A Novel Approach to Reduce the Gate and Sub-threshold Leakage in a Conventional SRAM Bit-Cell Structure at Deep-Sub Micron CMOS Technology

Neeraj Kr. Shukla
ITM University
HUDA Sectore-23A
Gurgaon (Haryana) India

R. K. Singh
VCT-Kumaon Engg. College
Dwarahat, Almora
Uttarakhand, India

Manisha Pattanaik
Deptt. of IT (VLSI Group)
ABV-IIITM
Gwalior (MP) India

ABSTRACT
In the age of scaled silicon technology to improve the functional efficiency of a CMOS design, the device geometry and device parameters are constantly scaled. The major factors of the power consumption due to continuous reduction of the oxide thickness (t_{OX}) is the gate leakage current in both the active and standby mode of the device and other is due to scaled supply voltage, the sub-threshold leakage current. This work proposes a novel approach called as the P3 SRAM Bit-Cell Scheme for the reduction of the active and standby leakage power through the gate and sub-threshold leakage reduction in the active and standby mode of the memory operation. As the gated transistor is of minimum feature size, so the area penalty is minimum in terms of a large memory and can be compromised. To the best of my knowledge, pMOS Gated-Ground and full-supply voltage body bias for pMOS transistor along with the PP SRAM bit-cell structure is used for the first time in the memory bit-cell design to reduce the power in 45nm CMOS technology at V_{DD} = 0.7V and 0.8V. In comparison with the Conventional and PP SRAM Bit-cells, the active power is achieved for Write Data '0' as 89.21% , 94.38%, and for Write Data '1' as 89.23%, 94.45%, respectively at V_{DD} = 0.7V. When the V_{DD} = 0.8V, the active power is achieved for Write Data '0' as 9.15%, 93.63%, and for Write Data '1' as 91.68%, 93.59%, respectively.

General Terms
Conventional SRAM Bit-Cell, Low-Power, Circuit Level Simulation.

Keywords
Deep-Sub Micron, Gate Leakage, Sub-threshold Leakage, PP-SRAM, Stacking Effect.

1. INTRODUCTION
The growing demand for high density VLSI circuits and exponential dependency of the leakage current on the oxide-thickness is becoming a major challenge in deep-sub-micron CMOS technology [1]. According to the ITRS-2003 (International Technology Roadmap), 90% of the chip-area will be occupied by the memory core by 2014 [2]. In the view of ITRS, the gate equivalent oxide thickness as low as 0.5nm for future CMOS technology. In other words, it demands the processor with high processing power, high performance, and low-power on-chip memory. There are several sources for the leakage current, i.e., the sub-threshold current due to low threshold voltage, the gate leakage due to very thin gate oxides, etc., [3]. In Fig.1, the leakage power from a high-performance microprocessor has been shown. It increases steadily, as the technology is scaled down [4].

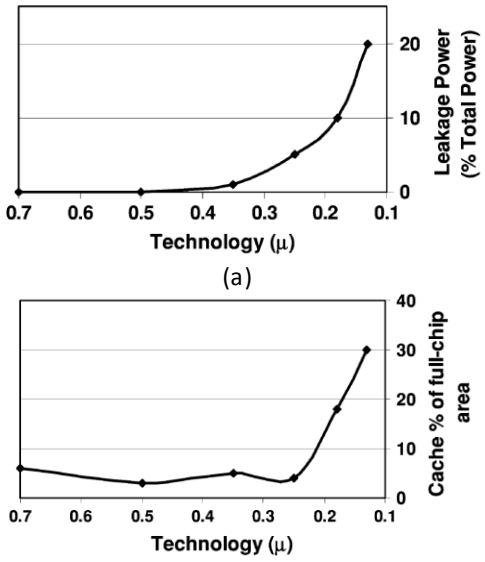

Fig. 1. (a) Leakage Power Percentage of Total Power. (b) Cache Area Percentage of Total Chip Area [4].

Portable mobile-multimedia applications have two operational modes, the active mode and standby mode. Cell phone, for example, have low activity factor, i.e., their idle time is linger in comparison with the device active time. Even at the idle time, these devices affected by the leakage power loss which reduces the battery service time. To address this challenge of power saving, researchers in the domain have reported work at the device level, circuit level, architecture level, and algorithm level. At the circuit level, existing designs need to be modified in such a way that it curbs the drawing of battery current when the device is in the operational and non-operational mode. So, based on the two modes of operation, power reduction techniques have been reported in two categories [5], the leakage control in the standby mode and leakage control in the active mode. In the standby mode, at the circuit level techniques, it cutoff the circuit from the power supply-rails while in the active

mode, certain design modifications, as adding additional transistors are used to minimize the leakage current during runtime.

In this work, we present a CMOS design technique that reduce the overall leakage power in the Conventional SRAM Bit-Cells. In the proposed design, we focus on the static and active power dissipation of the memory, i.e., when the memory is in the inactive mode (when the cell is fully 'ON' with no read/write operation, i.e., the hold state) and active mode (when the cell is active and read/write operation is being performed). The rest of the paper is organized as follows, In Section II, a basic overview of the SRAM Bit-Cell is presented. In Section III, the four major components of the leakage power as Reverse-Biased Junction Leakage, Gate-Induced Drain Leakage, Gate Direct Tunneling Leakage, and Sub-threshold leakage are considered. In section IV, a review of the related work is presented. In Section V, the proposed work on a Low-leakage SRAM Bit-Cell is presented which is followed by the Simulation work and Conclusions in Sections VI and VII, respectively.

2. THE CONVENTIONAL 6T-SRAM BIT-CELL

The Conventional SRAM (CV-SRAM) cell has Six MOS transistors ('4' nMOS and '2' nMOS), Fig.2. Unlike DRAM it doesn't need to be refreshed as the bit is latched in. It can operate at lower supply voltages and has large noise immunity. However, the six transistors of an SRAM cell take more space than a DRAM cell made of one transistor and one capacitor thereby increasing the complexity of the cell [6].

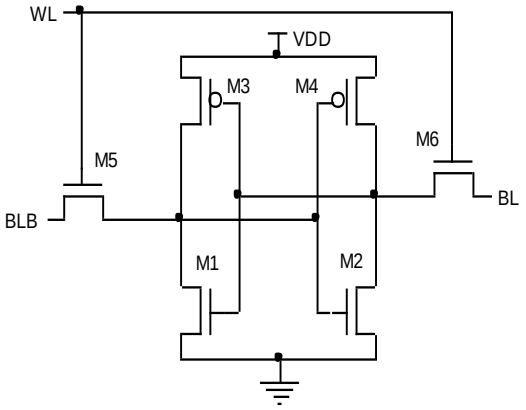

Fig. 2. 6T-CMOS SRAM Cell [6]

The SRAM Bit Cell

The memory bit-cell has two CMOS inverters connected back to back (M1, M3, and M2, M4). Two more pass transistors (M5 and M6) are the access transistors controlled by the Word Line (WL), Fig.2. The cell preserves its one of two possible states '0' or '1', as long as power is available to the bit-cell. Here, Static power dissipation is very small. Thus the cell draws current from the power supply only during switching. But ideal mode of the memory is becoming the main concern in the deep-sub-micron technology due to its concerns in the leakage power and data retentation at lower operating voltages.

The Operation of SRAM Bit-Cell

Although the two nMOS and pMOS transistors of SRAM memory bit-cell form a bi-stable latch, there are mainly the following three states of SRAM memory cell [7], the Write, Read, and Hold states.

A. Standby Operation (Hold)

When WL = '0', M5 and M6 disconnect the cell from Bit-Lines (BL and BLB). The two cross-coupled inverters formed by M1-M4 will continue to reinforce each other as long as they are disconnected from the outside world. The current drawn in this state from the power supply is termed as standby current.

B. Data Read Operation

Read cycle starts with pre-charging BL and BLB to '1', i.e., V_{DD}. Within the memory cell M1 and M4 are ON. Asserting the word line, turns ON the M5 and M6 and the values of Q and Q' are transferred to Bit-Lines (BL and BLB). No current flows through M6, thus M4 and M6 pull BL upto V_{DD}, i.e., BL = '1' and BLB discharges through M1 and M5. This voltage difference is sensed and amplified to logic levels by sense amplifiers.

C. Data Write Operation

The value to be written is applied to the Bit lines. Thus to write data '0', we assert BL=0, BLB = '1' and to write data '1', the BL = '1', BLB = '0', asserted when WL = '1'.

3. LEAKAGE POWER IN SRAM

Leakage is concerned when the device is in the off state. In broader sense, there are four main sources of leakage currents in a CMOS transistor, Fig.3.

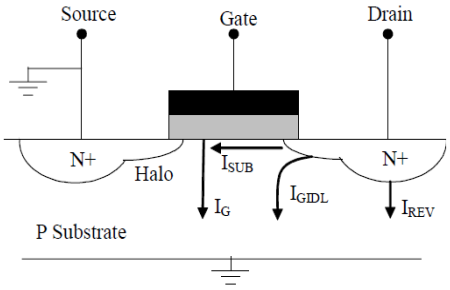

Fig.3: Leakage Current Components in an nMOS Transistor

3.1 Reverse-Biased Junction Leakage (I_{REV})

The junction leakage occurs from the source or drain to the substrate through the reverse biased diodes when a transistor is in the OFF mode. A reverse-biased pn junction leakage has two main components, one is due to the minority carrier diffusion/drift near the edge of the depletion region and the other is due to electron-hole pair generation in the depletion region of the reverse-biased junction. For Ex: in a CMOS inverter with low input voltage, the nMOS is OFF, the pMOS is ON and the output voltage is high. Subsequently, the drain-to-substrate voltage of the OFF nMOS transistor is equal to the supply

voltage (V_{DD}). This results in a leakage current from the drain to the substrate through the reverse-biased diode. If both n and p regions are heavily doped, Band-to-Band Tunneling (BTBT) dominates the pn junction leakage [8]. The junction leakage has a rather high temperature dependency (i.e., around 50–100x/100 ^{o}C).

3.2 Gate-Induced Drain Leakage (I_{GIDL})

The Gate Induced Drain Leakage (GIDL) is caused by high field effect in the drain junction of MOS transistors. In an nMOS transistor with grounded gate and drain potential at V_{DD}, the significant band bending in the drain allows the electron-hole pair generation through avalanche multiplication and band-to-band tunneling. A deep depletion condition is created since the holes are rapidly swept out to the substrate. At the same time, electrons are collected by the drain, resulting in GIDL current. This leakage mechanism is made worse by high drain to body voltage and high drain to gate voltage. Thinner oxide and higher supply voltage increase GIDL current. For Ex: with a V_{DG}=3V and T_{ox} = 4nm, there is roughly a 10 fold increase in the GIDL current when V_{DB} is increased from 0.8V to 2.2V.

3.3 Gate Direct Tunneling Leakage (I_G)

The gate leakage flows from the gate through the "leaky" oxide insulation to the substrate. In oxide layers thicker than 3–4nm, this kind of current results from the Fowler-Nordheim tunneling of electrons into the conduction band of the oxide layer under a high applied electric field across the oxide layer. For lower oxide thicknesses (which are typically found in 0.15µm and lower technology nodes), however, direct tunneling through the silicon oxide layer is the leading effect. Mechanisms for direct tunneling include electron tunneling in the conduction band (ECB), electron tunneling in the valence band (EVB), and hole tunneling in the valence band (HVB), among which ECB is the dominant one. The magnitude of the gate direct tunneling current increases exponentially with the gate oxide thickness T_{ox} and supply voltage V_{DD}. In fact, for relatively thin oxide thicknesses (in the order of 2-3 nm), at a V_{GS} of 1V, every 0.2nm reduction in T_{ox} causes a tenfold increase in I_G [9]. Gate leakage increases with temperature at only about 2x/100 ^{o}C.

3.4 Sub-threshold leakage (I_{SUB})

The Sub-threshold Leakage Current is the drain-to-source leakage current when the transistor is in the OFF mode. This happens when the applied voltage V_{GS} is less than the threshold voltage V_t of the transistor, i.e., weak inversion mode. Sub-threshold current flows due to the diffusion current of the minority carriers in the channel of Metal Oxide Semiconductor Field Effect Transistor (MOSFET). The Eq.1 relates sub-threshold current I_{SUB} with other device parameters.

$$I_{SUB} = I_o e^{\frac{V_{gs} - V_{tho} - \eta V_{ds} - \gamma V_{sb}}{nV_\theta}} \left(1 - e^{\frac{-V_{ds}}{V_\theta}}\right) \quad (1)$$

Where,

$$I_o = \mu Cox \left(\frac{W}{L}\right) V_\theta^2 e^{1.8}$$

As the supply voltage (V_{DD}) is being uniformly scaled down with successive technology nodes. The transistor delay is inversely proportional to the difference of supply and threshold voltage [10], the threshold voltage must also be scaled down proportionally with each technology node to maintain the circuit performance. This leads to an exponential increase in sub-threshold leakage current. As seen from Eq.1, increasing the threshold voltage (V_t) of the transistor is an effective way to reduce sub-threshold leakage.

4. A REVIEW OF RELATED WORK

In general, the SRAM is being considered as the systems of Low-Activity Factor. Researchers in the static memory design domain have reported several integrated circuit and architecture level approaches to address the power and performance in deep-sub-micron CMOS technologies.

In [11], G.Razavipore, et.al., proposed the PP SRAM Bit-Cell structure where the two nMOS access transistors has been replaced by the two High-V_t pMOS access transistors by considering the view that the gate leakage (gate oxide direct tunneling) current will be lower than that of the leakage offered by the nMOS transistors in the ideal mode of the bit-cell. It utilizes the Dual Threshold Voltage technology with Forward Body Biasing (FBB) to reduce the sub-threshold leakage without loosing the performance. As compared to the conventional 6T SRAM Bit-cell, this work reduces the total gate leakage current by 27% and the idle power by 37% with no access time degradation and an improvement in the SNM by 15% in 45nm CMOS technology at V_{DD} = 0.8V.

In [12], B. Amelifard, et.al., proposed an architecture of the SRAM Bit-cell by inserting an extra nMOS transistor in between the ground line and SRAM cell, called as the Gated-Ground technique to reduce the leakage power consumption in the high performance cache memories with single V_t (transistor threshold voltage) process. The turning OFF of the Gated-Ground transistor gives large reduction in leakage power. This technique requires no extra circuitry, i.e., the row decoder itself can be used to control the Gated-Ground transistor. At the simulation level on 100nm and 70nm, leakage power consumption is achieved at 16.5% and 27% in L1 cache and 50% and 47% reduction in L2 cache with less than 5% impact on execution time and within 4% increase in area overhead.

In [13], K. Nil, et.al., proposed Multiple Threshold CMOS Technique (MT-CMOS) to influence the leakage current. In the active mode of the operation of the memory cell, the low threshold voltage is preferred because of the higher performance (speed). However, in the standby mode of operation, high threshold voltage is useful for reduction of the leakage power. Therefore, if transistors can be set to different threshold voltages, most likely using Reverse-Body-Bias (RBB), the threshold voltage can be set according to different modes of operation of the memory. At the time of accessing the that is in the Standby mode, it has some overheads, as the threshold voltage must be returned to the proper level before the value can be read.

5. PROPOSED WORK – P3 SRAM BIT CELL

In this work, a novel architecture called as the P3 SRAM Bit-Cell Scheme has been proposed for the reduction of the active and standby leakage power through the gate and sub-threshold leakage reduction in the active and standby mode of the memory operation. In it a gated-ground pMOS transistor (typical) has been used with the PP SRAM Bit-Cell structure [11]. The two pMOS access transistors reduce the gate leakage (gate oxide direct tunneling) current. This current will be lower than that of the leakage offered by the nMOS transistors in the ideal mode of the bit-cell due to the high barrier width of the pMOS in comparison with the nMOS transistor. Full supply body biasing is being used to reduce the sub-threshold leakage without loosing the cells performance. As the gated transistor is of minimum feature size, so the area penalty is minimum in terms of a large memory and can be compromised. The extra pMOS transistor between the cell and ground, produces the stacking effect in conjunction with bit-cell. When the Gated-pMOS is turned off, opposes the leakage current flow through it.

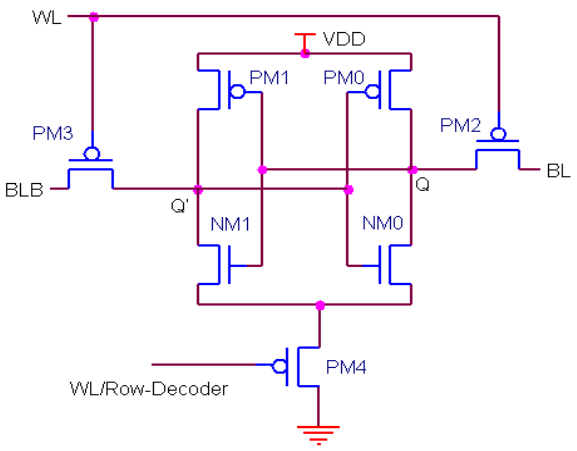

Fig. 4. The proposed P3 SRAM Bit-Cell

When WL = '0', the bit-cell behaves as the normal bit-cell and ready to perform the read/write operation in the active mode operation. When WL = '1', bit-cell is in the OFF state and puts the memory in the standby mode to hold data. To the best of my knowledge, pMOS Gated-Ground and full-supply voltage body bias for pMOS transistor along with the PP SRAM bit-cell structure is used for the first time in the memory bit-cell design to reduce the power in 45nm CMOS technology at V_{DD} = 0.7V and 0.8V.

6. SIMULATION WORK

The simulation work is being performed at the Cadence-Virtuoso Sch. and Spectre-AMS Designer platform at CMOS Technology, pdk 45nm for t_{ox} = 2.4nm, V_{th} = 2.2V, at V_{DD} = 0.7V, and 0.8V and T =25^{0}C when Active Data for Write '0' and '1' and Standby Stored Data '0' and '1', shown in Fig. 5-8.

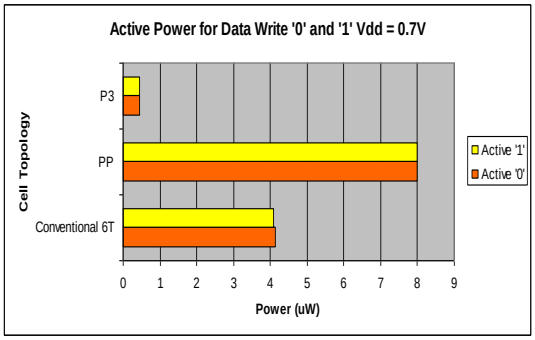

Fig. 5. Active Power Consumption when Write Data is '0' at V_{DD} = 0.7V.

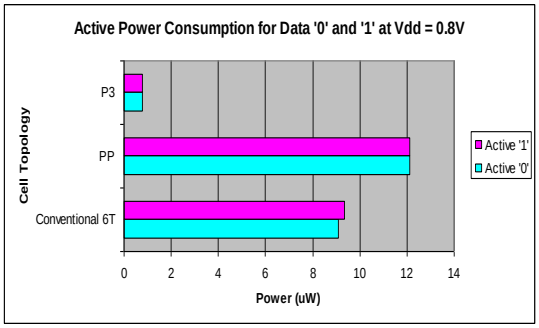

Fig. 6. Active Power Consumption when Hold Data is '1' at V_{DD} = 0.8V.

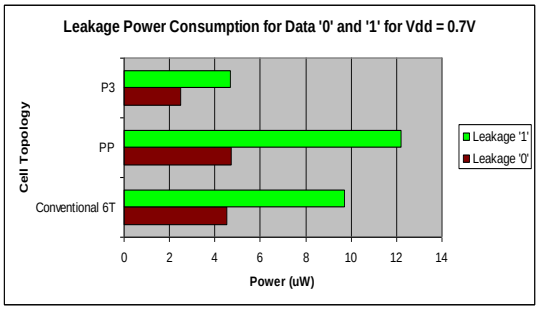

Fig. 7. Leakage Power Consumption when Hold Data is '0' at V_{DD} = 0.7V.

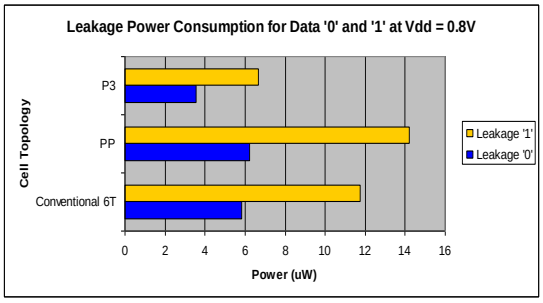

Fig. 8. Active Power Consumption when Write Data is '1' at V_{DD} = 0.8V.

7. CONCLUSIONS

In this paper, a novel structure of the SRAM Bit-Cell, called as Stacked-pMOS PP SRAM Bit-Cell (or simply called as P3-SRAM Bit-Cell) structure, is presented. The proposed bit-cell utilizes the Gated-Ground technique for transistor stacking, PP-SRAM Cell, and full-supply body-biasing to reduce the active and standby power in the memory. It is found that P3 scheme effectively reduces the leakage power while maintaining the cell performance at 45nm CMOS technology node with a little area penalty of gated transistor. As the gated transistor is of minimum feature size, so the area penalty is minimum in terms of a large memory. In comparison with the conventional and PP SRAM Bit-cells, the active power is achieved for Data write '0' as 89.21% , 94.38%, and for Data write '1' as 89.23%, 94.45%, respectively at V_{DD} = 0.7V. When the V_{DD} = 0.8V, the active power is achieved for Data write '0' as 9.15%, 93.63%, and for Data write '1' as 91.68%, 93.59%, respectively. Also, in comparison with the conventional and PP SRAM Bit-cells, the leakage power is achieved for Data Hold '0' as 44.30%, 46.67%, and for Data Hold '1' as 51.83%, 61.67%, respectively at V_{DD} = 0.7V. When the V_{DD} = 0.8V, the leakage power is achieved for Data Hold '0' as 38.82%, 42.84%, and for Data Hold '1' as 43.23%, 53.10%, respectively. The work is performed for t_{ox} = 2.4nm, V_{th} = 2.2V, Cadence pdk 45nm, at V_{DD} = 0.7V, and 0.8V at T =25^0C when Active Data for Write '0' and '1' and Standby Hold Data '0' and '1'.

8. ACKNOWLEDGEMENT

The authors are grateful to their respective organizations for their encouragement and support.

9. REFERENCES

[1] Neeraj Kr. Shukla, R.K. Singh, and Manisha Pattanaik, "Design and Analysis of a Novel Low-Power SRAM Bit-Cell Structure at Deep-Sub-Micron CMOS Technology for Mobile Multimedia Applications", International Journal of Advanced Computer Science and Applications (IJACSA), USA, Vol.2, Issue 5, May, 2011, pp. 43-49.

[2] International Technology Roadmap for Semiconductors. Online-Available at http://www.publicitrs.net

[3] K. M. Kao, et. al., "BSIM4 Gate Leakage Model Including Source-Drain Partition," in Proc. Int. Electron Devices Meeting, Dec. 2000, pp. 815–818.

[4] Kevin Zhang, Uddalak Bhattacharya, Zhanping Chen, Fatih Hamzaoglu, Daniel Murray, Narendra Vallepalli, Yih Wang,B. Zheng, and Mark Bohr, "SRAM Design on 65-nm CMOS Technology with Dynamic Sleep Transistor for Leakage Reduction" IEEE Journal of Solid-State Circuits, Vol. 40, No. 4, APRIL 2005, pp. 895-901.

[5] B.S. Deepaksubramanyan and Adrian Nu˜nez, "Analysis of Subthreshold Leakage Reduction in CMOS Digital Circuits", Proceedings of the 13th NASA VLSI Symposium, Post Falls, IDAHO, USA, June 5-6, 2007, pp 1-8.

[6] Neeraj Kr. Shukla, Shilpi Birla, R.K. Singh, and Manisha Pattanaik, "Speed and Leakage Power Trade-off in Various SRAM Circuits", International Journal of Computer and Electrical Engineering (IJCEE), Singapore, Vol.3, No.2, Apr. 2011, pp. 244-249.

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

[8] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. New York: Cambridge Univ. Press, 1998, ch. 2, pp. 94–95.

[9] K. Cao,W.-C Lee,W. Liu, X. Jin, P. Su, S. Fung, J. An, B. Yu, and C. Hu, "BSIM4 gate leakage model including source drain partition," in *Tech. Dig. Int. Electron Devices Meeting*, 2000, pp. 815–818.

[10] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S. H. Lo, G. Sai-Halasz, R.Viswanathan, and et al., "CMOS scaling into nanometer regime", Proc. of the IEEE, vol. 85, Apr. 1997, pp. 486–504.

[11] G. Razavipour, A. Afzali-Kusha, and M. Pedram, "Design and Analysis of Two Low-Power SRAM Cell Structures", IEEE Transaction on Very Large Scale Integration (VLSI) Systems, Vol. 17, No. 10, Oct. 2009, pp. 1551-1555.

[12] Behnam Amelifard, Farzan Fallah, and Massoud Pedram, "Reducing the Sub-threshold and Gate-tunneling Leakage of SRAM Cells using Dual-*Vt* and Dual-*Tox* Assignment", Proceeding of Design, Automation and Test in Europe, Munich, 2006. Date: 6-10 March 2006, pp. 1-6.

[13] K. Nil, et.al., " A Low-Power SRAM using Auto-Backgate-Controlled MT-CMOS", Proceedings of the International Symposium on Low-Power Electronics and Design, Aug. 1998, pp. 293-98.

ABOUT THE AUTHORS

[1]**Neeraj Kr. Shukla** (IEEE, IACSIT,IAENG, IETE, IE, CSI, ISTE,VSI-India), a Ph.D. Scholar at the UK Technical University, Dehradun (Uttarakhand) India is an Asst. Professor in the Department of Electrical, Electronics & Communication Engineering, ITM University, Gurgaon, (Haryana) India. He has received his M.Tech. (Electronics Engineering) and B.Tech. (Electronics & Telecommunication Engineering) Degrees from the J.K. Institute of Applied Physics & Technology, University of Allahabad, Allahabad (Uttar Pradesh) India in the year of 1998 and 2000, respectively. His main research interests are in Low-Power Digital VLSI Design and its Multimedia Applications, Open Source EDA, and RTL Design.

[2]**R.K. Singh** (IAENG, IE, ISTE), Professor in the Department of Electronics & Communication Engineering, VCT-Kumaon Engineering College, Dwarahat, Almora (UK) India. He is being honored with the Ph.D. in Electronics Engineering in the Year 2003 from the University of Allahabad, Allahabad (Uttar Pradesh), India. He has received his M.E. (Electronics & Control Engineering) in 1992 from BITS, Pilani, (Rajasthan) India. He has around thirty five research publications in the conferences and journals at national and international. He has also guided eight ME thesis. He has authored seven text-books in the field of VLSI Design, Basic Electronics, and Opto-Electronics. He has worked at various capacities in the

Academic domain such as, the Principle, Kumaon Engineering College, Dwarahat in the year 2003-04, Director (O), Directorate of Technical Education, Uttaranchal in the year 2005, and Joint Director, State Project Facilitation Unit, Dehradun for the World Bank TEQIP Project. Apart from his industrial experience, he contributed as a Scientist and Senior Scientist in Engineering Research at Central Electronics Engineering Research Institute (CEERI), Dehradun (UK) India, with a focus in Fibre Optics Communication and their subsequent application in optical devices to various other technologies and VLSI Design and Con Controls (P) Ltd. He is also the recipient of couple of prestigious awards, e.g., Rastriya Samman Puruskar, Jewel of India Award, Rastriya Ekta Award, Life Time Achievement Award, and Arch of Excellence Award. His current areas of interest are VLSI Design, Opto-Electronics and its applications.

[3]**Manisha Pattanaik** (IE, ISTE) has been honored the Ph.D. from Indian Institute of Technology (IIT) Kharagpur, India in the field of VLSI Design from the Department of Electronics and Electrical Communication Engineering in the year 2004. Currently she is an Associate Professor (VLSI Group) at ABV-India Institute of Information Technology & Management (ABV-IIITM), Gwalior, (Madhya Pradesh), India. She has been awarded various scholarships, e.g., National Scholarships, Merit Scholarships and MHRD Fellowships. She shared the responsibility in the capacity of referee for IEEE International Conferences on VLSI Design for two consecutive years, 2003-04. Her areas of interest are Leakage Power Reduction of Nano-Scale CMOS Circuits, Characterization of Logic Circuit Techniques for Low-Power/Low-Voltage and High performance analog & digital VLSI applications and CAD of VLSI Design.