Optimal Design of High Fan-In Multiplexers via Mixed-Integer Nonlinear Programming

Hsu-Wei Huang, Cheng-Yeh Wang and Jing-Yang Jou
Department of Electronics Engineering, National Chiao Tung University
Hsinchu 300, Taiwan, R.O.C.

Abstract - In this paper, a novel strategy for designing the heterogeneous-tree multiplexer is proposed. We build the multiplexer delay model by curve fitting and then formulate the heterogeneous-tree multiplexer design problem as a special type of optimization problem called mixed-integer nonlinear programming (MINLP). A new design parameter, the switch size in each stage, is introduced to improve the speed of the heterogeneous-tree multiplexer. The proposed strategy can determine the multiplexer architecture and the switch size in each stage simultaneously. Three optimization methods are provided to synthesize the heterogeneous-tree multiplexer according to the design specifications.

I. Introduction

High fan-in multiplexers are used extensively in many applications, including the column decoders of memories [6] and the resistor-chain digital-to-analog converters [7]. Traditionally, high fan-in multiplexers have been designed using the uniform approach [9], in which all switches are interconnected with the output node in common, or the binary-tree structure [10], which is a tree-like multistage structure. However, a more general architecture called the heterogeneous-tree multiplexer was recently proposed [1]. This new architecture has better speed performance than the traditional architectures. Design strategies based on different kinds of switch topology, such as transmission gate and tri-state buffer, are also proposed in [1, 2, 3].

This paper will introduce a new method that has several important advantages over previous methods. The delay formula of the heterogeneous-tree multiplexer is built by curve fitting, and then the heterogeneous-tree multiplexer design problem is formulated as a special type of optimization problem called mixed-integer nonlinear programming (MINLP). The important advantage of mixed-integer nonlinear programming is that the integer constraints can be set so that the design parameters of the heterogeneous-tree multiplexer can be determined properly. Reasonable design parameters can always be obtained directly and no rounding is required. Another feature of the proposed approach is that the proper multiplexer architecture and the switch size in each stage can be determined simultaneously, useful in reducing the multiplexer delay. In [2], only the multiplexer architecture that minimizes the multiplexer delay could be determined. In this work, we provide three optimization methods - delay minimization, delay minimization under area constraint, and area minimization under delay constraint - to help the designers to determine the multiplexer architecture and the switch size in each stage according to the design specifications.

The rest of this paper is organized as follows. Section II defines the problem. Section III describes our design strategies. Section IV presents the experimental results. Section V draws the conclusions.

II. Problem Formulation

The structure of the heterogeneous-tree multiplexer is shown in Fig. 1. The multiplexer has N inputs and is partitioned into k stages. In the i-th stage, the switches are combined into groups of S_i elements with a common output node, which represents an input to the subsequent stage. This paper considers the case in which the switches are implemented using tri-state buffer (Fig. 2).

The first stage includes N/S_1 groups of tri-state buffers and N/S_1 outputs. The second stage includes N/S_1S_2 outputs. The k-th stage includes $N/S_1S_2\ldots S_k$ outputs. The last stage has a single output that represents the multiplexer output, so the following constraint can be derived.

$$\prod_{i=1}^{k} S_i = N \tag{1}$$

In this paper, we will assume that the delay of the address decoder of the multiplexer can be neglected. The delay of the address decoder can be reduced by careful circuit design, for example, using superbuffer or BiCMOS driver [9, 10].

The delay of the heterogeneous-tree multiplexer depends strongly on the number of stages and the number of switches per group in each stage, so the problem as defined in the previous approaches [1, 2] is to determine the optimal number of stages and the optimal number of switches per group in each stage such that the multiplexer delay is minimal. In [1, 2, 3], the switch size is kept constant and assumed to be minimal. Here, another parameter - the switch size in each stage – is introduced. This work will consider the multiplexer architecture, which includes the number of stages and the number of switches per group in each stage, at the same time as the switch size in each stage. We define our problem as follow. The problem we want to solve is to determine the multiplexer architecture and the switch size in each stage such that the multiplexer meets the design specifications.

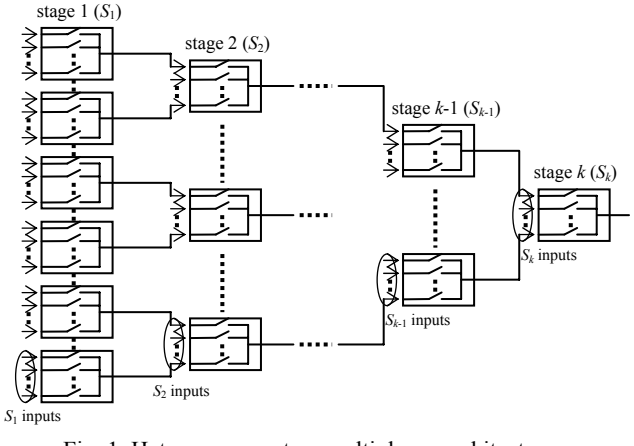

Fig. 1. Heterogeneous-tree multiplexer architecture

III. Proposed Design Strategy

This section introduces the proposed design strategy. The delay model to be used is first considered and then the concept of mixed-integer nonlinear programming is introduced. Next, the path delay model and the load capacitance model in each stage of the heterogeneous-tree multiplexer are constructed, and then the delay model of heterogeneous-tree multiplexer is derived. Finally, the objective functions and the constraint functions of the mixed-integer nonlinear programming for the design problem of the heterogeneous-tree multiplexer are proposed.

It is easy to show by exhaustive search that there exists a stage number k such that the heterogeneous-tree multiplexer meets the design specifications. Since k is a small integer, the proposed strategy will find the optimal value of k by exhaustive search. That is, every possible value of k will be tried in solving MINLP and then the optimal value of k will be selected such that the heterogeneous-tree multiplexer meets the design specifications.

A. Delay Modeling Approach

The delay model applied in this paper is the Convex Delay Model proposed in [4, 5]. This model allows us to capture the effects of varying the transistor sizes on the gate delay. The general form of the delay model is given by

$$Delay = \sum_{j=1}^{m} P_j \cdot \prod_{i=1}^{n} \left(x_i^{\triangle} + c_{ij}\right)^{\beta_{ij}} + C. \tag{2}$$

Here, the x_i's are characterization variables, and the c_{ij}'s, β_{ij}'s, C, and P_j's are referred to collectively as characterization constants. m is the parameter used to increase the characterization flexibility, and n is the number of characterization variables. The parameter $\triangle$ is set to either –1 or 1, depending on the variable. The problem of characterization is to determine appropriate values for the characterization constants by curve fitting.

B. Mixed-Integer Nonlinear Programming

A mixed-integer nonlinear programming is an optimization problem of the following form.

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & c_{Li} \leqq f_i(x) \leqq c_{Ui},\ i = 1, 2, \ldots, m \\ & b_{Lj} \leqq g_j(x) \leqq b_{Uj},\ j = 1, 2, \ldots, p \\ & x_{Lk} \leqq x_k \leqq x_{Uk},\ k = 1, 2, \ldots, n \\ & x_e \in Z,\ x_e \subseteq x_k \end{aligned}$$

where $f_0(x)$ is the objective function, $f_i(x)$ are the nonlinear constraint functions, $g_j(x)$ are the linear constraint functions, x_k are the variables, c_{Ui}, b_{Uj}, and x_{Uk} are the upper bounds, and c_{Li}, b_{Lj}, and x_{Lk} are the lower bounds.

C. Design Strategy I

This section presents the design strategy I, which only determines the multiplexer architecture and keeps the switch size minimal.

C.1 Delay of Heterogeneous-Tree Multiplexer

The switch delay model associated with design strategy I is defined as follows. As the equation shows, the switch delay depends only on the load capacitance of the switch.

$$Switch\ Delay = \sum_{j=1}^{m} P_j \cdot \left(C + c_j\right)^{\beta_j} + q, \tag{3}$$

where C is the load capacitance of the switch, and P_j, c_j, β_j and q are the characterization constants.

The path delay of the multiplexer is the sum of the switch delay in each stage of the heterogeneous-tree multiplexer, so the path delay model of the heterogeneous-tree multiplexer can be obtained as follows.

$$Path\ Delay = \sum_{i=1}^{k} \left[\sum_{j=1}^{m} P_j \cdot \left(C_i + c_j\right)^{\beta_j} + q \right], \tag{4}$$

where C_i is the load capacitance of stage i, and P_j, c_j, β_j and q are the characterization constants.

The switch associated with stage i drives the output capacitance C_{out} of the other (S_i-1) OFF switches that belong to the same group, as well as the input capacitance of the switch in the following stage, C_{in}. (See Fig. 3) Hence, the load capacitance of stage i is defined as

$$C_i = \left(S_i - 1\right) \cdot C_{out} + C_{in}, \tag{5}$$

where C_{out} is the output capacitance of the switch, and C_{in} is the input capacitance of the switch.

C.2 Design Optimization

This section presents the objective functions and constraint functions of mixed-integer nonlinear programming for three different optimization methods.

Substituting Eq. (5) into Eq. (4) yields the multiplexer delay equation as follows.

$$\sum_{i=1}^{k} \left[\sum_{j=1}^{m} P_j \cdot \left(\left(\left(S_i - 1\right) \cdot C_{out} + C_{in} \right) + c_j \right)^{\beta_j} + q \right] \tag{6}$$

The area of the multiplexer is defined as the number of switches in the multiplexer.

$$N + \left(\frac{N}{S_1}\right) + \left(\frac{N}{S_1 S_2}\right) + \cdots + \left(\frac{N}{S_1 S_2 \cdots S_{k-1}}\right) \tag{7}$$

The basic constraints are as follows.

$$S_i \geq 2 \ \ \& \ \ S_i \in Z \ \ , \ \ i = 1, 2, \ldots, k \tag{8}$$

Optimization Method Ⅰ: The multiplexer architecture whose delay is minimal can be obtained using this optimization method. The objective function of MINLP is defined by Eq. (6), and the constraint functions of MINLP are defined by Eqs. (1) and (8).

Optimization Method Ⅱ: The multiplexer architecture whose delay is minimal under area constraint can be obtained using this optimization method. The objective function of MINLP is defined by Eq. (6), and the constraint functions of MINLP are defined by Eqs. (1), (7) and (8).

Optimization Method Ⅲ: The multiplexer architecture whose area is minimal under delay constraint can be obtained using this optimization method. The objective function of MINLP is defined by Eq. (7), and the constraint functions of MINLP are defined by Eqs. (1), (6) and (8).

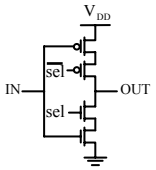

Fig. 2. Tri-state buffer

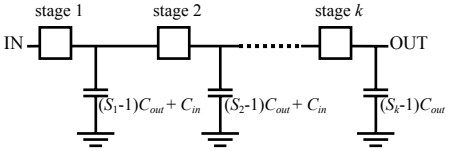

Fig. 3. Critical path model associated with design strategy I

D. Design Strategy II

This section presents the design strategy II, which determines the multiplexer architecture and the switch size in each stage simultaneously. The switch size can be different in each stage of the heterogeneous-tree multiplexer.

D.1 Delay of Heterogeneous-Tree Multiplexer

The switch delay model associated with design strategy II is defined as follow. As the equation shows, the switch delay depends on both the load capacitance and the size of the switch.

$$\textit{Switch Delay} = \sum_{j=1}^{m} P_j \cdot \left(W^{-1} + c_{1j}\right)^{\beta_{1j}} \left(C + c_{2j}\right)^{\beta_{2j}} + q, \tag{9}$$

where W is the NMOS transistor width of the switch, C is the load capacitance of the switch, and P_j, c_{1j}, β_{1j}, c_{2j}, β_{2j}, and q are the characterization constants.

The path delay model of the heterogeneous-tree multiplexer can be obtained as follows.

$$\textit{Path Delay} = \sum_{i=1}^{k} \left[\sum_{j=1}^{m} P_j \cdot \left(W_i^{-1} + c_{1j}\right)^{\beta_{1j}} \left(C_i + c_{2j}\right)^{\beta_{2j}} + q \right], \tag{10}$$

where W_i is the NMOS transistor width of the switch in stage i, C_i is the load capacitance of stage i, and P_j, c_{1j}, β_{1j}, c_{2j}, β_{2j}, and q are the characterization constants.

The load capacitance of stage i is defined as

$$C_i = \left(S_i - 1\right) \cdot C_{out_i} + C_{in_{i+1}}, \tag{11}$$

where C_{out_i} is the output capacitance of the switch in stage i, and $C_{in_{i+1}}$ is the input capacitance of the switch in stage i+1.

The gate capacitance and the source/drain capacitance of the MOSFET are proportional to the channel width of the MOSFET, so the input/output capacitance of the tri-state buffer is defined as follows.

$$C_{out_i} = a \cdot W_i + b, \tag{12}$$

$$C_{in_{i+1}} = c \cdot W_{i+1} + d, \tag{13}$$

where a, b, c, and d are the characterization constants, and W_i and W_{i+1} are the NMOS transistor width of switches in stages i and i+1, respectively.

D.2 Design Optimization

This section presents the objective functions and constraint functions of mixed-integer nonlinear programming associated with three different optimization methods.

Substituting Eq. (11) into Eq. (10) yields the following multiplexer delay equation.

$$\sum_{i=1}^{k} \left[\sum_{j=1}^{m} P_j \cdot \left(W_i^{-1} + c_{1j}\right)^{\beta_{1j}} \left(\left(\left(S_i - 1\right) \cdot \left(a \cdot W_i + b\right) + \left(c \cdot W_{i+1} + d\right)\right) + c_{2j}\right)^{\beta_{2j}} + q \right] \tag{14}$$

The area of the multiplexer is defined as the sum of the NMOS transistor widths of the switches in the multiplexer.

$$N \cdot W_1 + \left(\frac{N}{S_1}\right) \cdot W_2 + \left(\frac{N}{S_1 S_2}\right) \cdot W_3 + \cdots + \left(\frac{N}{S_1 S_2 \cdots S_{k-1}}\right) \cdot W_k \tag{15}$$

In the experiment, the range of NMOS transistor widths is 0.3 μm to 3 μm. In order to obtain reasonable design parameters, $10W_i$ must be a positive integer. So we can obtain the following constraints.

$$3 \geq W_i \geq 0.3 \;\; \& \;\; 10W_i \in Z \;\; , \;\; S_i \geq 2 \;\; \& \;\; S_i \in Z \;\; , \;\; i = 1, 2, \ldots, k \tag{16}$$

Optimization Method Ⅰ: This optimization method can determine the multiplexer architecture and the switch size in each stage such that the multiplexer delay is minimal. The objective function of MINLP is defined by Eq. (14), and the constraint functions of MINLP are defined by Eqs. (1) and (16).

Optimization Method Ⅱ: This optimization method can determine the multiplexer architecture and the switch size in each stage such that the multiplexer delay is minimal under area constraint. The objective function of MINLP is defined by Eq. (14), and the constraint functions of MINLP are defined by Eqs. (1), (15) and (16).

Optimization Method Ⅲ: This optimization method can determine the multiplexer architecture and the switch size in each stage such that the multiplexer area is minimal under delay constraint. The objective function of MINLP is defined by Eq. (15), and the constraint functions of MINLP are defined by Eqs. (1), (14) and (16).

IV. Experimental Results

This section presents the experimental results of the proposed approach. A 256-inputs multiplexer is designed and simulated using 0.18 μm CMOS process. In this experiment, the parameter m is set to 1. The output loading of the 256-inputs multiplexer is 0.003 pF. The experiment has two parts. In part A, the multiplexer is designed according to strategy I. In part B, the multiplexer is designed according to strategy II.

In the experiment, the curve fitting problem and the mixed-integer nonlinear programming problem are solved using MINOS [12] and Tomlab [11], respectively.

A. The Experimental Results of Design Strategy I

As the experimental results in Table I show, the predicted delay agrees well with the simulated result. The error rate between the estimated result and the HSPICE simulation result is less than 2%. Optimization method I yields the same result as obtained by the previous approach [2] although the technology adopted is different. That is, the multiplexer delay is minimal when k = 4 and $S_1 = S_2 = S_3 = S_4 = 4$. Using optimization method II, the multiplexer architecture that yields minimal delay under the area constraint of 290 switches is k = 3, S_1 = 16 and $S_2 = S_3 = 4$.

According to optimization method III, the multiplexer architecture that yields minimal area under the delay constraint of 0.65 *ns* is $k = 2$, $S_1 = S_2 = 16$. The experimental results reveal that the proposed approach can effectively determines the multiplexer architecture according to the design specifications.

B. The Experimental Results of Design Strategy II

References [2, 3] use only minimally sized switches and do not consider the transistor sizing problem. In this experiment, we try to show that the delay of the heterogeneous-tree multiplexer can be improved by considering the multiplexer architecture and the switch size in each stage simultaneously. As the experimental results presented in the Table II show, the minimal multiplexer delay obtained by the optimization method I of strategy II is 0.4527 *ns* and that obtained by the optimization method I of strategy I is 0.5278 *ns*. The minimal multiplexer delay can be reduced by almost 17% if the multiplexer architecture and the switch size in each stage are suitably selected using the optimization method I of strategy II. In comparison II, the multiplexer delay increases when the switch size in each stage is increased manually by a stage-ratio of three. Obviously, the heterogeneous-tree multiplexer delay will increase if the switch size in each stage is selected improperly. It is clearly that the proposed method can effectively reduce the multiplexer delay and has better performance than the approach proposed in [2].

As the experimental result in the Table III show, the predicted delay agrees well with the simulated result. The error rate between the estimated result and the HSPICE simulation result is less than 5%. In optimization method I, as desired, the minimal multiplexer delay can be obtained by determining the multiplexer architecture and the switch size in each stage simultaneously. Using optimization methods II and III, the multiplexer architecture and the switch size in each stage, meeting the design constraints, can be obtained.

V. Conclusions

In this paper, we propose a novel design strategy for designing a high fan-in heterogeneous-tree multiplexer. Curve fitting is used to build the multiplexer delay model, and then the heterogeneous-tree multiplexer design problem is formulated as a special type of optimization problem called mixed-integer nonlinear programming. The proposed design strategy can help designers to determine the multiplexer architecture and the switch size in each stage according to the design requirement, such as delay minimization, delay minimization under area constraint, or area minimization under delay constraint. The feature of mixed-integer nonlinear programming enables reasonable design parameters to be obtained directly, independently of the technology is used. Another important advantage is that the multiplexer delay can be improved by suitably selecting the multiplexer architecture and the switch size in each stage simultaneously.

Acknowledgements

This work was supported in part by the R.O.C. National Science Council under Grant No. NSC-91-2218-E-009-013.

TABLE I
Experimental Results of Design Strategy I

	Multiplexer Architecture	Constraint	Area	Delay (estimation)	Delay (HSPICE simulation)	Error
Optimization Method Ⅰ	S_i = (4, 4, 4, 4) W_i = (0.3, 0.3, 0.3, 0.3)	none	340 switches	0.5208 *ns*	0.5278 *ns*	1.3%
Optimization Method Ⅱ	S_i = (16, 4, 4) W_i = (0.3, 0.3, 0.3)	290 switches	276 switches	0.5717 *ns*	0.5797 *ns*	1.4%
Optimization Method Ⅲ	S_i = (16, 16) W_i = (0.3, 0.3)	0.65 *ns*	272 switches	0.6226 *ns*	0.6290 *ns*	1%

TABLE II
Speed Improvement associated with Design Strategy II

	Multiplexer Architecture	Constraint	Area	Delay (estimation)	Delay (HSPICE simulation)	Error
Optimization Method Ⅰ	S_i = (4 , 8 , 8) W_i = (3 , 1.2 , 0.6)	none	849.6 μm	0.4672 *ns*	0.4527 *ns* (improve 16.6%)	3.2%
Comparison Ⅰ	S_i = (4 , 4 , 4 , 4) W_i = (0.3, 0.3, 0.3, 0.3)	none	102 μm	0.5208 *ns*	0.5278 *ns*	1.3%
Comparison Ⅱ	S_i = (4 , 4 , 4 , 4) W_i = (0.3, 0.9, 2.7, 8.1)	none	210 μm	0.5912 *ns*	0.6065 *ns*	2.5%

TABLE III
Experimental Results of Design Strategy II

	Multiplexer Architecture	Constraint	Area	Delay (estimation)	Delay (HSPICE simulation)	Error
Optimization Method Ⅰ	S_i = (4 , 8 , 8) W_i = (3 , 1.2 , 0.6)	none	849.6 μm	0.4672 *ns*	0.4527 *ns*	3.2%
Optimization Method Ⅱ	S_i = (4 , 8 , 8) W_i = (1.3 , 0.8 , 0.6)	400 μm	388.8 μm	0.4743 *ns*	0.4796 *ns*	1.1%
Optimization Method Ⅲ	S_i = (16 , 16) W_i = (0.3 , 0.3)	0.65 *ns*	81.6 μm	0.6013 *ns*	0.6290 *ns*	4.4%

References

[1] Ming-Bo Lin, "On the Design of Fast Large Fan-In CMOS Multiplexers", *IEEE Transactions on Computer-Aided Design*, pp.963-967, 2000.

[2] M. Alioto and G. Palumbo, "Optimized Design of High Fan-In Multiplexers Using Switches with Driving Capability", *The 8th IEEE International Conference on Electronics, Circuits and Systems*, 2001.

[3] M. Alioto, G. Di Cataldo, and G. Palumbo, "Optimized Design of High Fan-In Multiplexers Using Tri-State Buffers", *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, pp.1500-1505, Oct. 2002.

[4] M. Ketkar, K. Kasamsetty, and S. Sapatnekar, " Convex Delay Models for Transistor Sizing", *37th Design Automation Conference*, pp.655-660, 2000.

[5] K. Kasamsetty, M. Ketkar, and S. Sapatnekar, "A New Class of Convex Functions for Delay Modeling and Its Application to the Transistor Sizing Problem", *IEEE Transactions on Computer-Aided Design*, pp.779-788, Jul. 2000.

[6] R. T. Howe and C. G. Sodini, *Microelectronics: An Integrated Approach*, Englewood Cliffs, NJ: Prentice-Hall, 1997.

[7] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*, New York: Wiley, 1997.

[8] J. Rabaey, *Digital Integrated Circuits (A Design Perspective)*, Prentice Hall, 1996.

[9] S. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits: Analysis and Design*, McGraw-Hill, 1996.

[10] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design: A System Perspective*, Addison-Wesley, 1993.

[11] Kenneth Holmstrom and Anders Goran, *User's Guide for Tomlab v4.0.5*, Tomlab Optimization Inc., February 28, 2003.

[12] Bruce A. Murtagh and Michael A. Saunders, *MINOS 5.5 User's Guide*, Report SOL 83-20R, Dept of Operations Research, Stanford University.