cell. The significant power saving is possible since the cell keeps the WRITE_BIT "high" instead of discharging when it writes a "0", which reduces the activity factor of the WRITE_BIT.

Table 1. Summarized simulation results

		6T-Tied	6T-ind	8T-Tied	8T-ind
Dynamic Power (W) ($V_{DD}=0.8V$)		1.42E-04	1.41E-04	6.78E-05	6.73E-05
Leakage Power ($V_{DD}=0.8V$)	Temperature 27°C	1.36E-06	1.06E-06	1.20E-06	1.08E-06
	Temperature 100°C	8.84E-06	8.34E-06	9.04E-06	8.63E-06
Write Delay(ps)		27.3ps	31.3ps	28.3ps	29.3ps
Read Delay (ps)		22.25ps	25.1ps	23.16ps	25.7ps
SNM (mV)		172	186	220	240
Read SNM (mV)		125	136	181	190

While the conventional 6T SRAM always discharges one of the bit lines to write data into the cell, the proposed 8T SRAM discharge the WRITE_BIT only when it writes a "1". As the probability of writing '0' increases, the power dissipation due to discharging of the bit line is reduced compared to the conventional case. Figure 3 shows the dynamic power consumption for different V_{DD} . As shown in Figure 3, the power saving of the 8T SRAM is manifest as V_{DD} increases because the dynamic power difference between the 6T SRAM and the proposed 8T SRAM increases exponentially as V_{DD} increases.

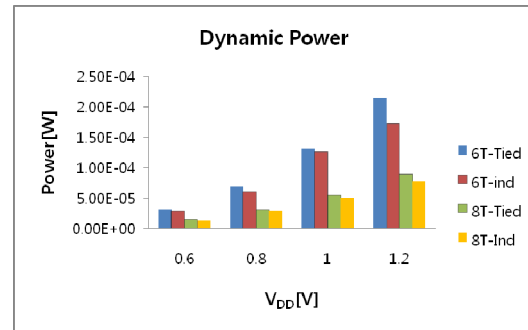


Figure 3. Dynamic power consumption versus V_{DD} variation

4.2 Leakage Power Consumption

Table 1 and Figure 4 shows the leakage power of the four different configurations. The proposed 8T FinFET SRAM shows a slightly higher leakage power because it has one more leakage current path from the READ_BIT, N4, N6. However, N4, N6 cause the stack effect; the N4 back gate is connected to GND to cause N4 to have a high V_t . Hence, the leakage current through the READ_BIT, N4, and N6 path is relatively small.

The difference in leakage current in all four configurations is less than 2%. In the 6T and 8T independent gate configurations, the leakage current is smaller because they are configured to have a high V_t by connecting the back gates of the PMOS to V_{DD} .

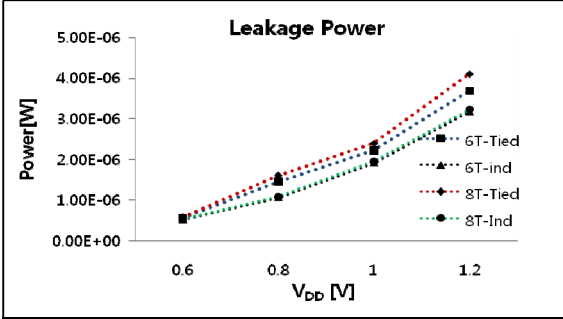


Figure 4. Leakage power consumption versus V_{DD} variation

4.3 Process Variations

The Static Noise Margin (SNM) is the widely used metric to measure the stability of SRAM bit-cells [6]. The static noise margin of a SRAM cell is defined as the minimum DC noise voltage necessary to flip the state of the cell. The voltage transfer curves (VTCs) of the back-to-back inverters in a bit-cell are used to measure the SNM [6]. By separating the Read and Write, a wider SNM can be achieved during read operation as shown in Table 1.

Monte Carlo based SNM simulation is used to study the impact of gate length and fin width on the cell read stability for the 8T-Tied and 8T-Ind cell configurations. The variations on physical sizes are assumed to be $3\sigma L_g = 3\sigma T_{si} = 10\%$ for L. The results are shown in Fig. 5. It is well known that the electrical characteristics of independent gate FinFET are more sensitive to process variations. As expected, the proposed 8T-Tied shows a larger spread than 8T-Ind.

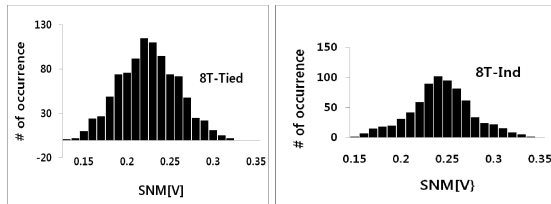


Figure 5. Bitcell read stability (measured by the static noise margin) distributions for 8T-Tied and 8T-Ind cell schemes

4.4 Read/Write Delay

At the cell level, the READ access time is determined by the time taken for the bitlines to develop a potential difference of at least 100mV. The READ time depends on the transistor sizes in the READ path. The READ delay for the proposed 8T SRAM cell is almost the same as the conventional cell due to the similar sizes of the path transistors. For WRITE

operation, the WRITE delay is defined as the time from the 50% activation of the WL to the time when Q_{bar} reaches 90% of its full swing. The WRITE delay is approximately equal to the propagation delay of inv2 (N2/P2) and inv1. As inv1 is only driving the diffusion capacitor of N5, it is then desirable to reduce the input capacitance of inv1 as much as possible to reduce the load on inv2. Therefore, the proposed 8T SRAM is only slightly slower than the 6T SRAM in WRITE operation.

5. Conclusion

This paper proposed a new 8T SRAM cell using FinFETs. This new SRAM cell has no feedback connection between the two back-to-back inverters when data is written and separates the write and read ports with 8 transistors. The proposed technique saves dynamic power by reducing the discharging activity during WRITE operation. Compared to the 6T Tied configuration, the proposed 8T SRAM saves power up to 48% and has 56% wider SNM during READ operation.

6. References

- [1] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K.E. Ismail, S.-H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S.P. Wong, "CMOS scaling into the nanometer regime," *Proceedings, IEEE*, 1997, vol. 85, no. 4, pp. 486-504.
- [2] R.V. Joshi, R.Q. Williams, E. Nowak, K. Kim, J. Beintner, T. Ludwig, I. Aller, and C. Chuang, "FinFET SRAM for high-performance low-power applications", *Proceeding of the 34th European Solid-State Device Research conference*, 2004, pp. 69 - 72.
- [3] X. Huang, et al., "Sub-50 nm P-channel FinFET," *IEEE TED*, 2001, vol. 48, pp. 880-886.
- [4] "Predictive Technology Model for 32 nm CMOS FinFET echnologies", <http://www.eas.asu.edu/~ptm/>
- [5] Sherif A. Tawfik Volkan Kursun, "Low-Power and Compact Sequential Circuits With Independent-Gate FinFETs", *IEEE Transactions on Electron Devices*, January 2008, Vol. 55, no. 1.
- [6] A. Bhavnaganwala, X. Tang, and J. D. Meindl, "The Impact of Intrinsic Device Fluctuations on CMOS SRAM Cell Stability," *IEEE Journal of Solid-State Circuits*, April 2001, Vol. 36, No. 4, pp. 658-665.