Probabilistic Analytical Framework to Minimize Expected Leakage by Employing A Dual V_{th} Design Technique

Rashmi Bahal, Shyam Akashe , Arun Agrawal

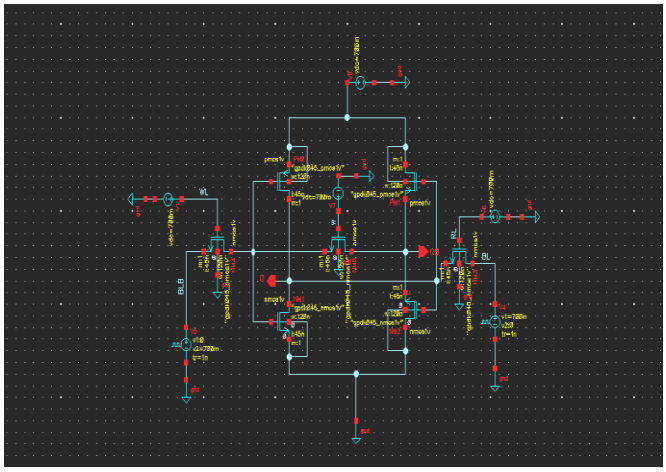
Abstract-The growing demand in the multimedia rich applications are motivating the low-power and high-speed circuit designer to work more closely towards the design issues arising from the design trade-offs in power and speed. This paper targets the modeling and simulation of leakage currents and its minimization approach by Dual Vt approach. We consider the optimal selection of Vth under a statistical model of threshold variation. Probabilistic analytical models are introduced to account for the impact of Vth uncertainty on leakage power and timing slack. Using this model, we show that the non-probabilistic analysis significantly underestimates the leakage power.

Index Terms- Dual Vth ,high-speed, leakage current, , Probabilistic analytical models

I. INTRODUCTION

With the number of transistors on a chip rapidly increasing, and the ever-increasing levels of integrated cache, leakage power management is indispensable for cost effective packaging and cooling solutions in high-end microprocessors. Leakage power is also a concern in low-end mobile system-on-chips where the low standby power feature is crucial. As a result of continued scaling of MOS transistor, a dramatic increase in the performance of VLSI ICs has been achieved. However, as a result of scaling, power dissipation due to leakage currents has also increased dramatically and is a major source of concern especially for low power applications. Till now the dominant leakage mechanism has been due to drain source sub threshold current. Assuming this leakage mechanism, a number of techniques have been proposed in literature for reducing the impact of leakage power dissipation such as gated-Vdd scheme Dual-Vt SRAM etc.

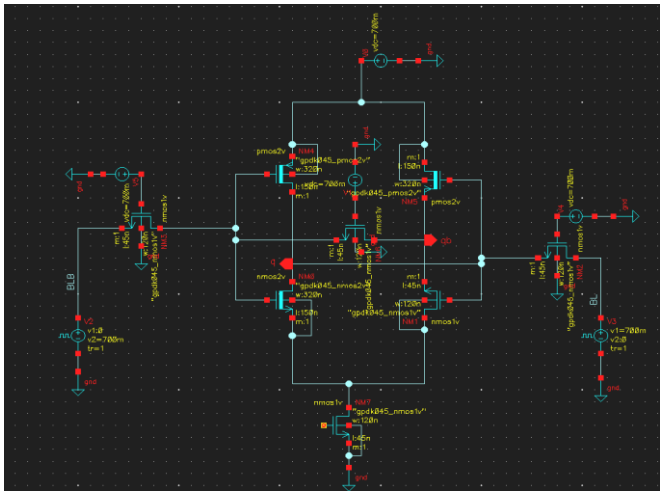
The above idea was applied to the previously discussed 7T Dual-Vt SRAM cell, as shown in figure 1. Due to complete isolation of read and write circuits, the transistors N4 and N5 are active only in reading process. Therefore, during standby and write operation SUBSL is set to zero and consequently, the mentioned transistors preserve their high threshold voltage values.So, the leakage power consumption during these phases is reduced. The major contribution of this paper is the analysis of the dual Vth design methodology in the presence of large variation in the threshold voltage. In this paper we derive a set of analytical models that allow the probabilistic analysis of leakage power within the dual-Vth design methodology. The equations that we derive can also be used to probabilistically describe the leakage power, in general. A specific issue that we seek to answer is the way in which the values of the two Vth can be selected in the presence of variability, which is, we seek to find the optimal separation between the nominal values of low and high Vth such that the overall power savings are optimized. With large Vth variability, the separation must be big enough to overcome the statistical noise.



(a)

Manuscript received December 19, 2011.

Rashmi Bahal,M.Tech VLSI Student, Electronics Department, Gwalior (M.P) India ,vlsi.rashmi@gmail.com
Shyam AKashe.,Associate Professor, Electronics Department, Gwalior (M.P) India ,shyam.akashe@itmuniversity.ac.in
Arun Agrawal,Assistant Professor, Electronics Department, Gwalior (M.P) India ,development_arun@rediffmail.com.



(b)

Fig 1.1(a) Novel 7T SRAM and (b) Dual Vt 7T SRAM Cell

II. LEAKAGE REDUCTION SCHEMES IN SRAM

In this section, we discuss about different type of leakage current present in MOS transistors. The amount of leakage current is increased with advancement of technology generation. Reduction in dimension, increase of doping, reduction of V_T etc causes the increase of leakage current in every generation. At the 90nm technology node, leakage power may make up 42% of total power [1] [2]. The primary reason for this increase in leakage power is the reduction of threshold voltage (Vth) of devices, which is causing an exponential increase in leakage current... So leakage reduction technique is the demand of modern age. Many researchers have been done in this area and many researches are going on in this direction. Without finding proper leakage reduction technique it is not possible to design a circuit properly.

2.1. By Dual threshold Voltage

Another approach, called Dual VTH CMOS, uses high-VTH devices on non-critical paths and low-VTH devices on the critical ones. This technique requires an algorithm that performs a search for the places where the high-VTH devices can be placed [11] For a logic circuit, a higher threshold voltage can be assigned to some transistors in noncritical paths so as to reduce the leakage current, while the performance is maintained due to the use of low threshold transistors in the critical path(s) .Dual-Vth assignment is an efficient technique for leakage reduction.

2.2.Transistor stacking effect

Sub threshold leakage current flowing through a stack of series-connected transistors reduces when more than one transistor in the stack is turned off. This effect is known as the stacking effect this takes place when more than one transistor in series is turned off. At that time it is found that there is a huge reduction in leakage current.

2.3 .By gating the supply voltage

The last approach considered is power supply gating, in which the power supply is shut down so that idle units do not consume leakage power. This can be done using "sleep" transistors. Normally, there is one sleep transistor per gate, but larger granularities are possible, which require fewer but larger devices. The problems with the technique are reduced performance and noise immunity if care is not exercised when designing the sleep transistors. If there is intention to provide support for Dynamic Voltage Scaling (DVS), MTCMOS clearly lacks the required power supply scaling capabilities, making it necessary to use another scheme. With this in mind, we have also considered voltage regulators that can both remove and scale the supply voltage to the circuit

III. REDUCING STATIC POWER BY USING PROBABILISTIC VTH MODEL

In this section, we develop a probabilistic model to describe static power reduction using a dual Vth design under Vth variability. We seek a model and an optimization strategy that will allow us to reduce leakage current (power), while ensuring that the clock frequency satisfies certain requirements. The way is to formulate the power reduction problem as a constrained optimization problem

$$\min P_{leak}\ , s.t.\ f_{clock} \geq f^*$$

However, in relation to the probabilistic model, we will treat *both Plea k* and delay probabilistically. Also, in our actual formulation, we normalize the power of a dual Vth design to that of the single Vth design, and employed it as a measure of static power optimization. The dual-to-single leakage power ratio (*R*) is given by:

$$R = P_{leak}\ /P_{single}$$

Now, by considering the probabilistic nature of our formulation, we will minimize the expected value of the power ratio, *R*, under the constraint that with a certain probability the frequency target will be met:

$$\min E\{R\}\ , s.t.\ P\{f_{clock}\ \ f^*\} \geq \alpha$$

In the following sections we derive the appropriate probabilistic models for power ratio and delay.

IV. MODEL OF STATIC POWER OPTIMIZATION

To measure the minimum static power achievable in a dual Vth design, we first model the ratio of static power with respect to the two threshold voltages. The amount of static power optimization is defined as the ratio of the static power of a dual Vth design compared to the static power in a single Vth design so this value is less than one.

First, we follow [1] and [2] to formulate the static power improvement without the variation in Vth, and then modify the formulation to include a probabilistic Vth model. In the model proposed by [1] and [2], the circuit is modeled as a collection of non-crossing combinational logic paths. Leakage power is reduced by assigning a subset of gate stages to higher Vth. To model the leakage power, we assume that an arbitrary fraction of the total paths gate stages can be assigned to higher Vth. Since the total transistor width and gate stages are largely proportional, we can model the power as a ratio of gate stages. Then the dual-to-single leakage power ratio(R) can be modeled by:

$$R = 1 - \frac{N_h}{N}\left[1 - 10^{\frac{v_{th}^h - v_{th}^1}{s}}\right] \quad (1)$$

Where s is the sub threshold swing, ^{l}Vth and ^{h}Vth are the low and high threshold voltages, N_h is the number of gate stages in the entire circuit set to the high threshold voltage, and N is the total number of gates stages in the circuit.

We now modify Eq. 1 to allow the probabilistic treatment of threshold voltage variation. We model threshold voltage as a random, Gaussian variable. Empirical evidence suggests that variation in threshold voltage can be modeled by normal (Gaussian) distribution [1]. Under the assumption that Vth is a Gaussian random variable, the power ratio R follows a log normal distribution. Although the fraction of gates assigned to the high threshold voltage, Nh N, is also a function of the threshold voltage, a simulation suggests that Eq. 2 is a reasonable approximation of E[R], which allows us to simplify the derivation. The expected value of a log-normally distributed function can be observed in closed form is given by

$$E[R] = 1 - \frac{N_H}{N}\left[1 - g\{V_{OS}\}\right] \quad (2)$$

Where $V_{OS} = V_{th}^h - V_{th}^l$ and

$$g[V_{os}] = exp\left[-V_{OS}.\ \frac{In(10)}{s} + \left\{\frac{\sigma^2 V_{th}}{2}\right\}.\left\{\frac{ln(10)}{s}^2\right\}\right] \quad (3)$$

Where $V_{OS} = E[V_{os}]$ and $\sigma_{V_{th}}^2 = \sigma_{V_{th}^h}^2 + \sigma_{V_{th}^l}^2$

Furthermore the variance is given by

$$\sigma_R^2 = \left[\frac{N_h}{N}\right]^2 \left[g[V_{OS}]\right]^2 \left[e^{\sigma^2 VTH\ \left(\frac{\ln(10)}{s}^2\right)} - 1\right] \quad (4)$$

Now the power ratio R signifies the ratio of the static power of a dual Vth design and that of a single Vth design taking into consideration, the statistical nature of the threshold voltage. In our work we analyze the effect of hVth and σVth on the power ratio. In the next section we derive a probabilistic model of path delay that allows us to express the ratio of gates stages set to high Vth and thereby allow us to measure the expected power ratio.

V. PROBABILISTIC CIRCUIT DELAY MODELING

In a dual Vth design methodology, the static power is minimized by setting the fraction of gates on paths with timing slack to a higherVth. This makes the paths slower but less leaky. In this section, we provide a probabilistic description of path delay degradation as a result of setting a portion of the gate stages on a path to a higher Vth.

Let $\gamma = {}^{d_h}/_{d_l}$ be the degradation of delay per gate stage, with d_h and d_l corresponding to the gate delay at high and low Vth. Using the alpha-power law model [15], we can show that this ratio can be described as

$$\gamma = \left[\frac{V_{dd} - V_{th}^l}{V_{dd} - V_{th}^h}\right]^2 \quad (5)$$

Where V_{dd} is the power supply voltage, and $_{l}$Vth and $_{h}$Vth are the low and high threshold voltages. We now describe the degradation of delay of a path when a certain fraction of its gates has been set to high V_{th}. Let

$$\Delta D = D_{lh} - D_l \quad (6)$$

be the change in delay of a path after a fraction of gates along the path is set to high V_{th}. Here, D_l is the delay of a path with all its gates at low $V_{th,}$ and D_{lh} is the delay of a path with a mixture of gates at low and high Vth. We also rely on the observation that, to a good approximation, path delay is proportional to the total path capacitance [1] and [2], which, we in turn approximate by gate stages. Combining (5) and (6), we can show that:

$$\Delta D = \sum_i^{n_h}(d^h - d^l) = \sum_i^{n_h} d^l(\gamma_i - 1) = d^i \sum_i^{n_h}(\gamma_i - 1) \quad (7)$$

where n is the number of gate stages along a path, and n_h is the number of gate stages assigned to high Vth.Due to the uncertainty in threshold voltage, path delay is needed to be modeled probabilistically. To do this we first derive the variance of path delay (Eq. 7). In order to simplify the derivation, it is convenient to include all Vth-induced variability into γ , that is, we assume that only the difference between the low and high threshold voltages is a random quantity.

This reduction can be easily arrived, and does not jeopardize the generality of our framework. We further assume that the individual threshold voltage variations are uncorrelated. Then,the path delay variance is given by

$$Var\{\Delta D\} = \sigma_{\Delta D}^2 = \text{Var}\{d\sum_i^{n_h} d^l(\gamma_i - 1) = d^l \sum_i^{n_h}(\gamma_i - 1) \quad (8)$$

Defining $V_T = V_{dd} - V_{th}^h$ and $V_{os} = V_{th}^h - V_{th}^l$ we can write this equation as

$$\gamma = \left[1 + \frac{V_{os}}{V_T}\right]^{\alpha} \quad (9)$$

To finish the derivation, we use the statistical delta-method to find the variance and mean of γ

$$\gamma = 1 + \alpha\frac{(V_{os} + \Delta V_{os})}{V_T} \quad (10)$$

The variance of γ, which follows from the statistical delta method, is given by

$$\text{Var}\{\gamma\} = \frac{\alpha^2}{V_T^2}\sigma_{Vth}^2 \quad (11)$$

and the mean of γ is given by

$$\gamma_0 = 1 + \alpha\,(V_{os}/V_T) \quad (12)$$

By modeling the variability of Vth in γ, the mean path delay is

$$E[\Delta D] = n_h d^l(\gamma_0 - 1) \quad (13)$$

and the variance of the path delay is

$$\text{Var}\{\Delta D\} = \sigma_{\Delta D}^2 \quad (14)$$

Having derived the mean and variance of the path delay, we go onto specify the delay constraints in the next section

VI. FINDING OPTIMAL VTH SEPARATION UNDER THE PROBABILISTIC MODELS

In this section we derive the analytical framework for finding the optimal value of the high Vth under the statistical description of path delays and power consumption. The objective is to minimize the expected power ratio, E[R] , under the timing constraint that no path delay exceeds the critical path delay. In order to make the analysis tractable, we use a simplified model in which the circuit consists of a collection of M non-crossing paths. In contrast to earlier approaches, this constraint is formulated probabilistically, in terms of the probability of meeting the timing constraint

$$P(ckt\ delay \le T) = [max\{D_1 - D_M\} \le T] = \alpha_d$$

Because we assume that the paths are non-crossing

$$P(ckt\ delay \le T) = \prod_{i=1}^{M}[D_i \le T] = \alpha_d$$

For a specified confidence level α_d we can compute the probability that every path does not exceed the timing constraint. This probability is given by

$$P\{(D_i \le T)\} = \alpha_d^{1/m} \quad (15)$$

The timing constraint that we enforce is that no path delay can be greater than the critical path delay, which is normalized to one. We can express this constraint by first letting Si be the amount of slack in a single path

$$S_i = (1 - D_i) \quad (16)$$

To insure that no path delay is greater than the critical path delay,we must satisfy

$$P[\Delta D_i \le S_i] = \alpha_c \quad (17)$$

Since we assume the path delays to be described by the Gaussian distribution, we can re-write (17) as

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$\Delta D_l^{\alpha} = E[\Delta D_l] + \Phi^{-1}(\alpha_c).\sigma_{\Delta Dl} \quad (18)$$

where φ is the cumulative distribution function of the standard normal distribution.

Any value of α_c can be used, for example, a convenient value is = 99.7% , such that $\Phi^{-1}(\alpha_c) = 3$

With the constraint on delay given in Eq. 17 we can now find the number of gates along a path to set to the high Vth. As follows from Eq. 1, the power ratio R is linearly proportional to
the number of gate stages that we can set to high Vth in the entire circuit. Note that

$$N_h = N.N_{ave}^h \quad (19)$$

where N_{ave}^h is the average ratio of gates set to high Vth per path, which can be found by evaluating

$$N_{ave}^h = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{n_h}{n}\right]i \quad (20)$$

where M is the number of paths in the circuit.

We can compute the value of (n_h/n_l) per path by considering the amount of timing slack per path. Using Eq. 13, 14, and 17we may re-write Eq. 18 as:

$$\Delta D_l^{\alpha} = n_{hd^1}(\gamma_0 - 1) + \Phi^{-1}(\alpha_c).d^1.\sqrt{n_h}.\frac{\alpha}{V_T}\ \sigma V_{th} \leq S_I \quad (21)$$

We can now set an analytical quadratic equation to find n_h per path as a function of Si. To quantitatively evaluate the average ratio of gates set to the higher Vth, we assume a specific shape of the initial path delay distribution, p(DL). First we use the triangular distribution that has been shown to be characteristic of many circuits [10]. To make the analysis simpler, the model normalizes all path delays to critical path delay. Then, integrating n_h(Si) for individual paths over the entire path delay distribution we get

$$N_{ave}^h = \frac{\int_0^1 n_h\ (si)\,p(Dl)\Delta D}{\int_0^1 n_{ip(Dl)\Delta D}} \quad (22)$$

where *ni* is the number of gate stages per path. The analytical integration of Eq. 22 is possible; however, it does not provide any particular insight into the nature of the problem. In our implementation, we used numerical integration to compute the value of for a distribution of specific form. Then, we can link it back to the original equation for the expected power ratio,

Eq. 2, such that the result is only a function of the Vth separation:

$$E[R] = 1 - N_{ave}^h\ (1 - g(V_{OS}))$$

This equation can now be used to explore the dependence of the expected power ratio on Vth separation, and for finding the optimal value of the higher Vth

VII. ANALYSIS AND RESULTS

We have implemented the analytical results developed in the earlier sections in a Cadence analog environment. The greater the variation in Vth the higher the value of high Vth has to be to minimize static power. Hence, the probabilistic model for Vth becomes more important as the variation in Vth gets larger, which is predicted to occur given current scaling trends.

Table 1. shows the measurements of leakage in Novel 7T SRAM Cell and Dual Vt 7T SRAM Cell

Process technology	45nm
Power supply voltage	0.7v
Temperature	27ºC
Novel 7T SRAM Cell	3.6846pA
Dual V_t 7T SRAM Cell	58.9aA

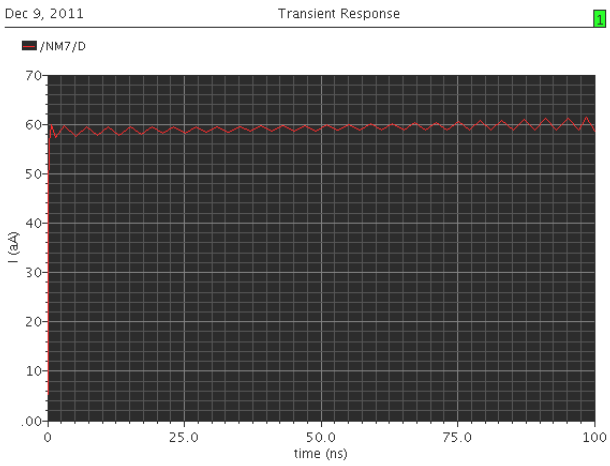


Fig 6.1 .Leakage Waveform in Dual Vt SRAM

VIII. CONCLUSIONS

In this paper we derive a probabilistic analytical framework to minimize expected leakage by employing a dual Vth design technique. By this analysis we observed that the non-probabilistic model severely underestimates the expected leakage current. We also observe that under variability a larger separation between the lower and higher Vth is required to achieve optimal leakage power reduction. All the results are obtained using Cadence tool.

ACKNOWLEDGMENT

The author would like to thank Institute of Technology and Management, Gwalior for providing the Tools and Technology for the work to be completed.

REFERENCES

[1] S. Narendra, D. Blaauw, A. Devgan, F. Najm, "Leakage issues in IC design: trends, estimation, and avoidance,"Proc. of ICCAD, 2003.

[2] J. Kao, S. Narendra, A. Chandrakasan, "Subthreshold Leakage Modeling and Reduction Techniques," ICCAD,2002, pp. 141-149.

[3] S. Sirichotiyakul, et al, "Stand-by power minimization through simultaneous threshold voltage selection and circuit sizing," DAC, 1999, pp. 436-441.

[4] Asenov, S. Kaya, JH. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," IEEE Transactions on Electron Devices, vol.49, no.1, Jan. 2002, pp.112-19.

[5] M. Orshansky, L. Milor, P. Chen, K. Keutzer, C. Hu, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Volume: 21 Issue: 5 , pp. 544-553, May 2002.

[6] W. Feller, "Probability Theory and its Applications," *John Wiley & Sons*, second ed., vol. II. 1971.

[7] T. Sakurai, A.R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, Volume: 25 Issue: 2, pp. 584-594, April 1990.

[8] J.D. Warnock, et al., "The circuit and physical design of the POWER4 microprocessor," *IBM J. of R&D*, Vol. 46, pp. 27-52, Jan. 2002.

[9] M. Ketkar and S. S. Sapatnekar, "Standby Power Optimization via Transistor Sizing and Dual Threshold Voltage Assignment," in Proc. International Conference on Computer-Aided Design, 2002, pp. 375-378.

[10] M. Mani, A. Devgan, and M. Orshansky, "An Efficient Algorithm for Statistical Minimization of Total Power Under Timing Yield Constraints," in Proc. Design Automation Conference, 2005, pp. 309-314.

[11] S. Mukhopadhyay, C. Neau, R. T. Cakici, A. Agarwal, C. H. Kim, and K. Roy,"Gate Leakage Reduction for Scaled Devices Using Transistor Stacking," IEEE Transactions on VLSI Systems, vol. 11, no. 4, pp. 716-730, 2003.

[12] S. Mukhopadhyay and K. Roy, "Accurate Modeling of Transistor Stacks to Effectively Reduce Total Standby Leakage in Nano-Scale CMOS Circuits," in Proc. Symposium on VLSI Circuits, 2003, pp. 53-56.

[13] D. Nguyen, A. Davare, M. Orshansky, D. Chinney, B. Thompson, and K. Keutzer,"Minimization of Dynamic and Static Power Through Joint Assignment of Threshold Voltages and Sizing Optimization," in Proc. of the International Symposium on Low Power Electronics and Design, 2003,

[14] P. Pant, R. K. Roy, and A. Chatterjee, "Dual-Threshold Voltage Assignment withTransistor Sizing for Low Power CMOS circuits," IEEE Transactions on VLSISystems, vol. 9, no. 2, pp. 390-394, April 2001.

[15] K. Zhang et al., "SRAM design on 65-nm CMOS technology with dynamic sleep transistor for leakage reduction," IEEE J. Solid-State Circuits, vol. 40, no. 4, Apr.2005, pp. 895-901.

[16] N. Azizi et al, "Low-leakage asymmetric-cell SRAM,"IEEE Trans. on VLSI Systems, vol. 11, no. 4, Aug. 2003, pp.701-715.

[17] C. H. Kim et al, "A forward body-biased low-leakage SRAM cache: device, circuit and architecture considerations," IEEE Trans. On VLSI Systems, vol. 13, no.3, Mar. 2005, pp. 349-357.

[18] C. Kim and K. Roy, "Dynamic Vt SRAM: a leakage tolerant cache memory for low voltage microprocessor," inProc. ISLPED, 2002, pp. 251–254.

[19] H. Qin et al, "SRAM leakage suppression by minimizing standby supply voltage," in Proc. of ISQED, 2004, pp.56-60.

[20] K. Zhang et al, "A 3-GHz 70Mb SRAM in 65nm CMOS technology with integrated column-based dynamic power Supply," in Proc. ISSCC, 2005, pp. 474–475.

AUTHORS PROFILE

Rashmi Bahal was born in 19th May 1987. She received her B.E from MITS, Gwalior in 2008. She is currently pursuing Mtech in VLSI DESIGN from ITM Gwalior. Her researches interests are VLSI Design, Low Power

Shyam Akashe was born in 22nd May 1976. He received his M.Tech from ITM, Gwalior in 2006. He is currently working as Associate Professor in Electronics & Instrumentation Engineering Department of Institute of Technology & Management, Gwalior. Currently, He is pursuing his Ph.D from Thapar University, Patiala on the topic of Low Power Memory Cell Design. His research interests are VLSI Design, Low Power, VLSI signal processing, FPGA Design and Communication System