

**Fast and Exact Simultaneous Gate and Wire Sizing  
by Lagrangian Relaxation**

Chung-Ping Chen, Chris C. N. Chu and D. F. Wong  
ccp@cs.utexas.edu, cnchu@cs.utexas.edu and wong@cs.utexas.edu  
Department of Computer Sciences, University of Texas at Austin, Austin, TX 78712.

October 13, 1997

### **Abstract**

This paper considers simultaneous gate and wire sizing for general VLSI circuits under the Elmore delay model. We present a fast and exact algorithm which can minimize total area subject to maximum delay bound. The algorithm can be easily modified to give exact algorithms for optimizing several other objectives (e.g. minimizing maximum delay or minimizing total area subject to arrival time specifications at all inputs and outputs). No previous algorithm for simultaneous gate and wire sizing can guarantee exact solutions for general circuits. Our algorithm is an iterative one with a guarantee on convergence to global optimal solutions. It is based on Lagrangian relaxation and “one-gate/wire-at-a-time” local optimizations, and is extremely economical and fast. For example, we can optimize a circuit with 13824 gates and wires in about 13 minutes using under 12 MB memory on an IBM RS/6000 workstation.

# 1 Introduction

Since the invention of integrated circuits almost 40 years ago, gate sizing has always been an effective technique to achieve desirable circuit performance. As technology continues to scale down, total number of gates and interconnects within a die grows over millions. In such increasingly dense integrated circuits, a significant portion of the total circuit delay comes from the interconnects. Therefore, developing efficient algorithms which can handle large scale gate and interconnect optimization problems are of great importance.

In the past, gate delay was the dominant factor in determining circuit performance. Thus, gate and transistor sizing have been extensively studied in the literature [5, 11, 14, 19]. As interconnect delay plays an increasingly important role in determining circuit performance, wire sizing has been an active research topic in the past few years [2, 4, 6, 8, 16, 18]. Since gate sizes affect wire-sizing solutions and wire sizes affect gate-sizing solutions, it is beneficial to simultaneously size both gates and wires. Very few results on simultaneous gate and wire sizing have been reported [2, 6, 7, 15, 17]. [7] studied simultaneous driver and wire sizing and [2] considered simultaneous wire and buffer sizing, but both works only apply to circuits that are of tree topology. For simultaneous gate and wire sizing for general circuits, [17] uses a least-square optimization technique, [15] employs a sequential quadratic programming approach, and [6] uses a greedy sizing technique in conjunction with dynamic programming. But none of these algorithms can guarantee to give exact solutions for objectives such as minimizing total area subject to maximum delay bound or minimizing maximum delay.

In this paper, we consider simultaneous gate and wire sizing for general VLSI circuits under the Elmore delay model. We present a fast and exact algorithm which can minimize total area subject to maximum delay bound. The algorithm can be easily modified to give exact algorithms for optimizing several other objectives (e.g. minimizing maximum delay or minimizing total area subject to arrival time specifications at all inputs and outputs). Our algorithm is an iterative one with a guarantee on convergence to global optimal solutions. It is based on Lagrangian relaxation and “one-gate/wire-at-a-time” local optimizations, and is extremely economical and fast. For example, we can optimize a circuit with 13824 gates and wires in about 13 minutes using under 12 MB memory on an IBM RS/6000 workstation.

The rest of this paper is organized as follows. In Section 2, we will introduce some notations and terminology that we will use in this paper. In Section 3, we will present our algorithm for the problem of minimizing total area subject to maximum delay bound. In Section 4, we will show how to modify our algorithm to minimize maximum delay, to handle arrival time specifications at all inputs and outputs, to consider power consumption and to use a more accurate gate model. In Section 5, experimental results to show the runtime and storage requirements of our algorithm is presented.

## 2 Preliminaries

In this section, we will define some notations and terminology that we will use in this paper.

For a general VLSI circuit, we can ignore all latches and optimize its combinational subcircuits.

Therefore, we will focus on combinational circuits below.

Given a combinational circuit with  $s$  input drivers,  $t$  output loads, and  $n$  gates or wire segments, the gate sizes or the segment widths are allowed to be varied in order to optimize some objective. For  $1 \leq i \leq s$ , let  $R_i^D$  be the driver resistance of the  $i$ th input driver. For  $1 \leq i \leq t$ , let  $C_i^L$  be the load capacitance of the  $i$ th output load. See Figure 1 for an illustration of a circuit.

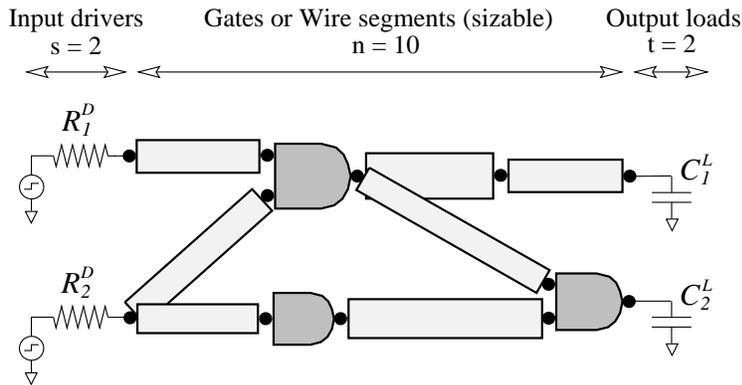


Figure 1: An illustration of a circuit.

A gate, a wire segment, or an input driver is called a *component*. In order to unify the notations that we will introduce later, imagine that two fictitious components are added to the circuit. The first one is called an output component which consists of all the  $t$  output loads. The second one is called an input component which connects to all the  $s$  input drivers. Let a *node* be a connection point between two components or the output point of the output component. Note that the output of each component should connect to a distinct node. So it is easy to see that there are  $n + s + 2$  components and  $n + s + 2$  nodes.

Let  $m = n + s + 1$ . We label the nodes by indexes  $0, \dots, m$  as follows. The node with index 0 is the output point of the output component. For  $1 \leq i \leq t$ , the node with index  $i$  is the one connecting to the  $i$ th output load. For  $t + 1 \leq i \leq n$ , the node with index  $i$  is a connection point among the gates and wire segments. The indexes are assigned in such a way that if node  $i$  and node  $j$  are connected to an input and the output of some component respectively, then  $i > j$ . For  $n + 1 \leq i \leq n + s$ , the node with index  $i$  is the one connecting to the  $(i - n)$ th input driver. The node with index  $m$  is the output point of the input component. It is not difficult to see that if we view the circuit as a directed acyclic graph, the node index assignment is a reverse topological ordering of the graph. We also label the components by indexes  $0, \dots, m$  such that the output of the component with index  $i$  is connected to node  $i$ . See Figure 2 for an illustration of the circuit in Figure 1 with fictitious components, node indexes and component indexes.

For  $0 \leq i \leq m - 1$ , let  $input(i)$  be the set of indexes of components directly connected to the input(s) of component  $i$ . For  $1 \leq i \leq m$ , let  $output(i)$  be the set of indexes of components directly connected to the output of component  $i$ . For example, for the circuit in Figure 2,  $input(0) = \{1, 2\}$ ,  $input(6) = \{8, 9\}$ ,  $output(6) = \{3, 4\}$ ,  $input(8) = \{11\}$ , and  $output(8) = \{6\}$ . Note that  $j \in input(i)$  if and only if  $i \in output(j)$ .

Let  $G$  be the set of component indexes of gates in the circuit. Let  $W$  be the set of component indexes

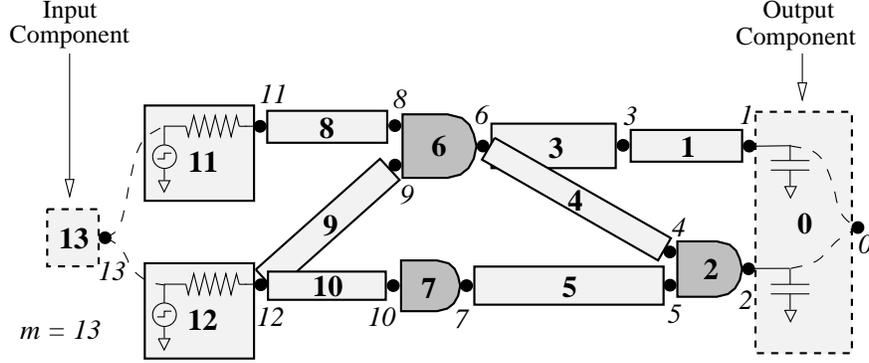


Figure 2: An illustration of a circuit in Figure 1 with node indexes and component indexes. The factitious input component and output component are also shown.

of wire segments in the circuit. For the circuit in Figure 2,  $G = \{2, 6, 7\}$  and  $W = \{1, 3, 4, 5, 8, 9, 10\}$ .

If  $i \in G$ , then let  $x_i$  be the gate size,  $r_i$  be the output resistance of the gate and  $c_i$  be the input capacitance of a pin of the gate. (To simplify the notations, we assume that the input capacitances of all input pins of a gate are the same. It is clear that all our results will still hold if this is not the case.) Let  $\hat{r}_i$  and  $\hat{c}_i$  be respectively the unit size output resistance and the input capacitance per unit size of gate  $i$ . Then  $r_i = \hat{r}_i/x_i$  and  $c_i = \hat{c}_i x_i$ . If  $i \in W$ , then let  $x_i$  be the segment width,  $r_i$  be the segment resistance and  $c_i$  be the segment capacitance. Let  $\hat{r}_i$ ,  $\hat{c}_i$  and  $f_i$  be respectively the unit width wire resistance, the wire area capacitance per unit width and the wire fringing capacitance of segment  $i$ . Then  $r_i = \hat{r}_i/x_i$  and  $c_i = \hat{c}_i x_i + f_i$ . For  $i \in G \cup W$ , let  $L_i$  and  $U_i$  be respectively the lower bound and upper bound of the value of  $x_i$ , i.e.  $L_i \leq x_i \leq U_i$ .

For the purpose of delay calculation, we model components as RC circuits. A gate is modeled as a switch-level RC circuit as shown in Figure 3. (For simplicity, we ignore the intrinsic delay of gates in the model. It is easy to see that all our results will still hold even if intrinsic delay is considered.) A wire segment is modeled as a  $\pi$ -type RC circuit as shown in Figure 4.

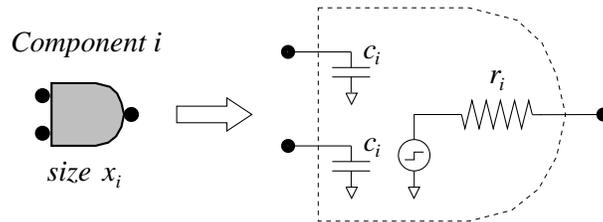


Figure 3: The model of component  $i$ , which is a gate, by a switch-level RC circuit. Note that  $r_i = \hat{r}_i/x_i$  and  $c_i = \hat{c}_i x_i$ , where  $\hat{r}_i$  and  $\hat{c}_i$  are respectively the unit size output resistance and the input capacitance per unit size of gate  $i$ . Although the gate shown here is a 2-input AND gate, the model can be easily generalized for any gate with any number of input pins.

Elmore delay model [10] is used for delay calculation. Basically, the Elmore delay along a signal path is the sum of the delays associated with the resistors in the path, where the delay associated with a resistor is equal to its resistance times its downstream capacitance. For our case, each component (except the 2 factitious components) contains a resistor. We label the resistors by indexes  $1, \dots, n+s$

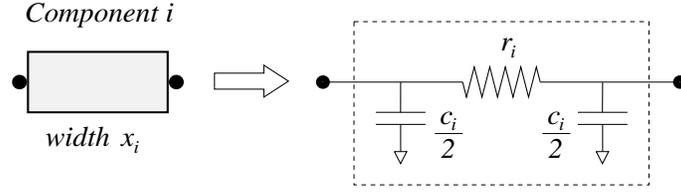


Figure 4: The model of component  $i$ , which is a wire segment, by a  $\pi$ -type RC circuit. Note that  $r_i = \hat{r}_i/x_i$  and  $c_i = \hat{c}_i x_i + f_i$ , where  $\hat{r}_i$ ,  $\hat{c}_i$  and  $f_i$  are respectively the unit width wire resistance, the wire area capacitance per unit width and the wire fringing capacitance of segment  $i$ .

such that resistor  $i$  is the one inside component  $i$ . For convenience, for  $n+1 \leq i \leq n+s$ , let  $r_i = R_{i-n}^D$  (i.e. the driver resistance of the  $(i-n)$ th input driver). So for  $1 \leq i \leq n+s$ , the resistance of resistor  $i$  is  $r_i$ . For  $1 \leq i \leq n+s$ , let  $C_i$  be the downstream capacitance of resistor  $i$ . Figure 5 shows the circuit in Figure 2 after replacing the components by the RC models. The resistance of each resistor is marked in the figure. Also, the regions corresponding to the downstream capacitances of resistor 5 and resistor 12 are shaded.

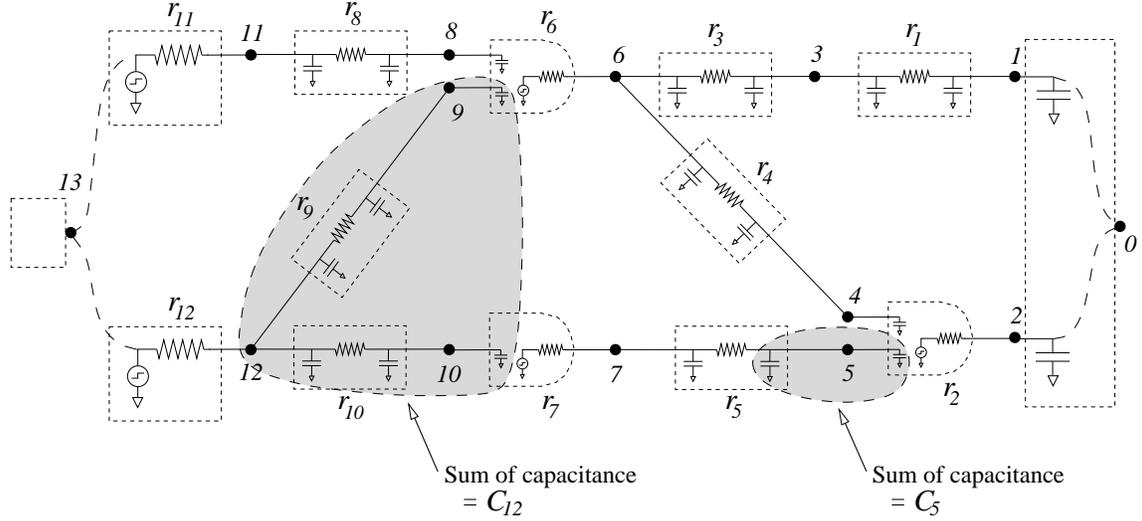


Figure 5: Illustration of the circuit in Figure 2 after replacing the gates and wire segments by the RC models. The resistance of each resistor is marked in the figure. Also, the regions corresponding to the downstream capacitances of resistor 5 and resistor 12 are shaded.

Let  $D_i = r_i C_i$  be the delay associated with resistor  $i$ . We represent a signal path passing through resistors  $i_1, \dots, i_k$  by the set  $p = \{i_1, \dots, i_k\}$ . Let  $P$  be the set of all possible paths from node  $m$  to node 0 (i.e. from an input driver to an output load). Then for any  $p \in P$ , the Elmore delay along path  $p$  is  $\sum_{i \in p} D_i$ .

### 3 Minimizing total area subject to maximum delay bound

In this section, we will solve the problem of minimizing the total component area with respect to component sizes  $x_1, \dots, x_n$  subject to the constraint that the maximum delay from any input driver to any output load is at most some constant  $A_0$  (i.e.  $A_0$  is a bound on the arrival time at node 0).

In Section 3.1, we will first show how to formulate the problem as a constrained optimization problem with a polynomial number of constraints. We call this formulation the primal problem ( $\mathcal{PP}$ ). Then we will show how to solve  $\mathcal{PP}$  by Lagrangian relaxation, which is a general technique for solving constrained optimization problems. We outline the basic idea of Lagrangian relaxation below. More details can be found in [1, 12, 13].

In Lagrangian relaxation, “troublesome” constraints are “relaxed” and incorporated into the objective function after multiplying them by constants called Lagrange multipliers, one multiplier for each constraint. For each fixed vector  $\lambda$  of the Lagrange multipliers introduced, we have a new optimization problem (which should be easier to solve because it is free of troublesome constraints) called the Lagrangian relaxation subproblem associated with  $\lambda$  ( $\mathcal{LRS}/\lambda$ ). It can be shown that there exists a vector  $\lambda$  such that the optimal solution of  $\mathcal{LRS}/\lambda$  is also the optimal solution of the original constrained optimization problem  $\mathcal{PP}$ . The problem of finding such a vector  $\lambda$  is called the Lagrangian dual problem ( $\mathcal{LDP}$ ). So if we can solve both  $\mathcal{LRS}/\lambda$  and  $\mathcal{LDP}$ , then the optimal solution of  $\mathcal{PP}$  will be given by  $\mathcal{LRS}/\lambda$  where  $\lambda$  is the optimal solution of  $\mathcal{LDP}$ .

In Section 3.2, we will show how  $\mathcal{PP}$  is relaxed to obtain the  $\mathcal{LRS}/\lambda$ . We will use the Kuhn-Tucker conditions (see [1] for a reference) to greatly simplify  $\mathcal{LRS}/\lambda$ . We called the simplified version  $\mathcal{LRS}/\mu$ . In Section 3.3, we will show how to solve  $\mathcal{LRS}/\mu$  (i.e.  $\mathcal{LRS}/\lambda$ ) for any fixed vector  $\mu$ . In Section 3.4, we will show how to solve the  $\mathcal{LDP}$  by the classical method of subgradient optimization.

#### 3.1 Problem formulation

For each  $i$ , the area of component  $i$  is proportional to its size  $x_i$ . Therefore, the total component area can be written as  $\sum_{i=1}^n \alpha_i x_i$  for some constants  $\alpha_1, \dots, \alpha_n$ . Then the problem of minimizing total area subject to maximum delay bound can be formulated directly as the mathematical program below:

$$\begin{aligned} & \text{Minimize} && \sum_{i=1}^n \alpha_i x_i \\ & \text{Subject to} && \sum_{i \in p} D_i \leq A_0 \quad \forall p \in P \\ & && L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

However, the number of possible signal paths from node  $m$  to node 0 (and hence the number of constraints in the mathematical program above) can be exponential in  $n$ . So this direct formulation is impractical.

This difficulty can be handled by the classical technique of partitioning the constraints on path delay into constraints on delay across components. We associate a variable  $a_i$  to each node  $i$ .  $a_i$  represents the arrival time at node  $i$  (i.e. the maximum delay from node  $m$  to node  $i$ ). Then it is not difficult to see that the mathematical program below, which we called the primal problem

( $\mathcal{PP}$ ), is equivalent to the mathematical program above:

$$\begin{aligned}
\mathcal{PP} : \quad & \text{Minimize} \quad \sum_{i=1}^n \alpha_i x_i \\
& \text{Subject to} \quad a_j \leq A_0 \quad j \in \text{input}(0) \quad /* \text{ outputs } */ \\
& \quad \quad \quad a_j + D_i \leq a_i \quad i = 1, \dots, n \text{ and } \forall j \in \text{input}(i) \\
& \quad \quad \quad D_i \leq a_i \quad i = n + 1, \dots, n + s \quad /* \text{ inputs } */ \\
& \quad \quad \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n
\end{aligned}$$

Note that the number of constraints in  $\mathcal{PP}$  is polynomial in  $n$  and  $s$ . Also note that for the problem  $\mathcal{PP}$ , the objective function is a posynomial [9] and the constraints can be rewritten in the form of posynomials. It is well known that under a variable transformation, the problem is convex. So problem  $\mathcal{PP}$  has a unique global minimum and no other local minimum. We will consider the formulation  $\mathcal{PP}$  in the following.

### 3.2 Lagrangian Relaxation

Following the Lagrangian relaxation procedure, we introduce a non-negative value called the Lagrange multiplier for each constraint on arrival time. For all  $j \in \text{input}(0)$  (i.e.  $j = 1, \dots, t$ ), we introduce  $\lambda_{j0}$  for the constraint  $a_j \leq A_0$ . For  $i = 1, \dots, n$  and for all  $j \in \text{input}(i)$ , we introduce  $\lambda_{ji}$  for the constraint  $a_j + D_i \leq a_i$ . For  $i = n + 1, \dots, n + s$ , we introduce  $\lambda_{mi}$  for the constraint  $D_i \leq a_i$ . Let  $\boldsymbol{\lambda}$  be a vector of all the Lagrange multipliers introduced. Let  $\boldsymbol{x} = (x_1, \dots, x_n)$  and  $\boldsymbol{a} = (a_1, \dots, a_{n+s})$ . Let

$$\begin{aligned}
L_\lambda(\boldsymbol{x}, \boldsymbol{a}) &= \sum_{i=1}^n \alpha_i x_i \\
&+ \sum_{j \in \text{input}(0)} \lambda_{j0} (a_j - A_0) \\
&+ \sum_{i=1}^n \sum_{j \in \text{input}(i)} \lambda_{ji} (a_j + D_i - a_i) \\
&+ \sum_{i=n+1}^{n+s} \lambda_{mi} (D_i - a_i)
\end{aligned} \tag{1}$$

Then the Lagrangian relaxation subproblem associated with the Lagrange multipliers  $\boldsymbol{\lambda}$  will be:

$$\begin{aligned}
\mathcal{LRS}/\boldsymbol{\lambda} : \quad & \text{Minimize} \quad L_\lambda(\boldsymbol{x}, \boldsymbol{a}) \\
& \text{Subject to} \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n
\end{aligned}$$

Let  $(\boldsymbol{x}^*, \boldsymbol{a}^*)$  be the optimal solution of  $\mathcal{PP}$ . By Kuhn-Tucker conditions, if the optimal solution of  $\mathcal{LRS}/\boldsymbol{\lambda}$  is also the optimal solution of  $\mathcal{PP}$ , then  $\boldsymbol{\lambda}$  must satisfy the conditions  $\frac{\partial L_\lambda}{\partial a_i}(\boldsymbol{x}^*, \boldsymbol{a}^*) = 0$  for  $1 \leq i \leq n + s$ . So we can consider those  $\boldsymbol{\lambda}$  only. We observe that the conditions on the Lagrange multipliers  $\boldsymbol{\lambda}$  can be used to greatly simplify the objective function  $L_\lambda(\boldsymbol{x}, \boldsymbol{a})$ , and hence the problem  $\mathcal{LRS}/\boldsymbol{\lambda}$ . In the following, we will first derive the conditions. Then we will show in Lemma 1 how to simplify the problem.

By rearranging (1), we can write

$$\begin{aligned}
L_\lambda(\mathbf{x}, \mathbf{a}) &= \sum_{i=1}^t \left( \lambda_{i0} - \sum_{j \in \text{input}(i)} \lambda_{ji} \right) a_i \\
&+ \sum_{i=t+1}^n \left( \sum_{k: i \in \text{input}(k)} \lambda_{ik} - \sum_{j \in \text{input}(i)} \lambda_{ji} \right) a_i \\
&+ \sum_{i=n+1}^{n+s} \left( \sum_{k: i \in \text{input}(k)} \lambda_{ik} - \lambda_{mi} \right) a_i \\
&+ \sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \left( \sum_{j \in \text{input}(i)} \lambda_{ji} \right) D_i - \sum_{j \in \text{input}(0)} \lambda_{j0} A_0 + \sum_{i=n+1}^{n+s} \lambda_{mi} D_i \\
&= \sum_{i=1}^{n+s} \left( \sum_{k \in \text{output}(i)} \lambda_{ik} - \sum_{j \in \text{input}(i)} \lambda_{ji} \right) a_i \\
&+ \sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \left( \sum_{j \in \text{input}(i)} \lambda_{ji} \right) D_i - \sum_{j \in \text{input}(0)} \lambda_{j0} A_0 + \sum_{i=n+1}^{n+s} \lambda_{mi} D_i \quad (2)
\end{aligned}$$

By setting  $\partial L_\lambda / \partial a_i = 0$ , we obtain the following conditions on  $\lambda$ .

Optimality Conditions on Lagrange Multipliers:

$$\sum_{k \in \text{output}(i)} \lambda_{ik} = \sum_{j \in \text{input}(i)} \lambda_{ji} \quad \text{for } 1 \leq i \leq n+s$$

Let  $\Omega_\lambda = \{\lambda \geq \mathbf{0} : \lambda \text{ satisfies the optimality conditions on Lagrange multipliers}\}$

**Lemma 1** For any  $\lambda \in \Omega_\lambda$ , solving  $\mathcal{LRS}/\lambda$  is equivalent to solving

$$\begin{aligned}
\mathcal{LRS}/\mu : & \text{Minimize } L_\mu(\mathbf{x}) \\
& \text{Subject to } L_i \leq x_i \leq U_i \quad i = 1, \dots, n
\end{aligned}$$

where  $\mu = (\mu_0, \dots, \mu_{n+s})$ ,  $\mu_i = \sum_{j \in \text{input}(i)} \lambda_{ji}$  for  $0 \leq i \leq n+s$ , and  $L_\mu(\mathbf{x}) = \sum_{i=1}^{n+s} \mu_i D_i + \sum_{i=1}^n \alpha_i x_i$ .

**Proof:** By substituting the optimality conditions on Lagrange multipliers back to (2), we get

$$\begin{aligned}
L_\lambda(\mathbf{x}, \mathbf{a}) &= \sum_{i=1}^n \alpha_i x_i + \sum_{i=1}^n \left( \sum_{j \in \text{input}(i)} \lambda_{ji} \right) D_i - \sum_{j \in \text{input}(0)} \lambda_{j0} A_0 + \sum_{i=n+1}^{n+s} \lambda_{mi} D_i \\
&= \sum_{i=1}^{n+s} \mu_i D_i + \sum_{i=1}^n \alpha_i x_i - \mu_0 A_0 \quad (3)
\end{aligned}$$

Note that  $L_\lambda(\mathbf{x}, \mathbf{a})$  in (3) no longer depends on  $\mathbf{a}$ . Also note that  $\mu_0 A_0$  is a constant. So if let  $L_\mu(\mathbf{x}) = \sum_{i=1}^{n+s} \mu_i D_i + \sum_{i=1}^n \alpha_i x_i$ , then minimizing  $L_\mu$  is the same as minimizing  $L_\lambda$ . After finding the

optimal  $\mathbf{x}$ , the optimal  $\mathbf{a}$  can be found by considering one by one the variables  $a_i$ 's in the order of decreasing  $i$ . For each  $a_i$ , we set it to the smallest possible value that satisfies the constraints of  $\mathcal{PP}$ . Hence the lemma follows.  $\square$

### 3.3 Solving $\mathcal{LRS}/\mu$

In this subsection, for any fixed  $\mu \geq \mathbf{0}$ , we will show how to solve  $\mathcal{LRS}/\mu$  optimally by a greedy algorithm based on iteratively re-sizing the gates and wire segments. Similar techniques have been successfully applied to some other wire or buffer sizing problems before (e.g. [3, 8]).

If we re-size component  $i$  (i.e. changing  $x_i$ ) while keeping the sizes of all the other components fixed, we say that it is a local re-sizing of component  $i$ . An optimal local re-sizing of component  $i$  is a local re-sizing that minimize  $L_\mu(\mathbf{x})$ .

For  $1 \leq i \leq n$ , let  $upstream(i)$  be the set of resistor indexes (excluding  $i$ ) on the path(s) from component  $i$  to the nearest upstream gate(s) or input driver(s). For example, for the circuit in Figure 5,  $upstream(1) = \{3, 6\}$  and  $upstream(6) = \{8, 9, 11, 12\}$ . Let  $R_i = \sum_{j \in upstream(i)} \lambda_j r_j$  (i.e.  $R_i$  is a weighted upstream resistance of component  $i$ ). For  $i \in W$ , let  $C'_i = C_i - \hat{c}_i x_i / 2$ , and for  $i \in G$  or for  $n+1 \leq i \leq n+s$ , let  $C'_i = C_i$ . Note that for  $1 \leq i \leq n+s$ ,  $C'_i$  is independent of  $x_i$ .

**Lemma 2** For  $1 \leq i \leq n$ ,  $L_\mu(\mathbf{x})$  can be written in the following form:

$$L_\mu(\mathbf{x}) = A_i(\mathbf{x})x_i + \frac{B_i(\mathbf{x})}{x_i} + E_i(\mathbf{x})$$

where  $A_i(\mathbf{x}), B_i(\mathbf{x})$  and  $E_i(\mathbf{x})$  are independent of  $x_i$ ,  $A_i(\mathbf{x}) = \hat{c}_i R_i + \alpha_i$  and  $B_i(\mathbf{x}) = \lambda_i \hat{r}_i C'_i$ .

**Proof:**

$$\begin{aligned} L_\mu(\mathbf{x}) &= \sum_{i=1}^{n+s} \mu_i r_i C_i + \sum_{i=1}^n \alpha_i x_i \\ &= \sum_{i \in G} \mu_i r_i C'_i + \sum_{i \in W} \mu_i r_i \left( C'_i + \frac{\hat{c}_i x_i}{2} \right) + \sum_{i=n+1}^{n+s} \mu_i r_i C'_i + \sum_{i=1}^n \alpha_i x_i \\ &= \sum_{i=1}^{n+s} \mu_i r_i C'_i + \sum_{i \in W} \frac{\mu_i \hat{r}_i \hat{c}_i}{2} + \sum_{i=1}^n \alpha_i x_i \end{aligned}$$

For any  $i$  between 1 and  $n$ ,  $\mu_i r_i C'_i = \mu_i \hat{r}_i C'_i / x_i$ . For any  $j \neq i$ , if  $j \notin upstream(i)$ , then  $\mu_j r_j C'_j$  is independent of  $x_i$ . If  $j \in upstream(i)$ , then  $\mu_j r_j C'_j = \mu_j r_j \hat{c}_i x_i +$  terms independent of  $x_i$ . So

$$\begin{aligned} L_\mu(\mathbf{x}) &= \left( \sum_{j \in upstream(i)} \mu_j r_j \hat{c}_i + \alpha_i \right) x_i + \frac{\mu_i \hat{r}_i C'_i}{x_i} + \text{terms independent of } x_i \\ &= (\hat{c}_i R_i + \alpha_i) x_i + \frac{\mu_i \hat{r}_i C'_i}{x_i} + \text{terms independent of } x_i \end{aligned}$$

Hence the lemma follows.  $\square$

**Lemma 3** Let  $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$  be a component-sizing solution. An optimal local re-sizing of component  $i$  is given by changing the size of component  $i$  to

$$x_i^* = \min \left( U_i, \max \left( L_i, \sqrt{\frac{B_i(\tilde{\mathbf{x}})}{A_i(\tilde{\mathbf{x}})}} \right) \right).$$

**Proof:** If we fix the size of component  $j$  to  $\tilde{x}_j$  for all  $j \neq i$ , and we change  $x_i$ , we can view  $L_\mu$  as a function of  $x_i$ , and by Lemma 2, it is given by

$$F(x_i) = A_i(\tilde{\mathbf{x}})x_i + \frac{B_i(\tilde{\mathbf{x}})}{x_i} + E_i(\tilde{\mathbf{x}})$$

Differentiating  $F$  with respect to  $x_i$ , we get

$$\frac{dF}{dx_i} = A_i(\tilde{\mathbf{x}}) - \frac{B_i(\tilde{\mathbf{x}})}{x_i^2}$$

Let  $\theta(\tilde{\mathbf{x}}) = \sqrt{B_i(\tilde{\mathbf{x}})/A_i(\tilde{\mathbf{x}})}$ . Note that

$$\begin{aligned} dF/dx_i &= 0 & \text{if } x_i = \theta(\tilde{\mathbf{x}}) \\ dF/dx_i &< 0 & \text{if } x_i < \theta(\tilde{\mathbf{x}}) \\ dF/dx_i &> 0 & \text{if } x_i > \theta(\tilde{\mathbf{x}}) \end{aligned}$$

Hence  $F(x_i)$  is decreasing when  $x_i < \theta(\tilde{\mathbf{x}})$ ,  $F(x_i)$  is increasing when  $x_i > \theta(\tilde{\mathbf{x}})$ , and  $F(x_i)$  is minimum at  $x_i = \theta(\tilde{\mathbf{x}})$ . If  $x_i$  is constrained to the range  $[L_i, U_i]$ , we consider three cases:

**Case 1:**  $\theta(\tilde{\mathbf{x}}) \in [L_i, U_i]$ .

In this case,  $F(x_i)$  is minimized when  $x_i = \theta(\tilde{\mathbf{x}})$ .

**Case 2:**  $\theta(\tilde{\mathbf{x}}) > U_i$ .

Then  $F(x_i)$  is decreasing in  $[L_i, U_i]$ . So  $F(x_i)$  is minimized when  $x_i = U_i$ .

**Case 3:**  $\theta(\tilde{\mathbf{x}}) < L_i$ .

Then  $F(x_i)$  is increasing in  $[L_i, U_i]$ . So  $F(x_i)$  is minimized when  $x_i = L_i$ .

Hence the lemma follows. □

$\mathcal{LRS}/\boldsymbol{\mu}$  can be solved by a greedy algorithm based on iteratively re-sizing the components. In each iteration, the components are examined one at a time; each time a component is re-sized optimally using Lemma 3 while keeping the sizes of the other components fixed. We call the algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$  and it is described below. Note that in order to use Lemma 3 to re-size component  $i$ , we need to compute  $R_i$  and  $C_i'$  first. Our algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$  computes  $C_i'$ 's and  $R_i$ 's incrementally by traversing the circuit in a reverse topological order (step 2) and in a topological order (step 3) respectively. So it is not difficult to see that each iteration of the algorithm takes only  $O(n)$  time.

Note that  $L_\mu(\mathbf{x})$  is a posynomial [9] in  $\mathbf{x}$ . It is well known that under a variable transformation, a posynomial is equivalent to a convex function. So  $L_\mu(\mathbf{x})$  has a unique global minimum and no other local minimum. We show in the following that algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$  always converges to the optimal solution.

**ALGORITHM SOLVELRS/ $\mu$ :****Output:**  $\mathbf{x} = (x_1, \dots, x_n)$  which minimizes  $L_\mu(\mathbf{x})$ 

1. **for**  $i := 1$  **to**  $n$  **do**  $x_i := L_i$
2. /\* Finding  $C'_i$  for  $1 \leq i \leq n$   
by traversing the circuit in a reverse topological order \*/  
**for**  $i := 1$  **to**  $t$  **do**  

$$C'_i := \begin{cases} C_i^L & \text{if } i \in G \\ C_i^L + f_i/2 & \text{if } i \in W \end{cases}$$
**for**  $i := t + 1$  **to**  $n$  **do**  

$$C'_i := \begin{cases} 0 & \text{if } i \in G \\ f_i/2 & \text{if } i \in W \end{cases}$$
**for all**  $k$  s.t.  $i \in \text{input}(k)$  **do**  

$$C'_i := \begin{cases} C'_i + \hat{c}_k x_k & \text{if } k \in G \\ C'_i + \hat{c}_k x_k + f_k/2 + C'_k & \text{if } k \in W \end{cases}$$
3. /\* Finding  $R_i$  and  $x_i$  for  $1 \leq i \leq n$   
by traversing the circuit in a topological order \*/  
**for**  $i := n$  **downto**  $1$  **do**  
 $R_i := 0$   
**for all**  $j \in \text{input}(i)$  **do**  

$$R_i := \begin{cases} R_i + \mu_j R_{j-n}^D & \text{if } n + 1 \leq j \leq n + s \\ R_i + \mu_j \hat{r}_j / x_j & \text{if } j \in G \\ R_i + \mu_j \hat{r}_j / x_j + R_j & \text{if } j \in W \end{cases}$$

$$x_i := \min \left( U_i, \max \left( L_i, \sqrt{(\mu_i \hat{r}_i C'_i) / (\hat{c}_i R_i + \alpha_i)} \right) \right)$$
4. Repeat step 2 and 3 until no improvement.

**Lemma 4** *If algorithm SOLVE\_LRS/ $\mu$  converges, then the solution is optimal to  $\mathcal{LRS}/\mu$ .*

**Proof:** Suppose the algorithm converges to  $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ . Then for  $1 \leq i \leq n$ , by Lemma 3,  $x_i^* = \min \left( U_i, \max \left( L_i, \sqrt{B_i(\mathbf{x}^*)/A_i(\mathbf{x}^*)} \right) \right)$ . Note that  $L_\mu(\mathbf{x})$  is a posynomial in  $\mathbf{x}$ , and that under the transformation  $x_i = e^{z_i}$  for  $1 \leq i \leq n$ , the function  $H(\mathbf{z}) = L_\mu(e^{z_1}, \dots, e^{z_n})$  is convex over  $\Omega_{\mathbf{z}} = \{\mathbf{z} : L_i \leq e^{z_i} \leq U_i, 1 \leq i \leq n\}$ . Let  $\mathbf{z}^* = (z_1^*, \dots, z_n^*)$  where  $x_i^* = e^{z_i^*}$  for  $1 \leq i \leq n$ . We now consider 3 cases:

**Case 1:**  $x_i^* = \sqrt{\frac{B_i(\mathbf{x}^*)}{A_i(\mathbf{x}^*)}}$ .

In this case, we have  $\frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) = 0$ . Thus

$$\frac{\partial H}{\partial z_i}(\mathbf{z}^*) = \frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) \frac{\partial x_i}{\partial z_i}(\mathbf{z}^*) = \frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) e^{z_i^*} = 0.$$

**Case 2:**  $x_i^* = L_i$ .

In this case,  $L_i \geq \sqrt{\frac{B_i(\mathbf{x}^*)}{A_i(\mathbf{x}^*)}}$ . We have  $\frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) \geq 0$  and  $z_i - z_i^* \geq 0, \forall \mathbf{z} \in \Omega_{\mathbf{z}}$ . Hence

$$\frac{\partial H}{\partial z_i}(\mathbf{z}^*)(z_i - z_i^*) = \frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) e^{z_i^*} (z_i - z_i^*) \geq 0, \forall \mathbf{z} \in \Omega_{\mathbf{z}}.$$

**Case 3:**  $x_i^* = U_i$ .

In this case,  $U_i \leq \sqrt{\frac{B_i(\mathbf{x}^*)}{A_i(\mathbf{x}^*)}}$ . We have  $\frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) \leq 0$  and  $z_i - z_i^* \leq 0, \forall \mathbf{z} \in \Omega_{\mathbf{z}}$ . Hence

$$\frac{\partial H}{\partial z_i}(\mathbf{z}^*)(z_i - z_i^*) = \frac{\partial L_\mu}{\partial x_i}(\mathbf{x}^*) e^{z_i^*} (z_i - z_i^*) \geq 0, \forall \mathbf{z} \in \Omega_{\mathbf{z}}.$$

So  $\frac{\partial H}{\partial z_i}(\mathbf{z}^*)(z_i - z_i^*) \geq 0$  for all  $i$  and for all  $\mathbf{z} \in \Omega_{\mathbf{z}}$ . Thus for any feasible solution  $\mathbf{x}$ ,

$$\begin{aligned} L_\mu(\mathbf{x}) - L_\mu(\mathbf{x}^*) &= H(\mathbf{z}) - H(\mathbf{z}^*) \\ &\geq \nabla H(\mathbf{z}^*)(\mathbf{z} - \mathbf{z}^*) \quad \text{as } H \text{ is convex} \\ &= \sum_{i=1}^n \frac{\partial H}{\partial z_i}(\mathbf{z}^*)(z_i - z_i^*) \\ &\geq 0 \end{aligned}$$

Therefore  $\mathbf{x}^*$  is the global minimum point. □

**Lemma 5** *The algorithm SOLVE\_LRS/ $\mu$  always converges.*

**Proof:** For any two vectors  $\mathbf{x}$  and  $\mathbf{x}'$ , we use  $\mathbf{x} \leq \mathbf{x}'$  to denote that  $x_i \leq x'_i$  for all  $i$ . Let  $\mathbf{x}^*$  be the optimal solution,  $\mathbf{x}$  be a feasible solution, and  $\mathbf{x}'$  be the solution after locally re-sizing a component of  $\mathbf{x}$ . If  $\mathbf{x} \leq \mathbf{x}^*$ , then we can prove that  $\mathbf{x} \leq \mathbf{x}' \leq \mathbf{x}^*$  (this is similar to the dominance property in [6]).

In step 1 of algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$ , we set  $x_i = L_i$  for all  $i$  initially. So we know that for all  $i$ ,  $x_i$  is non-decreasing for each local re-sizing, and is upper bounded by  $x_i^*$ . Hence the algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$  converges.  $\square$

By Lemma 4 and Lemma 5, we have the following theorem.

**Theorem 1** *For any fixed vector  $\boldsymbol{\mu} \geq \mathbf{0}$ , algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$  always converges to the optimal component-sizing solution of the problem  $\mathcal{LRS}/\boldsymbol{\mu}$ .*

Algorithm SOLVE\_LRS/ $\boldsymbol{\mu}$  runs in  $O(rn)$  time using  $O(n)$  storage, where  $n$  is the number of components and  $r$  is the number of iterations. We will observe that the number of iterations  $r$  is constant (i.e. the run time of SOLVE\_LRS/ $\boldsymbol{\mu}$  is linear) in practice.

### 3.4 Solving the $\mathcal{LDP}$

Define the function  $Q(\boldsymbol{\lambda}) =$  the optimal value of the problem  $\mathcal{LRS}/\boldsymbol{\lambda}$ . In this subsection, we will consider the Lagrangian dual problem:

$$\begin{aligned} \mathcal{LDP} : \quad & \text{Maximize} \quad Q(\boldsymbol{\lambda}) \\ & \text{Subject to} \quad \boldsymbol{\lambda} \in \Omega_\lambda \end{aligned}$$

As said in Section 3.1,  $\mathcal{PP}$  can be transformed into a convex problem. So Theorem 6.2.4 of [1] implies that if  $\boldsymbol{\lambda}$  is the optimal solution of  $\mathcal{LDP}$ , then the optimal solution of  $\mathcal{LRS}/\boldsymbol{\lambda}$  will also optimize  $\mathcal{PP}$ .

By Theorem 6.3.1 of [1],  $Q(\boldsymbol{\lambda})$  is a concave function over  $\boldsymbol{\lambda} \geq \mathbf{0}$ . However,  $\mathcal{LRS}/\boldsymbol{\lambda}$  is not differentiable in general. So methods like steepest descent, which depends on the gradient directions, are not applicable. The subgradient optimization method is usually used instead. The subgradient optimization method can be viewed as a generalization of the steepest descent method in which the gradient direction is substituted by a subgradient-based direction (see [1] for a reference).

Basically, starting from an arbitrary point  $\boldsymbol{\lambda}$ , the method iteratively moves from the current point to a new point following the subgradient direction. At step  $k$ , we first solve  $\mathcal{LRS}/\boldsymbol{\lambda}$  (by solving the simpler  $\mathcal{LRS}/\boldsymbol{\mu}$ ). Then for each relaxed constraint, we define the subgradient to be the right hand side minus the left hand side of the constraint, evaluated at the current solution. The subgradient direction is the vector of all the subgradients. We move to a new point by multiplying a step size  $\rho_k$  to the subgradient direction and adding it to  $\boldsymbol{\lambda}$ . After each time we moved, we project  $\boldsymbol{\lambda}$  back to the nearest point in  $\Omega_\lambda$  so that we can solve  $\mathcal{LRS}/\boldsymbol{\mu}$  instead of  $\mathcal{LRS}/\boldsymbol{\lambda}$  for the next iteration. The procedure is repeated until it converges.

It is well known (see Theorem 8.9.2 of [1] for example) that if the step size sequence  $\{\rho_k\}$  satisfies the conditions  $\lim_{k \rightarrow \infty} \rho_k = 0$  and  $\sum_{k=1}^{\infty} \rho_k = \infty$ , then the subgradient optimization method will always converge to the optimal solution

The description is summarized in the algorithm SOLVE\_LDP below.

**Theorem 2** *The algorithm SOLVE\_LDP always converges to the optimal solution of  $\mathcal{LDP}$ .*

**ALGORITHM SOLVE\_LDP:****Output:**  $\lambda$  which maximizes  $\mathcal{LRS}/\lambda$ 

1.  $k := 1$  /\* step counter \*/  
 $\lambda :=$  arbitrary initial vector in  $\Omega_\lambda$
2.  $\mu = (\mu_0, \dots, \mu_{n+s})$  where  $\mu_i = \sum_{j \in \text{input}(i)} \lambda_{ji}$   
Solve  $\mathcal{LRS}/\lambda$  by calling SOLVE\_LRS/ $\mu$  to solve  $\mathcal{LRS}/\mu$  and calculating  $a_1, \dots, a_{n+s}$  as described in the proof of Lemma 1.
3. /\* Move to a new  $\lambda$  by adjusting the Lagrange multipliers  $\lambda_{ji}$  \*/  
**for**  $i := 0$  **to**  $n + s$  **do**  
    **for all**  $j \in \text{input}(i)$  **do**  
        
$$\lambda_{ji} := \begin{cases} \lambda_{ji} + \rho_k(a_j - A_0) & \text{if } i = 0 \\ \lambda_{ji} + \rho_k(a_j + D_i - a_i) & \text{if } 1 \leq i \leq n \\ \lambda_{ji} + \rho_k(D_i - a_i) & \text{if } n + 1 \leq i \leq n + s \end{cases}$$
4. Project  $\lambda$  onto the nearest point in  $\Omega_\lambda$ .
5.  $k := k + 1$
6. Repeat step 2-5 until  $(\sum_{i=1}^n \alpha_i x_i - Q(\lambda)) \leq$  error bound.

We conclude Section 3 by giving the algorithm SGWS-LR (Simultaneous Gate and Wire Sizing by Lagrangian Relaxation) below.

**ALGORITHM SGWS-LR:****Output:** the optimal gate and wire sizing solution  $x$ 

1. Call SOLVE\_LDP to find the optimal  $\lambda$
2.  $\mu = (\mu_0, \dots, \mu_{n+s})$  where  $\mu_i = \sum_{j \in \text{input}(i)} \lambda_{ji}$
3. Call SOLVE\_LRS/ $\mu$  to find the optimal  $x$

**Theorem 3** *For simultaneous gate and wire sizing, the problem of minimizing total area subject to maximum delay bound can be solved optimally by SGWS-LR.*

## 4 Extensions

In Section 3, the objective of our problem is the total component area and the constraint is on the maximum delay from any input to any output (i.e. the arrival time at node 0). In this section, we will extend our approach to handle problems with other objectives and with other constraints. In Section 4.1, we will treat the maximum delay as the objective and show how to minimize it. We will also see that the problem of minimizing maximum delay subject to total area bound is easy to handle. In Section 4.2, instead of assuming that all the input signals arrive at time 0 and all the output signals have a single bound on the arrival time, we allow different arrival time specifications on the input and output signals. In Section 4.3, we will see that power can be handled similarly as area. In Section 4.4, we will show that a more accurate gate model can be used.

For all the extensions, we will see only slight modifications to our algorithm presented in Section 3 are needed. Moreover, convergence to global optimal solutions is still guaranteed. Actually, it is not difficult to see that any combination of the problem in Section 3 or the extensions can be handled similarly. For example, we can optimally solve the problem of minimizing power subject to bounds on area and on maximum delay from any input to any output.

#### 4.1 Minimizing Maximum Delay

Instead of having a constant bound  $A_0$  for the arrival time at node 0, we introduce one more variable  $a_0$  to represent the arrival time at node 0, and we want to minimize  $a_0$ . As in Section 3.1, by partitioning the constraints on path delay into constraints on delay across components, the problem of minimizing maximum delay by simultaneous gate and wire sizing can be formulated as the mathematical program below:

$$\begin{aligned} \mathcal{PP} : \quad & \text{Minimize} \quad a_0 \\ & \text{Subject to} \quad a_j \leq a_0 \quad j \in \text{input}(0) \quad /* \text{ outputs } */ \\ & \quad \quad \quad a_j + D_i \leq a_i \quad i = 1, \dots, n \text{ and } \forall j \in \text{input}(i) \\ & \quad \quad \quad D_i \leq a_i \quad i = n + 1, \dots, n + s \quad /* \text{ inputs } */ \\ & \quad \quad \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

If all the constraints on arrival time are relaxed, then the Lagrangian relaxation subproblem associated with the Lagrange multipliers  $\lambda$  will be:

$$\begin{aligned} \mathcal{LRS}/\lambda : \quad & \text{Minimize} \quad L_\lambda(\mathbf{x}, \mathbf{a}) \\ & \text{Subject to} \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

$$\text{where } L_\lambda(\mathbf{x}, \mathbf{a}) = a_0 + \sum_{j \in \text{input}(0)} \lambda_{j0}(a_j - a_0) + \sum_{i=1}^n \sum_{j \in \text{input}(i)} \lambda_{ji}(a_j + D_i - a_i) + \sum_{i=n+1}^{n+s} \lambda_{mi}(D_i - a_i).$$

As before, by Kuhn-Tucker conditions, we have the following optimality conditions on Lagrange multipliers.

$$\begin{aligned} 1 &= \sum_{j \in \text{input}(0)} \lambda_{j0} \\ \sum_{k \in \text{output}(i)} \lambda_{ik} &= \sum_{j \in \text{input}(i)} \lambda_{ji} \quad \text{for } 1 \leq i \leq n + s \end{aligned}$$

Then for  $\lambda$  satisfying the conditions,  $\mathcal{LRS}/\lambda$  can be simplified to

$$\begin{aligned} \mathcal{LRS}/\mu : \quad & \text{Minimize} \quad L_\mu(\mathbf{x}) \\ & \text{Subject to} \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

$$\text{where } \mu = (\mu_0, \dots, \mu_{n+s}), \mu_i = \sum_{j \in \text{input}(i)} \lambda_{ji} \text{ for } 0 \leq i \leq n + s, \text{ and } L_\mu(\mathbf{x}) = \sum_{i=1}^{n+s} \mu_i D_i.$$

It is easy to see that  $\mathcal{LRS}/\mu$  can be solved optimally by the iterative local re-sizing algorithm in Section 3.3 and the corresponding  $\mathcal{LDP}$  can be solved optimally by the subgradient optimization

method as described in Section 3.4. Therefore the problem of minimizing maximum delay can also be solved optimally by our approach.

In fact, the problem of minimizing maximum delay subject to area bound can also be optimally solved by our Lagrangian relaxation approach. The constraint on area can be relaxed and incorporated into the objective function as well. The function  $L_\lambda(\mathbf{x}, \mathbf{a})$  will be of the same form as the one in Section 3.2.

## 4.2 Arrival Time Specifications on Inputs and Outputs

In Section 3, we assume that all the input signals arrive at time 0 and we want to bound the arrival time at the outputs uniformly by a single constant  $A_0$ . We show in this subsection that different arrival time specifications on the input and output signals can be easily handled. We demonstrate the idea by considering the problem of minimizing total area subject to different arrival time constraints at inputs and outputs.

For  $n + 1 \leq i \leq n + s$ , let  $A_i$  be the arrival time specification of the input signal at the  $(i - n)$ th input driver. For  $1 \leq j \leq t$ , let  $A_j$  be the arrival time requirement on the output signal at the  $j$ th output load. Then the problem can be formulated as follows:

$$\begin{aligned} \mathcal{PP} : \quad & \text{Minimize} \quad \sum_{i=1}^n \alpha_i x_i \\ & \text{Subject to} \quad a_j \leq A_j \quad j \in \text{input}(0) \quad /* \text{ outputs } */ \\ & \quad \quad \quad a_j + D_i \leq a_i \quad i = 1, \dots, n \text{ and } \forall j \in \text{input}(i) \\ & \quad \quad \quad A_i + D_i \leq a_i \quad i = n + 1, \dots, n + s \quad /* \text{ inputs } */ \\ & \quad \quad \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

If all the constraints on arrival time are relaxed, then the Lagrangian relaxation subproblem associated with the Lagrange multipliers  $\boldsymbol{\lambda}$  will be:

$$\begin{aligned} \mathcal{LRS}/\boldsymbol{\lambda} : \quad & \text{Minimize} \quad L_\lambda(\mathbf{x}, \mathbf{a}) \\ & \text{Subject to} \quad L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

$$\begin{aligned} \text{where } L_\lambda(\mathbf{x}, \mathbf{a}) = & \sum_{i=1}^n \alpha_i x_i + \sum_{j \in \text{input}(0)} \lambda_{j0}(a_j - A_j) + \sum_{i=1}^n \sum_{j \in \text{input}(i)} \lambda_{ji}(a_j + D_i - a_i) \\ & + \sum_{i=n+1}^{n+s} \lambda_{mi}(A_i + D_i - a_i). \end{aligned}$$

Again, by Kuhn-Tucker conditions, we have the following optimality conditions on Lagrange multipliers.

$$\sum_{k \in \text{output}(i)} \lambda_{ik} = \sum_{j \in \text{input}(i)} \lambda_{ji} \quad \text{for } 1 \leq i \leq n + s$$

So for  $\boldsymbol{\lambda}$  satisfying the conditions, we can simplify  $L_\lambda(\mathbf{x}, \mathbf{a})$ :

$$L_\lambda(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^{n+s} \mu_i D_i + \sum_{i=1}^n \alpha_i x_i + \sum_{i=n+1}^{n+s} \lambda_{mi} A_i - \sum_{j \in \text{input}(0)} \lambda_{j0} A_j$$

So the Lagrangian relaxation subproblem can be formulated in exactly the same form as the problem  $\mathcal{LRS}/\mu$  in Section 3.2.  $\mathcal{LRS}/\mu$  and  $\mathcal{LDP}$  can be solved as before. Therefore even with different arrival time specifications on inputs and outputs, the problem can still be solved optimally by our approach.

### 4.3 Power Consideration

For each  $i$ , the power consumption of component  $i$  is proportional to its size  $x_i$ . Therefore, the total power consumption can be written as  $\sum_{i=1}^n \beta_i x_i$  for some constants  $\beta_1, \dots, \beta_n$ . It is of the same form as the total component area. So it is easy to see that it can be handle in exactly the same way as component area.

### 4.4 More Accurate Gate Model

For higher precision timing requirements, more accurate gate models are desirable. Although in Section 2, we model a gate as a switch-level RC circuit with a resistance proportional to the gate size, better gate models can be easily integrated into our algorithm. We now show an example of using precharacterized function as the delay model for gates.

The following precharacterized delay function  $D_i()$  and output slope function  $T_i()$  can capture the input slope effect as well as the diffusion capacitance effect to the delay of gate  $i$ :

$$\begin{aligned} D_i(x_i, t_i, C_i) &= \hat{s}_i + \hat{p}_i t_i + \hat{q}_i x_i + \frac{\hat{r}_i}{x_i} C_i, \\ T_i(x_i, t_i, C_i) &= \bar{s}_i + \bar{p}_i t_i + \bar{q}_i x_i + \frac{\bar{r}_i}{x_i} C_i, \end{aligned}$$

where  $x_i$  is the gate size,  $t_i$  is the input rise or fall time of gate  $i$ ,  $C_i$  is the capacitance load,  $\hat{s}_i, \hat{q}_i, \hat{r}_i, \bar{s}_i, \bar{q}_i$  and  $\bar{r}_i$  are precharacterized coefficients. It is not difficult to see that while keeping the size of other components fixed, the input slope  $t_i$  is a linear function of  $x_i$  since  $\hat{c}_i x_i$  contributes only linear term to its parents' capacitance load. Hence the delay of gate  $i$  can be rewritten as follows:

$$D_i(x_i, t_i, C_i) = \hat{s}_i' + \hat{q}_i' x_i + \frac{\hat{r}_i}{x_i} C_i$$

where  $\hat{s}_i' = \hat{p}_i(\bar{s}_j + \bar{p}_j t_j + \bar{q}_j x_j)$ ,  $\hat{q}_i' = \hat{q}_i + \frac{\bar{r}_j \hat{c}_i}{x_i}$ , and component  $j$  is the parent of component  $i$ . It is not hard to see that after the substitution,  $A_i(\mathbf{x}) = \hat{c}_i R_i + \alpha_i + \hat{q}_i'$ . Hence our algorithms in Section 3 will still convergences to the optimal solution under this modification.

## 5 Experimental Results and Concluding Remark

We implemented our algorithms in an RS/6000 workstation. Table 1 shows the experimental results on adders of different sizes ranging from 8 bits to 512 bits. Number of gates range from 120 to 15360. Number of wires range from 96 to 12288 (note that the number of wires here means the

number of sizable wire segments). The total number of sizable components range from 216 to 21648. The stopping criteria of our algorithm is the solution is within 1% of the optimal solution. The lower bound and upper bound of the size of each gate are 1 and 100 respectively. The lower bound and upper bound of the width of each wire are 1 and 3  $\mu\text{m}$  respectively.

Table 1 shows the runtime and storage requirements of our algorithm. For circuit with 13824 sizable components, the runtime and storage requirements of our algorithm are about 13 minutes and 12 MB only. For circuit with 27648 sizable components, the runtime and storage requirements of our algorithm are about half an hour and 23 MB. The maximum delays for the solution of minimum gate and wire sizes and for our algorithm are also listed.

Figure 6 and Figure 7 show the runtime and storage requirements of our algorithm. Figure 6 shows that the runtime increases roughly three times when the circuit size is doubled. Hence the empirical runtime of our program are about  $n^{\log 3 / \log 2} \approx n^{1.6}$ . Figure 7 shows that the ratio of the storage versus the circuit size of our algorithm is close to linear. The storage requirement for each sizable component is about 0.8 KB.

Figure 8 shows the area versus delay tradeoff curve of a 16-bit adder. In our experiment, we observe that to generate a new point in the area and delay tradeoff curve, SOLVE\_LDP converges in only about 5 iterations. It is because the  $\lambda$  of the previous point is a good approximation for that of the new point and hence the convergence of SOLVE\_LDP is fast. As a result, generating these tradeoff curves requires only a little bit more runtime but provides precious information.

Figure 9 shows the convergence sequence of our algorithm SOLVE\_LDP on a 128-bit adder. It shows that our algorithm converges smoothly to the optimal solution. The solid line represents the upper bound of the optimal solution and the dotted line represents the lower bound of it. The lower bound values comes from the optimal value of  $\mathcal{LRS}/\lambda$  at current iteration. Note that the optimal solution is always inbetween the upper bound and the lower bound. So these curves provide useful information about the distance between the optimal solution and the current solution, and help users to decide when to stop the programs.

Finally, we would like to point out that our Lagrangian relaxation approach can be adapted to solve the transistor sizing problem. We observe that the transistor sizing problem is very similar to the gate sizing problem. Although transistor sizing has been extensively studied, we believe the Lagrangian relaxation approach is much more efficient than current techniques. For example, the most recent algorithm that can guarantee exact transistor sizing solutions is [19]. The largest test circuit in [19] has 832 transistors and the reported runtime and memory are 9 hours (on a Sun SPARCstation 1) and 11 MB, respectively. Note that for a problem of similar size (834), our approach only needs 7 seconds of runtime and 1.15 MB memory (see Table 1).

## References

- [1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, Inc., second edition, 1993.
- [2] Chung-Ping Chen, Yao-Wen Chang, and D. F. Wong. Fast performance-driven optimization for buffered clock trees based on Lagrangian relaxation. In *Proc. IEEE Intl. Conf. on Computer-*

Circuit Name	Circuit Size			Maximum Delay (ns)		Runtime (min:sec)	Memory (MB)
	# Gates	# Wires	Total	Min. Size	Our Alg.		
adder (8 bits)	120	96	216	8.55	4.70	0:01	0.48
adder (16 bits)	240	192	432	17.23	8.12	0:02	0.76
adder (32 bits)	480	384	864	33.36	16.00	0:07	1.15
adder (64 bits)	960	768	1728	66.07	31.90	0:15	1.75
adder (128 bits)	1920	1536	3456	131.51	63.70	0:39	2.82
adder (256 bits)	3840	3072	6912	262.43	127.32	3:05	5.37
adder (512 bits)	7680	6144	13824	524.08	256.21	13:09	11.83
adder (1024 bits)	15360	12288	27648	1047.53	508.95	36:12	22.92

Table 1: The runtime and storage requirements of our algorithm on test circuits of different sizes.

*Aided Design*, pages 405–408, 1996.

- [3] Chung-Ping Chen and D. F. Wong. A fast algorithm for optimal wire-sizing under Elmore delay model. In *Proc. IEEE ISCAS*, volume 4, pages 412–415, 1996.
- [4] Chung-Ping Chen, Hai Zhou, and D. F. Wong. Optimal non-uniform wire-sizing under the Elmore delay model. In *Proc. IEEE Intl. Conf. on Computer-Aided Design*, pages 38–43, 1996.
- [5] M. A. Cirit. Transistor sizing in CMOS circuits. In *Proc. ACM/IEEE Design Automation Conf.*, pages 121–124, 1987.
- [6] Jason Cong and Lei He. An efficient approach to simultaneous transistor and interconnect sizing. In *Proc. IEEE Intl. Conf. on Computer-Aided Design*, pages 181–186, 1996.
- [7] Jason Cong and Cheng-Kok Koh. Simultaneous driver and wire sizing for performance and power optimization. In *Proc. IEEE Intl. Conf. on Computer-Aided Design*, pages 206–212, 1994.
- [8] Jason Cong and Kwok Shing Leung. Optimal wiresizing under the distributed Elmore delay model. *IEEE Trans. Computer-Aided Design*, 14(3):321–336, March 1995.
- [9] R. J. Duffin, E. L. Peterson, and C. Zener. *Geometric Programming – Theory and Application*. John Wiley & Sons, Inc., NY, 1967.
- [10] W. C. Elmore. The transient response of damped linear network with particular regard to wideband amplifiers. *J. Applied Physics*, 19:55–63, 1948.
- [11] J. P. Fishburn and A. E. Dunlop. TILOS: A posynomial programming approach to transistor sizing. In *Proc. IEEE Intl. Conf. on Computer-Aided Design*, pages 326–328, 1985.
- [12] M. L. Fisher. An application oriented guide to lagrangian relaxation. *Interfaces*, 15(2):10–21, March–April 1985.
- [13] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison Wesley, second edition, 1984.
- [14] David P. Marple and Abbas El Gamal. Optimal selection of transistor sizes in digital VLSI circuits. In *Proc. 1987 Stanford Conf.*, pages 151–172, 1987.

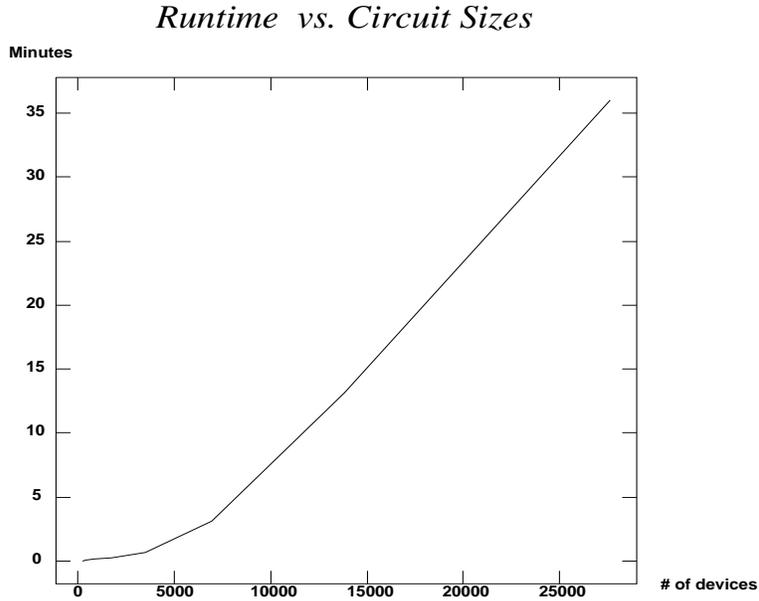


Figure 6: *The runtime requirement of our algorithm versus circuit size.*

- [15] N. Menezes, R. Baldick, and L. T. Pileggi. A sequential quadratic programming approach to concurrent gate and wire sizing. In *Proc. IEEE Intl. Conf. on Computer-Aided Design*, pages 144–151, 1995.
- [16] N. Menezes, S. Pullela, F. Dartu, and L. T. Pileggi. RC interconnect syntheses—a moment fitting approach. In *Proc. IEEE Intl. Conf. on Computer-Aided Design*, pages 418–425, 1994.
- [17] N. Menezes, S. Pullela, and L. T. Pileggi. Simultaneous gate and interconnect sizing for circuit level delay optimization. In *Proc. ACM/IEEE Design Automation Conf.*, pages 690–695, 1995.
- [18] Sachin S. Sapatnekar. RC interconnect optimization under the Elmore delay model. In *Proc. ACM/IEEE Design Automation Conf.*, pages 387–391, 1994.
- [19] Sachin 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. Computer-Aided Design*, 12(11):1621–1634, November 1993.

### Storage vs. Circuit Sizes

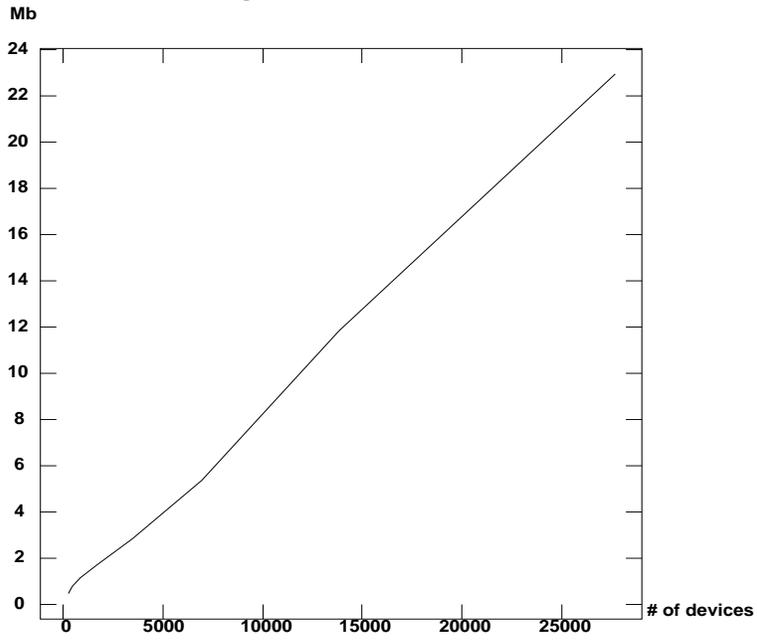


Figure 7: The storage requirement of our algorithm versus circuit size.

### Area vs. Delay Tradeoff Curve

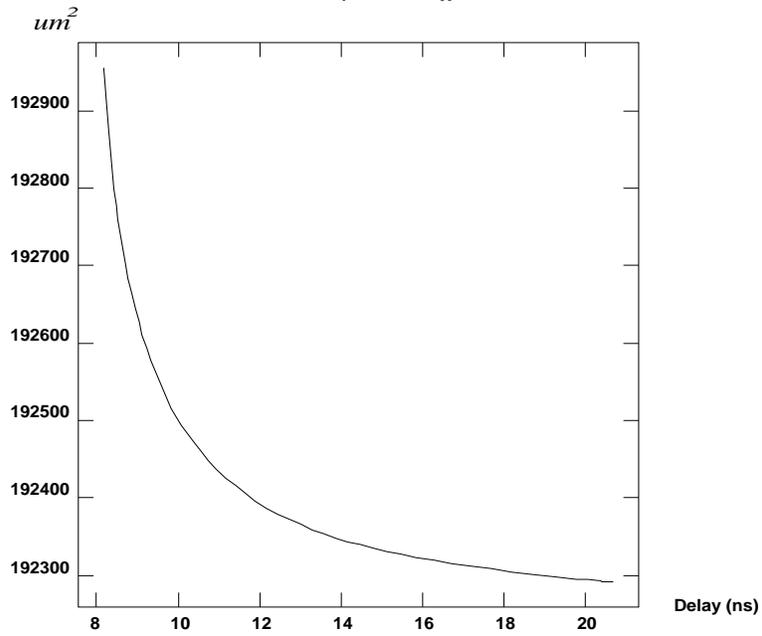


Figure 8: The area versus delay trade off curve for a 16-bit adder.

### The Convergence Curve

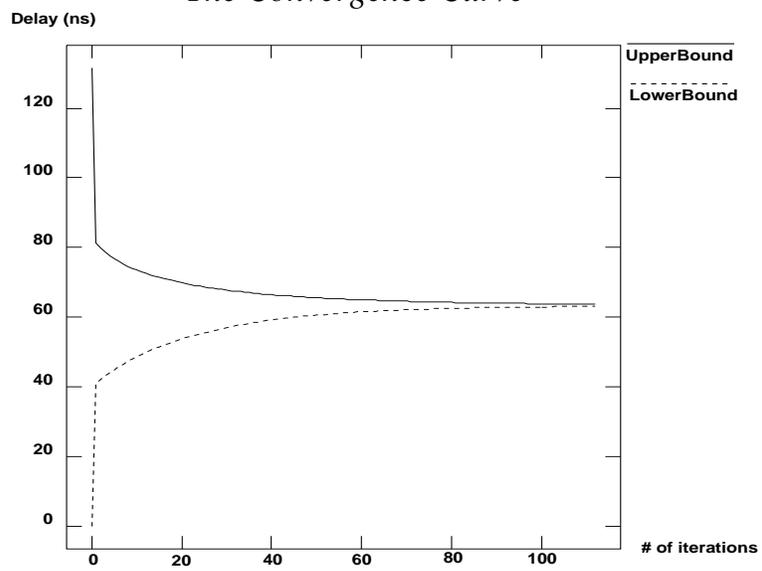


Figure 9: The convergence sequence for a 128-bit adder.