Low-Power Repeater Insertion With Both Delay and Slew Rate Constraints

Yuantao Peng Xun Liu
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695
{ypeng, xunliu}@ncsu.edu

ABSTRACT

In this paper, a novel repeater insertion algorithm is presented to minimize the power dissipation of interconnect trees under given timing budgets and slew rate constraints. In contrast to traditional bottom-up dynamic programming approaches, the proposed algorithm combines a Lagrangian relaxation framework and a graph-based search method to derive possible solutions in a top-down fashion. As a result, it is capable of analyzing repeater slew rates efficiently. In addition, our scheme incorporates accurate circuit models and is therefore able to capture the precise delay and slew rate information, leading to high-quality interconnect designs.

We have applied our scheme to interconnects of different topologies and various timing and slew rate constraints. Experimental results demonstrate the effectiveness of our approach in comparison with previous low-power repeater insertion schemes. Under tight timing constraints, our scheme can always derive repeater insertion solutions that meet both delay and slew rate requirements, whereas other schemes often fail. Under loose timing constraints, our algorithm achieves up to a 23% average power dissipation reduction for different interconnects specifications with shorter runtimes.

Categories and Subject Descriptors

J.6 [**Computer-aided design (CAD)**]: Generic CADD

General Terms

Algorithms

Keywords

Repeater Insertion, Interconnect, Slew Rate, Low Power

1. INTRODUCTION

The repeater insertion technique has been widely used to reduce global interconnect delays. The number of repeaters on a single chip is expected to increase significantly [20]. Consequently, repeater power reduction schemes are crucial in designing future low-power battery-based VLSI chips. The power minimization of repeaters must satisfy several constraints. Specifically, repeater power savings cannot be achieved at the cost of causing timing violations. In addition, the signal slew rate has become a critical design metric in ensuring signal integrity of deep-submicron chips. Signal slew rate constraints therefore must also be considered in algorithms for repeater power reduction.

This paper presents a novel low-power repeater insertion algorithm. The proposed algorithm inserts repeaters into an interconnect tree routed on multiple metal layers so that the total repeater power is minimized and the given delay and signal slew rate constraints are satisfied. Our scheme combines a Lagrangian relaxation framework and a graph-based search method. Unlike bottom-up dynamic programming (DP) approaches, our algorithm derives low-power repeater insertion solutions in a top-down fashion, from the interconnect driver to sinks. Therefore, it can capture the signal slew rates of repeaters efficiently. Moreover, our scheme uses highly accurate timing models generated from SPICE simulations. As a result, it reduces the positive timing slacks for maximal power savings.

We have applied our scheme to a set of interconnect trees with various delay and slew rate constraints to demonstrate its effectiveness. For comparison purposes, we have also implemented a DP based scheme and a repeater sizing based scheme and applied them to the same set of interconnect designs. Experimental results show that our schemes can always generate low-power repeater insertion solutions that meet both delay and signal slew rate constraints. Whereas the other two schemes often fail to derive any valid solutions when the delay constraints are tight. Under loose timing constraints, when the DP-based scheme can derive valid repeater insertion solutions, our scheme can achieve up to 23% power reduction with shorter average runtimes. Compared with the repeater sizing scheme, our scheme can achieve up to 82% power savings.

The rest of our paper contains 5 sections. Section 2 presents background information about repeater insertion and describes our timing and power models. The problem of low-power repeater insertion under given delay and slew rate constraints is formulated in Section 3. Section 4 presents our Lagrangian relaxation based algorithm in details. Section 5 presents experimental results. Section 6 summarizes our paper.

2. BACKGROUND

In this section, we first introduce the general repeater insertion problem. We then briefly review the related research. Finally, we present our timing and power models.

2.1 Repeater Insertion

Figure 1 illustrates the structure of an interconnect tree. The driver s is a buffer that sends data onto the global interconnect. The sinks t_i, $i = 1, 2, \ldots, n$, are buffers driven by the interconnect and can be modeled as capacitors. The interconnect tree is made of

DAC2006, July 24–28, 2006, San Francisco, California, USA.

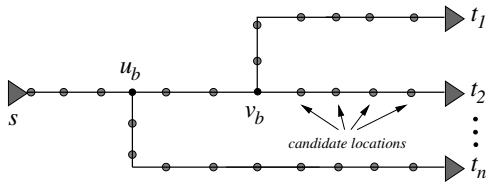

Figure 1: Interconnect tree structure.

several tree branches. Each tree branch b connecting vertices u_b and v_b is routed by a sequence of two-pin uniform interconnects with possibly different RC characteristics. Every uniform two-pin interconnect can be modeled by its length, and resistance and capacitance per unit length. Possible locations, or *candidate locations*, at which repeaters can be inserted are shown as gray dots on the interconnect tree.

Repeaters are often represented by their widths. Feasible repeater widths are finite and form a repeater library. The derivation of a repeater insertion solution for an interconnect is to select several positions from all candidate locations and place repeaters of possibly different widths at these positions. Apparently, different repeater insertion solutions have different electrical properties, e.g., delay and power. The optimization of repeater insertion is to find the solution that optimizes a specific objective, e.g., power reduction, while satisfying certain constraints, e.g., timing targets.

2.2 Previous Research

Repeater insertion has been investigated extensively in the literature [4, 18]. Several circuit models have been proposed to compute the delay or power dissipation of repeaters such as the switch-level RC model [7], the generalized model considering slew rate [13], and the moment matching model [2, 15]. Various design objectives are used such as delay minimization [3, 11, 17], power minimization [9, 12, 16], and cross-coupling noise reduction [1, 5]. Repeater insertion algorithms can be classified as analytical approaches [6] and DP-based approaches [10, 13]. Analytical approaches are usually fast but only applicable to 2-pin interconnects with uniform RC characteristics. DP-based approaches are widely used since they can handle realistic interconnects with practical design constraints such as discreteness of repeater locations. The joint application of the DP-based approach and an analytical solver has been proposed recently for both fast and practical repeater insertion [14].

Early low-power repeater insertion schemes often use the Elmore delay model and linear repeater model to compute the interconnect and repeater delays [1, 10]. These models can overestimate delay values, leading to large positive timing slacks and power degradation. A DP-based low-power repeater insertion algorithm with accurate delay models, i.e., moment-matching method and k-factor gate model, has been proposed in [2]. However, it assumes that the input signal slew rates of all repeaters are the same, leading to delay estimation inaccuracy. The repeater power minimization scheme in [13] uses a generalized repeater delay model which considers the input signal slew rates. It does not take any slew rate constraints into consideration, however. A graph based repeater insertion scheme has been proposed in [8] to handle the signal slew rate constraints. However, it is used for delay minimization and can only be applied to two-pin nets. Low-power repeater insertion schemes that consider both delay and signal slew rate constraints are still elusive.

2.3 Timing and Power Models

Timing models for delay and slew rate computation are critical components of low-power repeater insertion algorithms. Model accuracy is crucial since inaccurate timing models can lead to solutions with constraint violations. On the other hand, since delays and signal slew rates will be frequently computed during the derivation of the low-power repeater insertion solutions, models with high computational complexities will result in prohibitively long runtimes. To ensure high accuracy and low complexity, we use analytical timing models that are derived empirically by extensive SPICE simulations and curve fittings.

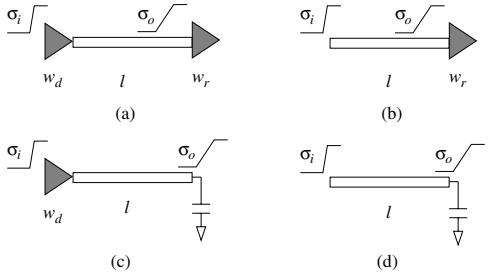

Figure 2: Four cases of a single repeater-interconnect stage.

Our models are created for single repeater-interconnect stages as shown in Figure 2. All repeater insertion solutions of any interconnect tree can be partitioned into a set of single repeater-interconnect stages. Figure 2(a) shows a typical repeater-interconnect stage. Such a stage contains a driver with size w_d, an interconnect segment with length l, and a receiver with size w_r. The interconnect has uniform RC characteristics. There are also three special cases for the repeater-interconnect stage as shown in Figure 2(b)–(d). Specifically, there may not be a driver at the beginning point of the interconnect. Second, when the receiver does not exist, a capacitance is used at the end of the interconnect to represent the downstream load. All four cases of the repeater-interconnect stage can be represented by a three-dimensional vector (w_d, l, w_r). In cases that the driver does not exist, the corresponding w_d is set to infinitely large, i.e., ∞. If there is no receiver, a virtual repeater whose input capacitance is equal to the downstream load is used as a receiver.

Every signal is modeled as a finite ramp whose slope is characterized by signal slew rate, i.e., time between the 10% and 90% of the signal swing. Two analytical models are generated to compute the propagation delay d_o of a repeater-interconnect stage from the input of the driver to the input of the receiver and the signal slew rate σ_o at the input of the receiver, respectively. The inputs of both models include the stage specification (w_d, l, w_r) and the signal slew rate σ_i at the input of the driver. A fourth order complete polynomial is chosen as the function template for both models. Therefore, the signal delay d_o and signal slew rate σ_o are calculated as:

$$d_o(w_d, \sigma_i, l, w_r) = \sum_{\substack{i+j=4 \\ 0 \le i,j \le 4}} \alpha_{i,j}(w_d, \sigma_i) l^i w_r^j$$

$$\sigma_o(w_d, \sigma_i, l, w_r) = \sum_{\substack{i+j=4 \\ 0 \le i,j \le 4}} \beta_{i,j}(w_d, \sigma_i) l^i w_r^j \qquad (1)$$

where $\alpha_{i,j}(w_d, \sigma_i)$ and $\beta_{i,j}(w_d, \sigma_i)$ are coefficients and unique for each (w_d, σ_i) pair.

The characterization procedure of our timing models is straightforward. The four-dimensional space (σ_i, w_d, l, w_r) is evenly sampled first. For any (σ_i, w_d, l, w_r), the repeater-interconnect stage (w_d, l, w_r) is simulated in SPICE with an input signal of slew rate σ_i to derive the delay d_o and slew rate σ_o. For special cases when no repeater exists at the beginning of the interconnect, a voltage

source is used instead. Curve-fittings are then used to derive the coefficients $\alpha_{i,j}(w_d,\sigma_i)$ and $\beta_{i,j}(w_d,\sigma_i)$ for all (i,j). Experimental results show that our model for delay computation has a 2% error on the average in comparison with SPICE simulations. Our model for signal slew rate computation has a 4% error on the average.

Power dissipation of repeaters needs to be estimated accurately since it is the optimization objective of low-power repeater insertion algorithms. Repeater power includes switching power, leakage power, and short circuit power. Both switching power and leakage power of repeaters are linearly proportional to repeater widths. Short-circuit power is only a small fraction of the total power [**?**] and is also monotonic with respect to repeater width. As result, the total repeater width is used to approximate the total repeater power in the rest of this paper.

3. PROBLEM FORMULATION

The problem of low-power repeater insertion for given timing and signal slew rate constraints (LPRI$_{ts}$) is described as follows:

Problem LPRI$_{ts}$: *Let $T=(B,V)$ be an interconnect tree with n sinks, where V is the set of vertices and B is the set of interconnect branches that connect two vertices. The source of the tree is denoted as s and sinks of the tree are denoted as t_i, $i=1,2,\ldots,n$. The buffer sizes of both the source driver and sink receivers are given and cannot be changed. Furthermore, let* $\mathbf{W}$ *be the repeater library containing possible repeater widths, and let P_b be the set of candidate repeater locations along branch $b\in B$. Given the required arrival time q_i for each sink i and the signal slew rate constraint ς for all repeaters, find the repeater width $w_{b,j}\in\{\mathbf{W}\cup 0\}$, $j=1,\ldots,|P_b|$ for each location $p_{b,j}$ in any $P_b, b\in B$ to*

$$\begin{aligned}
&\textit{Minimize:} && \sum_{b\in B} W_b \\
&\textit{Subject to:} && \forall k\in V, \exists a_k \geq 0\,, && (2)\\
& && a_{u_b}+d_b \leq a_{v_b}\,, && (3)\\
& && \forall v_b\in\{t_i | i=1,2,\ldots,n\},\ a_{v_b}=q_i\,, && (4)\\
& && \forall w_{b,j}\neq 0,\ \sigma_{b,j}\leq\varsigma\,, && (5)
\end{aligned}$$

where $W_b=\sum_{j=1,2,\ldots,|P_b|} w_{b,j}$ denotes total widths of repeaters on b, u_b and v_b are the beginning and end points of b, and $w_{b,j}=0$ indicates that no repeater is placed at location $p_{b,j}$. The parameter d_b represents the signal delay from u_b to v_b.

Intuitively, a_k represents the required arrival time (RAT) at vertex k, $\sigma_{b,j}$ represents the input signal slew rate of a repeater inserted at location $p_{b,j}$ when $w_{b,j}\neq 0$. The constraints (2)–(4) ensure that all the delay constraints are satisfied. The constraints (5) ensure that the signal slew rates at the inputs of all repeaters meet the slew rate constraint.

It is very challenging to solve Problem LPRI$_{ts}$ directly because it contains a large number of constraints. Lagrangian relaxation method [21, 14] is therefore applied to convert Problem LPRI$_{ts}$ into a dual problem with less unknowns and constraints. Specifically, a Lagrangian relaxation function is created by introducing a set of Lagrangian multipliers as:

$$L(\vec{W},\vec{a},\vec{\sigma},\vec{\lambda}) = \sum_{b\in B} W_b + \sum_{b\in B}\lambda_b(a_{u_b}+d_b-a_{v_b}) + \sum_{w_{b,j}\neq 0}\lambda_{\sigma_{b,j}}(\sigma_{b,j}-\varsigma)$$

where $\vec{\lambda}$ is a vector containing all Lagrangian multipliers λ_b and $\lambda_{\sigma_{b,j}}$, $\vec{W}$ includes all repeater widths, $\vec{a}$ contains the RATs at all vertices, and $\vec{\sigma}$ denotes the signal slew rates at the inputs of all repeaters. The equivalent dual problem can then be described as:

Dual problem: *Given an interconnect tree $T=(B,V)$ of n sinks, possible repeater locations $P=\{P_b | b\in B\}$, a repeater library* $\mathbf{W}$*, timing constraints at all sinks* $\mathbf{q}$ *and the signal slew rate constraint* ς*, for every location $p_{b,j}$ in each P_b, find a repeater width $w_{b,j}\in\{\mathbf{W}\cup 0\}$ to*

$$\begin{aligned}
&\textit{Maximize:} && Q_{\vec{\lambda}}\\
&\textit{Subject to:} && \lambda_b = \sum_{b'\in des(b)}\lambda_{b'}\\
& && \forall b\in B\,,\ \lambda_b\geq 0\\
& && \forall w_{b,j}\neq 0\,,\ \lambda_{\sigma_{b,j}}\geq 0
\end{aligned}$$

where $Q_{\vec{\lambda}}$ is defined as the minimal value of $L(\vec{W},\vec{\sigma},\vec{\lambda})$.

4. REPEATER INSERTION ALGORITHM

As a special case of the general Lagrangian problem, the dual problem of low-power repeater insertion under delay and slew rate constraints can be solved by a general framework. Specifically, the framework iteratively performs two tasks: deriving a repeater insertion solution to maximize $Q_{\vec{\lambda}}$ for a given $\vec{\lambda}$ and updating $\vec{\lambda}$ using an existing solution. It produces the optimal solution when convergence is reached. Liu *et al.* used such a framework to solve the problem of minimal repeater insertion under timing constraints [14]. In particular, they applied a modified DP-based algorithm to derive the optimal repeater insertion for any $\vec{\lambda}$ and used the subgradient method for updating $\vec{\lambda}$. Recently, Peng *et al.* replaced the subgradient method by the ellipsoid method to achieve fast convergence of $\vec{\lambda}$ [19].

When slew rate constraints are considered in repeater power reduction, i.e. the problem addressed in this paper, schemes in [14, 19] cannot be applied. Specifically, the number of Lagrangian multipliers $\lambda_{\sigma_{b,j}}$ introduced due to signal slew rate constraints is equal to the number of repeaters inserted in current repeater insertion solution, which can vary during the iteration of the framework. Therefore, neither subgradient nor ellipsoid method can be used to update $\lambda_{\sigma_{b,j}}$, which are part of $\vec{\lambda}$. In addition, DP based repeater insertion approaches generate solutions in a bottom-up fashion, i.e., from the interconnect sinks to the driver. On the other hand, the computation of signal slew rate at any location on the interconnect requires upstream signal information. As a result, DP based schemes cannot achieve both high accuracy and high efficiency in capturing signal slew rates.

We next present our algorithm to solve the dual of Problem LPRI$_{ts}$. The novelty of our scheme is two-fold. First, we partition the interconnect trees into branches and solve them in a top-down fashion, from the interconnect driver to sinks. Second, we modify the Lagrangian function by introducing a penalty function to eliminate $\lambda_{\sigma_{b,j}}$. As a result, the dimension of $\vec{\lambda}$ becomes constant, leading to the easy update of $\vec{\lambda}$.

4.1 Interconnect Tree Partitioning

As presented in Section 2, our timing models are derived for two-pin repeater-interconnect stages. As a result, to derive delays and slew rates for signals on interconnect trees, we must partition these trees into two-pin branches. Each branch may contain multiple candidate repeater locations and be routed by a sequence of interconnects with possibly different RC characteristics. After the partitioning, for each two-pin interconnect branch, we need to calculate the input signal slew rate at the beginning of the branch and equivalent load repeater at the end of the branch.

Figure 3 illustrates the procedure of tree partitioning. On the left, a simple tree topology is shown that forks at vertex v. The closest repeaters on paths $v\rightsquigarrow g$ and $v\rightsquigarrow h$ are w_g and w_h, respectively.

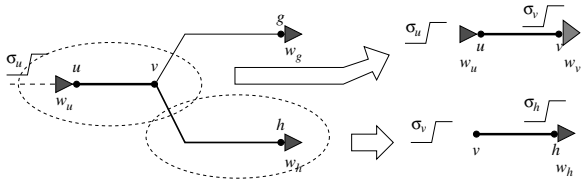

Figure 3: Equivalent receiver and input slew rate derivation.

This tree is divided at v into three parts. For the branch along the path $u \rightsquigarrow v$, a virtual repeater w_v is added to model the downstream capacitive load. Specifically, the width of w_v is calculated as C_v/C_o, where C_o is the input capacitance per gate width and C_v is the sum of all downstream capacitance from v, including gate capacitance of w_g and w_h and interconnect capacitance of $v \rightsquigarrow g$ and $v \rightsquigarrow h$. Note that repeater w_v is only used to facilitate the derivation of repeater insertion solution of the branch along $u \rightsquigarrow v$, not a real repeater actually inserted.

The calculation of input slew rates of branches $v \rightsquigarrow g$ and $v \rightsquigarrow h$ are straightforward. In particular, we analyze and solve all branches in a top-down order, i.e., following a preorder tree traverse. Since the end point of a branch b is the beginning point of its child branches, the input slew rates of the child branches are equal to the slew rate at the end point of b, which can be computed once the repeater insertion solution on b is derived.

4.2 Graph Model for Two-pin Branches

We next describe our scheme of calculating the minimal Lagrangian function $L(\vec{W}, \vec{\sigma}, \vec{\lambda})$ of a given $\vec{\lambda}$ for a two-pin interconnect branch. Specifically, we model the branch as a direct acyclic graph (DAG) and conduct a shortest path algorithm on the DAG.

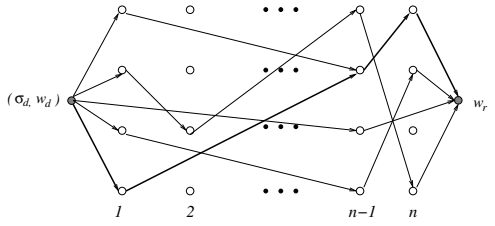

Figure 4: Graph model for two-pin nets

Figure 4 illustrates the DAG generation for an interconnect branch. In particular, the equivalent driver[1] and receiver are represented as the leftmost and rightmost dots on the DAG, respectively. The widths of the driver and receiver, and the input signal slew rate of the driver are known and denoted as w_d, w_r, and σ_d, respectively. Without the loss of generality, the branch is assumed to contain n candidate locations. For each candidate location $i \in \{1,2,\ldots,n\}$, a column of nodes are introduced. Each node represents a repeater insertion *scenario* (σ, w) , where σ is the input slew rate of the repeater, w is the repeater width. The signal slew rate is discretized into K values and the repeater width is from a repeater library of M repeaters. As a result, there are $M*K$ nodes, or scenarios, in each column. Note that the equivalent receiver on the DAG represents a set of scenarios with repeater width w_r and all possible input signal slew rates.

Edges are inserted as follows. Any scenario $(\sigma_{i,m,k}, w_{i,m,k})$, $i \in \{1,2,\ldots,n\}$, $m \in \{1,2,\ldots,M\}$, $k \in \{1,2,\ldots,K\}$ represents the insertion of a repeater with width m and input slew rate k at location i. If its downstream closest repeater has width $m' \in \{1,2,\ldots,M\}$ and is located at $i' \in \{1,2,\ldots,n\}$, the input slew rate of the downstream repeater can be computed using our timing models. If the resulting slew rate is closest to the discretized slew rate $k' \in \{1,2,\ldots,K\}$, an edge is then used to connect $(\sigma_{i,m,k}, w_{i,m,k})$ and $(\sigma_{i',m',k'}, w_{i',m',k'})$. All the nodes are examined in a top-down manner, i.e., from left to right in Figure 4, for the possible insertion of edges. Since each node in column i can only have $(n-i)M$ different downstream neighbor configurations, the total number of edges and the time complexity for edge insertion are $O(n^2M^2K)$. After the DAG is built, any repeater insertion solution of the interconnect branch can be represented as a path from the driver to the receiver.

[1]In case that there is not a driver, the timing models created for the circuit in Figure 2(b) can be used.

4.3 Penalty Function to Eliminate λ_σ

Each path in the DAG represents a unique repeater insertion solution for the corresponding interconnect branch. If edge weights are set to the corresponding increases of the Lagrangian function value, the minimization of the Lagrangian function value can be achieved by finding the shortest path in the DAG. Specifically, The weight of an edge $e_{i_k,j_l}, i < j, k, l \in \{1,2,\ldots,M*K\}$ between a scenario $(\sigma_{i,k}, w_{i,k})$ at candidate location i and a scenario $(\sigma_{j,l}, w_{j,l})$ at candidate location j, is

$$\Gamma(e_{i_k,j_l}) = w_{j,l} + \lambda_b \tau(e_{i_k,j_l}) + \lambda_{\sigma_{j,l}}(\sigma_{j,l} - \varsigma) \,, \qquad (6)$$

where λ_b denotes the Lagrangian multiplier for the delay constraint of current interconnect branch b, $\lambda_{\sigma_{j,l}}$ denotes the Lagrangian multiplier for the slew rate constraint at location j and $\tau_{e_{i_k,j_l}}$ denotes the signal propagation delay between inputs of repeaters at i and j.

Although the weight assignment in (6) can be used to minimize the Lagrangian function for a given $\vec{\lambda}$, it has some drawbacks. In particular, it contains the Lagrangian multipliers $\lambda_{\sigma_{j,l}}$, which cannot be updated efficiently. We propose a new way of calculating the edge weights in the DAG by introducing a penalty function that eliminate $\lambda_{\sigma_{j,l}}$. Specifically, our new edge weight is:

$$P(e_{i_k,j_l}) = (w_{j,l} + \lambda_b \tau(e_{i_k,j_l}))(1 + \lambda_b \tau(e_{i_k,j_l}) \Theta(\sigma_{j,l})) \qquad (7)$$

where function $\Theta(\sigma_{j,l})$ is defined as:

$$\Theta(\sigma_{j,l}) = \begin{cases} 0 & \text{if } \sigma_{j,l} \le \varsigma \\ \sigma_{j,l} - \varsigma & \text{if } \sigma_{j,l} > \varsigma \end{cases}$$

The new edge weight expression has several additional favorable properties in addition to helping eliminate $\lambda_{\sigma_{j,l}}$. First, when no slew violation exists, $P = \Gamma$. As a result, the Lagrangian function value calculated using P is the same with that using Γ, maintaining the optimality of the result. When slew violations exist, the penalty function Θ increases the value of P, providing a similar effect as that of the third term $\lambda_{\sigma_{j,l}}(\sigma_{j,l} - \varsigma)$ in (6).

4.4 Algorithm Summary

The pseudocode of our low-power repeater insertion algorithm under delay and slew rate constraints (RDS) is shown in Figure 5. With the top-down solution computation strategy and elimination of Lagrangian multipliers for slew constraints, our algorithm is able to apply the general Lagrangian framework. Specifically, the remaining Lagrangian multipliers $\vec{\lambda}$ are initialized in Line 1. The **repeat** loop in Lines 2–9 is then used to iteratively compute the optimal repeater widths for all candidate locations. In particular, the interconnect branches are traversed in preorder. For each branch b, the corresponding equivalent driver and receiver are derived in Line 5. Our graph-based two-pin net solver G2P is then applied to derive the repeater insertion solution that minimizes the modified Lagrangian function for b. Since the repeater insertion solution of b

will affect the equivalent receiver computation for the parent branch of b, the **for** loop in Lines 4–6 needs to be repeated till the resulting Lagrangian function converges. In Line 8, the $\vec{\lambda}$ is updated based on the repeater insertion results of all branches. If a timing violation occurs at a receiver of the interconnect tree, the corresponding Lagrangian multiplier will be increased. Algorithm RDS returns the final solution in Line 9 when $\vec{\lambda}$ converges. Note that although the convergence of algorithm RDS is not theoretically guaranteed, convergences are reached for all the test cases.

```
RDS(T)
1    initialize λ⃗ = λ⃗_0
2    repeat
3        repeat
4            foreach (b ∈ B in preorder tree traverse)
5                derive equivalent driver and receiver for b
6                G2P(b)
7            if (Lagrangian function converges) then break
8        update λ⃗ according to the timing slacks
9        if (λ⃗ converges) then return all w_{b,j}
```

Figure 5: Algorithm RDS

```
G2P(b)
1    insert equivalent driver scenario (σ_d, w_d)
2    foreach (σ_k, w_k) ∈ column i, where i ∈ {0, 1, ..., n}
3        generate all output edges with weights
4    perform shortest path search
5    return optimal repeater insertion solution
```

Figure 6: Subroutine G2P

The pseudocode of our graph-based two-pin net solver G2P is shown in Figure 6. For a given interconnect branch b, subroutine G2P constructs the DAG and performs a shortest path algorithm. Specifically, Line 1 initializes the driver scenario (σ_d, w_d). Lines 2–3 insert all weighted edges. After the DAG is built, Line 4 finds the shortest path from the driver to the receiver. Line 5 returns the optimal solution.

5. EXPERIMENT SETUP AND RESULTS

We have applied our scheme RDS to various interconnect trees to demonstrate its effectiveness. Specifically, our interconnect trees were generated using a TSMC 0.18μm technology. The total number of sinks ranged from 2 to 30. The length of tree branches ranged from 800μm to 2400μm. The distance between adjacent repeater candidate locations was 400μm.

The characterization procedure in Subsection 2.3 was used to derive our timing models. Specifically, the driver size w_d and receiver size w_r varied from $10u$ to $400u$ with a granularity of $10u$. The interconnect length l between the driver and the receiver ranged from 0μm to 2400μm with a granularity of 400μm. The input signal slew rate σ_i varied from 100ps to 1000ps with a granularity of 100ps. A curve fitting algorithm with minimal mean-square error criterion was used to compute all the coefficients in our timing models.

In our experiments, a repeater library containing repeaters from $10u$ to $200u$ with a granularity of $10u$ was used, where u is the minimal repeater width. The timing targets ranged from $0.85\tau_{elm}$ to $1.15\tau_{elm}$, where τ_{elm} is the minimal Elmore delay of the interconnect tree that can be achieved by repeater insertion. For each design, 10 timing targets evenly distributed in the above range were used. The signal slew rate constraint was set to 400ps.

For comparison purposes, we have implemented two low-power repeater insertion schemes. The first one, called DPA, is a DP-based scheme with the moment matching method and k-factor model for interconnect and repeater delay computation [2]. To handle the signal slew constraints, an upper bound of the effective load capacitance is introduced for each repeater in the library. Such a bound, derived by SPICE simulations and unique to the slew rate constraint ς, ensures that, when the corresponding repeater drives any effective capacitive load below the bound, the resulting slew rate is always less than ς. During the execution of DPA, any solution that contains a repeater whose load is greater than the corresponding bound is eliminated from the solution set.

The second comparison scheme, called SIZ, is based on repeater sizing. Specifically, SIZ first performs a DP algorithm with the Elmore delay model to derive a low-power repeater insertion solution with only the delay constraints. The actual delays and signal slew rates are then computed using our proposed timing models. Since no slew rate is considered, such a solution often has slew rate violations. SIZ then checks all repeaters in a postorder traverse. Whenever a repeater with slew violation is found, SIZ increases the size of the parent repeater until the violation is eliminated or the maximal repeater in the library is chosen. Since the size of interconnect driver s cannot be changed, a buffer chain with a tapping ratio of 2.7 is used at the output of s to remove any slew violation and reduce the signal delay.

Table 1 shows the experimental results from 40 interconnects. Columns 2–3 show the numbers of sinks and candidate locations in these nets. Our scheme RDS can derive valid solutions for all nets under any timing and slew rate constraints according to SPICE simulations. DPA produces results with no slew rate violations but fails to meet delay constraints in some cases. The numbers of delay violations V_{DPA}^{T} for each interconnect net out of total 10 runs are shown in Column 4. SIZ fails to eliminate either delay or slew rate violations completely. The numbers of delay violations V_{SIZ}^{T} and slew rate violations V_{SIZ}^{S} out of 10 runs are listed in Columns 5–6. When both RDS and DPA can derive valid solutions, the power dissipations of the results, P_{RDS} and P_{DPA}, are compared. The average power reductions of RDS over DPA, i.e., $\Delta_{DPA} = (P_{DPA} - P_{RDS})/P_{DPA}$, for each net with different timing targets are listed in Column 7. Column 8 shows the similar power comparison $\Delta_{SIZ} = (P_{SIZ} - P_{RDS})/P_{SIZ}$ between RDS and SIZ. As can be seen from the table, our scheme can achieve up to 23% and 82% average power reduction in comparison with other two schemes.

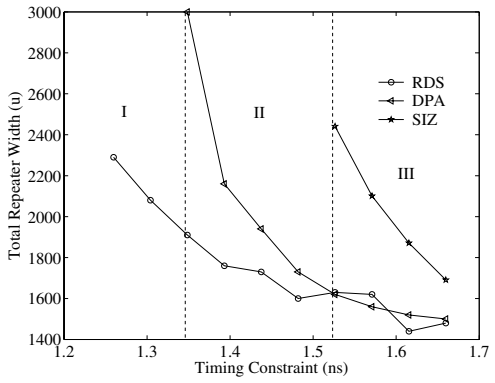

Figure 7: Total Repeater Width vs. Timing Constraints

Figure 7 reveals the relationship between the total repeater width and the timing constraints for the three schemes. As can be seen, when the timing constraints are tight, as shown by region I in the figure, RDS can derive valid solutions, whereas DPA and SIZ cannot. When the timing targets are in the middle range, i.e., region II,

Table 1: Experimental results.

Net	Sink	Loc	V^T_{DPA}	V^T_{SIZ}	V^S_{SIZ}	Δ_{DPA} (%)	Δ_{SIZ} (%)
1	12	75	2	5	5	6.43	38.81
2	30	216	2	6	4	9.74	10.91
3	28	201	2	5	5	8.65	16.98
4	19	142	2	6	5	6.29	18.21
5	11	85	3	5	4	6.15	37.68
6	6	41	2	5	4	8.63	58.14
7	14	91	2	5	5	8.11	37.00
8	21	156	2	5	5	8.77	21.04
9	16	120	2	5	4	7.68	31.68
10	9	66	2	5	3	9.07	45.58
11	8	49	2	5	5	6.52	50.36
12	17	127	2	5	5	4.25	25.79
13	15	111	3	5	4	7.11	38.62
14	14	109	2	5	5	7.28	29.10
15	22	161	2	5	6	6.49	13.49
16	12	93	2	5	5	3.40	26.53
17	15	113	2	5	6	6.00	29.32
18	25	182	3	5	6	3.69	15.79
19	18	138	2	6	4	7.46	18.43
20	13	102	2	5	5	6.24	27.31
21	4	29	2	6	3	19.84	67.10
22	19	141	2	5	5	0.80	16.76
23	26	189	3	5	4	5.33	21.49
24	17	129	2	6	4	7.51	24.76
25	11	85	2	5	5	8.09	34.73
26	29	238	3	5	4	1.54	22.02
27	20	147	2	5	5	0.88	14.77
28	16	119	2	5	4	8.11	29.25
29	24	178	2	5	5	8.34	20.59
30	11	69	2	6	5	6.39	32.27
31	18	134	2	5	4	3.97	26.54
32	23	171	2	5	5	6.09	19.29
33	27	196	2	5	6	8.96	18.23
34	2	11	2	6	0	23.06	82.62
35	10	62	2	5	4	8.48	42.33
36	5	34	2	6	3	8.66	55.30
37	13	93	2	6	5	11.82	32.40
38	3	20	2	6	2	16.80	71.45
39	7	50	2	5	5	8.81	52.98
40	20	131	2	5	5	0.63	19.88
Ave	-	-	2	5	5	7.6	32.4

SIZ still fails to derive any valid solutions. Although DPA can derive valid solutions in this region, RDS achieves significant power reductions compared with DPA. When the timing target is very loose, i.e., in region III, both DPA and RDS delivers valid repeater insertion solutions with comparable power dissipation. The solutions of SIZ, though valid, have significant power degradation in comparison with those of DPA and RDS. The average runtime of RDS is in seconds, 37% shorter than that of DPA.

6. CONCLUSION AND FUTURE WORK

This paper presents a low-power repeater insertion scheme for global interconnect trees under given timing and slew rate constraints. Our scheme combines a Lagrangian relaxation framework and a graph-based search method to derive the repeater insertion solutions in a top-down fashion. Consequently, our scheme can capture the accurate signal slew rates of repeaters efficiently. Moreover, due to the incorporation of accurate delay and signal slew rate models, our schemes can produce valid repeater insertion results with substantial power savings.

Experimental results demonstrate the effectiveness of our scheme. In comparison with other low-power repeater insertion schemes, for high performance VLSI designs with tight timing constraints, our scheme can deliver violation-free repeater insertion solutions, while the other schemes fail to derive solutions that meet the timing and/or slew rate constraints. When the timing constraints are loose, our scheme can achieve up to 23% average power dissipation reduction for different nets of various timing constraints with shorter average runtimes.

In our implementation, we ignore the short-circuit power and use total repeater width to approximate total repeater power. More accurate power models can be used to increase the power estimation accuracy. Furthermore, better effective capacitance estimation can also be incorporated into our scheme to improve the solution quality. In addition, the repeater library used in our algorithm only contains non-inverting repeaters. It is an interesting future work to extend our scheme to handle both inverting and non-inverting repeaters.

7. REFERENCES

[1] C. Alpert, A. Devgan, and S. T. Quay. Buffer insertion for noise and delay optimization. *IEEE Trans. CAD*, 18(11):1633–1645, Nov. 1999.

[2] C. J. Alpert, A. Devgan, and S. T. Quay. Buffer insertion with accurate gate and interconnect delay computation. In *DAC*, June 1999.

[3] C. J. Alpert, J. Hu, S. S. Sapatnekar, and P. G. Villarrubia. A practical methodology for early buffer and wire resource allocation. *IEEE Trans. CAD*, 22(5), May 2003.

[4] H. B. Bakoglu. *Circuits, Interconnects, and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.

[5] C. P. Chen and N. Menezes. Noise-aware repeater insertion and wire sizing for on-chip interconnect using hierarchical moment-matching. In *DAC*, June 1999.

[6] C.-C. N. Chu and D. F. Wong. Closed form solution to simultaneous buffer insertion/sizing and wire sizing. In *ISPD*, Apr. 1997.

[7] J. Cong, L. He, C. K. Koh, and P. H. Madden. Performance optimization of VLSI interconnect layout. *Integration, the VLSI Journal*, 21(1):1–94, Jan. 1996.

[8] Y. Gao and D. F. Wong. A graph based algorithm for optimal buffer insertion under accurate delay models. In *DATE*, Mar. 2001.

[9] G. S. Garcea, N. P. van der Meijs, and R. H. Otten. Simultaneous analytical area and power optimization for repeater insertion. In *ICCAD*, Nov. 2003.

[10] L. Ginneken. Buffer placement in distributed RC-tree networks for minimal Elmore delay. In *ISCAS*, 1990.

[11] N. Hedenstierna and K. O. Jeppson. CMOS circuit speed and buffer optimization. *IEEE Trans. CAD*, 6(2):270–280, Feb. 1987.

[12] P. Kapur, G. Chandra, and K. C. Saraswat. Power estimation in global interconnect and its reduction using a novel repeater optimization methodology. In *DAC*, June 2002.

[13] J. Lillis, C. K. Cheng, and T.-T. Y. Lin. Optimal wire sizing and buffer insertion for low power and a generalized delay model. *JSSC*, 31(3):437–447, Mar. 1996.

[14] I.-M. Liu, A. Aziz, and D. F. Wong. Meeting delay constraints in DSM by minimal repeater insertion. In *DATE*, Mar. 2000.

[15] N. Menezes and C.-P. Chen. Spec-based repeater insertion and wire sizing for on-chip interconnect. In *Proc. VLSI Design*, 1999.

[16] A. Nalamalpu and W. P. Burleson. A practical approach to DSM repeater insertion: Satisfying delay constraints while minimizing area and power. In *IEEE Inter. ASIC/SOC Conference*, Sept. 2001.

[17] M. Nekili and Y. Savaria. Optimal methods of driving interconnections in VLSI circuits. In *ISCAS*, May 1993.

[18] R. Otten. Global wires harmful? In *ISPD*, Apr. 1998.

[19] Y. Peng and X. Liu. Freeze: Engineering a fast repeater insertion solver for power minimization using the ellipsoid method. In *DAC*, June 2005.

[20] P. Saxena, N. Menezes, P. Cocchini, and D. A. Kirkpatrick. Repeater scaling and its impact on CAD. *IEEE Trans. CAD*, 23(4):451–463, Apr. 2004.

[21] J. F. Shapiro. *Mathematical Programming: Structures and Algorithms*. Wiley-Interscience Publication, 1979.