Low-Power FinFET Circuit Synthesis Using Multiple Supply and Threshold Voltages

PRATEEK MISHRA, ANISH MUTTREJA, and NIRAJ K. JHA
Princeton University

According to Moore's law, the number of transistors in a chip doubles every 18 months. The increased transistor-count leads to increased power density. Thus, in modern circuits, power efficiency is a central determinant of circuit efficiency. With scaling, leakage power accounts for an increasingly larger portion of the total power consumption in deep submicron technologies (>40%).

FinFET technology has been proposed as a promising alternative to deep submicron bulk CMOS technology, because of its better scalability, short-channel characteristics, and ability to suppress leakage current and mitigate device-to-device variability when compared to bulk CMOS. The subthreshold slope of a FinFET is approximately 60mV which is close to ideal.

In this article, we propose a methodology for low-power FinFET based circuit synthesis. A mechanism called TCMS (Threshold Control through Multiple Supply Voltages) was previously proposed for improving the power efficiency of FinFET based global interconnects. We propose a significant generalization of TCMS to the design of any logic circuit. This scheme represents a significant divergence from the conventional multiple supply voltage schemes considered in the past. It also obviates the need for voltage level-converters. We employ accurate delay and power estimates using table look-up methods based on HSPICE simulations for supply voltage and threshold voltage optimization. Experimental results demonstrate that TCMS can provide power savings of 67.6% and device area savings of 65.2% under relaxed delay constraints. Two other variants of TCMS are also proposed that yield similar benefits. We compare our scheme to extended cluster voltage scaling (ECVS), a popular dual-V_{dd} scheme presented in the literature. ECVS makes use of voltage level-converters. Even when it is assumed that these level-converters have zero delay, thus significantly favoring ECVS in time-constrained power optimization, TCMS still outperforms ECVS.

Categories and Subject Descriptors: B.6.1 [**Logic Design**]: Design Styles—*Combinational logic*; B.7.0 [**Integrated Circuits**]: General; G.1.6 [**Numerical Analysis**]: Optimization—*Linear programming*; J.6 [**Computer-Aided Engineering**]: Computer-Aided Design (CAD)

General Terms: Algorithms, Performance, Design

Additional Key Words and Phrases: Low-power, TCMS, linear programming, synthesis

ACM Reference Format:

Mishra, P., Muttreja, A, and Jha, N. K. 2009. Low-power FinFET circuit synthesis using multiple supply and threshold voltages. ACM J. Emerg. Technol. Comput. Syst. 5, 2, Article 7 (July 2009), 23 pages. DOI = 10.1145/1543438.1543440 http://doi.acm.org/10.1145/1543438.1543440

This work was supported by SRC under contract No. 2007-HJ-1602.

Author's address: P. Mishra, Electrical Engineering Department, Princeton University, Princeton NJ 08544; email: pmishra@princeton.edu.

DOI 10.1145/1543438.1543440 http://doi.acm.org/10.1145/1543438.1543440

1. INTRODUCTION

Technology scaling has resulted in continual improvement in the performance of digital circuits. The accompanying reduction in the supply voltage, V_{dd}, reduces the switching power consumption quadratically. A reduction in V_{dd} also necessitates the reduction in threshold voltage V_{th} to maintain the gate drive strength $(V_{dd} - V_{th})$. The reduction in V_{th} with each technology generation leads to an exponential increase in leakage current. Also, the number of transistors in a chip increases exponentially, resulting in an increased power density. Thus, power consumption has become a major concern for chip designers because of the increased packaging and cooling costs as well as potential reliability problems. Therefore, power efficiency has assumed increased importance. This paper explores how circuits based on FinFETs, an emerging transistor technology that is likely to supplement or supplant bulk CMOS at the 32nm node or beyond, can be made power-efficient.

The steady miniaturization of metal-oxide-semiconductor field-effect transistors (MOSFETs) with each new generation of CMOS technology has provided us with improved circuit performance and cost per function over several decades. However, continued transistor scaling will not be straightforward in the sub-32nm regime because of fundamental material and process technology limits [King 2005]. The main challenges for sub-32nm gate length regime are two-fold: (a) minimization of leakage current, and (b) reduction in the device-to-device variability to increase yield [Bernstein et al. 2003]. Several innovative device structures, such as ultra thin-body silicon-on-insulator (SOI) and FinFETs, have been proposed to address the challenges being posed by continued scaling. These devices have increased surface-to-volume ratio which improves the device's electrostatics, resulting in better short-channel characteristics. FinFETs (Figure 1) have emerged as the best solution for next generation MOSFETs because of better scalability and ease of fabrication. FinFETs with gate lengths down to 10nm have already been demonstrated with excellent control of short-channel effects and less than 0.5ps intrinsic delay [Yu et al. 2002]. Digital logic circuits implemented in FinFETs have been shown to be much more power-efficient than the same circuits implemented in bulk CMOS at the same gate length [Swahn and Hassoun 2006].

However, beyond the technology-driven benefits, circuits can also benefit from the dual-gate structure of FinFETs to further optimize power and performance. Such a structure provides us with the ability to control the V_{th} of one gate by applying a voltage bias at the other gate. This property leads to easy V_{th} controllability of FinFETs. Since V_{th} controls the speed as well as leakage current of a transistor, its controllability can be a powerful tool for circuit optimization. A new circuit synthesis style based on multiple supply and threshold voltages is presented in this article. The synthesis style is dependent on TCMS, which is an innovative way to control V_{th} of connected-gate FinFETs. TCMS is based on the principle that in an overdriven gate, delay and subthreshold current can be reduced simultaneously. TCMS is explained in a greater detail in Section 3. In classical multiple supply voltage schemes, logic gates on the critical path are typically assigned high supply voltage while the

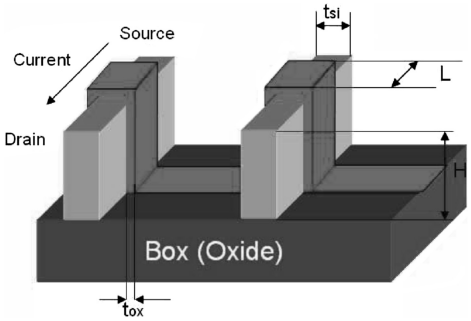

Fig. 1. Multifin FinFET.

gates on the noncritical paths are connected to a low supply voltage in order to reduce power consumption while maintaining circuit performance [Usami and Horowitz 1995; Usami et al. 1998]. In addition to multiple supply voltage design techniques, lowering the V_{th} can maintain high performance while lowering the supply voltage. Unfortunately, this leads to an exponential increase in the leakage current, which has become an important concern in low-voltage high-performance designs [Roy et al. 1999].

Using our TCMS scheme, one can sharply diverge from the way circuits have been designed in the past. This scheme does not make use of a lower supply voltage, and thus no lowering of V_{th} is required to maintain performance. It uses a nominal and a higher supply voltage. A possible consequence of an increased supply voltage is an increased V_{th}. Thus the leakage power can be reduced drastically. We employ a set of three supply voltages: A nominal supply voltage (V_{dd}^{L}), a slightly higher supply voltage (V_{dd}^{H}), and a slightly negative supply voltage (V_{ss}^{H}). The scheme is based on the principle that in an overdriven gate (a gate which is driven by an input voltage that is higher than its supply voltage) the delay and subthreshold leakage can be reduced simultaneously. A preliminary version of this work was presented in Mishra et al. [2008]. We make the following contributions in this article:

(1) We propose the TCMS scheme for arbitrary logic circuits which uses multiple supply and threshold voltages to reduce circuit power consumption.
(2) We discuss a library consisting of inverters, and two-input NAND and NOR gates based on the TCMS scheme. The library consists of seven different types of inverters, 25 different types of NAND and NOR gates.
(3) We extend a linear programming–based optimization methodology for implementing the TCMS scheme based on multiple supply and threshold voltages for delay-constrained power optimization.
(4) Experimental results show that the application of TCMS to a set of benchmarks reduces power consumption on an average by 67.6% at 30% slack.
(5) We propose two variants of the TCMS scheme. The first uses dual supply and threshold voltages to reduce circuit power consumption. The second uses a TCMS scheme with a single V_{th}. These schemes also result in significant power savings.

(6) We also implement ECVS using the linear programming framework and show that, even under an optimistic scenario, the power saving obtained by ECVS on an average is lower than that obtained by the TCMS scheme.

The article is organized as follows. In Section 2, we review related work. In Section 3, we discuss the TCMS principle, which forms the basis for the scheme presented in this article. In Section 4, we discuss gate library design using TCMS. In Section 5, we discuss the power optimization methodology and the implementation of ECVS. In Section 6, we present the experimental results and conclude in Section 7.

2. RELATED WORK

Next, we review prior work in the area of FinFET circuit design. FinFETs have been used in a variety of innovative ways in digital and analog circuit designs. However, the property which has been exploited the most is the ability to control the two gates of a FinFET independently. The independent gates can be used to merge parallel transistors to reduce circuit power and area [Mahmoodi et al. 2004; Datta et al. 2007]. Another principle, which has often been employed, is the use of a back-gate voltage bias to modulate front gate V_{th}. In Joshi et al. [2007], both a forward bias to reduce V_{th} while performing Read/Write operations in an SRAM and a reverse bias to reduce the leakage power in the stand-by mode was used. In Wei et al. [1997], Beckett [2005], Chiang et al. [2005], and Zhang et al. [2005], various circuits employing back-gate voltage bias to control subthreshold leakage were presented. Though a large part of previous work has been devoted to the design of specific circuits using FinFETs, some researchers have also explored the area of FinFET circuit synthesis. A logical effort based algorithm for gate sizing using FinFETs was presented in Swahn and Hassoun [2006]. Different logic design styles, leading to low leakage, based on independent control of FinFET gates were studied in Cakici et al. [2005] and Muttreja et al. [2007]. A tool that directly translates CMOS netlists to FinFET netlists was presented in Nowak et al. [2004]. The downside of the use of independently controllable FinFET gates is that its fabrication requires an extra processing step. This also complicates the layout due to extra wiring required to feed the bias to the back gate.

In this article, low-power FinFET circuits are synthesized using the TCMS concept. TCMS is also based on the ability to control FinFET V_{th}. However, no voltage biases are fed to the back gate in order to control the V_{th} of the front gate. In fact, the TCMS scheme is applied to shorted-gate FinFETs (in which the front and back gates are shorted). TCMS was only applied to global interconnects in Muttreja et al. [2008]. We propose a significant generalization of the TCMS concept to the synthesis of any logic circuit. This approach is very different from existing multiple supply and threshold voltage power optimization schemes. In existing schemes, the gates on the critical path operate at the higher V_{dd}, that is, nominal supply voltage, or lower V_{th} to meet the performance requirements, and the gates on the noncritical paths operate at the lower V_{dd} or higher V_{th}, thereby reducing the overall power consumption without performance degradation. In contrast to the earlier approach, in our scheme, both a nominal V_{dd} and a higher

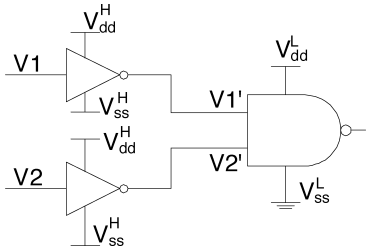

Fig. 2. The principle of TCMS.

V_{dd} as well as a nominal V_{th} and a higher V_{th} are deployed on critical as well as noncritical paths.

3. THE PRINCIPLE OF TCMS

The V_{th} at the front (back) gate of a FinFET can be controlled not only through process-related engineering such as (a) controlling the number of dopant atoms in the channel, and (b) using different values for the gate work-function, but also through the application of a voltage at the other gate. A general relationship between the threshold voltage ($V_{th_{g_f}}$) of the front gate (g_f) and applied voltage bias V_{g_bs} on the back gate (g_b) is given in Trivedi et al. [2007]. However, for the purpose of this work, the following simple relationship among $V_{th_{g_f}}$, V_{g_bs}, $V_{th_{g_b}}$ (V_{th} of the back gate), and $V^0_{th_{g_f}}$ (minimum observed value of $V_{th_{g_f}}$) would suffice:

$$V_{th_{g_f}} \approx \begin{cases} V^0_{th_{g_f}} - \delta(V_{g_bs} - V_{th_{g_b}}) & \text{if } V_{g_bs} < V_{th_{g_b}}, \\ V^0_{th_{g_f}} & \text{otherwise} \end{cases} \quad (1)$$

where δ is a positive quantity whose value depends upon the ratio of gate and body capacitances. If the FinFET is operated in a shorted-gate mode, the V_{th} of both gates respond simultaneously to the change in the voltage at the other gate. This happens because when the back gate is in depletion mode, charge coupling occurs between the front and back gates. However, when the back gate is in strong inversion mode, the free carriers effectively screen the back gate electric field, making $V_{th_{g_f}}$ independent of V_{g_bs}.

TCMS exploits the fact that in an overdriven FinFET, the delay and subthreshold leakage can be reduced at the same time. Figure 2 is used to further illustrate this point. In this figure, the supply voltages for the inverters are V^H_{dd} and V^H_{ss}, and for the NAND gate they are V^L_{dd} and V^L_{ss}. A possible set of values of V^H_{dd}, V^H_{ss} and V^L_{dd} are 1.08V, −0.08V and 1.0V. V^L_{ss} is assumed to be tied to ground. In the remainder of this paper, it is assumed that any logic gate connected to V^H_{dd} is also connected to V^H_{ss} and similarly for the lower supply voltages. Let $V1$ and $V2$ in Figure 2 be held at logic 0. This would lead to a logic 1 at $V1'$ and $V2'$. Thus, the N-type FinFETs in the NAND gate will be conducting and the P-type FinFETs will be leaking. Both the subthreshold leakage current and, delay can be controlled through the control of FinFET V_{th}. In this case, it can be seen that the N-type FinFETs experience a bias voltage of 1.08V,

which is higher than the normal gate drive of 1.0V. On the other hand, P-type FinFETs are reverse biased by −0.08V. Thus, the N-type FinFETs experience an increased gate-to-source voltage, compared to the case when they are driven by a supply voltage of 1.0V. The increased drive strength of N-type FinFETs results in a reduction in the falling delay of the NAND gate. The applied bias causes the V_{th} of the P-type FinFETs to be increased, thereby resulting in a lower subthreshold leakage. In addition, a negative gate-to-source bias on the P-type FinFETs further brings down their subthreshold leakage current. Similarly, the application of a logic 1 at the circuit inputs leads to a reduction in the leakage current in the N-type FinFET and improvement in the drive strength of the P-type FinFET. The use of TCMS-style logic gates in circuit synthesis is explained in greater detail in the next section.

TCMS is based on the principle that an N-type (P-type) FinFET experiences an overdrive when it is conducting, and simultaneously a P-type (N-type) FinFET experiences a reverse-biased voltage which leads to very low subthreshold currents. TCMS can provide considerable power savings despite the use of an increased V_{dd}. In conventional multiple supply voltage schemes, power savings can be attributed to the use of a lower V_{dd} on noncritical paths, which results in lower leakage and dynamic power dissipation. However, in TCMS, power savings can mainly be attributed to the reduction in the leakage current. Although the dynamic and leakage power may slightly increase for gates operating at the higher supply voltage, this is far outweighed by the reduction in leakage power in overdriven gates.

We performed HSPICE simulations on an overdriven N-type FinFET using the predictive technology model (PTM) for 32nm FinFETs. PTMs are available from Zhao and Cao [2006], and have also been used for all other HSPICE experiments reported in this paper. These models have been verified against manufactured 32nm FinFETs [Zhao and Cao 2007] and have been widely used for circuit simulations [Wang et al. 2006; Muttreja et al. 2007]. Figure 3 shows the simulation results. In the simulation, the drain of the N-type FinFET was tied to V_{dd}^L and the source terminal to ground. The gate voltage level varied between V_{dd}^H and V_{ss}^H. Thus, the N-type FinFET was forward-biased when V_{dd}^H was applied at its gate and reverse-biased when V_{ss}^H was applied. Let I_{on}^L (I_{off}^L) and I_{on}^H (I_{off}^H) denote the on-currents (off-currents) through the FinFET at normal drive ($V_{gs} = V_{dd}^L$) and overdrive ($V_{gs} = V_{dd}^H$), respectively. As shown in Figure 3, I_{on}^H exceeds I_{on}^L by 3.4%. On the other hand, I_{off}^H is almost 5X smaller than I_{off}^L. The large reduction in the subthreshold current in an overdriven FinFET is the key to the large power savings in TCMS schemes.

4. LIBRARY DESIGN USING TCMS

The concept of TCMS was illustrated through its application to a NAND gate in the previous section. However, TCMS can be extended to any logic gate based on shorted-gate FinFETs. This is explained next.

Consider the two-input NAND gate shown in Figure 4. The power supply voltages for the NAND gates are V_{dd}^L and V_{ss}^L. Consider the two inputs a and b. They may be the outputs of a high-V_{dd} gate or a low-V_{dd} gate (a high-V_{dd}

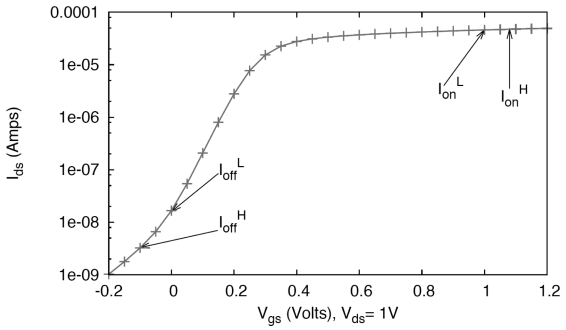

Fig. 3. Simulated I_{ds} -V_{g_fs} characteristics for an overdriven 32nm N-type FinFET.

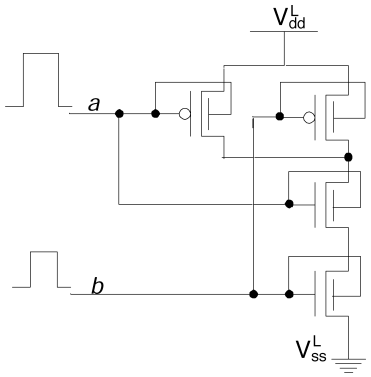

Fig. 4. NAND gate employing the TCMS principle.

gate is connected to V_{dd}^H and V_{ss}^H, and a low-V_{dd} gate to V_{dd}^L and V_{ss}^L). During circuit synthesis, when this gate is embedded in a larger circuit, it might so happen that a is the output of a high-V_{dd} gate and b comes from a low-V_{dd} gate or *vice versa*. Suppose the former is true. Thus, the FinFETs connected to input a follow the TCMS principle explained above. FinFETs connected to input b cannot employ the TCMS principle because there is no gate-to-source voltage difference to exploit.

On the other hand, if the power supply voltages for the NAND gate are V_{dd}^H and V_{ss}^H, then the FinFETs connected to input a will not be able to take advantage of the TCMS principle. Also, input b is from the output of a low-V_{dd} gate and is driving a high-V_{dd} gate. This results in an increased leakage current because the P-type FinFET is forward-biased by $V_{dd}^H - V_{dd}^L$.

To avoid this problem, a level-converter may be used to restore the signal to V_{dd}^H. These level-converters may be combined with a flip-flop, as in the clustered voltage scaling (CVS) technique [Usami and Horowitz 1995], to minimize the power for voltage level restoration. In an "asynchronous" approach, such as ECVS [Usami et al. 1998], level-converters may be inserted between logic gates connected to V_{dd}^L and V_{dd}^H. In such schemes, the power and delay overheads for the level-converters are large.

In the case of TCMS, using level-converters is not an attractive option, because power savings are obtained through the use of overdriven gates, the frequent use of which necessitates frequent level conversion. However, level conversion can be built into logic gates without requiring the use of level-converters [Diril et al. 2005], through the use of a high-V_{th} FinFET at the inputs of high-V_{dd} gates that need to be driven by a low-V_{dd} input voltage. FinFET V_{th} may be controlled through a number of mechanisms. For example, there are several process-related options to statically control the V_{th} of a FinFET, for example, channel doping, gate workfunction engineering, or asymmetrical double gates [Chang et al. 2000].

The first step towards evaluating the utility of the TCMS principle for arbitrary logic circuits involves the design of technology libraries, consisting of high-V_{dd} cells, low-V_{dd} cells, low-V_{dd} cells that are being driven by high-V_{dd} cells, and high-V_{dd} cells that are being driven by low-V_{dd} cells. All these cells have to be characterized at both high-V_{th} and low-V_{th}. Thus, the design variables that need to be targeted are supply voltage, input gate voltage, and threshold voltage. Hence, for a two-input NAND gate of a given size, we have five design variables: supply voltage, gate input voltage for input a, gate input voltage for input b, V_{th} for FinFETs connected to input a, and V_{th} for FinFETs connected to input b. If the V_{th} of a P-type FinFET connected to an input is high (low), then the corresponding N-type FinFET connected to the same input also has a high (low) V_{th}. It can be easily seen that 32 two-input NAND gates of a particular size are possible, because of the five design variables. For example, one type of NAND gate may have a high supply voltage (V_{dd}^H,V_{ss}^H), a low input a gate voltage, a high input b gate voltage, a high V_{th} for FinFETs connected to input a, and a low V_{th} for FinFETs connected to input b. Let 1 denote the case when either a high supply voltage or a high input gate voltage or a high V_{th} is used. Similarly, let 0 denote when either a low supply voltage or a low input gate voltage or a low V_{th} is used. Using this convention, the example NAND gate can be termed nand10110. The first 1 in nand10110 denotes a high supply voltage; thereafter, 0 denotes a low input a gate voltage, third 1 denotes a high input b gate voltage, the fourth 1 represents the high V_{th} for input a, and the fifth 0 represents a low V_{th} for input b. Thus, 32 NAND gate modes are possible ranging from nand00000 to nand11111. However, certain combinations of design variables are not allowed: a logic gate with a high supply voltage and low input gate voltages cannot employ low-V_{th} transistors as this will lead to a large leakage current, as explained earlier. Thus, nand10000, nand10001, nand10010, nand10100, nand10101, nand11000, and nand11010 are not allowed. This leads to 25 NAND gate modes instead of 32. Similarly, there are 25 NOR gate modes. Since the inverter is a one-input gate, it has three design variables: supply voltage, input gate voltage and V_{th}. This leads to seven valid modes for inverters. For each NAND, NOR and inverter mode, we include five sizes: X1, X2, X4, X8, and X16. The library is characterized by simulating the delay, leakage, and short-circuit power consumption of each constituent cell in HSPICE. Transistor capacitance is also measured using HSPICE. To model interconnect delay and load, fanout and size-dependent wire load models were obtained by scaling the wire characteristics available as part of a 130nm

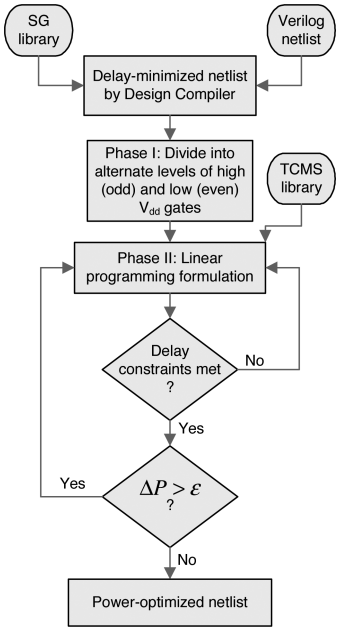

Fig. 5. Power optimization flow.

technology library, according to the method presented in Sylvester and Keutzer [1998].

5. POWER OPTIMIZATION METHODOLOGY

In this section, the methodology for implementing the TCMS scheme, via the use of multiple supply and threshold voltages, for delay-constrained power optimization is presented. Our power optimization flow is shown in Figure 5. The optimization methodology uses a two-phase strategy to find the circuit with the best power consumption. In the initialization phase, called Phase I, the logic netlist is first mapped to low-V_{dd} gates with low V_{th}. The circuit is then levelized into alternate levels of high-V_{dd} and low-V_{dd} gates. In Phase II, an extension of linear programming–based gate sizing algorithm [Chinnery and Keutzer 2005] is applied to the netlist obtained from Phase I. We also illustrate our methodology through its application to a small benchmark (ISCAS'85 c17) in this section. In addition, we discuss the implementation of ECVS using the linear programming framework described in Phase II and illustrate its impact through application to c17.

5.1 Optimization Flow

The power minimization flow is shown in Figure 5. It starts by mapping the logic netlist to low-V_{dd} gates with low-V_{th} and finding its delay-minimized configuration, using Synopsys Design Compiler. The library of low-V_{dd} gates with low-V_{th} is referred as the SG library because of its use of shorted-gate FinFETs. Thereafter, in Phase I, the circuit is divided into alternate levels of high-V_{dd}

and low-V_{dd} gates. The gates at odd levels are changed to high-V_{dd} gates with high-V_{th}. The gates at even levels are changed to other modes of low-V_{dd} gates to maintain circuit consistency, as will be clearer later. Next, in Phase II, a linear programming based algorithm is used to assign gate sizes and modes to the mapped circuit by selecting cells from the TCMS library. Cell selection is based on the algorithm presented in Chinnery and Keutzer [2005]. The linear programming formulation can be used for reducing both delay and power in the circuits. The iteration terminates when all the delay constraints are met and the change in power consumption between successive iterations is less than some pre-specified percent. We give details of the optimization flow next.

5.2 Phase I: Initialization of the Circuit

Recall that the pair of high (low) supply voltages are denoted by V_{dd}^H and V_{ss}^H (V_{dd}^L and V_{ss}^L). The high and low threshold voltages are denoted by V_{th}^H and V_{th}^L, respectively. First, the circuit is synthesized by mapping it to low-V_{dd} gates with low threshold voltage V_{th}^L. During the initialization procedure, the circuit is levelized. The level of each primary input is defined to be 0. The level of a gate G, denoted as $l(G)$, can be calculated by $l(G) = 1 + max_{i\epsilon\{1,2,\ldots,FN\}} l(GI_i)$ where GI_i is the ith fanin of gate G and FN is the gate fanin. Next, all the gates located at an odd level in the initial netlist are replaced by high-V_{dd} gates, with V_{th}^H at FinFETs connected to input a if this input arrives from an even level, that is, input a is the output of a low-V_{dd} gate. On the other hand, if the input arrives from an odd level, the threshold voltage of the FinFETs can, in general, be allowed to be either V_{th}^L or V_{th}^H. However in our approach, we replace it by V_{th}^H to reduce the initial leakage power. Note that this V_{th}^H assignment can be changed to V_{th}^L in Phase II if the optimization algorithm deems it necessary. The gates at an even level are replaced with other modes of low-V_{dd} gates to maintain circuit consistency, as mentioned earlier. We next illustrate the initialization phase through an example.

Consider the circuit shown in Figure 6. Initially, the circuit is synthesized using V_{dd}^L and V_{th}^L, that is, all the NAND gates and inverters are of the form nand00000 and inv000, respectively. Thereafter, as explained earlier, the inverters of size X4 at level 1 are replaced with other inverters from the TCMS cell library. The replaced cells have size X4, but their supply voltages are (V_{dd}^H, V_{ss}^H) and their threshold voltages are V_{th}^H, i.e., the replaced inverters are of mode inv101. Similarly, the NAND gate at level 3 is replaced with a high-V_{dd} NAND gate which employs V_{th}^H as the threshold voltage, i.e., it has the nand10111 mode. The NAND gate at level 2 is replaced with nand01011, and the inverter at level 4 is replaced with inv011. This is done so that modes of the gates at an even level are consistent with the circuit topology.

When a gate is changed from a low-V_{dd} gate to a high-V_{dd} gate, it is not necessary that both of its inputs will come from an even level and will thus be low-V_{dd} signals. It might so happen that one of the inputs comes from an odd level and is the output of a high-V_{dd} gate. This explains the need for 25 different NAND and NOR gate modes and seven different inverter modes in the cell library. The circuit is divided into alternate levels of high-V_{dd} and low-V_{dd}

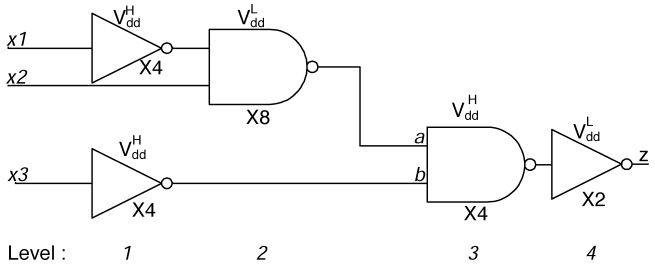

Fig. 6. Example circuit.

gates, to make use of the TCMS scheme which is based on the principle of a high-V_{dd} gate driving a low-V_{dd} gate. This also leads to a low-V_{dd} gate driving a high-V_{dd} gate. However, as explained earlier, in such a situation, V_{th}^{H} at the inputs is used to reduce leakage currents.

In Phase II the circuit, initialized in the above fashion, is fed to a gate sizing algorithm. This is described next.

5.3 Phase II: Linear Programming Formulation

Circuit sizing algorithms often perform a search amongst the various candidate cells available for each gate to select the cell with the best power-delay sensitivity. Let $\triangle P$ represent the reduction in power and $\triangle D$ the degradation in delay, if an alternate cell is used. The ratio $\frac{\triangle P}{\triangle D}$ is the power-delay sensitivity. Such a cell is then used to replace the gate. However, as shown in Chinnery and Keutzer [2005], such decisions can be quite suboptimal. The major advantage of the linear programming approach is that it leads to an analysis of how changing each gate affects the gate it has a path to. We next review the gate sizing algorithm presented in Chinnery and Keutzer [2005] and discuss enhancements we have made to it for implementing the TCMS scheme.

The linear programming formulation is an iteration based algorithm. In each iteration, it selects the best cells for any number of gates in the circuit, based on the power-delay sensitivity. To reduce power, the cell with maximum reduction in power for a given increase in delay is chosen. To reduce delay, the cell with maximum delay decrease for the corresponding increase in power is chosen. When an alternative cell is chosen, the level of the existing gate and the input gate voltages play an important role. If the existing gate is at an odd level and the voltage at input a (b) is high (low), it can only be replaced by gates which have a high supply voltage and high (low) voltage at input a (b). The same is true for gates at even levels. The free design variables are the threshold voltages and gate sizes. The linear programming formulation is able to select alternative cells for any number of gates in the circuit during each iteration. It uses a cell choice variable γ_v for each gate. γ_v denotes whether an alternative cell has been chosen or not. γ_v varies continuously in the range [0, 1]. A value of γ_v greater than a threshold value indicates an alternative cell should be used, else not. In Chinnery and Keutzer [2005], the threshold value chosen is 0.99. We found that such a high threshold value greatly impairs the chances of a cell being replaced. We found empirically that a threshold value of 0.6 works

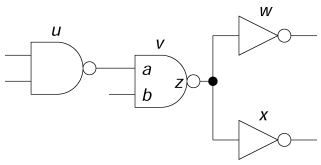

Fig. 7. A circuit to illustrate delay constraints.

better. An alternate cell for gate v (Figure 7[1]) is then chosen by minimizing power among various candidate cells for which $d'_v \leq d_v + \gamma_v \triangle d_v$, where d'_v is the delay through v after a cell change. At the end of an iteration of the algorithm, all the gates whose alternative cells have a γ_v value greater than 0.6 are replaced with alternative cells. Equation (2) gives the objective used to optimize power in the linear programming formulation. Delay constraints at individual gates (for the circuit in Figure 7) and at the circuit outputs are given by Equations (3) and (4), respectively.

$$\min \left(\sum_{v \in V} \gamma_v \triangle P_v \right) \tag{2}$$

$$\begin{aligned} t_{vw,rise} \geq t_{uv,fall} &+ d_{uv,rise} \\ &+ \gamma_v(\triangle t_{uv,fall,v} + \triangle d_{uv,rise,v} + \beta_{vw,fall} \triangle s_{uv,rise,v}) \\ &+ \sum_{x \in fanout(v), x \neq w} \gamma_x(\triangle d_{uv,rise,x} + \beta_{vw,\ fall} \triangle s_{uv,rise,x}) \end{aligned} \tag{3}$$

$$\max_{v \in outputs} \{t_{v,rise}, t_{v,fall}\} \leq T_{max} \tag{4}$$

In these equations, all timing arcs are assumed to have a negative polarity; that is, a falling input causes a rising transition at the output if the output changes. $\triangle P_v$ is the change in power due to changing gate v. $t_{uv,fall}$ is the falling arrival time at gate v from gate u. $t_{vw,rise}$ is the rising arrival time at gate w from gate v. $d_{uv,rise}$ is the delay from the signal on uv to the output of v rising. $\triangle t_{uv,fall,v}$ is the change in $t_{uv,fall}$ due to the cell of v changing. $\triangle d_{uv,rise,x}$ ($\triangle d_{uv,rise,v}$) and $\triangle s_{uv,rise,x}$ ($\triangle s_{uv,rise,v}$) are the changes in delay and slew, respectively, of this timing arc if cell x (v) changes. $\beta_{vw,fall}$ is a sensitivity term that determines how delay $d_{vw,fall}$ is impacted by $\triangle s_{uv,rise}$. T_{max} is the maximum allowed signal arrival time at circuit outputs. Term $\triangle t_{uv,fall,v}$ was added to the original formulation in Chinnery and Keutzer [2005]. It was found empirically that this term helps to better model the effect of a change in fanout load of gate u on its delay.

As mentioned earlier, the sizing algorithm proceeds by selecting alternative cells for any number of gates in the circuit depending upon the value of γ_v. However, after each power optimization iteration, there might be a violation of the delay constraints provided for the circuit. This may happen because the algorithm replaces several gates at once and the delay constraints are based on individual gates changing, and simultaneous changes in a gate and its fanin

[1]Reproduced from [Chinnery and Keutzer 2005].

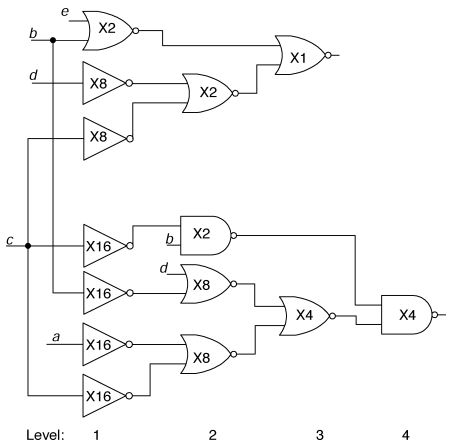

Fig. 8. Delay-minimized netlist obtained using Design Compiler.

are not modeled. If the arrival time constraint (ATC) at the outputs is violated, delay minimization is performed using the delay minimization version of the algorithm mentioned earlier. However, the objective function and the criteria for selecting an alternative cell changes. A detailed analysis of the delay minimization framework is provided in Chinnery and Keutzer [2005]. The iteration terminates when the ATC has been met and the power reduction from one iteration to the next is less than some pre-specified percent.

Next, we show the application of our methodology to the smallest ISCAS'85 benchmark c17.

5.4 Application of Methodology to c17

We synthesized the power-optimized netlist for ISCAS'85 benchmark c17 at 130% ATC, that is, with a slack of 30% relative to the delay-minimized version, using the methodology illustrated earlier. The set of high supply voltages used were 1.08V and −0.08V. The nominal set of supply voltages were 1.0V and ground. These supply voltages were chosen by fixing the nominal set of supply voltages and experimenting with various sets of high supply voltages. The two threshold voltages for N-type FinFETs were 0.29V and 0.45V and those for P-type FinFETs were −0.25V and −0.40V. The switching activity at each primary input was set to 0.1.

c17 is initially mapped to low-V_{dd} gates with low-V_{th} and the delay-minimized logic netlist is obtained, as shown in Figure 8. Thereafter, the power-optimized netlist is obtained at 130% ATC. We achieved 50.3% power reduction for this circuit. The initial power consumption of the delay-minimized netlist was 301.35μW (leakage power: 28.02μW, dynamic power: 273.33μW). In the power-optimized netlist, leakage power reduces by 92.8% and dynamic power by 45.9%, and, hence, the total power consumption reduces to 149.87μW. The cells chosen by our methodology for c17 are shown in Figure 9.

The reasoning behind why the leakage power could be reduced by almost 92.8% is as follows. The cells in the optimized netlist either have a high or low

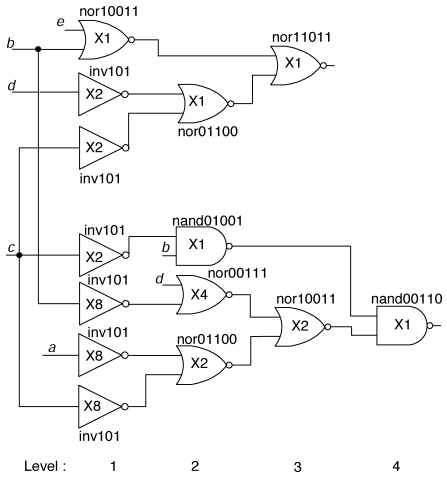

Fig. 9. Power-minimized netlist obtained using the TCMS principle.

V_{th}. If they have a low V_{th}, they are being driven by high-V_{dd} gates and thus, as explained in Section 3, the leakage power of these gates reduces significantly. To illustrate the point, consider the NOR and NAND gates at levels 2 and 4 in Figure 9. The mode of NOR and NAND gates in level 2 are nor01100, nor00111, and nand01001, while the mode of the NAND gate in level 4 is nand00110. Even when these gates employ a low V_{th}, they consume low leakage power because they are driven by high-V_{dd} gates. The gates at an odd level employ a high V_{th} because they are driven by low-V_{dd} gates. The high V_{th} results in low leakage power consumption in these cells. The increase in supply voltage of the cells at odd levels tends to increase the leakage power dissipation in these cells. However, the reduction in leakage power obtained by the increase in the V_{th} of these cells outweighs the increase in leakage power caused due to the use of a higher supply voltage. The dynamic power reduces in the power-optimized netlist because there is a large reduction in the area (and hence capacitance) as compared to the delay-minimized netlist (as can be seen by the gate sizes shown in Figures 8 and 9). The cells at the odd levels are high-V_{dd} gates. The dynamic power consumption of these cells tends to increase due to an increase in their supply voltage, but tends to decrease due to a reduction in the area of the cells they drive. The power consumption of the cells at even levels decreases if there is a reduction in the area of the cells they drive. In Figure 9, there is a decrease in the size of all except one (in which case the size is the same) cell in the netlist. Thus, the dynamic power of the low-V_{dd} cells decreases. The dynamic power of high-V_{dd} cells also decreases in most cases because the reduction in area outweighs the increase in supply voltage. The total number of fins in the delay-minimized netlist is 538 while the total number of fins in the power-optimized netlist is only 216.

5.5 Comparison to Conventional Multiple-V_{dd} Approach

In traditional multiple-V_{dd} approaches, voltage level-converters are required to feed a high supply voltage gate from a low supply voltage gate. Since these

supply voltages are different from those used in TCMS, let us denote them as HIGH-V_{dd} and LOW-V_{dd} gates, respectively. For example, HIGH-V_{dd} = 1.0V and LOW-V_{dd} = 0.7V are often used. There are two major conventional multiple-V_{dd} approaches that have been published in the literature. The first one is the "synchronous" multiple-V_{dd}, also known as CVS, methodology [Usami and Horowitz 1995] which has level-converters at the outputs of combinational logic only. Level-converters may be combined with flip-flops, known as level-converting flip-flops, to reduce the delay and power overhead attached with a level-converter.

The second methodology is the "asynchronous" multiple-V_{dd} scheme, also known as ECVS [Usami et al. 1998], in which asynchronous level-converters are used. These level-converters allow any gate in a path to be changed to be a LOW-V_{dd} gate, provided the path has sufficient slack. Since there is no restriction on the assignment of LOW-V_{dd} gate, ECVS can theoretically achieve greater power savings compared to CVS. Thus, we compare the TCMS scheme with the ECVS methodology.

The ECVS scheme is also implemented using the linear programming framework described earlier. The nominal set of supply voltages were 1.0V and ground. The other set of supply voltages used were 0.7V and ground. The V_{th} for N-type (P-type) FinFET was set to 0.29V (-0.25V). A new library (LOW-V_{dd} library) was created for the lower supply voltage (0.7V, ground). The library was characterized at the above V_{th}. The input gate voltage varies between 0.7V and ground, that is, no high input gate voltage is used when the LOW-V_{dd} library is characterized. When a HIGH-V_{dd} gate feeds a LOW-V_{dd} gate, the rising delay value can be directly obtained from the library. However, the falling delay value reduces because the input gate voltage increases from V_{dd}^{L} to V_{dd}^{H}. To circumvent this problem, we reduce the falling delay value by a fixed fraction [Srivastava and Sylvester 2003] whenever a HIGH-V_{dd} gate feeds a LOW-V_{dd} gate. Similarly, a HIGH-V_{dd} library was characterized with supply voltages (1.0V, ground). Only these two libraries are required for the ECVS methodology.

Initially, the circuit is mapped to the HIGH-V_{dd} library to obtain the delay-minimized netlist, and then the ECVS methodology is applied. We synthesized the power-optimized netlist for c17 at 130% ATC, using ECVS. The netlist is shown in Figure 10. The cells marked as inv and nand are LOW-V_{dd} gates, while the cells marked as inv_h and nor_h are HIGH-V_{dd} gates. The power consumption of the ECVS power-optimized netlist is $176.69\mu W$. The dynamic power reduces by 40.2% and the leakage power by 53.1% in the power-optimized netlist. On the other hand, leakage power reduces by 92.8% in the TCMS scheme. There is a larger reduction in leakage power in the TCMS scheme because of the negative gate-to-source voltage resulting from the TCMS principle and also due to the other set of high-V_{th} employed. Although ECVS employs LOW-V_{dd} gates which decrease the dynamic power consumption quadratically, still, dynamic power reduces by a larger margin in the TCMS scheme because of the greater reduction in area obtained. The total number of fins in ECVS power-optimized netlist is 282 which is 30.1% higher as compared to the TCMS scheme. Out of 14 gates in the ECVS circuit, three are mapped to LOW-V_{dd} gates. Note that the power-optimized netlist does not have any level-converters because there is

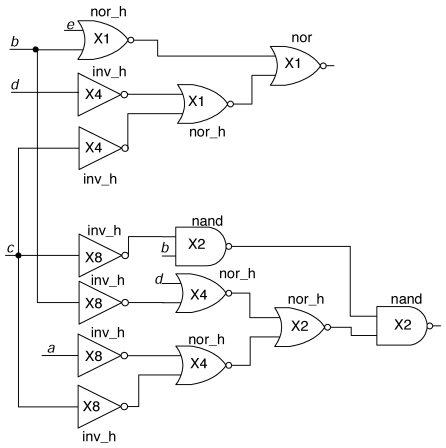

Fig. 10. Power-minimized netlist obtained using ECVS.

no LOW-V_{dd} gate driving a HIGH-V_{dd} gate. For larger circuits, however, ECVS would have to incur delay and power overheads of level-converters.

6. EXPERIMENTAL RESULTS

We present experimental results for power optimization on the ISCAS'85 benchmark suite in this section. The cell libraries were characterized using HSPICE based on the PTM [Zhao and Cao 2006] at 70^{o}C for 32nm FinFETs. This operating temperature was chosen because FinFETs suffer significantly from self-heating, and thermal simulations yield a temperature near 70^{o}C when the switching activity is 0.1, which is what is assumed for the experimental results here.

The same set of supply and threshold voltages were used that are mentioned in Section 5.4. The average runtime for sizing the benchmarks on a shared server farm using eight dual-core 64-bit AMD Opteron processors, running Red Hat Linux 4.0, was few CPU hours.

The input switching activities were propagated through the entire circuit to obtain the switching activities at different nodes using Synopsys Design Compiler.

The circuits were initially mapped to low-V_{dd} gates with low-V_{th} and their delay-minimized configurations were obtained. Thereafter, power was optimized using our methodology at various ATCs.

We present the experimental results in Table I for 130% ATC. Column 1 lists the ISCAS'85 benchmarks. Major column 2 presents dynamic, leakage and total power of the delay-minimized version. Major column 3 presents the power results for the TCMS scheme assuming (V_{dd}^H, V_{ss}^H) of (1.08V,−.08V). As we can see, the leakage power reduces by 95.8% and dynamic power by 53.3%, providing a total power reduction of 67.6% on an average when compared to the delay-minimized netlist. The FinFET area reduces by 65.2% on an average, as shown in Table II.

Table I. Power Savings Using the TCMS Scheme

Design	Power Consumption (μW)											
	Delay-minimized			TCMS (1.08V and −0.08V)			TCMS (Single V_{th})			Dual-V_{dd}		
	Dynamic	Leakage	Total	Dynamic	Leakage	Total	Dynamic	Leakage	Total	Dynamic	Leakage	Total
c432	679.15	546.38	1225.53	345.08	24.26	369.34	382.53	25.89	408.42	372.93	31.16	404.09
c499	9174.54	1757.26	10931.80	3973.30	95.37	4068.67	4357.34	100.63	4457.97	4181.26	118.74	4300.00
c880	1499.76	921.35	2421.11	768.27	33.90	802.17	827.37	35.08	862.45	802.10	27.14	829.24
c1355	8015.13	1583.94	9599.07	3414.07	80.98	3495.05	3805.77	88.01	3893.78	3564.44	94.86	3659.30
c1908	2810.42	1290.01	4100.43	1276.65	58.35	1335.00	1435.42	62.95	1498.37	1373.25	62.49	1435.74
c3540	2198.84	2073.06	4271.90	1198.03	53.25	1251.28	1273.48	90.67	1364.15	1262.22	69.07	1331.30
c5315	5545.22	2929.66	8474.88	2889.58	126.10	3015.68	3047.51	129.72	3177.23	2956.05	136.97	3093.02
c6288	11662.20	9749.32	21411.52	5479.86	394.28	5874.14	5806.66	397.43	6204.09	5792.84	362.09	6154.93
c7552	10836.80	5781.22	16618.02	5133.74	234.35	5368.09	5357.89	213.67	5571.56	5170.33	204.72	5375.05
Total	52422.06	26632.20	79054.26	24478.58	1100.84	25579.42	26293.97	1144.05	27438.02	25475.42	1107.24	26582.67
Savings	0	0	0	53.3%	95.8%	67.6%	49.8%	95.7%	65.3%	51.4%	95.8%	66.3%

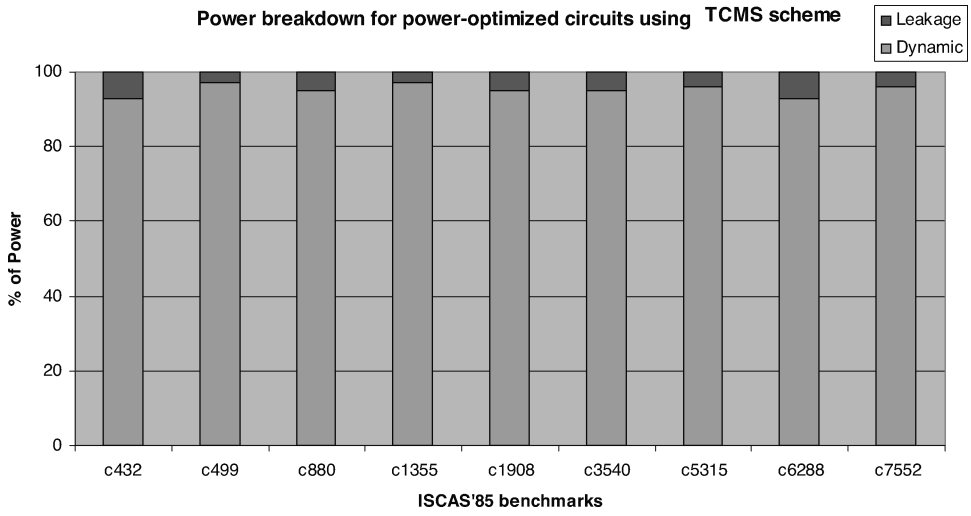

Fig. 12. Power breakdown for power-optimized circuits.

TCMS scheme, because of the greater area reduction obtained in the latter. It was also observed that in the power-optimized netlists obtained using the dual-V_{th} TCMS scheme, most of the cells employed high-V_{th}. This further explains the similar reductions in leakage power achieved by the two techniques.

Even though TCMS leads to a substantial power reduction, a limitation is the need to lay out an additional V_{ss}^{H} line. This limitation can be addressed by using the double-supply/double-ground grid suggested in Popovich et al. [2005]. Another way to address this limitation is to replace V_{ss}^{H} with V_{ss}^{L}, i.e., just use one V_{ss} line instead of two. This would decrease the power reduction possible. However, since the TCMS principle will still be applicable to the P-type FinFETs in the circuit, the power reduction would still be appreciable. Therefore, we performed experiments with dual supply voltages V_{dd}^{L}, V_{dd}^{H} and a single ground line V_{ss}^{H}. We refer to this as the dual-V_{dd} scheme. The results are shown in major column 5 in Table I. As expected, the overall power savings decreases slightly from 67.6% to 66.3%. The dynamic power consumption is slightly higher because the fin-count in the dual-V_{dd} scheme is higher than the fin-count in the TCMS scheme (see Table II). However, the leakage power consumption is almost similar across all the benchmarks. This is true because when a low-V_{dd} gate drives a high-V_{dd} gate in the TCMS scheme, the gate-to-source voltage difference increases the leakage current exponentially. This is counteracted by the use of a high-V_{th} in high-V_{dd} gates. However, in the dual-V_{dd} scheme, there is no gate-to-source voltage difference when a low-V_{dd} gate drives a high-V_{dd} gate and the output of the low-V_{dd} gate is low, due to the use of a single ground line. This leads to exponential savings in leakage power consumption of high-V_{dd} gates for the above case. However, when a high-V_{dd} gate drives a low-V_{dd} gate, there is an exponential amount of power savings in the low-V_{dd} gates due to the TCMS principle. In the dual-V_{dd} scheme, these power savings can only come from P-type FinFETs. The two counteracting effects in the dual-V_{dd} scheme thus lead to similar power savings to the TCMS scheme.

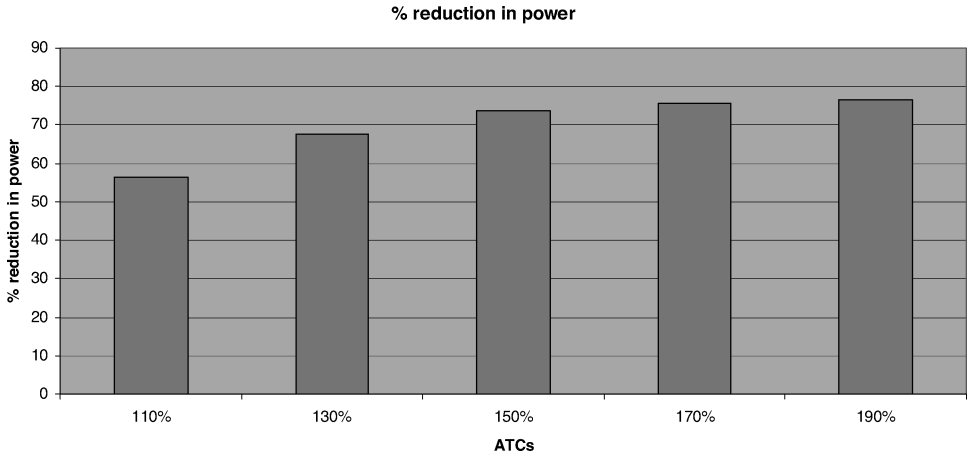

Fig. 13. Reduction in power consumption at various ATCs.

Next, we consider trends in average power savings across ISCAS'85 benchmarks at successively relaxed ATCs. As expected, the average total power savings increases from 56% to 76% (Figure 13). This happens because at relaxed ATCs, the linear programming algorithm has more overall slack to allocate to individual gates. This shows that the proposed TCMS based optimization methodology can effectively utilize the increased slack to reduce power consumption in circuits. The results at 110% ATC show the impact of TCMS under very stringent timing constraints. The average total power savings in this scenario is smaller (56%) because of the reduced availability of gate slacks. However, there is still a significant reduction in average power savings making this technique suitable for high-performance applications as well.

The TCMS scheme was also compared with the ECVS methodology. Table III presents dynamic, leakage and total power and area of the power-optimized netlists using ECVS at 130% ATC. The leakage power reduces on an average by 88.7% and the dynamic power reduces on an average by 50.7% when compared to the delay-minimized netlist. The FinFET area reduces by 61.2% on an average, as shown in Table III. We do not consider the delay overhead due to level-converters in this work. Considering this overhead would have reduced the slack available and thus reduced the power savings obtainable while meeting the delay constraints. Thus, the reported power savings for this scheme are quite optimistic.

The leakage and dynamic power of ISCAS'85 benchmarks (except c7552) reduce by a smaller amount when ECVS is applied, as compared to the TCMS optimization methodology, in spite of the fact that the delay overhead of level-converters is not considered. The efficacy of ECVS is the result of a large replacement of HIGH-V_{dd} gates with LOW-V_{dd} gates, as shown in Figure 14. On an average, 86% of the HIGH-V_{dd} gates are replaced by LOW-V_{dd} gates. Although the fin-count reduces by a smaller margin in ECVS, the dynamic power reduction is comparable to the TCMS because of the quadratic reduction in dynamic power provided by the LOW-V_{dd} gates. However, as pointed out earlier,

Table III. Power Savings Using ECVS

	ECVS Scheme			
Design	Dynamic (μW)	Leakage (μW)	Total (μW)	Area (No. of fins)
c432	417.31	68.12	485.43	4987
c499	4002.47	294.11	4296.58	16567
c880	865.59	86.40	951.99	8173
c1355	3588.80	234.94	3823.74	14437
c1908	1380.27	147.06	1527.33	11758
c3540	1697.14	228.63	1925.77	19347
c5315	3439.03	328.07	3767.10	25840
c6288	5700.75	1055.93	6756.68	85328
c7552	4765.77	545.35	5311.12	42776
Total	25857.13	2988.61	28845.74	229213
Savings	50.7%	88.7%	63.5%	61.2%

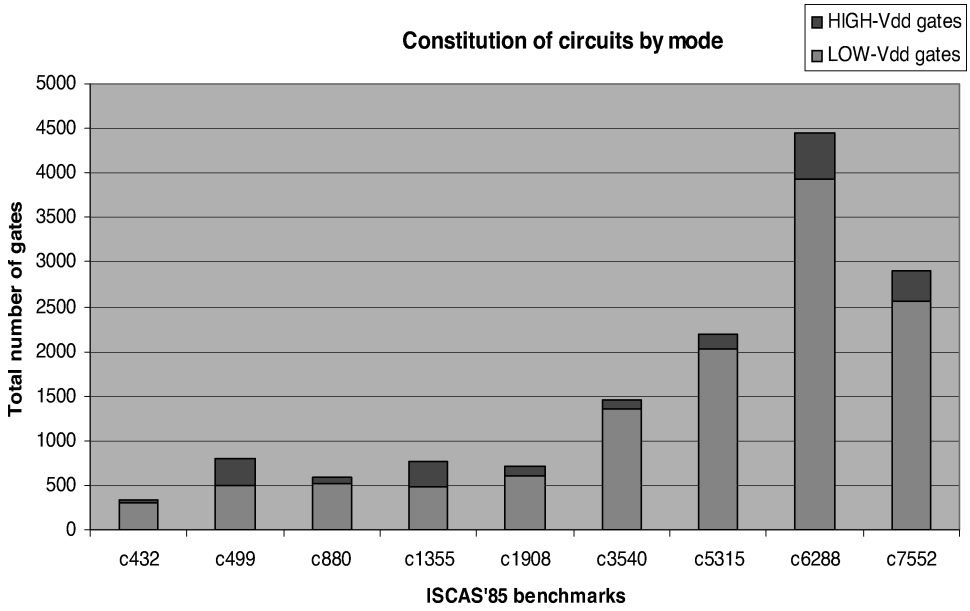

Fig. 14. Constitution of circuits by mode in ECVS circuits.

if the delay of the level-converters is included in the formulation, substantially fewer HIGH-V_{dd} gates will be replaced by LOW-V_{dd} gates. This will result in significantly decreased dynamic power savings from the ECVS scheme. Note that leakage power savings of ECVS would also go down significantly if much fewer HIGH-V_{dd} gates were replaced by LOW-V_{dd} gates.

Also, currently we perform experiments at 70^0C. However, in modern microprocessors, the operating temperature can be as high as 110^0C. Furthermore, FinFETs suffer from increased self-heating because in a FinFET the channel is surrounded by silicon dioxide, which has lower thermal conductivity compared to bulk silicon [Su et al. 1994]. Since, leakage power increases exponentially with a temperature increase, the fraction of leakage power relative to total

power will also increase drastically with a rise in temperature. In this case, the power savings obtained using the TCMS scheme will increase further because of its ability to reduce the leakage power effectively.

7. CONCLUSIONS

In this article, we proposed a synthesis scheme to reduce the power consumption of FinFET based circuits. This scheme is based on TCMS which is able to reduce both delay and subthreshold current in a logic gate simultaneously. The efficacy of the scheme was demonstrated on a set of ISCAS'85 benchmarks. We also proposed a variant of the TCMS scheme known as dual-V_{dd} scheme which addresses the problem of power supply layout imposed by TCMS scheme. The dual-V_{dd} scheme also promises high power savings under relaxed delay constraints. In addition, we showed that switching from a dual-V_{th} scheme to a single-V_{th} scheme does not adversely impact power savings much. Finally, we compared our optimization methodology with the conventional ECVS scheme and showed that the power savings obtained using the TCMS scheme exceeds the power savings obtained using ECVS.

REFERENCES

BECKETT, P. 2005. Low-power circuits using dynamic threshold voltage devices. In *Proceedings of the Great Lakes Symposium on VLSI*. 213–216.

BERNSTEIN, K., CHUANG, C.-T., JOSHI, R. V., AND PURI, R. 2003. Design and CAD challenges in sub-90nm CMOS technologies. In *Proceedings of the International Conference on Computer-Aided Design*. 129–136.

CAKICI, T., MAHMOODI, H., MUKHOPADHYAY, S., AND ROY, K. 2005. Independent gate skewed logic in double-gate SOI technology. In *Proceedings of the International SOI Conference*. 83–84.

CHANG, L., TANG, S., KING, T.-J., BOKOR, J., AND HU, C. 2000. Gate length scaling and threshold voltage control of double-gate MOSFETs. In *Proceedings of the International Electronic Device Meeting*. 719–722.

CHIANG, M.-H., KIM, K., TRETZ, C., AND CHUANG, C.-T. 2005. Novel high-density low-power logic circuit techniques using DG devices. *IEEE Electron. Dev. Lett. 52*, 10, 2339–2342.

CHINNERY, D. AND KEUTZER, K. 2005. Linear programming for sizing, V_{dd} and V_{th} assignment. In *Proceedings of the International Symposium on Low Power Electronics and Design*. 149–154.

DATTA, A., GOEL, A., CAKICI, R. T., MAHMOODI, H., LAKSHMANAN, D., AND ROY, K. 2007. Modeling and circuit synthesis for independently controlled double-gate FinFET devices. *IEEE Trans. Comput.-Aid. Des. 26*, 11, 1957–1966.

DIRIL, A. U., DHILLON, Y. S., CHATTERJEE, A., AND SINGH, A. D. 2005. Level-shifter free design of low-power dual supply voltage CMOS circuits using dual threshold voltages. *IEEE Trans. VLSI Syst. 13*, 9, 1103–1107.

JOSHI, R. V., KIM, K., WILLIAMS, R. Q., NOWAK, E. J., AND CHUANG, C.-T. 2007. A high-performance, low-leakage, and stable SRAM row-based back-gate biasing scheme in FinFET technology. In *Proceedings of the International Conference on VLSI Design*. 665–672.

KING, T.-J. 2005. FinFETs for nanoscale CMOS digital integrated circuits. In *Proceedings of the International Conference Computer-Aided Design*. 207–210.

SU, L. T., CHUNG, J. E., ANTONIADIS, D. A., GOODSON, K. E., AND FLIK, M. I. 1994. Measurement and modelling of self-heating in SOI nMOSFET's. *IEEE Electron. Dev. Lett. 41*, 69–75.

MAHMOODI, H., MUKHOPADHYAY, S., AND ROY, K. 2004. High-performance and low-power domino logic using independent gate control in double-gate SOI MOSFETs. In *Proceedings of the International SOI Conference*. 67–68.

MISHRA, P., MUTTREJA, A., AND JHA, N. K. 2008. Evaluation of multiple supply and threshold voltages for low-power circuit synthesis. In *Proceedings of the IEEE International Symposium on Nanoscale Architectures*. 77–84.

MUTTREJA, A., AGARWAL, N., AND JHA, N. K. 2007. CMOS logic design with independent gate FinFETs. In *Proceedings of the International Conference on Computer Design*. 560–567.

MUTTREJA, A., MISHRA, P., AND JHA, N. K. 2008. Threshold voltage control through multiple supply voltages for power-efficient FinFET interconnects. In *Proceedings of the International Conference on VLSI Design*. 220–227.

NOWAK, E. J., ALLER, I., LUDWIG, T., KIM, K., JOSHI, R. V., CHUANG, C.-T., BERNSTEIN, K., AND PURI, R. 2004. Turning silicon on its edge. *IEEE Circ. Dev. Mag. 20*, 1, 20–31.

POPOVICH, M., FRIEDMAN, E. G., SOTMAN, M., AND KOLODNY, A. 2005. On-chip power distribution grids with multiple supply voltages for high-performance integrated circuits. In *Proceedings of the Great Lakes Symposium on VLSI*. 2–7.

ROY, K., WEI, L., AND CHEN, Z. 1999. Multiple-V_{dd} and multiple-V_{th} CMOS (MVCMOS) for low-power applications. In *Proceedings of the International Symposium on Computer Architecture*. 366–370.

SRIVASTAVA, A. AND SYLVESTER, D. 2003. Minimizing total power by simultaneous V_{dd}/V_{th} assignment. In *Proceedings of the Asia South Pacific Design Automation Conference*. 400–403.

SWAHN, B. AND HASSOUN, S. 2006. Gate sizing: FinFETs vs 32nm bulk MOSFETs. In *Proceedings of the Design Automation Conference*. 528–531.

SYLVESTER, D. AND KEUTZER, K. 1998. Getting to the bottom of deep submicron. In *Proceedings of the International Conference on Computer-Aided Design*. 203–211.

TRIVEDI, V. P., FOSSUM, J. G., AND ZHANG, W. 2007. Threshold voltage and bulk inversion effects in nonclassical CMOS devices with undoped ultra-thin bodies. *Solid-State Electron. 1*, 170–178.

USAMI, K. AND HOROWITZ, M. 1995. Clustered voltage scaling technique for low-power design. In *Proceedings of the International Symposium on Low-Power Electronics and Design*. 3–8.

USAMI, K., IGARASHI, M., MINAMI, F., ISHIKAWA, T., KANZAWA, M., ICHIDA, M., AND NOGAMI, K. 1998. Automated low-power technique exploiting multiple supply voltages applied to a media processor. *IEEE J. Solid-State Circ. 33*, 3, 463–472.

WANG, F., XIE, Y., BERNSTEIN, K., AND LUO, Y. 2006. Dependability analysis of FinFET circuits. In *Proceedings of the Symposium on Emerging VLSI Technologies and Architectures*. 399–404.

WEI, L., CHEN, Z., AND ROY, K. 1997. Double-gate dynamic threshold voltage (DGDT) SOI MOSFETs for low-power high-performance designs. In *Proceedings of the International SOI Conference*. 82–83.

YU, B., CHANG, L., SHIBLY, A. ET AL. 2002. FinFET scaling to 10nm gate length. In *Proceedings of the International Electronic Device Meeting*. 251–254.

ZHANG, W., FOSSUM, J. G., MATHEW, L., AND DU, Y. 2005. Physical insights regarding design and performance of independent-gate FinFETs. *IEEE Elect. Dev. Lett. 52*, 10, 2189–2206.

ZHAO, W. AND CAO, Y. 2006. New generation of predictive technology model for sub-45nm design exploration. In *Proceedings of the International Symposium on Quality of Electronic Design*. 585–590. `http://www.eas.asu.edu/~ptm`.

ZHAO, W. AND CAO, Y. 2007. Predictive technology model for nano-CMOS design exploration. *ACM J. Emerg. Techn. Comput. Syst. 3*, 1, 1–17.

Received October 2008; revised February 2009; accepted March 2009 by Iris Bahar