Design of 16nm SRAM Architecture

[1,2]Iuri A. C. Gomes, [1,2]Cristina Meinhardt, [1,3]Paulo F. Butzen
iualbandes@hotmail.com, {cristinameinhardt, paulobutzen}@furg.br

[1]Centro de Ciências Computacionais - Universidade Federal do Rio Grande - FURG
[2]PPGC – Universidade Federal do Rio Grande do Sul – UFRGS
[3]PGMICRO – Universidade Federal do Rio Grande do Sul - UFRGS

Abstract

This work aims to design, validate, analyze and compare two blocks of SRAM memory architecture using 6T bit cell. One block use High Performance (HP) transistor type and the other Low Power (LP) type. Both models are from a 16nm predictive technology. Each block is composed of four main circuits: Bit Cell, Sense-Amplifier, Pre-Charge and Write Driver. The two designed architectures are validate through electrical simulation and the obtained results show the HP architecture with better performance and the LP architecture with better robustness.

1. Introduction

Static Random Access Memories (SRAMs) are critical components in the memory hierarchy of modern integrated circuits. They are the most used memory for cache systems. Cache systems are intermediate memories in the memory hierarchy, placed between the main memory (usually DRAM – Dynamic Random Access Memory) and the processor. They are designed to work in the same or very close processor frequency. In several applications, the embedded SRAMs can occupy the majority of the chip area and contain hundreds of millions of transistors [1]. SRAMs need to be not only fast but also reliably, i.e, it must be stable and robust to the system work properly.

The scaling of MOSFET transistors has been the strategy adopted by the semiconductor industry to keep following the Moore Law [2]. As the process technology continues to scale deeper into the nanometer region, the stability of embedded SRAM cells is a growing concern[1]. The reduction in technology dimensions brings many benefits to SRAM. The main benefits can be considered the improvement in performance, smaller memory blocks area and lower power consumption. However, miniaturization causes the memory cells to be more vulnerable to noise. In other words, nanometer SRAMs loses in stability.

The main objective of this work is the design of two SRAM architecture using predictive transistors models of 16nm, one block using High Performance (HP) transistors and other Low Power (LP) transistors. All architectures are composed by four circuits: Bit Cell, Sense-Amplifier, Pre-Charge and Write Driver. Section 2 describes the used methodology and the design considerations of the circuits used in this work. Section 3 presents and discusses the obtained results. Finally, Section 4 presents the conclusions of main aspects observed in the project.

2. Design and Methodology

As mentioned previously, the proposed SRAM is composed by four circuits: Write Driver, Bit-Cell, Sense-Amplifier and Pre-charge circuit. These blocks and the full architecture were described in SPICE netlists using a 16nm predictive technology [3]. The electrical simulations were performed in NGSPICE [4], 0.7V was used as nominal voltage. This architecture has been validated according to the performance of the READ, WRITE and HOLD operations, checking the correct behavior of an SRAM.

2.1. Bit-Cell

There are several bit-cells proposals for SRAM in literature, and they are usually named according to the number of transistors that form the architecture, each with its performance characteristics, power consumption and arrangement of transistors. Figure 1 shows the most common memory bit cell, the SRAM 6T. As the name implies, it is composed by six transistors. It was chosen to be the SRAM bit cell used in this work. The most important characteristics of this cell are the stability and simple transistor arrangement. In this arrangement, the four internal transistors of the circuit (Q1 - Q4) create the two crossed CMOS inverters responsible for storing one bit. The transistors Q5 and Q6 (access transistors) define when the Bit-Cell can be accessed by an operation.

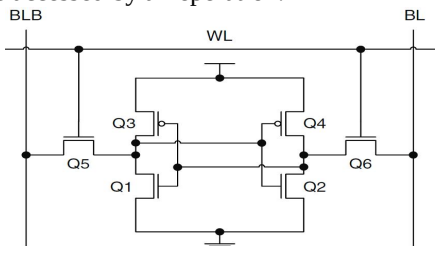

Fig.1 – Six transistors bit cell – SRAM 6T

The transistor sizing of the memory cell influences the performance and reliability of SRAM. The design adopted for a cell should follow two main rules for proper operation of SRAM [1], [5-7]:

- The NMOS transistors of the pull-down network (Q1 and Q2) must be approximately equal to or stronger than the access transistors (Q5 and Q6). This rule guarantees the stability of the read operation.
- The transistors PMOS from the pull-up network (Q3 and Q4) should be approximately equal to or weaker than the access transistors (Q5 and Q6). This rule ensures the exchange of the value stored in the Bit-Cell during the write operation.

For all transistors are adopted the minimum L of the technology, 16nm. Observing the rules presented above, the conclusion regarding the W of the transistors are:

$$\boldsymbol{W_{nmos} \geq W_{acs} \geq W_{pmos}}$$

For 16nm technology, used in this work, the following considerations are defined:

- Stronger pull-down network: Wnmos = 80nm.
- Weaker pull-up network: Wpmos = 40nm.
- Medium strength access transistors: Wacs= 60nm.

2.2. Write Driver

The Write Driver is responsible for writing a specific value in the bitcell. The circuit has the function of charging or discharging the bitlines to the desired bit be written in the memory cell. The circuit schematic of write driver designed in this work is presented in fig. 2.

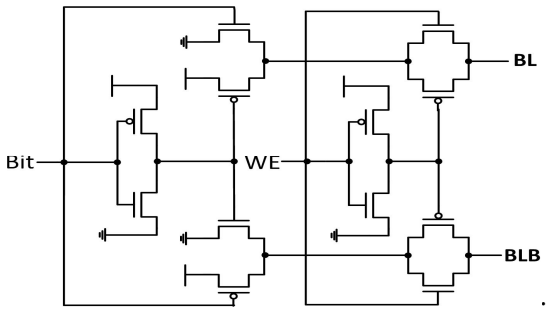

Fig.2: Write Driver

This circuit has two input signals, the signal representing the bit value to be written in the memory cell, and the control signal write enable (WE). The control signal WE function is to allow or not the access to the bitlines by the Write Driver. When WE is on, the Write Driver imposes in the bitlines the required voltage values for writing the desired value in the Bit-Cell.

In this project, the transistors width of W = 1um is used for all transistors. This size is necessary because there is only one circuit for each column of the proposed architecture. Since the capacitance of each bit line that has to be charged is large, these transistors width have to be large.

2.3. Pre-charge

The pre-charge of the bitlines is a very important factor for the correct functioning of the SRAM. To perform the read operation it is necessary that the bitlines are charged at the same voltage. After the writing and reading operations, one of the bitlines is discharged generating an undesired voltage difference between the bitlines. After these operations, the pre-charge circuit has the function to equalize the bitlines voltages to the supply voltage. The pre-charge circuit is depicted in fig. 3.

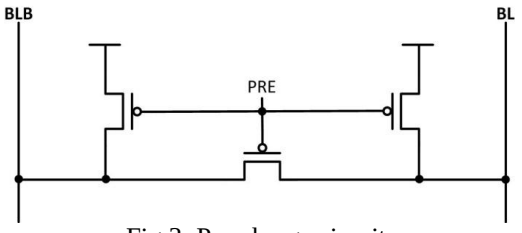

Fig.3: Pre-charge circuit

In this circuit, a transistor width of 500nm is used. As in Write Driver, this large size was used because the circuit has to charge the big bitlines capacitances.

2.4. Sense-Amplifier

The sense-amplifier (SA) is a circuit of great importance in the SRAM architecture. During the read operation one of the bitlines discharges while the other bitline remains at supply voltage. This slow discharge is due to the large bitline capacitance and the access transistor of the bit cell is small. For this reason, the SA is used to amplify a small difference between the bitline voltages values to digital levels. A significant portion of the SRAM performance during the read operation is due to the Sense - Amplifier circuit.. One of the most common types of SA is showed in the fig. 4.This circuit is known as latch-type SA [1]. The latch-type SA was chosen as the read circuit for this work. A transistor width of W = 100nm have been used in the circuit.

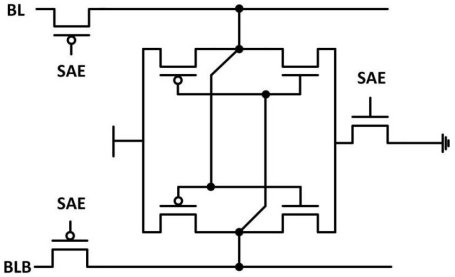

Fig.4: Latch-type Sense-Amplifier

2.5. SRAM Architecture and Validation Procedure

The project uses 256 bit cells for one column in the memory block. The 256 memory cells are implemented through one “under evaluation” SRAM 6T cell and the simulation of other 255 memory cells through two large size bit cells that have transistor width 127 times bigger than the “under evaluation” bit cell. One of these large size bit cells store the value “0” and and the other one the logic value “1”. They are used to emulate capacitances and leakage currents. Fig. 5 illustrates the structure described above. The follow sections describe in more details the considerations used to design each block.

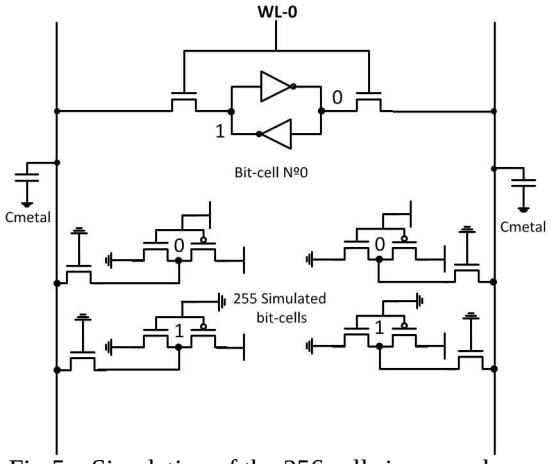

Fig.5 – Simulation of the 256 cells in one column

The last procedure before doing the performance and stability tests was to validate each block individually. To validate the bit-cell, it is necessary to analyze if the 6T circuit is able to store a bit. To validate the write-driver the writing ability is analyzed, i.e, we have tested the capacity to write the correct bit value in the bit-cell. The validation of the pre-charge circuit regarded its ability to charge the bit-lines to supply. The Sense-Amplifier validation is done in two steps: First It is analyzed if the bit returned by the READ operation was correct; secondly, it is analyzed if the bit stored in the bit-cell did not change. After the validations the blocks were integrated creating the final architecture.

3. Results

The first analysis regards the functionality of the two memory architecture designs. . These architectures were tested through the validation of theirs READ, WRITE and HOLD operations. The READ operation returns the correct value in the bit-cell without destructing the cell stability. The WRITE operation writes the desired value in the cell and the HOLD operation keep the bit stored. After this analysis, we could conclude that HP SRAM architecture and also LP SRAM architecture were in fact working properly.

The second factor to be analyzed was the performance of the architectures. Tab. 1 presents the performance time for the READ and WRITE operation for the LP and HP architectures. As we can see, the HP architecture performance is almost 20 times better than the LP architecture in both operations. The main reason for those results lies in difference of the Vth of the HP and LP transistors. The HP technology has a lower Vth, so it switches faster than the LP technology.

Tab.1 - Performance Results

Arquitetura	Write	Read
HP	128ps	106ps
LP	2629ps	2348ps

The third and last analysis regarded the SRAM stability and robustness. We calculated the Static-Noise Margin for both architectures during the READ and HOLD operation. Tab. 2 shows that LP architecture is more robust and stable than the HP architecture. We can also see that the READ operation is the most vulnerable moment in both architectures. The reason for those results also lies in the Vth difference of the LP and HP transistors. The higher Vth of the LP technology makes it more robust and stable than the HP transistors.

Tab.2 – Static Noise Margin (SNM) Results

Arquitetura	SNM READ	SNM HOLD
HP	41.4mV	77.4mV
LP	176.1mV	224.3mV

4. Conclusion and Future Works

The validation of both architectures (HP and LP) demonstrated that is possible to design a 16nm SRAM architecture that has WRITE, READ and HOLD operations working properly.

Further analysis showed that the HP architecture had a better performance than the LP architecture. The HP SRAM was twenty times faster than the LP SRAM regarding the READ and WRITE operations.

The stability analysis indicated that the LP circuit was more robust than the HP circuit. The LP circuit SNM was 4.5 times better than the HP SNM for the READ operation. The SNM for the HOLD operation of the LP architecture was 2.9 times better than the HP architecture.

Further analysis can be done to this work: Evaluate the power consumption and the area occupied by the architecture. Optimize the sizing of the devices in the architecture. Evaluate other types of SRAM bit-cell like the 7T, 8T.

5. Referências

[1] Pavlov, Andrei and Sachdev, Manoj. *CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies.* s.l. : Springer, 2008.

[2] Moore, Gordon E. Cramming more components onto integrates circuits. *Eletronics Magazine.* 1965.

[3] Predective Technology Model (PTM). *http://ptm.asu.edu/.* [Online]

[4] Ngspice circuit simulator. *http://ngspice.sourceforge.net/.*

[5] Masood Qazi, Mahmut E. Sinangil, Anantha P. Chandrakasan. *Challenges and Directions for Low-Voltage SRAM.* 2011.

[6] Pedroni, Volnei A. *Eletrônica Digital Moderna e VHDL.* s.l. : Campus, 2010.

[7] Alorda, B., et al. *Static-Noise Margin Analysis during Read Operation of 6T SRAM Cells.* 2009.