UDSM (Ultra-Deep Sub-Micron)-Aware Post-Layout Power Optimization for Ultra Low-Power CMOS VLSI*

Kyu-won Choi and Abhijit Chatterjee
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA 30332
{kwchoi,chat}@ece.gatech.edu

ABSTRACT

In this paper, we propose an efficient approach to minimize total power (switching, short-circuit, and leakage power) without performance loss for ultra-low power CMOS circuits in nanometer technologies. We present a framework for combining supply/threshold voltage scaling, gate sizing, and interconnect scaling techniques for power optimization and propose an efficient heuristic algorithm which ensures that the total slack budget is maximal and the total power is minimal in the presence of back end (post-layout-based) UDSM effects. We have tested the proposed algorithms on a set of benchmark circuits and some building blocks of a synthesizable ARM core. The experimental results show that our polynomial-time solvable strategy delivers over an order of magnitude savings in total power without compromising performance.

Categories and Subject Descriptors

B.7.2 Hardware [Integrated Circuits]: Design Aids-simulation.

General Terms: Algorithms.

Keywords: Low-power design, nanometer design, time slack distribution, and device and interconnect co-optimization.

1. INTRODUCTION

As the scale of integration expands, more transistors, faster and smaller than their predecessors, are being packed into smaller chips. The steady growth in clock frequency and processing capacity per chip has increased power dissipation dramatically. ITRS predicts that by 2010 over one billion transistors will be integrated into a single monolithic die. The total number of transistors integrated on a chip will exceed the human population by 2013 and will even become larger than the total number of neurons in a human brain by 2015. This reflects the rapid progress of technology and the need for technology scaling, especially for low power because the steady growth in clock frequency and processing capacity per chip has increased power dissipation tremendously and power storage technology has not improved significantly to compensate. Therefore, power-aware design techniques need immediate attention [1].

As technology sizes continue to decrease (with features below 0.1 micron), many new effects are being observed due to the use of ultra-deep sub-micron (UDSM) technologies. The significant UDSM effects are caused by i) *leakage power increase* due to short-channel effects and ii) *on-chip interconnect limits* which potentially threatens to decelerate or halt the historical progression of the semiconductor industry because the miniaturization of interconnects, unlike transistors, does not enhance their performance. For the past two decades the driving force for integrated circuits has been scaling of both the devices and interconnects [2,3,4,5]. We believe that the most effective approach to power optimization in UDSM designs is to consider both device and interconnect concurrently throughout the entire design process from the RTL level to layout design according to the need of the design complexity and the accuracy.

In our approach, we co-optimize device and interconnect parameters using efficient USDM-aware models and heuristics. Layout-based on-chip interconnect effects due to electrical, thermal, and mechanical stresses in a multilevel interconnect stack and MOSFET's short channel effects are considered to minimize the total (switching, short-circuit, and leakage) power. The optimization problem is state below.

Given: i) A logic or gate net-list from synthesizable applications, ii) device/interconnect technology and parasitics from initial layout, iii) a required operational clock frequency, iv) activity profiles at each input, and v) delay/area constraints.

Minimize: $Total\_Power\ (Vdd, Vth, Wgate, Wint)$

Subject to:
$Delay(Vdd, Vth, Wgate, Wint) \leq Tspec$
$Vdd = Vdd\text{-}global,\ \forall gate$
$Vth = Vth\text{-}global,\ \forall gate$
$\quad$ *or (Vth-high or Vth-low for dual Vth,* $\forall$ *gate i)*
$Maxsize(i) \geq Wgatei \geq Minsize(i),\ \forall gate\ i$
$Maxsize(j) \geq Wintj \geq Minsize(j),\ \forall int\ j$

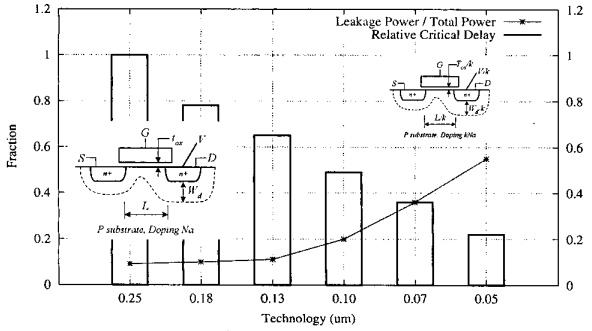
Determine: i) The optimal supply voltage *Vdd*, ii) the threshold voltage *Vthi* of each MOSFET, iii) the channel width *Wgatei* of each MOSFET, and iv) the interconnect width *Wintj* of each interconnect such that the sum of the static, dynamic and short-circuit components of energy consumption is minimized while allowing operation at the desired clock frequency *fc*.

2. MOTIVATION AND RATIONALE

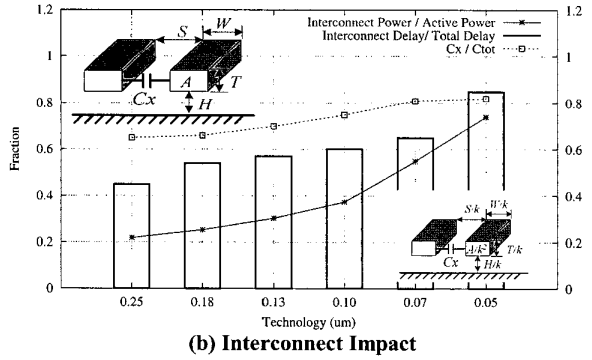
UDSM Effects for Power and Delay: Significant UDSM effects in terms of power and delay mainly caused by low-threshold voltage and high density-interconnections are shown in Figure 1.

* This work was supported by NSF (ITR) under grant #CCR 022-0259.

ISLPED'03, August 25–27, 2003, Seoul, Korea.



(a) Leakage Impact



(b) Interconnect Impact

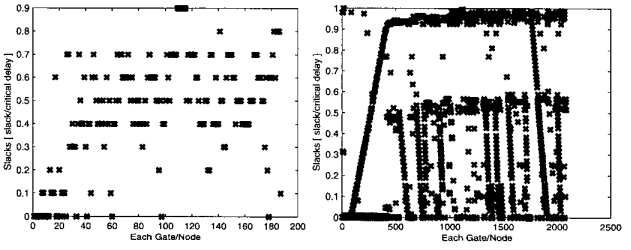
Figure 1. UDSM Effects for 12-Banchmark Circuits

Table 1. Summary of Technology Parameters [ITRS,MOSIS]

Technology (nm)		250	180	130	100	70	50
Min. Vdd (v)		2.0	1.8	1.5	1.3	0.9	0.6
Vth-high (v)		1.0	0.9	0.75	0.65	0.45	0.3
Vth-low (v)		0.5	0.45	0.375	0.325	0.225	0.1
Tox (A°)		60	45	35	30	20	12
Wgate (um)		5.75	4.1	3.0	2.3	2.8	3.9
fc (GHz)		1.0	1.2	1.6	2.0	2.5	3.0
Levels		6	6	7	8	9	9
Poly	H (um)	0.2	0.15	0.13	0.1	0.07	0.07
	W (um)	0.25	0.18	0.13	0.1	0.07	0.05
	S (um)	0.25	0.18	0.13	0.1	0.07	0.05
	R (ohm)	4	5.3	6.2	8	11.4	11.4
M1-2	H (um)	0.5	0.46	0.34	0.26	0.2	0.14
	W (um)	0.30	0.23	0.17	0.13	0.1	0.07
	S (um)	0.30	0.23	0.17	0.13	0.1	0.07
	T (nm)	650	500	360	320	270	210
	R (ohm)	0.044	0.048	0.065	0.085	0.11	0.16
M3-4	H (um)	2.0	2.0	1.2	1.0	0.6	0.6
	W (um)	1.0	1.0	0.6	0.5	0.3	0.3
	S (um)	1.0	1.0	0.6	0.5	0.3	0.3
	T (nm)	900	900	900	900	900	900
	R (ohm)	0.011	0.011	0.018	0.0224	0.036	0.036
M5-6	H (um)	2.5	2.5	2.0	2.0	1.5	1.5
	W (um)	2.0	2.0	1.0	1.0	0.75	0.75
	S (um)	2.0	2.0	1.0	1.0	0.75	0.75
	T (nm)	1400	1400	900	900	900	900
	R (ohm)	0.009	0.009	0.011	0.011	0.015	0.015
M7-8	H (um)	-	-	2.5	2.5	2.4	2.4
	W (um)	-	-	2.0	2.0	1.2	1.2
	S (um)	-	-	2.0	2.0	1.2	1.2
	T (nm)	-	-	1400	1400	900	900
	R (ohm)	-	-	0.009	0.009	0.0094	0.0094
M9	H (um)	-	-	-	-	2.5	2.5
	W (um)	-	-	-	-	2.0	2.0
	S (um)	-	-	-	-	2.0	2.0
	T (nm)	-	-	-	-	1400	1400
	R (ohm)	-	-	-	-	0.009	0.009
Via (M1-M2)	size (um)	0.5	0.36	0.26	0.2	0.14	0.1
	R (ohm)	0.46	0.69	0.95	1.43	2.16	3.27
k		3.3	2.7	2.3	2.0	1.8	2.5

We demonstrate those effects with 12-typical benchmark circuits (MCNC91, ISCAS89) by using the parameters in Table 1. As shown in Figure 1(a), leakage power increases significantly as the technology feature size becomes smaller. Interconnect delay and power have become dominating factors in determining circuit performance in UDSM designs mainly due to: i) increased routing densities that have led to shrinking interconnect pitches and ii) aspect ratios that attempt to keep the resistance of these narrower interconnects constant as shown in Figure 1(b). Both have resulted in an increase in the capacitance per unit length, particularly due to interlayer coupling capacitance. Interconnect scaling approaches have been introduced to optimize the interconnect structure of each net in terms of interconnect topology, wire width and spacing, buffer locations and sizes to meet performance and signal reliability requirements [3].

Slack Analysis: The first observation of this research is that the logic modules on non-critical paths in digital circuits typically have high timing slack which may allow the increase of their delay without violating the timing constraints. For most circuits, logic modules on critical paths only account for a small portion of all modules. Figure 2(a) and Figure 2(b) show the slack map for the s298 benchmark circuit and for the ARM ALU-64, respectively. In these figures, x-axis is each gate and y-axis shows slack percentage of the circuit after technology mapping. With typical 12-benchmark circuits, statistics show that the number of gates on critical paths accounts for only about 12 % of the total number of gates, while more than 60 % of gates have their slack larger than 25% of the critical path delay. This potentially provides much room for power reduction without speed loss of the circuits.



(a) Slack Map for S298 **(b) Slack Map for ALU-64**

Figure 2. Slack Map

Low-power Design Maintaining Speed: In general, low-power optimizations that do not compromise performance of a system are dependent on time slack calculation and surplus delay (slack budget) distribution among the circuit modules. Time slack is measured as the difference between the signal required time and the signal arrival time at the primary output of each module. Figure 3 shows a key simple rationale for gate level power optimization. The more the slack assigned to a power-hungry gate/interconnect, the higher is the potential to save power by slowing down the gate/interconnect. The slowing down is achieved by optimally adjusting the module's supply voltage, threshold voltage and by resizing the gates and interconnects in the module to minimize total power, dynamic and also leakage power.

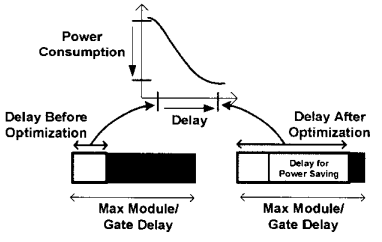


Figure 3. Gate-level Power Optimization Rationale

Why Post-Layout Based? : Performance-driven logic synthesis followed by performance-driven layout [4,5,9] has become a necessity for designing high performance circuits. However, this loosely coupled two-phase timing optimization methodology has serious limitations for deep-submicron-based designs in which interconnection delays become the dominant factor in determining the circuit speed. Accurate information regarding interconnection delays is not available during the logic synthesis phase and the interconnection delays are just roughly estimated in this phase. Therefore, errors in estimation could result in a logic design far-off from optimal.

3. DELAY AND ENERGY MODEL

We used reasonably accurate UDSM-aware MOSFET and interconnect models (below 10% errors compared to HSPICE level 49 model) for delay and power, compared with most of the literature [15,16,17].

3.1 MOSFETs in UDSM

Since Shockley's square law model is not accurate for UDSM MOSFETs due to velocity saturation, short channel effects, mobility degradation, and series resistance, new MOSFET models have been developed [15]. Traditional long-channel MOS theory states that device current in the saturation mode of operation is proportional to the square of gate drive (Vdd-Vth) and inversely proportional to channel length. At ultra-short channel lengths most carriers travel at maximum saturated velocity, Vsat, throughout the channel, which nearly eliminates the impact of channel length on current. We used a recent version of the alpha-power law model [15,16,17] in this research.

3.2 Interconnects in UDSM

For UDSM technologies, interconnect models must be carefully monitored and controlled since the impact of interconnect on system performance is increasing. The interconnect capacitance in UDSM dominates the total net capacitance due to; i) increased routing densities that have led to shrinking wiring pitches and ii) interconnect aspect ratios that attempt to keep the resistance of these narrower wires constant. Much research has been performed on modeling of the interconnect coupling effect in the time [16] and also frequency domain. In general, frequency domain approaches are ideal for measurement of high-frequency phenomena such as skin effect but unsuitable for recreating the actual environment in real designs [17]. We used post-layout time-domain models for the interconnect so that the analysis results give easily measurable delay and energy, rather than hard-to-evaluate frequency response.

3.3 Component Delay (Gate+Interconnect) Model

We use a transregional model for estimating the worst-case signal propagation delay through a gate. The delay model has been derived using an extension of the alpha-power law saturation drain current model to the subthreshold region. The drain current model incorporates effects of high-field and quasi-ballistic (velocity overshoot) carrier transport in the MOSFET channel. The delay model consists of four major components: i) the delay due to switching MOSFETs, ii) the distributed interconnect RC delay, iii) the time of flight delay, and iv) the delay component due to the non-zero rise time of the input signal. These definitions of gate delay and interconnect resistance delay allow the definition of arrival times and required times at the input and output of a gate in the network, which are used for defining time slack.

3.4 Component Energy (Gate+Interconnect) Model

The equations used to compute the dynamic and static energy dissipation of a gate have been presented and analyzed in a recent work [15]. It is assumed that the gates are simple multi-input gates with symmetric series or parallel pull-up and pull-down MOSFET configurations. Contributions of subthreshold leakage through the MOSFET channel as well as the leakage across the device drain junctions to static dissipation are included.

4. PROPOSED APPROACH

We believe that the most effective approach to power optimization in UDSM designs is to consider both device and interconnect designs at the gate level after initial layout in terms of optimization complexity, simulation efficiency, and estimation accuracy. The proposed design flow of our approach is shown in Figure 4.

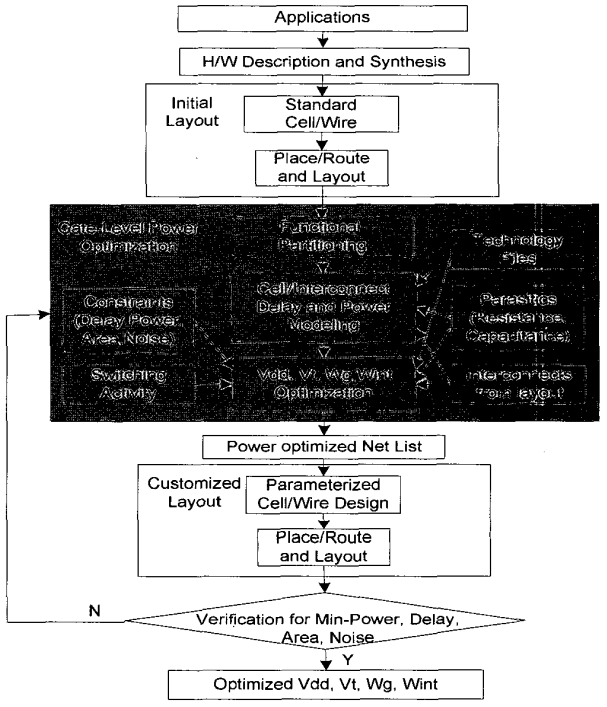


Figure 4. Power Optimization Procedure

4.1 Pre-procedure

Partitioning: For simulation run-time efficiency and power optimization effectiveness, we used a topological depth-based partitioning [7] which ensures the minimization of the delay skew between sub-modules, and constrains maximum sub-module size (or fan-out size). First of all, labeling of each circuit node is conducted according to topological order. Then, according to the maximum depth and maximum size constraints, the whole flattened gate-level digital circuit is partitioned into sub-modules. The complexity of this algorithm is $O(b^m)$, where b is the branching factor (i.e., average fan-out number) and m is the maximum topological depth.

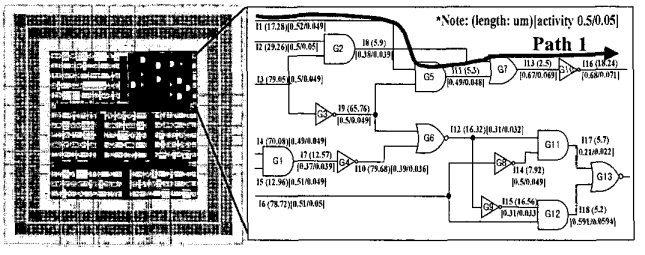
Activity Profiling: We used *Monte Carlo simulation* for activity profiling of each module/sub-module. This simulation based approach is capable of handling various device models, different

circuit design styles, but takes very long simulation time. However, our hierarchical partitioning approach partially compensates for this problem.

Initial Layout: We placed and routed the synthesized circuits by using feasible technology library and then scaled the device/interconnect dimensions such as oxide thickness, effective channel length/width, and interconnect length /width according to the scaling factor in [1,17]. For some geometric and electrical parameters such as interconnect capacitance/resistance, W/S/T/H of each metal layer, Vdd/Vth range of each technology, etc..., we followed the 2001 edition of the ITRS projections [1].

Path Enumeration: After obtaining the post-layout gate-level net lists, we enumerated all possible input-output paths of the circuits.

Figure 5 shows the pre-procedure for a benchmark circuit. Using this approach, it is possible to compute the path delays as a sum of the device and interconnect delays using the UDSM models described earler.



(a) Layout and Partitioning (b) Post-Layout Gate-Level Circuit

Path 1: I(1)-G(5)-I(11)-G(7)-I(13)-G(10)-I(16)
Path 2: I(2)-G(2)-I(8)-G(7)-I(13)-G(10)-I(16)
Path 3: I(3)-G(2)-I(8)-G(7)-I(13)-G(10)-I(16)
Path 4: I(3)-G(3)-I(9)-G(5)-I(11)-G(7)-I(13)-G(10)-I(16)
Path 5: I(3)-G(3)-I(9)-G(6)-I(12)-G(11)-I(17)-G(13)-I(19)
Path 6: I(3)-G(3)-I(9)-G(6)-I(12)-G(9)-I(15)-G(12)-I(18)-G(13)-I(19)
Path 7: I(4)-G(1)-I(7)-G(4)-I(10)-G(6)-I(12)-G(11)-I(17)-G(13)-I(19)
Path 8: I(4)-G(1)-I(7)-G(4)-I(10)-G(6)-I(12)-G(9)-I(15)-G(12)-I(18)-G(13)-I(19)
Path 9: I(5)-G(1)-I(7)-G(4)-I(10)-G(6)-I(12)-G(11)-I(17)-G(13)-I(19)
Path 10: I(5)-G(1)-I(7)-G(4)-I(10)-G(6)-I(12)-G(9)-I(15)-G(12)-I(18)-G(13)-I(19)
Path 11: I(6)-G(8)-I(14)-G(11)-I(17)-G(13)-I(19)
Path 12: I(6)-G(12)-I(18)-G(13)-I(19)

(c) Path Enumeration

Figure 5. Pre-procedure for c499

4.2 Proposed Heuristic

In general, there are several methods to solve non-linear optimization problems for VLSI circuit optimization; i) exhaustive search, ii) pseudo-polynomial time Lagrangian and Dynamic Programming algorithms, iii) approximation or randomization, and iv) heuristic-based methods. In this paper, a heuristic-based approach for solving the device and interconnect co-optimization problem is proposed. For better understanding of our proposed algorithm, we illustrate our algorithm with the example circuit in Figure 5(b). Figure 6 shows the power and the delay for the example circuit of Figure 5(b) as a basis before optimization with 0.05 um technology (see Table 1 for the parameters used)

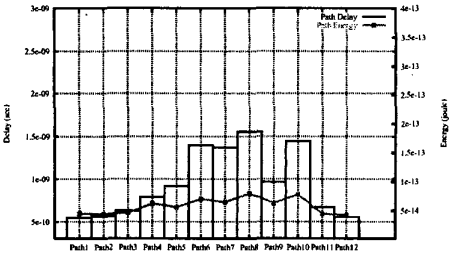


Figure 6. Power and Delay Basis for an Example Circuit in Figure 5(b) (without optimization)

Step 1: Set all parameters for minimum power. In this phase, all the paths show largest delay within the specified technology parameter ranges and violate the delay constraints. Figure 7 shows the power and the delay for the above example circuit.

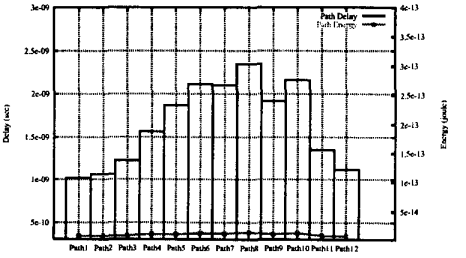


Figure 7. Technology Mapping for Minimum Power

Step 2: Select delay violation paths that violate delay constraints according to delay values. Path 5, 6,7,8,9,10 are the violation paths in this example as shown in Figure 8.

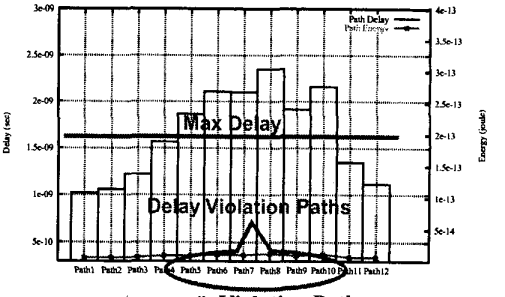


Figure 8. Violation Paths

Step 3: Determine the amount of delay to be decreased for each component on the violation paths by using the EPDR heuristic for minimizing power. The path effort is number of path which includes the node on each path. For example, for node G(6) in Figure 5(c), the paths that include the node G(6) on their paths are path 5,6,7,8,9,10. Therefore, the path-effort of the node G(6) is 6.

EPDR Heuristic: EPDR (Energy*Path-effort and Delay Ratio) Paradigm: When the ratio (*module energy*path effort*) / (*module delay)* for each module is the same, the total energy consumption is minimal. Therefore the surplus time slacks should be assigned for each module according to the cost function of $\frac{E_1 \cdot P_1}{d_1} = \frac{E_2 \cdot P_2}{d_2} = \ldots = \frac{E_n \cdot P_n}{d_n}$ (1,..., n: module number) for minimum power.

Rationale: The component which has more violation paths and less energy consumption possibility due to the delay reduction should have higher priority to reduce some portion of the amount of the violation delay than others.

Validation: An experiment was conducted to see if the EPDR heuristic holds up when supply voltage, threshold voltage and device/interconnect sizing are modulated to control the power consumption of the two component (G6-I12 and G11-I17) of the circuit in Figure 5(b). The graph in Figure 9 shows that the total power consumption of the circuit is minimal when the value of the metric is close to 1. This shows that the EPDR paradigm holds up when our proposed power optimizations are applied to complex device and interconnect models.

$$metric = \frac{E_{(G6-I12)} \cdot Path_{(G6-I12)} \Big/ E_{(G11-I17)} \cdot Path_{(G11-I17)}}{D_{(G6-I12)} \Big/ D_{(G11-I17)}}$$

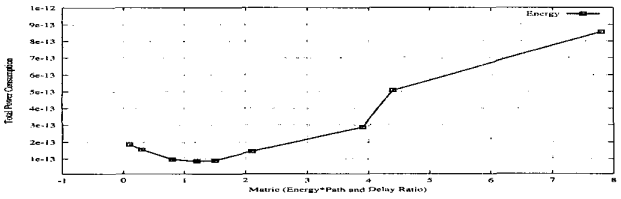


Figure 9. EPDR Validation

Step 4: Find Vdd, Vth, Wg, Wint of each module to minimize the increase of the power for each module on the violated paths when the delay of each module is reduced to the amount of the delay from Step 3. The cost function to be minimized is $\frac{\Delta Power(Vdd,Vth,Wg,Wint,\eta)}{\Delta Delay(Vdd,Vth,Wg,Wint)}$. Eqn. (1) below shows the cost function used in this paper.

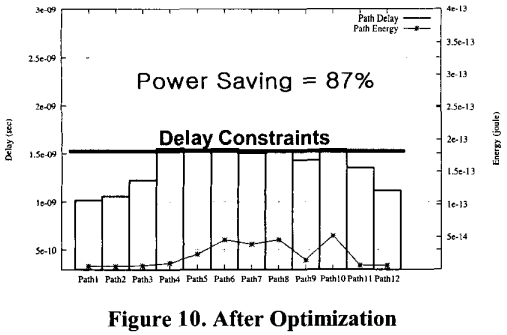


Figure 10. After Optimization

Step 5: Repeat Step 2-4 until all paths satisfy the delay bound. Figure 10 shows the power optimization result for the example circuit.

5. EXPERIMENTAL RESULTS

We developed a simulation frame work with C/C++/STL and Perl on Ultra-80 Unix machine for the hierarchical power optimization. Also, we used off-the-shelf commercial tools (synopsis and cadence) for the RTL description, the functional verification, and the logic synthesis / initial layout of the target system. A few arithmetic modules from ARM Core and ISCAS89/MCNC91 benchmark circuits are used for the experimental demonstration. For the range of the technology parameter values, the 2001 updated version of ITRS (International Technology Roadmap for Semiconductors), the MOSIS (Integrated Circuit Fabrication service) parameter test results with our scaling techniques which is proven by HSPICE.

Lowering the supply voltage causes the dynamic component of the dissipation to reduce quadratically. However, at very low values of the supply voltage, the threshold voltage must be reduced considerably, causing the leakage dissipation to increase exponentially. An increase in device width contributes to larger static dissipation and to some extent prevents the dynamic power component from reducing quadratically. In addition, interconnect scaling is closely related to the device scaling and quite important for this optimization, especially for nanometer technologies. Therefore, the total of the static and the dynamic components of dissipation is minimized by a unique choice of supply voltage, threshold voltage, device width, and interconnect scaling values. Table 2 and Table 3 demonstrate the effectiveness of the optimization with the proposed design flow (for the technology parameters, refer the Table 1). Average 19.519x savings in total power without performance loss is achieved over traditional circuit design method. Since our optimization algorithm experiments returned quite small threshold voltage values, we performed experiments to determine the impact of the threshold voltage variation due to process fluctuations on the amount of power savings possible in Figure 11(a), As the threshold voltage fluctuations increase from 0 to 50%, the power savings drop from 16-25x to 6-12x. Thus even for large threshold voltage fluctuations we can get substantial power savings. Figure 11(b) shows the clock skew impact for power savings. In Figure 11(c), we can see the normalized total power of the circuits for various percentages of low-Vth gates when we use dual-threshold voltages. The results indicate that as the percentage is increased, total power saving is converged. Finally, we compared our proposed approach with a globally optimal solution to the technology mapping problem. However, the staggering complexity of the problem prevents us from arriving at any optimal solution for comparison. For example, let there be Q quantizations of Vdd, Vth, Wg, and Wint. For a circuit with N components, there are $S = Q^{(2N+2)}$ combinations of the parameters over the entire circuit. Even for small circuits this is a large number (Q=20, N=6, $S=20^{14}$). Figure 11(d) shows the comparison with exhaustive optimal solution, randomly averaged case, and our proposed technique for s27. It shows that our algorithm performs reasonable well and fast. In addition, our experiments show that judiciously chosen partitioning also can significantly help to reduce the time complexity, maintaining similar power optimization results.

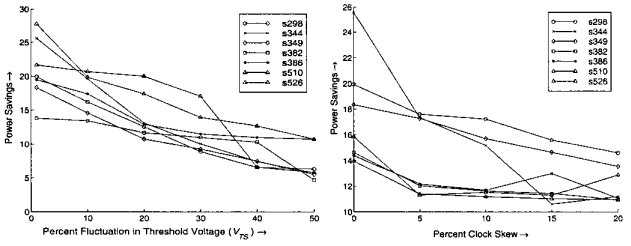
$$\frac{\Delta E_{Total_i}}{\Delta T_{d_i}} = \frac{\frac{1}{2}\eta_i V_{ddi}^2 (1 + K_{Short-Circuit_i}) \cdot \left[w_i \{ C_{PD_i} + (f_{ii} - 1) C_{m_i} \} \sum_{j=1}^{fa_i} \left((Ca_{ij} \cdot Wint_{ij} + 2Cf_{ij}) \cdot l_{ij} + \frac{Cx_{ij} \cdot lab_{ij}}{d_{ab}} \cdot \left[\frac{1}{1 - \frac{Wa_{ij} + Wb_{ij}}{2}} \right] \right) \right] + V_{ddi} W_{gate_i} I_{off} / f_c}{\left[\left[\frac{1}{2} - \frac{1 - \frac{V_{thi}}{V_{ddi}}}{1 + \alpha} \right] \max_{j \in (1, f_{ii})} \{ t_{d_{ij}} \} + \frac{V_{ddi} / 2}{I_{Dsw} - f_{ii} \beta I_{off}} \right] \cdot \left[C_{PD_i} + \frac{1}{w_{gate_i}} \sum_{j=1}^{fa_i} \left((Ca_{ij} \cdot Wint_{ij} + 2Cf_{ij}) \cdot l_{ij} + \frac{Cx_{ij} \cdot lab_{ij}}{d_{ab}} \cdot \left[\frac{1}{1 - \frac{Wa_{ij} + Wb_{ij}}{2}} \right] \right) \right] + \max_{j \in (1, f_{ii})} \left[R_{int_{ij}} \cdot \left((Ca \cdot Wint + 2Cf) \cdot l + \frac{Cx \cdot lab}{d_{ab}} \cdot \left[\frac{1}{1 - \frac{Wa + Wb}{2}} \right] \right) + \frac{L_{int_{ij}}}{v_{int}} \right] + \frac{1}{2} C_{m_i} V_{ddi} \sum_{j=1}^{fa_i - 1} \frac{1}{I_{Dsw}(j)}} \quad (1)$$

Table 2. Before Optimization (Tech.: 0.05um, Vdd:0.65, Vth:0.3)

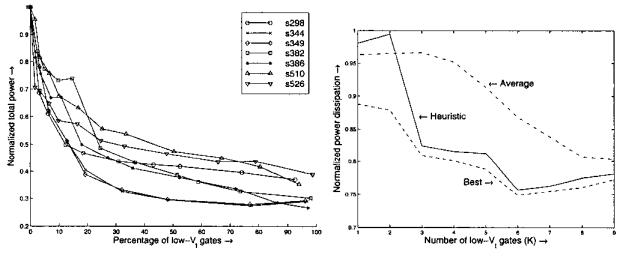
System Module	Gates / Depth	Delay (ns)	Input Activity	Average ϖ_g ϖ_{int}	Energy Dissipation			
					Leakage	Switching	Short-ckt	Total
4 - Full Adder	106/48	2.50	0.5	11.9, 0.2	2.51E-12	2.58E-11	1.27E-12	2.96E-11
			0.05	11.9, 0.2	2.51E-12	2.55E-13	1.29E-14	2.78E-12
16 - Look ahead	1838/ 81	3.5	0.5	15.6, 0.2	1.78E-11	4.51E-10	5.50E-11	5.24E-10
			0.05	15.9, 0.2	1.78E-11	4.54E-12	5.48E-13	2.28E-11
64 - ALU	3417/ 226	7.9	0.5	6.1, 0.2	1.68E-10	2.60E-09	1.69E-10	2.93E-09
			0.05	6.1, 0.2	1.68E-10	1.12E-10	1.69E-12	2.82E-10
s298	286/18	1.02	0.5	3.3, 0.2	2.88E-12	8.50E-12	1.40E-13	1.15E-11
			0.05	3.7, 0.2	2.88E-12	2.88E-12	1.50E-15	2.96E-12
s344	229/28	2.39	0.5	5.9, 0.2	6.89E-12	3.76E-11	5.82E-13	4.51E-11
			0.05	5.9, 0.2	6.89E-12	3.77E-13	5.68E-15	7.27E-12
s386	426/23	2.37	0.5	7.9, 0.2	8.90E-14	2.88E-11	5.89E-13	2.95E-11
			0.05	7.5, 0.2	8.90E-14	3.03E-11	5.68E-15	3.97E-13
s526	596/18	2.77	0.5	5.2, 0.2	1.12E-12	3.03E-11	1.18E-12	3.26E-11
			0.05	5.1, 0.2	1.12E-12	3.14E-13	5.79E-15	1.44E-12
c6288	2406/ 129	5.29	0.5	4.7, 0.2	9.13E-10	1.90E-09	3.87E-10	3.20E-09
			0.05	4.3, 0.2	9.13E-10	2.77E-10	3.87E-12	1.19E-09

Table 3. After Optimization (Tech.: 0.05um)

System Module	Gates / Depth	Delay (ns)	Input Activity	Vdd, Vth	Average ϖ_g ϖ_{int}	Total Energy	Power Reduction	Area Reduction
4 - Full Adder	106/48	2.50	0.5	0.65, 0.1	5.21, 0.39	1.21E-11	2.45x	11.7%
			0.05	0.625, 0.12	3.16, 0.24	5.81E-13	4.77x	16.4%
16 - Look ahead	1838/ 81	3.5	0.5	0.625, 0.1	7.1, 0.14	2.12E-11	24.7x	1.8%
			0.05	0.625, 0.1	6.16, 0.15	1.70E-12	13.4x	4.4%
64 - ALU	3417/ 226	7.9	0.5	0.625, 0.1	7.81, 0.15	1.09E-10	26.8x	-2.1%
			0.05	0.625, 0.1	3.16, 0.18	2.95E-11	9.55x	5.2%
s298	286/18	1.02	0.5	0.65, 0.1	2.21, 1.70	9.25E-13	12.4x	11.7%
			0.05	0.625, 0.12	3.7, 1.98	1.13E-13	26.2x	16.4%
s344	229/28	2.39	0.5	0.625, 0.1	4.1, 1.90	2.30E-12	19.6x	1.3%
			0.05	0.625, 0.1	5.16, 2.05	8.37E-13	8.6x	14.4%
s386	426/23	2.37	0.5	0.625, 0.1	7.81, 1.51	6.97E-13	42.2x	1.1%
			0.05	0.625, 0.1	3.16, 0.20	1.445E-14	27.3x	6.0%
s526	596/18	2.77	0.5	0.65, 0.1	5.21, 1.70	1.39E-12	23.4x	11.7%
			0.05	0.625, 0.12	3.16, 1.99	9.74E-14	14.7x	16.4%
c6288	2406/ 129	5.29	0.5	0.625, 0.1	3.1, 1.69	1.10E-10	29.1x	1.3%
			0.05	0.625, 0.1	4.1, 1.65	4.51E-11	26.4x	14.4%
						Average	19.519x	



(a) Vth fluctuations Impact **(b) Clock Skew Impact**



(c) Power vs. percentage of low-Vth **(d) Relative performance**

Figure 11. Process Fluctuation Impacts and Effectiveness of Optimization

6. CONCLUSION AND FUTURE WORK

This paper presents an efficient design flow and a novel EPDR based heuristic algorithm that can solve very complicated non-linear optimization problem in polynomial time for low power CMOS circuits. Consequently experimental results demonstrate that the proposed power optimization strategy yields reductions in total power by a factor from 2.4x to 42.2x over (average 19.5x) without optimization case in nanometer (below 0.1 micron) technologies, maintaining the circuit speed and the proposed heuristic approach presents run-time saving around 50% or more across few functional sub-modules comparing with exhaustive search case. Future work will include application-specific and architecture- driven issues with this UDSM-aware post-layout technology mapping techniques because assorted software, algorithmic, architectural techniques can significantly help to minimize the total switching activity for a target system.

7. REFERENCES

[1] International Technology Roadmap for Semiconductors 2001 Edition, Executive Summary, (http://public.itrs.net/Files/2001ITRS/Home.htm).

[2] J. Meindl, "Low power microelectronics: retrospect and prospect," *Proceedings of the IEEE*, vol. 83, pp. 619-635, April 1995.

[3] J. Cong, "An interconnect-centric design flow for nanometer technologies," *Proceedings of the IEEE*, vol. 89, pp. 505-528, April 2001.

[4] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-power CMOS digital design," *IEEE Journal of Solid-State Circuits*, vol. 27, pp. 473-484, April 1992.

[5] J.M. Rabaey and M. Pedram, *Low Power Design Methodologies*, Kluwer Academic Publishers, 1996, pp 21-64,130-160.

[6] Pakaj Pant, Vivek De, and Abhijit Chatterjee, "Simultaneous power Supply, threshold voltage, and transistor size optimization for low-power operation of CMOS circuits," *IEEE Transactions on VLSI Systems*, vol. 6, pp. 538-545, Dec 1998.

[7] Kyu-won Choi and Abhijit Chatterjee, "HA^{2}TSD: Hierarchical time slack distribution for ultra-low power CMOS VLSI," *Proc. of the International Symposium on Low Power Electronics and Design,* pp.207-212, 2002.

[8] Kyu-won Choi and Abhijit Chatterjee, "PA-ZSA (Power Aware Zero Slack Algorithm): A graph based timing analysis for ultra low-power CMOS VLSI," *Proc. of PATMOS'2002*, pp. 178-187, Sep. 2002.

[9] K. Roy and S.C. Prasad, *Low-Power CMOS VLSI Circuit Design,* John wiley & Sons, Inc., 2000, pp 201-252.

[10] T. Kobayashi and T. Sakurai, "Self adjusting threshold voltage scheme (SATS) for low voltage high speed operation," *IEEE CICC*, 1994, pp.271-.

[11] S. Mutoh, "1-V Power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 847-, April 1992.

[12] A. Fariborz, "A dynamic threshold voltage MOSFET (DTMOS) for ultra-low voltage operation," IEDM Tech., 1994, pp809-.

[13] L. Wei, Z. Chen, and K.Roy, "Double gate dynamic threshold voltage (DGDT) SOI MOSFETs for low power high performance designs," IEEE SOI conference, 1997, pp. 82-83.

[14] S.S. Sapatnekar, V.B. Rao, P.M. Vaidya, and S Kang, "An exact solution to the transistor sizing problem for CMOS circuits using convex optimization," *IEEE Trans. On CAD of Integrated Circuits and Systems*, vol. 12, no. 11, pp. 1621-1634, September 1993.

[15] T. Sakurai and A.R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE Journal Solid-State Circuits*, vol. 25, pp. 584-594, Apr. 1990.

[16] T. Sakurai, "Closed-form expression for interconnect delay, coupling, and crosstalk in VLSI," *IEEE Transactions on Electron Devices*, vol. 40, pp. 118-124, Jan. 1993.

[17] D. Sylvester and K. Keutzer, "Getting to the bottom of deep submicon," *Proc. of ICCAD*, pp. 203-211, 1998.