

DESIGN OF 1024*16 CM8 ULTRA LOW VOLTAGE SRAM WITH SELF TIME POWER REDUCTION TECHNIQUE

Akanksha Tyagi

Electronics and Communication, Sharda University, (India)

ABSTRACT

Technology scaling has enabled us to integrate both memory and logic circuits on a single chip. However, the performance of embedded memory and especially SRAM (Static Random Access Memory) that is widely used in the industry as on the on-chip memory cache in ultra low voltage applications can adversely affect the speed and power efficiency of the overall system. This report discusses the design techniques to realize input/output circuits which are used to access SRAM cell based memory array in ultra low voltage applications, to overcome the cell's variations. It also explains the variability problems in a SRAM bit-cell and many approaches to address them. The column decoder/multiplexer, the write driver circuit, the data output circuit, and the sense amplifier is discussed and implemented at transistor level using a six-transistor (6T) SRAM cell. Self time techniques have been implemented to optimize the power and access speed of SRAM.

Keywords: *Embedded memory, power efficiency, Self time techniques, SRAM, Technology scaling, ultra low voltage.*

I. INTRODUCTION

Fast low power SRAMs have become a critical component of many VLSI chips. This is especially true for microprocessors, where the on-chip cache size are growing with each generation to bridge the increasing divergence in the speeds of the processor and the main memory. Simultaneously, 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. This work explores the design of SRAMs, focusing on optimizing delay and power. While process and supply scaling remain the biggest drivers of fast low power designs, this thesis investigates some circuit techniques which can be used in conjunction to scaling to achieve fast, low power operation.

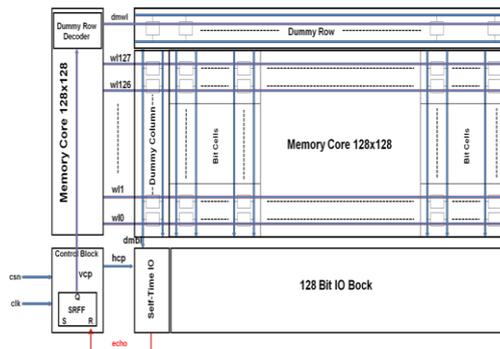
II. SELF TIMING THE SRAM CORE

Memory timing circuits need a delay element which tracks the bitline 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 bitline, while still providing a full swing

output. This technique for achieving this uses a dummy column and row in the RAM to time the flow of signals through the core.

In general, the speed of access to various rows is not identical. Clearly rows closest to the sense amplifier should give the fastest access time. Similarly columns closest to the word line drivers are enable first. To use a pulsed word line to its best advantage, we should tailor the width of pulse according to the access time of the RAM.

The technique for achieving this uses a “dummy column” in the RAM to time the flow of signals through the core. A dummy column is an additional column of bit cells and self timed IO block placed at the side farthest from the word drivers. Bit cells in the dummy column are forced to a known state by shorting one of the internal nodes to a given voltage.



III. OPERATION TECHNIQUE

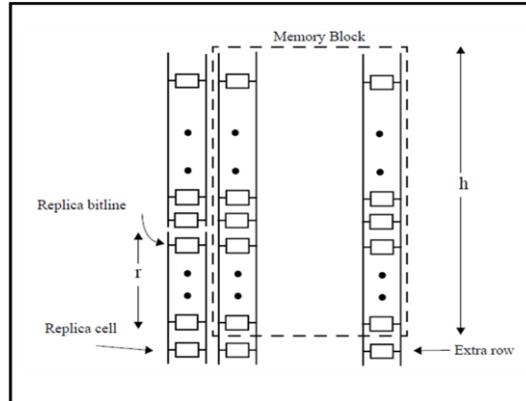
The sequence of operations that occur is as follows: The SR flip-flop is set and the word line is triggered. Cells along the row are enabled with the dummy column being the last cell enabled. By the time the sense amplifier corresponding to the dummy cell generates a low signal signal, the rest of the columns would have been sensed. The low signal from the sense amplifier resets the SR flip-flop and turns off the word line. This method handles the case of non uniform access time across the rows. The dummy column often adds insignificant overhead to the entire RAM. Consequently it is often preferred technique for pulsing the word line. This circuit is also at times termed word line kill circuitry.

3.1 Replica Delay Element Based on Capacitance Rationing

Memory timing circuits need a delay element which tracks the bitline 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 bitline, while still providing a full swing output. This section uses a normal memory cell driving a short bitline, but it uses a number of memory cells connected to a replica of the full bitline. The short bitline’s capacitance is set to be a fraction of the main bitline capacitance. The value is determined by the required bitline swing for proper sensing. For the clocked voltage sense amplifiers we use the minimum bitline swing for correct sensing is around a tenth of the supply.

An extra column in memory block is converted into the replica column by cutting its bitline pair to obtain a segment whose capacitance is the desired fraction of the main bitline. The replica bitline has a similar structure to the main bitlines in terms of the wire and diode parasitic capacitances. Hence its capacitance ratio to the main bitlines is set purely by the ratio of the geometric lengths, r/h . The replica memory cell is programmed to always

store a zero so that, when activated, it discharges the replica bitline. The delay from the activation of the replica cell to the 50% discharge of the replica bitline tracks that of the main bitline very well over different process corners. The delays can be made equal by fine tuning of the replica bitline height using simulations. The replica structure takes up only one additional column and hence has very little area overhead.



3.2 Implementation Environment

The cadence tool is used to develop the schematic and layout of the IO block. The design is done at 180nm technology node and 1V power supply (V_{DD}). The whole chip is characterized with HSIM provided by Synopsys tool.

The functional operating range and the data retention range is tabulated in the Table 1 and Table 2.

TABLE 1

Functional Operating Range

| Operating Range | Min | Max | Unit |
|-------------------|-----|-----|--------------------|
| Supply Voltage | 0.8 | 2.0 | V |
| Temperature Range | -40 | 125 | $^{\circ}\text{C}$ |
| Process Corners | SS | FF | - |

TABLE 2

Data Retention Range

| Operating Range | Min | Unit |
|-------------------|-----|--------------------|
| Supply Voltage | 0.8 | V |
| Temperature Range | 125 | $^{\circ}\text{C}$ |
| Process Corners | FF | - |

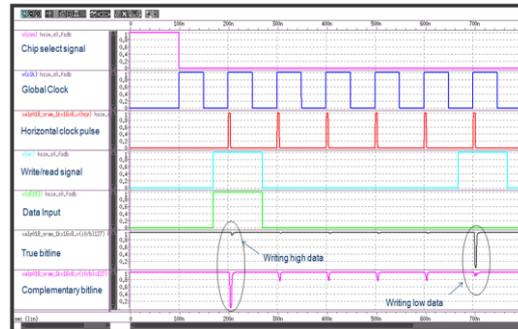
The aspect ratio W/L is taken in the range

$$0.6 < W/L < 1.8.$$

IV. MEASUREMENT RESULTS

The result shows the writing of high and low data on last bitlines $bl127$ and $blx127$. During write operation the write/read signal wr is high and data is available at the input node $d[15]$. When the hcp rises, the bit-line pre-charge signal de-asserts leaving bitlines $bl127$ and $blx127$ in un-driven state. Since both the bitlines have been

pre-charged to the V_{DD} level, both will stay in pre-charged state until the write occurs. The word-line (not shown) is then asserted to allow the horizontal clock pulse hcp to pull down the bitline $blx127$ to the ground, forcing a high logic value into the memory array. Once the content of the memory cell is fully stored in the memory cell, the write is succeeded and the word-line and the hcp are de-asserted. The pre-charge is again asserted to force both bitlines at V_{DD} .



The latch used in the write driver circuit is used to meet the data setup and hold time requirements. Before clock signal clk turns high, data should be presented at $d[i]$ for some time, which is the *data setup time* prior write. After the write operation, wr will turn low. Before $d[i]$ can change, the signal on $d[i]$ should already stay at least for a certain time, which is called the *data hold time* after write. The data setup time is calculated at the node N3 while the data hold time is calculated at node N2. The method used to find the data setup and hold time is given below:

$$\text{Data setup time} = \text{Data delay} - \text{Clock delay};$$

$$\text{Data hold time} = \text{Clock delay} - \text{Data delay}.$$

The data delay in the data setup and hold time is measured at different nodes. For data setup time, the data delay is the time between the $0.5V_{DD}$ of the data input $d[i]$ and the $0.2V_{DD}$ of rising data (or $0.8V_{DD}$ of falling data) available at node N3. The clock delay is the time between the $0.5V_{DD}$ of clk and $0.8V_{DD}$ of internally generated clock pc . While for data hold time, the data delay is the time between the $0.5V_{DD}$ of the data input $d[i]$ and the $0.2V_{DD}$ of rising data (or $0.8V_{DD}$ of falling data) available at node N2. The clock delay is same as in data setup time.

V. CONCLUSION

In this 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. The 8-bit IO block structure for the said SRAM is sketched. 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.

Finally the design of 128-bit IO block is presented . 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).A conclusion section must be included

and should indicate clearly the advantages, limitations, and possible applications of the paper. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

- [1] RF Silicon Technology Pvt. Ltd. (All the confidential specification provided by the company under Non Disclosure Agreement)
- [2] A. P. Chandrakasan, et. al., "Low-Power CMOS Digital Design", *IEEE Journal of Solid State Circuits*, vol. 27, no. 4, pg 473-484, April 1992.
- [3] Janm Rabaey , Anantha Chandrakasan , Borivoje Nikolic . "Digital Integrated Circuits", Second Edition 2004.
- [4] Tegze P.Haraszti, "CMOS memory circuits", Kluwer Academic Publishers", 2000, pg 165-275.
- [5] Vishwani D. Agrawal. CMOS SRAM Circuitry Design and Parametric Test in Nano-Scaled Technologies, edited by Pavlov, Andrei and Sachdev, Manoj. Location: Springer, January 01, 2008.
- [6] Yuh-Kuang, Tseng *Industrial Research and Technology Institut*, Chapter 49.
- [7] Meixner and J. Banik. Weak write test model: An SRAM cell stability design for test technique. In Proc. IEEE International Test Conference (ITC), pages 1043- 1052, November 1997.
- [8] James B Kuo, Jea-Hong Lou, "Low-Voltage CMOS VLSI Circuits", pg 235-343.
- [9] K. Itoh, *VLSI Memory Chip Design*. Springer-Verlag, 2001.
- [10] A. Agarwal, B. Paul, S. Mukhopadhyay, and K. Roy, "Process variation in embedded memories: Failure analysis and variation aware architecture," *IEEE J. Solid-State Circuits*, vol. 40, pp. 1804-1813, 2005.