

Low Power 8T SRAM Using 32nm Independent Gate FinFET Technology

Young Bok Kim

Dept. of Electrical and
Computer Engineering
Northeastern University
Boston, MA, USA
youngbok@ece.neu.edu

Yong-Bin Kim

Dept. of Electrical and
Computer Engineering
Northeastern University
Boston, MA, USA
ybk@ece.neu.edu

Fabrizio Lombardi

Dept. of Electrical and
Computer Engineering
Northeastern University
Boston, MA, USA
lombardi@ece.neu.edu

ABSTRACT

In this paper, new SRAM cell design methods for FinFET technology are proposed. One of the most important features of FinFET is that the independent front and back gate can be biased differently to control the current and the device threshold voltage. By controlling the back gate of FinFET, SRAM cell can be designed for minimum power consumption. This paper proposes a new 8T (8 transistors) SRAM structure that reduces dynamic power in writing operation and provides wider SNM (Static Noise Margin). Using the new FinFET based 8T SRAM cell, dynamic power consumption is reduced about 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.

I. INTRODUCTION

Conventional bulk CMOS scaling beyond 45nm is severely constrained by short channel effects and vertical gate insulator tunneling [1]. Double-gate FinFET technology [2] has been proposed as a very promising candidate to circumvent the conventional bulk CMOS scaling constraint by changing the device structure in such a way that MOSFET gate length can be scaled further even with thicker oxide, which makes it possible to continue scaling beyond the limit of the conventional bulk CMOS.

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 essential to develop a low power SRAM design technique for the new device such as FinFET.

II. FinFET Technology

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

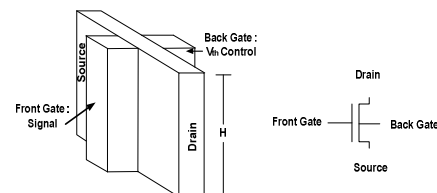


Figure 1: FinFET Structure and Symbol

Figure 1 shows the structure of multi-fin double-gate FinFET device. Current flow is parallel with the wafer plane. The thickness t_{si} of the single fin equals to the silicon channel thickness. Each fin provides the width of the device, and H is the height of the each fin. FinFET circuit behavior is studied using PTM (Predictive Technology Model) of the 32 nm CMOS FinFET technologies [4]. One 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 governed by the process, but also it can be controlled by the back gate voltage (V_{Gb}). This is similar to the body effect in the 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 the DGM. Therefore, it is possible to modulate the threshold voltage of FinFET by biasing the two gates independently [5].

III. Proposed 8T FinFET SRAM Cell

A. Write/Read Operations

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

All the Read_Bit lines are pre-charged before any READ operation. During read operation, transistor N5 is turned on by setting W_bar signal high and the READ_ROW(RD) is “high” to turn on N6. When Q2=“0”, the N4 is off making the READ_BIT voltage not change from the pre-charged value, which means the cell data Q2 holds “0”. On the other hand, If Q2= “1”, the transistors

N4 and N6 are turned on. In this case, due to charge sharing, the READ_BIT voltage will be dropped about 100~200mV, which is enough to be detected in the sense amplifier.

B. Front and back gate configuration of FinFET

The operation of writing “1” is stable because NMOS transistor N3 can pass “0” faithfully

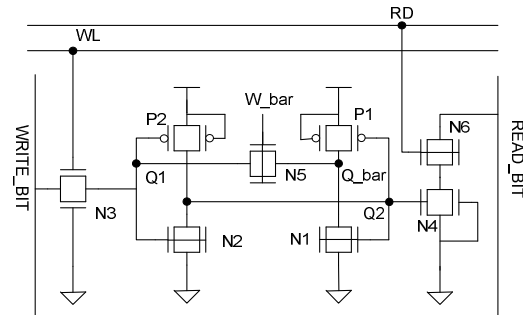


Figure 2: The proposed 8T SRAM structure with independent gate control.

On the other hand, when writing “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 this voltage drop, the transistor N2 and P2 must be designed as a low-skew inverter to ensure Q2 to be solid ground level to represent “0” state.

A low-skewed inverter has a weaker PMOS transistor. If the PMOS FinFET back gate is connected to the V_{DD} , the current is reduced making the V_{th} (threshold voltage) of the transistor P2 higher than that of the case where front and back gate are tied together. By connecting back gate to the V_{DD} in PMOS, the leakage current also can be reduced.

Let’s suppose 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 source of the transistor N5. The degraded voltage at Q1 may turn on the transistor P2 slightly causing short circuit current through transistors P2 and N2. To overcome this problem, the low skewed inverter (N2 and P2) mentioned for writing “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 gate and back gate must be connected together as shown in Figure 2.

To implement a low skewed inverter with transistors N2 and P2, transistor ratio of N2 to P2 should be at least 2 to have a solid ground level at Q2. However, by connecting the back gates of the FinFETs appropriately, the P2 and N2 sizes can be same. That is, if the back gate of P2 is connected to V_{DD} , and the front and back gate of N2 are tied together, then the current ratio N2/P2 can be more than 2. This means that the inverter transistor sizes N2/P2 can be smaller than 2 if the FinFET's threshold voltages are controlled by connecting the back gates appropriately. Transistor ratio N3/P2 of 1.3, N1/P2 of 3, and low V_{th} of the transistor N5 guarantees a stable READ operation when Q_bar stores "0". However, if the similar approach to N2/P2 sizing is used to optimize 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 gate and back gate of N1 are tied together, and the back gate of P2 is connected to VDD, the transistor N1 needs to be only 1.5 times larger than transistor P2 to satisfy the relationships among N1, N3, and P2.

Combining the threshold voltage controllability of the FinFET using appropriate back gate connection and transistor sizing techniques, the proposed 8T SRAM cell can accomplish low power consumption due to smaller node capacitance and tuning V_{th} at the minimal cost of the area overhead.

IV. SIMULATION RESULTS

In this paper, total 4 different SRAM cells are designed; Each 6T SRAM and 8T SRAM cell is 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 independent gate control which connects the back gates of the NMOS(PMOS) transistors to GND(V_{DD}) to reduce the leakage current. And the proposed 8T SRAM (8T-Ind) configuration is shown in Figure 2.

The technology parameters of the FinFETs used in this research 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 = 2×10^{20}

cm^{-3} , Work functions (N-FinFET) = 4.5eV, Work functions (P-FinFET) = 4.9eV, and V_{DD} = 0.8 V.

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

A. Dynamic Power Consumption

The proposed 8T SRAM achieves 48% writing power saving while maintaining the cell performance, read/write delay, and stability of the conventional cell. The power saving comes from the fact that the cell keeps WRITE_BIT "high" in stead of discharging when it writes "0", which reduces the activity factor of the WRITE_BIT.

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

B. Leakage Power Consumption

Table 1 shows the leakage power of the four different configurations. The proposed 8T FinFET SRAM shows slightly higher leakage power because it has one more leakage current path. The READ_BIT, N4, N6 path make additional leakage current path. However, N4, N6 cause stack effect and N4 back gate is connected to GND that makes N4 have high- V_i . As a result, the leakage current through the READ_BIT, N4, and N6 path is relatively small. The difference of the leakage current in all of the four configurations is less than 2%. In 6T and 8T independent gate configuration

cases, the leakage current is smaller than others because they are configured to have high V_t by connecting the back gates of the PMOS to V_{DD} .

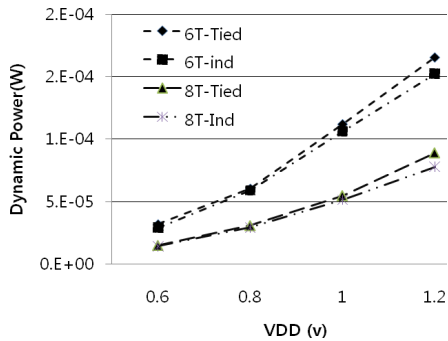


Figure 3: The Dynamic Power consumption with V_{DD} variation

C. Cell Stability

Figure 4 shows the Static Noise Margin (SNM) difference between the conventional 6T SRAM and the proposed 8T SRAM. Static Noise Margin is the standard metric to measure the stability in SRAM bit-cells [6]. The static noise margin of 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 bit-cell are used to measure SNM [6]. Separating the Read and Write bit offers wider SNM during read operation as shown in Figure 4. When reading the stored data, only READ_BIT affects inv1 (N1/P1) output voltage. Consequently, this fact makes the cell hard to flip.

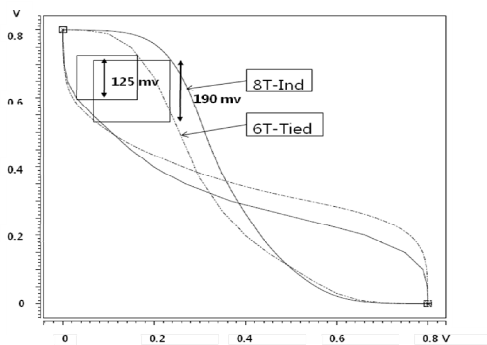


Figure 4: Static noise margin SNM for conventional front and back gate tied 6T SRAM(6T-Tied) and proposed cell (8T-Ind)

D. Read/Write Delay

The READ access time at the cell level is determined by the time taken for the bitlines to develop a potential difference of at least 100mV.

The read time depends on the READ path's transistors' sizes. The proposed 8T SRAM cell's

READ delay is almost same as the conventional cell's due to the similar sizes of 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} becomes 90% of its full swing. The write delay is approximately equal to the propagation delay of inv2 (N2/P2) and inv1. Because inv1 is only driving the diffusion capacitor of N5, it is desirable to reduce the input capacitance of inv1 as much as possible to reduce the load on inv2. The proposed 8T SRAM is little slightly slower than 6T SRAM in writing operation because of this reason.

IV. CONCLUSION

This paper proposes a new 8T SRAM using FinFET. The new SRAM cell cuts off the feedback connection between the two back-to-back inverters in the SRAM cell when data is written and separates write and read port with 8 transistors. The proposed technique saves dynamic power by reducing discharging frequency during write operation. Compared to 6T Tied configuration, the proposed 8T SRAM saves power up to 48% and obtains 56% wider SNM during read operation at the minimal cost of 2% leakage power and 3% delay increase. As the cells are more frequently accessed, the dynamic power saving is linearly increases.

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," *Proceeding . IEEE*, vol. 85, no. 4, pp. 486–504, 1997 .
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*, pp. 69 – 72, 2004
3. X. Huang, et al., "Sub-50 nm P-channel FinFET," *IEEE TED*, vol. 48, pp. 880-886, 2001
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*, Vol. 55, no. 1, January 2008
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*, Vol. 36, No. 4, pp. 658-665, April 2001.