

Practical Repeater Insertion For Low Power: What Repeater Library Do We Need?

Xun Liu Yuantao Peng

Dept. of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695
{xunliu, ypeng}@ncsu.edu

Marios C. Papaefthymiou

Dept. of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109
marios@eecs.umich.edu

ABSTRACT

In this paper, we investigate the problem of repeater insertion for low power under a given timing budget. We propose a novel repeater insertion algorithm to compute the optimal repeater number and width in the discrete solution space, as defined by a given repeater library. Using our algorithm, we show that rounding the solution under the continuity assumption to the closest discrete solution candidate may result in suboptimal designs, or it may even fail to find an existing solution. Given a certain tolerance to the degradation of repeater power dissipation, we address two practical and highly important questions: (1) How coarse could the repeater size granularity be? (2) What range should the repeater size be in?

Experimental results demonstrate the high effectiveness of the proposed scheme and provide valuable insights into repeater library design. Our approach achieves up to 23% power reduction in comparison to rounding-based approaches. With a 4% power degradation tolerance, repeater size granularity as coarse as 8λ can be used, reducing the library size by more than 87%. For interconnects with various wire lengths and timing targets, our investigation reveals that the range of optimal repeater sizes for low-power is limited, indicating that a low-cost small-size repeater library, if well designed, is adequate to provide high quality repeater insertion solutions.

Categories and Subject Descriptors

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

General Terms

Design

Keywords

Interconnect, Repeater Insertion, Low power

1. INTRODUCTION

Due to the ever increasing chip dimension and decreasing transistor feature size, repeaters are inserted into global interconnects among circuit blocks to reduce signal delays. As a result, the number of repeaters is expected to exceed 1 million in deep submicron VLSI systems [9]. To alleviate the ensuing design complexity, libraries containing a limited number of various-size repeaters are often designed. The huge number of repeaters also results in high

power dissipation, which can contribute up to 20% of total chip power dissipation [17, 18]. Consequently, the development of a power-efficient repeater insertion methodology is essential.

This paper investigates the problem of repeater power reduction. Two main contributions are presented. First, a novel algorithm is proposed to derive a repeater insertion solution with minimum power dissipation under a given timing budget, considering the discreteness of repeater sizes. Starting from a solution estimate, our scheme performs a search that is guaranteed to find the discrete optimal solution or report that no solution exists. The proposed scheme deviates from other repeater insertion methods by its provably low computation complexity and the use of highly accurate timing and power models that capture detailed circuit physics and practical design issues such as short circuit power and repeater location flexibility.

The second contribution of this paper is that, using the proposed approach, we provide valuable insights into the design of repeater libraries. Specifically, we address the issue of selecting the repeaters for a library by computing the proper granularity and range of repeater sizes.

We have implemented our low-power repeater insertion algorithm into a software tool called RIS and used it to design global interconnects of different lengths in $0.18\mu\text{m}$ technology using various timing budgets. Experimental results show that solutions derived from RIS dissipate up to 23% less power than those calculated by rounding the solutions under the continuity assumption to the closest discrete solution candidates. We have also used RIS to improve repeater library design. For the $0.18\mu\text{m}$ technology used in this paper, RIS can reduce the repeater library size by more than 87% with less than 4% average power degradation.

A plethora of research on repeater insertion has been reported in the literature [2, 16]. Several repeater models are proposed such as the switch-level RC model [5], the linearized form of the Shichman-Hodges equations [20], and the short-channel α -power law model [1]. Early repeater insertion algorithms ignore geometric constraints due to layout and assume that repeaters can be placed at any location derived from the mathematical analysis [4, 8, 12]. The concept of feasible region is proposed in [6]. A well-designed global interconnect must be able to meet its timing budget as long as the repeaters are placed within their feasible regions. The worst-case delay increase due to the repeater location deviation is studied in [14]. Though repeater insertion is originally applied to reduce signal delays of global interconnects, the minimization of repeater power consumption has become equally important in nanometer system designs. Repeater insertion with minimum power subject to timing constraints is investigated in [7, 15] under the assumption of continuously variable repeater widths. The combination of repeater insertion with other techniques, such as flip-flop insertion [11] and wire sizing [10], for power reduction is also proposed. Repeater insertion with a given library with discrete repeater widths

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC'04, June 7–11, 2004, San Diego, California, USA.

Copyright 2004 ACM 1-58113-828-8/04/0006 ...\$5.00.

has been addressed in [12, 13, 19]. The research on the design of repeater libraries is limited, however.

The remainder of this paper has seven sections. Section 2 presents background material on the delay and power models of repeaters. Section 3 formulates the problem of repeater insertion for low power with discrete widths. Our repeater insertion algorithm is presented in Section 4. Section 5 gives the power comparison results and demonstrates the high effectiveness of our approach over rounding-based schemes. Our study on repeater library design is described in Section 6. In Section 7, we use the proposed scheme to assess the adverse power impact of limiting the number of repeaters inserted to be even. We also analyze the power degradation due to repeater location deviation. Section 8 summarizes our paper.

2. REPEATER DELAY AND POWER

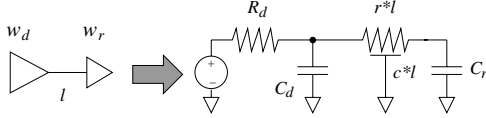


Figure 1: Repeater RC model.

The model for repeater and interconnect in a single stage is presented in Figure 1. The output resistance R_d and capacitance C_d of the driver are computed as R_o/w_d and $C_o * w_d$, respectively, where w_d is the driver width and R_o and C_o are the output resistance and capacitance per unit width. The interconnect is modeled as a distributed RC line with r and c being the resistance and capacitance per unit length. The interconnect length is l . The load capacitance C_r is calculated as $C_i * w_r$, where w_r is the receiving buffer width and C_i is the input capacitance per unit width. The delay of this stage τ can be computed as follows [3]:

$$\tau = \log_e 2 (R_d C_d + (R_d + r l) C_r + R_d c l + \frac{1}{2} r c l^2). \quad (1)$$

Repeater power dissipation includes switching power, leakage power, and short-circuit power. For the single stage as shown in Figure 1, the total power consumption P_{stage} is computed as follows [3]:

$$P_{stage} = k_1 (C_d + C_r + c l) + k_2 w_d + k_3 w_d \tau, \quad (2)$$

where k_1 , k_2 , and k_3 are constants that are determined by the given technology and design specifications:

$$\begin{aligned} k_1 &= \alpha V_{DD}^2 f_{clk}, \\ k_2 &= \frac{3}{2} I_{off_n} W_{nmin}, \\ k_3 &= \alpha V_{DD} W_{nmin} I_{short-circuit} f_{clk} \log_e 3 / \log_e 2. \end{aligned}$$

In the definition of k_1 , k_2 , and k_3 , α is the signal activity, V_{DD} is the power supply, f_{clk} is the clock frequency, W_{nmin} is the NMOS transistor width of an minimum-size inverter, I_{off_n} and $I_{short-circuit}$ are the leakage and short-circuit current per unit transistor width.

3. PROBLEM FORMULATION

Figure 2 shows our global interconnect structure. The driver is a buffer *within* the circuit block that sends data onto the global interconnect. The load is a buffer *within* the receiver block and can be modeled as a capacitor. The sizes of the driver and load are pre-designed, equal to w_d and w_r , respectively, and cannot be changed. The widths of all repeaters are equal, set to w . The distance between the driver and the first repeater is l_1 and the distance between the

last repeater and the load is l_2 . The distance between two adjacent repeaters is $l = (L - l_1 - l_2)/(n - 1)$, where n is the number of repeaters, and L is the total wire length.

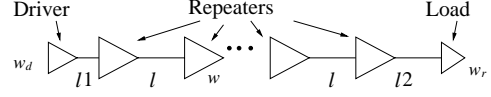


Figure 2: Interconnect model.

In the actual system layout, if the assigned location of a repeater R is occupied by other logic cells, R must be placed at some other location. We define *repeater location flexibility* (RLF) f as the maximum distance that every repeater along a global interconnect can deviate from its assigned location without causing any timing violation. A global interconnect should be designed to allow a positive RLF. To model the performance degradation due to RLF, we use the structure as shown in Figure 3 to compute the signal delay and power dissipation of the interconnect, in which each repeater is moved as far as possible and the moving directions of adjacent repeaters are opposite.

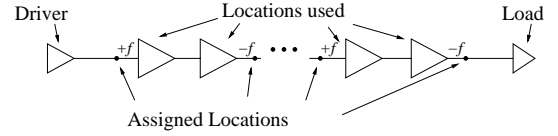


Figure 3: Repeater insertion with a flexibility f .

We assume that the number of repeaters is even so that signals are not inverted during transmission. Therefore, besides the first and last stage, there are $n/2$ stages with interconnect length $l - 2f$ and $n/2 - 1$ stages with interconnect length $l + 2f$. The signal delay τ_{total} of the entire global interconnect can be calculated as:

$$\tau_{total} = \tau_{1st} + \frac{n}{2} \tau_{l-2f} + (\frac{n}{2} - 1) \tau_{l+2f} + \tau_{last}, \quad (3)$$

where

$$\begin{aligned} \tau_{1st} &= \log_2 e (R_o C_o + (R_o/w_d + r(l + f)) C_i w \\ &\quad + R_o c (l + f)/w_d + \frac{1}{2} r c (l + f)^2), \\ \tau_{l-2f} &= \log_2 e (R_o C_o + (R_o/w + r(l - 2f)) C_i w \\ &\quad + R_o c (l - 2f)/w + \frac{1}{2} r c (l - 2f)^2), \\ \tau_{l+2f} &= \log_2 e (R_o C_o + (R_o/w + r(l + 2f)) C_i w \\ &\quad + R_o c (l + 2f)/w + \frac{1}{2} r c (l + 2f)^2), \\ \tau_{last} &= \log_2 e (R_o C_o + (R_o/w + r(l_2 + f)) C_i w_r \\ &\quad + R_o c (l_2 + f)/w + \frac{1}{2} r c (l_2 + f)^2). \end{aligned}$$

Similarly, the total power dissipation can be calculated as follows:

$$P_{total} = P_{1st} + \frac{n}{2} P_{l-2f} + (\frac{n}{2} - 1) P_{l+2f} + P_{last}, \quad (4)$$

where

$$\begin{aligned} P_{1st} &= k_1 (C_o w_d + C_i w + c(l + f)) + k_2 w_d + k_3 w_d \tau_{1st}, \\ P_{l-2f} &= k_1 (C_o w + C_i w + c(l - 2f)) + k_2 w + k_3 w \tau_{l-2f}, \\ P_{l+2f} &= k_1 (C_o w + C_i w + c(l + 2f)) + k_2 w + k_3 w \tau_{l+2f}, \\ P_{last} &= k_1 (C_o w + C_i w_r + c(l_2 + f)) + k_2 w + k_3 w \tau_{l_2+f}. \end{aligned}$$

The problem of *repeater insertion with discrete widths for low power* is formulated as follows:

Problem RIDW: Let L and D be the length and the timing budget of the interconnect, respectively, and let $W = \{w_1, w_2, \dots, w_k\}$ be possible widths of repeaters. Given a semiconductor technology $(R_o, C_o, C_i, V_{DD}, I_{off}, W_{nmin}, I_{short-circuit})$ and a global interconnect design specification $(c, r, \alpha, f_{clk}, w_d, w_r, f)$, compute an even integer number n , a width $w \in W$, and the locations $l1$ and $l2$ of the first and last repeaters, such that the interconnect delay $\tau_{total} \leq D$ and the power dissipation P_{total} is minimized.

4. HEURISTIC SOLUTION

Solving Problem RIDW is a challenging task because Equation (3) and Equation (4) are not linear and have four discrete or continuous unknowns, n , w , $l1$, and $l2$. In this section, we present our heuristic solver. Specifically, we first derive $l1$ and $l2$ based on empirical design knowledge. We then describe a 2-step scheme to compute n and w . We also argue that the computation of n and w is optimal.

4.1 Estimating $l1$ and $l2$

It is well known that, for long interconnects, the repeater size w is usually much larger than that of the interconnect driver w_d . Therefore, the optimal $l1$ is expected to be very small. In our heuristic, we set $l1$ to zero. In fact, in case of speed optimization, w can be 1–2 orders of magnitude larger than w_d . As a result, intermediate size repeaters should be inserted to reduce delay [3, 11]. When the timing budget is not tight and power minimization is the objective, however, the optimal w is only several times larger than w_d as shown in Section 6. Therefore, the insertion of intermediate size repeaters is not necessary.

We next derive the estimate for $l2$. In repeater insertion for low power, compared to delay minimization, the segment length l between repeaters is large and the repeater width w is small [3, 7]. As a result, the interconnect capacitance cl is much larger than the gate capacitance difference $c_o w - c_o w_r$. If $c_o w - c_o w_r$ is ignored, the last stage should be designed similarly to the stages between repeaters. Consequently, we set $l2 = l$, which is equal to L/n , since $l1 = 0$.

After the derivation of $l1$ and $l2$, the target locations of the repeaters can be determined by L and n . It is worth mentioning that, since we take into account of a positive RLF f in the computation of τ_{total} and P_{total} , the actual positions of the repeaters can deviate from their target places by up to f .

4.2 Solving for w and n

With $l1 = 0$ and $l2 = l = L/n$, Equation (3) and Equation (4) can be rewritten as

$$\tau_{total} = \bar{\alpha} \cdot (w, 1/w, n, 1/n, w/n, 1)^t, \quad (5)$$

$$P_{total} = \bar{\beta} \cdot (w, w^2, w/n, w^2/n, wn, 1)^t, \quad (6)$$

where $\bar{\alpha}$ and $\bar{\beta}$ are constant vectors whose elements are solely determined by the problem formulation. Although Equation (5) and Equation (6) are much simpler than Equation (3) and Equation (4), solving for n and w so that $\tau_{total} \leq D$ and P_{total} is minimized remains challenging due to the non-linear form of Equation (5) and Equation (6) and the discreteness of n and w . We next present an effective and efficient searching scheme to solve this problem.

Our search procedure first calculates the starting point in the 2-dimensional space defined by n and w , assuming that both variables are continuous. Figure 4 shows the solution space with contours of delay and power. Each circle c represents the solutions with a

constant delay D_c . The solutions inside c have lower delay values than D_c . Each line represents the solutions with the same power dissipation, with the lines close to the bottom-left corner having lower values. It is clear that, for continuous n and w , the optimal solution under a given timing budget D is the point of tangency on the corresponding circle, which satisfies

$$\tau_{total} = D. \quad (7)$$

We can then use Lagrange's method to calculate the continuous optimal solution. We define $F = P_{total} + \gamma(\tau_{total} - D)$. We set $\partial F / \partial n = 0$ and $\partial F / \partial w = 0$ and eliminate γ , resulting in the following equation:

$$0 = \bar{\theta} \cdot (n^4 w, n^3 w^2, n^3 w, n^3, n^2 w^2, n^2 w, n w^3, n w^2, n w, n, w^3, w^2, w)^t, \quad (8)$$

where $\bar{\theta}$ is a constant vector whose elements are solely determined by the problem formulation. Equations (7) and (8) can be numerically solved together to derive the optimal continuous solution (n_c, w_c) for the minimum power under the timing constraint D .

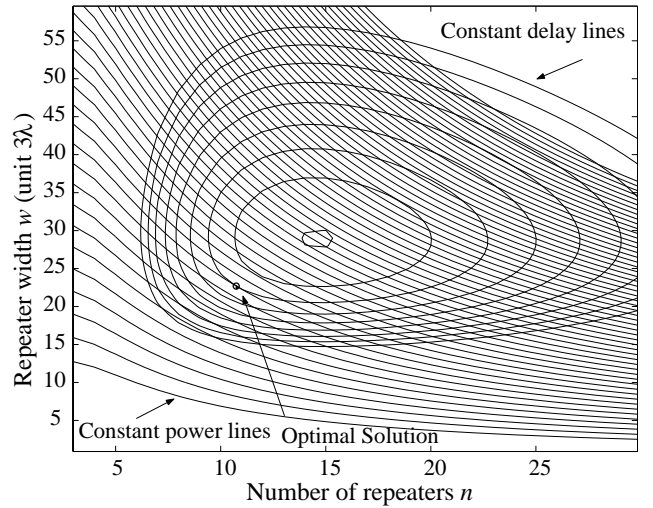


Figure 4: Delay and power contours.

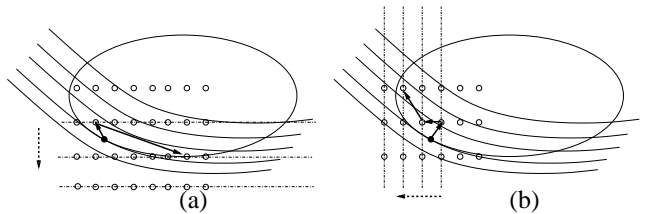


Figure 5: Solution search example.

The ensuing search procedure can be explained using Figure 5. The circle represents the solution contour of the given timing budget D , and the lines represent constant power contours. The solid dot represents the continuous optimal solution (n_c, w_c) from Lagrange's method. The hollow dots represent the discrete solution candidates. Two search paths made of discrete solutions are traversed sequentially. Figure 5(a) shows the downward search. Starting from the repeater width just above (n_c, w_c) , a horizontal line moves downward, crossing the solutions with the same repeater width. (If no feasible discrete solution above (n_c, w_c) exists, the starting location is the largest repeater width.) For all the solution candidates on each line within the timing circle, only the leftmost candidate is searched. The solid arrows point to the points on the

```

SEARCH( $n_c, w_c, D$ )
1  $P_{min} = \infty, (n_{opt}, w_{opt}) = \text{unknowns}$ 
2  $w_{cur} = \max\{w_i \in W \mid w_i \geq w_c\}$  ?
    $\min\{w_i \mid w_i \geq w_c, w_i \in W\} : \max\{w_i \in W\}$ 
3  $n_{cur} = \min\{2k \mid 2k \geq n_c, k \text{ is an integer}\}$ 
4 while(line  $w = w_{cur}$  intersects circle  $\tau = D$ )
5    $n_{left} = \text{Left}(\text{segment } w = w_{cur} \text{ within circle } \tau = D)$ 
6    $n_{right} = \text{Right}(\text{segment } w = w_{cur} \text{ within circle } \tau = D)$ 
7   if ( $\exists k, n_{left} \leq 2k \leq n_{right}, k \text{ is an integer}$ )
8      $\text{update\_store}(\text{the most left discrete solution on the segment})$ 
9    $w_{cur} = \max\{w_i \mid w_i < w_{cur}, w_i \in W\}$ 
10 while(line  $n = n_{cur}$  intersects circle  $\tau = D$ )
11    $w_{bot} = \text{Bottom}(\text{segment } n = n_{cur} \text{ within circle } \tau = D)$ 
12    $w_{top} = \text{Top}(\text{segment } n = n_{cur} \text{ within circle } \tau = D)$ 
13   if ( $\exists w_i, w_i \in W, w_{bot} \leq w_i \leq w_{top}$ )
14      $\text{update\_store}(\text{the lowest discrete solution on the segment})$ 
15    $n_{cur} = n_{cur} - 2$ 
16 if ( $P_{min} \neq \infty$ )
17   return ( $P_{min}, n_{opt}, w_{opt}$ )
18 return false

```

Figure 6: Our search scheme.

```

UPDATE_STORE( $n_{input}, w_{input}$ )
1 if ( $P_{min} > \text{Power}(n_{input}, w_{input})$ )
2    $P_{min} = \text{Power}(n_{input}, w_{input})$ 
3    $(n_{opt}, w_{opt}) = (n_{input}, w_{input})$ 

```

Figure 7: Subroutine `update_store`.

search path. The search terminates when the next line does not intersect the timing circle any more. Figure 5(b) shows the second part of the search. Starting from the solution column *just* on the right of (n_c, w_c) , a vertical line sweeps to the left until no intersection is possible. The search path, marked by the solid arrows, consists of the lowest discrete solutions on the line segments within the timing circle. The final solution is chosen as the one that is on either path and dissipates the least amount of power.

The pseudocode of our search scheme is described in Figure 6. It calculates the optimal discrete solution (n_{opt}, w_{opt}) and the corresponding minimal power dissipation P_{min} for a given timing budget D . Our search first initializes P_{min} , n_{opt} , and w_{opt} in line 1. It then sets the current search locations, w_{cur} and n_{cur} , in lines 2–3. Lines 4–9 describe the downward search. During each iteration, a new solution candidate on the path is derived and evaluated using the subroutine `update_store`, which calculates the current best solution as shown in Figure 7. The search toward the left is given in lines 10–15, in which solutions on the path are also used to update the current best solution. After both paths are searched, the best discrete solution is returned in line 17 or no-solution is reported.

The optimality of the solution computed by our search scheme can be proved in a straightforward manner. For two solutions on the same horizontal line, the left solution always dissipates less. Similarly, for two solutions on the same vertical line, the lower one always consumes less power. Consequently, we only use the leftmost or the lowest solution candidate on each line segment to update P_{min} . The detailed proof is omitted due to space limitations.

Although we assume n is an even integer to avoid signal inversion, our scheme can be easily extended to handle the case in which n can be any integer. The only modifications are changing line 3 in Figure 6 to $n_{cur} = \lceil n_c \rceil$ and reducing the moving step size of n_{cur} from 2 to 1. We will discuss the impact of limiting n to even numbers on power dissipation in Section 7.

```

RIS( $D, L$ )
1  $(n_c, w_c) = \text{LagrangeSolver}(D, L)$ 
2 if ( $n_c < 0$ )
3   return false
4 return Search( $n_c, w_c, D$ )

```

Figure 8: Algorithm RIS.

4.3 Heuristic summary

The pseudocode of our repeater insertion solver RIS is given in Figure 8. In line 1, we derive the continuous optimal solution n_c and w_c using Lagrange’s method in conjunction with a numerical solver. If this step fails to compute a solution that meets the timing requirement D , RIS returns false in line 3. Otherwise RIS performs the local search in line 4 and returns the discrete optimal solution or reports no solution exists.

The runtime complexity of RIS is $O(n_D + w_D)$, where n_D is the number of various n within the timing circle and w_D is the number of different w within the timing circle, since each iteration in *Search* is performed in $O(1)$ time and solving two polynomial functions numerically in *LagrangeSolver* takes constant time. For a given interconnect, the values of n_D and w_D increase with D , since the size of the timing contour increases with loose timing constraints. In practice, however, RIS usually finishes in constant time. Such a high efficiency is due to the fact that the starting point of our search (n_c, w_c) is a bottom-left tangent point of the timing circle, and we only search downward and leftward from it.

5. PERFORMANCE EVALUATION

This section evaluates the performance of Algorithm RIS. Several repeater libraries have been used in our experiments. The size of a repeater is represented by the width of its NMOS transistor. The width of its PMOS transistor is always set twice as large as that of the NMOS transistor. All repeater libraries contain the minimum size repeater, whose NMOS transistor size is 3λ . For a library with a repeater size granularity g , it contains repeaters of size $\{3\lambda, 3\lambda + g, 3\lambda + 2g, \dots\}$.

Several global interconnects in 0.18 μm technology with lengths ranging from 7 mm to 10 mm were designed. The delay target was set to 1.52 ns and the RLF was 400 μm . The driver width w_d was set to 4 times as wide as the minimum size repeater. The load was chosen as a minimum size repeater. The interconnect was chosen to be copper wire on the metal 4 layer. The circuit parameters of repeaters and interconnects, e.g., unit length wire capacitance and unit width gate capacitance, were extracted from a TSMC technology file and calibrated using HSPICE simulations.

To demonstrate the effectiveness of our approach, we have compared Algorithm RIS with a scheme which rounds the continuous estimate to a close discrete solution. Specifically, this rounding approach checks the four discrete solution candidates around (n_c, w_c) , as shown in Figure 9, and picks the solution that satisfies the timing budget D and dissipates the least power. Note that this rounding method still has to use our Lagrange’s method based solver to calculate (n_c, w_c) .

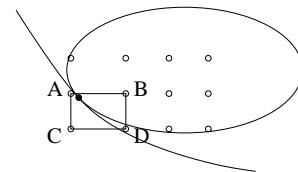


Figure 9: Rounding approach.

Figure 10 shows the comparison results with repeater granularity g being 1λ and 8λ . As expected, RIS outperforms the rounding scheme in all cases. Table 1 gives the average and maximum percentages of power reduction $(P_{round} - P_{RIS})/P_{round}$ using libraries with granularities from 1λ to 32λ . Although in many cases the power savings it achieves are comparable with those of the rounding scheme, Algorithm RIS achieves substantial power savings in certain cases. For example, when granularity is 32λ , RIS achieves up to 23% power reduction over rounding. Furthermore, in a number of cases, RIS finds the optimal solution, whereas the rounding scheme fails to yield any valid solution at all. The runtime for a single interconnect design is less than 0.12 seconds on a 2.2 GHz PC with 1GB memory.

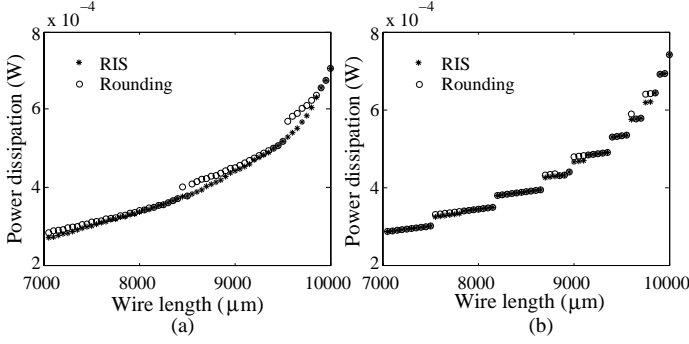


Figure 10: Comparison between RIS and rounding approach, with the repeater width granularity (a) 1λ and (b) 8λ .

Table 1: Power improvement of RIS

Repeater Width Granularity (λ)	1	2	4	8	16	32
Mean Power Reduction (%)	3	2	1	1	1	2
Max Power Reduction (%)	8	6	6	4	8	23

6. REPEATER LIBRARY DESIGN

In this section, we investigate the problem of repeater library design by deriving the proper granularity and range of the repeater widths.

6.1 Repeater width granularity

We used the same experimental setup as the one in Section 5. We applied RIS to design global interconnects of various lengths and with different timing budgets, using repeater libraries with width granularity g ranging from 1λ to 32λ .

The design results are shown in Figure 11, in which the solid line represents the ideal case results, i.e., continuous repeater width. As can be seen, power degradation deteriorates with the increase of g . The average and maximum power increase percentages with respect to the ideal case are given in Table 2. This result shows that, although a coarse granularity of 32λ can increase power dissipation by up to 28%, changing g from 1λ to 8λ increases power dissipation by only 3% on average. Such a granularity increase, however, reduces the library size by 87.5%, resulting in a significant reduction of the repeater library design time and cost.

Table 2: Power increase due to repeater size granularities

Repeater Width Granularity (λ)	1	2	4	8	16	32
Mean Power Increase (%)	1	2	3	4	6	12
Max Power Increase (%)	4	4	5	9	16	28

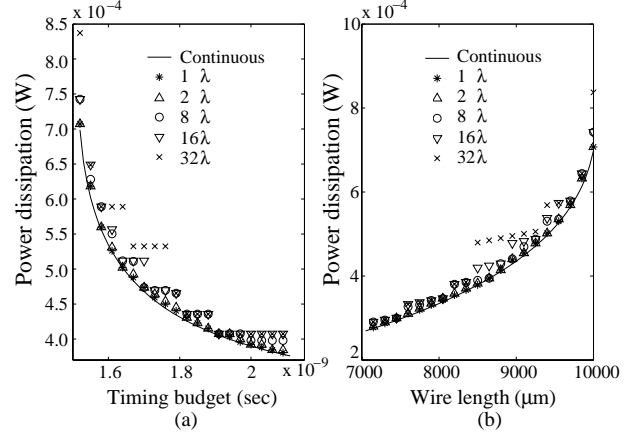


Figure 11: Minimum power of interconnect designs using repeater libraries of various granularities. (a) $L=10$ mm (b) $D=1.52$ ns.

6.2 Repeater width range

This section analyzes another key parameter of repeater library design: the repeater width range. Figure 12 shows the optimal discrete repeater widths from the experimental results in Subsection 6.1, using repeater libraries with 1λ and 8λ granularities. The range of repeater widths is quite limited, from 20λ to 80λ . Repeaters with sizes outside of this range do not need to be included in the repeater library. From Figure 12, if the granularity is set to 8λ , only 8 repeaters need to be designed, resulting in a very low design cost.

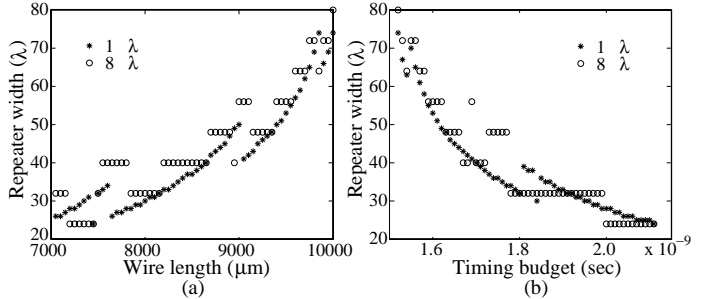


Figure 12: Optimal repeater width. (a) $D=1.52$ ns (b) $L=10$ mm.

It is important to point out that the optimal granularity and range of repeater widths will change for various technologies, fabrication processes, and interconnect specifications. Our analysis can thus be performed using the corresponding device and design parameters to derive the new granularity and range of the repeater widths.

7. INTERCONNECT DESIGN ISSUES

In this paper, we assume the number of repeaters to be even so that signals are not inverted during transmission. Although this constraint facilitates the system design, we need to analyze whether it substantially affects the power dissipation of the interconnect designs. We performed our experiments twice, with the granularity of n set to 2 and 1, corresponding to the cases with and without the even repeater number constraint, respectively. The results are shown in Figure 13. Interestingly, the average power dissipation difference between the two cases is below 2% on the average, indicating that limiting n to even numbers is acceptable. (The design scheme using non-inverting repeaters, which are cascaded inverters, is not analyzed, since it is inferior to the approach using simple inverters [11].)

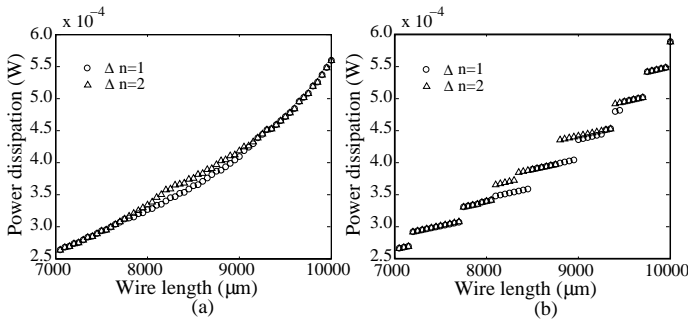


Figure 13: Power degradation due to the even-repeater-number constraint. Repeater granularity is (a) 1λ (b) 8λ .

We subsequently investigate the power impact of RLF. Figure 14 shows the results of a 10-mm interconnect design, in which the repeater width granularity is set to 1λ . For large RLF, tight timing budgets cannot be satisfied and, therefore, their power dissipation values are not shown. From Figure 14, the power dissipation increases substantially with the increase of RLF, especially under tight timing budgets. As a result, repeater planning for RLF reduction is crucial for low-power and high performance system designs.

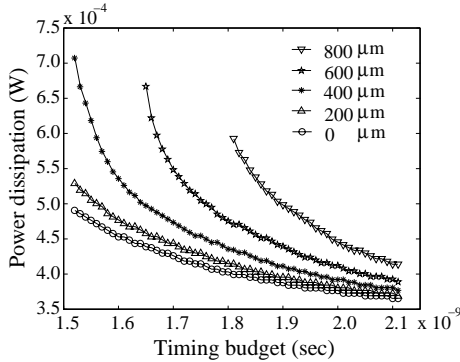


Figure 14: Power effect of repeater location flexibility.

8. CONCLUSION

In this paper, we present a novel and efficient procedure for low-power repeater insertion for global interconnects that captures detailed physical and design issues such as short-circuit power and repeater location flexibility. Our approach considers the discreteness of the repeater number and width. Specifically, starting from a solution estimate derived by Lagrange's method, our algorithm performs a search for the optimal solution in the discrete solution space. This search runs extremely fast and guarantees to reach the optimal discrete solution. Contrary to the repeater insertion schemes that assume the continuous repeater width, our scheme only derives the repeater sizes that are available in the repeater library and, therefore, does not suffer from performance loss due to solution rounding. Experimental results validate the high effectiveness of our scheme. Our repeater insertion solver can achieve up to 23% power reduction in comparison to the schemes that rely on rounding. Furthermore, for some design cases, our scheme succeeds in finding the optimum, whereas the rounding based schemes fail to find any solution that meets the timing requirement.

The second contribution of this paper is the analysis of key issues in repeater library design, leading to the answer of what sizes of repeaters should be included in a repeater library. Using our repeater insertion solver RIS, we can derive the proper granularity and range of the repeater size and, therefore, design a low-cost

small-size repeater library without significant performance degradation. We have shown that, by choosing the right granularity, the size of the repeater library can be reduced by 87% with only less than 4% average power increase for interconnect designs. Furthermore, on the interconnect design aspect, we have discovered that no significant power dissipation increase is caused by limiting the repeater number to be even, whereas large repeater location flexibility can lead to a substantial power increase.

9. REFERENCES

- [1] V. Adler and E. G. Friedman. Repeater design to reduce delay and power in resistive interconnect. *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, 45(5):607–616, May 1998.
- [2] H. B. Bakoglu. *Circuits, Interconnects, and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.
- [3] K. Banerjee and A. Mehrotra. A power-optimal repeater insertion methodology for global interconnects in nanometer designs. *IEEE Trans. VLSI Systems*, 49(11):2001–2007, Nov. 2002.
- [4] C.-C. N. Chu and D. F. Wong. Closed form solution to simultaneous buffer insertion/sizing and wire sizing. In *Inter. Symp. on Physical Design*, Apr. 1997.
- [5] 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.
- [6] J. Cong, T. Kong, and D. Z. Pan. Buffer block planning for interconnect-Driven floorplanning. In *Inter. Conf. on Computer-Aided Design*, Nov. 1999.
- [7] G. S. Garcea, N. P. van der Meijs, and R. H. Otten. Simultaneous analytical area and power optimization for repeater insertion. In *Inter. Conf. on Computer-Aided Design*, Nov. 2003.
- [8] M. Kang, W. Dai, T. Dillinger, and D. LaPotin. Delay bounded buffered tree construction for timing driven floorplanning. In *Inter. Conf. on Computer-Aided Design*, Nov. 1997.
- [9] P. Kapur, G. Chandra, and K. C. Saraswat. Power estimation in global interconnect and its reduction using a novel repeater optimization methodology. In *Design Automation Conference*, June 2002.
- [10] R. Li, D. Zhou, J. Liu, and X. Zeng. Power-optimal simultaneous buffer insertion/sizing and wire sizing. In *Inter. Conf. on Computer-Aided Design*, Nov. 2003.
- [11] W. Liao and L. He. Full-chip interconnect power estimation and simulation considering concurrent repeater and flip-flop insertion. In *Inter. Conf. on Computer-Aided Design*, Nov. 2003.
- [12] J. Lillis, C. K. Cheng, and T.-T. Y. Lin. Optimal wire sizing and buffer insertion for low power and a generalized delay model. In *Inter. Conf. on Computer-Aided Design*, Nov. 1995.
- [13] H. Zhou, D. F. Wong, I. M. Liu and A. Aziz. Simultaneous routing and buffer insertion with restrictions on buffer locations. *IEEE Trans. Computer Aided Design of Integrated Circuits and System*, 19(7):819–824, 2000.
- [14] A. Nalamalpu and W. P. Burleson. Repeater insertion in deep sub-micron CMOS: Ramp-based analytical model and placement sensitivity analysis. In *International Symposium on Circuits and Systems*, May 2000.
- [15] A. Nalamalpu and W. P. Burleson. A practical approach to DSM repeater insertion: Satisfying delay constraints while minimizing area and power. In *IEEE International ASIC/SOC Conference*, Sept. 2001.
- [16] R. Otten. Global wires harmful? In *Inter. Symp. on Physical Design*, Apr. 1998.
- [17] D. Sylvester and K. Keutzer. Getting to the bottom of deep submicron. In *Inter. Conf. on Computer-Aided Design*, Nov. 1998.
- [18] D. Sylvester and K. Keutzer. A global wiring paradigm for deep submicron design. *IEEE Trans. CAD*, 19(2):242–252, Feb. 2000.
- [19] L. P. P. van Ginneken. Buffer placement in distributed rc-tree networks for minimal elmore delay. In *Proc. Intl. Symposium on Circuits and Systems*, 1990.
- [20] C. Y. Wu and M. Shiao. Delay models and speed improvement techniques for RC tree interconnections among small geometry CMOS VLSI. *Journal of Solid-State Circuits*, 25(10):1247–1256, Oct. 1990.