STUDY OF SRAM AND ITS LOW POWER TECHNIQUES

Anu Tonk¹, Meenu Rani Garg²

Department of Electronics Engineering, YMCA University, Faridabad, Haryana

ABSTRACT

This paper discusses the basic operations of SRAM such as write, read and hold. These operations are performed with help of Tanner tools at .18µm technology. The paper also discusses the low power design techniques for SRAM. There is a four type low power technique discussed here for SRAM. One is the Half-swing Pulse-mode techniques in which a Half-swing Pulse-mode gate family is used that in turn uses reduced input signal swing without sacrificing performance and saves the power. Second is a memory bank partitioning, in which memory array is partitioned to enhance the speed and to reduce the power. Third is the Quiet Bit line architecture in which the voltage of bit line stays as low as possible. To prevent the excessive full-swing charging on the bit line, one-side driving scheme for write operation is used and for read precharge free-pulling scheme is used to keep all bit lines at low voltages at all times. Fourth is the Pulsed Word line and Reduced Bit line Swing in which voltage at bit lines is reduced.

Keywords: SRAM, Half Swing Pulse Mode Technique, Quiet-Bitline Architecture, Pulsed Word line and Reduced Bit line Swing

1. INTRODUCTION

Static Random Access Memories (SRAMs) are an important component of microprocessors and system-on-chips. SRAMs are used as large caches in microprocessor cores and serve as storage in various IPs on a system-on-chip. SRAMs used in high performance microprocessors and graphics chips have high speed requirements. At the same time, SRAMs used in application processors which go into mobile, handheld and consumer devices have very low power requirements. Since SRAMs serve as large storage on these chips, it’s very important to get maximum density out of these. Traditionally a large number of SRAM bit cells, up to 1024 in some cases, are connected to a common bit-line to get the highest density and array efficiency.
1.1 6T SRAM Cell Operations

1.1.1 Hold / Standby mode: If the word line is not asserted, the access transistors M5 and M6 disconnect the cell from the bit lines. The two cross coupled inverters formed by M1 – M4 will continue to reinforce each other as long as they are connected to supply.

1.1.2 Write mode: The write cycle begins by applying the value to be written to the bit lines and asserting the WL.

- Write ‘0’ and hold ‘0’: To write logic ‘0’, ‘0’ has to be applied to BL and ‘1’ has to be applied to BLbar. WL is then asserted. Bi-stable Latching Circuitry is used to store each bit. When BL is given logic ‘0’, it will be fed to ‘A’, the PMOS will be closed and it pulls the Vdd to give the output as logic ‘1’ at ‘Abar’. For holding ‘0’, WL is made zero.

![Schematic of 6T SRAM for Write ‘0’ and hold ‘0’](image-url)
Write ‘0’ and hold ‘0’: To write logic ‘1’, ‘1’ has to be applied to BL and ‘0’ has to be applied to BLbar. WL is then asserted. Bi-stable Latching Circuitry is used to store each bit. When BL is given logic ’1’, it will be fed to ‘A’, the NMOS will be closed and it pulls down the output to logic’0’ at ‘Abar’. For holding ‘0’, WL is made zero which disconnects the BL and BLbar from the latch circuit.

Figure 1.2 Waveform for Write ‘0’ (0-50ns) and hold ‘0’ (50-100ns)
Figure 1.3 Schematic of 6T SRAM for Write ‘1’ and hold ‘1’

Figure 1.4 Waveform for Write ‘1’ (0-50ns) and hold ‘1’ (50-100ns)
1.1.3 Read mode: Both the BL and BLbar are precharged to $V_{dd}$, then asserting the WL, enabling both access transistors. Values stored in A and Abar are transferred to bit lines.

- **Write ‘0’ and Read ‘0’**: The read cycle is started by precharging both the bit lines to a logical ‘1’, then asserting the WL, enabling both the access transistors. The values stored in A and Abar are transferred to the bit lines by leaving BLbar at its precharged value and discharging BL to a logical ‘0’ and BL is pulled towards Vdd, a logical ‘1’.

![Schematic of 6T SRAM for Write ‘0’ and Read ‘0’](image1)

**Figure 1.5** Schematic of 6T SRAM for Write ‘0’ and Read ‘0’

![Waveform for Write ‘0’ (0-50ns) and Read ‘0’ (50-100ns)](image2)

**Figure 1.6** Waveform for Write ‘0’ (0-50ns) and Read ‘0’ (50-100ns)
Write ‘1’ and Read ‘1’: The read cycle is started by precharging both the bit lines to a logical ‘1’, then asserting the WL, enabling both the access transistors. The values stored in A and Abar are transferred to the bit lines by leaving BL at its precharged value and discharging BLbar to a logical ‘0’ and BL is pulled towards Vdd, a logical ‘1’.

Figure 1.7 Schematic of 6T SRAM for Write ‘1’ and Read ‘1’

Figure 1.8 Waveform for Write ‘1’ (0-50ns) and Read ‘1’ (50ns-100ns)
2. LOW POWER TECHNIQUES FOR SRAM

In any low power application, reducing operating voltage is always an excellent method to lower dynamic power dissipation. However, the performance and other system requirements typically limit the operating voltage and we have to find other techniques. In SRAM, the following low power techniques can be applied.

2.1 Half-Swing Pulse-Mode Technique

Without affecting the performance low power SRAM can be designed using Half Swing Pulse Mode gate family that uses reduced input signal swing. These gates are applicable in SRAM decoder and write circuits because it decreases the power by reducing signal swing on the high capacitance predecode lines, write bus lines, and bit lines. In [3] they used dual-CMOS. To reduce the power dissipation we can keep the supply voltage low or we can reduce the supply voltage but by reducing the supply voltage transistor threshold voltage also lowered and it affect the performance. To overcome from this problem they used multiple-threshold CMOS (MT-CMOS) and variable threshold CMOS (VT-CMOS). MT-CMOS used in decode and peripheral logic.

![Figure 1.9: Half-Swing Pulse Mode AND gate](image)

To prevent the significant leakage current high- \( V_t \) device is used, while to provide the good performance low-\( V_t \) device is used. \( V_t \)-CMOS controls the transistor threshold by varying the bias of the well and/or substrate. The power can be saved by operating the bit lines from \( V_{dd}/2 \) rather than \( V_{dd} \). To use \( V_{dd}/2 \) half swing pulse-mode gate family can be used but it has some disadvantages because it needs level-conversion and reduced gate overdrive at receiving gates.
2.2 Memory Bank Partitioning

Another technique that can be used to improve the speed is to partition a memory array into smaller memory banks so that only the addressed bank is activated. Interestingly, this technique can also improve the power efficiency of the SRAM because the word line capacitance being switched and the number of bit cells activated are reduced. The partitioning can be applied hierarchically to produce smaller memory banks. However, at some point the delay and power overhead associated with the memory bank decoding circuit dominates and further partitioning becomes harmful. Typically, two to eight bank partitions are appropriate.

2.2.1 Quiet-Bitline Architecture

By quiet it is meant that the voltage of the bit lines stay as low as possible at all times. The immediate advantage is that all charging/discharging power associated with the bit lines can be eliminated. There are two different methods of functioning for write and read operations. For write operation, one side driving scheme is used because it limits the excessive full-swing charging on the bit lines. In this method, a strong “0” signal is forced into the cell being accessed from one side, while other side is floating. For read operation, pulling scheme is used. It operates in four steps: bit line equalization, word line activation, and bit line pulling and finally sense amplification. For a read operation, the bit line charging power can be reduced quiet significantly if the aggressive word line pulse control is used to limit the bit line swing to around 100-200mV. However, all bit lines are usually forced to have full swings during a write operation, mainly with a view to enable a quick cell flipping. A one side driving scheme in which only strong “0” signal is forced into a bit line or bit line-bar for the write operation, while leaving the other side floating. And we remove the recharging for both read and write operations. The combination of these two methods leads to a design with quiet bit lines.

2.2.2 Pulsed Word line and Reduced Bit line Swing

In a conventional SRAM design, the bit lines are allowed to swing from rail to rail during a read operation. To conserve power, the voltage swing at the bit lines needs to be limited. One way to achieve this is to isolate the memory cells from the bit lines after a successful differential sensing. This activity prevents the memory cells from changing the bit line voltage further. In some cases, the bit line sense amplifiers also need to be isolated from the bit lines after sensing. This prevents the bit line capacitance from having a large voltage swing, thus saving power. This method requires pulsed signals at the word lines and the sense amplifiers.

3. CONCLUSION

We have surveyed various low power techniques for SRAM. Out of these techniques, Half-swing Pulse-mode technique in which half-swing pulse-mode gate is used which is well suited for decreasing the power in SRAM peripherals circuits like decoder, and write circuit by reducing the signal swing on high-capacitance precode lines, write bus lines and bit lines.

Quiet bit line architecture gives 84.4% power reduction compared to self-designed baseline low power-SRAM. SRAM cell using body-Bias technique can work under 0.3V supply voltage. If we use this technique, it increases the SRAM area. Write margin is not degraded.
REFERENCES


