Gate Sizing: FinFETs vs 32nm Bulk MOSFETs*

Brian Swahn *and* Soha Hassoun
Tufts University
Medford, MA 02155
swahn@ece.tufts.edu, soha@cs.tufts.edu

ABSTRACT

FinFET devices promise to replace traditional MOSFETs because of superior ability in controlling leakage and minimizing short channel effects while delivering a strong drive current. We investigate in this paper gate sizing of finFET devices, and we provide a comparison with 32nm bulk CMOS. Wider finFET devices are built utilizing multiple parallel fins between the source and drain. Independent gating of the finFET's double gates allows significant reduction in leakage current. We perform temperature-aware circuit optimization by modeling delay using temperature-dependent parameters, and by imposing constraints that limit the maximum allowable number of parallel fins. We show that finFET circuits are superior in performance and produce less static power when compared to 32nm circuits.

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

General Terms: Design

Keywords: FinFET, thermal modeling, gate sizing

1. INTRODUCTION

The scaling of the MOSFET transistor has delivered astronomical increases in transistor density and performance, leading to more chip functionality at higher speeds. The main roadblock to continued success is the leakage phenomenon. Increased leakage stems from decreased oxide thicknesses, higher substrate dopings, and decreased channel lengths. A lowered threshold voltage to obtain better performance at lowered operating voltages further exacerbates the leakage problem. Moreover, the continued-shrinking proximity of the source and drain reduces the effective control of the gate over the channel, accentuating DIBL, drain-induced-barrier lowering. The ITRS predicts that static power dissipation per device will surpass the dynamic power dissipation by 2007.

To reduce leakage while scaling performance, the 2003 ITRS predicts using strained silicon channels, ultra-thin single-gate FETs, and metallic gates. The ITRS predicts ultimately moving towards double-gate (and multi-gate) devices. From a circuit perspective, the double-gate devices operate in a manner similar to MOSFETs.

*This research was supported by NSF grant CCF-0429921, and a gift from the Altera Corp.

DAC 2006, July 24–28, 2006, San Francisco, California, USA.

When a potential larger than the threshold voltage, V_T, is applied between the gates of the DG device and source, current flows from the drain to the source. The double gates, however, allows modulating the channel from two sides instead of one. The two gates together strongly influence the channel potential, combating the drain impact, and leading to the better ability to shut off the channel current. DIBL is thus reduced, and the swing is improved.

While several double-gate transistor variations have been investigated, recent research efforts have focused on finFETs, illustrated in Figure 1(a). The finFET, originally dubbed as the folded-channel MOSFET [10], consists of a narrow vertical fin that sticks up from the wafer surface (z-direction). Source and drain terminals are built at opposite ends of the fin. The gate, whose width is twice the fin height, drapes over the fin. The current flow in the channel is parallel to the plane of the wafer, and thus the label quasi-planar despite what appears to be a non-planar fin. Other multi-gated planar and non-planar devices have been proposed, however, finFETs, are a likely contender because of cost-effective manufacturing, the natural alignment of the double gate, and routability of the gate.

Figure 1(a) illustrates key geometric parameters for a finFET. The distance between the source and drain is referred to as the gate length, L_{gate}. The fin height, H_{fin}, is uniform for all fins on chip. A larger height complicates processing and causes defects. The oxide thickness between the side gates and the fin is t_{ox}. The oxide thickness between the top gate and the fin is t_{ox-top}. W_{fin} is the fin thickness. Fin engineering (balancing height, fin thickness, oxide thickness, and channel length) is essential in minimizing the leakage current, I_{off}, and maximizing the on current, I_{on}.

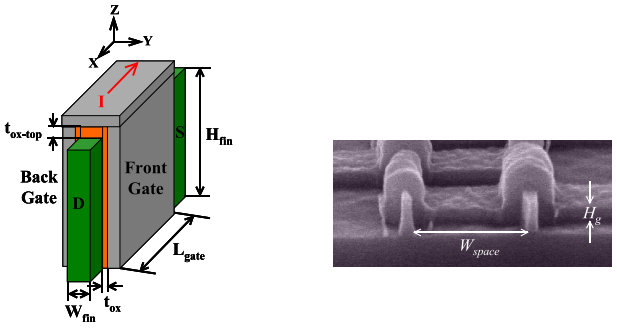
We investigate in this paper issues in finFET sizing and independent gate biasing of the front and back gates while considering thermal constraints. FinFET sizing is challenging because wider devices are created using multiple fins. Device-width quantization thus must be considered [3]. FinFET sizing then consists of finding the optimal number of parallel fin for each gate in the circuit. Independent biasing of one of the finFET gates provides lower leakage yet performance is affected. Independent biasing must thus be judiciously used. We use an NLP-based heuristic to solve the sizing problem, and we provide a detailed comparison with 32nm bulk CMOS devices. For this investigation, we design our base device to have the following characteristics, typical of recent manufactured finFETs: $L_{gate} = 45nm$; $H_{fin} = 65nm$; $W_{fin} = 10nm$; $t_{ox} = 1.6nm$. We use the UFDG SPICE models [8] to generate our finFET data and the 32nm Predictive Technology Modeling (PTM) models [14, 17] to generate our 32nm bulk data.

We begin with a discussion of issues of sizing finFETs vs 32nm devices. We then present the problem formulation and the solution. We conclude with experimental results.

2. ISSUES IN SIZING: FINFETS VS 32NM

2.1 Discrete vs. Continuous Sizing

The width of a finFET (i.e. the width of the area controlled by the gates), is defined as: $W_{gate} = 2 \times H_{fin}$. Wider finFETs are formed



(a) FinFET device. (b) Multi-fin device [3].

Figure 1: FinFET device geometries.

by draping the gate across multiple fins between the source and drain region, as shown in Figure 1(b) [3]. These fins are tightly laid out to minimize area and gate resistance. Fin sizing is then a discrete optimization problem, with an n-fin device delivering about n times the current delivered by a single-fin device, and introducing n times the load due to a single-fin device.

For traditional MOSFETs, the issue of discrete vs. continuous transistor sizing has been thoroughly investigated. While library gate sizes are discrete, continuous optimization techniques are presumably faster and they are often applied. The solution is then converted to discrete sizes. Provided a rich set of library, the performance impact of using this method is within 2%-7% or less [7,9] of discrete sizing. Furthermore, optimal gate sizes found using continuous techniques can be constructed by combining gates from a standard library with high accuracy [9]. As we will see in Section 3, we propose an approximating heuristic to solve the finFET sizing problem while using a non-linear programming formulation and solver to find optimal sizes for the devices in the 32nm circuits.

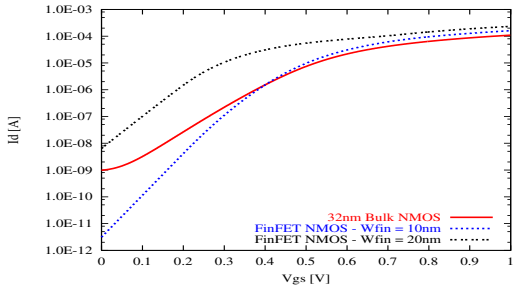


Figure 2: FinFET and 32nm I-V characteristics.

2.2 I_{on} and I_{off}

Figure 2 illustrates the on and off currents for a minimally sized n-type 32nm device vs. an n-type finFET ($W_{fin} = 10nm$). FinFETs provide more *on* current than the 32nm devices, resulting in faster switching times. Bulk CMOS has several leakage mechanisms: subthreshold leakage, gate leakage, reverse-biased junction, band-to-band tunneling (BTBT), and gate-induced drain leakage [13]. In a finFET, there are two leakage mechanisms: the subthreshold current and gate leakage [2]. Figure 2 also shows I_{on} and I_{off} for different fin thicknesses; both I_{on} and I_{off} increase as the fin thickness increases. The increase in I_{on} is due to a reduction in parasitic source/drain resistances, while decreased gate control over the channel causes an increase in I_{off} .

2.3 Gate vs. Body Biasing

For bulk CMOS, energy reduction is achieved via adaptive body biasing (ABB) and multiple-threshold MOSFETs. ABB research showed that adjusting the body voltage is an effective method for post-silicon tuning to reduce V_T variability under process variation, reducing leakage current and thus energy, and tripling the accepted die count in the highest frequency bin [16]. ABB was shown to be most effective when combined with adaptive supply voltage control.

FinFETs can provide some operating bias similar to that achieved via ABB. The front and back gates of a finFET can be made electrically independent (no top oxide is deposited). By applying different potentials to the gates, both the threshold voltage of the device and the leakage current can be changed. Fried and his colleagues describe such a device [6]. The independent biasing of a multi-fin device slows the device and results in added net area to allow for routing the two gate signals.

2.4 Temperature Dependence

While finFETs provide excellent electrostatic characteristics, they suffer from significant self-heating. The small and confined dimensions of the fin reduce the thermal conductivity (which increases the thermal resistance) of the device due to reduced phonon mean free path [11]. Heat transport out of the device is hindered, and the device temperature rises. FinFET thermal problems are further exacerbated with the construction of wider finFETs. In this work we assume a specific operating temperature. To ensure that none of the multi-fin finFETs operate at a higher temperature, we place a thermal sizing constraint. Our sizing algorithm thus is thermally-aware, and it must find appropriate finFET sizes that are equal to or less than the maximum number of parallel fins. For the finFETs, it is the temperature of the source that impacts the current generation [12].

3. GATE SIZING PROBLEM

Our goal is to contrast 32nm and finFET technologies in circuit design, and to compare area, delay, and power tradeoffs. Given a finFET circuit, we investigate how to best select the gate width (number of parallel fins within a finFET device), biasing, and fin width under delay and thermal constraints. Our objective is power minimization as finFETs promise reduced power consumption. Our problem formulation will assume a known operating temperature. I_{on} and I_{off} are therefore functions of the user-specified operating temperature. In a processor, for example, the instruction issue unit is the hottest functional unit and it typically maintains a steady state operating temperature. Our sized biased circuit is to operate correctly at a specific temperature while minimizing power. We first provide a formulation of our temperature-dependent delay constraints, and then describe the power optimization objective. We then formalize our problem and provide a solution.

3.1 Thermally-Dependent Delay Modeling

Delay constraints are imposed by setting the delay at each register input and at each primary output (PO) node to be less than T_{clk}, the clock period. The delay at the output of gate k, d_{k_O}, is obtained by adding d_k, the delay of gate k, to the maximum delay at gate k's inputs while considering all gates j in the fanin of k. That is,

$$d_{k_O} = d_k + \max_{j \in FI(k)} (d_{j_O}) \quad (1)$$

Logical effort provides a simplified means of modeling gate delays. The technology-independent delay of each gate has two components: effort (or stage) delay, and parasitic delay. The effort delay is computed as $g \times h$. The logical effort g describes the relative ability of a gate topology to deliver current compared to an inverter which has a logical effort of 1. The electrical effort, h, is the ratio of the gate's output to input capacitance $\frac{C_{out}}{C_{in}}$. The parasitic delay, p, indicates the delay due to self loading, and it can be expressed as $\alpha \times p_{inv}$, where α is multiplicative factor dependent on the gate type, and p_{inv} is the delay associated with an inverter's self loading. A technology parameter multiplier, τ, allows the technology-independent delay to become a technology dependent one.

The input capacitance of a gate is expressed in terms of $C_g \times n_i$, where $3 \times C_g$ is gate capacitance of a minimum size inverter, assuming a 2:1 p:n ratio of our gates. n_i specifies the multiplicative sizing factor over a minimum size inverter. We thus place integer restrictions on n_i. The ratio of the output to input capacitance of a gate, $\frac{C_{out}}{C_{in}}$, is proportional to the ratio of n_j to n_i, where n_j represents the effective number of fins that gate i is driving[1]. The delay equation for a gate i driving gate j is thus:

$$\begin{aligned} d_i &= \tau \cdot (g \cdot h + p), \ \ where \\ g &= logical\ effort\ of\ gate \\ h &= \frac{C_{out}}{C_{in}} = \frac{n_j \cdot C_g}{n_i \cdot C_g} = \frac{n_j}{n_i} \\ p &= \alpha \cdot p_{inv}, \ \ where\ \alpha\ depends\ on\ gate\ type \\ n_i &\in \{1,2,3,...,n_{i_{max}}\} \end{aligned} \tag{2}$$

The latter constraint limits the maximum multiplicative sizing factor to ensure that the device source temperatures will not rise above a user-specified temperature.

In our problem formulation, we wish to express gate delay as a function of *temperature (T), fin width*, and *biasing*, in addition to its drive and load (n_i and n_j). We first fix the temperature at which we wish the circuit to run. Thus, in the delay equation above, both τ and p_{inv} become a function of the temperature. Even when considering a temperature range, the maximum temperature should be considered as it yields the worst delay. Race conditions, however, must be verified using the worst case short delays at the lowest temperatures.

To achieve accurate delay estimates, the logical effort delay model must be calibrated for a process as discussed by Sutherland et al. [15]. Model calibration involves finding τ and p_{inv} for different temperatures, fin thickness, and biasing values. We allow only one biasing value equal to one third the power supply to avoid significant degrading of I_{on}. We used a 5-stage finFET ring oscillator and SPICE to compute τ and p_{inv} for various temperatures, fin thickness, and biases.

3.2 Power Objective

We investigate how finFET sizing impacts active vs. standby power. Leakage and short circuit current in the following discussion are assumed to be a function of the gate size, load, temperature, biasing state, and fin thickness.

Active power refers to when the circuit performing a computation, and it is the sum of:

$$\begin{aligned} P_{dyn} &= \alpha \cdot C_{out} \cdot V_{DD}^2 \cdot f \\ P_{sc} &= \alpha \cdot n_i \cdot I_{sc} \cdot t_{sc} \cdot V_{DD} \cdot f \end{aligned} \tag{3}$$

Dynamic power, P_{dyn}, is due to charging and discharging the circuit's capacitance, C_{out}. It is a function of the switching frequency, f, the power supply value, V_{DD}, and a switching factor, α, that reflects the portion of the load being switched each clock cycle. The short circuit power, P_{sc}, is due to both nfinFET and pfinFET being on simultaneously and having a direct path between the power supply and ground. It is a function of the short circuit current, the supply voltage, n_i, f, and α. Our power model does not reflect glitching, which potentially could increase power consumption.

Power dissipation during a standby state is dependent on specific standby vectors that are applied to the circuit. By assigning a low leakage state to the output nodes of registers, significant leakage current can be saved during standby mode. Forcing the output to a high or low state requires additional logic to control the state of the register while enabling the register to resume normal operation once awakened. Leakage power for a gate in this case is a function of a specific leakage state s_i. The standby leakage power is thus:

$$P_{leak-standby} = n_i \cdot s_i \cdot I_{leak} \cdot V_{DD} \tag{4}$$

3.3 Problem Formulation

Total power can be expressed as a weighted sum of dynamic and standby leakage power for all gates in the circuit. The user can set the relative weights to favor one optimization over the other. Our problem is then: Given a minimum size circuit, an operating temperature, weighting factors, we find for each gate n_i, W_{fin_i}, and $bias_i$ to minimize the power dissipation. More formally, our problem becomes:

$$\begin{aligned} Minimize:\ & P_{total} \\ Subject\ to:\ & d_{PO} \le T_{clk} \ \ \forall PO \\ & d_{k_O} = d_k + \max_{j \in FI(k)} (d_{j_O}) \\ & 1 \le n_i \le n_{max}; n_i \in \mathbf{Z}^* \\ & 1 \le W_{fin_i} \le W_{fin_{max}}; W_{fin_i} \in \mathbf{R}^* \\ & bias_i \in \{1,0\} \end{aligned} \tag{5}$$

3.4 Solution

Our problem formulation contains both continuous variables (W_{fin_i} and the delay variables), and integer variables (n_i, and $bias_i$). It is non-linear because the electrical effort contains the ratio of output to input capacitance, or n_i to n_j. The problem is thus MINLP, a Mixed-Integer Non-Linear Programming problem. A MINLP solver will thus find an *exact* and *globally* optimal solution. The run time of MINLP, however, may prove prohibitive for larger circuits. One alternative is to utilize a faster NLP (Non-Linear Programming) solver, and then convert the continuous solutions into integer ones using rounding. Boyd recommends a gradual conversion of the variables for mixed-integer geometric programming problems [4]. We utilize a three-step approach that uses NLP. First, the NLP solver is run and we round n_i into integer solutions. With fixed n_i, values, the NLP solver is run again and we round and fix $bias_i$ values. Finally, we run the NLP solver and find values for the continuous variables, W_{fin_i}. In our experiments (not reported in this paper), we have found that this technique leads to low error and improved run time over MINLP.

4. EXPERIMENTAL RESULTS

We focus in this section on a baseline experiment to compare the performance of finFETs against 32nm devices. Because I_{on} is higher for the finFETs, a finFET inverter is about three times faster than a 32nm one. Thus, a 32nm circuit is unable to achieve the same performance as a finFET circuit, even with aggressive sizing. For our experiment, we thus set the delay constraint for both the finFET and 32nm circuits to be 40% of the initial clock period obtained using static timing analysis for a minimum-size circuit.

The finFET optimization problem was solved using our NLP-based heuristic while the 32nm optimization problem was solved using NLP. We utilized the CONOPT package from the General Algebraic Modeling System (GAMS) [5] to solve each problem using our NLP-based heuristics. All our experiments were run on Pentium 4 machine running at 2.4GHz with 1GB of memory. We used the MCNC benchmarks, which were mapped to 2-input nands, 2-input nors, and inverters. While some of the examples are small they are representative of a single combinational stage. Each such stage can be optimized in parallel with others in a larger circuit. We used a SAT solver to generate the minimum leakage state vectors [1]. In our results, we placed a 15-fin maximum finFET device width to limit the device temperature to 60°C.

Table 1 summarizes the results of gate sizing of several finFET circuits to minimize active power. The first and second columns list the circuit's name and number of gates respectively. The following

[1]To simplify the presentation, we do not include the wire delay. However, in our experiments, $C_{out} = n_j \cdot C_g + C_{wire}$, where C_{wire} values were randomly generated.

Name	# Gates	Area[μm^2]	P_{total}[mW]	$\%P_{dyn}/P_{total}$	$\%P_{sc}/P_{total}$	$P_{standby}$[nW]	% Biased	% $W_{fin-nominal}$	% $n_i = 1$
alu2	487	313.56	7.28	46.4	53.6	0.15	5.1	71.7	82.5
C1908	817	576.49	16.28	44.9	55.1	0.43	0.9	61.6	75.5
s9234.1	1,547	980.17	31.71	51.2	48.8	0.45	2.5	75.4	87.3
dsip	3,509	1,928.86	51.39	52.1	47.9	0.17	6.4	92.6	96.9
s15850.1	4,560	2,331.44	43.16	57.0	43.0	0.72	9.1	82.5	96.2
mm30a	1,974	1,054.15	7.71	49.3	50.7	0.35	11.6	79.6	90.2
C432	303	207.96	5.11	44.4	55.6	0.11	8.6	58.7	78.2
s298	153	101.78	6.89	52.7	47.3	0.03	0.0	80.4	83.7
t481	5,284	3,097.95	115.25	45.8	54.2	0.59	0.1	93.2	92.1
C3540	1,609	980.97	19.13	44.9	55.1	0.72	2.1	62.2	87.8
s1196	626	389.60	16.13	46.7	53.3	0.28	1.4	68.4	85.9
Avg.	1,897.18	1087.54	29.09	48.67	51.33	0.36	4.35	75.12	86.94

Table 1: FinFET active power optimization.

Ckt	Norm. Delay	Norm. Area	Norm. P_{total}	$\%P_{dyn}$	$\%P_{sc}$	Norm. $P_{standby}$	% W_{min}
alu2	4.66	5.144	0.251	49.5	50.5	311.456	45.8
C1908	4.59	8.880	0.443	43.5	56.5	476.622	30.7
s9234.1	4.53	3.870	0.210	57.2	42.8	205.723	54.5
dsip	4.19	2.003	0.147	65.7	34.3	259.779	91.7
s15850.1	4.16	2.425	0.173	67.6	32.4	143.812	85.8
mm30a	4.60	26.914	0.973	40.6	59.4	1,241.383	68.2
C432	4.61	5.818	0.278	47.4	52.6	351.615	34.3
s298	4.33	2.686	0.191	62.5	37.5	138.022	58.2
t481	4.47	2.421	0.154	57.2	42.8	169.044	78.3
C3540	4.63	3.992	0.221	50.5	49.5	234.704	49.7
s1196	4.51	4.066	0.224	52.6	47.4	241.516	47.4
Avg.	4.48	6.2	0.3	54.03	45.97	343.06	58.6

Table 2: 32nm active power optimization.

columns report area in μm^2, total active power, P_{total}, in milliwatts, and then the contributions of switching power, P_{dyn}, and short circuit power, P_{sc}, as a percentage of P_{total}. Standby power, $P_{standby}$, is reported in nanowatts. The last three columns respectively describe the percentage of gates that were biased, that had minimum fin thickness, and that were of minimum size.

Table 1 shows the following: (a) dynamic power is split evenly between short circuit current and dynamic switching, (b) dynamic power is in mW while static power is in nW, (c) on average, 4% of the gates are biased, (d) on average, 75% of the devices have the minimum fin width, and (e) the majority of gates are of minimum size.

Table 2 shows the results of optimizing the same circuits to minimize active power using 32nm devices. The first column lists the circuit's name. The second, third and fourth columns list delay, area and total power, all normalized to the finFET circuits in Table 1. The next two columns report the contributions of switching power and short circuit power respectively. The last two columns show the normalized standby power and the percentage of devices that had a minimum width after optimization.

The 32nm device circuits exhibit an average of $\sim$4.5x increase in delay when compared to the finFET circuits, due to a lower I_{on}. The delay differences must be taken into account when comparing total power as total power is mainly composed of dynamic and short circuit power. The finFET circuits operate at higher clock speeds, resulting in larger total power. If the same delay constraints between the two circuits was used, the finFET circuits would exhibit lower total active power. However, since the 32nm device circuits are several times slower than the finFET circuits, the resulting *sized* finFET circuits would be minimum sized. Standby power increases by two to three orders of magnitude over the finFETs. This increase in standby power can be attributed to the large difference in I_{off} between the finFET devices and 32nm devices, as was shown in Figure 2.

5. CONCLUSION

This paper provides the first detailed comparison of finFET and 32nm bulk circuits. In addition, the paper shows how to use thermal information to guide circuit optimization problems. We investigated how finFET sizing, independent-gate biasing, and fin width can be used to achieve performance at minimum dynamic power. Our results show that finFETs outperform 32nm bulk devices and encourage further research into more efficient heuristic finFET sizing techniques and further detailed experiments.

6. ACKNOWLEDGMENTS

We would like to thank Fadi Aloul for sharing his SAT solver.

7. REFERENCES

[1] F. Aloul, S. Hassoun, K. Sakallaha, and D. Blaauw. "Robust SAT-Based Search Algorithm for Leakage Power Reduction". In *Power And Timing Modeling, Optimization and Simulation*, 2002.

[2] H. Ananthan, A. Bansal, and K. Roy. "FinFET SRAM – Device and Circuit Design Considerations". In *International Symposium on Quality Electronic Design*, pages 511–6, 2004.

[3] K. Bernstein, C. Chaung, R. Joshi, and R. Puri. "Design and CAD Challenges in Sub-90nm CMOS Technologies". In *Proc. Int. Conf. on Computer Aided Design*, pages 129–36, 2003.

[4] S. Boyd and L. Vandenberghe. *"Convex Optimization"*. Cambridge University Press, 2004.

[5] GAMS Development Corp. "The Solver Manuals", 2005.

[6] D. Fried, E. Nowak, J. Kedzierski, J. Dusterr, and K. Kornegay. "A Fin-Type Independent-Double-Gate NFET". *Device Research Conference*, pages 45–6, 2003.

[7] J. Grodstein, E. Lehman, H. Harkness, B. Grundmann, and Y. Watanabe. "A Delay Model for Logic Synthesis of Continuously Sized Networks". In *Proc. Int. Conf. on Computer Aided Design*, pages 458–62, 1995.

[8] SOI Group. "UFDG MOSFET Model (Ver. 3.1)". *University of Florida, Gainesville, FL 32611*, 2006.

[9] R. Haddad, L. van Ginneken, and N. Shenoy. "Discrete Drive Selection for Continuous Sizing". In *Proc. IEEE Int. Conf. on Computer Design*, pages 110–5, 1997.

[10] D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu. "A folded-channel MOSFET for deep-sub-tenth micron era". *International Electron Devices Meeting 1998. Technical Digest*, pages 1032–4, 1998.

[11] Y. Ju and K. Goodson. "Phonon Scattering in Silicon Films of Thickness Below 100 nm". *Applied Physics Letters*, 74:3005–7, 1999.

[12] M. Lundstrom and Z. Ren. "Essential Physics of Carrier Transport in Nanoscale MOSFETs". *IEEE Trans. on Electron Devices*, 49(1):133–41, January 2002.

[13] S. Mukhopadhyay, A. Raychowdhury, and K. Roy. "Accurate Estimation of Total Leakage Current in Scaled CMOS Logic Circuits Based on Compact Current Modeling". In *Proc. Design Automation Conf*, pages 169–74, 2003.

[14] *"Predictive Technology Model (PTM), http://www.eas.asu.edu/~ptm"*, 2006.

[15] I. Sutherland, B. Sproull, and D. Harris. *"Logical Effort: Designing Fast CMOS Circuits"*. Morgan Kaufmann Publishers Inc., 1999.

[16] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De. "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage". *IEEE Journal of Solid-State Circuits*, 37:1396, 2002.

[17] W. Zhao and Y. Cao. "New Generation of Predictive Technology Model for Sub-45nm Design Exploration". *International Symposium on Quality Electronic Design*, pages 585–90, 2006.