Design of Low Power SRAM using Hierarchical Divided Bit-line Approach in 180-nm Technology

Mohd. Sayeeduddin Habeeb [#1],

Lecturer

King Khalid University

Saudi Arabia

Md, Salahuddin [#2]

Asst. Professor

Azad College of Engineering and Technology

Hyderabad, India

Abstract— **Fast low power SRAMs have become a critical component of many VLSI chips. This is especially true for microprocessors, where the on-chip cache sizes 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. Thus, a significant effort has been invested in reducing the power of CMOS RAM chips using circuit and architectural techniques. This paper presents a design using hierarchical divided bit-line approach for reducing active power in SRAMs by 40-50% and access time at the expense of 5-10% increase in the number of transistors when compared to Conventional SRAM. A Hierarchical divided bit line approach is chosen to implement a 1Kb SRAM memory on 0.18 micron CMOS technology using CADENCE design tool.**

Keywords— SRAM Cell, Power, Hierarchical divided bit line approach, bit-line capacitance.

I. INTRODUCTION

Static random access memory (SRAM), being the fastest of the currently high volume manufactured memory families, continues to be a critical component for a multitude of electronic applications. Several techniques have been proposed to reduce the power consumption of SRAMs .The read power is reduced by limiting the swing voltages of bit lines and data bus to small voltages during read cycles. However, the SRAM consumes much larger power during write cycles than read cycles due to the full swing property in the bit lines and data bus during write cycles. If the same voltage swing is allowed, then power consumed during read or write is directly proportional to the capacitance of bit-line. Bit-line capacitance can be reduced by the proposed divided bit-line approach, where the number of transistors connected to the bit-line is reduced by combining two or more SRAM cells. Combination of two or more SRAM cells results in division of bit-line into several sub bit lines. These sub bit-lines are further combined to obtain two or more levels of hierarchy The basic idea behind this extension is to make use of these sub bit-lines, which have much smaller capacitance, to develop a limited voltage swing on the bit-line. This division of bit-line into hierarchical sub bit-lines results in reduction of bit-line capacitance, which reduces active power.

II. SRAM CELL

Memory cells are the key components of any SRAM unit. An SRAM cell can store one bit of data. An SRAM cell comprises two back-to-back connected inverters forming a latch and two access transistors. Access transistors serve for read and write access to the cell.

An SRAM cell offers the following basic properties. [3]

(a)The data read operation should not destroy the stored information in the SRAM cell.

(b)The cell should allow modification of the stored information during the data write phase.

Read Operation

Consider the data read operation first, assuming that logic "0" is stored in the cell. The voltage levels in the CMOS SRAM cell at the beginning of the "read" operation are depicted in Figure 1. Here, the transistors M2 and M5 are turned off, while the transistors Ml and M6 operate in the linear mode. Thus, the internal node voltages are V1 = 0 and V2 = VDD before the cell access (or pass) transistors M3 and M4 are turned on. The active transistors at the beginning of the data-read operation are highlighted in Figure 1.
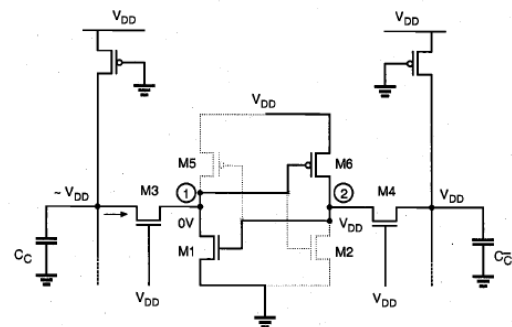

Figure 1 Voltage levels in the SRAM cell at the beginning of the "read" operation

After the pass transistors M3 and M4 are turned on by the row selection circuitry, the voltage level of column C will not show any significant variation since no current will flow through M4. On the other half of the cell, however, M3 and Ml will conduct a nonzero current and the voltage level of column C will begin to drop slightly. The voltage difference is fed to sense amplifier.

Write Operation

Consider the write "0" operation, assuming that logic "1" is stored in the SRAM cell initially. Figure 2 shows the voltage levels in the CMOS SRAM cell at the beginning of the data write operation. The transistors M1 and M6 are turned off, while the transistors M2 and M5 operate in the linear mode. Thus, the internal node voltages are V = *VDD* and *V2*= 0 *V* before the cell access (or pass) transistors M3 and M4 are turned on.
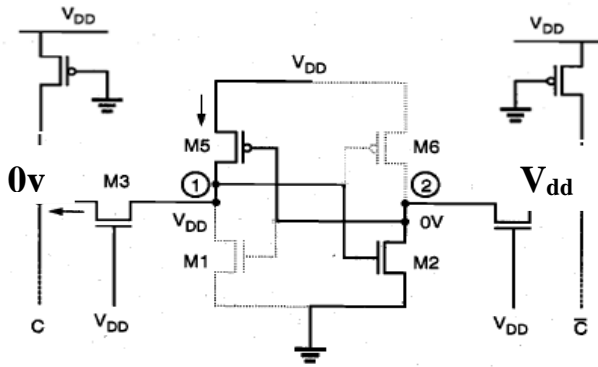
Figure 2 Voltage levels in the SRAM cell at the beginning of the "write" operation.

The column voltage VC is forced to logic "0" level by the data-write circuitry; thus, it may be assumed that VC is approximately equal to 0 V. Once the pass transistors M3 and M4 are turned on by the row selection circuitry, we expect that the node voltage V2 remains below the threshold voltage of Ml. Consequently, the voltage level at node 2 would not be sufficient to turn on Ml. To change the stored information, i.e., to force V, to 0 V and V2 to *VDD,* the node voltage V1, must be reduced below the threshold voltage of M2, so that M2 turns off first. Note that a symmetrical condition also dictates the aspect ratios of M6 and M4.

III. CONVENTIONAL SRAM

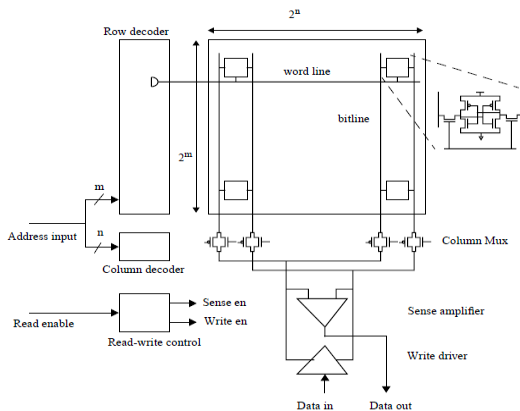The Figure 3 illustrates the Conventional SRAM memory design consisting of array of SRAM memory cells.

Figure 3 Conventional SRAM Memory Design

In a read operation, the bit lines start pre charged to some reference voltage usually close to the positive supply. When word line turns high, the access transistors connected to the cell node storing a '0' starts discharging the bit line, while the complementary bit line remains in its pre charged state, thus resulting in a differential voltage being developed across the bit line pair. SRAM cells are optimized to minimize the cell area, and hence their cell currents are very small, resulting in a slow bit line discharge rate. To speed up the RAM access, sense amplifiers are used to amplify the small bit line signal and eventually drive it to the external world During a write operation, the write data is transferred to the desired columns by driving the data onto the bit line pairs by grounding either

the bit line or its complement. If the cell data is different from the write data, then the '1' node is discharged when the access transistor connects it to the discharged bit line, thus causing the cell to be written with the bit line value.

IV HIERARCHICAL DIVIDED BIT LINE APPROACH

Divided Bit-Line Approach

Power consumption in SRAMs, for a normal read cycle, is given by the below expression

$$P = Vdd \times I_{dd} \qquad (1)$$
$$I_{dd} = (mIact\Delta t + C_{pt} Vint) f + Idcp \qquad (2)$$

where, Vdd is an external supply voltage, Idd is the total current, I act is the effective active current, Vint is an internal supply voltage, CPT is the total capacitance of the peripheral circuits, I_{dcp} is the total static current, m is the number of columns and f is the operation frequency. This equation is based on the fact that in SRAMs, holding current is very small and decoder charging current is negligible because of NAND decoders. To reduce the total power consumption, active current should be reduced as it dominates the total current. Active current is the current that flows during word line activation, i.e., during charging or discharging of bit-line capacitance. This active current is directly proportional to bit-line capacitance [6].

Here the approach presented reduces I_{dd} by reducing the active current, I act .The total effective charging current owing through a bit-line, during a read operation, can be expressed as

$$I_{eff} = C_{eff} \times \Delta V/\Delta T \qquad (3)$$

where C_{eff} is the effective bit-line capacitance, ΔV is the voltage swing and ΔT is the word line activation time. Similarly, expression for power can be written

$$P= (I_{eff} \times V \times \Delta T) \times f \qquad (4)$$

The bit-line capacitance is mainly composed of the drain capacitance of the pass transistors of the SRAM cell and metal capacitance of bit-line.
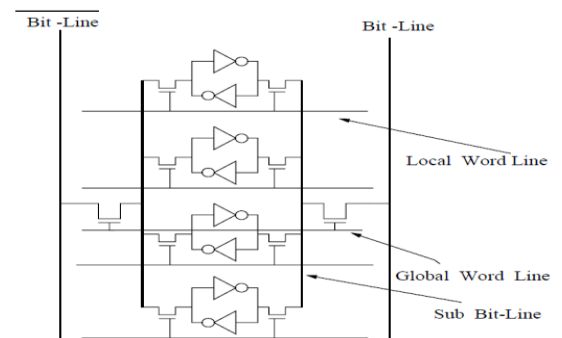

Figure 4 Divided Bit Line Approach

To reduce this capacitance, drain capacitance and metal capacitance should be reduced.

Theoretical Basis

In this subsection, we provide a theoretical basis for our approach and derive an optimal value for the number of SRAM cells to be combined [16].
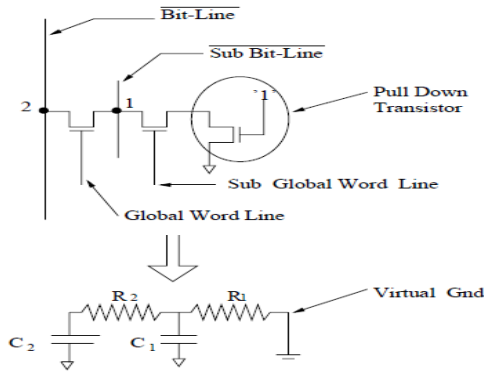
Figure 5 R-C model for Pass transistors reading '1' stored in SRAM cell

Figure 5 shows the modeling of pass-transistors as RC chain. As bit-lines are always precharge before reading, analysis is performed for bit line (or bit-line) which has to be pulled down to read a `1' (or a `0'). The access time Delay can be written in terms of parameters TP, TD2 and TR2 as

$$T_{d2} - T_{r2} + T_{r2} \ln \{T_{r2}/T_p[V(t)]\} \leq T_{delay} \leq T_p - T_{r2} + \ln \{T_{d2}/T_p[v(t)]\} \quad (5)$$

Where V (t) is the normalized voltage with respect to supply and is given by $1-(\Delta V/v)$.

The parameters T_P, T_{D2}, T_{R2} are given by following equations are obtained from Elmore delay model

$$T_p = T_{d2} = R_1C_1 + (R_1 + R_2) C2 \quad (6)$$
$$T_{r2} = R_1^2C_1/ (R1 + R2) C2 + (R_1+R_2) C1 \quad (7)$$

Where R1 is the resistance of the pass-transistor 1 and R2 is the resistance of the pass-transistor 2. C1 and C2 are the capacitances at node 1 and at node 2, respectively, in above Fig 5.

Let us assume that the number of rows in the memory array is N and the number of cells combined in the divided bit-line is denoted by M. Then, the capacitances C1 and C2 can be obtained as

$$C1 = C (M + 1)/N \quad (8)$$
$$\text{For } N \geq 1; C = C (M)/N \quad (9)$$
$$C2 = C/M + 0.1 \times C \quad (10)$$

where C denotes the original drain capacitance of N rows. Metal capacitance contribution to total bit-line capacitance is assumed to be 10% of the total drain capacitance. To make a first order approximation, the resistances of two pass-transistors are assumed to be equal. From the above equations and assumptions the delay is found as

$$T_{delay} = RC \{M/N + 2/M + 0.2\} \times \ln (1/V (t)) \quad (11)$$

Using this approximated delay we can write normalized delay as

$$T_{delay} \text{ (normalized)} = 1/1.1 \times [(M/N) + (2/M) + 0.2] \quad (12)$$

Now, an optimal value of M for minimum power consumption is found. The optimal Value for number of cells to be combined is given by differentiating (12) w.r.t 'M'.

Where N is number of rows, Δv is voltage swing.
$$M_{optimal} = \sqrt{[(N/2) \times (\Delta V/v)]} \quad (13)$$

Power consumption, as a function of N and M, can be expressed as

$$\text{Power} = f (M) = \{C_2 \times \Delta v \times V + 2 \times C_1 \times V^2\} \times f$$
$$= [\{C/M + 0.1 \times C\} \times \Delta v \times V + 2 \times C(M/N) \times V^2] \times f \quad (14)$$

Hierarchical Divided Bit-Line Approach

In high density SRAMs, the number of sub bit lines will increase and even with divided bit-line architecture, bit-line capacitance can be significant. From this point of view, hierarchical divided bit-line architecture is developed [2]. Figure. 6 shows the concept of hierarchical divided bit-line approach.

In this architecture, the bit line is divided into more than two levels. Let us assume that the total number of levels in the hierarchy is L and at each level I, the number of blocks combined to form a new block is Mi [5]. Then capacitance, C_i at each node is given by

$$C_I = C (M_i+1)/N \quad (15)$$
$$C_L = C \times 1/(\prod_{I=1}^{L-1}) \quad (16)$$

Using these capacitances and again making the assumption that resistance values of all pass transistors are equal, the expression for active power can be written as

$$\text{Power} = f (M1, .M_{L-1}; L)$$
$$= 2V^2C/N \times \sum_{i=1} M_i + C\Delta VV/ M_i + 0.1 C\Delta VV \quad (17)$$

Similarly the expression for delay of hierarchical bit line can be written as

$$T_{delay} = RC \times {}^{L-1}\sum_{i=1} {}_l[M_I/N] + L/(\prod_{I=1}^{L-1} M_I) + 0.1 \times L \quad (18)$$
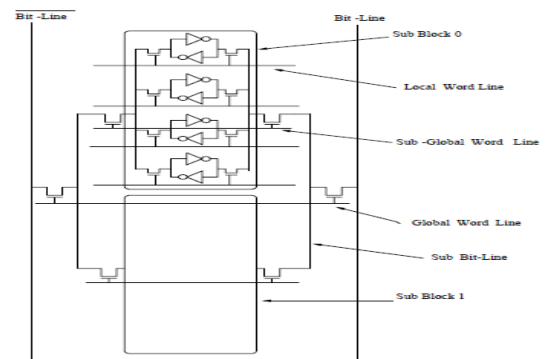

Figure 6 Hierarchical divided bit Line Approach

To obtain the optimum parameter for the number of blocks combined at each level, we differentiate (16) partially with respect to all Mi's and solve the following equation

$$\partial f/\partial M_i = 0 \text{ for } i-1 \text{ to } L-1 \quad (19)$$

The solution to the above equation gives the optimum values for Mi's, which are given by

$$M_1 = M_2 = ... = M_{L-1} = (N/2 \times \Delta v/v)^{1/L} \quad (20)$$

By substituting these values for Mi's expression for Power is obtained as

$$\text{Power} = (2V^2C/N) L \times (N/2 \times \Delta v/v)^{1/L} \quad (21)$$

V PERFORMANCE ANALYSIS

The power consumption for the SRAM memory design using various approaches is compared in this section. The Figure 7 illustrates the power consumption during write operation for 64-bit SRAM memory design using DBL and HDBL.
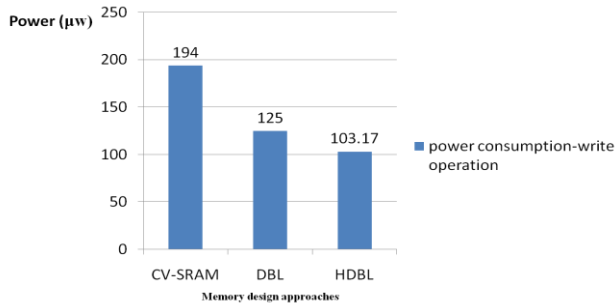

Figure 7 Power consumption – SRAM memory (64-bit)

The power consumed during the write operation is measured in µw. The DBL approach saves approximately 35% of power when compared to CV-SRAM memory design. The HDBL saves around 46% of power during write operation. The Figure 8 illustrates the power consumption during both read/write operation for 1KB SRAM memory design using various approaches. The power consumed is measured in µw.
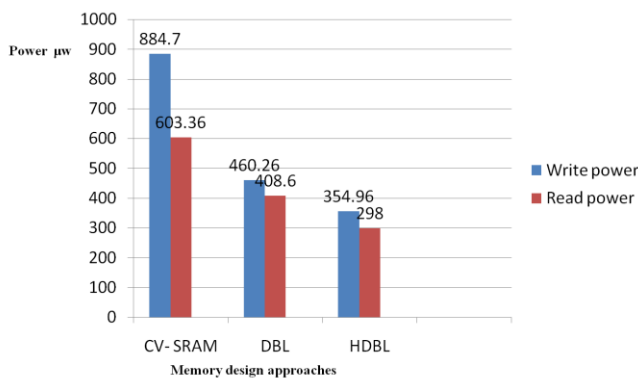
Figure 8 Power consumption –Memory design (1KB)

Bit Line Capacitance Analysis
Bit line capacitance is the major concern for power consumption in SRAM.
In general the drain capacitance C_{drain} = 0.5fF-1fF, Wire capacitance, C_{wire} = 0.2fF/micron of wire. The bit line capacitance in general is given by
C bit = (source/drain cap + wire cap + contact cap) x no. of cells in column

1) For Conventional SRAM the bit line capacitance is as shown below
Bit line capacitance/column = 16fF
Total bit line capacitance = 512fF

2) For divided bit line approach the capacitance across bit line is divided as C1 and C2 where
C1 = C (M/N) and C2 = [C/M] +) 0.1C
Where N= number of rows, M= no of cells combined and C= original drain capacitance.
Bit line capacitance /Column is given by = 10.56fF
Total bit line capacitance = 337.92fF.

3) For hierarchical divided bit line approach the capacitance C_i indicates the capacitance at each node i, and C_L indicates capacitance at each level given by
$C_i = C [M_i/N]$ Where N=number of rows and Mi = No of cells combined
Bit line capacitance/Column = 8.0624fF, Total bit line capacitance = 257.99fF.

Bit line Capacitance Comparison

Approaches	SRAM	DBL	HDBL
Bit-line capacitance/column	16fF	10.56fF	8.0624fF
Total bit line capacitance	512fF	337.92fF	257.99fF

This illustrates the reduction in the value of capacitance in DBL and HDBL approach when compared to conventional SRAM.

VI SIMULATION RESULTS

SRAM Cell
The Schematic entry and layout of the propose hierarchical divided bit line approach is done using Cadence tool on 180nm. Figure 9 shows the SRAM cell schematic entry. The design of the cell involves the selection of transistor sizes for all six transistors to guarantee proper read and write operations.
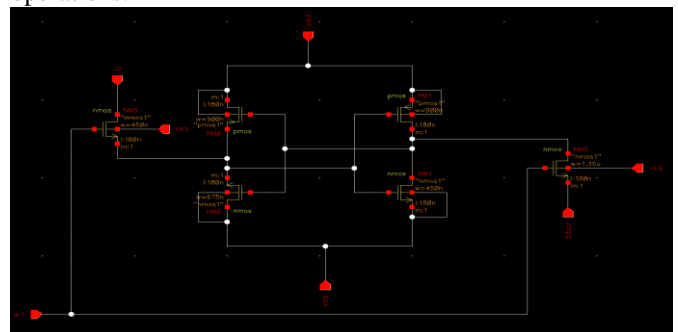

Figure 9 SRAM Cell Schematic Entry

The Figure 10 illustrates the Layout view of the 6T CMOS SRAM cell. The area of the SRAM cell is 14 µm. The power consumed is 6.426µw.
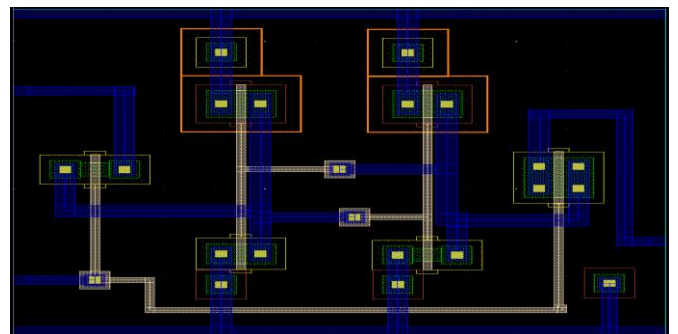

Figure 10 SRAM Cell Layout

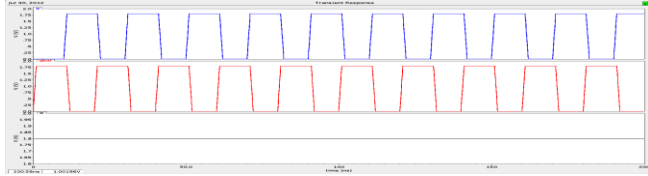The Figure 11 illustrates the simulation results of the SRAM cell.

Figure 11 SRAM Cell Simulation Result

Sense Amplifier Circuit

The sense amplifier used is the latch based sense amplifier shown in figure 5.3(a). The sense amplifier circuit consists of Sense enable signal which is enabled during read operation. The output consists of Out and Out bar pins in order to access the data from the cell selected with the help of decoder circuits.
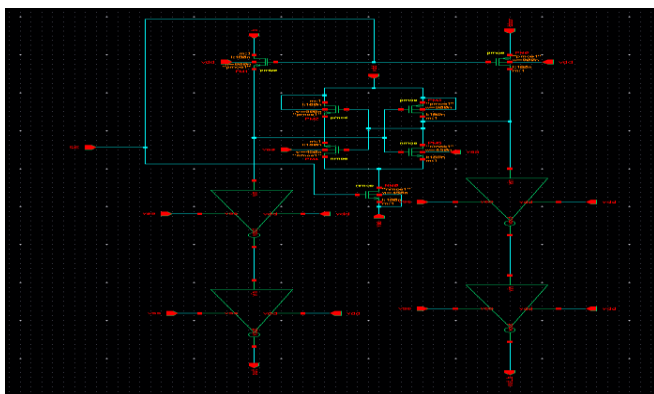

Figure 12 Sense Amplifier Schematic

The Figure 13 illustrated the layout sketch of the sense amplifier circuit schematic. The area of the Sense amplifier circuit is 1.5 μm. The power consumption is 99.01μw. The Figure 14 illustrates the simulated results of the sense amplifier circuit for which the schematic and layout sketch shown . The X axis is labeled as Time-ns. The Y axis consists of the parameters sense enable (SE) signal, Out, out bar.

Write Driver Circuit

Simplified write circuitry for the SRAM memory operation is shown below in Figure 15. The input signal is viewed only when the write enable signal is enabled.
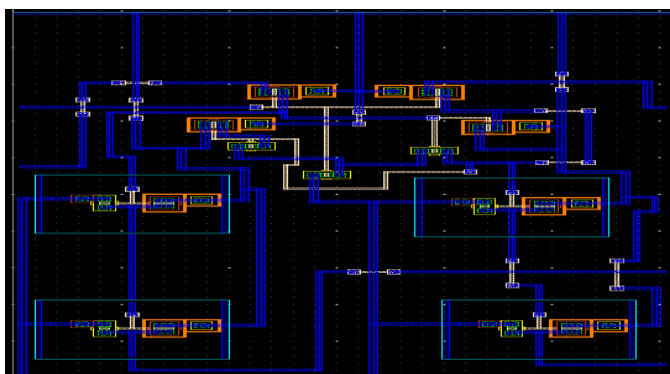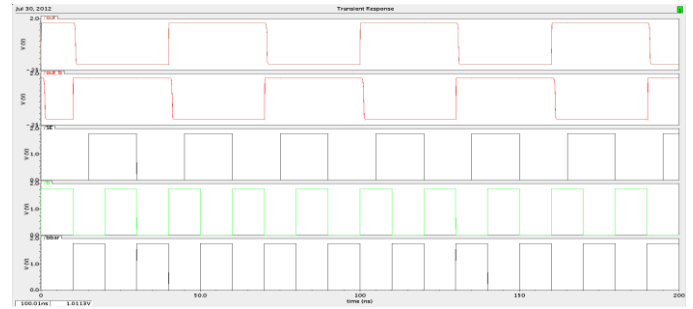

Figure 13 Sense Amplifier Layout

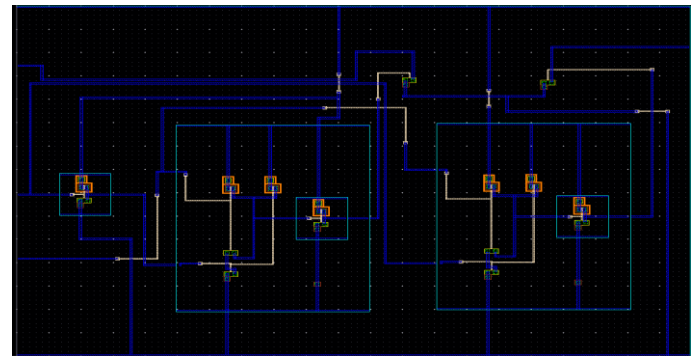Figure 14 Sense Amplifier Simulated Waveform

Figure 15 Write Driver Circuit Layout

Conventional SRAM memory design (1Kb)

The Figure 16 illustrates the layout sketch for SRAM memory design of 1-Kb.The area consumed by the memory design is around 3085μm.the power consumption during write operation is 884.7μw.
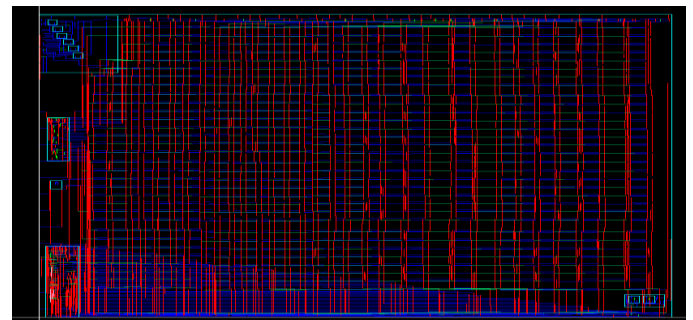

Figure 16 CV SRAM Memory design layout – 1Kb

Memory design using divided bit line approach (1Kb)

The Figure 17 illustrates the memory design of 1-Kb using divided bit line approach. The area consumed by the memory design using divided bit line approach is around 3784μm.The power consumption during write operation is 460.26μw.
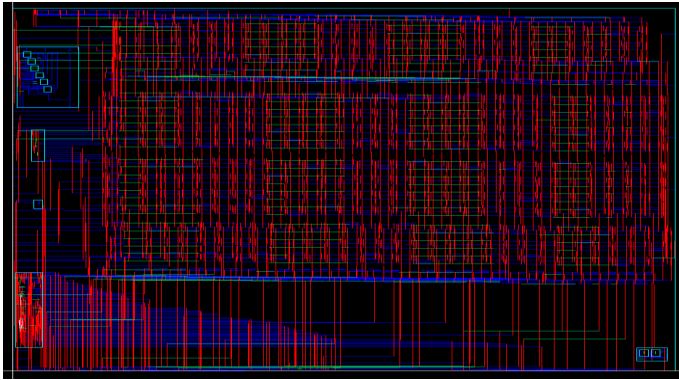
Figure 17 SRAM memory using divided bit line approach layout sketch– 1KB

Memory design using hierarchical divided bit line approach (1Kb)

The Figure 18 illustrates the memory design of 1Kb using HDBL approach. The Figure 19 illustrates memory design using Hierarchical divided bit line approach layout sketch – 1Kb.The area consumed by the memory design using hierarchical divided bit line approach is 4715μm.The power consumed during write operation is 354μw.
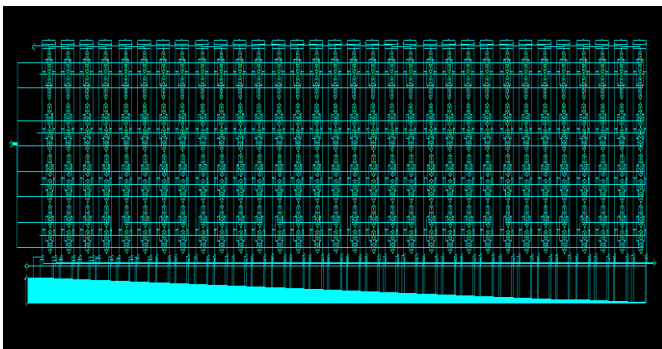

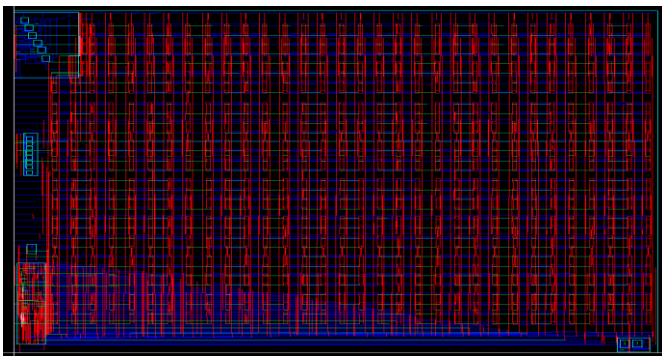Figure 18 SRAM memory using Hierarchical divided bit line approach – 1Kb

Figure 19 SRAM memory using Hierarchical divided bit line approach layout sketch – 1Kb

The Figure 20 and Figure 21 illustrates the write and read operation of memory – 1Kb using HDBL.
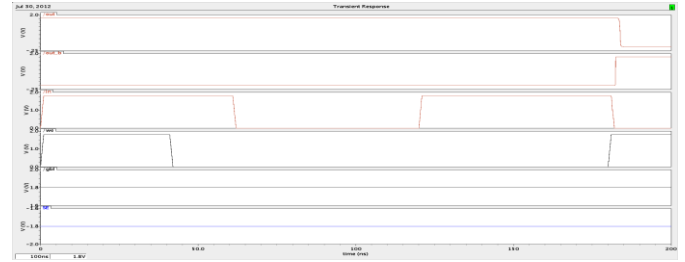

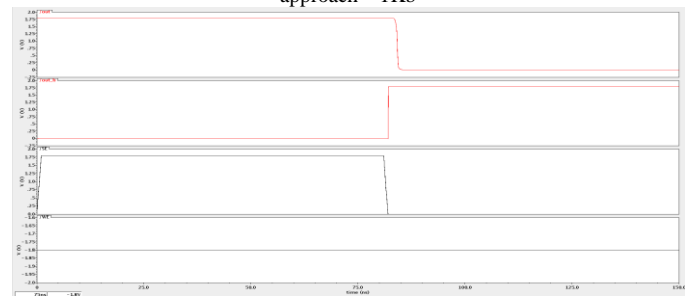Figure 20 SRAM memory write operation using Hierarchical divided bit line approach – 1Kb

Figure 21 SRAM memory read operation using Hierarchical divided bit line approach – 1Kb

VII CONCLUSION

This paper presents a 6T-based SRAM memory, and the various design techniques used to overcome power dissipation factor. A basic 6T SRAM structure is chosen for designing the SRAM bit cell which is used further for memory design, after this low power design techniques such as divided bit line and hierarchical divided bit line approach are implemented on the SRAM memory design. The conventional SRAM (CV-SRAM) and SRAM using Hierarchical divided Bit line Approach is implemented for the comparisons. The architecture of the CV-SRAM is almost the same as that of the other HDBL - SRAM except that the CV-SRAM uses the conventional bit line whereas the other SRAM uses the hierarchical bit line. Using this approach the power consumption is reduced at the expense of slight increase in the number of transistors when compared to Conventional SRAM. Thus the power consumed is decreased by 50-60% using HDBL and 30-40% using DBL successfully. The area consumed by the memory design of 1Kb is around 3085μm and for the memory design using divided bit line approach area consumed is around 3284 μm and for memory design using hierarchical divided bit line approach it is 3615μm. Thus the memory design using SRAM cell is implemented by various approaches for reducing power around 50-60% at the expense of 10-20% increase in area.

REFERENCES

[1] Martin Margala University of Alberta "Low Power SRAM circuit Design''.

[2] C.M.R. Prabhu, Ajay Kumar Singh "A proposed SRAM cell for low power consumption during write operation" Multimedia University, Melaka, Malaysia.

[3] R.H. Dennard ,"Evolution of the MOSFET Dynamic RAM - a Personal view ,"*IEEE Transactions on Electron Devices*", vol. ed-31, no. 11, Nov. 1984.

[4] David A. Hodges, Horace G. Jackson, A. Saleh "Analysis and Design of Digital Integrated Circuits in Deep Submicron Technology".

[5] Arash Azizi-Mazrea Mohammad T. Manzuri Shalmani, Hamid Barati, and Ali Barati "Delay and Energy Consumption Analysis of Conventional SRAM" World Academy of Science, Engineering and Technology vol .37 2008

[6] Byung-Do Yang and Lee-Sup Kim "A Low Power SRAM using Hierarchical Bit Line and Local Sense Amplifiers "*IEEE J. Solid - State Circuits*, vol. 40, no6, pp. 1366-1376.

[7] K.W. Mai, T. Mori, B. S. Amrutur, R. Ho, B.Wilburn, M. A. Horowitz,I. Fukushi, T. Izawa and S. Mitarai, "Low-power SRAM design using half-swing pulse-mode techniques," IEEE Solid-State Circuits, pp. 1659–1671, Nov. 1998.

[8] K. Kanda, H. Sadaaki, and T. Sakurai, "90% write power saving SRAM using sense amplifying memory cell," *IEEE J. Solid State Circuits*, Jun. 2004.

[9] B. S. Amrutur and M. A. Horowitz, "A replica technique for word line and sense control in low-power SRAM'S ," *IEEE J. Solid State Circuits*, vol. 33, pp. 1208–1219, Aug 1998 .

[10] H. Mizuno and T. Nagano, "Driving source-line cell architecture for sub-1-V high-speed low Power applications," *IEEE J. Solid State Circuits* , vol. 31, pp. 552 – 557, Apr. 1996.

[11] Amrutur, B. and Horowitz, M. (1994), "Techniques to reduce power in fast wide memories", *Proc. Symp. Low Power Electronics*, October, pp. 92-3.

[12] Bhardwaj, M., Min, R. and Chandrasekaran, A.P. (2001), "Quantifying and enhancing power awareness of VLSI systems", *IEEE Trans. VLSI systems*, Vol. 9 No. 6, pp. 757-72.

[13] Chang, Y., Lai, F. and Yang, C. (2004), "Zero-aware asymmetric SRAM cell for reducing cache power in writing zero", *IEEE Trans. VLSI Systems*, Vol. 12 No. 8,pp. 827-36.

[14] Geens, P. and Dehaene, W. (2005), "A small granular controlled leakage reduction system for SRAMs", Solid-State Electronics, Vol. 49, pp. 1176-782.

[15] J Rubinstein "Signal Delay in RC Tree Networks", *IEEE Transactions on Computer-Aided Design,* Vol. CAD-2, No. 3, pp. 202-211, July 1983.

[16] Sharifkhani, M. and Sachdev, M. (2007), "Segmented virtual ground architecture for low power embedded SRAM", *IEEE Trans. VLSI Systems*, Vol. 15 No. 2, pp. 196-205.

[17] Y. Somasekhar, (1996), "On the design of adiabatic SRAM's", TR-ECE 96-4, April. Yeo, K .S. and Roy, K - 2005.

[18] K. Itoh, K. Sasaki, and Y. Nakagome " Trends in Low Power RAM Circuit Technologies", Proceedings of the IEEE, Vol. 83, No. 4, April 1995.

[19] Sreerama Reddy G.M, P. Chandrasekhar Reddy "Design and VLSI Implementation of 8 Mb Low Powers SRAM in 90nm" Euro Journals Publishing, Inc. 2009 European Journal of Scientific Research Vol.26 No.2 (2009), pp.305-314.

AUTHOR'S PROFILE:

Mr. Mohammed Sayeeduddin Habeeb completed his Masters of Engineering specializing in Communication Engineering from Osmania Univesity, India. He completed his research thesis at DOFI, RCI, DRDO laboratory (Ministry of Defence) and designed Programmable Pre Modulation Filter in Avionic application. After completing his Master's studies he worked as Asst. Professor at Shadan College Of Engineering and Technology, India. Since 2012 he is working as a lecturer in King Khalid University, Abha. His research interests are Design and analysis of digital filters, digital signal processing and there optimization methods

MD SALAHUDDIN received the B.Eng. in Electronics and Communication Engineering from Osmania University, Hyderabad, AP, INDIA in 2010, and the M.Tech. in VLSI System Design from JNT University, Hyderabad, TS, INDIA in 2012. His research interests include Wireless Communication, Microelectronics, Analog VLSI, Image processing, Embedded Systems.

.