



Design and Implementation of Low Leakage Power SRAM System Using Full Stack Asymmetric SRAM

Rajlaxmi Belavadi¹, Pramod Kumar.T¹, Obaleppa. R. Dasar², Narmada. S², Rajani. H. P³

PG Student, Department of ECE, KLE Dr. MSSCET, Belgaum, Karnataka, India¹

PG Student, Department of ECE, VTU Belgaum, Karnataka, India²

Professor & Head of the Department TCE, KLE Dr. MSSCET, Belgaum, Karnataka, India³

Abstract: This paper explores the design and analysis of Static random Access Memory (SRAMs), focusing on optimizing delay and power. To address sub threshold leakage issue full stack approach is used. The full stack technique reduces leakage power to a great extent. The full stack technique is applied to SRAM cell in asymmetric manner in order to realise still higher power reduction. This work compares performance of SRAM using full stack approach with that of conventional 6T-SRAM design. The impact of temperature and different process corners on the performance of full stack design is also analysed. The static power and dynamic power measurement is done using Cadence Virtuoso ADE Visualization and Analysis XL Browser and XL Calculator. The layout is drawn and verified for DRC, LVS and RC extraction using Cadence Assura Tool.

Keywords: 10T SRAM, Full Stack, Dual Vth, Low Power.

I. INTRODUCTION

SRAM is a type of semiconductor memory which does not need to be periodically refreshed. With scaling down of technology, the feature sizes have shrunk more and more and miniaturization at chip level has occurred. But as a trade off, the demand for power has also increased. SRAM continues to be a critical component in microelectronics applications. Leakage is a serious problem particularly for SRAM. With shrinking technologies the threshold voltage is decreasing and leakage current of SRAM is increasing.

SRAMs contribute to a significant portion of the total system power dissipation. Caches, tag arrays, register files, branch table predictors, instruction windows, translation look aside buffers are common examples of microprocessor modules in which SRAMs are used. For the CMOS feature sizes below 180nm, leakage power dissipation has become an overriding concern for VLSI circuit designers. Power consumption of CMOS consists of dynamic and static components. Dynamic power is consumed when transistors are switching, and static power is consumed regardless of transistor switching [1].

Since cache memory cells remain in idle state most of the time static power is becoming an overriding concern in submicron technologies. This work focuses on static power reduction using stack principle. Fig.1 shows the structure of stack approach. When the two transistors are turned off together, induced reverse bias between the two transistors results in sub-threshold leakage current reduction. However, divided transistors increase delay significantly and could limit the usefulness of the approach [2]. In Fig.1 single transistor is divided into two transistors. The stack effect occurs when two or more stacked transistors are turned off together; the result is reduced leakage power consumption. The leakage through two series OFF transistor is much lower than that of a single transistor because of the stack effect.

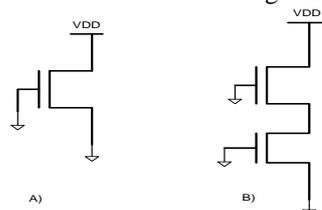


Fig.1 Stack approach



II. MEMORY ARCHITECTURE

Fig.2 shows a typical structure of Memory architecture [3]. It is primarily composed of the following sub-blocks: address decode logic, memory core, read column logic, write column logic, read control, and write control logic. While the generic structure of SRAMs is usually the same, SRAMs typically differ from each other in their size, organization of the memory core (in terms of number of rows and columns). SRAMs usually support read and write operations [6]. For these operations, the row decoder selects the appropriate wordline corresponding to the input address thereby activating a row in the memory array.

For a read operation, the precharged bitlines either retain charge or discharge depending on the data stored in the memory core cells selected by the wordline. The sense-amplifier in the read logic detects the changes in the voltage on the bitlines and the appropriate data is multiplexed to the data output. The read control logic controls the signals to the sense-amplifiers and bitline precharge logic [3].

For a write operation, the sense-amplifiers are isolated and the write buffers in write logic drive the bitlines in accordance with the data to be written into the memory location corresponding to the write address. After a read/write has been performed, the bitlines are precharged to supply voltage (referred to as precharge phase) thereby getting ready for another read/write in the next cycle. Typically, in a SRAM clock cycle, while read/write is performed in the first phase (referred to as read/write phase) of the clock cycle, precharge is performed in the second phase. Bitline precharge is done independent of the operation in the first phase of the clock cycle. If there is no operation being performed in a clock cycle, all the wordlines remain deactivated (logic LOW) and the bitlines stay precharged (logic HIGH). This no operation phase is referred as idle phase. The leakage current in SRAMs vary within a clock cycle depending on the phase of the operation being performed, since different transistors would be in off state during different operations[3].

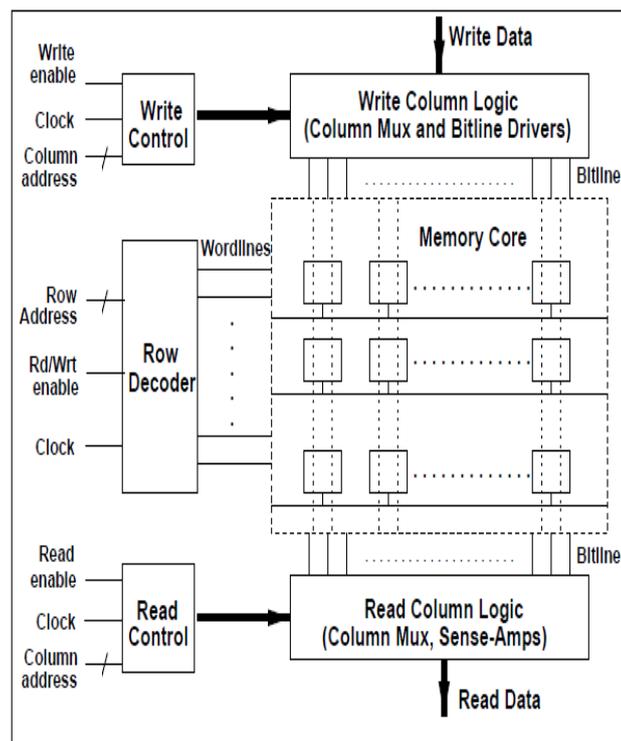


Fig.2 Memory system Architecture



III. 6T STANDARD SRAM CELL

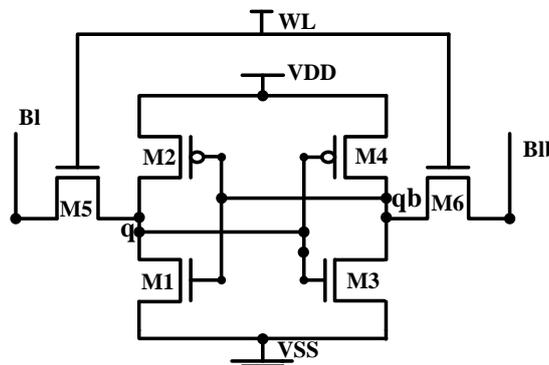


Fig.3 6T SRAM cell

Data 0 or 1 is stored in SRAM cell, using two cross coupled inverters. This storage cell has two stable states **0** and **1** which is reinforced because of cross coupling. Two additional *access* transistors serve to control the access to the storage cell during read and write operations. So a typical SRAM cell is a six transistor structure. A 6T SRAM cell requires a careful device sizing to ensure read stability, write margin and data retention in standby modes. The Fig.3 shows a typical 6T SRAM cell. Access to the cell is enabled by the word line which controls the two access transistors M5 and M6. These device inturn controls whether the cell should be connected to the bit lines. Bit lines are used for both read and write operations. For read stability, M1 transistor is required to be much larger than M5 transistor to make sure that the node between M1 and M5 does not flip. During write operation, bit lines overpower cell with a new value. High bit lines must not overpower inverters during read operation. So, M2 is designed to be weaker than M5 [4].

IV. MEMORY SYSTEM USING FULL STACK 10T SRAM CELLS

A. FULL STACK 10T SRAM CELL:

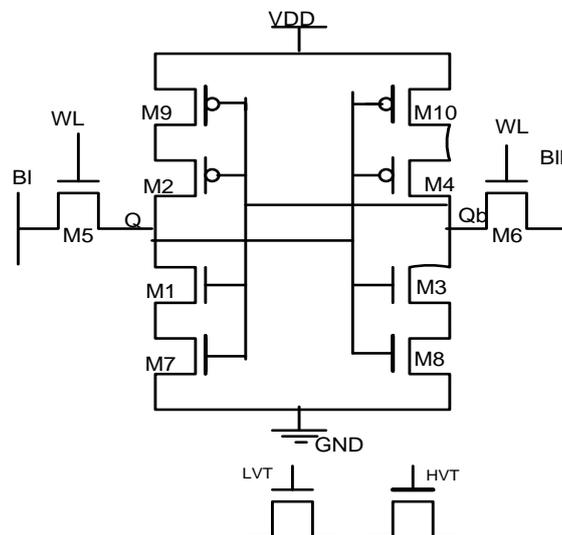


Fig.4 Full stack SRAM cell

The transistors M1 to M6 constitute the standard 6T SRAM cell. Stacking is applied to this cell both in the pull down and pull up section. Transistors M7 and M8 are connected as Nmos stacking transistors to pull down transistors M1 and M3. Pmos stacking transistors M9 and M10 are connected to pull up load transistors M2 and M4. During cell inactive state, the node at '0' value is connected to ground through two series connected off transistors. This reduces leakage current due to stack effect. Similarly the leakage current flowing through off transistors in the pull up path also gets reduced due to stacking. The node storing '1' gets charged by larger resistance which reduces active current and active power consumption. The power can be further lowered by using high Vt transistors.



In this cell both high-Vt and normal-Vt transistors are used. So that optimized delay and power can be observed. Transistor sizing for SRAM can be approached in two ways. For the function of SRAM cell, read & write stability needs to be guaranteed. In read stability, M1 transistor is required to be much larger than M5 transistor to make sure that node between M1 and M5 transistors must not flip. When in write mode, bit lines (B1 or B1b) overpower cell with new value. However, high bit lines must not overpower inverters during read operation. That results in the determination of sizing M2 transistor weaker than M5 transistor. The stacking transistors are taken as minimum size transistors [5].

B. PERIPHERAL CIRCUITS

- a) **Address decoding:** The 8x8 SRAM system has storage capacity of 8words of 8bit each. To address these words, 3:8 row decoder is used. And for 16x16 SRAM system, 4:16 row decoders are used. Lyon-Schediwy decoder [6] which uses lesser number of transistors than the regular AND/NAND decoder is used for address generation. Since Lyon-Schediwy decoder is a faster decoder, and uses lesser number of transistors compare to regular decoder, the power dissipation is greatly reduced. The 3:8 row decoder is shown in Fig.5 and Fig.6 depicts 4:16 decoder.

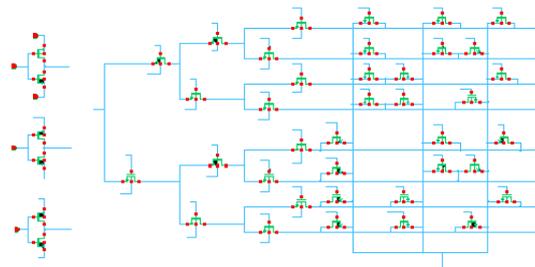


Fig.5 3:8 row decoder

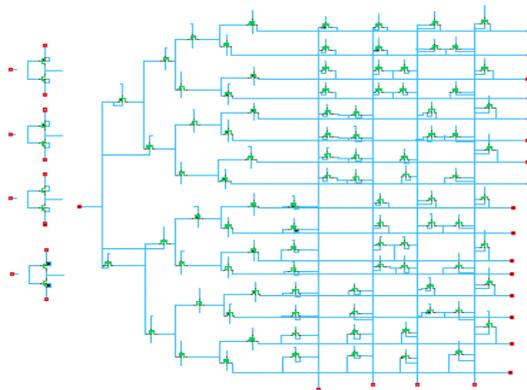


Fig.6 4:16 row decoder

- b) **Precharging:** The function of the pre charge circuit is to charge the bit lines and inverse bit-lines. The pre-charge enables the bit-lines to be charged high at all times except during write and read cycle. Whenever the precharge signal PRE_C is applied to the equalizing transistors, it simply equalizes the voltage levels on B and B_bar [7].

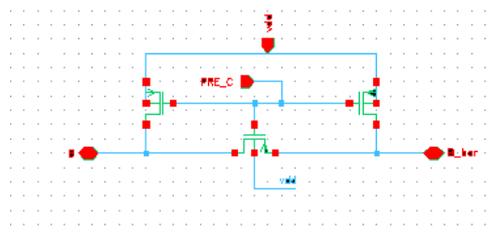


Fig.7 Pre charge circuit

- c) **Data Write Operation:** The write circuit writes the data from the port and on getting the write enable signal, performs the write operation. Each of the 8 columns of the memory array has one data write circuit as shown



in fig. 4. The data write circuit consists of two inverters and an AND gate implemented by pass-transistors. The data write circuitry writes data and its complement onto the bit-lines when activated by write-enable (WE) signal. The data and its complement are written onto the individual nodes Q and QN of the selected word through the access transistors of the SRAM cell.

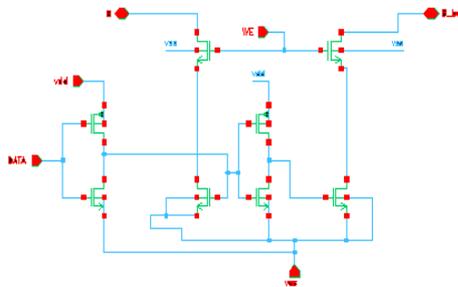


Fig.8 write driver

- d) **Data Read Operation:** The data written into the SRAM cell is retained as long as the power is present. During read operation the sense amplifier enable (SAE) signal is applied to the sense amplifier as shown in fig. 9. This activates the sense amplifier for read operation only for Short sense duration. At the same time the column is isolated from the bit-lines by using signal Ymux. This causes one of the bit-lines Bl/Blb to discharge from the precharged value. This creates a differential voltage on the bit-lines which is sensed by the sense amplifier and amplified to the full extent. The data and its complement are reflected on the output lines READ and READ_BAR[7].

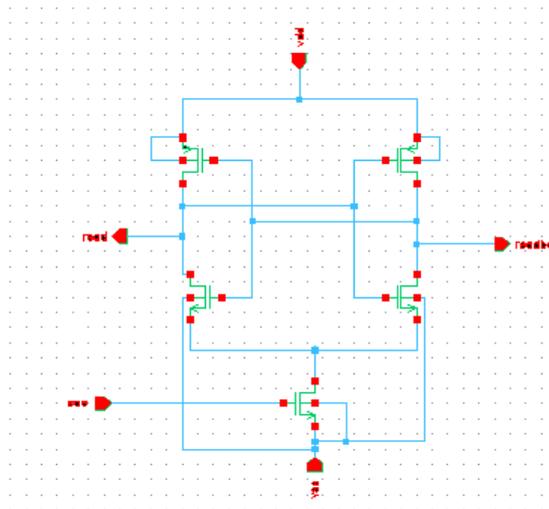


Fig.9 sense amplifier

C. LOW POWER MEMORY SYSTEM FOR ARRAYS

Combining 64-10T SRAM cells, address decoder the precharge circuitry, data write circuitry, and sense amplifiers, 8x8 SRAM array is designed in 90nm CMOS technology files. Similarly 16x16 SRAM array also designed using 256-10T SRAM cells and appropriate peripheral circuitry. The complete setup is data write and read for memory array system is shown in Fig.10. This block diagram shows all the different peripheral circuits combined with the static RAM cells, to form a complete working SRAM 8 x 8 array for write and read operation. Fig.10 shows the block diagram of SRAM memory with all input signals; precharge, write enable, sense amplifier enable, 16 word lines and 16 input data bits. SRAM 16x16 array is shown in Fig.11.

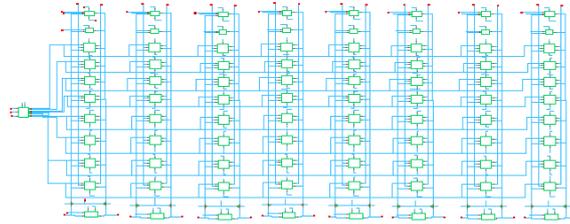


Fig.10 10TSRAM8x8 array

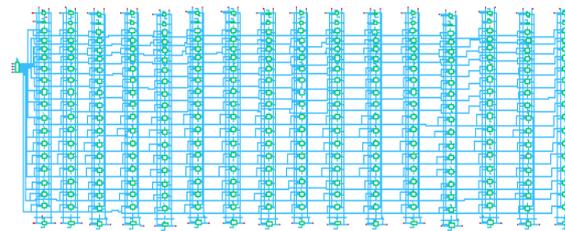


Fig.11 10TSRAM16x16 array

V. SIMULATION AND RESULTS

All simulations were done in Cadence design environment, using GPDK 90nm technology files. Power dissipation is calculated across the different process corners. As the temperature increases power dissipation also increases and different temperature values are taken with a power supply voltage of 1.1V. Static power dissipation is measured during write ‘1’ and write ‘0’ operations as well as during cell inactive state. When compared to the total power dissipation of single 6T standard SRAM cell, it is observed that the single Full stack SRAM cell total power reduced by 9.5X times. A comparison of the Access Time, Static and Total Power Dissipation in the Full stack SRAM cell with 6T standard SRAM cell is presented in Table. 3. For 8 x 8 SRAM Array System, the Total Power Dissipation, when compared to 8 x 8 6T standard SRAM array system, it is observed that the total power dissipation of Full stack SRAM 8 x 8 array system has reduced considerably by 6X.

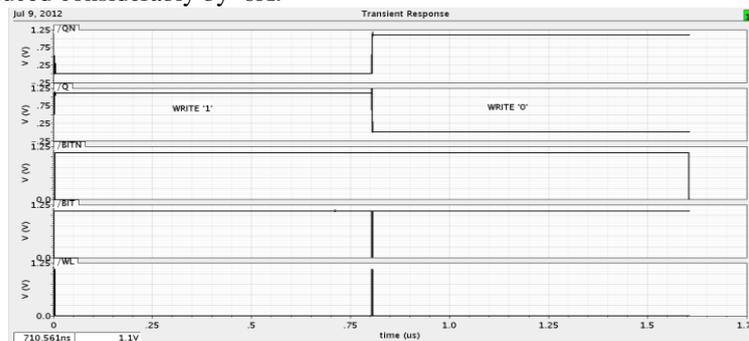
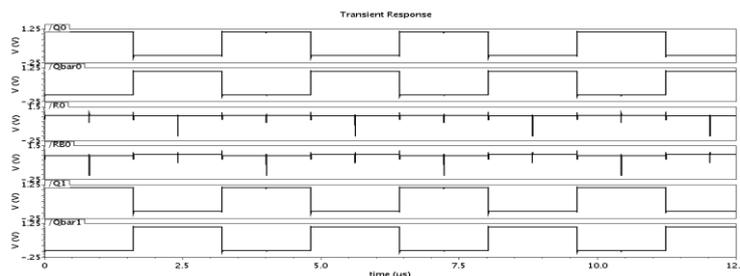


Fig. 12: Simulated output of single Full stack SRAM cell



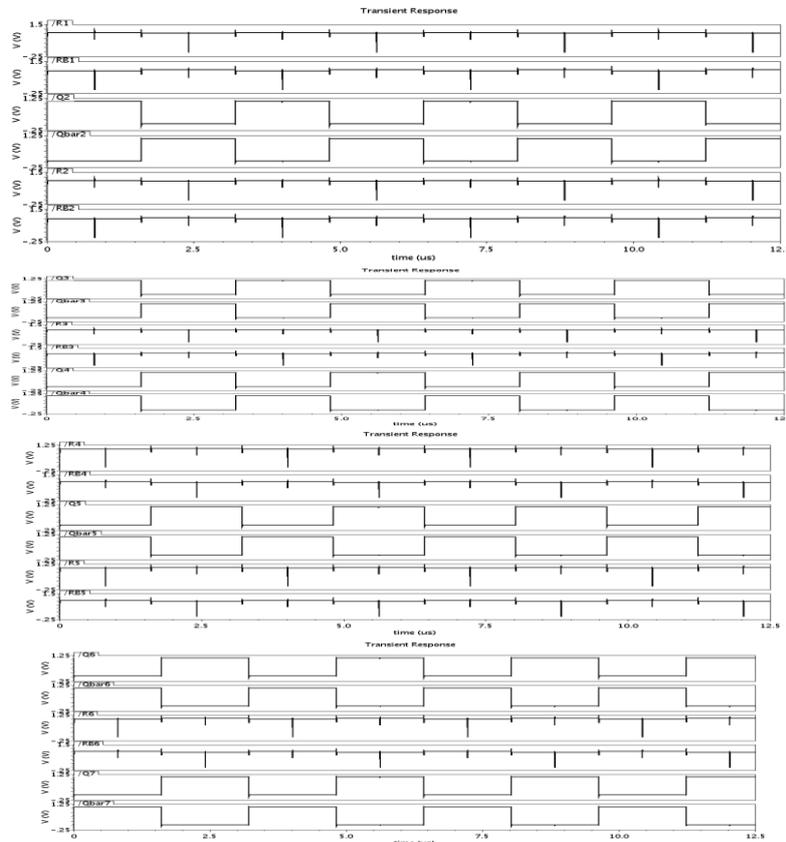


Fig.13: Memory data written/read for data 11110000

Table 1: Total Power dissipation for different process corners for single Full stack SRAM cell

Process Corners	Total Power Dissipation(nW)
Slow-Slow	5.22
Slow-Fast	8.06
Fast-Slow	7.805
Fast-Fast	13.05

Table 2: Static, Total Power dissipation, for Single Full stack SRAM cell at diff temperatures (Vdd=1.1V)

Temperature(degree Celsius)	Total Power Dissipation (nW)	Static Power (nW)
-50	7.541	1.518
-35	7.781	1.572
-25	7.896	1.609
0	8.154	1.701
27	8.406	1.809
35	8.478	1.843
50	8.608	1.912
75	8.829	2.047
100	9.118	2.270



Table 3: Static, Total Power dissipation, and Access Time for Single SRAM cells (V_{dd}=1.1V)

Single SRAM cell	Static Power (nW)		Total Power Dissipation (nW)	Access Time (pS)
	Hold state	Write '1' state		
Conventional 6T SRAM cell	43.5	43.5	76.8	42.34
Full stack SRAM cell with Hvt	1.80	1.80	8.406	73.70
Full stack SRAM cell without Hvt	3.130	3.130	22.83	61.04

Table 4: Access Time and Total Power dissipation for 8 x8 SRAM array systems (V_{dd}=1.1V)

8x8 SRAM array system	Total power dissipation(mW)
6T standard SRAM 8 x 8 array	6.163
Pre Layout Full stack SRAM 8 x 8 Array	1.323
Post Layout Full stack SRAM 8 x 8 array	1.226

The layout was drawn and verified for DRC, LVS and RC extraction using Cadence Assura Tool. Having successfully made layout of single bit Full stack SRAM, layout for 8 x 8 memory system of size 8words x 8bits including all the memory cells and peripherals circuits is made using *Cadence Assura tool*. Pre layout and Post layout Access Time, Read Time and Total Power dissipation are measured and tabulated in Table 4.

VI. LAYOUT

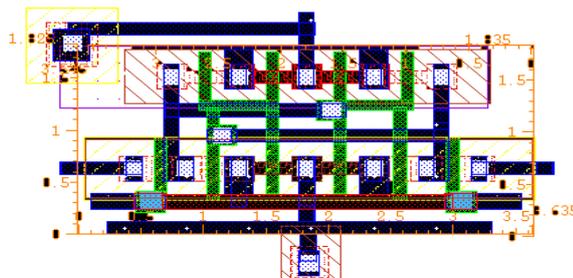


Fig 13: Layout of single cell Full stack SRAM with Hvt

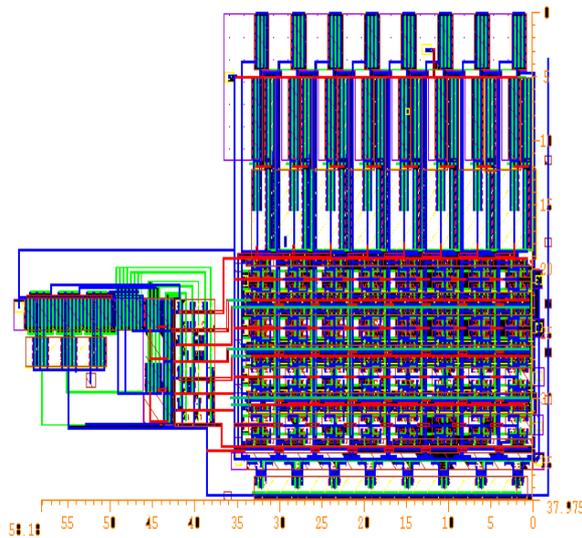


Fig 14: Layout of 64 bit Full stack SRAM with Hvt

VII. CONCLUSION

In this work leakage power is reduced, at the expense of additional stacking transistors. Using full stack approach SRAM performance improved without losing read and write ability. The design and implementation of the SRAM memory done in cadence environment in 90nm process technology. The full stacking has reduced power considerably.

REFERENCES

- [1] Weste and Harris “*CMOS VLSI Design- A circuits and systems perspective*”
- [2] M. Powell, S.-H. Yang, B. Falsafi, K. Roy and T. N. Vijaykumar, “Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep submicron Cache Memories,” International Symposium on Low Power Electronics and Design, pp. 90-95, July 2000.
- [3] Abhishek Mathur, Arun Jayachandran, Ramya Venumbaka “Low Leakage SRAM design using sleep transistor stack”
- [4] Zhengya Zhang, Zheng Guo “*Active Leakage Control with Sleep Transistors and Body Bias*”
- [5] JOON-SUNG YANG / GAHNGSOO MOON DEC. 9, 2005 “32K-BIT SLEEPY SRAM”. Benton H. Calhoun, Frank A. Honore, Anantha P. Chandrakasan- “*A Leakage Reduction Methodology for Distributed MTCMOS*”
- [6] N. Weste and K. Eshragian. *Principles of CMOS VLSI Design, A Systems Perspective*. Addison-Wesley Publishing Company, Reading, CA, 1998.
- [7] Andrei Pavlov, Manoj Sachdev CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies. Process-Aware SRAM Design and Test.