

DESIGN OF LOW POWER 8 BIT SRAM ARCHITECTURE USING LEAKAGE FEED BACK WITH STACK & SLEEP STACK WITH KEEPER

V.G. Santhi Swaroop^{**}, B.Murali Krishna^{*}, M.Vijaya Bhaskar^{**},
B.Raghu kanth^{**}, V.SAI PRAVEEN^{**}

^{*} Assistant professor, ECE Department, K.L.University, Vijayawada, India.

^{**} M.Tech -VLSI Students, VLSI, ECE Department, K.L.University, Vijayawada, India.

Abstract— The major concern in Very Large Scale Integration (VLSI) circuit design is power consumption and the challenging aspect for VLSI designers is to reduce the power consumption. International technology roadmap for semiconductors (ITRS) reports that “leakage power dissipation” may come to dominate total power consumption. One of the main reasons causing the leakage power increase is that increase of sub-threshold leakage power. Some leakage current reduction techniques like sleep approach, stack, zigzag & some new techniques like ,sleepy-stack, leakage feedback approach and sleepy keeper techniques and , after that to combine the advantages of above written techniques, proposed two novel approaches , named “LEAKAGE FEEDBACK WITH STACK (LFS)” & “SLEEP STACK WITH KEEPER (SSK)” which reduces leakage current while saving exact logic state.

In this paper we design 8 x 8 S-RAM by using the “Leakage Feedback with Stack” & “Sleep Stack with Keeper” leakage current reduction techniques. The proposed circuits were designed in 0.18 μ m CMOS/VLSI technology within Micro-Wind tool, and measure power consumption for two design approaches, “Sleep-Stack with keeper approach” achieves up to nearly 50% less power consumption than existing counterparts.

Key-words: low power design, leakage reduction, sleep, stack, sleepy-stack keeper, Leakage Feedback with Stack.

I. INTRODUCTION

Static RAMs are used extensively in modern processors as on chip memories due to their large storage density and small access latency. Low power on-chip memories have become the topic of substantial research as they can account for almost half of total CPU dissipation, even for extremely power-efficient designs. Power dissipation which was

previously considered an issue only in portable devices is rapidly becoming a significant design constraint in many system designs. Dynamic power has been a predominant source of power dissipation till recently. However, static power dissipation is becoming an significant fraction of the total power. Static power is the power dissipated in a design in the absence of any switching activity and is defined as the product of supply voltage and leakage current. The absolute and the relative contribution of leakage power to the total system power is expected to further increase in future technologies because of the exponential increase in leakage currents with technology scaling. The International Technology Roadmap for Semiconductors (ITRS) predicts that leakage power would contribute to 50% of the total power in the next generation processors. Therefore, it is important for system designers to get an early estimate of leakage power to meet the challenging power constraints.

There are several VLSI techniques to reduce leakage power. Each technique provides an efficient way to reduce leakage power, but disadvantages of each technique limit the application of each technique. We implement a new approach, thus providing a new choice to low-leakage power VLSI designers. Previous techniques are summarized and compared with our new approach presented in this paper.

II. PREVIOUS WORK

In addition to sub-threshold leakage, another contributor to leakage power is gate-oxide leakage power due to the tunneling current through the gate-oxide insulator. Since gate-oxide thickness will be reduced as the technology decreases, in nano-scale technology, gate-oxide leakage power may be comparable to sub-threshold leakage power if not handled properly. The most well-known traditional approach is the sleep approach [2][3]. In the sleep approach, both (i) an additional “sleep” PMOS transistor is placed between VDD and the pull-up

network of a circuit and (ii) an additional "sleep" NMOS transistor is placed between the pull-down network and GND. These sleep transistors turn off the circuit by cutting off the power rails. Fig 1 shows its structure. The sleep transistors are turned on when the circuit is active and turned off when the circuit is idle. By cutting off the power source, this technique can reduce leakage power effectively.

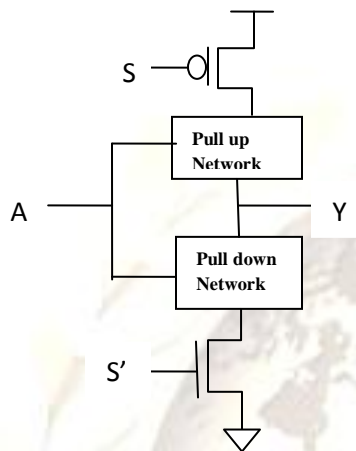


Fig.1 Sleepy approach.

The sleepy stack technique divides existing transistors into two half size transistors like the stack approach [6][7]. Then sleep transistors are added in parallel to one of the divided transistors. Fig 3 shows its structure. During sleep mode, sleep transistors are turned off and stacked transistors suppress leakage current while saving state. Each sleep transistor, placed in parallel to the one of the stacked transistors, reduces resistance of the path, so delay is decreased during active mode. $W/L=4.5$

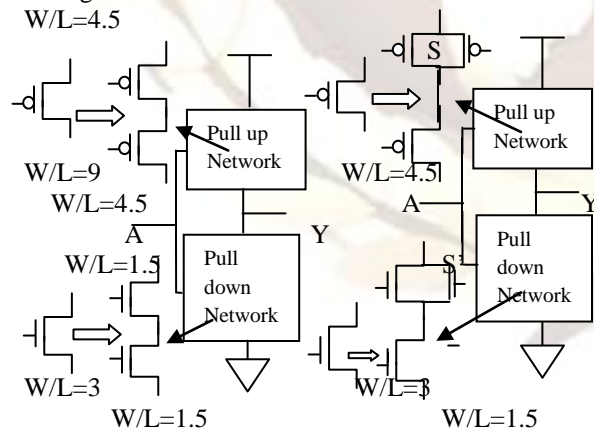


Fig.2 Stack approach Fig.3 Sleepy Stack approach

However, area penalty is a significant matter for this approach since every transistor is replaced by

three transistors and since additional wires are added for S and S', which are sleep signals.

Another technique for leakage power reduction is the stack approach, which forces a stack effect by breaking down an existing transistor into two half size transistors [5]. Fig 2 shows its structure. When the two transistors are turned off together, induced reverse bias between the two transistors results in sub-threshold leakage current reduction.

The leakage feedback approach is based on the sleep approach. However, the leakage feedback approach uses two additional transistors to maintain logic state during sleep mode, and the two transistors are driven by the output of an inverter which is driven by output of the circuit. As shown in Fig 4, a PMOS transistor is placed in parallel to the sleep transistor (S) and a NMOS transistor is placed in parallel to the sleep transistor (S'). The two transistors are driven by the output of the inverter which is driven by the output of the circuit.

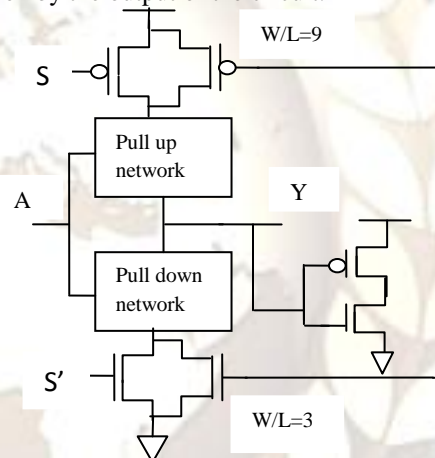


Fig.4 Leakage feedback approach

During sleep mode, sleep transistors are turned off and one of the transistors in parallel to the sleep transistors keep the connection with the appropriate power rail.

The basic problem with traditional CMOS is that the transistors are used only in their most efficient, and naturally inverting, way: namely, PMOS transistors connect to VDD and NMOS transistors connect to GND. It is well known that PMOS transistors are not efficient at passing GND; similarly, it is well known that NMOS transistors are not efficient at passing VDD. However, to maintain a value of '1' in sleep mode, given that the '1' value has already been calculated, the sleepy keeper approach uses this output value of '1' and an NMOS transistor connected to VDD to maintain output value equal to '1' when in sleep mode.

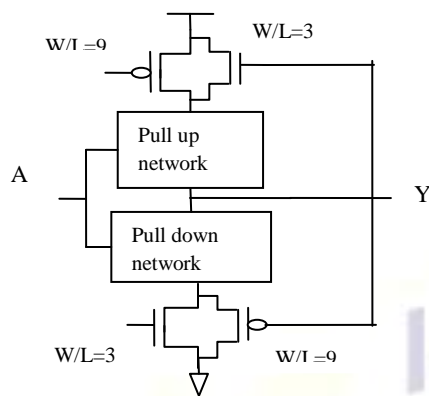


Fig.5 Sleep Keeper approach

As shown in Fig 5, an additional single NMOS transistor placed in parallel to the pull-up sleep transistor connects VDD to the pull-up network. When in sleep mode, this NMOS transistor is the only source of VDD to the pull-up network since the sleep transistor is off. Similarly, to maintain a value of '0' in sleep mode, given that the '0' value has already been calculated, the sleepy keeper approach uses this output value of '0' and a PMOS transistor connected to GND to maintain output value equal to '0' when in sleep mode. As shown in fig. 5, an additional single PMOS transistor placed in parallel to the pull-down sleep transistor is the only source of GND to the pull-down network which is the dual case of the output '1' case explained above. For the sleep, sleepy stack and leakage feedback approaches, sleepy keeper approach, dual Vth technology can be applied to obtain greater leakage power reduction.

III. MOTIVATION

The art of power analysis and optimization of integrated circuits used to be a specialty in analog circuit design. Power dissipation of VLSI chips is traditionally a neglected subject. In the past the device density and operating frequency were low enough that it was a constraining factor in the chips. As the technology varies, more transistors, faster and smaller than their predecessors, which leads to the growth in operating frequency and processing per capacity leads to increase in power consumption. There are two types of power dissipation in CMOS Circuits: Dynamic and Static. Dynamic power is caused by switching activities of the circuit and most significant source of dynamic power dissipation in CMOS circuits is the charging and discharging of the capacitance. Static Power dissipation is related to the logical states of the circuits rather than switching activities. In CMOS logic, leakage current is the only source of static power dissipation.

Currently, sub-threshold leakage seems to be the dominant contributor to overall leakage power [8]. Another possible contributor to leakage power is gate-oxide leakage. A possible solution widely reported is the potential use of high k (high dielectric constant) gate insulators [9]. In any case, this papers targets reduction of the sub-threshold leakage component of static power consumption; other approaches should be considered for reduction of gate oxide leakage. Do please note, however, that all results reported in this paper include all sources of leakage power. With application of dual threshold voltage (V_{th}) techniques, the sleep, zigzag and sleepy stack approaches result in orders of magnitude subthreshold leakage power reduction [7] but in this paper, we are not using dual V_{th} approach. The major advantage of the sleepy stack approach over the sleep and zigzag approaches is that the sleepy stack approach saves exact logic state. However, the sleepy stack approach carries a non-trivial penalty: each transistor in the original, base case, traditional CMOS design results in three transistors in the sleepy stack equivalent. The goal of our new approach is to achieve the benefit of all above written techniques, and now we propose two novel approaches, named "leakage feedback with stack" & "sleep stack with keeper" which reduces leakage current while saving exact logic state.

IV. LEAKAGE FEED BACK WITH STACK & SLEEP STACK WITH KEEPER

This section describes new leakage reduction techniques in which we call the first one "leakage feedback with stack (LFS)" approach and other is "sleep-stack with keeper". This section explains the structure of the leakage feedback with stack approach and sleep-stack with keeper.

In this technique i.e. leakage feedback with stack, we are combining taking advantage of two techniques i.e. leakage feedback approach due to less transistor than sleepy-stack in which we replaces each transistor in base case into three transistors, and ultra low power technique i.e. Stack approach, here we are combining these two techniques. This is shown in Fig. 6. In another approach i.e. sleep-stack with keeper, we are combining the three different low power leakage reduction techniques i.e. sleep transistors, stack approach with keeper as shown in Fig. 7.

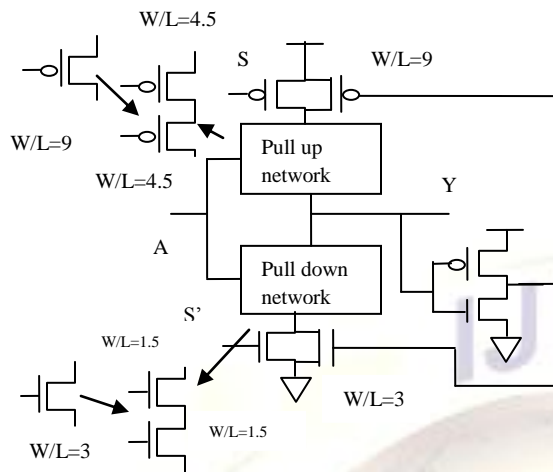


Fig.6 Leakage feedback with Stack

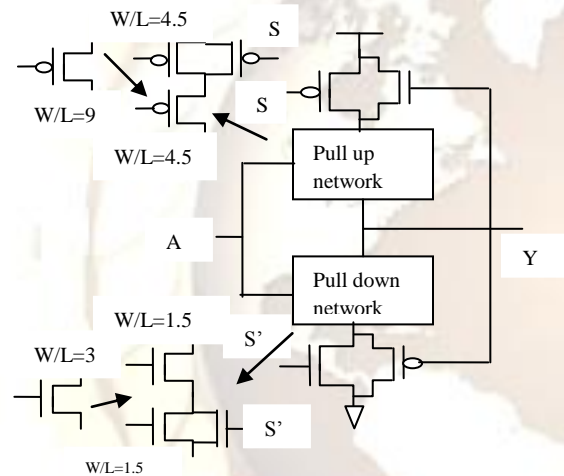


Fig.7 Sleep Stack with Keeper

V. SRAM

Static Random Access Memory (SRAM) to be one of the most fundamental and vitally important memory technologies today. Because they are fast, robust, and easily manufactured in standard logic processes, they are nearly universally found on the same die with microcontrollers and microprocessors. Due to their higher speed SRAM based Cache memories and System-on-chips are commonly used. Due to device scaling there are several design challenges for nanometer SRAM design. Low power SRAM design is crucial since it takes a large fraction of total power and die area in high performance processors. A SRAM cell must meet the requirements for the operation in submicron/nano ranges. The scaling of CMOS technology has significant impacts

on SRAM cell random fluctuation of electrical characteristics and substantial leakage current.

A. 6T SRAM Cell

The schematic of SRAM cell is shown in the Fig.8. It has 2 pull up PMOS and 2 NMOS pull down transistors as two cross coupled inverters and two 2 NMOS access transistors to access the SRAM cell during Read and Write operations. Both the bit lines (BL and BLB) are used to transfer the data during the read and write operations in a differential manner. To have better noise margin, the data signal and its inverse is provided to BL and BLb respectively. The data is stored as two stable states, at storing points VR and VL, and denoted as 0 and 1.

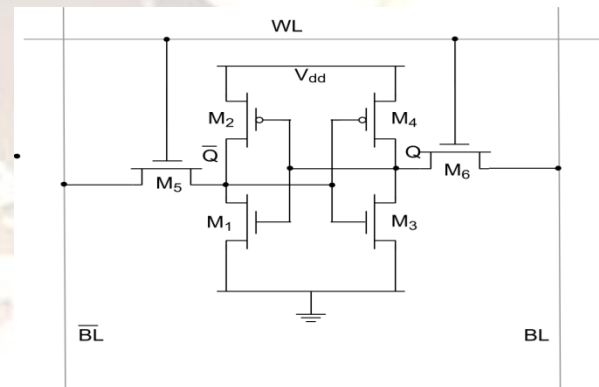


Fig.8 6T SRAM Cell

The heart of the SRAM cell is a dual inverter latch. Each end of the latch holds a value that's the complement of the other side. There are different BL and BLb conditions that represent the process of reading and writing into the cell.

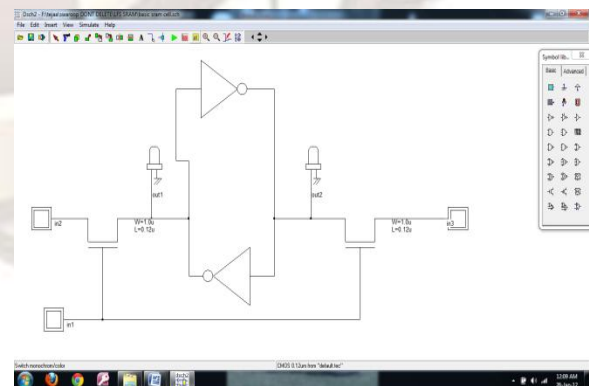


Fig.9 Schematic for 6T SRAM Cell

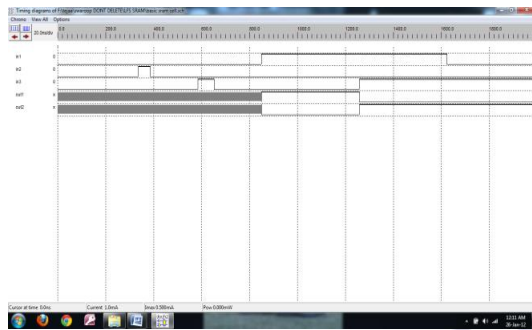


Fig.10 Simulations for 6T SRAM Cell

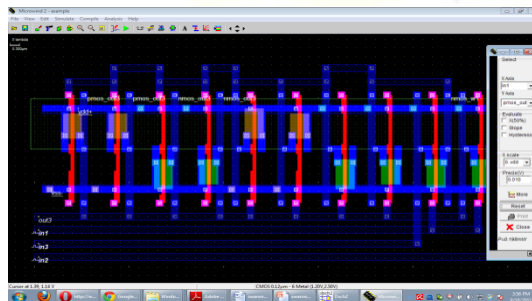


Fig.11 Layout for 6T SRAM Cell

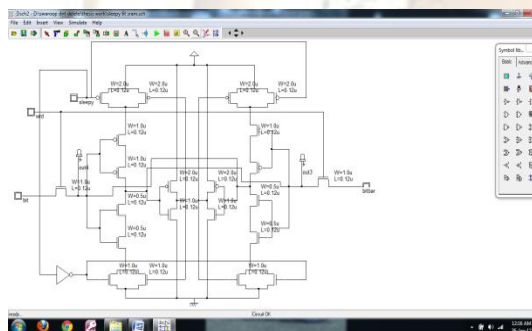


Fig.12 Schematic for 6T SRAM Cell using Leakage feedback with Stack

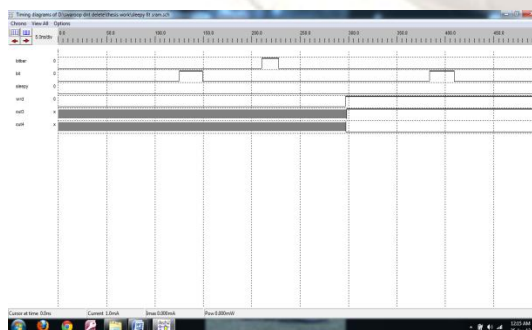


Fig.13 Simulations for 6T SRAM Cell using Leakage feedback with Stack

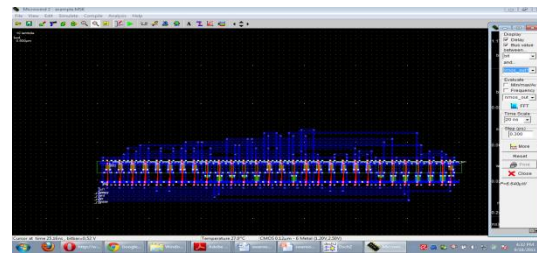


Fig.14 Layout for 6T SRAM Cell using Leakage feedback with Stack

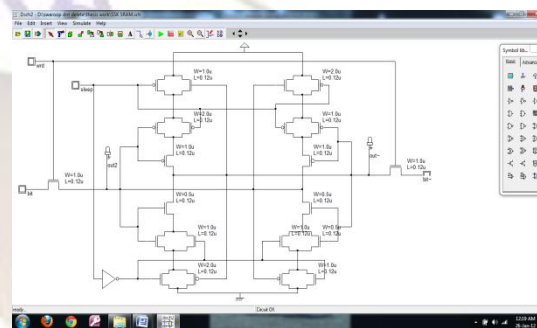


Fig.15 Schematic for 6T SRAM Cell using SSK

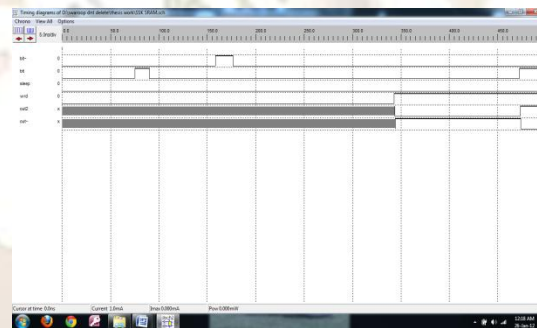


Fig.16 Simulations for 6T SRAM Cell using SSK

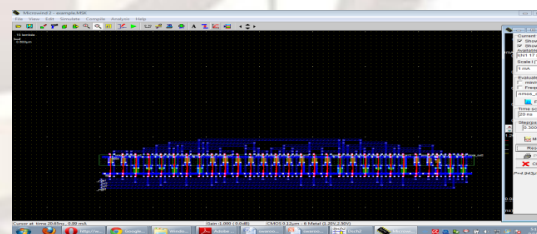


Fig.17 Layout for 6T SRAM Cell using SSK

VI. SRAM ARCHITECTURE

A general SRAM Architecture and its peripherals are shown in Fig.18. The SRAM array consists of compact rows and columns of bit cells. For small caches, it is possible to place a word of data in a row, however, in large memories because of

space limitation, it is necessary to arrange several words of data in each row. Cells of each column share the same bitlines. Before the read access, the bitlines are precharged to a known value by the precharge circuits. The row decoders are used to select a row in the array. Depending on the mode of operation, storage cells in the row are connected the common bitlines and either the stored data in the cell is read by sense amplifiers or overwritten by the write circuits. For larger memories, multiple blocks of the same array are used such that an extra address generator called block address decoder is required.

A. Sense amplifier

A sense amplifier circuit is used to read the data from the cell. In addition, it helps reduce the delay times and minimizes power consumption in the overall SRAM chip by sensing a small difference in voltage on the bit lines.

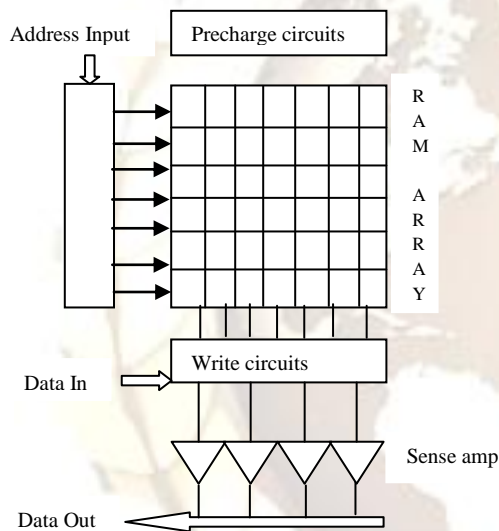


Fig 18. General SRAM Architecture

It has been optimized to a form that uses up the least number of transistors while still being reliable at high speeds. It is also much simpler than the small-signal (differential sensing) circuit. For having high performance SRAMs, it is essential to take care of the read speed both in the cell-level design and in the design of a clever sense amplifier. Sense amplifiers are one of the most critical circuits in the periphery of CMOS memories. Their performance strongly influences both memory access time and overall memory power consumption. High density memories commonly come with increased bitline parasitic capacitances. These large capacitances slow down voltage sensing and makes bitline voltage swings energy-consuming, which result in slower more power hungry memories. Need for larger memory

capacity, higher speed, and lower power dissipation impose trade offs in the design of sense amplifier.

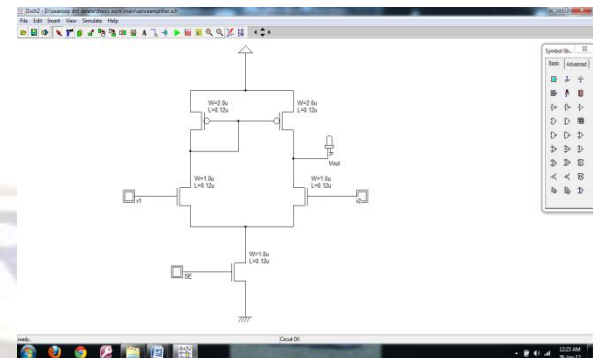


Fig 19. Schematic for sense amplifier

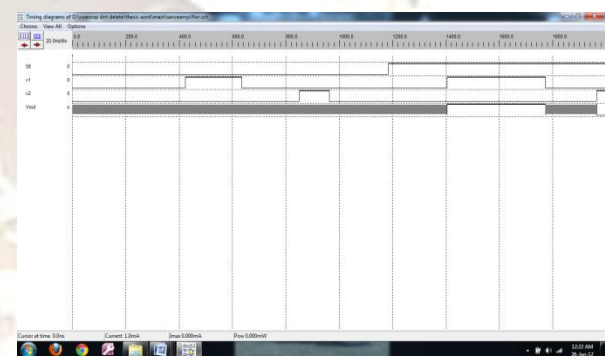


Fig 20. Simulations for Sense amplifier

B. Row Decoder

A row decoder is used to decode the given input address and select the wordline. When performing a write or read operation only one of the row is selected and 8 bits of data is transmitted. There are 8 rows and row contained 8 cells each. The row decoder selects one of those rows, depending on the 3 bit address given to it. In order to design an 8x8 SRAM a 3x8 decoder is used. Number of wordline equals to the number of rows in the SRAM cell array.

The decoder selects 1 of 8 wordlines, with respect to the input address. The output of the decoder is fed to a 2-input AND. This AND is the wordline driver. This AND supports a large capacitance on the wordline. Each cell loads the wordline with two transistors. Therefore, in the design there would be 16 transistors per wordline forming a large capacitance on the wordline. Other input to this AND is the Clock. Only when both Clock and decoder output signals are enabled, the AND enables a wordline to the rows of SRAM cell arrays. In a typical SRAM design, the output from the decoder would directly enable the wordline. This

AND was introduced in the design to achieve a clock enabled design. The schematic for 3x8 decoder is shown in Fig 21.

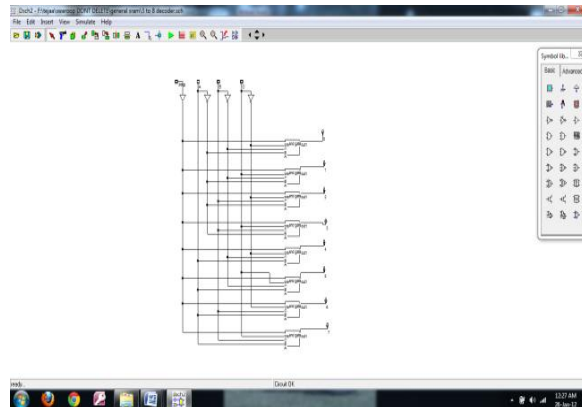


Fig 21. Schematic for 3x8 Decoder

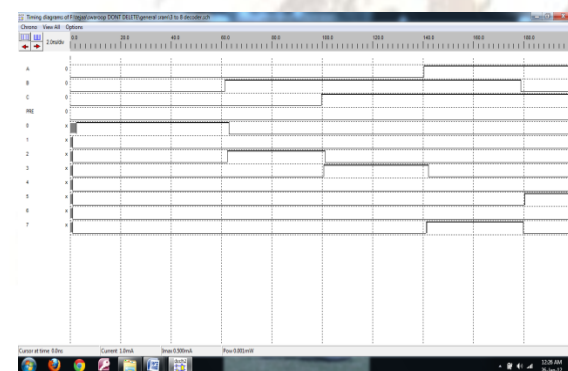


Fig 22. Simulations for 3x8 Decoder

C. 8bit SRAM Architecture

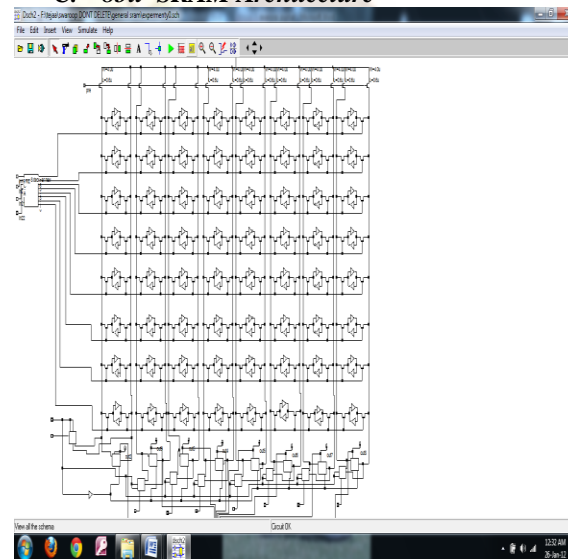


Fig 23. Schematic for 8x8 SRAM Architecture

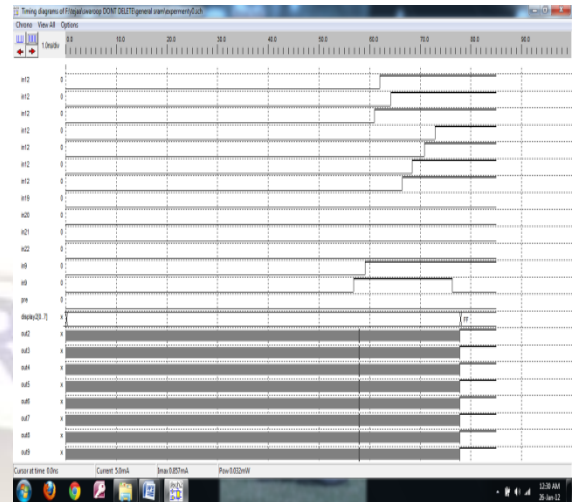


Fig 24. Simulations for 8x8 SRAM Architecture

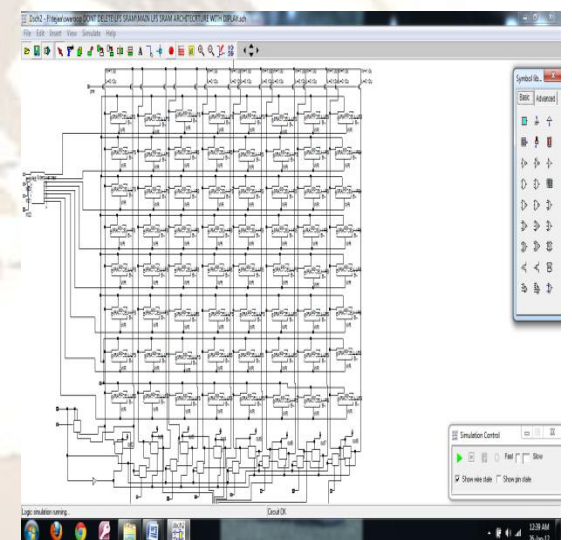


Fig 25. Schematic for SRAM Architecture using Leakage feedback with Stack

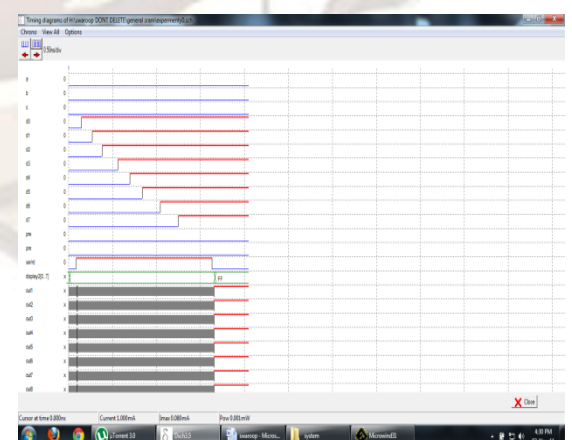


Fig 26. Simulations for Architecture using Leakage Feedback with Stack

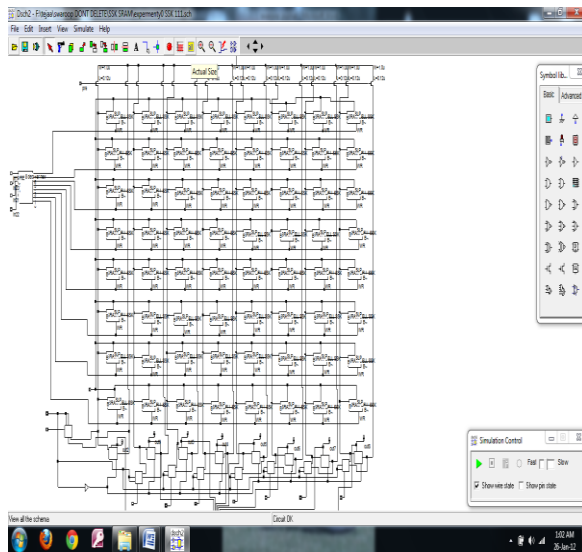


Fig 27. Schematic for 8x8 SRAM Architecture using Sleep Stack with keeper

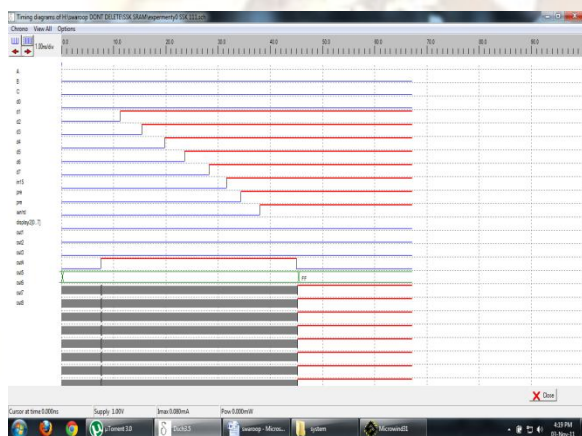


Fig 28. Simulations for 8x8 SRAM Architecture using Sleep Stack with Keeper

Comparison between Basic SRAM and SRAM using LFS technique

Topology	Power dissipation in basic SRAM (mW)	Power dissipation in SRAM using LFS(mW)
50 nm	16.22E-03	0.404E-03
70nm	65.35E-03	1.979E-03
90nm	0.121	8.466E-03
120nm	0.198	6.640E-03
180nm	0.524	49.31E-03

Table 1. power consumption of basic SRAM and SRAM using LFS Technique

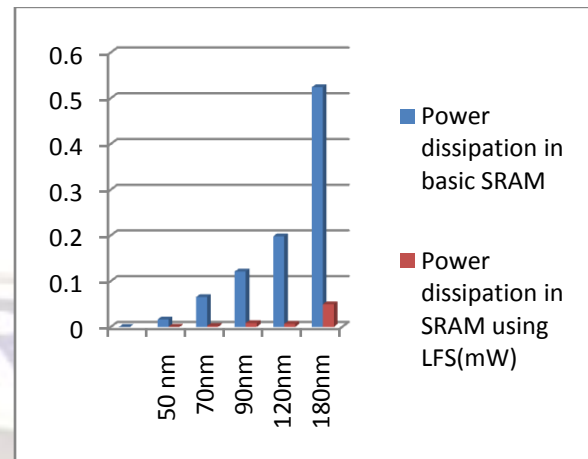


Fig 29. Power consumption graph for basic SRAM and SRAM using LFS in different technologies

Comparison between Basic SRAM and SRAM using SSK technique

Topology	Power dissipation in basic SRAM (mW)	Power dissipation in SRAM using SSK(mW)
50 nm	16.22E-03	0.375E-03
70nm	65.35E-03	1.237E-03
90nm	0.121	3.684E-03
120nm	0.198	4.934E-03
180nm	0.524	31.60E-03

Table 2. power consumption of basic SRAM and SRAM using SSK technique

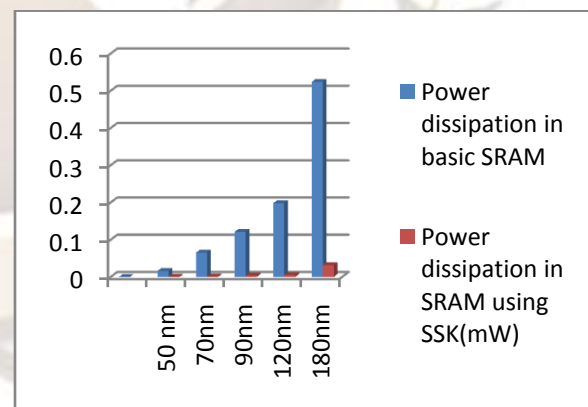


Fig 30. Power consumption graph for basic SRAM and SRAM using SSK in different technologies

VII. Conclusion

In this paper we design 8 bit SRAM Architecture by using the "Leakage Feedback with Stack" & "Sleep Stack with Keeper" leakage current

reduction techniques. The proposed circuits were designed in 0.18um CMOS/VLSI technology. In this paper we observed that the proposed technique "sleepy stack with keeper" have low power consumption when compared to the other low power techniques and having delay and area overhead. Similarly other proposed technique "leakage feedback with stack" is also having low power consumption when compared with existing low power reduction techniques

Based on simulations result with a SRAM Architecture circuit, we find that "Sleep-Stack with keeper approach" achieves up to 60 % less power consumption.

Hence it is concluded that the proposed SRAM Architecture is used for low power designs and these designed techniques are used for high performance and low power applications.

REFERENCES

- [1] New Low power techniques: Leakage Feedback with Stack & Sleep Stack with Keeper. Int'l Conf. on Computer & Communication Technology | ICCCT'10 |.
- [2] International Technology Roadmap for Semiconductors by Semiconductor Industry Association, <http://public.itrs.net>, 2007.
- [3] S. Mutoh et al., "1-V Power Supply High-speed Digital Circuit Technology with Multi-threshold-Voltage CMOS," IEEE Journal of Solid-State Circuits, Vol. 30, No. 8, pp. 847-854, August 1995.
- [4] M. Powell, S.-H. Yang, B. Falsafi, K. Roy and T. N. Vijay Kumar, "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.
- [5] Z. Chen, M. Johnson, L. Wei and K. Roy, "Estimation of Standby Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks," International Symposium on Low Power Electronics and Design, pp. 239-244, August 1998.
- [6] J. Park, "Sleepy Stack: a New Approach to Low Power VLSI and Memory," Ph.D. Dissertation, School of Electrical and Computer Engineering, Georgia Institute of Technology, 2005.
- [7] J.C. Park, V. J. Mooney III and P. Pfeifferberger, "Sleepy Stack Reduction of Leakage Power," Proceeding of the International Workshop on Power and Timing Modeling, Optimization and Simulation, pp. 148-158, September 2004.
- [8] Jan M Rabaey, Anantha Chandrakasan and Borivoje Nikolic, Digital Integrated Circuits, A Design Perspective, Second Edition, Pearson Education, (2002).
- [9] A 2.2GHz 32x4 bit 6T-SRAM Design in 45nm CMOS Xi Chen, Ting Zhu, Harun Demircioglu.
- [10] Designing an SRAM array at the 90nm CMOS tech node Shrivathsa Bhargav and Jaime Peretzman ELEN 4321 – Digital VLSI circuits Columbia University, Fall 2007.



G Santhi Swaroop Vemana was born in A.P, India. He received **B.TECH** degree in **Electronics and Communication Engineering** from Jawaharlal Nehru technological university in 2008. He worked as OFC engineer at united telecoms ltd at GOA during 2009-2010. Presently he is pursuing M.Tech VLSI Design in KL University. His research interests include FPGA Implementation, Low Power Design.



B.Muralikrishna was born in AP, India. He received his, Diploma degree in Electronics & Communication Engineering from SIR C.R.Reddy Polytechnic, Eluru, A.P., Affiliated to the SBTE &T in 2004, B.Tech degree in Electronics & Communications Engineering from Nimra College of Engineering & Technology, Vijayawada, A.P., India, Affiliated to the JNTU Hyderabad in 2007, M.Tech degree in VLSI Design from GITAM University Vishakhapatnam, A.P India, in 2010. Presently he is working as Assistant Professor, Department of ECE

in K.L.University, Guntur, AP, India. He had one year experience in Viraj IT solutions PVT Ltd. as Software Engineer during 2007-2008. He had worked as Teaching Assistant in GITAM University, Vishakhapatnam during 2009-2010. Where he has been engaged in the teaching, research and development of Low-Power VLSI, Design for Testability, CPLD's & FPGA Architectures, Embedded Systems and Fault Tolerance.



Vijaya Bhaskar Madivada was born in A.P, India. He received the **B.TECH** degree in **Electronics & Communications Engineering** from Jawaharlal Nehru Technological University in 2010. Presently he is pursuing M.Tech VLSI Design in KL University. His research interests include FPGA Implementation, Low Power Design.



Raghu kanth . B was born in A.P, India. He received the **B.TECH** degree in **Electronics & Communications Engineering** from Jawaharlal Nehru Technological University in 2007. Presently he is pursuing M.Tech VLSI Design in KL University. His research interests in analog vlsi low power vlsi and testing.



Sai Praveen V was born in AP, India. He received **B.TECH.** in **Electronics & Communications Engineering** from Prasad V. Potluri Siddhartha Institute Of Technology in 2009. Presently he is pursuing M.Tech VLSI design in KL University. His research interest in Low Power VLSI.