LECTOR: A Technique for Leakage Reduction in CMOS Circuits

Narender Hanchate, *Student Member, IEEE,* and Nagarajan Ranganathan, *Fellow, IEEE*

***Abstract*—In CMOS circuits, the reduction of the threshold voltage due to voltage scaling leads to increase in subthreshold leakage current and hence static power dissipation. We propose a novel technique called LECTOR for designing CMOS gates which significantly cuts down the leakage current without increasing the dynamic power dissipation. In the proposed technique, we introduce two leakage control transistors (a p-type and a n-type) within the logic gate for which the gate terminal of each leakage control transistor (LCT) is controlled by the source of the other. In this arrangement, one of the LCTs is always "*near its cutoff voltage*" for any input combination. This increases the resistance of the path from V_{dd} to ground, leading to significant decrease in leakage currents. The gate-level netlist of the given circuit is first converted into a static CMOS complex gate implementation and then LCTs are introduced to obtain a leakage-controlled circuit. The significant feature of LECTOR is that it works effectively in both active and idle states of the circuit, resulting in better leakage reduction compared to other techniques. Further, the proposed technique overcomes the limitations posed by other existing methods for leakage reduction. Experimental results indicate an average leakage reduction of 79.4% for MCNC'91 benchmark circuits.**

***Index Terms*—Deep submicron, leakage power, power optimization, transistor stacking.**

I. INTRODUCTION

POWER dissipation is an important consideration in the design of CMOS VLSI circuits. High power consumption leads to reduction in the battery life in the case of battery-powered applications and affects reliability, packaging, and cooling costs. The main sources for power dissipation are: 1) capacitive power dissipation due to the charging and discharging of the load capacitance; 2) short-circuit currents due to the existence of a conducting path between the voltage supply and ground for the brief period during which a logic gate makes a transition; and 3) leakage current. The leakage current consists of reverse-bias diode currents and subthreshold currents. The former is due to the stored charge between the drain and bulk of active transistors while the latter is due to the carrier diffusion between the source and drain of the OFF transistors.

The short-circuit power dissipation can be reduced to 10% of total power dissipation by designing the circuit to have equal input and output rise/fall edge times [1]. The power dissipation resulting from switching activity is the dominant component for technology processes with feature size larger than 1 μm. With technology processes maturing toward the deep-submicron regime, the feature sizes of the transistors are becoming smaller, thereby reducing the load capacitances. The reduction in feature size also forces a reduction in the supply voltage. The voltage scaling technique takes benefit of the quadratic dependence of switching power on supply voltage for dynamic power savings. However, this technique pays a penalty for the operation of the circuit by increasing the delay drastically as supply voltage approaches the threshold voltage V_t of the devices [2]. In order to facilitate voltage scaling without affecting the performance, threshold voltage has to be reduced. In general, the ratio between the supply voltage and the threshold voltage should be at least 5, so that the performance of CMOS circuits is not affected [3]. This also leads to better noise margins and helps to avoid the hot-carrier effects in short-channel devices [4].

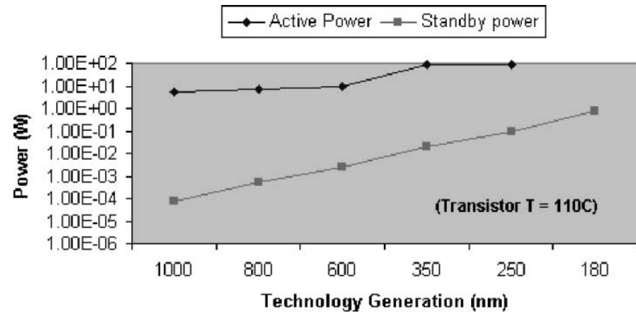


Fig. 1. Voltage supply and threshold voltage scaling trend.

Scaling down of threshold voltage V_t results in exponential increase of the subthreshold leakage current [5]. The supply voltage and threshold voltage scaling trends for Intel's microprocessor process technologies are discussed in [6]. It can be seen from Fig. 1 that the leakage power is only 0.01% of the active power for 1-μm technology, while it is 10% of the active power for 0.1-μm technology. There is a fivefold increase in leakage power as the technology process advances to a new generation. Projecting these trends, it can be seen that the leakage power dissipation will equal the active power dissipation within a few generations. Hence, efficient leakage power reduction methods are very critical for the deep-submicron and nanometer circuits.

In this paper, we describe a new leakage power reduction technique called LECTOR (LEakage Control TransistOR) for designing CMOS circuits. The rest of the paper is organized as follows. Section II describes briefly the prior works on leakage power reduction and their limitations. Section III introduces the transistor models used for estimating the leakage power. Our design strategy and an approach for minimizing the area overhead

Manuscript received December 11, 2002; revised August 20, 2003. This paper was recommended by Associate Editor Dr. R. Secareanu.

The authors are with the Nanomaterials and Nanomanufacturing Research Center, Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620 USA (e-mail: narender@csee.usf.edu; ranganat@csee.usf.edu).

Digital Object Identifier 10.1109/TVLSI.2003.821547

1063-8210/04$20.00 © 2004 IEEE

are described in Sections IV and V, respectively. Results are presented in Section VI, followed by conclusions in Section VII.

II. Related Work

Numerous methods for leakage power control have been reported in the literature. The work in [7] makes use of the dependence of the leakage current on the input vector to the gate. With additional control logic, the circuit is put into a low-leakage standby state when it is idle and restored to the original state when reactivated. Reactivation state forces the need to remember the original state information before going to low-leakage standby state. This requires special latches, thereby increasing the area of the circuit by about five times in the worst case [8]. Also, the amount of time for which the unit remains in idle state should be long enough so that the dynamic power consumed in forcing the circuit to low-leakage state and the leakage power dissipated in the standby state together is less than the leakage power without the technique.

Another technique for leakage power control is power gating, which turns off the devices by cutting off their supply voltage [9], [10]. This technique makes use of a bulky NMOS and/or PMOS device (sleep transistor) in the path between the supply voltage and ground. The sleep transistor is turned on when the circuit is active and turned off when the circuit is in idle state with the help of sleep signal. This creates virtual power and ground rails in the circuit. Hence, there is a significant detrimental effect on the switching speed when the circuit is active. The identification of the idle regions of the circuit and the generation of the sleep signal need additional hardware capable of predicting the circuit states accurately. This additional hardware consumes power throughout the circuit operation even when the circuit is in an idle state to continuously monitor the circuit state and control the sleep transistors.

The use of multiple threshold voltage CMOS (MTCMOS) technology for leakage control is described in [11] and [12]. The transistors of the gates are at low threshold voltage and the ground is connected to the gate through a high-threshold voltage NMOS gating transistor. The logical function of a gating transistor is similar to that of a sleep transistor. The existence of reverse conduction paths tend to reduce the noise margin or in the worst case may result in complete failure of the gate [4]. Moreover, there is a performance penalty since high-threshold transistors appear in series with all the switching current paths. A variation of MTCMOS technique is the Dual V_t technique, which uses transistors with two different threshold voltages. Low-threshold transistors are used for the gates on the critical path and high-threshold transistors are used for those not in the critical path [4], [13], [14]. In both MTCMOS and Dual V_t methods, additional mask layers for each value of threshold voltage are required for fabricating the transistors selectively according to their assigned threshold voltage values. This makes the fabrication process complex.

In addition to these limitations, the techniques discussed above suffer from turning-on latency, that is, when the idle subsections of the circuit are reactivated, they cannot be used immediately because some time is needed before the subcircuit returns to its normal operating condition. The latency for power gating is typically a few cycles, and for Dual V_t technology, is much higher [15]. Also, these techniques are not effective in controlling the leakage power when the circuit is in active state.

In [16], the authors use the concept of forced stacks for leakage control. Forced stacking introduces an additional transistor for every input of the gate in both N- and P-networks. This ensures that two transistors are OFF instead of one for every OFF-input of the gate and hence makes a significant savings on the leakage current. However, the loading requirements for each input introduced by the forced stacking reduces the drive current of the gate significantly. This results in a detrimental impact on the speed of the circuit.

In [10], a blend of sleep transistors and the stacking effects are used to reduce leakage power. This method identifies a circuit input vector for which the leakage current of the circuit is the lowest possible. The sleep signal controlled transistors are inserted away from the critical path where only one transistor is OFF when low-leakage input vector is applied to the circuit. Hence, this technique is input-vector dependent. Moreover, as this technique uses sleep transistors, it needs additional hardware for controlling them. This additional hardware consumes power in both idle and active states of the circuit.

In this work, we develop a new technique for leakage control in CMOS circuits. The proposed technique avoids the problems associated with the techniques discussed above.

III. Preliminaries

In this section, we briefly describe the models used in this work for estimating power dissipation for short-channel MOSFETs. The leakage current calculation is not straightforward due to the highly nonlinear behavior of the drain current of the device with respect to source/drain voltages. We have used the Berkeley Short-Channel IGFET (BSIM) Predictive Technology Model to estimate the leakage power dissipation, as it fits well with HSPICE simulations [5]. In the BSIM model, the threshold voltage V_t is expressed as

$$V_t = V_{FB} + \phi_s + k_1\sqrt{\phi_s} - k_2\phi_s - \eta V_{dd} \qquad (1)$$

where V_{FB} is the flatband voltage, ϕ_s is twice the Fermi potential, k_1 and k_2 represent the nonuniform doping effect, and η models the drain-induced barrier lowering (DIBL) effect, an undesirable punch-through current flowing between the source and drain below the surface of the channel. The leakage current for NMOS transistors operating in weak-inversion region (i.e., $V_{gs} = 0$) is given by

$$I_s = I_0 \exp\left(\frac{(V_{gs} - V_t)}{nV_T}\right)\left(1 - \exp\left(-\frac{V_{ds}}{V_T}\right)\right) \qquad (2)$$

where V_T is the thermal voltage and is given by q/kT, n is the subthreshold slope coefficient, and $I_0 = \mu_0 C_{\text{ox}}(W_{\text{eff}}/L_{\text{eff}})V_t^2 e^{1.8}$. Equation (2) gives a simple method for estimating the leakage current in a single NMOS transistor. A similar expression for the leakage current in a single PMOS transistor can be obtained.

In most CMOS logic design styles, the gates consist of series-parallel networks of PMOS and NMOS transistors. The

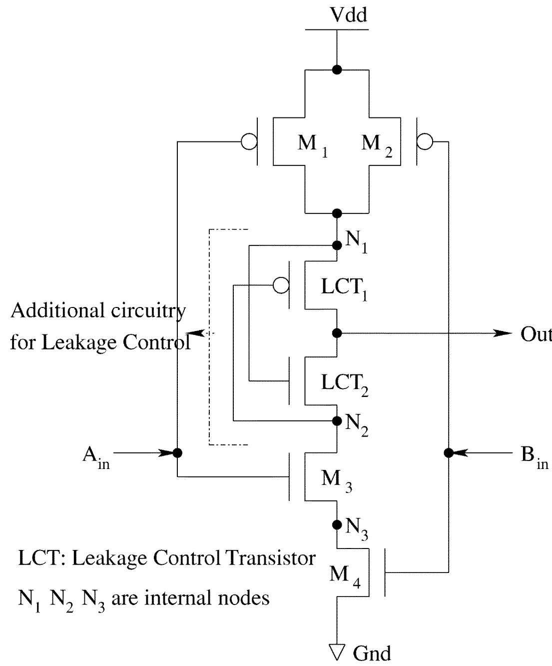
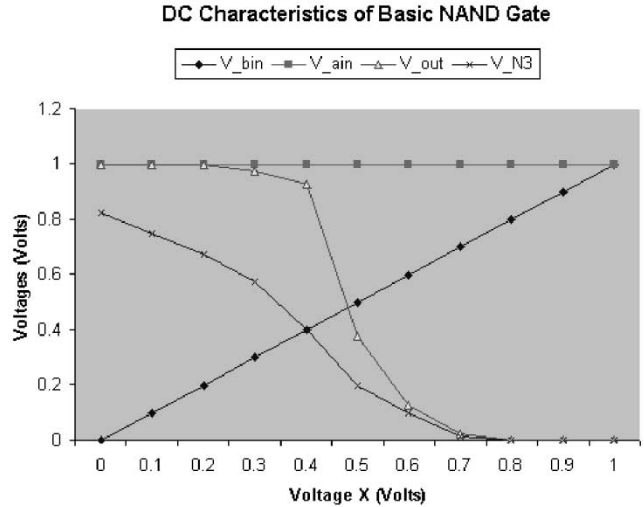
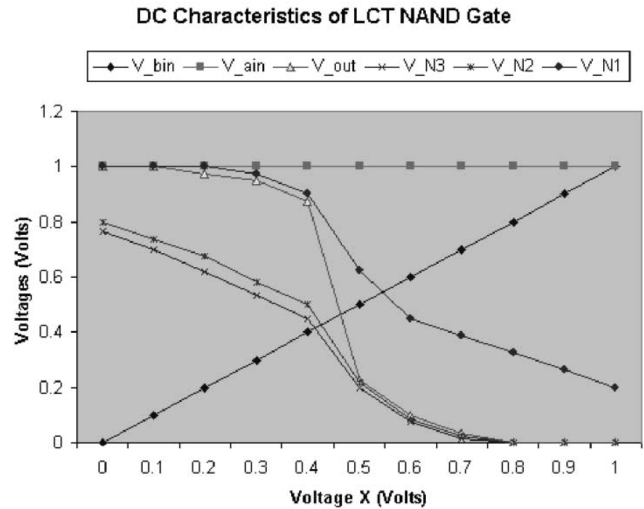


Fig. 2. Two input LCT NAND gates. The leakage control transistors LCT_1 and LCT_2 are inserted between nodes N_1 and N_2 and they act as self-controlled stacked transistors.



(a)



(b)

Fig. 3. DC characteristics of two-input NAND gate. (a) Characteristics of basic NAND gate. (b) Characteristics of LCT NAND gate. It can be observed that the output voltage variation is similar in both cases.

leakage current contributed by the MOS transistors connected in parallel is the sum of the currents through the individual transistors. However, in the case of transistors connected in series, the evaluation of leakage current becomes complex due to its nonlinear characteristics. The analysis of the current through a stack of transistors when they operate in subthreshold region is as given in [3]:

$$I_{s1}{:}I_{s2}{:}I_{s3} = 1.8\exp\left(\frac{\eta V_{dd}}{nV_t}\right){:}1.8{:}1 \tag{3}$$

where $I_{s1}(i = 1, 2, 3)$ is the leakage current for i stacked MOS transistors and I_{s1} is given by (2). Equation (3) shows that the leakage current of the MOS network can be expressed as a function of a single MOS transistor. If the number of stacked MOS transistors is more than three, the leakage current is very small and can be neglected.

IV. LECTOR TECHNIQUE

The basic idea behind our approach for reduction of leakage power is the effective stacking of transistors in the path from supply voltage to ground. This is based on the observation made in [10], [16], [17] that "a state with more than one transistor OFF in a path from supply voltage to ground is far less leaky than a state with only one transistor OFF in any supply to ground path." In our method, we introduce two leakage control transistors (LCTs) in each CMOS gate such that one of the LCTs is *near its cutoff* region of operation. We illustrate our LEakage Control TransistOR technique (LECTOR) with the case of a NAND gate. A CMOS NAND gate with the addition of two leakage control transistors is shown in Fig. 2 (we later refer to it as the LCT NAND gate).

Two leakage control transistors LCT_1 (PMOS) and LCT_2 (NMOS) are introduced between the nodes N_1 and N_2 of the pull-up and pull-down logic of the NAND gate. The drain nodes of the transistors LCT_1 and LCT_2 are connected together to form the output node of the NAND gate. The source nodes of the transistors are connected to nodes N_1 and N_2 of pull-up and pull-down logic, respectively. The switching of transistors LCT_1 and LCT_2 are controlled by the voltage potentials at nodes N_2 and N_1 respectively. This wiring configuration ensures that one of the LCTs is always *near its cutoff region*, irrespective of the input vector applied to the NAND gate. This can be seen from the dc characteristics shown in Fig. 3, obtained from HSPICE simulations. Fig. 3(a) shows the dc characteristics of the unmodified NAND gate and Fig. 3(b) shows that of the LCT NAND gate, when the input A_{in} is fixed at 1 V and B_{in} is varied from 0 to 1 V.

Consider the dc characteristics of the LCT NAND gate. When $A_{\text{in}} = 1$ V and $B_{\text{in}} = 0$ V, the voltage at the node N_2 is 800 mV. This voltage is not sufficient to turn LCT_1 completely to

TABLE I
STATE MATRIX OF TWO-INPUT LCT NAND GATE

Transistor Reference	Input Vector - (A_{in}, B_{in})			
	(0,0)	(0,1)	(1,0)	(1,1)
M_1	On state	On state	Off state	Off state
M_2	On state	Off state	On state	Off state
LCT_1	Near Cut-Off state	Near Cut-Off state	Near Cut-Off state	On state
LCT_2	On state	On state	On state	Near Cut-Off state
M_3	Off state	Off state	On state	On state
M_4	Off state	On state	Off state	On state

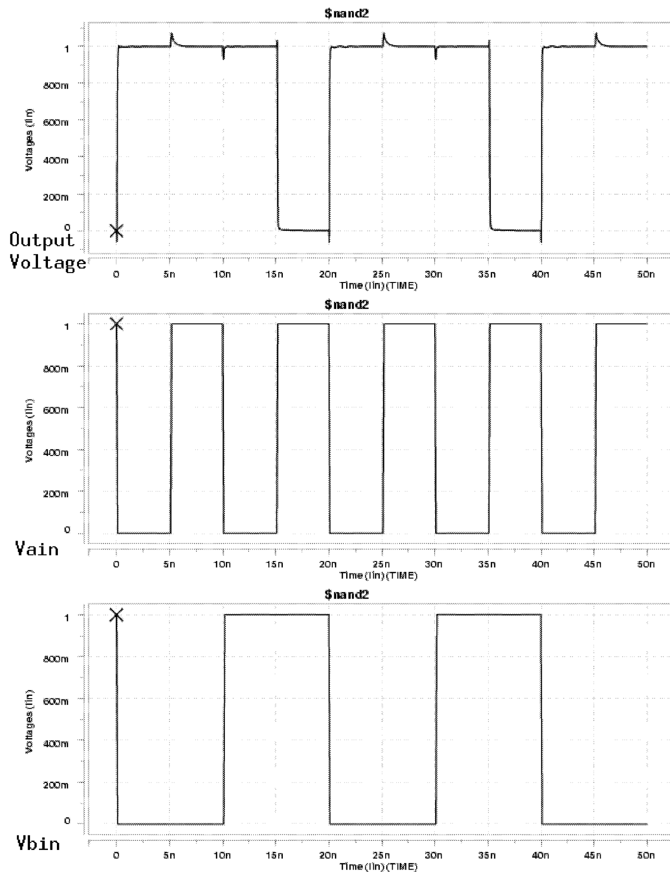


Fig. 4. Transient characteristics of two-input LCT NAND gate using HSPICE. (The x axis shows simulation time in nanoseconds and the y axis shows the voltage levels in millivolts).

OFF state. Hence, the resistance of LCT_1 will be lesser than its OFF resistance, allowing conduction. Even though the resistance of LCT_1 is not as high as its OFF state resistance, it increases the resistance of V_{dd} to ground path, controlling the flow of leakage currents, resulting in leakage power reduction. Similarly, when $A_{\text{in}} = 1$ V and $B_{\text{in}} = 1$ V, the voltage of node N_1 is 200 mV, operating the transistor LCT_2 near its *cutoff* region. The state of the transistors for all possible combinations of input vectors for the LCT NAND gate are tabulated in Table I. Thus, the introduction of LCTs increases the resistance of the path from V_{dd} to ground. This also increases the propagation delay of the gate. To reduce this hostile effect, the transistors of LCT gate are sized such that the propagation delay is equal to its conventional counter part.

TABLE II
THRESHOLD VOLTAGES OF MOS MODELS USED

Technology Process	Threshold Voltage of NMOS (V_{tn})	Threshold Voltage of PMOS (V_{tp})
70nm	0.2V	-0.22V
100nm	0.2607V	-0.303V
130nm	0.332V	-0.3499V
180nm	0.3999V	-0.42V

Fig. 4 shows the transient curves obtained by HSPICE at various nodes of the LCT NAND gate simulated for 70-nm technology at 1-V supply voltage. It can be observed from the curves that the LCT NAND produces exact output logic levels. We have also tested the LECTOR with process variations of 180, 130, and 100 nm technologies at 1-V supply voltage. The threshold voltages used by these BSIM predictive models are tabulated in Table II. We have observed that LECTOR produces slightly weak logic levels (around 960 mV of logic high and 35 mV of logic zero) for 180-nm and 130-nm process technologies. But it produces exact logic levels for 70-nm and 100-nm process variations. The reason for this behavior of LECTOR could be the difference between the supply voltage and the threshold voltage of a process variation. When we increase the supply voltage to 1.8 V, LECTOR produces exact logic levels for all the four process variations.

The transistors of the LCT NAND gate are sized to study its effect on the propagation delay of the gate. Y-LCT gates are obtained by sizing the widths of only the leakage control transistors LCT_1 and LCT_2 to Y times the width of other PMOS and NMOS transistors of the gate, respectively. Also, we have tried to adjust the widths of all the transistors of the LCT NAND gate such that its propagation delay is almost equal to that of a conventional NAND gate. (We refer to this gate as Iso-LCT NAND gate.) The widths of PMOS and NMOS transistors, except LCT transistors, used in simulation of Y-LCT gates are set to two and three times the minimum feature width of the respective process variation.

The leakage power dissipation and propagation delay values for two-input NAND gate using 100-nm and 70-nm technology processes are tabulated in Table III. It can be observed from the Tables II and III that as the threshold voltage decreases, which is the case as we move toward deep submicron, our technique produces better leakage power reductions. Hence, the use

TABLE III
LEAKAGE POWER FOR TWO-INPUT NAND GATE

100nm Process Technology, Supply Voltage = 1V						
NAND gate type	Leakage Power Dissipation in Watts for Input Vector (0,0)	(0,1)	(1,0)	(1,1)	Delay in (ps)	% Average Savings
Conventional	1.228e-10	9.117e-10	5.356e-10	2.241e-09	13.53	-
LCT	1.180e-10	5.542e-10	4.477e-10	1.539e-09	18.79	30.20%
1.5-LCT	1.182e-10	5.788e-10	4.533e-10	1.593e-09	17.12	28.01%
2-LCT	1.194e-10	5.985e-10	4.606e-10	1.621e-09	15.64	26.54%
2.5-LCT	1.197e-10	6.114e-10	4.678e-10	1.651e-09	15.39	25.22%
Iso-LCT	1.224e-10	6.394e-10	4.872e-10	1.704e-09	13.59	22.51%
	70nm Process Technology, Supply Voltage = 1V					
Conventional	6.450e-10	5.600e-09	3.817e-09	1.091e-08	15.16	-
LCT	6.065e-10	3.808e-09	3.622e-09	5.564e-09	21.40	35.12%
1.5-LCT	6.108e-10	3.900e-09	3.650e-09	5.742e-09	19.09	33.70%
2-LCT	6.146e-10	3.961e-09	3.667e-09	5.872e-09	18.36	32.71%
2.5-LCT	6.189e-10	4.016e-09	3.675e-09	6.060e-09	17.99	31.48%
Iso-LCT	6.314e-10	4.125e-09	3.689e-09	6.196e-09	15.23	30.18%

Note: Y-LCT = LCT transistors are Y times the widths of their respective type.

of LECTOR enhances further reduction of the supply voltage along with the threshold voltage, thereby favoring technology advancements.

In the technique discussed in [10], the sleep transistors have to be able to isolate the power supply and/or ground from the rest of the transistors of the gate. Hence, they need to be made bulkier dissipating more dynamic power. This offsets the savings yielded when the circuit is idle. Moreover, this technique is input-vector dependent and needs additional circuitry to monitor and control the switching of sleep transistors. Thus, it consumes power in both active and idle states. In comparison, LECTOR is vector independent and the required control signals are generated within the gate. Forced stacks [16] have 100% area overhead. In comparison, LECTOR requires exactly two additional transistors for every path from supply voltage to ground irrespective of the logic function realized by the gate. The loading requirements of the gate with forced stacks are huge and depend on the number of additional transistors added. In comparison, the loading requirements with LCTs is much lower and is a constant. Hence, the performance degradation is insignificant in the case of LECTOR, and we overcome the major drawback faced by forced stack technique.

Table IV shows the effect of noise on various LCT gates and its corresponding conventional gates. Noise analysis was performed by creating a SPICE deck of the circuits. The experiment setup is shown in Fig. 5 and simulating using HSPICE. Our experiments are based on 70-nm process parameters and 25°C temperature. A switching waveform is applied at node A to generate a noise waveform on net N. The waveform on net N is denoted as $X(t)$, response on the fanout of net N as $Y(t)$ and the transient sensitivity as $S(t)$ [Fig. 5(a)]. If the waveform of the net N is slightly changed by a constant voltage D, i.e., $X_1(t) = X(t) + D$, then the response of the fanout of net N is $Y_1(t)$ [Fig. 5(b)]. The transient sensitivity $S(t)$ is defined as

TABLE IV
NOISE ANALYSIS OF VARIOUS GATES

CMOS Gate	Conventional Gate Max Noise	Conventional Gate Sensitivity	LCT Gate Max Noise	LCT Gate Sensitivity
Inverter	0.3651	-0.2744	0.4008	-0.1517
NAND2	0.2786	-0.1986	0.3427	-0.2025
NOR2	0.4129	-0.2241	0.4830	-0.2543
AOI22	0.3560	-0.2840	0.3829	-0.3642
OAI22	0.3658	-0.2864	0.4122	-0.2938

the ratio of the change in the output voltage to the change in the input voltage when the input is changed by a very small dc offset, given by (4)

$$S(t) = \lim_{D\to 0} \frac{Y_1(t) - Y(t)}{X_1(t) - X(t)} = \lim_{D\to 0} \frac{Y_1(t) - Y(t)}{D}. \quad (4)$$

The sensitivity reported in Table IV is the maximum value of the transient sensitivity. The sensitivity gives a measure of how the fanout of net N reacts to the noise waveform on net N and indicates whether the noise is amplified or attenuated. If sensitivity of a net is greater than one, then the net is said to have a noise failure. It can be seen from Table IV that the LCT gates show stability to noise injections and the noise sensitivity values are very close to that of conventional gates. This is important to show that proposed LECTOR technique is stable to noise fluctuations.

V. REDUCING AREA OVERHEAD IN IMPLEMENTATION OF LECTOR

A method to control the leakage power in a two-input NAND gate was described above. If we extend this idea to all the basic

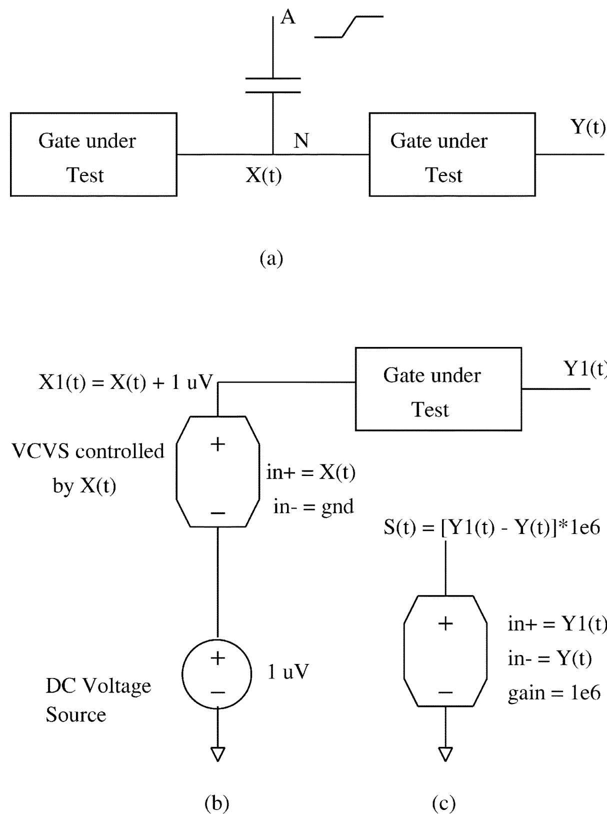


Fig. 5. Setup used for performing the noise analysis using HSPICE.

TABLE V
AREA OVERHEAD FOR A LCT GATE

CMOS Gate Type	Number of Transistors in UnModified Gate	LCT Gate
2-input NAND	4	6
2-input NOR	4	6
2-input AND	6	10
2-input OR	6	10
3-input NAND	6	8
3-input NOR	6	8
3-input AND	8	12
3-input OR	8	12

logic gates, then the area overhead on the entire circuit will increase up to an upper bound of 60% depending on the logic gates used. The area overhead on various logic gates implemented using LCTs is tabulated in Table V. Here, we present a generalized approach which reduces the area overhead of the overall circuit along with significant leakage reduction.

Consider a four-input AND-OR-inverter (AOI22) gate. Fig. 6 shows two different implementations of the gate. The first implementation uses two two-input AND gates, one two-input OR gate, and an inverter, requiring a total of 20 transistors for its realization. A second implementation shows an equivalent compound/complex gate realization. A compound gate is formed by the combination of series and parallel MOS structures with complementary pull-up and pull-down logic [18]. The required series and parallel combinations of transistors are generated by analyzing the relevant Karnaugh map for both n- and p-logic structures and using DeMorgan's theorems. As these gates are built on static CMOS technology, they are called static CMOS complex gates (SCCG) [18]. SCCG implementation of an AOI22 gate needs eight transistors and hence is an efficient method for its realization. We have chosen to use SCCG gates for LECTOR as it not only saves area but also is well suited for inserting leakage control transistors. This is because SCCG gates have exactly one common node which connects the pull-up and pull-down circuits of the gate. Hence, in this gate, all the paths from the supply voltage to ground would pass through this node. Inserting LCTs at this node would control the leakage currents flowing through all the paths in the gate, thereby saving significantly.

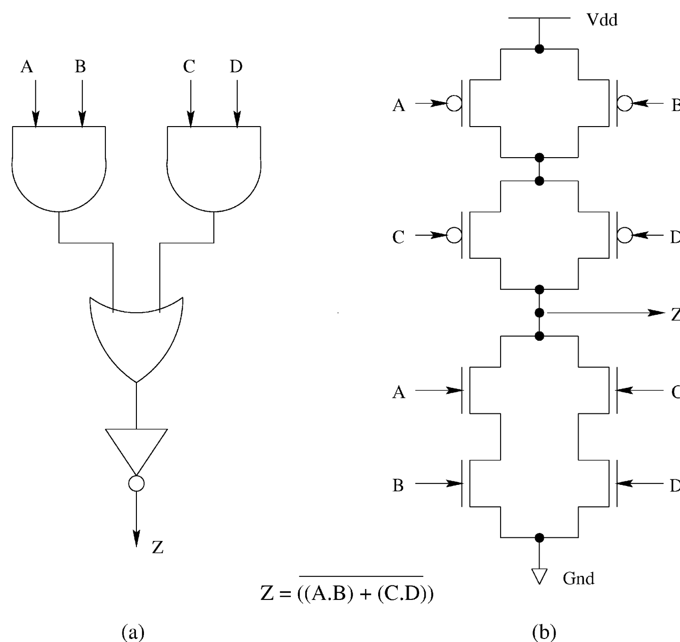


Fig. 6. Two implementations of AND-OR-inverter gate.

The limitation with SCCG gates is that if the number of transistors in series exceeds an upper limit in any path of pull-up or pull-down logic, then there is a hostile effect on the propagation delay of the gate. Typically, this upper bound can be safely fixed to three or even four transistors. The leakage power in the AOI22 gate can be controlled by adding the two LCTs between the pull-up and pull-down logic of the implementation shown in the Fig. 6(b). Fig. 7 shows the general scheme for converting any SCCG gate to leakage-controlled gate.

Any CMOS logic circuit can be expressed in terms of its gate-level netlist and hence is the starting point of our design methodology. Fig. 8 describes the sequence of steps for obtaining a leakage-controlled circuit from the gate-level netlist. The gate-level netlist is fed as input to a script, which generates the corresponding Berkeley Library Interchange Format (BLIF). The Sequential and Combinational circuit Synthesis System (SIS tool) [19] or the TRAnsistor Binding Using COmplex gates (TRABUCO) [20] tool can be used for performing technology mapping of the BLIF netlist. Technology mapping comprises

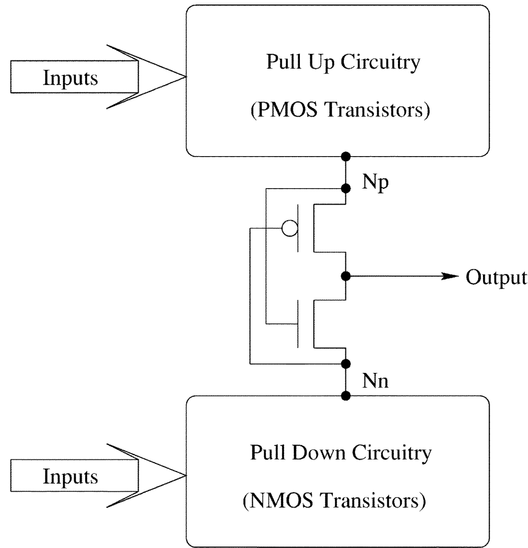


Fig. 7. Generalized structure for leakage-controlled gates.

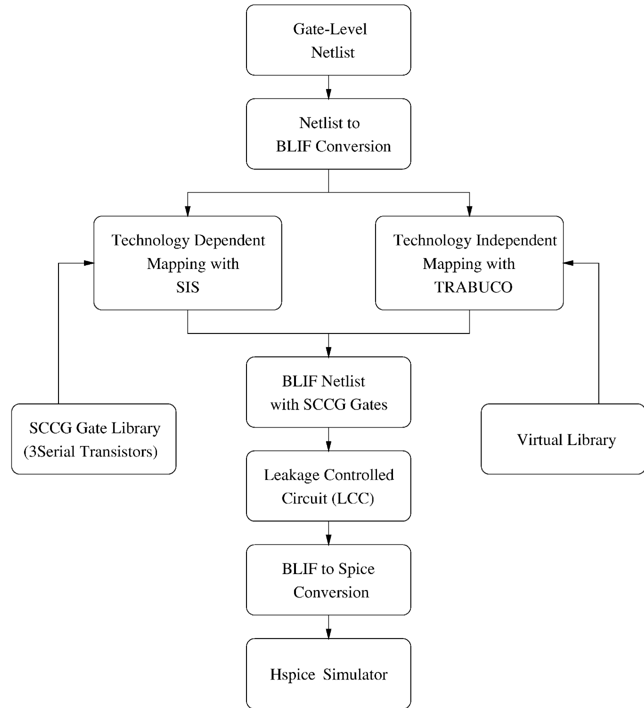


Fig. 8. General design flow.

of covering phase and matching phase. SIS is a technology-dependent mapping tool and hence uses a standard set of gates predefined in its library for performing technology mapping.

A library of SCCG gates with a maximum of two series transistors each was created. The constraint on the maximum number was induced in the topology for performance reasons. The covering phase in the SIS tool is performed by representing the netlist in the form of a tree called the parent tree. The matching is done by dividing the parent tree into subtrees. Each subtree is then replaced with the logic gate in the library whose tree representation matches efficiently [19].

TRABUCO, on other hand, is a technology-independent mapping tool, which maps the netlist on to a virtual library of SCCG gates. The tool creates virtual library of SCCG gates dynamically and hence does not need any library support. The set of SCCG gates used by the tool depends on the number of serial transistors input by the user. The covering scheme used by this tool is based on the dynamic tree covering approach similar to the 0-1-knapsack problem [20]. This approach generates a minimal number of multi-input SCCGs. The circuit with SCCG gates obtained by SIS or TRABUCO tool from the input netlist is then converted to a LECTOR-style circuit by introducing LCTs as shown in Fig. 7. The LECTOR circuit is then converted to SPICE format with the help of a script for simulation with HSPICE simulator.

VI. Experimental Results

The LECTOR technique was implemented and tested on MCNC'91 benchmark circuits. First, the MCNC'91 benchmark netlists were converted to BLIF using a script. SIS tool takes the BLIF input and performs technology mapping to SCCG gates. SIS tool provides various scripts, which aid in mapping the BLIF to the needs of the user. *Read_blif* script was used for reading the BLIF netlist by the SIS tool. *Read_library* script was invoked to read library of gates from a file in *genlib* format. SIS tool performs the covering phase of technology mapping by tree-covering algorithm and the logical decomposition of gates using the *tech_decomp* script. The matching phase of the technology mapping was completed with the map script optimized for area.

The number of transistors in the path from V_{dd} to ground influences the operation of LCT gates. As the number of transistors in the path increases, the body effect becomes prominent, resulting in incorrect logic switching. So we have restricted the number of transistors on any path from V_{dd} to ground of LCT gates to six (three PMOS and three NMOS including the LCTs). Hence, SCCG library with gates constrained to two series transistors was used for generating LECTOR-style circuits. The eight different SCCG gates which can be realized with two series transistors are inverter, buffer, two-input NAND and NOR gates, two-input AOI and OAI gates (AOI22, AOI21, OAI22, OAI21). After mapping the BLIF to the SCCG library, these gates are replaced with the leakage-controlled gates, designed by the adding two LCTs to SCCG gates, to obtain a LCT MCNC Circuit. The addition of LCTs would make the number of series transistors in LCT gates to three.

In order to make a fair comparison with the circuit implemented without LECTOR, we created a library of gates containing at most three series transistors. This library contains an inverter, buffer, two- and three-input NAND, NOR, OR, AND gates, and two-input XOR gate. SIS tool is used to map the BLIF netlist on to this library to obtain an unmodified MCNC circuit (we refer to this as U-MC circuit later). We have analyzed the critical path delays of both LCT MCNC (referred to as LCT-MC) circuit and the corresponding U-MC circuit using HSPICE. The transistors of LCT gates present on the critical path of the LCT-MC circuit are sized such that the critical path delay is kept almost equal to that of the U-MC circuit. This helps to make a fair and direct comparison of LECTOR and U-MC circuits with respect to power dissipation. The HSPICE simulator was preferred for measuring the leakage power due to its accuracy. Simulations

TABLE VI
EXPERIMENTAL RESULTS FOR MCNC '91 BENCHMARKS

MCNC '91 Circuit	Leak Pwr (μW) in μW		Dyn. Pwr (in μW) in μW)		Total Pwr (μW) in μW		Normalized area	% Leakage savings by	
	U-MC	LCT-MC	U-MC	LCT-MC	U-MC	LCT-MC		Stacks [10] with 16% delay overhead	LECTOR with no delay overhead
I1	1.159	0.156	1.122	1.094	1.275	0.291	1.21	67.34%	86.54%
I2	2.305	0.735	2.420	2.661	2.375	0.807	1.14	61.61%	68.11%
I3	1.383	0.419	1.063	1.025	1.451	0.712	1.18	61.97%	69.70%
I4	2.356	0.632	2.043	1.995	2.436	0.756	1.12	55.03%	73.17%
I5	4.625	0.475	5.905	5.945	5.473	0.732	1.19	83.51%	89.72%
I6	6.906	1.912	18.62	18.85	12.70	6.495	1.13	60.74%	72.31%
I7	8.933	3.126	25.62	25.59	17.95	9.248	1.12	51.39%	65.01%
I8	30.05	5.038	59.06	58.34	41.08	10.56	1.08	74.32%	83.23%
I9	21.90	2.897	38.79	37.24	25.75	8.982	1.12	85.15%	86.77%
I10	40.47	5.842	54.58	52.13	43.31	11.23	1.15	83.39%	85.56%
								Dual V_T [14] with upto 64% delay overhead	LECTOR with no delay overhead
C432	1.395	0.672	6.970	6.889	1.923	1.260	1.17	28.83%	51.82%
C499	3.469	1.444	16.38	16.03	5.037	2.990	1.15	22.96%	58.37%
C880	6.141	1.154	9.472	9.502	7.972	2.493	1.18	82.67%	81.21%
C1355	8.089	1.672	16.08	15.93	9.754	3.104	1.11	21.50%	79.32%
C1908	19.61	1.926	28.43	27.10	24.53	3.613	1.13	84.92%	90.18%
C2670	52.17	2.845	70.45	69.59	65.70	4.338	1.19	90.25%	94.54%
C3540	64.79	3.852	81.65	76.49	73.41	4.588	1.14	83.36%	94.05%
C5315	82.58	4.826	99.21	97.60	88.50	5.892	1.12	91.56%	94.16%
C6288	163.7	9.725	520.4	502.2	390.2	30.23	1.10	61.75%	94.05%
C7552	323.2	10.24	576.2	566.6	393.3	22.96	1.08	90.90%	96.86%

The critical path delays of U-MC and LCT-MC circuits are made equal by sizing LCT gates on critical path. The leakage power reduction results indicated for techniques [10] and [14] have high delay and area penalties whereas those indicated for LECTOR are with zero delay penalty and much lesser area penalty. Normalized area is the ratio of layout areas of LCT-MC and U-MC circuits.

were performed assuming 70-nm fabrication process parameters generated using the BSIM predictive models with temperature set to 25 °C.

Leakage power dissipation was measured by exciting both circuits with the same set of randomly generated input vectors. After applying one input vector, the circuit is made to wait long enough before exciting it with the next input vector. This allows the circuit's switching activity to die down, after which the power dissipated is due to leakage currents. We measured the power during this time period and averaged it over all the input vectors to obtain the average leakage power dissipations for each circuit. We also measured the average total power dissipated in each case. To measure the dynamic power dissipation, we have driven both the circuits with same excitation such that the switching activity is very high. The input vectors are fed at a faster rate not allowing the circuit to settle down. Hence, in this case, leakage power dissipation will be minimal.

The experimental results are listed in Table VI. The first column lists the MCNC'91 benchmark circuits. The average leakage power dissipation (in microwatts) for U-MC and LCT-MC circuits are given in columns two and three, respectively. Columns six and seven provide the average total power dissipation (in microwatts) for the respective circuits. Both the leakage and total power estimates were obtained by averaging the circuit simulations over ten different sets of 500 randomly generated (with a probability of 0.5) input patterns. The input patterns were triggered with long time intervals between them so as to have minimum circuit switching activity. Thus, it was ensured that the major contributor to the power dissipation is leakage power.

The dynamic power estimates (in microwatts) obtained for U-MC and LCT-MC circuits are indicated in columns four and five, respectively. Ten different sets of 1000 randomly (with a probability of 0.5) generated input patterns are used to estimate dynamic power in both cases. The inputs were applied in quick succession so as to maintain high switching activity in the circuit. Hence the major contributor to the power dissipation in this case is dynamic power. The normalized area of LCT-MC circuits with respect to its corresponding U-MC circuits is given in column eight. We created standard cell layouts of all the gates

used in synthesis of both LCT-MC and U-MC circuits. Cadence Silicon Ensemble was used for placement and routing of the netlists using the standard cell layouts. The layout area required by the circuits were measured to calculate the area overheads of the LCT-MC circuits with respect to U-MC circuits.

The results show an average reduction of 79.4% in average leakage power dissipation with an average area overhead of 14%. It should be noted that the leakage savings obtained is without any significant sacrifice in dynamic power and with zero delay penalty. Columns nine and ten of Table VI compares the leakage savings obtained for MCNC'91 benchmarks using the LECTOR and the techniques from [10] and [14]. We have chosen the technique based on transistor stacks [10] as it works on the same basic idea as LECTOR. The dual threshold voltage technique combined with delay balancing and retiming, proposed in [14], is chosen for comparison because of its significant results in dual threshold voltage technique category. For circuits C432–C7552, we compare our results with dual threshold voltage technique [14] and for circuits I1–I10, we compare with transistor stacks [10]. It should be noted that leakage reduction results shown in [10] suffer from delay and area penalties. The leakage control transistors in transistor stacks technique are sized to 30% of total sizing required to achieve zero delay penalty, and hence results in average delay overhead of 16%. The area overhead due to additional transistors and control circuitry is about 46.3% to 50%. Also, the savings indicated for transistor stacks technique are obtained on application of the *low-leakage input vectors* and neglect the dynamic power dissipations due to control circuitry. In general, the dynamic power dissipation due to control circuitry is very significant and may overshadow the savings in leakage power depending on the switching activity of control logic. In the case of Dual V_T technique [14], the leakage savings given are with respect to benchmark circuits without performing technology mapping. When technology mapping is performed before calculating the leakage reduction, the leakage savings will be significantly less for these circuits than that shown in Table VI. Also, the results given are based on the maximum allowable delays for the gates with the threshold voltage scaled to as high as 0.5 V_{dd}. This would result in delay penalty of anywhere between 5% to 64%, as pointed out in [10]. In spite of these dissimilarities in calculations of leakage power savings, we can observe from Table VI that LECTOR provides significantly better results.

VII. Conclusions

The scaling down of device dimensions, supply voltage, and threshold voltage for achieving high performance and low dynamic power dissipation has largely contributed to the increase in leakage power dissipation. With deep-submicron and nanometer technologies, the leakage current becomes more critical in portable systems where battery life is of prime concern. We have presented an efficient design methodology for reducing the leakage power in CMOS circuits. LECTOR yields better leakage reduction as the threshold voltage decreases and hence aids in further reduction of supply voltage and minimization of transistor sizes. Unlike other leakage control techniques, LECTOR does not need any control circuitry to monitor the states of the circuit. Hence, LECTOR avoids the sacrifice of obtained leakage power reduction in the form of dynamic power consumed by the additional circuitry to control the overall circuit states. Experimental results show that our technique yields average leakage reduction of about 79.4% for the same critical path delay in comparison to the unmodified circuit for MCNC'91 benchmarks with an area overhead of 14%.

Acknowledgment

The authors wish to thank the anonymous reviewers for all their comments that helped in improving the quality of the manuscript.

References

[1] H. J. M Veendrick, "Short circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits ," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 468–473, Aug. 1984.

[2] A. P. Chandrakasan and R. W. Brodersen, "Minimizing power con-sumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, pp. 498–523, Apr. 1995.

[3] R. X. Gu and M. I. Elmasry, "Power dissipation analysis and op-timization for deep submicron CMOS digital circuits," *IEEE J. Solid-State Circuits*, vol. 31, pp. 707–713, May 1999.

[4] Q. Wang and S. Vrudhula, "Static power optimization of deep sub-micron CMOS circuits for dual V_T technology," in *Proc. ICCAD*, Apr. 1998, pp. 490–496.

[5] B. J. Sheu, D. L. Scharfetter, P. K. Ko, and M. C. Jeng, "BSIM: Berkeleyshort-channel IGFET model for MOS transistors," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 558–566, Aug. 1987.

[6] S. Thompson, P. Packan, and M. Bohr, "MOS scaling: Transistor challenges for the 21st century," *Intel Technol. J.*, vol. Q3, 1998.

[7] J. P. Halter and F. Najm, "A gate-level leakage power reduction method for ultralow-power CMOS circuits," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1997, pp. 475–478.

[8] D. Duarte, Y.-F. Tsai, N. Vijay Krishnan, and M. J. Irwin, "Evaluating run-time techniques for leakage power reduction," in *7th Proc. ASP-DAC*, 2002, pp. 31–38.

[9] M. D. 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," in *Proc. IEEE ISLPED*, 2000, pp. 90–95.

[10] M. C. Johnson, D. Somasekhar, L. Y. Chiou, and K. Roy, "Leakage control with efficient use of transistor stacks in single threshold CMOS," *IEEE Trans. VLSI Syst.*, vol. 10, pp. 1–5, Feb. 2002.

[11] J. Kao, A. Chandrakasan, and D. Antoniadis, "Transistor sizing issues and tool for multi-threshold CMOS technology," in *Proc. 34th DAC*, 1997, pp. 409–414.

[12] C. Gopalakrishnan and S. Katkoori, "Resource allocation and binding approach for low leakage power," in *Proc. IEEE Int. Conf. VLSI Design*, Jan. 2003, pp. 297–302.

[13] L. Wei, Z. Chen, M. Johnson, and K. Roy, "Design and optmization of low voltage high performance dual threshold CMOS circuits," in *Proc. 35th DAC*, 1998, pp. 489–492.

[14] V. Sundarajan and K. K. Parhi, "Low power synthesis of dual threshold voltage CMOS VLSI circuits," *Proc. IEEE ISLPED*, pp. 139–144, 1999.

[15] S. Rele, S. Pande, S. Onder, and R. Gupta, "Optimizing static power dissipation by funtional units in superscalar processors," in *Proc. Int. Conf. Compiler Construction*, Apr. 2002, pp. 261–275.

[16] S. Narendra, S. Borkar, V. De, D. Antoniadis, and A. P. Chandrakasan, "Scaling of stack effect and its application for leakage reduction," *Proc. IEEE ISLPLED*, pp. 195–200, Aug. 2001.

[17] S. Sirichotiyakul, T. Edwards, C. Oh, R. Panda, and D. Blaauw, "Duet: An accurate leakage estimation and optimization tool for dual-V_t circuits," *IEEE Trans. VLSI Syst.*, vol. 10, pp. 79–90, Apr. 2002.

[18] N. H. E. Weste and K. Eshraghian, Eds., *Principles of CMOS VLSI Design: A Systems Prespective*, 2 ed. Reading, MA: Addison-Wesley, 1993.

[19] E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. R. Stephan, R. K. Brayton, and A.-S. Vincentelli, Eds., *SIS: A System for Sequential Circuit Synthesis*. Berkeley, CA: Univ. of California Press, 1992.

[20] A. I. Reis, "Covering strategies for library free technology mapping," in *Proc. XII Integrated Circuits and Systems Design*, 1999, pp. 180–183.

Narender Hanchate (S'01) received the B.E. degree in electronics and communication engineering from Osmania University, Hyderabad, India, in 2000 and the M.S. degree in computer science and engineering from the University of South Florida, Tampa, in 2003, where he is currently working toward the Ph.D. degree in computer science and engineering.

From 2000 to 2001, he was a Design Engineer in the field of VLSI and Embedded systems with HCL Technologies Ltd., Chennai, India. His research interests include the development of methodologies for low-power synthesis, interconnect optimization, and crosstalk analysis of deep-submicron circuits.

Mr. Hanchate is a student member of the IEEE Computer Society.

Nagarajan Ranganathan (M'88–SM'92–F'02) received the B.E. (Honors) degree in electrical and electronics engineering from Regional Engineering College, Tiruchirapalli, University of Madras, India, in 1983 and the Ph.D. degree in computer science from the University of Central Florida, Orlando, in 1988.

He is currently a Professor in the Department of Computer Science and Engineering and the Nanomaterials and Nanomanufacturing Research Center at the University of South Florida, Tampa. During 1998–1999, he was a Professor at the University of Texas at El Paso. His research interests include VLSI design, design automation, computer architecture, and parallel processing. He has developed many special-purpose VLSI chips for computer vision, image processing, pattern recognition, data compression, and signal processing applications. He has published over 175 papers in reputed journals and conferences and is a co-owner of five U.S. patents.

Dr. Ranganathan is a member of the IEEE Computer Society and IEEE Circuits and Systems Society. He has served on the editorial boards for *Pattern Recognition* (1993–1997), *VLSI Design* (1994-present), IEEE TRANSACTIONS ON VLSI SYSTEMS (1995–1997), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–1999), and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (1997–2000). He was the Chair of the IEEE Computer Society Technical Committee on VLSI during 1997–2001. He served as the Steering Committee Chair of the IEEE TRANSACTIONS ON VLSI SYSTEMS during 2001–2002 and is serving as the Editor-in-Chief for 2003–2004. He was elected as a Fellow of IEEE in 2002 for his contributions to algorithms and architectures for VLSI systems. Recently, he received the Theodore and Venette Askounes-Ashford Distinguished Scholar Award from the University of South Florida, Tampa.