Reliable Crosstalk-Driven Interconnect Optimization

IRIS HUI-RU JIANG
National Chiao Tung University
SONG-RA PAN
University of California, Santa Barbara
YAO-WEN CHANG
National Taiwan University
and
JING-YANG JOU
National Chiao Tung University

As technology advances apace, crosstalk becomes a design metric of comparable importance to area and delay. This article focuses mainly on the crosstalk issue, specifically on the impacts of physical design and process variation on crosstalk. While the feature size shrinks below $0.25 \ \mu m$, the impact of process variation on crosstalk increases rapidly. Hence, a crosstalk insensitive design is desirable in the deep submicron regime. In this article, *crosstalk sensitivity* is referred to as the influence of process variation on crosstalk in a circuit. We show that the lower bound of crosstalk sensitivity grows quadratically, while that of crosstalk increases linearly. Therefore, designers should also consider crosstalk sensitivity, when optimizing other design objectives such as crosstalk, area, and delay. According to our modeling, these objectives are all in posynomial forms, and thus the multi-objective optimization problem can optimally be solved by Lagrangian relaxation. Experimental results show that our method is effective and efficient. For instance, a circuit of 2856 gates and 5272 wires is optimized using 13-minute runtime and 2.8-MB memory on a Pentium III 1.0 GHz PC with 256-MB memory. In particular, by relaxing Lagrange multipliers to the critical paths, it takes only two iterations for all solutions to converge to the global optimal, which is much more efficient than related previous work. This relaxation scheme provides a key insight into the rapid convergence in Lagrangian relaxation.

---

1. INTRODUCTION

As the feature size shrinks into the deep submicron regime, crosstalk is becoming a new design challenge of comparable importance to area and timing. This article focuses mainly on the crosstalk issue, specifically on the impacts of physical design and process variation on crosstalk.

In the deep submicron era, the interconnect delay starts to dominate the delay in a circuit. Gate delay declines as technology progresses while local interconnect delay remains the same and global interconnect increases quadratically. In addition, the growing coupling capacitance magnifies the effective loading of interconnect and induces noise to interfere signal propagation. Moreover, interconnect may become a bottleneck of the continuation of Moore's law, which has perfectly forecasted the legend of semiconductor industry so far [Semiconductor Industry Association 1999]. This trend forces designers to endeavor after interconnect optimization. Recent literature intensively focuses on crosstalk, mainly depending on the coupling capacitance between wires. The coupling information can be completely extracted after detailed routing. The typical techniques to reduce the coupling capacitance include buffer insertion [Alpert et al. 1998; Zhang and Sapatnekar 2004], wire permutation [Gao and Liu 1993], wire perturbation [Saxena and Liu 1999], wire shielding [Rabaey 1996], wire sizing [Jiang et al. 2000], and gate sizing [Hashimoto et al. 2002; Becer et al. 2003; Jiang et al. 2000; Sinha et al. 2004; Sinha and Zhou 2004]. Since wire and gate sizing can be done by incremental changes, they are suitable for post-layout optimization.

On the other hand, designers have to consider the impact of *subwavelength lithography* when the feature size becomes smaller than the wavelength of the light shining through the mask [Kahng and Pati 1999]. Subwavelength lithography could cause variations in the dimensions of components. This type of process variation may create considerable unexpected circuit behavior [Rabaey 1996] and, in the worst case, could offset the optimization done by physical design. Hence, a reliable design of insensitivity to process variation is desirable.

Considering crosstalk and process variation together, this article first raises an issue on *crosstalk sensitivity*, which reflects the influence of process variation on the crosstalk in a circuit. Crosstalk sensitivity is measured by the first derivative of crosstalk with respect to wire width, thus it could be considered during the wire sizing stage in post-layout optimization. In Section 3, we derive the formula for crosstalk sensitivity. Based on our formula, as technology scales down, *the lower bound* of crosstalk sensitivity increases quadratically, while *the lower bound* of crosstalk increases linearly. This fact shows that process variation affects crosstalk more than physical design does. Currently, most research attention is directed to crosstalk minimization; however, our work reveals that
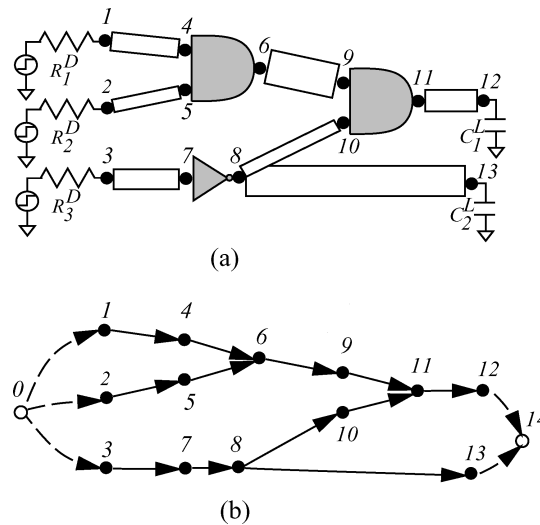
Fig. 1. (a) A combinational circuit with 3 primary inputs, 2 primary outputs, 3 gates and 7 wires, where the gate and wire sizes can be varied. (b) The corresponding circuit graph, where two artificial nodes, 0 and 14, are added.

a reliable design with crosstalk insensitivity is also desirable in deep submicron technology.

When technology advances apace, designers are simultaneously challenged by multiple objectives, including crosstalk, crosstalk sensitivity, area, and delay. Our modeling for these objectives shows that they can simultaneously be handled by gate and wire sizing in post-layout optimization. According to our modeling, these design metrics are all in posynomial forms [Hillier and Lieberman 1990], and thus the multi-objective optimization problem can optimally be solved by the Lagrangian relaxation method. Moreover, our method can easily be extended to other objectives, for example, energy in high-performance circuitry. The experimental results show that our method is effective and efficient. For instance, a circuit of 2856 gates and 5272 wires is optimized using 13-minute runtime and 2.8-MB memory. We note that all solutions rapidly converge to the global optimal after only two iterations by relaxing Lagrange multipliers to the critical paths of the benchmark circuits. The relaxation scheme provides a key insight into the rapid convergence in Lagrangian relaxation. To the best knowledge of authors, this kind of efficiency has never been reported in related previous work.

2. CIRCUIT INTERPRETATION

A digital circuit can be divided into combinational and sequential parts. After applying peripheral retiming [Malik et al. 1991], we can perform optimization on the combinational part to achieve our design objectives. Hence, this article focuses on combinational circuits and interprets them in the way similar to that adopted in Chen et al. [1999].

Given a combinational circuit with s primary inputs, t primary outputs, and n gates and/or wires as shown in Figure 1(a), we construct the corresponding
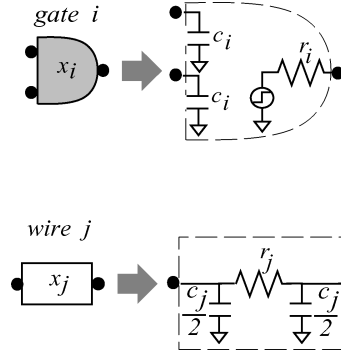
Fig. 2. A gate/wire is modelled as a combination of RC elements. A gate is the loading of its upstream, but the driver of its downstream. A wire is represented by the π model.

circuit graph as depicted in Figure 1(b). Each primary input or primary output has a respective corresponding input driver or output load. A *component* is a circuit element that may be a gate, a wire, or an input driver. A node is located at the output of a component which either connects two components or links one primary output to one output load. A circuit graph $H = (V, E)$ is a directed acyclic graph. The set of nodes $V = G \cup W \cup R \cup \{\tilde{s}\} \cup \{\tilde{t}\}$ contains the set G of gates, the set W of wires, the set R of input drivers, as well as two artificial nodes—the source $\tilde{s}$ and the sink $\tilde{t}$. On the other hand, the set E of edges represents the connections between nodes. An edge (i, j), an ordered pair, links node i to node j if data flow from node i to node j. Additional edges are added to connect $\tilde{s}$ to input drivers and link primary outputs to $\tilde{t}$. The index of each node is done by the topological sort [Cormen et al. 2001]. We set the indices of $\tilde{s}$ and $\tilde{t}$ as 0 and $m = n + s + 1$ respectively. In addition, $input(i) = \{j | (j, i) \in E\}$, and $output(i) = \{j | (i, j) \in E\}$.

Figure 2 shows the analytical models for gates and wires used throughout this paper. We choose the π model [Rabaey 1996] to approximate wire behavior. For a gate i of size x_i, the resistance r_i is $\hat{r}_i/x_i$, and the capacitance c_i is $\hat{c}_i x_i$, where $\hat{r}_i$ and $\hat{c}_i$ are its unit-size resistance and capacitance. For a wire j of size x_j, the resistance r_j is $\hat{r}_j/x_j$, and the capacitance c_j is $\hat{c}_j x_j + f_j + 2C_{cj}$, where $\hat{r}_j$ is its unit-size resistance, and $\hat{c}_j$, f_j, and $2C_{cj}$ are its respective unit-size, fringing, and worst case coupling capacitance. We will detail the calculation of the coupling capacitance in Section 3.1. Later, by incorporating the coupling capacitance into the wire capacitance, we can directly consider the coupling effect on delay, and even on power. In addition, an input driver $i, 1 \leq i \leq s$, is treated as a gate whose x_i is *always* 1 and $\hat{r}_i$ equals R_i^D. With the circuit model, a combinational circuit is transformed into a network with resistors and capacitors, as illustrated in Figure 3. In the transformed circuit, for $1 \leq i \leq n+s$, $upstream(i)$ is the set of all the nodes except i on i's upstream paths; similarly, $downstream(i)$ is the set of i and all the nodes on its downstream paths. The Elmore delay [Elmore 1948] is used as the delay of a component; the delay D_i of node i is $r_i C_i$, where C_i is the downstream capacitance. In the circuit graph H of a circuit, each node i is associated with all the above parameters. Thus, we can optimize a circuit through manipulating the corresponding circuit graph.
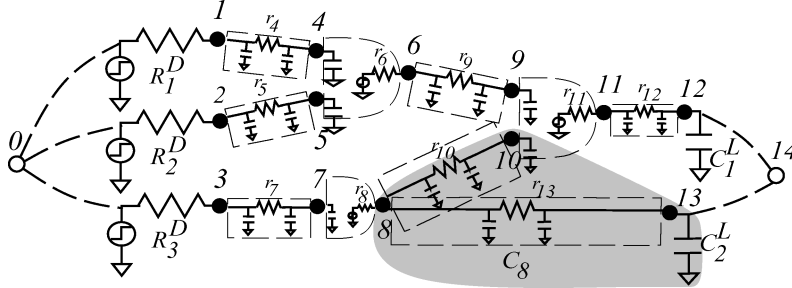
Fig. 3.   A circuit is transformed to an RC network. As an example, the delay $D_8 = r_8 C_8$, where C_8 is the sum of all the capacitors in the shaded area.

$\hat{f}_{ij}$ is the unit−length fringing capacitance between i and j
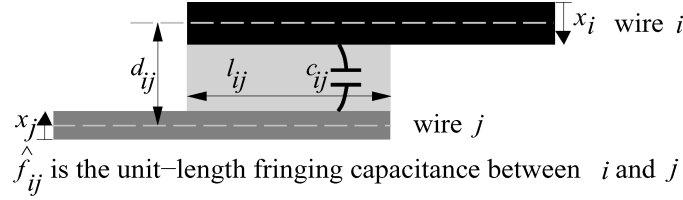
Fig. 4.   Two neighboring wires has coupling capacitance.

3. CROSSTALK AND CROSSTALK SENSITIVITY

In this section, we will discuss the formulae for crosstalk and crosstalk sensitivity.

3.1 Crosstalk-Coupling Capacitance

In this article, the coupling capacitance is used as the quantity of crosstalk. Figure 4 gives an instance where a coupling capacitance exists between two parallel wire segments i and j probably belonging to different routing trees. The coupling capacitance c_{ij} between two neighboring wires i of size x_i and j of size x_j is directly proportional to the overlap length l_{ij} but inversely proportional to the center-to-center distance d_{ij}.

$$c_{ij} = \frac{\hat{f}_{ij} l_{ij}}{d_{ij} - \frac{x_i + x_j}{2}}$$
$$= \left(\frac{\hat{f}_{ij} l_{ij}}{d_{ij}} \right) \left(1 - \frac{x_i + x_j}{2 d_{ij}} \right)^{-1} (>0). \tag{1}$$

As can be seen in Eq. (1), wire sizing affects crosstalk, thus causing variation on delay and disturbance on signal integrity. The first term in Eq. (1), $\hat{f}_{ij} l_{ij}/d_{ij}$, is a constant extracted by technology files and design circuits, while the second term, $(1-(x_i+x_j)/2d_{ij})^{-1}$, could be varied by wire sizing. Let $x = (x_i + x_j)/2d_{ij}$, the second term becomes $(1-x)^{-1}$, $0 < x < 1$. It can be seen that c_{ij} is a positive quantity, and its lower bound $\tilde{c}_{ij}$ is $\hat{f}_{ij} l_{ij}/d_{ij}$. As a result, the crosstalk lower bound $\tilde{c}_{ij}$ is inversely proportional to wire spacing, thus increasing linearly with technology advancing. $(1-x)^{-1}$ can be expressed by Taylor series $\sum_{n=0}^{\infty} x^n$.

If $(1-x)^{-1}$ is approximated by the first k terms in the series, then the *error ratio* is x^k. When $k = 2$,

$$c_{ij} \approx \tilde{c}_{ij}\left(1 + \frac{x_i + x_j}{2d_{ij}}\right). \tag{2}$$

The *neighborhood* $N(i)$ of wire i is defined as the set of its adjacent wires; the *dominating index* $I(i)$ of $N(i)$ of wire i is defined as the set of adjacent wires with the indices greater than i. For instance, if wires 7 and 4 are adjacent to wire 5, then $N(5) = \{7, 4\}$ and $I(5) = \{7\}$. Hence, the worst case coupling capacitance $2C_{ci}$ between wire i and its neighbors is $2\sum_{j \in N(i)} c_{ij}$. By Eq. (2), the capacitance c_i of wire i is

$$\begin{aligned}
c_i &= \hat{c}_i x_i + f_i + 2C_{ci} = \hat{c}_i x_i + f_i + 2\sum_{j \in N(i)} c_{ij} \\
&= \hat{c}_i x_i + f_i + 2\sum_{j \in N(i)} \tilde{c}_{ij}\left(1 + \frac{x_i + x_j}{2d_{ij}}\right) \\
&= \left(\hat{c}_i + 2\sum_{j \in N(i)} \hat{c}_{ij}\right)x_i \\
&\quad + \left(f_i + 2\sum_{j \in N(i)} \frac{\hat{f}_{ij}l_{ij}}{d_{ij}}\left(1 + \frac{x_j}{2d_{ij}}\right)\right).
\end{aligned} \tag{3}$$

In Eq. (3), we consider the worst-case coupling capacitance $2C_{ci}$. If the switching behavior of wires is available, the worst case coupling capacitance $2C_{ci}$ in wire capacitance c_i can be substituted by the effective coupling capacitance [Jiang et al. 2000].

On the other hand, by incorporating the coupling capacitance into the wire capacitance, we can directly consider the coupling effect on delay, and even on power. Note that Eq. (3) is posynomial (positive polynomial) [Hillier and Lieberman 1990], an important property to guarantee the optimality of our algorithm.

3.2 Crosstalk Sensitivity

As indicated in Figure 4, the influences of x_i and x_j on c_{ij} are in the same direction: the larger x_i and x_j, the larger c_{ij}. Consequently, the *crosstalk sensitivity* ς_{ij} of c_{ij} is defined as the *superposition* of the first derivatives of c_{ij} with respect to x_i and x_j.

$$\begin{aligned}
\varsigma_{ij} &\equiv \left|\frac{\partial c_{ij}}{\partial x_i} + \frac{\partial c_{ij}}{\partial x_j}\right| \\
&= \frac{\partial c_{ij}}{\partial x_i} + \frac{\partial c_{ij}}{\partial x_j} \\
&= \left(\frac{\hat{f}_{ij}l_{ij}}{d_{ij}^2}\right)\left(1 - \frac{x_i + x_j}{2d_{ij}}\right)^{-2} \quad (>0).
\end{aligned} \tag{4}$$

Equation (4) reveals that the crosstalk sensitivity ς_{ij} is also a positive quantity; moreover, its lower bound $\tilde{\varsigma}_{ij}$ is $\hat{f}_{ij}l_{ij}/d_{ij}^2$. The crosstalk sensitivity lower bound $\tilde{\varsigma}_{ij}$ is *quadratically* proportional to the inverse of wire spacing. Hence, process variation affects crosstalk in a quadratic fashion. Crosstalk sensitivity should be an increasingly important design metric in the deep submicron regime.

4. THE GATE AND WIRE SIZING PROBLEM

A generic optimization problem in the gate and wire sizing stage is described as follows.

$$
\begin{aligned}
\mathcal{M} : \textit{Minimize} \quad & D_{\max} \ / * \textit{max delay} * / \\
\textit{Subject to} \quad & D \leq D_{\max}, \ / * \textit{timing relationship} * / \\
& X \leq X^B, \ / * \textit{crosstalk constraints} * / \\
& S \leq S^B, \ / * \textit{crosstalk sensitivity constraints} * / \\
& A \leq A^B, \ / * \textit{area constraints} * / \\
& \textit{size constraints}.
\end{aligned}
$$

Problem $\mathcal{M}$ tries to minimize the critical delay $D_{\max}$ under timing, crosstalk, crosstalk sensitivity, area, as well as size constraints. The following will detail the objective and these constraints.

4.1 The Objective and Constraints

In order not to traverse total paths (that may grow exponentially in the graph size), each node i, $1 \leq i \leq m$, is associated with its arrival time a_i. The objective function is to minimize the critical delay, and is equivalently to minimize the arrival time a_m of sink. The timing relationship between components are subject to the timing constraints. Thus,

$$
\begin{aligned}
a_j &\leq a_m, \ j \in input(m), \\
a_j + D_i &\leq a_i, \ s+1 \leq i \leq n+s, j \in input(i), \\
D_i &= a_i, \ 1 \leq i \leq s.
\end{aligned}
$$

The crosstalk for each pair of adjacent wires i and j is bounded by X_{ij}^B. Hence, we have, by Eq. (1),

$$
c_{ij} \leq X_{ij}^B.
$$

$$
\Rightarrow \tilde{c}_{ij} < \tilde{c}_{ij} \left(1 - \frac{x_i + x_j}{2d_{ij}} \right)^{-1} \leq X_{ij}^B.
$$

$$
\Rightarrow \frac{x_i + x_j}{2d_{ij}} \leq 1 - \frac{\tilde{c}_{ij}}{X_{ij}^B}. \tag{5}
$$

As can be seen, the crosstalk bound X_{ij}^B must be larger than or equal to the crosstalk lower bound $\tilde{c}_{ij}$. If technology is scaled down, $\tilde{c}_{ij}$ will increase in a linear fashion.

On the other hand, for each pair of adjacent wires i and j, the crosstalk sensitivity is constrained by S_{ij}^B. By Eq. (4),

$$\varsigma_{ij} \leq S_{ij}^B.$$

$$\Rightarrow \tilde{\varsigma}_{ij} < \tilde{\varsigma}_{ij} \left(1 - \frac{x_i + x_j}{2d_{ij}}\right)^{-2} \leq S_{ij}^B.$$

$$\Rightarrow \frac{x_i + x_j}{2d_{ij}} \leq 1 - \sqrt{\frac{\tilde{\varsigma}_{ij}}{S_{ij}^B}}. \tag{6}$$

The crosstalk sensitivity bound S_{ij}^B needs to be greater than or equal to the crosstalk sensitivity lower bound $\tilde{\varsigma}_{ij}$. As technology advancing, $\tilde{\varsigma}_{ij}$ thus increases in a quadratic fashion. In other words, the crosstalk sensitivity issue may become more significant than the crosstalk one in the deep submicron era.

By Inequalities (5) and (6), we have

$$\frac{x_i + x_j}{2d_{ij}} \leq \min\left(1 - \frac{\tilde{c}_{ij}}{X_{ij}^B}, 1 - \sqrt{\frac{\tilde{\varsigma}_{ij}}{S_{ij}^B}}\right).$$

$$\Rightarrow x_i + x_j \leq 2d_{ij} \min\left(1 - \frac{\tilde{c}_{ij}}{X_{ij}^B}, 1 - \sqrt{\frac{\tilde{\varsigma}_{ij}}{S_{ij}^B}}\right).$$

$$\Rightarrow x_i + x_j \leq \chi_{ij}^B. \tag{7}$$

Hence, in the gate and wire sizing stage, designers should pursue not only crosstalk minimization but crosstalk insensitivity also. Note that, though not presented here, if the variation on each gate/wire size caused during fabrications is provided, the above crosstalk and crosstalk sensitivity constraints can easily be extended to the case with a crosstalk bound to the sum of crosstalk and crosstalk variation. The crosstalk variation can be calculated by

$$v_{ij} \equiv \frac{\partial c_{ij}}{\partial x_i}\triangle x_i + \frac{\partial c_{ij}}{\partial x_j}\triangle x_j,$$

where $\triangle x_i$ and $\triangle x_j$ are the respective size variations on wires i and j. By some calculation, $c_{ij} + v_{ij} \leq X_{ij}^B$ can be simplified as the same form as Inequality (7). Therefore, the properties derived in the succeeding sections still hold for the extended formulation.

Although many objectives have to be attained, area is still an important design concern. The area occupied by each gate/wire i is given by $\square_i x_i$, where $\square_i$ is the quantity of unit size. Since the sizes of input drivers are fixed, we focus on the remaining area occupied by gates and wires, which is restricted to A^B.

$$\sum_{i=s+1}^{n+s} \square_i x_i \leq A^B.$$

Given a technology file and circuit, the lower bounds of gate/wire sizes are always related to the feature size, and the upper bounds are limited by the

Table I. Technology Scaling Down Z Times Benefits
Gates but Harms Wires

Parameter	Scaling Factor
Dimensions	$1/Z$
Area per device	$1/Z^2$
Chip size	Z_c
Intrinsic gate delay	$1/Z$
Local interconnect RC delay	1
Global interconnect RC delay	$Z^2 Z_c^2$
Crosstalk lower bound	Z
Crosstalk sensitivity lower bound	Z^2

performance concern. Thus, we have the following constraints for all gates and wires.

$$L_i \leq x_i \leq U_i, s + 1 \leq i \leq n + s.$$

We summarize how the aforementioned design characteristics change as technology scales down Z times. Table I reveals the profitable effects of scaling—the speed of gates increases in a linear fashion, while the area declines in a quadratic fashion. In contrast, the table also indicates the harmful impacts of scaling—the delay of wires does not decline; moreover, the crosstalk grows in a linear fashion, and the crosstalk sensitivity increases in a quadratic fashion. As shown in Table I, crosstalk sensitivity should be a new comer after crosstalk which may play an even important role in future technology.

4.2 Problem Formulation

We substitute the objective and constraints derived in the preceding subsection for those in Problem $\mathcal{M}$ as follows.

$$
\begin{aligned}
\mathcal{P} : Minimize \quad & a_m \\
Subject\ to \quad & a_j \leq a_m & j \in input(m) \\
& a_j + D_i \leq a_i & s + 1 \leq i \leq n + s, \\
& & j \in input(i) \\
& D_i = a_i & 1 \leq i \leq s \\
& x_i + x_j \leq \chi_{ij}^B & i, j \in W \\
& \sum_{i=s+1}^{n+s} \Box_i x_i \leq A^B & \\
& L_i \leq x_i \leq U_i & s + 1 \leq i \leq n + s.
\end{aligned}
$$

5. LAGRANGIAN RELAXATION

As given in the problem $\mathcal{P}$, the objective and constraints are in posynomial (positive polynomial) forms [Hillier and Lieberman 1990]. This property guarantees $\mathcal{P}$ can optimally be solved by the Lagrangian relaxation method. We relax the constraints into the objective function by introducing one Lagrange multiplier to each constraint. Let $\mathbf{x} = (x_{s+1}, \ldots, x_{n+s})$ and $\mathbf{a} = (a_1, \ldots, a_m)$. As a result, the

Lagrangian function is given by

$$L_{\lambda,\gamma,\eta}(\mathbf{x}, \mathbf{a}) = a_m + \sum_{j \in\ input(m)} \lambda_{jm}(a_j - a_m)$$

$$+ \sum_{i=s+1}^{n+s} \sum_{j \in\ input(i)} \lambda_{ji}(a_j + D_i - a_i)$$

$$+ \sum_{i=1}^{s} \lambda_{0i}(D_i - a_i)$$

$$+ \sum_{i \in W} \sum_{j \in I(i)} \gamma_{ij}\left(x_i + x_j - \chi_{ij}^B\right)$$

$$+ \eta \left(\sum_{i=s+1}^{n+s} \square_i x_i - A^B \right).$$

By Kuhn–Tucker conditions [Winston 1994], $\frac{\partial L_{\lambda,\gamma,\eta}}{\partial a_i}(\mathbf{x}, \mathbf{a}) = 0$, we have Theorem 5.1.

THEOREM 5.1. *The optimality conditions on Lagrange multipliers are given by*
(1) $\sum_{j\ \in\ input(m)} \lambda_{jm} = 1$.
(2) $\sum_{k\ \in\ output(i)} \lambda_{ik} = \sum_{j\ \in\ input(i)} \lambda_{ji}$, $for\ 1 \leq i \leq n + s$.

PROOF. Rearranging terms for $L_{\lambda,\gamma,\eta}(\mathbf{x}, \mathbf{a})$, we have

$$L_{\lambda,\gamma,\eta}(\mathbf{x}, \mathbf{a}) = \left(1 - \sum_{j \in input(m)} \lambda_{jm} \right) a_m$$

$$+ \sum_{i=1}^{n+s} \left(\sum_{k \in ouput(i)} \lambda_{ik} - \sum_{j \in input(i)} \lambda_{ji} \right) a_i$$

$$+ \sum_{i=1}^{n+s} \left(\sum_{j \in input(i)} \lambda_{ji} \right) D_i$$

$$+ \sum_{i \in W} \sum_{j \in I(i)} \gamma_{ij}\left(x_i + x_j - \chi_{ij}^B\right)$$

$$+ \eta \left(\sum_{i=s+1}^{n+s} \square_i x_i - A^B \right). \tag{8}$$

We apply Kuhn–Tucker conditions, $\frac{\partial L_{\lambda,\gamma,\eta}}{\partial a_i}(\mathbf{x}, \mathbf{a}) = 0$, the theorem follows. □

If we define $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ and $\mu_i = \sum_{j \in input(i)} \lambda_{ji}$, under the optimality conditions, Eq. (8) can be rewritten as

$$L_{\mu,\gamma,\eta}(\mathbf{x}) = \sum_{i=1}^{n+s} \mu_i D_i + \sum_{i \in W} \sum_{j \in I(i)} \gamma_{ij}\left(x_i + x_j - \chi_{ij}^B\right)$$

$$+ \eta \left(\sum_{i=s+1}^{n+s} \square_i x_i - A^B \right), \tag{9}$$

where $D_i = r_i C_i$ and C_i is i's downstream capacitance.

For any vector $\boldsymbol{\lambda}$ satisfying the optimality conditions in Theorem 5.1, the corresponding Lagrangian relaxation subproblem $\mathcal{LRS}$ of the problem $\mathcal{P}$ is formulated as follows:

$$\mathcal{LRS} : Minimize \quad L_{\mu,\gamma,\eta}(\mathbf{x})$$
$$Subject \ to \quad L_i \leq x_i \leq U_i \ \ s+1 \leq i \leq n+s.$$

By $\frac{\partial L_{\mu,\gamma,\eta}}{\partial x_i}(\mathbf{x}) = 0$, we have the optimal sizing for each gate or wire as follows:

THEOREM 5.2. *Let* $\tilde{\mathbf{x}} = (\tilde{x}_{s+1}, \ldots, \tilde{x}_{n+s})$ *be a solution. For* $s + 1 \leq i \leq n + s$, *the optimal sizing*

$$x_i^* \ = \ \min(U_i, \max(L_i, opt_i)), \ where$$
$$opt_i \ = \ \sqrt{\frac{O_1}{O_2 + O_3 + O_4}},$$
$$O_1 \ = \ \mu_i \hat{r}_i C_i',$$
$$O_2 \ = \ \eta \square_i,$$
$$O_3 \ = \ \sum_{k \, \in \, upstream(i)} \mu_k r_k \left(\hat{c}_i + 2 \sum_{j \in N(i)} \hat{c}_{ij} \right),$$
$$O_4 \ = \ \sum_{j \in N(i)} \mu_j r_j \hat{c}_{ij} + \gamma_{ij},$$
$$C_i' \ = \ \begin{cases} C_i - \frac{\hat{c}_i + 2 \sum_{j \in N(i)} \hat{c}_{ij}}{2} x_i & if \ i \in W, \\ C_i & otherwise. \end{cases}$$

PROOF. C_i' is the portion of downstream capacitance C_i which is independent of the size x_i. In terms of C_i', Eq. (9) can be rewritten in the following:

$$L_{\mu,\gamma,\eta}(\mathbf{x}) \ = \ \sum_{i=1}^{n+s} \mu_i r_i C_i'$$
$$+ \sum_{i \in W} \mu_i \hat{r}_i \frac{\hat{c}_i + 2 \sum_{j \in N(i)} \hat{c}_{ij}}{2}$$
$$+ \sum_{i \in W} \sum_{j \in I(i)} \gamma_{ij} \left(x_i + x_j - \chi_{ij}^B \right)$$
$$+ \eta \left(\sum_{i=s+1}^{n+s} \square_i x_i - A^B \right).$$

We extract the terms dependent on x_i.

$$L_{\mu,\gamma,\eta}(\mathbf{x}) \ = \ \frac{\mu_i \hat{r}_i C_i'}{x_i} + \sum_{k \, \in \, upstream(i)} \mu_k \hat{r}_k \left(\hat{c}_i + 2 \sum_{j \in N(i)} \hat{c}_{ij} \right) x_i + \sum_{j \in N(i)} \mu_j r_j \hat{c}_{ij} x_i$$
$$+ \sum_{j \in N(i)} \gamma_{ij} x_i + \eta \square_i x_i + terms \ independent \ of \ x_i.$$

Subroutine: LRS (Lagrangian Relaxation Subroutine)
Input: the circuit graph H and Lagrange multipliers $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, η
Output: $\mathbf{x} = (x_{s+1}, ..., x_{n+s})$ which minimizes $L_{\mu,\gamma,\eta}(\mathbf{x})$
S1. $x_i \leftarrow L_i, \forall\, s+1 \le i \le n+s$.
S2. Compute $C'_i, \forall\, s+1 \le i \le n+s$
　　　by traversing H in the reverse topological order.
S3. **for** $i = s+1$ **to** $n+s$ **do**
　　　　　$x_i \leftarrow \min(U_i, \max(L_i, opt_i))$.
S4. Repeat **S2-S3** until no improvement.

Algorithm: OS (Optimal Sizing)
Input: the circuit graph H
Output: $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$, η which maximize $\min L_{\mu,\gamma,\eta}(\mathbf{x})$
A1. $k \leftarrow 1$;
　　　$\boldsymbol{\lambda} \leftarrow$ an arbitrary vector in the optimality conditions;
　　　$\boldsymbol{\gamma} \leftarrow$ an arbitrary vector;
　　　$\eta \leftarrow$ an arbitrary positive number;
A2. $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)$, where $\mu_i = \sum_{j \in input(i)} \lambda_{ji}$.
A3. Call **LRS** and compute $a_1, ... a_m$.
A4. Adjust multipliers λ_{ji}'s, γ_{ij}'s, η
　　　by the sub-gradient method.
A5. Project $\boldsymbol{\lambda}$ onto the nearest point in the optimality condition.
A6. $k \leftarrow k + 1$.
A7. Repeat **A2-A6** until $(a_m - L_{\lambda,\gamma,\eta}(\mathbf{x})) \le$ error bound.

Fig. 5.   The optimal sizing algorithm.

The minimum $L_{\mu,\gamma,\eta}$ occurs when $\frac{\partial L_{\mu,\gamma,\eta}}{\partial x_i}(\mathbf{x}) = 0$; thus the theorem follows.   □

It can be shown that there exists a vector of Lagrange multipliers such that the optimal solution of $\mathcal{LRS}$ is also the optimal solution of the original problem $\mathcal{P}$. The problem to find such a vector is the Lagrangian dual problem:

$$\mathcal{LDP} : \textit{Maximize}\quad Q(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \eta)$$
$$\textit{Subject to}\quad \boldsymbol{\lambda}\ \textit{in the optimality conditions},$$

where

$$Q(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \eta) = \min L_{\lambda,\gamma,\eta}(\mathbf{x}).$$

By Theorems 5.1 and 5.2, we propose the **OS** algorithm shown in Figure 5 to solve Problem $\mathcal{LDP}$ optimally. At the beginning, **A1** sets γ and η to arbitrary positive numbers and assigns an arbitrary positive vector in the optimality conditions to $\boldsymbol{\lambda}$. In **A2**, $\boldsymbol{\mu}$ is then calculated with respect to $\boldsymbol{\lambda}$. **A3** solves the Lagrangian relaxation subproblem $\mathcal{LRS}$. Lagrange multipliers are then adjusted by the sub-gradient method in **A4**. **A5** projects the new multipliers onto the nearest point in the optimality conditions; our projection scheme is relaxing Lagrange multipliers to the critical paths. Hence, the algorithm can focus on the critical paths in the next iteration. Moreover, the relaxation method remedies a group of constraints in each iteration thus reducing the number of iterations. It is shown in the experimental results that this projection strategy leads to very fast convergence. **A6** updates the iteration counter. We repeat the above process until the solution converges within the error bound (see **A7**).

6. EXTENSION TO OTHER OBJECTIVES

Our formulation can also be extended to other objectives such as energy in high-performance circuitry. We demonstrate how the energy constraint can be incorporated into our formulation in this section. Extensions to other objectives can be considered similarly. For a given technology file and circuit, *energy* is the product of power consumption and delay, which is generally a constant. This type of energy can thus reflect the energy consumed by the circuit per low-to-high or high-to-low transition. Let V_{DD} be the supply voltage. If the overall energy in a circuit is subject to E^B, we have

$$\sum_{i=s+1}^{n+s} c_i V_{DD}^2 \leq E^B.$$

This inequality is also in a posynomial form; thus, without loss of optimality, it can be incorporated into the optimization problem solved in the preceding section. Let δ be the Lagrange multiplier for the energy constraint. Accordingly, the quantity of opt_i in Theorem 5.2 can be modified as follows.

$$opt_i = \sqrt{\frac{O_1}{O_2 + O_3 + O_4 + O_5}}, \ where$$

$$O_5 = \delta V_{DD}^2 \left(\hat{c}_i + 4 \sum_{j \in N(i)} \hat{c}_{ij} \right).$$

7. EXPERIMENTAL RESULTS

We implemented our algorithm and tested on the MCNC93 benchmark circuits on a Pentium III 1.0 GHz PC with 256 MB memory. The technology parameters used in our experiments are as follows. The supply voltage is 2.5 V. The resistance and capacitance of a unit-width inverter are 4.73 kΩ and 8.8 fF respectively, and the resistance, capacitance, and fringing capacitance of a unit-width wire are 5.3 Ω, 2.06 fF, and 102.6 fF respectively. The respective lower and upper bounds of a gate are 0.36 μm and 5 μm, while those bounds of a wire are 0.36 μm and 1.8 μm. The initial gate/wire size is set to as the average of the lower bound and the upper bound. The error bound used in the experiments is set as 0.1%.

Table II lists the names (Ckt Name) of the circuits, numbers of gates (#G) and wires (#W) in the circuits, total numbers of components (All), crosstalk (Xtalk), crosstalk sensitivity (Xtalk Sens.), area (Area), delay (Delay), numbers of iterations (ite), runtimes (time (measured by minute:second)), and storage requirements (mem). The improvement (Impr) is calculated by $\frac{Initial-Final}{Initial} \times 100\%$. The experimental results show that our method is effective and efficient. Table II reveals that while crosstalk is on average improved 72.75%, crosstalk sensitivity is on average improved 89.37%. The respective improvements on area and delay are 68.15% and 91.65%. Further, our method converges very fast and its storage requirement is quite small. For instance, a circuit of 2856 gates and 5272 wires is optimized using 13-minute runtime and 2.8-MB memory. We

Table II. Experimental Results on the MCNC93 Benchmark Circuits. The Initial Gate/Wire Size is the Average of the Lower Bound and Upper Bound

Ckt Name	Ckt Size			Xtalk (fF)		Xtalk Sens. (fF/μm)		Area (kμm^2)		Delay (ns)		ite	Time (m:s)	Mem (MB)
	#G	#W	All	Initial	Final	Initial	Final	Initial	Final	Initial	Final			
c432	217	408	663	370.39	95.44	1622.80	155.97	56.96	17.78	386.45	31.04	2	00:01	1.12
c499	253	491	787	426.76	120.84	1809.51	208.06	67.73	23.51	314.58	24.01	2	00:01	1.22
c880	390	714	1166	615.13	173.57	2599.26	299.20	98.04	33.03	337.47	26.27	2	00:03	1.21
c1355	650	1200	1893	1064.46	283.12	4640.17	469.55	166.02	52.51	359.35	28.38	2	00:10	1.39
c1908	390	826	1251	742.84	195.93	3239.36	327.38	114.70	36.18	376.44	29.80	2	00:04	1.30
c2670	924	1736	2819	1473.02	418.20	6154.96	700.62	239.38	78.51	464.04	38.62	2	00:18	1.59
c3540	1038	2187	3277	1915.42	506.60	8207.23	826.86	304.98	90.38	435.16	35.75	2	00:34	1.76
c5315	1778	3694	5652	3153.84	886.35	13351.34	1471.83	509.12	165.16	386.35	39.43	2	02:00	2.29
c6288	2856	5272	8162	4567.58	1218.15	19453.80	1978.36	725.77	219.64	689.87	64.50	2	13:11	2.80
c7552	2247	4344	6799	3698.97	1034.57	15557.38	1739.88	598.27	189.37	675.42	59.64	2	07:17	2.39
Impr	—			72.75%		89.37%		68.15%		91.65%			—	

note that all solutions rapidly converge to the global optimal after only two iterations by relaxing Lagrange multipliers to the critical paths of the circuits. To the best knowledge of authors, this kind of efficiency has never been reported in related previous works.

8. CONCLUSION

This article has raised a new issue—crosstalk sensitivity, which is an important new design metric in deep submicron technology. We have optimally solved a multi-objective optimization problem by Lagrangian relaxation. The experimental results show that our method is very efficient and effective. Our projection scheme, relaxing Lagrange multipliers to the critical paths, provides a crucial insight into effectively adjusting Lagrange multipliers, which is a key ingredient to Lagrangian relaxation.

For the impact of process variation on crosstalk, crosstalk sensitivity was measured by the first derivative of crosstalk with respect to wire width. For the same pattern of wires, the process variation in the region with dense wires is much larger than that in the region with sparse wires. We could extend the formula of crosstalk sensitivity to consider routing densities in the future.

REFERENCES

ALPERT, C. J., DEVGAN, A., AND QUAY, S. T.   1998.   Buffer insertion for noise and delay optimization. In *Proceedings of ACM/IEEE Design Automation Conference* (San Francisco, CA). ACM, New York, 362–367.

BECER, M., BLAAUW, D., ALGOR, I., PANDA, R., OH, C., ZOLOTOV, V., AND HAJJ, I. N.   2003.   Post-route gate sizing for crosstalk noise reduction. In *Proceedings of ACM/IEEE Design Automation Conference* (Anaheim, CA). ACM, New York, 954–957.

CHEN, C.-P., CHU, C. C. N., AND WONG, D. F.   1999.   Fast and exact simultaneous gate and wire sizing by lagrangian relaxation. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst. 18*, 7 (July), 1014–1025.

CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L., AND STEIN, C.   2001.   *Introduction to Algorithms, 2nd Ed.* McGraw Hill/The MIT Press, Cambridge, MA.

ELMORE, W. C.   1948.   The transient response of damped linear networks with particular regard to wide band amplifiers. *J. Appl. Phy. 19*, 1, 55–63.

GAO, T. AND LIU, C. L.   1993.   Minimum crosstalk channel routing. In *Proceedings of IEEE/ACM International Conference on Computer-Aided Design* (San Jose, CA). IEEE Computer Society Press, Los Alamitos, CA, 692–696.

HASHIMOTO, M., TAKAHASHI, M., AND ONODERA, H.   2002.   Crosstalk noise optimization by post-layout transistor sizing. In *Proceedings of ACM International Symposium on Physical Design* (Del Mar, CA). ACM, New York, 126–130.

HILLIER, F. S. AND LIEBERMAN, G. J.   1990.   *Introduction to Operations Research, 5th Ed.* McGraw Hill, New York.

JIANG, I. H.-R., CHANG, Y.-W., AND JOU, J.-Y.   2000.   Crosstalk-driven interconnect optimization by simultaneous gate and wire sizing. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst. 19*, 9 (Sept.), 999–1010.

KAHNG, A. B. AND PATI, Y. C.   1999.   Subwavelength optical lithography: Challenges and impact on physical design. In *Proceedings of ACM International Symposium on Physical Design* (Monterey, CA). ACM, New York, 112–119.

MALIK, S., SENTOVICH, E., BRAYTON, R. K., AND SANGIOVANNI-VINCENTELLI, A.   1991.   Retiming and resynthesis: Optimization of sequential networks with combinational techniques. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst. 10*, 1 (Jan.), 84–94.

RABAEY, J. M.   1996.   *Digital Integrated Circuits: A Design Perspective*. Prentice-Hall, Englewood, Cliffs, N.J.

SAXENA, P. AND LIU, C. L.   1999.   Crosstalk minimization using wire perturbations. In *Proceedings of ACM/IEEE Design Automation Conference* (New Orleans, LA). ACM, New York, 100–103.

SEMICONDUCTOR INDUSTRY ASSOCIATION   1999.   *National Technology Roadmap for Semiconductors*.

SINHA, D. AND ZHOU, H.   2004.   Gate sizing for crosstalk reduction under timing constraints by lagrangian relaxation. In *Proceedings of IEEE/ACM International Conference on Computer-Aided Design* (San Jose, CA). IEEE Computer Society Press, Los Alamitos, CA, 14–19.

SINHA, D., ZHOU, H., AND CHU, C. C. N.   2004.   Optimal gate sizing for coupling-noise reduction. In *Proceedings of ACM International Symposium on Physical Design* (Phoenix, Arizona). ACM, New York, 176–181.

WINSTON, W. L.   1994.   *Operations Research: Applications and Algorithms, 3rd Ed.* Thomson Publishing.

ZHANG, T. AND SAPATNEKAR, S. S.   2004.   Simultaneous shield and buffer insertion for crosstalk noise reduction in global routing. In *Proceedings of IEEE International Conference on Computer Design* (San Jose, CA). IEEE Computer Society Press, Los Alamitos, CA, 93–98.