Joint Sizing and Adaptive Independent Gate Control for FinFET Circuits Operating in Multiple Voltage Regimes Using the Logical Effort Method

Xue Lin, Yanzhi Wang, and Massoud Pedram
University of Southern California
Los Angeles, CA, 90089, US
{xuelin, yanzhiwa, pedram}@usc.edu

ABSTRACT
FinFET has been proposed as an alternative for bulk CMOS in current and future technology nodes due to more effective channel control, reduced random dopant fluctuation, high ON/OFF current ratio, lower energy consumption, etc. Key characteristics of FinFET operating in the sub/near-threshold region are very different from those in the strong-inversion region. This paper first introduces an analytical transregional FinFET model with high accuracy in both sub- and near-threshold regimes. Next, the paper extends the well-known and widely-adopted logical effort delay calculation and optimization method to FinFET circuits operating in multiple voltage (sub/near/super-threshold) regimes. More specifically, a joint optimization of gate sizing and adaptive independent gate control is presented and solved in order to minimize the delay of FinFET circuits operating in multiple voltage regimes. Experimental results on a 32nm Predictive Technology Model for FinFET demonstrate the effectiveness of the proposed logical effort-based delay optimization framework.

Categories and Subject Descriptors
B.7.1 [**Integrated Circuits**]: Types and Design Styles.

General Terms
Performance, Design.

Keywords
FinFET; delay optimization; sub/near-threshold; independent gate control; logical effort.

1. INTRODUCTION
Aggressive voltage scaling from the traditional super-threshold region to the sub/near-threshold region has been shown effectiveness in reducing energy consumption of digital circuits [1][2][3]. It is especially beneficial for applications with relaxed performance requirements, such as wireless sensor processing, medical monitoring. Authors of [4][5] proved the existence of the minimum energy point (MEP), which is the optimal supply voltage V_{DD} to minimize energy consumption of the digital circuits. They showed that the MEP for CMOS circuits typically occurs in the near-threshold region.

FinFET devices, a kind of special quasi-planar double gate (DG) devices, have been proposed as an alternative for the bulk CMOS when technology scales beyond the 32nm technology node [6][7]. It is proved that FinFET devices can enhance the energy efficiency, ON/OFF current ratio, and soft-error immunity compared with bulk CMOS counterparts. Figure 1 shows the MEP of a 20-stage FinFET inverter chain with different activity factors. We observe that the MEP of FinFET circuits lies in the subthreshold region, which is typically lower than that of the bulk CMOS circuits. Therefore, the FinFET outperforms bulk CMOS in the ultra-low power designs by allowing for higher voltagescalability. On the other hand, another important operation point of FinFET circuits, the optimal V_{DD} to minimize the energy-delay product, typically lies in near-threshold regime.

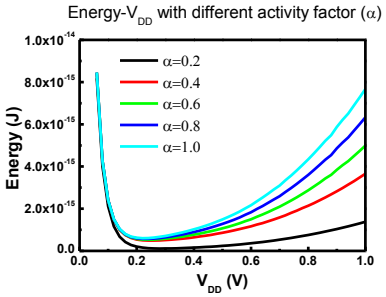

Fig. 1. MEP of a 20-stage FinFET inverter chain.

Figure 2 shows a double-gate FinFET device, where each fin contains two gates: a *front gate* and a *back gate*. Each fin is essentially the parallel connection of the *front-gate-controlled FET* and the *back-gate-controlled FET*, both with width *H* equal to the height of each fin. One unique feature of FinFET devices is the independent gate control, i.e., the front gate and the back gate of each fin can be controlled by different signals, which enables more energy saving margin and flexible circuit designs [10]. Previous work [9] utilized independent gate control for FinFETs in the pull-down network of an SRAM cell to keep the ~20 pA/um standby power budget, whereas the authors of [10][11] studied joint gate sizing and negative biasing on the back gate of FinFET devices and showed significant power reduction.

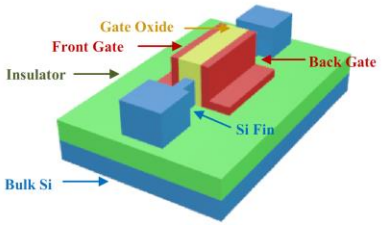

Fig. 2. Double-gate FinFET device.

Many burst-mode applications require high performance for brief time periods between extended sections of low performance operation [12]. Digital circuits supporting such burst-mode applications should work in both sub/near-threshold regime and super-threshold regime for brief time periods. A FinFET circuit optimized for strong-inversion regime may not be optimal for sub/near-threshold regime, and vice versa. Therefore, we need a better understanding of FinFET models in different voltage domains. First, we notice that the conventional FinFET models are expressed in a piecewise fashion with a breakpoint at or near the threshold voltage, separating the super-threshold regime where the α–power law model [16] applies and the sub-threshold regime

where the exponential dependency model [8] applies. We extend the empirical model from [15] for FinFET to provide a unified FinFET model covering both sub- and near-threshold regions.

Based on the accurate FinFET modeling, we employ the logical effort delay calculation and optimization method [14] for FinFET circuits operating in multiple voltage domains. The key of logical effort method is the derivation of the sizes of NMOS and PMOS transistors in a minimum-size inverter that achieves equal rise and fall times and using this as a template to derive the values of the logical effort and parasitic delays of complex logic gates. The authors of [13] extended the logical effort method to the bulk CMOS circuits in subthreshold region. Different from bulk CMOS devices, for the FinFET devices, widths are quantized into units of the fins and large width can only be obtained by using multiple fins. Due to this discretization effect of device width, it is difficult to derive FinFET inverters with equal rise and fall times as the templates in the logical effort method and therefore the logical effort method loses some simplicity and generality on the FinFET circuits. We solve this problem by employing adaptive independent gate control.

In this paper, we first derive the different FinFET templates with equal rise and fall times for all three operation regimes. We use adaptive independent gate control, i.e., applying different back gate biasing voltages at different supply voltage levels, for the FinFET template inverter design. Next, we derive the logical effort and parasitic delay values of arbitrarily sized (possibly with asymmetric rise and fall times) FinFET gates with independent gate control for all the operation regimes with respect to the corresponding template inverters. Using the extension of the logical effort-based delay optimization framework, we perform a joint optimization of gate sizing and adaptive independent gate control on FinFET circuits so that they can operate robustly in all the operation regimes, i.e., with the minimum *weighted delay* in all of sub-, near-, and super-threshold regimes. We propose a dynamic programming-based method to find the near-optimal solution of this problem in polynomial time complexity. Experimental results on HSpice simulation using 32nm Predictive Technology Model (PTM) for FinFETs [17] verify that the proposed improved logical effort-based optimization method provides a performance enhancement of up to 29.69% compared to the conventional method.To the best of our knowledge, this is the first paper that extends the logical effort delay calculation and optimization method for FinFET circuits operating in multiple voltage domains.

2. INDEPENDENT GATE CONTROL

FinFET devices show better suppression of the short channel effects, lower energy consumption, higher supply voltage scaling ability, and higher ON/OFF current ratio compared with the bulk CMOS counterparts [7][8]. In addition, the structure of FinFET allows for fabrication of separate front and back gates. In this structure, each fin is essentially the parallel connection of the *front-gate-controlled FET* and the *back-gate-controlled FET*, both with width *H* equal to the height of the fin. A double-gate fin in FinFET circuits has the following two connection modes: *double-gate mode* (both the front and back gates of the fin are tied to the same control signal) and *Independent-gate mode* (the front and back gates are tied to different control signals).

Independent gate control makes it possible to apply different voltages to the front and back gates of a single fin, and thereby, allowing for more flexible circuit designs. Due to capacitor coupling of the front gate and back gate, the threshold voltage of the front-gate-controlled FET varies in response to the back-gate biasing, and vice versa. Under relatively small back-gate biasing voltage, a linear relationship between the change of the threshold voltage and the back-gate biasing voltage is observed (suppose that we consider N-type FETs):

$$\frac{dV_{th}}{dV_{BN}} = -\frac{C_{oxb} \cdot C_{si}}{C_{oxf} \cdot (C_{oxb} + C_{si})} \quad (1)$$

where C_{si}, C_{oxf}, and C_{oxb} are the body capacitance, front gate capacitance, and back gate capacitance, respectively; V_{BN} is the bias voltage applied to the back gate of the N-type fin. Eqn. (1) shows that increasing the negative bias voltage on the back gate of the N-type fin (i.e., decreasing the back-gate biasing voltage) results in the increase of V_{th} of the front-gate-controlled N-type FET and therefore an exponential decrease of the leakage power. Figure 3 shows the relationship between the threshold voltage of the front-gate-controlled FET and the back gate biasing voltage of the N-type FinFET from HSpice simulation.

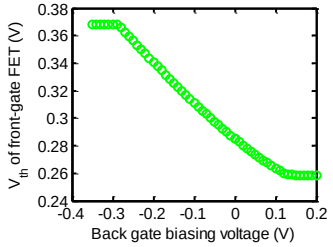

Fig. 3. Threshold voltage of the front-gate-controlled N-type FET v.s. back gate biasing voltage.

References [9][10][11] proposed and applied different implementation modes of FinFET logic gates to exploit the unique feature of independent gate control. We illustrate in Figure 4 the examples of a unit-sized inverter that achieves approximately equal rise and fall times in the super-threshold regime. In Mode (a) of Figure 4, the parallel P-type FETs are merged together, and the back gate of the N-type fin is tied to ground to achieve approximately equal rise and fall times. In Mode (b), forward or negative biasing is applied to the back gate of the N-type fin. In Mode (c), forward or negative biasing is applied to the back gate of P-type fin and the back gate of the N-type fin is tied to ground.

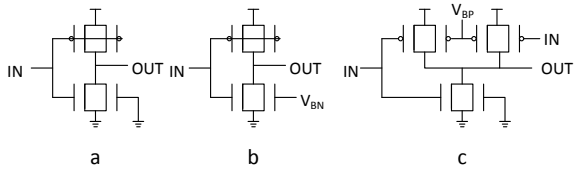

Fig. 4. Unit-sized inverters that achieve approximately equal rise and fall times in the super-threshold regime.

3. EMPIRICAL FINFET MODELING IN THE SUB/NEAR-THRESHOLD REGIME

The drain current I_{ds} of an N-type FinFET (say, the front-gate-controlled FET in an N-type fin) operating in the subthreshold regime obeys an exponential dependency on the gate drive voltage V_{gs} and drain-to-source voltage V_{ds}, as given by:

$$I_{ds} = I_0 \frac{W}{L} \cdot e^{\frac{V_{gs}+\lambda V_{ds}-V_{th}}{n \cdot v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}}\right) \quad (2)$$

Where I_0 is a technology-dependent parameter, λ is the drain voltage dependence coefficient (similar to but much smaller than the DIBL coefficient for bulk CMOS devices), n is the subthreshold slope factor, and v_T is the thermal voltage $\frac{kT}{q}$.

Figure 5 (green dots) plots the simulated curve of I_{ds} v.s. V_{DD} (we set $V_{gs} = V_{ds} = V_{DD}$) on a semi-logarithmic scale for an N-type FinFET with a threshold voltage V_{th} of 0.29 V. We can

observe that the curve is nearly straight for $V_{DD} < V_{th}$, corresponding to the exponential I-V relationship in the subthreshold region. The curve rolls off when $V_{DD} > V_{th}$. We extend the method of [15] for FinFETs to provide a unified transregional model covering both sub- and near-threshold regimes. In this model, the drain current I_{ds} is given as:

$$I_{ds} = I_0 \frac{W}{L} \cdot e^{\frac{(V_{gs}+\lambda V_{ds}-V_{th})-a\cdot(V_{gs}+\lambda V_{ds}-V_{th})^2}{n\cdot v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}}\right) \quad (3)$$

Where a is an empirical fitting parameter. We can extract the values of parameters I_0, a, and n from HSpice simulation.

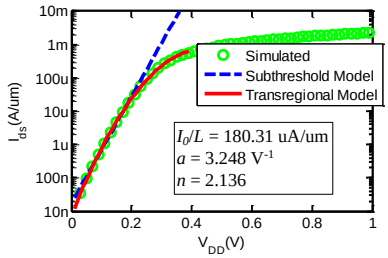

Fig. 5. The simulated and model curves of I_{ds} v.s. V_{DD}.

Over the sub- and near-threshold operation range of 0.05 V to 0.35 V ($V_{th} = 0.29$ V for N-type FinFETs), the transregional model (the red curve in Figure 5) results in an average error of 4.27% and a maximum error of 8.83% compared with HSpice simulation. The extracted parameters I_0, a, and n are labeled in Figure 5. If the parameter a is forced to be 0, we can reduce to the subthreshold model (2) (the blue curve in Figure 5). Over the subthreshold operation range of 0.05 V to 0.20 V, it results in an average error of 6.59% and a worst-case error of 15.65%. We can observe from Figure 5 that the transregional model (3) is even more accurate than the subthreshold model (2) in the subthreshold region. In summary, the transregional model provides accurate FinFET modeling in both sub- and near-threshold regions.

4. FINFET TEMPLATE INVERTER

Logical effort based delay calculation [14] is a simple and effective way to both estimate and optimize the delay of digital CMOS circuits. In this section, we first discuss the sizing of a FinFET template inverter to achieve equal rise and fall times without independent gate control. We will show that this is often infeasible. Next, we will derive the template inverter with equal rise and fall times using adaptive independent gate control.

4.1 Sizing of FinFET Template Inverter

For a FinFET inverter, let W_P and W_N denote the total widths of the P-type FETs and N-type FETs whose gates are tied to the input signal. In this subsection, we derive the W_P: W_N ratio that achieves equal rise and fall times for the template inverter at different V_{DD} level. In the super-threshold regime, the desirable W_P: W_N ratio is around 2.24, obtained by HSpice simulations.

In the sub/near-threshold regime, let $V_{gs} = V_{ds} = V_{DD}$, and then we have $e^{\frac{-V_{ds}}{v_T}} = e^{\frac{-V_{DD}}{v_T}} \approx 0$. We derive the desirable W_P: W_N ratio in the sub/near-threshold regime analytically using (3):

$$\frac{W_P}{W_N} = \frac{I_{0,N} \cdot e^{\frac{(V_{DD}+\lambda_N V_{DD}-V_{th,N})-\alpha_N\cdot(V_{DD}+\lambda V_{DD}-V_{th,N})^2}{n_N\cdot v_T}}}{I_{0,P} \cdot e^{\frac{(V_{DD}+\lambda_P V_{DD}-V_{th,P})-\alpha_P\cdot(V_{DD}+\lambda V_{DD}-V_{th,P})^2}{n_P\cdot v_T}}} \quad (4)$$

Using the 32nm PTM for FinFETs [17], the derived W_P: W_N ratios under different V_{DD} values for the subthreshold regime ($V_{DD} = 0.2$ V), near-threshold regime ($V_{DD} = 0.3$ V), and super-threshold regime ($V_{DD} = 1$ V) are listed in Table 1. We observe that the analytical results match the simulation results well.

Table 1. W_P: W_N ratios under different V_{DD} levels.

V_{DD} (V)	0.2	0.3	1.0
W_P: W_N(calculated)	1.01	1.69	---
W_P: W_N(simulated)	1.00	1.76	2.24

Please note that in the actual sizing of FinFET devices, the exact desirable W_P: W_N ratio may not be achieved because of the device-width quantization. Therefore, we propose an alternative method to derive the template inverter with equal rise and fall times using adaptive independent gate control under different operation regimes.

4.2 FinFET Template Inverter Design Using Independent Gate Control

In this subsection, we derive the FinFET template inverter with equal rise and fall times using adaptive independent gate control, i.e., applying different back gate biasing voltages at different V_{DD} levels. In this paper, we assume that forward or reverse independent gate control is only applied to the N-type FETs to reduce the complexity andthe area overhead. Hence, the template inverter is implemented as shown in Figure 4 (b). In the following, we derive the biasing voltage V_{BN} given V_{DD} in order to achieve balanced rise and fall times.

When $V_{DD} = 0.3$ V, the FinFET circuit operates in the near-threshold regime. Since the desirable W_P: W_N ratio is 1.76, we need to apply forward independent gate control ($V_{BN} > 0$) to achieve equal rise and fall times. Let $V_{th,N}(V_{BN})$ denote the threshold voltage of the N-type front-gate-controlled FET given biasing voltage V_{BN} on the back gate. We define $V_{th,P}(V_{BP})$ similarly. We derive the desirable V_{BN} value, denoted by $V_{BN,near}^{tem}$, through solving Eqn. (5). Please note that the currents of both the front-gate-controlled FET and the back-gate-controlled FET should be accounted for since forward independent gate control is applied.

$$I_{0,N} \frac{W}{L} \cdot e^{\frac{\left(V_{DD}-V_{th,N}(V_{BN})\right)-\alpha_N\cdot\left(V_{DD}-V_{th,N}(V_{BN})\right)^2}{n_N\cdot v_T}} + I_{0,N} \frac{W}{L} \cdot e^{\frac{\left(V_{BN}-V_{th,N}(V_{DD})\right)-\alpha_N\cdot\left(V_{BN}-V_{th,N}(V_{DD})\right)^2}{n_N\cdot v_T}} = 2 \cdot I_{0,P} \frac{W}{L} \cdot e^{\frac{\left(V_{DD}-V_{th,P}(V_{BP}=0)\right)-\alpha_P\cdot\left(V_{DD}-V_{th,P}(V_{BP}=0)\right)^2}{n_P\cdot v_T}} \quad (5)$$

Similarly, we obtain the bias voltage V_{BN} value when $V_{DD} = 0.2$ V, denoted by $V_{BN,sub}^{tem}$, for equal rise and fall times.

On the other hand, when $V_{DD} = 1$ V, the FinFET circuit operates in the super-threshold regime. Since the desirable W_P: W_N ratio is 2.24, we need to apply reverse independent gate control ($V_{BN} < 0$) to achieve equal rise and fall times. The desirable V_{BN} value is denoted by $V_{BN,super}^{tem}$ in this case. Results are shown in Table 2.

Table 2. V_{BN} values for the template inverter to achieve equal rise and fall times under different operation regimes.

$V_{DD} = 0.2$ V	$V_{DD} = 0.3$ V	$V_{DD} = 1.0$ V
$V_{BN,sub}^{tem} = 0.08$ V	$V_{BN,near}^{tem} = 0.03$ V	$V_{BN,super}^{tem} = -0.26$ V

5. LOGICAL EFFORT FOR FINFET CIRCUITS AND APPLICATION

Logical effort-based delay calculation and optimization method relies on the computation of the logical effort and parasitic delay values of logic cells. More specifically, the gate delay is modeled as $d = ghb + p$, where g is the *logical effort*,

*h*is the *electrical effort*, *b* is the *branching factor* that accounts for off-path capacitance, and *p* is the *parasitic delay*. Logical effort is defined as the ratio of the input capacitance of a gate to that of an inverter delivering the same amount of output current (related to its resistance.) The electrical effort represents the ratio of output capacitance to input capacitance. The ghb product is called the stage effort. The parasitic delay is defined as the delay of a gate driving no load. This value is set by the parasitic capacitance. In Section 5.1, we will derive the logical effort and parasitic delay values of arbitrarily sized (possibly asymmetric) FinFET gates with independent gate control for all the operation regimes.

Next, we will apply the proposed FinFET logical effort method to delay optimization for FinFET circuits operating in multiple operation regimes. In Section 5.2, we will introduce a joint optimization of gate sizing and adaptive independent gate control on a FinFET circuit so that it can achieve minimum *weighted delay* during its operation in multiple voltage regimes, based on the proposed FinFET logical effort method.

5.1 Logical Effort of FinFET Gates with Independent Gate Control

Consider the three FinFET template inverters. Let C_g and C_d denote the gate capacitance and diffusion capacitance of a single (front-gate-controlled or back-gate-controlled) FET with width H. We have $C_g: C_d \approx 4:3$ for the 32 nm PTM for FinFETs. Let R_{sub}, R_{near}, and R_{super} denote the effective resistances of the pull-down and pull-up network of the three template inverters in the corresponding operation regimes, respectively.

Consider a (possibly asymmetric) FinFET inverter with y parallel connected N-type fins and z parallel connected P-type fins. We assume that forward or reverse independent gate control is only applied to the N-type FETs in order to reduce the area overhead, and of course, the bias voltage is the same for all the FinFET gates in the same circuit block. The front gate and back gate of every P-type fin in the circuit block are tied together to the input signal, as shown in Figure 4 (b), so as to reduce the circuit design complexity. Let $V_{BN,sub}$, $V_{BN,near}$, and $V_{BN,super}$ denote the applied bias voltages on the N-type FETs operating in the subthreshold, near-threshold and super-threshold regimes, respectively. The bias voltage values can be positive (forward independent gate control) or negative (reverse independent gate control.) Let $s_{sub} \cdot R_{sub}$, $s_{near} \cdot R_{near}$, and $s_{super} \cdot R_{super}$ denote the effective resistances of an N-type fin in the three operation regimes, respectively, when the effect of independent gate control is considered. The values of s_{sub}, s_{near}, and s_{super} are functions of $V_{BN,sub}$, $V_{BN,near}$, and $V_{BN,super}$, respectively. Note that $s_{sub} = 1$ when $V_{BN,sub} = V_{BN,sub}^{tem}$. The effective resistances of the pull-down network of the inverter in the three regimes are given by $s_{sub} \cdot R_{sub}/y$, $s_{near} \cdot R_{near}/y$, and $s_{super} \cdot R_{super}/y$, respectively.

For such a FinFET inverter with specific values of y, z, s_{sub}, s_{near}, and s_{super}, we derive its logical effort and parasitic delay values for all of the three operation regimes. As shown in Figure 6, we use different template inverters for different operation regimes. Let $g_{f,sub}$ denote the falling logical effort of the FinFET inverter in the subthreshold regime with respect to the subthreshold template inverter. Similarly, we define $g_{f,near}$ and $g_{f,super}$. We calculate $g_{f,sub}$ using:

$$g_{f,sub} = \frac{s_{sub} \cdot R_{sub}/y \cdot (y+2z)C_g}{R_{sub} \cdot 3C_g} = \frac{(y+2z)s_{sub}}{3y} \qquad (6)$$

Figure 6 summarizes all the logical effort and parasitic delay values. Similarly, we also derive the logical effort and parasitic delay values of 2-input NAND gate and 2-input NOR gate for the sub/near/super-threshold regimes. A stack of more than two FinFETs may not be favored in sub/near-threshold operation. Thus gates with fan-in values larger than two are not considered. Details are omitted due to space limitation.

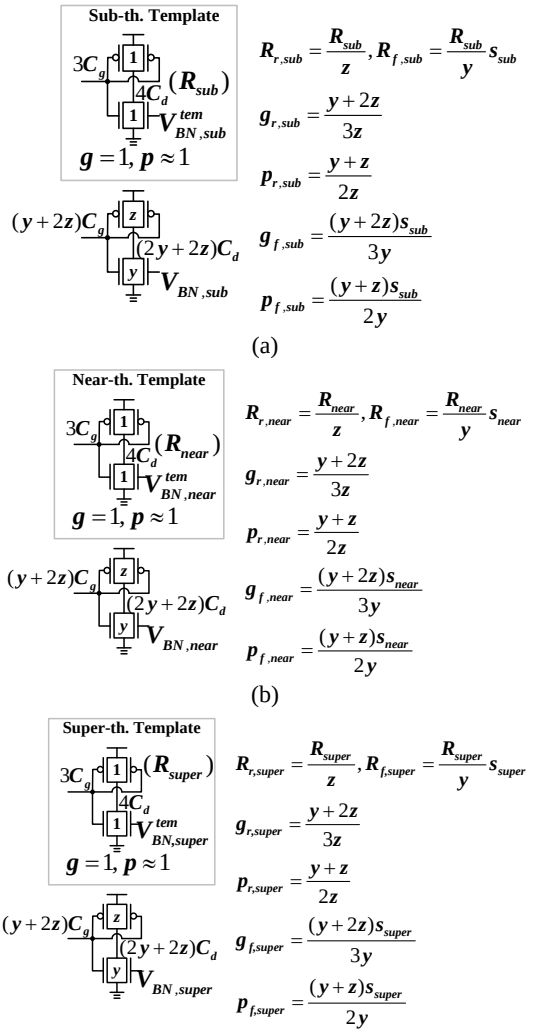

Fig. 6. Logical effort and parasitic delays of an arbitrarily sized inverter in sub/near/super-threshold regions.

5.2 Joint Optimization for FinFET Circuits

We use the super buffer as an example to demonstrate the proposed joint optimization framework that optimizes a FinFET circuit such that it has reasonable delays in all of sub/near/super-threshold regions. First, we define the weighted delay as the performance metrics. Let F_{sub}, F_{near}, and F_{super} denote the portions of cycles when the super buffer operates in the sub/near/super-threshold regimes, respectively. We have $F_{sub} + F_{near} + F_{super} = 1$. The weighted delay of the super buffer is

$$F_{sub}\frac{delay_{sub}}{delay_{sub,ref}} + F_{near}\frac{delay_{near}}{delay_{near,ref}} + F_{super}\frac{delay_{super}}{delay_{super,ref}} \qquad (7)$$

where $delay_{sub}$, $delay_{near}$, and $delay_{super}$ are the delay values of the super buffer in the sub/near/super-threshold operations, respectively. In Eqn. (7), we normalize $delay_{sub}$ by $delay_{sub,ref}$,

which is the delay of a super buffer optimized for subthreshold operation only. Similarly, $delay_{near}$ is normalized by $delay_{near,ref}$, and $delay_{super}$ is normalized by $delay_{super,ref}$. If we define the weighted delay as $F_{sub} \cdot delay_{sub} + F_{near} \cdot delay_{near} + F_{super} \cdot delay_{super}$ instead of Eqn. (7), minimizing the weighted delay of the super buffer is almost equivalent to minimizing the delay of the super buffer for subthreshold operation only, since $delay_{sub}$ is orders-of-magnitude larger than $delay_{near}$ and $delay_{super}$.

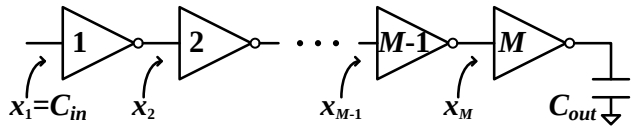

Fig. 7. An *M*-stage super buffer.

Figure 7 shows a FinFET super buffer with M stages and each i-th inverter is comprised of y_i parallel connected N-type fins and z_i parallel connected P-type fins. y_i's and z_i's are integer values. The input capacitance and output capacitance are given by C_{in} and C_{out}, respectively. Let $x_i (1 \le i \le M)$ denote the input capacitance of the i-th inverter in the super buffer. We have $x_1 = C_{in}$ and $x_{M+1} = C_{out}$. Each $x_i = (y_i + 2z_i)C_g$ and takes discrete values. Let $D_{r,sub}$ and $D_{f,sub}$ denote the rise and fall delays, respectively, of the super buffer in the subthreshold operation with respect to the subthreshold template inverter. Similarly, we define $D_{r,near}$ and $D_{f,near}$, and $D_{r,super}$ and $D_{f,super}$. These values can be calculated as follows based on the logical effort and parasitic delay values derived in Section 5.1. We provide in (8) and (9) the calculation of $D_{r,sub}$ and $D_{f,sub}$, respectively, as an example.

$$D_{r,sub} = \sum_{i=1}^{M/2}\left(\frac{x_{2i+1}}{x_{2i}} \cdot \frac{(y_{2i} + 2z_{2i})}{3z_{2i}} + \frac{(y_{2i} + z_{2i})}{2z_{2i}}\right) + \sum_{i=1}^{M/2}\left(\frac{x_{2i}}{x_{2i-1}} \frac{(y_{2i-1} + 2z_{2i-1})s_{sub}}{3y_{2i-1}} + \frac{(y_{2i-1} + z_{2i-1})s_{sub}}{2y_{2i-1}}\right) \quad (8)$$

$$D_{f,sub} = \sum_{i=1}^{M/2}\left(\frac{x_{2i}}{x_{2i-1}} \cdot \frac{(y_{2i-1} + 2z_{2i-1})}{3z_{2i-1}} + \frac{(y_{2i-1} + z_{2i-1})}{2z_{2i-1}}\right) + \sum_{i=1}^{M/2}\left(\frac{x_{2i+1}}{x_{2i}} \cdot \frac{(y_{2i} + 2z_{2i})s_{sub}}{3y_{2i}} + \frac{(y_{2i} + z_{2i})s_{sub}}{2y_{2i}}\right) \quad (9)$$

Note that $D_{r,sub}$ and $D_{f,sub}$ are calculated with respect to the subthreshold template inverter, and they are relative delay values. $D_{r,near}$, $D_{f,near}$, $D_{r,super}$, and $D_{f,super}$ are calculated in the similar way. Therefore, minimizing the following Eqn. (10) is equivalent to minimizing Eqn. (7). Then, the joint optimization problem for the minimum weighted delay is formulated as follows.

Given: F_{sub}, F_{near}, F_{super}, C_{in} and C_{out}.

Find: The optimal values of M, s_{sub}, s_{near}, s_{super}, y_i's and z_i's.

Minimize:

$$F_{sub} \cdot \max\{D_{r,sub}, D_{f,sub}\} + F_{near} \cdot \max\{D_{r,near}, D_{f,near}\} + F_{super} \cdot \max\{D_{r,super}, D_{f,super}\} \quad (10)$$

Subject to the following constraints:

The bias voltage constraints:

$$s_{sub,min} \le s_{sub} \le s_{sub,max} \quad (11)$$

$$s_{near,min} \le s_{near} \le s_{near,max} \quad (12)$$

$$s_{super,min} \le s_{super} \le s_{super,max} \quad (13)$$

The balancing constraints for each inverter:

$$BC_{sub,min} \le \frac{z_i s_{sub}}{y_i} \le BC_{sub,max} \quad (14)$$

$$BC_{near,min} \le \frac{z_i s_{near}}{y_i} \le BC_{near,max} \quad (15)$$

$$BC_{super,min} \le \frac{z_i s_{super}}{y_i} \le BC_{super,max} \quad (16)$$

We propose a dynamic programming-based algorithm to find the near-optimal solution of the joint optimization problem. The near-optimal solution is comprised of an outer loop and an inner loop. The outer loop finds the near-optimal values of M, s_{sub}, s_{near}, and s_{super}, using optimization algorithms such as the *gradient descent algorithm*. The inner loop finds the near-optimal values of y_i's and z_i's with given values of M, s_{sub}, s_{near}, and s_{super}. In the inner loop, we maintain matrices $\mathbf{D}_{r,sub}$, $\mathbf{D}_{f,sub}$, $\mathbf{D}_{r,near}$, $\mathbf{D}_{f,near}$, $\mathbf{D}_{r,super}$, $\mathbf{D}_{f,super}$, $\mathbf{Y}$, and $\mathbf{Z}$. As an example, $\mathbf{D}_{r,sub}(i, x_{i+1})$ stores the near-optimal rise delay from inverter 1 to inverter i in the sub-threshold regime, with given input capacitance x_{i+1} of the $(i+1)$-th inverter. Similarly, we define the values stored in matrices $\mathbf{D}_{f,sub}$, $\mathbf{D}_{r,near}$, $\mathbf{D}_{f,near}$, $\mathbf{D}_{r,super}$, and $\mathbf{D}_{f,super}$.

Suppose that we want to find the near-optimal delay values from inverter 1 to inverter i with given x_{i+1}. We already know the values of $\mathbf{D}_{r,sub}(i-1, x_i)$ (and also other matrices) for various x_i values. Therefore, we only need to find the optimal values of y_i and z_i in this calculation (we have $x_i = (y_i + 2z_i)C_g$). We provide more details in the following. When y_i and z_i are given, the fall delay of inverter i in the subthreshold regime with given x_{i+1} is calculated by:

$$D_{f,sub,i}(y_i, z_i, x_{i+1}) = \frac{x_{i+1}}{x_i} \cdot \frac{(y_i + 2z_i)s_{sub}}{3y_i} + \frac{(y_i + z_i)s_{sub}}{2y_i} \quad (17)$$

Similarly, we define and calculate $D_{r,sub,i}(y_i, z_i, x_{i+1})$, $D_{f,near,i}(y_i, z_i, x_{i+1})$, $D_{r,near,i}(y_i, z_i, x_{i+1})$, $D_{f,super,i}(y_i, z_i, x_{i+1})$, and $D_{r,super,i}(y_i, z_i, x_{i+1})$. We find the optimal values of y_i and z_i such that the following weighted delay is minimized:

$$F_{sub} \cdot \max\begin{Bmatrix}\mathbf{D}_{r,sub}(i-1, x_i) + D_{f,sub,i}(y_i, z_i, x_{i+1})\\ \mathbf{D}_{f,sub}(i-1, x_i) + D_{r,sub,i}(y_i, z_i, x_{i+1})\end{Bmatrix} + F_{near} \cdot \max\begin{Bmatrix}\mathbf{D}_{r,near}(i-1, x_i) + D_{f,near,i}(y_i, z_i, x_{i+1})\\ \mathbf{D}_{f,near}(i-1, x_i) + D_{r,near,i}(y_i, z_i, x_{i+1})\end{Bmatrix} + F_{super} \cdot \max\begin{Bmatrix}\mathbf{D}_{r,super}(i-1, x_i) + D_{f,super,i}(y_i, z_i, x_{i+1})\\ \mathbf{D}_{f,super}(i-1, x_i) + D_{r,super,i}(y_i, z_i, x_{i+1})\end{Bmatrix} \quad (18)$$

And store the optimal y_i and z_i values in $\mathbf{Y}(i, x_{i+1})$ and $\mathbf{Z}(i, x_{i+1})$, respectively. We calculate the values of $\mathbf{D}_{r,sub}(i, x_{i+1})$, $\mathbf{D}_{f,sub}(i, x_{i+1})$, $\mathbf{D}_{r,near}(i, x_{i+1})$, etc. correspondingly. Details are omitted. We perform *trace back* [18] after filling out the matrices up to the last inverter of the super buffer, and find the near-optimal y_i and z_i values for every inverter.

The procedure of the proposed near-optimal algorithm is omitted due to space limitation. This joint optimization framework can be applied to other circuit structures, for example, critical path sizing. We will show the effectiveness of this framework on both a super buffer and an n-input AND function (a common circuit structure in Memory decoders).

6. EXPERIMENTAL RESULTS

We test our joint optimization framework with the super buffer structure first. We perform simulations on the 32nm PTMfor FinFETs. The supply voltages in sub/near/super-threshold regimes are 0.2V, 0.3V, and 1V, respectively. For given

F_{sub}, F_{near}, F_{super}, C_{in} and C_{out} values, we find the near-optimal values of M, s_{sub}, s_{near}, s_{super}, y_i's and z_i's using the dynamic programming based algorithm. Then we simulate the optimized super buffer using HSpice. M iscorresponding to the number of stages. s_{sub}, s_{near}, ands_{super} are translated into back-gate biasing voltages in sub/near/super-threshold regimes, respectively. Note that in one operation regime, all the N-type FinFETs in the super buffer share one common back-gate biasing voltage. y_i's and z_i's are the numbers of fins in the N-type and P-type FinFETs of the i-th stage inverter. Baseline is a super buffer designed using inverters in the form of Figure 4 (a), where the numbers of fins in the pull-up and pull-down networks are equal, and optimized for the minimum weighted delay using the conventional logical effort method. Then we compare the weighted delay of the super buffer optimized with the proposed method and that of the baseline. The results are summarized in Table 3. The weighted delay of the proposed super buffer is normalized by that of the baseline. The proposed improved logical effort-based optimization method provides a performance enhancement of up to 29.54% over the conventional method.

Table 3. The normalized weighted delay values of the proposed super buffer and the baseline.

Experimental Setup			Weighted Delay	
C_{in}	C_{out}	Stage Num.	Proposed	Baseline
6	100	2	0.7829	1
6	500	4	0.7624	1
6	1000	4	0.7315	1
6	3000	6	0.7046	1
3	3000	6	0.7673	1
3	10000	6	0.7216	1

We further test our joint optimization framework with the n-input AND function, which is commonly used in Memory decoders and can be realized with a NAND-NOR structure. For given F_{sub}, F_{near}, F_{super}, C_{in} and C_{out} values, we find the near-optimal values of s_{sub}, s_{near}, s_{super}, y_i's and z_i's using the dynamic programming based algorithm. The number of stages M is determined by the input number n, and therefore M is no longer an optimization variable. The baseline is an$n$-input AND function optimized for the minimum delay in the near-threshold operation only (the back-gate biasing is set as zero.) Then we compare the weighted delay of the AND function optimized with the proposed method to that of the baseline. The results are summarized in Table 4, showing that the proposed optimization method provides a performance enhancement of up to 29.69% over the baseline method.

Table 4. The normalized weighted delay values of the proposed AND function and the baseline.

	Proposed	Baseline
64-input AND	0.7076	1
256-input AND	0.7031	1

7. CONCLUSION

This is the first paper that presented a new logical effort calculation and optimization framework of FinFET circuits operating in all of the subthreshold, near-threshold, and super-threshold regimes. The characteristics of FinFETs operating in the sub/near-threshold regime are very different from those in the strong-inversion region. First, we introduced an analytical transregional FinFET model with high accuracy in both sub- and near-threshold regimes. Based on the accurate model, we proposed: (i) derivation of different FinFET template inverters for all three operation regimes to achieve equal rise and fall times, utilizing independent gate control; (ii) calculation of logical effort and parasitic delay values of arbitrarily sized FinFET gates with independent gate control for all operation regimes. We extended the logical effort method for delay optimization of FinFET circuits operating in all three operation regimes. We proposed and solved a joint optimization of gate sizing and adaptive independent gate control in order to achieve this goal.

8. ACKNOWLEDGMENTS

This research is sponsored in part by grants from the Defense Advanced Research Projects Agency and the National Science Foundation.

9. REFERENCES

[1] R. Dreslinski, M. Wiekowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-threshold computing: reclaiming Moore's law through energy efficient integrated circuits," *Proc. of IEEE*, 2010.

[2] D. Markovic, C. Wang, L. Alarcon, T. Liu, and J. Rabaey, "Ultralow-power design in near-threshold region," *Proc. of IEEE*, 2010.

[3] A. Wang and A. Chandrakasan, "A 180 mV FFT processor using circuit techniques,"*ISSCC*, 2004.

[4] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and practical limits of dynamic voltage scaling,"*DAC*, 2004.

[5] B. H. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE J. Solid-State Circuits*, 2005.

[6] E. J. Nowak, I. Aller, T. Ludwig, K. Kim, R. V. Joshi, C.-T. Chuang, K. Bernstein, and R. Puri, "Turning silicon on its edge," *IEEE Circuits and Devices Magazine*, 2004.

[7] T. Sairam, W. Zhao, and Y. Cao, "Optimizing FinFET technology for high-speed and low-power design", *GLSVLSI*, 2007.

[8] F. Crupi, M. Alioto, J. Franco, P. Magnone, M. Togo, N. Horiguchi, and G. Groeseneken, "Understanding the basic advantages of bulk FinFETs for sub- and near-threshold logic circuits from device measurements," *IEEE Trans. on Circuits and Systems II*, 2012.

[9] T. Cakici, K. Kim, and K. Roy, "FinFET based SRAM design for low standby power applications,"*ISQED*, 2007.

[10] A. Muttreja, N. Agarwal, and N. K. Jha, "CMOS logic design with independent gate FinFETs,"*ICCD*, 2007.

[11] J. Ouyang and Y. Xie, "Power optimization for FinFET-based circuits using genetic algorithms,"*ISOCC*, 2008.

[12] B. H. Calhoun, A. Wang, N. Verma, and A. Chandrakasan, "Sub-threshold design: the challenges of minimizing circuit energy," *ISLPED*, 2006.

[13] J. Keane, H. Eom, T.-H. Kim, S. Sapatnekar, and C. Kim, "Subthreshold logical effort: a systematic framework for optimal subthreshold device sizing,"*DAC*, 2006.

[14] I. Sutherland, B. Sproull, and D. Harris, *Logical Effort: Designing Fast CMOS Circuits*. San Francisco, CA: Morgan Kaufmann,1999.

[15] D. M. Harris, B. Keller, J. Karl, and S. Keller, "A transregional model for near-threshold circuits with application to minimum-energy operation,"*ICM*, 2010.

[16] T. Sakurai and R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, 1990.

[17] W. Zhao and Y. Cao, "New generation of Predictive Technology Model for sub-45nm early design exploration," *IEEE Transactions on Electronic Devices*, 2006.

[18] R. Rivest, T. Cormen, C. Leiserson, An introduction to algorithms. *MITPress*, 1990.