Design and Analysis of IO Blocks for 1024*16 CM8 SRAM

Md Tarique Ali, Kamil Hasan
Department of Electronics and Communication Engineering, Alfalah School of Engineering & Technology, Maharishi Dayanand University, Rohtak, Haryana, India

Abstract- SRAM design is very crucial since it takes a large fraction of total power and die area in high-performance processors. The performance of embedded memory and its peripheral circuits can adversely affect the speed and power of the overall system. This paper explores the design of SRAM focusing on optimizing delay, reducing power and layout area. The key to low power operation of the design is self-timed architecture, multi stage decoding and full custom approach. A 1024x16 SRAM is designed at UMC 180 nm technology.

Keywords - Low power SRAM, Self-timed, Bit-line Fractioning, Multi-Stage decoding

1. INTRODUCTION
Low-power SRAM has become more important due to the high demands on the handheld devices. The active power of the SRAM is mainly consumed in bit lines and data lines because the SRAM charges and discharges the high capacitive bit lines and data lines in read and write cycles [1]. As the bit width of SRAM becomes larger for high-performance applications, the power consumption in bit lines and data lines continues to increase. Therefore, power dissipation has become an important consideration both due to the increased integration and operating speeds, as well as due to the explosive growth of battery operated appliances. Considerable attention has been paid to the low-power and high-performance design. To reduce the power consumption the first technique is to reduce the active duty cycle of the memory operation using self-timed architecture. An internal clock pulse with reduced Ton (On time) is generated which controls all the memory operation. Second technique of power reduction is to use multi-stage row and column decoding which reduces the power consumption as well as it also improves the timing characteristics of memory.

Figure 1: Architecture of 1024x16CM8 SRAM

Fig. 1 shows the architecture of 1024x16CM8 SRAM memory chip. The SRAM memory system design has been done based on array-structured memory architecture at 180nm technology node with a nominal supply voltage of 1V. A 10 bit address bus (addr[9:0]) is required for 1024 memory locations. In addition, there is 16 bit data input bus (d[15:0]), 16 bit data output bus (q[15:0]), active low chip selection signal (csn), write enable (wr), supply voltage (VDD), and ground (VSS).
The address signals (addr[9:0]) are divided into two groups. One group (addr[9:3]) is used for row decoding (word lines) and the other group (addr[2:0]) is used for column mux decoding. Based on the seven row address bits, the row decoder produces 27 = 128 horizontal word lines. With three column address bits, the column decoder generates 8 select lines for 8 bit Input/output block. This 8 bit multiplexer along with column Mux, pre-charge cells, sense amplifier and read/write circuitry forms 8 bit Input/output block. There are sixteen 8 bit IO blocks connected horizontally to generate 128 vertical bit lines. The array produced by the intersections of the 128 horizontal word lines and the 128 vertical bit lines is the 1024x16x8 memory cell array. When the chip select (csn) signal is high, the chip is in idle mode. When the chip select (csn) signal is low, the chip is accessed. Write enable signal is kept ‘1’ for write mode and it is kept ‘0’ for read mode of memory operation. The data fed to input port (d[i]) is written to the location defined by the address bus. The data from the memory location defined by the address bits is written on output port q[i]. Note that, during write operation d[i] is measured on the bit line of the designated memory cell and during read operation the output data is available on the q[i].

II. LOW POWER TECHNIQUES

Concept of Power Reduction Using Self-Timed Memory Design

The primary technique used for power reduction is self-timed architecture. Memory timing circuits need a delay element which tracks the bit-line delay but still provide a large swing signal which can be used by the subsequent stages of the control logic. The key to building such a delay stage is to use a delay element which is a replica of the memory cell connected to the bit-line, while still providing a full swing output. This technique uses a dummy column and dummy row in the RAM to control the flow of signals through the core. This section explores the self-timed technique for the SRAM. The circuit diagram of self-timed IO block is shown in figure 2.

The technique for achieving this uses a “dummy column” in the RAM to control the flow of signals through the core. A dummy column is an additional column of bit-cells. Bit-cells in the dummy column are forced to a known state by shorting one of the internal nodes to a given voltage.

![Figure 2: Self Time I/O Block](image)

When hcp is low the dummy bit line dmbl is connected to one input of the NAND gate G1 followed by an inverter I1. The other input of the G1 is connected to memory enable signal which is high when the chip is selected. Hence it will have a high echo (reset) signal. If a rising edge of the hcp occurs, the dmbl will get discharged through the dummy row and we will have a low echo signal. This low echo signal resets the flip circuit in control block and kills the corresponding word-line shown in figure 3.

Concept of Power Reduction Using Multi-Stage Decoding

The performance (power and speed) of static CMOS decoder is based on its architecture, the number of transistors, fan-in and the loading on the address buffer. The input buffer drive the interconnect capacitance of address lines and also the input capacitance of the NAND gates. By using the two stage decoder architecture the number on transistor, fan reduced. As a result both speed and power are optimized. The logical diagram of a two-level 7 to 128 decoder is shown in Figure 4.
III. CONCLUSION

In this paper we looked at design of IO blocks for 1024x16CM8 SRAM. The SRAM access path is split into two portions: the row decoders and the read data path. Techniques to optimize both of these were discussed. In Chapter 3 we sketched out the 8-bit IO block structure for the said SRAM. Optimal decoder implementations result when the decoder, excluding the predecoder, is implemented as a binary tree. This minimizes the power dissipation as only the smallest number of long decode wires transition. With the predecoder the total path effort becomes independent of the exact partitioning of the decode tree, which will allow the SRAM designer to choose the best memory organization, based on other considerations. The key to high speed in the SRAM data path is to reduce the signal swings in the high capacitance nodes like the bitlines and the data lines. Clocked voltage sense amplifiers are essential for obtaining low sensing power, and accurate generation of their sense clock is required for high speed operation. An energy efficient way to obtain low voltage swings in the bitlines is to limit the word line pulse width, by controlling...
the pulse width of the block select signal. The pulse widths are regulated by the aid of a replica delay element which consists of a replica memory cell and a replica bitline and hence tracks the delay of the memory cell over a wide range of process and operating conditions. Two variations of the technique were discussed in Chapter 4. This is a very simple and robust technique with very low area overhead and is easily applicable to a wide variety of SRAM designs. This method can also be used to design a low swing data line operation for both the read and write accesses, by pulsing the control signal which gates the drivers of these lines. The pulse width is controlled by discharging a replica line which has the same capacitance as the worst case data line. This replica line is also used to trigger the amplifiers at the receiving end of these lines. We finally presented a technique to achieve low bitline write power, by using small voltage swings on the bitlines during the write operation. The memory cell structure is modified such that the cells can be used as latch type sense amplifier to amplify and store the small swing write data presented on the bitlines. Finally the design of 128-bit IO block is presented in Chapter 5. The IO block is integrated with the memory core, the row decoder and the control unit and tested at different PVT conditions. Specifically the characterization is done for read access time at different load and clock slope. Based on the results shown in this report, the design memory array achieved a successful read and write operation at (1V and 250C).

REFERENCES
[1] RF Silicon Technology Pvt. Ltd. (All the confidential specification provided by the company under Non Disclosure Agreement)