A Novel Technique to Reduce Write Delay of SRAM Architectures

SWAPNIL VATS AND R.K. CHAUHAN*
Department of Electronics and Communication Engineering
M.M.M. Engineering College, Gorakhpur-273 010, U.P.
INDIA
*Email: rkchauhan27@gmail.com

Abstract:- This paper presents a novel circuit technique for improving the write delay of an SRAM cell. The technique is common for all the SRAM architecture. It utilizes a PMOS between power supply rail and the SRAM cell and an NMOS between SRAM cell and ground. The simulation results for write delay, SNM and power dissipation were presented with and without application of proposed technique on two different SRAM architectures. Significant improvements on write delay and power dissipation were noticed for the proposed modified SRAM architectures with less impact on SNM.

Key Words:- SRAM, SNM, NMOS, PMOS, DRAM, Butterfly curve.

1 Introduction

With unprecedented developments in semiconductor technology, it has become possible to make semiconductor memories of various types and sizes. These memories have become popular due to their small size, low cost, high speed, high reliability, and ease of expansion of the memory size**.** In fact these semiconductor memories provide a major advantage of digital over analog systems. These memories facilitate the storage of large quantities of digital information and data for large or short periods. This memory capability is what makes digital systems so versatile and adaptable to many situations**.** The amount of memory required in a particular system depends on the type of the application but in general, the number of transistors for the information (data) storage function is much larger than the number of transistors used for logic operations and other purposes. The ever-increasing demand for large data storage capacity has driven the fabrication technology and memory development toward more compact design rule and, consequently, toward higher data storage densities. Thus, the maximum realizable data storage capacity of single-chip semiconductor memory arrays approximately doubles every two years. On-chip memory arrays have become widely used subsystems in many VLSI circuits, and commercially available single-chip read/write memory capacity has reached few gigabits. This trend toward higher memory density and large storage capacity will continue to push the leading edge of digital systems design.

The R/W memory is commonly called *Random Access Memory* (RAM). Unlike sequential-access memories such as magnetic tapes, in RAMs any cell can be accessed with nearly equal access time. The stored data is volatile; i.e., the data stored is lost when the power supply voltage is turned off. In the last few decades random access memories (RAMs) have been the driving force behind the rapid development of CMOS technology. Based on operation type of individual data storage cells, RAMs are classified into two main categories. *Dynamic RAMs* (DRAM) and *Static RAMs* (SRAM). The DRAM cell consists of a capacitor to store binary information, 1 (high voltage) or 0 (low voltage), and a transistor to access the capacitor. Cell information (voltage) is degraded mostly due to a junction leakage current at the storage node. Therefore the cell data must be read and rewritten periodically (refresh operation) even when memory arrays are not accessed. On the other hand, SRAM cell is consists of a latch, therefore, the cell data is kept as long as the power is turned on and refresh operation is not required. Due to the advantage of low cost and high density, DRAM is widely used for the main memory in personal and mainframe computers, and engineering workstations, and memory in hand-held devices due to high densities and low power consumption.

DRAMs were first developed by Intel in the 1970s with a capacity of 1kb and in the last decade these capacities have reached to few Gb [3]. As battery-powered hand-held systems become popular, the power consumption and reliability of SRAM cell is a major concern in VLSI chip design.

This concern has driven most of the research work in the SRAM towards the improvement of reliability and power dissipation of SRAM architectures as evident from [5]-[9]. As shown in [5] by *Calhoun* and *Chandrakasan* the threshold variation has the greatest impact on SNM. In [6] again an efficient technique is presented for stability improvement. Similarly [8] investigates the considerations in the design of SRAM cells on SOI. In [9] a comparison regarding radiation robustness is presented between 8T and 6T SRAM architecture.

The growing gap between the Micro Processor Unit (MPU) cycle time and DRAM access time demands the introduction of several levels of caching in modern data processors. SRAM memories are preferred for this cache memories in personal computers MPUs. These are represented as L1 and L2 on chip embedded SRAM cache memories. SRAM cell's design is therefore a prime importance in today's IC design. To increase SRAMs operating speed along with its reliability is a major concern for researchers. Focus has been primarily made on the reliability concern that shows improvement in SRAM parameters like static noise margin (SNM), and power dissipation in the last decade [1]-[8].

Besides reliability and power dissipation of SRAM cell, delay in the operation of SRAM cells is one of the prime concerns which need to be addressed before the design of any SoC. This delay can be broadly categorised as the read delay and write delay. This paper presents a novel technique which can be used for reducing the write delay of an SRAM cell. The paper also considers the effect of proposed technique on power dissipation and SNM of two different SRAM cells architectures. All the SRAM architectures used in this work are symmetrical architectures and simulations were carried out using 120nm technology. A work which compares the noise margin and delays of 6T-SRAM architecture with two unsymmetrical 8T and 9T-SRAM architectures on 90 nm technology node is reported by Athe and S. Dasgupta. [7]

2 CMOS SRAM Cell Design

In the present study 6T-SRAM cell architecture is considered as the core element and it is shown as fig.1. In an SRAM cell design, some criteria must be

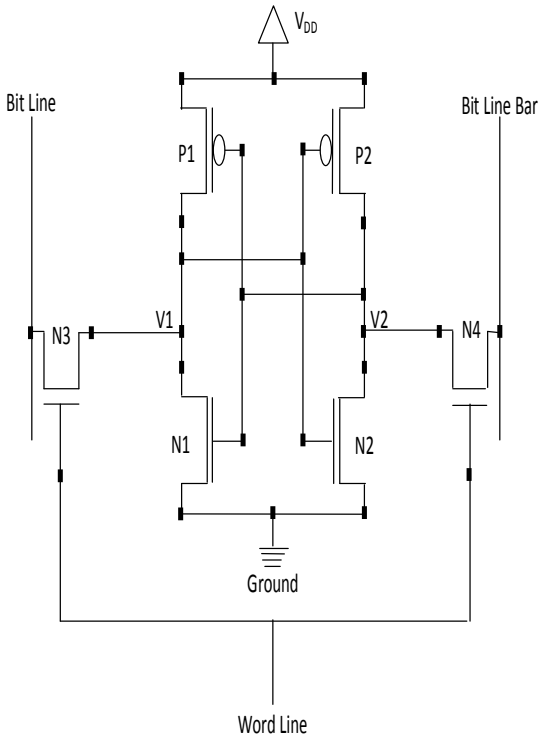

Figure 1: Basic 6T-SRAM architecture

Firstly consider the data-read operation. It is assumed here that logic "0" is stored in the cell. In the 6T-SRAM read operation, both the bit line and bit line bar is precharged to a high voltage level and the access transistors (N_3 & N_4 here) are made 'ON' by raising the voltage at the word line. The charge stored in the SRAM cells are sensed by data read circuitry which is responsible for detecting the small voltage drop in the voltage level of precharged bit line capacitance and amplifying it as a stored "0". The key design issue for the data-read operation is then to make the voltage V_1 not to exceed the threshold voltage of N_2 $\left(V_{T,n}\right)$, so that the transistor N_2 remains turned off during the read phase, i.e.,

$$V_{1,\max} \le V_{T,n} \qquad \ldots(1)$$

We can assume that, after the access transistors are turned on, the voltage level at bit line remains approximately equal to V_{DD}. Hence, N_3 operates in saturation while N_1 operates in the linear region.

$$\frac{k_{n,3}}{2}\left(V_{DD}-V_1-V_{T,n}\right)^2=\frac{k_{n,1}}{2}\left(2\left(V_{DD}-V_{T,n}\right)V_1-V_1^2\right). \qquad (2)$$

Combining this equation with equation (1) and taking the threshold voltages of N_1 and N_2 equal, results in:

$$\frac{k_{n,3}}{k_{n,1}} = \frac{(W/L)_3}{(W/L)_1} < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} - 2V_{T,n})^2} \quad (3)$$

Now the write "0" operation is considered, assuming that logic "1" is stored in the SRAM cell initially [1]. In this case the bit line voltage level is forced to logic "0" level by data-write circuitry and then access transistors are made 'ON', which changes the initial value of voltage V_1 to logic "0" level. To change the stored information, i.e., to force V_1 to logic "0" level and V_2 to V_{DD}, the node voltage V_1 must be reduced below the threshold voltage of N_2 $(V_{T,n})$, so that N_2 turns off first. When $V_1 = V_{T,n}$, the transistor N_3 operates in the linear region while P_1 operates in saturation region. Thus we can write:

$$\frac{k_{p,5}}{2}(0 - V_{DD} - V_{T,p})^2 = \frac{k_{n,3}}{2}(2(V_{DD} - V_{T,n})V_{T,n} - V_{T,n}^2) \quad (4)$$

Rearranging this condition results in:

$$\frac{k_{p,5}}{k_{n,3}} < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} + V_{T,p})^2} \quad (5)$$

$$\frac{(W/L)_5}{(W/L)_3} < \frac{\mu_n}{\mu_p} \cdot \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} + V_{T,p})^2} \quad (6)$$

2.1 Proposed Technique to Reduce Write-Delay of SRAM Cells

The general idea of the proposed technique is presented in fig.2. In this technique a PMOS and an NMOS pass transistors are used at the two extremes of SRAM cell. The gate of PMOS is always grounded where as the gate of NMOS is given supply of V_{DD} and SRAM cell is placed in between these two pass transistors. The beauty of this technique is that it can be applied to any SRAM architectures without affecting its reliability.

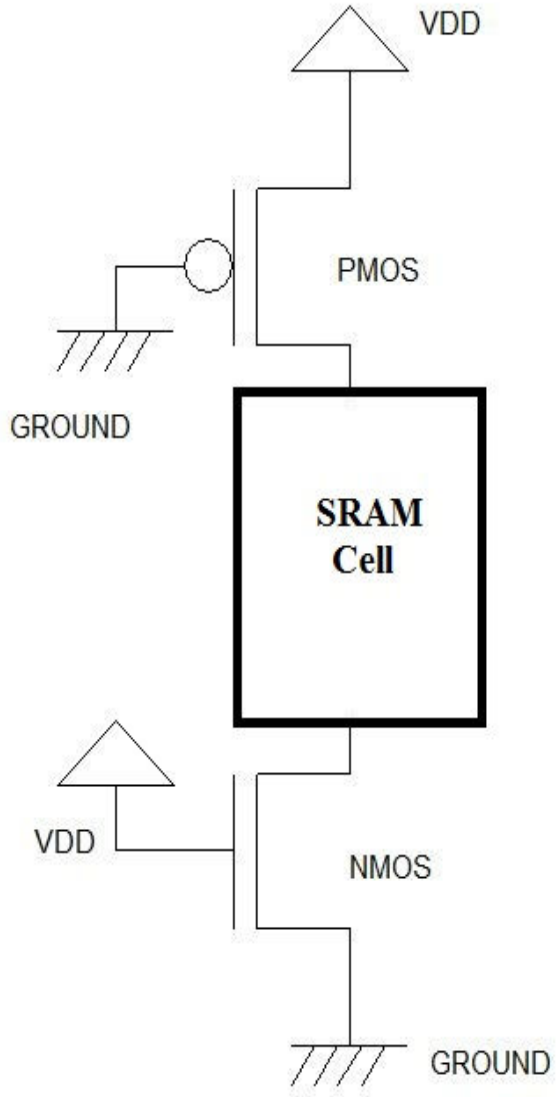

Fig. 2: Proposed technique for reducing write delay in SRAMs

2.2 Proposed technique applied on basic 6T-SRAM architecture

The architecture of basic 6T-SRAM is modified as per the proposed technique shown in fig.2. The architecture of modified 6T is shown in fig. 3a. The

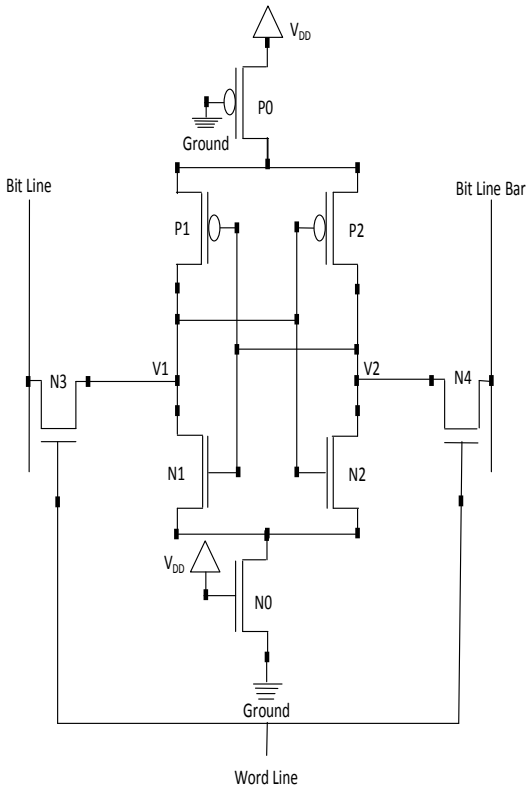

Fig. 3a: Modified 6T SRAM architecture

threshold voltages and W/L ratios of all the NMOS and PMOS transistors in modified architecture were identical. Fig.3b shows the butterfly curves for the 6T and modified 6T SRAM architecture. In which the solid lines represents the butterfly curve for 6T-SRAM architecture and dotted line represents the butterfly curve for modified 6T-SRAM architecture. In these butterfly curves both the horizontal and vertical axes represent the voltages in volts. In both the axes 1 unit is taken equal to .01 volts. Fig.3c and fig.3d shows the simulation results for 6T and its modified version respectively. In the simulation results the horizontal axis represents the timing in nano seconds and the vertical axis represents the voltage level in volts. It can be observed from butterfly curve that on applying the proposed technique on basic 6T-SRAM the degradation in SNM is quite low and was found to be 3.77%. But the improvements seen in write delay and power dissipation were quite significant and about 12.76% and 15.20% respectively over basic 6T-SRAM architecture.

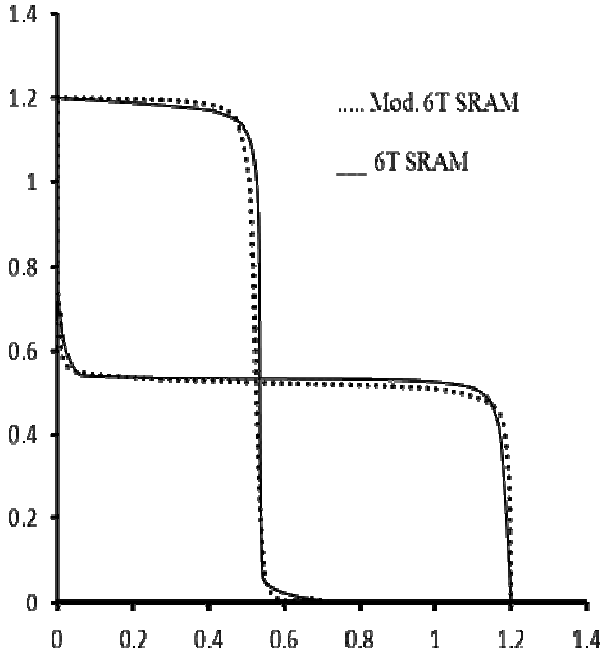

Fig.3b: Butterfly curves for
6T and 8T (modified 6T) SRAM architecture
Scale: Horizontal Axis- 1 Unit = .01 Volts
Vertical Axis- 1 Unit = .01 Volts

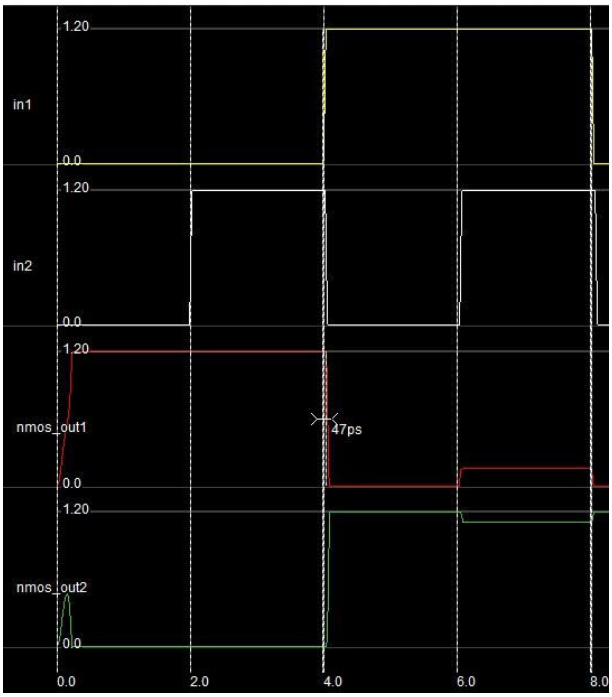

Fig.3c: Simulation result for basic
6T-SRAM architecture
Horizontal Axis- Time in Nano Seconds
Vertical Axis- Voltage level in Volts

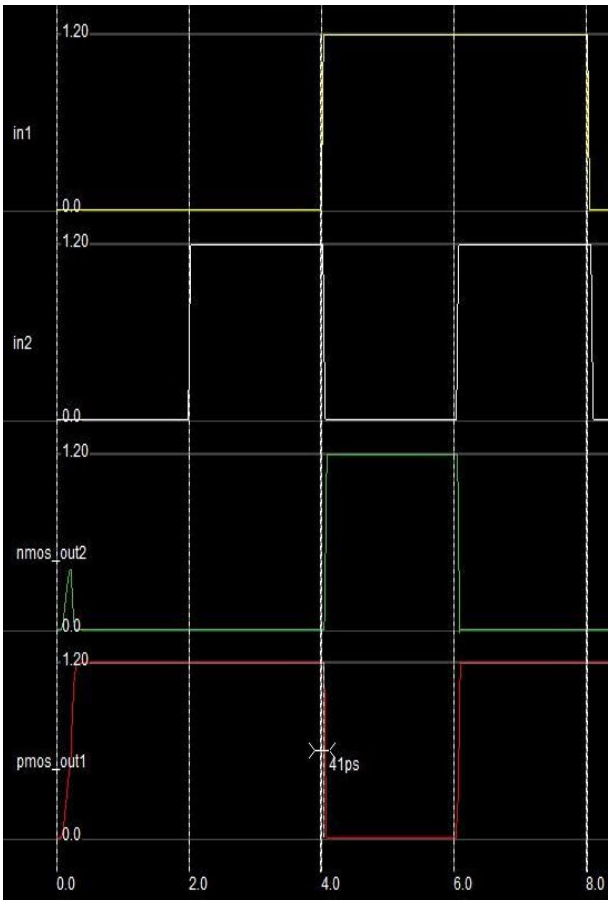

Fig. 3d: Simulation result of modified 6T-SRAM architecture
Horizontal Axis- Time in Nano Seconds
Vertical Axis- Voltage level in Volts

2.3 Proposed technique applied on 10T SRAM architecture [2]

The proposed technique (as shown in Fig.2) is now applied to another SRAM architecture so as to access the effectiveness of proposed technique. The circuit under consideration is 10T SRAM given by Sheshadri and Houston [2] as shown in fig. 4a. The device parameters taken for simulation in the circuits (10T as well as modified 10T SRAM obtained after applying proposed technique) have identical V_T and W/L. Fig. 4b shows the modified 10T SRAM architecture. Fig.4c shows the butterfly curves for the 10T and modified 10T SRAM architecture. In which the solid lines represents the butterfly curve for 10T-SRAM architecture and dotted line represents the butterfly curve for modified 10T-SRAM architecture. In these butterfly curves both the horizontal and vertical axes represent the voltages in volts. In both the axes 1 unit is taken equal to .01 volts. Fig.4d and Fig.4e shows the simulation results for 10T and its modified version respectively. In the simulation results the horizontal axis represents the timing in nano seconds and the vertical axis represents the voltage level in volts. The result shows significant improvement in write delay of modified 10T-SRAM, about 24.56% (fig 4e) over 10T-SRAM (fig 4d), with little degradation seen in SNM about 3.00% (fig 4c). Significant improvement in power dissipation was also noticed. Power dissipation reduces by 26.52%.

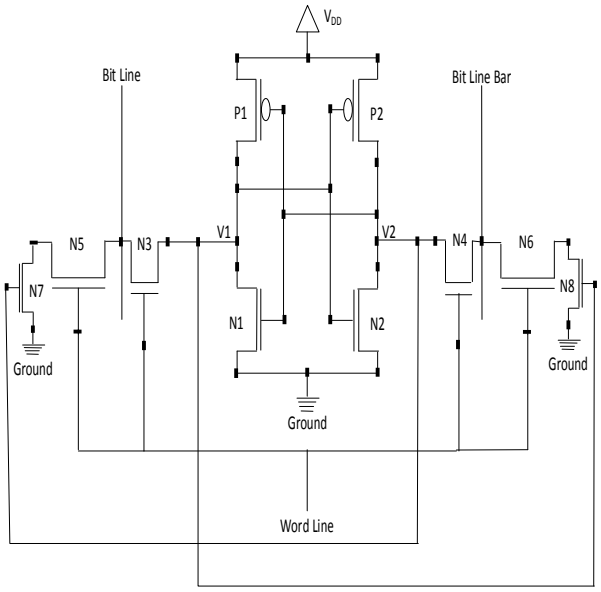

Fig. 4a: 10T SRAM architecture proposed by Sheshadri and Houston [2]

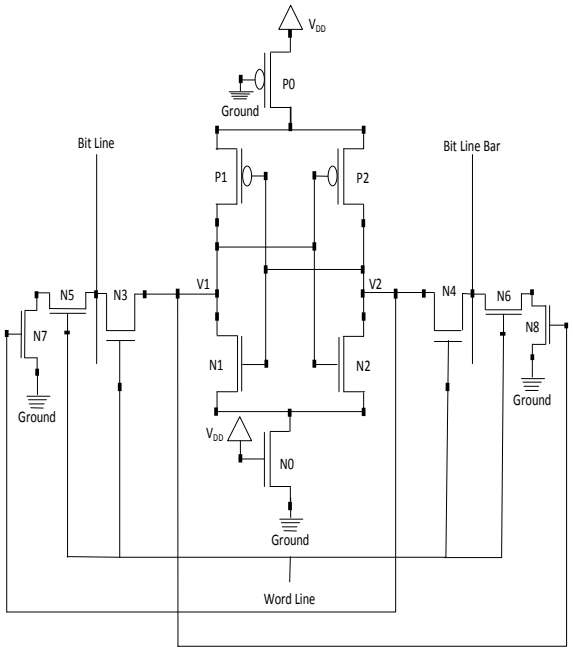

Fig. 4b: Modified 10T SRAM architecture

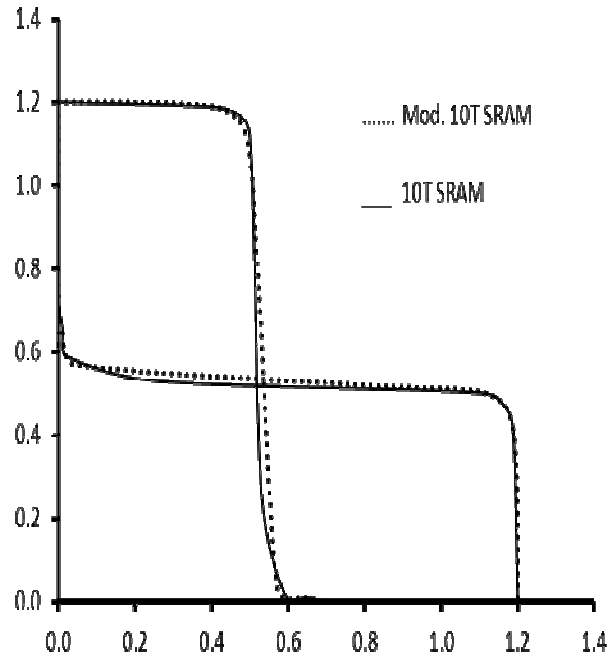

Figure 4c: Butterfly curve of 10T and modified 10T SRAM architecture
Scale: Horizontal Axis- 1 Unit = .01 Volts
Vertical Axis- 1 Unit = .01 Volts

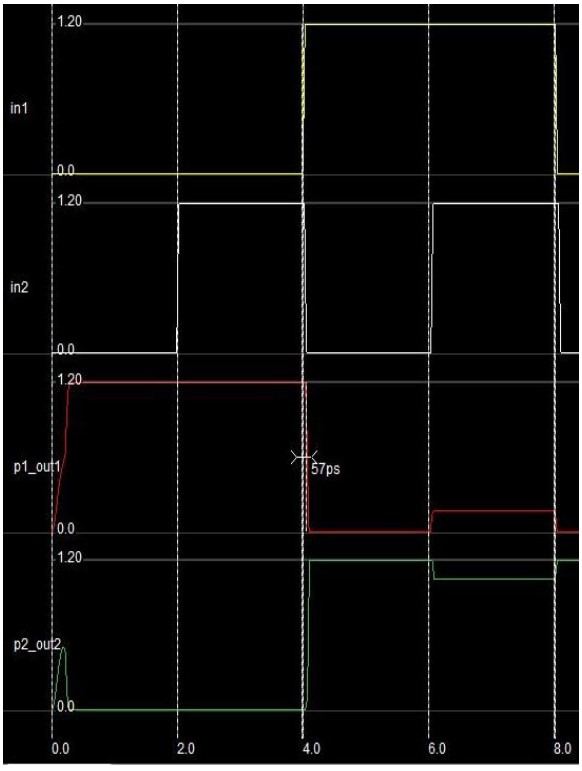

Fig. 4d: Simulation result for 10T-SRAM architecture with same V_T and W/L NMOSs
Horizontal Axis- Time in Nano Seconds
Vertical Axis- Voltage level in Volts

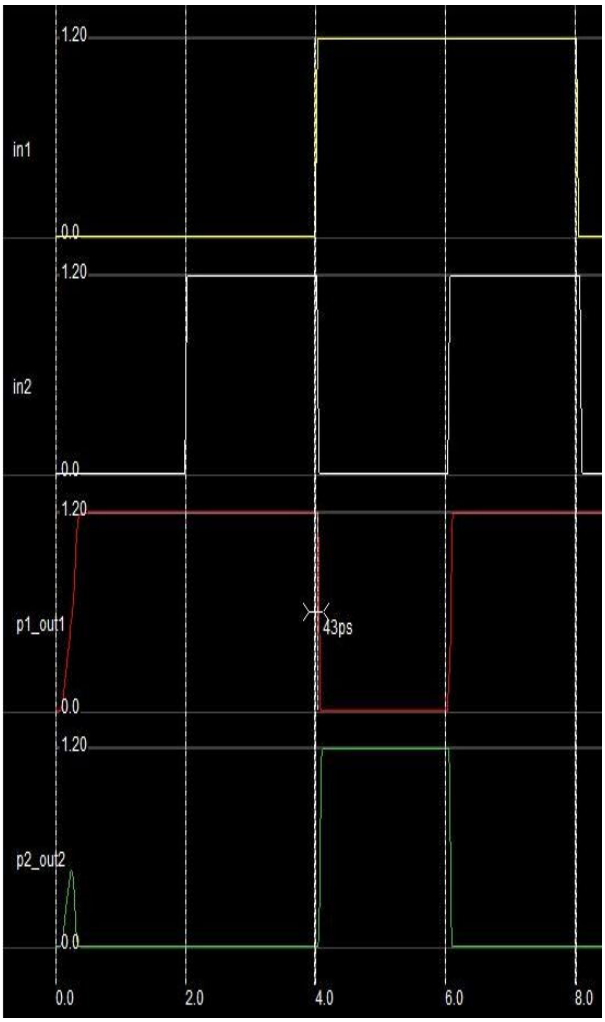

Fig. 4e: Simulation result of modified 10T-SRAM
Horizontal Axis- Time in Nano Seconds
Vertical Axis- Voltage level in Volts

2.4 An 8T- SRAM architecture with transmission Gates in place of access transistor

In this section the proposed technique is applied on an SRAM cell architecture which utilizes the transmission gates in place of access transistors connected to its word line. Improvements in static and dynamic performance are obtained when the switches are implemented with CMOS transmission gates. Researchers are working with transmission gates at different positions in SRAM architecture [8]. Fig. 5a shows the SRAM cell architecture which utilizes the transmission gates in place of access transistors connected to its word line. While simulating this architecture the threshold voltages and W/L ratios of all the NMOS and PMOS are identical.

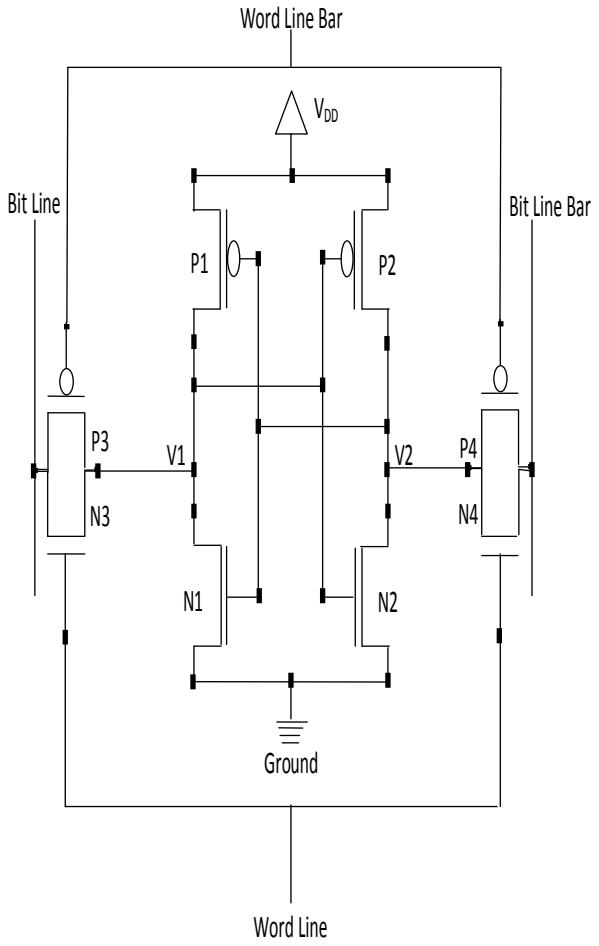

Fig. 5a: 8T-SRAM architecture with transmission gate in place of access transistor

The proposed technique for write delay improvement is applied on the above 8T-SRAM architecture resulting in a modified 8T-SRAM architecture which uses transmission gate in place of access transistors as shown in fig. 5b. Fig. 5c presents the butterfly curve for 8T and modified 8T with transmission gate. In which the solid lines represents the butterfly curve for 8T-SRAM architecture and dotted line represents the butterfly curve for modified 8T-SRAM architecture. In these butterfly curves both the horizontal and vertical axes represent the voltages in volts. In both the axes 1 unit is taken equal to .01 volts. Fig. 5d and fig. 5e represents the simulation results for 8T SRAM (with transmission gate in place of NMOS access transistor) and modified 8T SRAM architecture respectively. In the simulation results the horizontal axis represents the timing in nano seconds and the vertical axis represents the voltage level in volts. Here again the threshold voltages and W/L ratios of all the NMOS and PMOS are identical. The dotted line in fig. 5c represents butterfly curve for the modified 8T SRAM architecture and continuous line represents butterfly curve for 8T SRAM architecture. The SNM for 8T SRAM is .53 volt and for modified 8T is .51 volts (Fig. 5c) which represents the degradation in SNM is of 3.77% From simulation results in fig. 5d and fig. 5e, the improvement in write delay is about 8.89% while improvement in power dissipation is not significant in this particular case where transmission gates are used in place of NMOS access transistors.

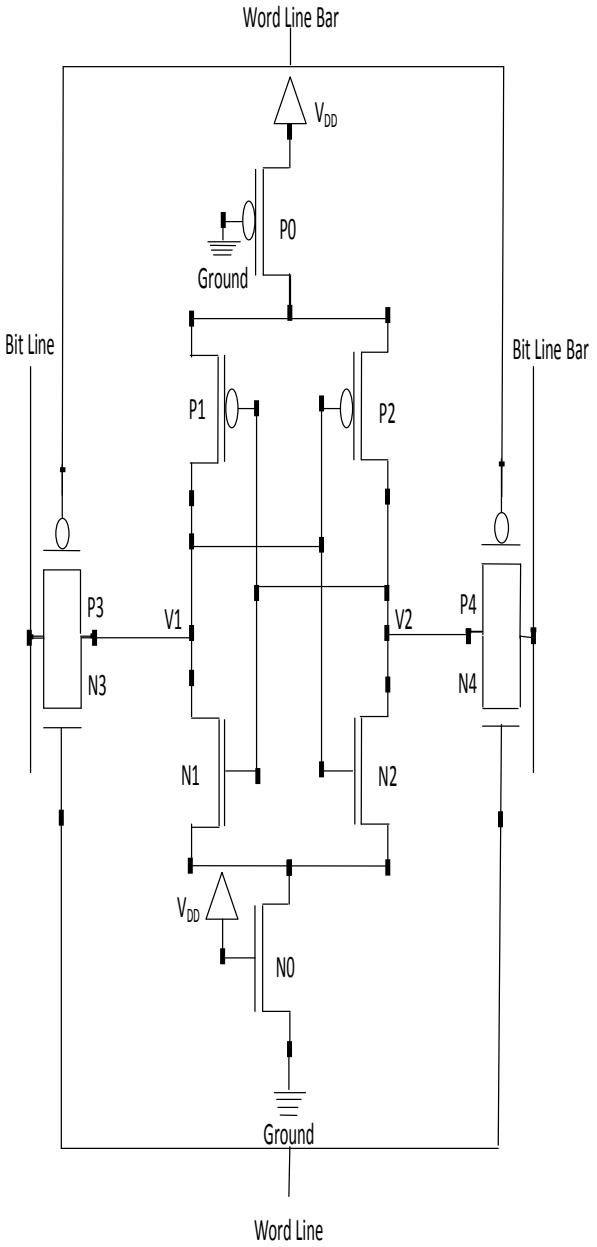

Fig. 5b: Modified 8T SRAM architecture with transmission gate in place of access transistor

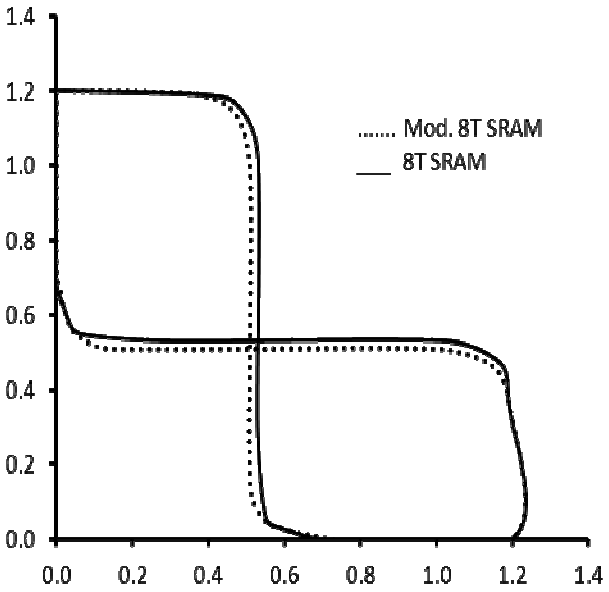

Fig. 5c: Butterfly curve for 8T-10T (modified 8T with transmission gate) SRAM architecture with transmission gate in place of access transistor
Scale: Horizontal Axis- 1 Unit = .01 Volts
Vertical Axis- 1 Unit = .01 Volts

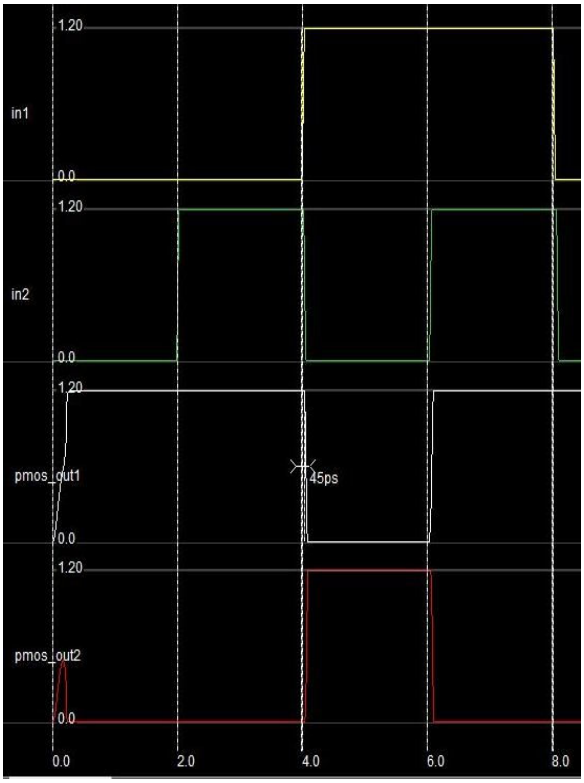

Fig. 5d: Simulation result for 8T-SRAM architecture with transmission gate in place of access transistor
Horizontal Axis- Time in Nano Seconds
Vertical Axis- Voltage level in Volts

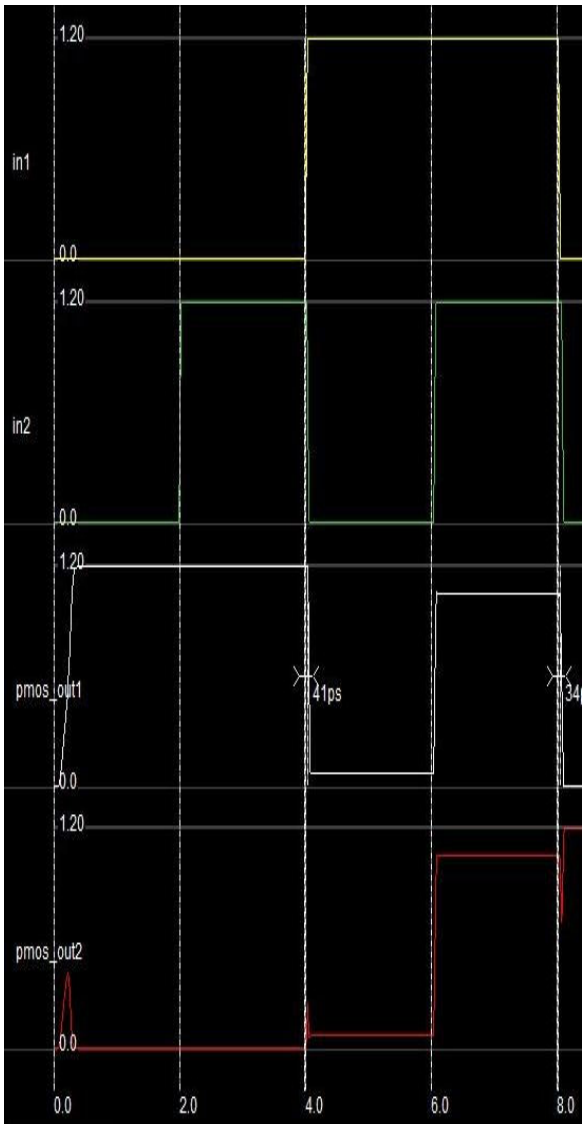

Fig. 5e: Simulation results for 10T-SRAM (modified 8T with transmission gate) architecture with transmission gate in place of access transistor
Horizontal Axis- Time in Nano Seconds
Vertical Axis- Voltage level in Volts

The results mentioned in the Table-1 were obtained after simulation for various SRAM architectures using 120 nm technology. The three important parameters obtained from simulation were SNM, write delay and power dissipation.

The analysis of simulation results reveals that the proposed technique provides write delay improvement in both the cases of 6T and 10T-SRAMs and hence one can say that it will show significant improvement in other cases also. The chip area consumed by the two transistors will be

insignificant as it can be used for several SRAM cells connected in parallel.

Table.1: Performance parameter of basic SRAM architecture and its modified version (as per the proposed technique)

S.No	SRAM Arch.	Write Delay (in ps)	SNM (inV)	Power Dissipation (in µW)
1.	Basic 6T-SRAM	47	.53	10.837
2.	Modified 6T-SRAM	41	.51	9.189
3.	10T-SRAM architecture	57	.50	17.992
4.	Modified 10T-SRAM architecture	43	.49	13.219
5.	8T-SRAM architecture with transmission gates in place of access transistors	45	.53	14.911
6.	Modified 8T-SRAM architecture with transmission gates in place of access transistors	41	.51	14.90

3 Conclusion

This work proposes a new technique for reducing write delay of an SRAM cell. The proposed technique is then applied on three different SRAM architectures. First circuit on which this technique was applied was basic 6T SRAM where as the other circuit was 10T SRAM given by Sheshadri and Houstan. The third architecture on which the proposed technique is applied is one in which the transmission gates are used in place of access transistors. Simulation results for all three cases, before and after the application of proposed technique, were obtained and listed in Table-1. The three fold advantage associated with this technique as evident from the above study is (a) The amount of write delay improvement is significant and (b) Reduction in power dissipation were significant compared to the degradation of SNM (c) The application of proposed technique is independent of the SRAM architectures. Although the application of proposed technique increases the area required by the SRAM architecture but reduction in speed and power dissipation can be helpful in the applications where area occupied by SRAM architectures is not major concern and a fast and power efficient SRAM architecture is the prime requirement. All the SRAM architectures used in the present work are symmetrical SRAM architectures. As a future scope of this technique it will be interesting to see the behavior of proposed technique on non symmetrical SRAM architectures.

References:

[1]. "*CMOS Digital Integrated Circuits-Analysis and Design"* by Sung-Mo (Steve) Kang, Yusuf Leblebici Tata McGraw-Hill New Delhi (Third Edition) (2003).

[2]. Anand Seshadri and Theodore W. Houston, "The Dynamic Stability of a 10T SRAM Compared to 6T SRAMs at the 32nm Node Using an Accelerated Monte Carlo Technique," *IEEE conference (2008)*. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04695924

[3]. Andrei Pavlov, *"CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies"* Process Aware SRAM Design and Test, Springer ISBN 978-1-4020-8362-4, (2008).

[4]. Evert Seevinck, Frans J. List, and Jan Lohstroh, "Static-Noise Margin Analysis of MOS SRAM Cells," *IEEE Journal of Solid-State Circuits*, Vol. SC-22, No. 5, pp. 748-754, (1987).

[5]. Benton H. Calhoun, and Anantha P. Chandrakasan, "Static Noise Margin Variation for Sub-threshold SRAM in 65-nm CMOS,"

[6]. *IEEE Journal of Solid-State Circuits*, Vol. 41, NO. 7, pp. 1673-1679, (July 2006).

[7]. Yeonbae Chung, Seung-Ho Song, "Implementation of low-voltage static RAM with enhanced data stability and circuit speed," *Microelectronics Journal* 40, pp. 944-951. (2009).

[8]. Paridhi Athe and S. Dasgupta, "A Comparative Study of 6T, 8T and 9T Decanano SRAM cell," *IEEE Symposium on Industrial Electronics and Applications (ISIEA)*, Kuala Lumpur, Malaysia, pp. 889-894, (2009).

[9]. He Wei Zhang Zheng – xuan, Zhang En-xia, Yu Wen-jie, Tian Hao, Wang Xi, "Practical considerations in the design of SRAM cells on

SOI," *Microelectronics Journal* 39, pp. 1829–1833, (2008).