| PAPER   *Special Section on VLSI Design and CAD Algorithms* |

Post-Layout Transistor Sizing for Power Reduction in Cell-Base Design

Masanori HASHIMOTO[†a)] *and* **Hidetoshi ONODERA**[†], *Regular Members*

SUMMARY   We propose a transistor sizing method that downsizes MOSFETs inside a cell to eliminate redundancy of cell-based circuits as much as possible. Our method reduces power dissipation of detail-routed circuits while preserving interconnects. The effectiveness of our method is experimentally evaluated using 3 circuits. The power dissipation is reduced by 75% maximum and 60% on average without delay increase. Compared with discrete cell sizing, the proposed method reduces power dissipation furthermore by 30% on average.
key words: *transistor sizing, low power design, cell-base design, post-layout optimization, gate sizing*

1.   Introduction

Cell-base design has a well-established framework for the development of ASICs, and has been widely adopted. On the other hand, cell-based circuits inherently contain redundancy, for example, in power dissipation. In this paper, we propose a post-layout transistor sizing method for power reduction. Our method aims to reduce the redundancy of cell-base design and to obtain high performance circuits close to full-custom quality while keeping the cell-base design framework. We downsize MOSFETs inside a cell continuously, and generate the corresponding cell layout on the fly. The cell layout generation system used in our method does not change the location of input and output pins while the transistor widths inside a cell are varied [1]. Exploiting this feature, we can optimize detail-routed circuits, without any modifications of interconnects, using the precise wire capacitance values extracted from the detail-routed circuits.

Many transistor sizing methods for delay and power optimization have been proposed [2]–[6]. These methods need to derive the delay time of each cell at any MOSFET size. References [2]–[4] utilize Elmore delay model. In this delay model, we can get the optimal solution of the problem formulated using a simple variable-transformation method. However, the accuracy of the delay model is not high enough, and hence the optimized circuits may violate the delay constraints. In Refs. [5], [6], the cell delay is approximated as a linear function of the cell size, and transistor sizing is formu-

lated as a linear optimization problem. This method also can obtain the optimal solution of the formulated problem. However, the linearization of the cell delay may introduce errors in timing analysis.

Recently, the delay time due to wire capacitance occupies a considerable part of the total circuit delay. Many of the previous transistor sizing methods [2], [3], [5], [6] concentrate on circuit-level optimization, and the consideration on layout is not enough. When the optimization result is applied to the layout, routing is affected, i.e. wire capacitances in the resulting layout become different from the initial values. The variation of wire capacitance may cause a violation of delay constraints. In Ref. [4], transistor sizing, re-routing and compaction techniques are performed to the circuit repeatedly for better consideration on layout. In a DSM process, coupling capacitances between adjacent interconnects in the same metal layer or two successive metal layers become dominant. The accurate capacitance evaluation of all the interconnects influenced by re-routing and compaction becomes computationally intensive and hence the repeated evaluation inside the optimization loop may become impractical.

Our method handles detail-routed circuits designed in a cell-base design style. Our method downsizes MOSFETs inside a cell for power reduction without any modifications of wiring using accurate values of wire capacitance. We use a cell layout generation system called VARDS [1] that can generate cell layout with variable transistor width while keeping the location of terminals unchanged. In order to get the accurate cell delay time, our method utilizes four-dimensional look-up tables with four variables; gate widths of PMOS and NMOS transistors, input transition time, and load capacitance.

This paper is organized as follows. Section 2 explains the post-layout transistor sizing method. Cell layout generation, cell delay model, and transistor sizing algorithms are discussed. Section 3 demonstrates some experimental results. Finally, Sect. 4 concludes the discussion.

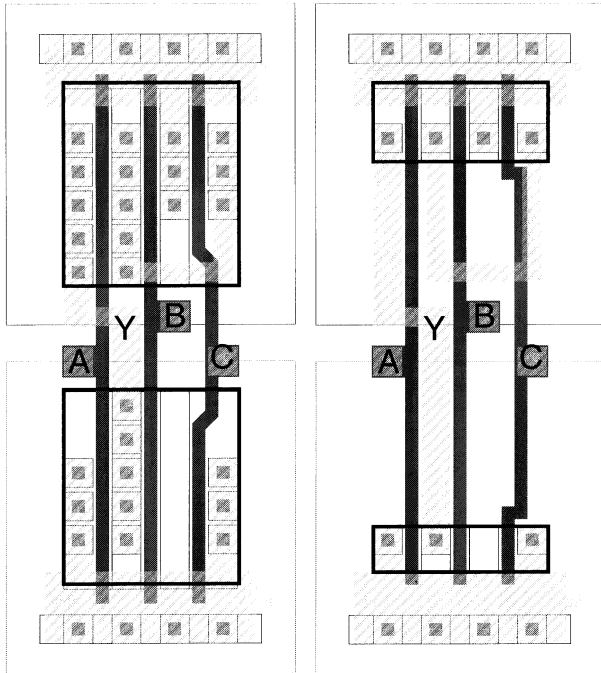## 2.   Post-Layout Transistor Sizing

In this section, we explain a transistor sizing method for power reduction preserving interconnects. We first discuss cell layout generation for post-layout transis-

(a) All transistor widths are the maximum.

(b) PMOS and NMOS transistor widths are different.

Fig. 1 Examples of AOI21 cell layout.

Fig. 2 Derivation of cell delay.

tor sizing. Next, we show a cell delay model that can calculate delay time for any PMOS and NMOS transistor sizes. Then, the noise margin constraints that guarantee the correct behavior of the circuits are discussed. Finally, we explain a transistor sizing algorithm for power reduction.

2.1 Cell Layout Generation

In order to apply the optimization result to the layout without any modifications of interconnects, the following features are required for cell layout generation.

- Each transistor width can be varied easily and flexibly.
- The location of each pin is fixed even when transistor widths are varied.

The fixed locations of input/output pins are needed to preserve interconnects. A cell layout generation system VARDS, which satisfies the above two requirements, has been proposed [1]. Figure 1 shows an example of AOI21 cells whose height is 13 interconnect pitches. The AOI21 cell in Fig. 1(a) is generated such that all transistor widths are the maximum. Figure 1(b) is an example that PMOS and NMOS transistor widths are different.

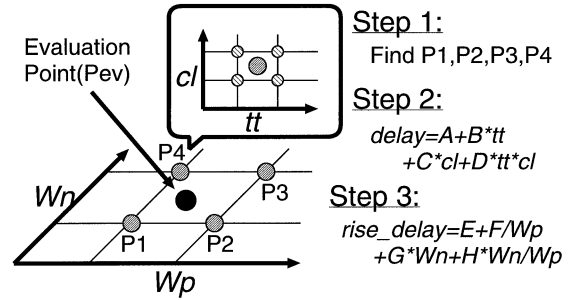## 2.2 Cell Delay Model

In the proposed method, PMOS and NMOS transis-

tors inside a cell are resized separately. Yet the transistor sizes of PMOS inside a cell are the same. Similarly, the sizes of NMOS are the same. Our method hence requires a cell delay model that has four variables, $W_p, W_n, tt,$ and cl, where $W_p(W_n)$ is the gate width of PMOS (NMOS) transistor, tt is the transition time of the input signal, and cl is the capacitive load. We build four-dimensional look-up tables with four variables $W_p, W_n, tt,$ and cl beforehand using a circuit simulator. Cell delay time is derived from the look-up tables using the following two-stage interpolation (Fig. 2). In the case of a multi-stage cell, we divide the cell into single-stage cells, and calculate the delay time of each single-stage cell. A single-stage cell is defined as a cell that consists of a PMOS block and an NMOS block connected at the output node. For example, AND cell is divided into NAND and INV.

Step1: Find four neighboring points (P_1, P_2, P_3, P_4) around the evaluation point (P_{ev}), in two-dimensional W_p-W_n space.

Step2: Calculate the delay time at each point of P_1, P_2, P_3, P_4 using Eq. (1) in two-dimensional tt-cl space.

Step3: Interpolate rise/fall delay time using Eq. (2/3) in W_p-W_n space from the four values at P_1, P_2, P_3, P_4 calculated at **Step2**.

$$delay = A + B \cdot tt + C \cdot cl + D \cdot tt \cdot cl, \tag{1}$$

$$rise\_delay = E + F \cdot \frac{1}{W_p} + G \cdot W_n + H \frac{1}{W_p} \cdot W_n, \tag{2}$$

$$fall\_delay = I + J \cdot W_p + K \cdot \frac{1}{W_n} + L \cdot W_p \frac{1}{W_n}, \tag{3}$$

$$energy = M + N \cdot W_p + O \cdot W_n + P \cdot W_p \cdot W_n, \tag{4}$$

where, $A, B, ..., P$ are coefficients to be determined such that the four values of the neighboring points are assigned to each interpolation equation. The transition time of the output signal is calculated similarly. In the case of the dissipated energy, Eq. (4) is used for the interpolation at **Step3**.

2.3 Noise Margin Constraints

Adequate amounts of noise margins are important to ensure the correct behavior of the circuits. The noise margins are defined as $NM_H = V_{OH} - V_{IH}$ and $NM_L = V_{IL} - V_{OL}$, where V_{OH} is the minimum HIGH output voltage, V_{IH} is the minimum HIGH input voltage, V_{IL} is the maximum LOW input voltage, and V_{OL} is the maximum LOW output voltage. The detailed definition of the noise margin is found in Ref. [7]. The noise margin depends on the ratio β_R, which is expressed as β_n/β_p, where $\beta_{n(p)}$ is the n(p)-device transconductance. We calculate the range of β_R that guarantees proper noise margins. The upper bound $\beta_{R(max)}$ can be derived from the following two equations [8], [9].

$$V_{IL} = \frac{2V_{out} - V_{DD} + V_{Tp} + \beta_{R(max)}V_{Tn}}{1 + \beta_{R(max)}}, \qquad (5)$$

$$\beta_{R(max)}(V_{IL} - V_{Tn})^2 = -(V_{out} - V_{DD})^2 + 2(V_{IL} - V_{DD} - V_{Tp})(V_{out} - V_{DD}), \qquad (6)$$

where V_{out} is the output voltage. Similarly, the lower bound $\beta_{R(min)}$ can be obtained from the following two equations.

$$V_{IH} = \frac{\beta_{R(min)}(2V_{out} + V_{Tn}) + V_{DD} + V_{Tp}}{1 + \beta_{R(min)}}, \qquad (7)$$

$$\beta_{R(min)}[2(V_{IH} - V_{Tn})V_{out} - V_{out}^2] = (V_{IH} - V_{DD} - V_{Tp})^2, \qquad (8)$$

where V_{Tp}, V_{Tn} are the threshold voltages of PMOS and NMOS transistors. We resize PMOS and NMOS transistors for power reduction within the range of $\beta_{R(min)} < \beta_R < \beta_{R(max)}$.

2.4 Transistor Sizing Algorithm

We devise a transistor sizing algorithm for power reduction based on sensitivity calculation. Our algorithm executes iterative optimization that decreases δ_{size} gradually, where δ_{size} is a variable that represents the amount of transistor width reduced in a single iteration.

Step1: Set δ_{size} to an initial value.
Step2: If δ_{size} is smaller than a pre-defined value, the optimization procedure finishes.
Step3: At each cell, evaluate the sensitivity, i.e. the amount of power reduction when the transistor widths decrease by δ_{size}. If the violations of noise margin or transition time constraints occur, sensitivity calculation is not performed.
Step4: Select the cell with the best sensitivity. If there are no cells with positive sensitivity, halve δ_{size} and go back to **Step2**.

Step5: Decrease the transistor widths of the selected cell by δ_{size}, and update the timing information of the cells affected by the downsizing. If delay violation occurs, cancel the downsizing.
Step6: Find the cell with the next best sensitivity. If there are no cells with positive sensitivity, go back to **Step3**. Otherwise, go back to **Step5**.

First, the above algorithm is executed for power reduction such that PMOS and NMOS transistors are resized simultaneously with the same β_n/β_p ratio. We next optimize power dissipation resizing PMOS and NMOS transistors independently, and we then get the final optimization result.

3. Experimental Results

In this section, some experimental results are shown. We first demonstrate the accuracy of the cell delay model based on look-up tables. We next show the power optimization results.

We generate cell layouts using VARDS [1] in a $0.35\,\mu$m process with three metal layers. The cell height is 13 interconnect-pitches, and the size ratio of PMOS and NMOS transistors is 1. In transistor sizing, we downsize MOSFETs within the range that VARDS can generate cell layouts. The maximum transistor width of standard driving-strength ($\times 1$) cells is $6.2\,\mu$m, and the value of W/L is 15.5. The transistor width can be reduced to $0.9\,\mu$m. Reference [10] reports that the optimal value of W/L around 20. The transistor width of our library is smaller than the reported value.

3.1 Accuracy of Cell Delay Model

We first examine the accuracy of the cell delay model. We use INV, 2-input NAND and 2-input NOR cells of standard driving-strength ($\times 1$) for this experiment. In the case of NAND and NOR cells, we evaluate the characteristics of the input pin that is close to the output terminal. We compare the delay time derived by the interpolation in Sect. 2.2 with the delay time evaluated by circuit simulation at the following 6561 points. The gate widths of PMOS and NMOS transistors (W_p, W_n) are varied to 0.9, 1.2, 1.5, 2.0, 2.5, 3.2, 4.0, 5.0, and $6.2\,\mu$m, respectively. The layout parameters of MOSFETs, such as length, width, area of diffusions, and perimeter of diffusions, are extracted from the cell layouts generated by VARDS [1]. The evaluation points of the input transition time (tt) are 0.02, 0.125, 0.25, 0.375, 0.5, 0.65, 0.8, 1.0, and 1.2 ns, also the points of load capacitance (cl) are 0.005, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.35 and 0.5 pF. The combinations of W_p and W_n that make the noise margin smaller than $0.25V_{DD}$ are excluded. When the absolute value of the delay time is extremely small, the relative error becomes meaninglessly large while absolute error is sufficiently small. We

Table 1 Average error of cell delay model based on look-up tables.

Cell	Transition	Variables of Interpolation		
		$W_p, W_n,$ tt, cl	tt, cl $(W_p, W_n$ fixed)	W_p, W_n $(tt, cl$ fixed)
INV	rise	0.003 ns 1.9%	0.002 ns 1.4%	0.001 ns 1.0%
	fall	0.004 ns 1.3%	0.002 ns 0.9%	0.002 ns 0.4%
NAND2	rise	0.003 ns 2.1%	0.002 ns 1.5%	0.001 ns 0.9%
	fall	0.005 ns 1.0%	0.002 ns 0.6%	0.003 ns 0.4%
NOR2	rise	0.002 ns 1.2%	0.001 ns 0.8%	0.001 ns 0.6%
	fall	0.005 ns 1.2%	0.002 ns 0.7%	0.003 ns 0.5%

Table 2 Error range of cell delay model based on 4-dimensional look-up tables.

Cell	Transition	Error Range (ns)
INV	rise	-0.029–0.008
	fall	-0.033–0.002
NAND2	rise	-0.030–0.009
	fall	-0.063–0.006
NOR2	rise	-0.028–0.009
	fall	-0.034–0.001

hence do not calculate the relative error when the delay time is less than 0.01 ns. The size of look-up tables is $5 \times 5 \times 5 \times 5$. The evaluation points of input transition time in look-up tables are 0.02, 0.25, 0.5, 0.8, 1.2 ns, and the points of load capacitance are 0.005, 0.05, 0.1, 0.2, 0.5 pF. The points of W_p and W_n in look-up tables are 0.9, 1.5, 2.5, 4.0 and 6.2 μm.

Table 1 shows the error of the cell delay model. The average error of the delay time calculated from 4-dimensional look-up tables of W_p, W_n, tt, and cl is about 2%. The interpolation error of the delay time derived in W_p-W_n space is comparable with the error calculated in tt-cl space. This means that the interpolation in W_p-W_n space does not deteriorate the accuracy of cell delay so much. Compared with the interpolation in tt-cl space, the average error of our 4-dimensional table model increases by 0.5%. Table 2 shows the error range of our cell delay model. The maximum error of overestimation is smaller than that of underestimation. The maximum overestimation of our cell delay model is below 0.01 ns.

3.2 Power Optimization Results

We show the results of power optimization. The circuits used for the experiments are an ALU in a DSP for mobile phone [13] (dsp_alu) and the circuits included ISCAS85 and LGSynth93 benchmark sets (C3540, alu4, C7552, des). These circuits are synthesized under two different constraints [11]: minimizing the circuit delay, and minimizing the circuit area. Also two transition

time constraints, 0.5 ns and 1.0 ns are given. Thus, each circuit is synthesized under four different constraints in total. The layouts of the circuits synthesized for minimizing the circuit delay are generated using a commercial placement and routing tool [12] and a logic synthesis tool [11] as follows:

1. The cells are placed with the timing-driven placement option.
2. The circuit delay is minimized by cell (discrete) sizing and buffer insertion with the back-annotated information of cell placement. The result is applied to the layout using the ECO (Engineering Change Order) technique.
3. The global routing and track assignment are executed.
4. The circuit delay is minimized by cell sizing and buffer insertion with the back-annotated information of global routing and track assignment. The layout is modified by the ECO.
5. The detail routing is performed.
6. The power dissipation is minimized by cell sizing keeping the circuit delay. The result is applied to the layout by the ECO.

The layouts of the circuits synthesized for minimizing the circuit area are generated as follows:

1. The cells are placed with the timing-driven placement option.
2. The global routing, track assignment and detail routing are executed.
3. The power dissipation is minimized by cell sizing. The layout is modified by the ECO.

We utilize the wire capacitance values extracted from the layouts for transistor sizing. The circuit scale is 1619 to 12858 cells. The cell library used for generating initial circuits includes six varieties in driving-strength for INV and BUF ($\times 1$, $\times 2$, $\times 3$, $\times 4$, $\times 6$ and $\times 8$). In the case of NAND2, NAND3, AND2, AND3, NOR2, NOR3, OR2, OR3, AOI21, OAI21 cells, there are four varieties ($\times 1$, $\times 2$, $\times 3$, $\times 4$). The circuit delay time is evaluated by a transistor-level static timing analysis tool [14], and the power dissipation is estimated by a transistor-level power simulator [15]. The input patterns are randomly generated with a transition probability of 0.5. The number of applied patterns is 100, which is the adequate number for power estimation at circuit level [16]. The cycle time of the input patterns is 100 ns.

We optimize power dissipation under the delay constraints of the initial circuits' delay time, i.e. the circuits are optimized keeping the circuit delay times unchanged. The initial value of δ_{size} in the optimization algorithm (Sect. 2.4) is 12.4 μm, and the termination value is 0.1 μm. The constraints that the noise margin is larger than $0.25V_{DD}$ are given. In order to examine the effectiveness of continuous transistor siz-

Table 3 Power optimization results (cell height: 13 interconnect pitches).

Circuit	Constraints		Initial Circuits		Cell (Discrete) Sizing			Proposed Method				#cells
	Tran. Time (ns)	Design†	Delay (ns)	Power (mW)	Delay (ns)	Power (mW)	Power Reduction (%)	Delay (ns)	Power (mW)	Power Reduction (%)	CPU Time (s)	
C7552	0.5	F	4.3	13.6	4.3	9.8	28	4.5	3.8	72	360	1986
		MA	6.5	13.1	7.1	9.3	29	6.5	3.3	75	177	1691
	1.0	F	3.2	12.2	3.5	7.0	43	3.5	3.6	70	360	2044
		MA	5.2	8.8	5.8	3.3	63	5.3	2.3	74	117	1619
des	0.5	F	3.4	14.7	3.4	11.3	23	3.4	6.4	56	1010	3515
		MA	4.2	12.3	4.6	9.5	23	4.4	4.9	60	630	2924
	1.0	F	2.6	13.6	2.9	9.8	28	3.0	6.4	53	971	3439
		MA	3.6	9.3	4.2	5.5	41	3.7	4.0	57	436	2864
dsp_alu	0.5	F	9.2	89.4	9.1	75.9	15	9.8	44.2	51	22501	12858
		MA	18.9	87.0	19.6	75.3	13	18.9	44.7	49	16249	11994
	1.0	F	6.8	69.9	7.4	48.4	31	8.1	33.2	53	10262	12679
		MA	15.3	59.1	17.5	40.4	32	16.7	27.2	54	5011	10963
Average	-	-	-	-	-	-	31	-	-	60	-	-

Design Constraints†: Fastest (F) or Minimum-Area (MA).

ing, the circuits are also optimized by cell (discrete) sizing as follows. We add weak driving-strength cells ($\times 0.15$, $\times 0.5$) to the standard cell library. The transistor widths of $\times 0.15$ and $\times 0.5$ cells are $0.9\,\mu m$ and $3.1\,\mu m$. We minimize power dissipation by cell sizing, keeping the circuit delay [11].

Table 3 shows the power optimization results. "CPU Time" represents the CPU time required for power optimization on an Alpha Station. In the case of cell sizing, power dissipation is reduced by 31% on average. The weak driving-strength cells are effective in power reduction. The amount of power reduction, however, is small compared with the proposed method. Our method reduces power dissipation by 75% maximum and 60% on average. The continuous transistor sizing and tuning the ratio of PMOS and NMOS widths contribute to further power reduction. The power reduction in small circuits is larger than the one in large circuits, because large circuits usually have heavier wire load. In the case of the largest circuit dsp_alu, the power dissipation is reduced by about 50%.

In some circuits, the circuit delay evaluated by a transistor-level timing analyzer [14] increases although the delay time calculated by the table-base static timing analysis is unchanged. One reason of this delay increase is that the path-balanced circuits become sensitive to the error of cell delay model [17]. Here, we explain this problem of delay increase briefly. The circuit delay, which is the maximum path delay time in a circuit, $D_{circuit}$ is represented as follows.

$$D_{circuit} = \max_i D_i \quad (i = 1, 2, ..., n), \qquad (9)$$

where D_i is the path delay time of the i-th path, and n is the number of the paths in the circuit. Suppose D_i fluctuates due to the error of cell delay model. Intuitively speaking, when we choose the maximum sample from a large population, the probability that we can find the large sample becomes high. The proposed method equalizes the delay times of many paths, which

corresponds to the increase of the substantial population. Therefore the circuit delay of the optimized circuit becomes larger than that of the initial circuit, though the accuracy of cell delay model is the same. Further examination of the reasons is required, considering the accuracy of the delay calculation tool as well.

The proposed method resizes PMOS and NMOS transistors independently. We evaluate the effect of this independent sizing. In our algorithm, we first resize PMOS and NMOS transistors simultaneously keeping the PMOS and NMOS size ratio (Phase 1). After the optimization of Phase 1, PMOS and NMOS transistors are downsized separately (Phase 2). Table 4 lists the amount of power reduction at Phase 1 and Phase 2. The independent sizing of PMOS and NMOS transistors reduces power dissipation furthermore by 10% maximum and 5% on average.

We next show the power optimization results when the initial circuits are generated using a low-power cell library. The cell-height of this low-power library is 9 interconnect pitches, and the standard transistor size is $3.4\,\mu m$. The varieties of driving strength for INV and BUF are $\times 1$, $\times 2$, $\times 3$, $\times 4$, $\times 6$ and $\times 8$. In the case of other cells, the varieties of $\times 1$, $\times 2$, $\times 3$, and $\times 4$ are included. The delay time of each initial circuit is given to the optimization procedure as the delay constraint. The results are shown in Table 5. Even when the low-power cell library is used for initial circuits, our method reduces power dissipation by more than 40% on average.

Hereafter we examine the optimization result in detail of des circuit generated for minimizing circuit delay under the transition time constraint of $0.5\,ns$. This circuit is designed using the cell library whose cell height is 13 interconnect pitches. Figure 3(a) shows a part of the initial layout. Figure 3(b) corresponds to the transistor-sized layout of the same location. The transistor sizes inside cells become different in instance by instance. PMOS and NMOS transistors inside each

Table 4 Power reduction at each optimization phase.

Circuit	Constraints		Phase 1[‡]		Phase 2[‡]	
	Tran. Time (ns)	Design[†]	Power Reduction (%)	CPU Time (s)	Power Reduction (%)	CPU Time (s)
C7552	0.5	F	68	260	4	100
		MA	71	123	4	54
	1.0	F	66	254	4	106
		MA	72	87	2	30
des	0.5	F	51	587	5	423
		MA	54	375	6	255
	1.0	F	47	606	6	365
		MA	53	262	4	174
dsp_alu	0.5	F	43	10902	8	11599
		MA	39	7492	10	8757
	1.0	F	49	6473	4	3789
		MA	49	3391	5	1620
Average	-	-	55	-	5	-

Design Constraints[†]: Fastest (F) or Minimum-Area (MA).
Phase 1[‡]: PMOS and NMOS transistors are resized simultaneously keeping the same ratio of PMOS and NMOS sizes.
Phase 2[‡]: PMOS and NMOS transistors are resized independently after Phase 1.

Table 5 Power optimization results (cell height: 9 interconnect pitches).

Circuit	Constraints		Power Reduction (%)
	Transition Time (ns)	Design	
C7552	0.5	Fastest	54
		Min-Area	56
	1.0	Fastest	51
		Min-Area	56
des	0.5	Fastest	38
		Min-Area	40
	1.0	Fastest	40
		Min-Area	41
dsp_alu	0.5	Fastest	35
		Min-Area	30
	1.0	Fastest	36
		Min-Area	35
Average	-	-	43

cell are resized separately. Also the routing is perfectly preserved. Our method generates cell layouts on the fly according to the optimization results, and replaces cells without any interconnect modifications.

We next show the distributions of transistor widths in the optimized circuit (Fig. 4). The transistor width of a standard driving-strength (×1) cell is $6.2\,\mu$m, and the transistor width can be reduced to $0.9\,\mu$m. Many MOSFETs are downsized close to the lower limit of $0.9\,\mu$m. Compared with PMOS transistors, the gate widths of NMOS transistors are small. The sum of PMOS gate widths is 14.0 mm, which is 24% larger than the sum of NMOS gate widths (11.3 mm).

Figure 5 expresses the slack distributions of the initial and optimized circuits. By transistor sizing, the number of the cells with 0 or almost 0 slack increases drastically. The sum of slack in the optimized circuit is 1259 ns, whereas the sum of slack in the initial circuit is 3261 ns. The total slack is reduced by 61%.

We then demonstrate the capacitance reduction in the circuit (Fig. 6). Our method does not modify any interconnects, so wire capacitance does not change. The gate capacitance of MOSFETs is reduced by 70%, which results in 49% reduction of the total capacitance.

We finally show the peak current reduction. We apply 100 input patterns, and evaluate the peak current at each time-step within a cycle. Figure 7 indicates the peak current of the initial and optimized circuits. The horizontal axis represents the time within a cycle of 3.4 ns. The peak current is reduced by 66%. Path-balancing effect of our method contributes to the peak current reduction, as well as gate capacitance reduction. The transition timing of each cell is well distributed throughout a cycle. Reducing the peak current is effective to avoid IR drop problem. Also, the current reduction is a useful way to evade electromigration. Thus, our method can increase the tolerance to IR drop and electromigration problems, and contribute to high-reliability LSI design.

3.3 Effectiveness of Interconnect Preservation

The proposed method optimizes a detail-routed circuit without any wiring modifications. We verify the effectiveness of the interconnect preservation. In a conventional transistor sizing method, the layout is modified using an ECO (Engineering Change Order) technique in order to preserve the placement and wiring as much as possible. But a certain amount of variation in wire capacitance is not avoidable.

We examine the effect of this capacitance variation statistically. We assume that the wire capacitance varies according to a normal distribution $N(m, \sigma)$ because of interconnect modifications, i.e. ECO. The mean m is the initial value used in transistor sizing, and the standard deviation σ is 20% of the initial value.

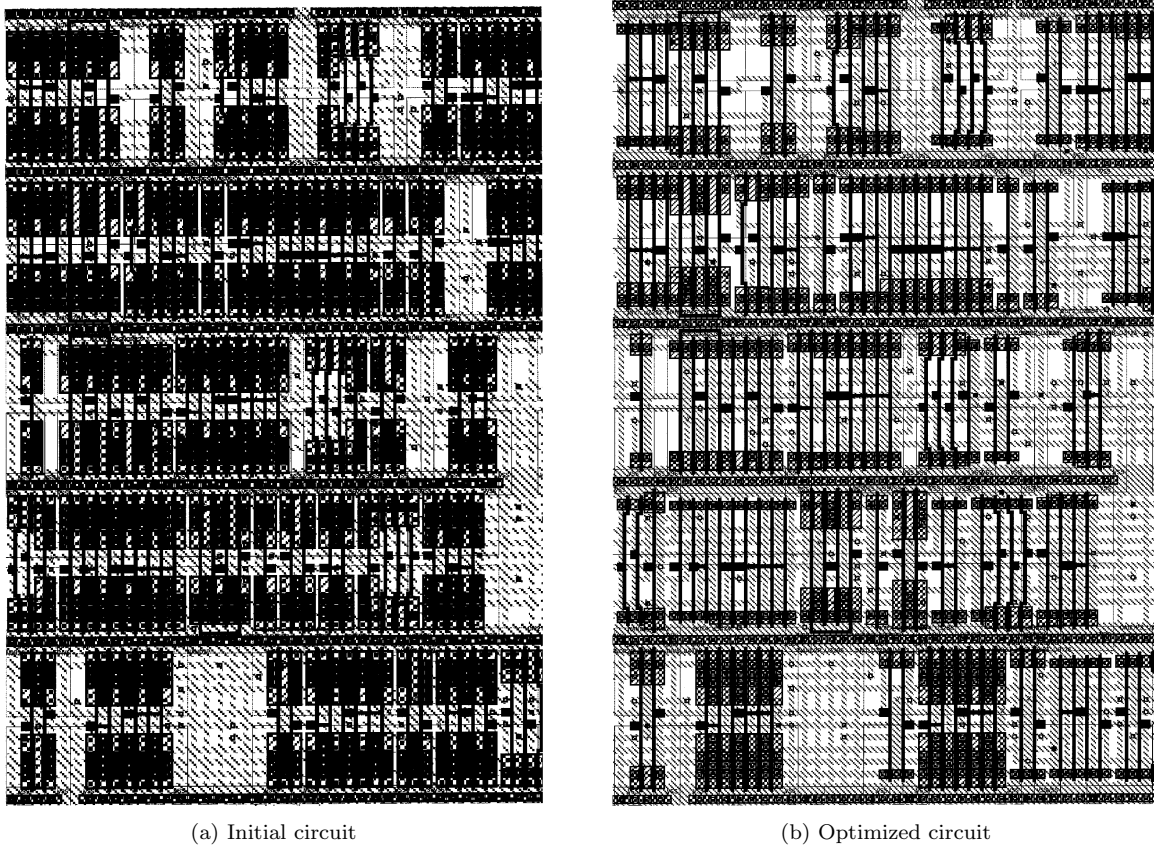(a) Initial circuit

(b) Optimized circuit

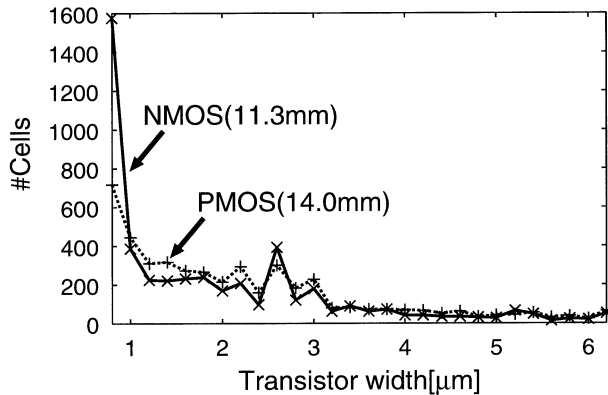**Fig. 3**   A part of layout (des, fastest, transition time constraint 0.5 ns).

Fig. 4   Distribution of transistor widths (des, fastest, transition time constraint 0.5 ns).
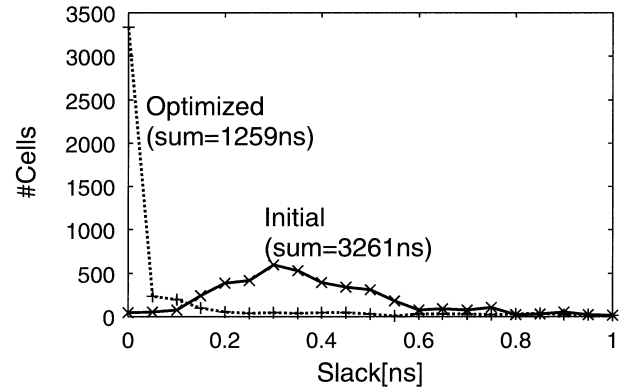

Fig. 5   Distribution of slack (des, fastest, transition time constraint 0.5 ns).

This fluctuation model is a simple assumption and it is not a realistic one based on the practical ECO behavior. The delay distribution is obtained using a Monte Carlo technique. We assign each interconnect capacitance randomly according to the given distribution, and evaluate the circuit delay using a static timing analysis technique. This process corresponds to one delay evaluation, and the total number of delay evaluation is 10,000. Figure 8 shows the delay variation in the optimized des circuit. As you see, the interconnect modi-

fications increase the circuit delay, although the mean of the interconnect capacitance m is the same with the initial capacitance. The circuit whose delay time is the same with the initial circuit (3.4 ns) can be hardly obtained. The circuit delay of "mean+3σ" is 3.7 ns, which is larger than the delay without wiring modifications by 9%. As for each path delay, the delay times of all the paths do not increase, or rather the delay times of some paths decrease. However the circuit delay is increased by max operation, because the circuit delay is
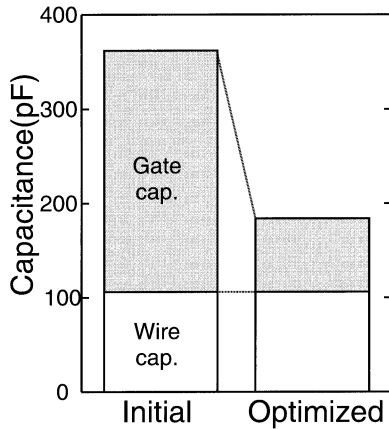
Fig. 6 Capacitance reduction (des, fastest, transition time constraint 0.5 ns).
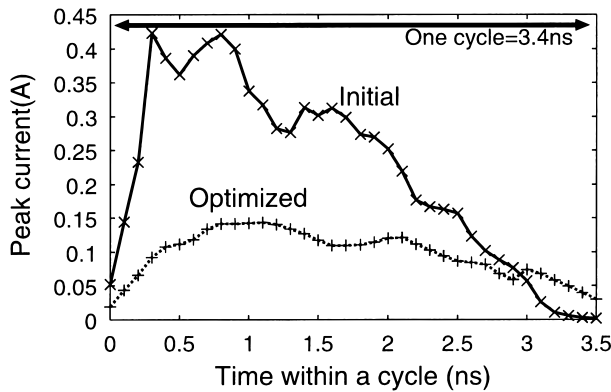

Fig. 7 Peak current reduction (des, fastest, transition time constraint 0.5 ns).
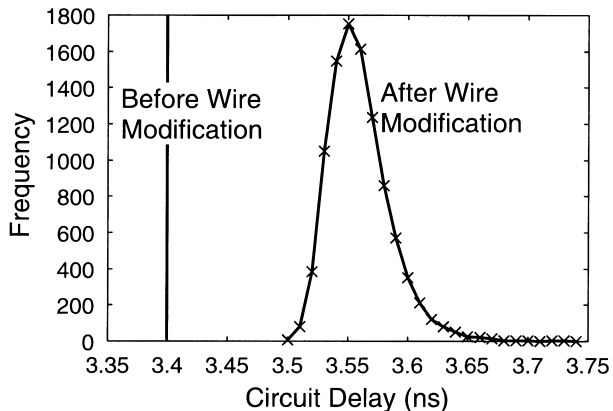

Fig. 8 Delay variation caused by interconnect modifications (des, fastest, transition time constraint 0.5 ns).

define as the maximum path delay (Eq. (9)). On the other hand, the proposed method can avoid this delay increase, thanks to the interconnect preservation.

4. Conclusion

We propose a power reduction method that downsizes MOSFETs in a cell without any interconnect modifications. The effectiveness of our method is experimentally verified using 3 benchmark circuits. The power dissipation is reduced by 75% maximum and 60% on average without delay increase. The amount of power reduction by our method is 30% larger than the usual discrete sizing that uses some small cells. We verify that our method also contributes to high-reliability LSI design. Our future work is to construct a methodology that can guarantee the circuit delay time with the assistance of statistical static timing analysis [18].

Acknowledgments

References

[1] T. Hashimoto and H. Onodera, "Layout generation of primitive cells with variable driving strength," Proc. SASIMI, pp.122–129, 2000.

[2] J.P. Fishburn and A.E. Dunlop, "Tilos: A posynomial programming approach to transistor sizing," Proc. ICCAD, pp.326–328, 1985.

[3] S.S. Sapatnekar, V.B. Rao, P.M. Vaidya, and S.M. Kang, "An exact solution to the transistor sizing problem for CMOS circuits using convex optimization," IEEE Trans. CAD, vol.12, pp.1621–1634, Nov. 1993.

[4] M. Yamada, S. Kurosawa, R. Nojima, N. Kojima, T. Mitsuhashi, and N. Goto, "Synergistic power/area optimization with transistor sizing and wire length minimization," IEICE Trans. Electron., vol.E78-C, no.4, pp.441–446, April 1995.

[5] M.R. C.M. Berkelaar and J.A. G. Jess, "Gate sizing in MOS digital circuits with linear programming," Proc. EDAC, pp.217–221, 1990.

[6] G. Chen, H. Onodera, and K. Tamaru, "An iterative gate sizing approach with accurate delay evaluation," Proc. ICCAD, pp.422–427, 1995.

[7] N. Weste and K. Eshraghian, Principles of CMOS VLSI Design A Systems Prespective, Addison-Wesley, 1985.

[8] J.P. Uyemura, Funcamentals of MOS Digital Integrated Circuits, Addison-Wesley, 1988.

[9] H.Y. Chen and S.M. Kang, "A new circuit optimization technique for high performance CMOS circuits," IEEE Trans. CAD, vol.10, no.5, pp.670–677, May 1991.

[10] D. Sylvester and K. Keutzer, "Getting to the bottom of deep submicron," Proc. ICCAD, pp.203–211, 1998.

[11] Design Compiler Reference Manual. Synopsys, Inc., CA, 1999.

[12] ApolloII Reference Manual. Avant!, CA, 1999.

[13] T. Iwahashi, T. Shibayama, M. Hashimoto, K. Kobayashi, and H. Onodera, "Vector quantization processor for mobile video communication," Proc. ASIC/SOC Conf., pp.75–79, 2000.

[14] PathMill Reference Manual. Synopsys, Inc., CA, 1999.

[15] PowerMill Reference Manual. Synopsys, Inc., CA, 1999.

[16] D. Brand and C. Visweswariah, "Inaccuracies in power estimation during logic synthesis," Proc. ICCAD, pp.388–394, 1996.

[17] M. Hashimoto and H. Onodera, "Delay increase by performance optimization," Proc. ISCAS, vol.5, pp.379–382, 2001.

[18] M. Hashimoto and H. Onodera, "A performance optimization method by gate resizing based on statistical static timing analysis," IEICE Trans. Fundamentals, vol.E83-A, no.12, pp.2558–2568, Dec. 2000.

Masanori Hashimoto   received the B.E., M.E. and Ph.D. degrees in Communications and Computer Engineering from Kyoto University, Kyoto, Japan, in 1997, 1999, and 2001, respectively. Since 2001, he has been an Instructor in Department of Communications and Computer Engineering, Graduate School of Informatics, Kyoto University. He received the IEEE Solid-State Circuit Society Japan Chapter Award in 1999. His research interest includes computer-aided-design for digital integrated circuits. He is a member of IEEE and IPSJ.

Hidetoshi Onodera   received the B.E., and M.E., and Dr.Eng. degrees in Electronic Engineering from Kyoto University, Kyoto, Japan, in 1978, 1980, 1984, respectively. Since 1983 he has been an Instructor (1983–1991), an Associate Professor (1992–1998), a Professor (1999–) in the Department of Communications and Computer Engineering, Graduate School of Infomatics, Kyoto University. His research interests include computer-aided-design for integrated circuits, and analog and mixed analog-digital circuits design. He is a member of the Information Processing Society of Japan, and IEEE.