Effect of Wire Delay on the Design of Prefix Adders in Deep-Submicron Technology

Zhijun Huang and Miloš D. Ercegovac
Computer Science Department
University of California at Los Angeles
Los Angeles, CA 90095
{zjhuang, milos}@cs.ucla.edu

Abstract

This paper investigates the wire delay effect on the design of prefix adders when the technology moves from 250nm to 70nm. The simulation is based on parameters from NTRS'97 and uses an analytical wire delay model that considers the fanout effect and the distributive nature of wire capacitance and resistance.

Simulation results show that wire delay exceeds logic delay and dominates the critical path delay of prefix adders in many cases. For a given technology, the wire delay contribution increases steadily as the adder width increases. As the feature size decreases, however, the wire delay contribution decreases slowly. The simulation data also imply that there is little need to consider wire resistance. On the other hand, the effect of wire coupling capacitance plays a critical role in prefix adders' performance.

1. Introduction

Among various binary adder architectures, a large family of prefix adders is particularly attractive because all of them have minimum logic depth and very efficient implementations [8][9][14] (Brent-Kung prefix adder [2] is an exception since it allows the logic depth of the structure to increase). An important problem in prefix adder design is how to further improve performance since the prefix structure has already had the minimum logic depth. Among those factors affecting the performance, the fanout and wire length are often the deterministic factors. As VLSI technology moves into the deep sub-micron (DSM) domain, the wiring problem becomes even more significant and, in many cases, dominates in both area and delay optimization [5][6][13]. Recently, Choe and Swartzlander [3] showed that larger multipliers would suffer from wire delays if they were uniformly scaled down into deep sub-micron region. In the technology roadmap given by the semiconductor industry, however, the metal aspect ratio keeps increasing to avoid the problems in uniform scaling. It is interesting and meaningful to look into the wire effects on arithmetic circuits under practical technology scaling parameters.

Based on technology parameters on the roadmap NTRS'97 [13], this paper investigates the wire delay effect on the performance of prefix adders when the technology moves from 250nm to 70nm. Three structures have been selected to cover the range of the prefix adder family with minimum logic depth: Ladner-Fisher structure [10], Knowles structure [8] and Kogge-Stone structure [9]. The range of datapath width under consideration is from 16 bits to 128 bits. The simulation uses an analytical wire delay model that considers the fanout effect and the distributive nature of wire capacitance and resistance.

The rest of the paper is organized as follows. Section 2 formulates the general delay calculation in prefix adder structures. Section 3 introduces an analytical delay estimation model used for the wire delay analysis. Section 4 shows the simulation results and the implications. Conclusions are given in Section 5.

2. Critical path delay in prefix adders

Two key properties, associativity and idempotency, make the prefix formulation of addition very flexible. Even for prefix adders with the same minimum logic depth, different structures can be developed for various area and speed requirements [8][14]. In this study, three prefix structures with the same minimum depth are chosen: Ladner-Fisher adder, Knowles adder and Kogge-Stone adder. They spread over the range of area-speed tradeoffs in prefix adder design [8]. These adder architectures are shown in Figure 1, 2 and 3, respectively. All graphs are drawn in the traditional prefix description style: black nodes depict nodes performing computation logic and white nodes represent nodes with no logic or only some buffers. The functions of computation nodes are shown in Figure 4.

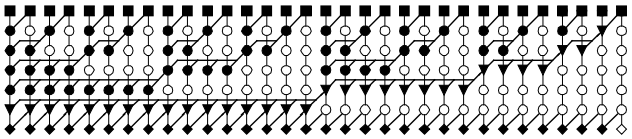


Figure 1. Ladner-Fisher adder

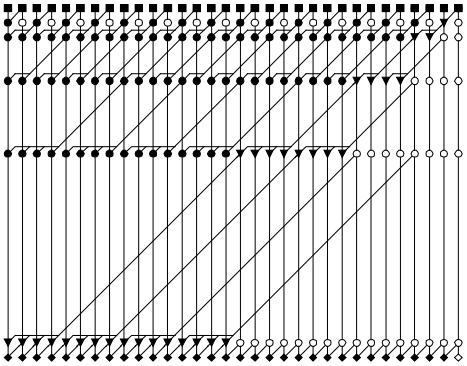


Figure 2. Knowles adder

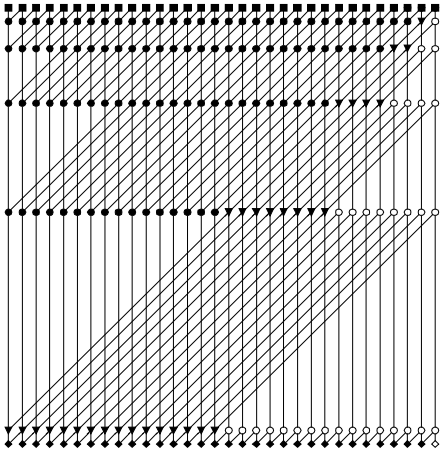
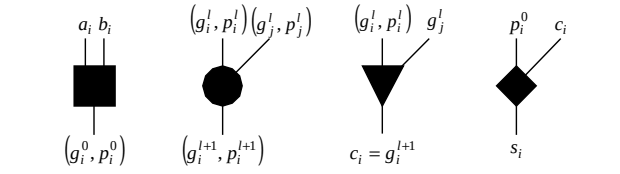


Figure 3. Kogge-Stone adder



$g_i^0 = a_i b_i$ $\quad g_i^{l+1} = g_i^l + p_i^l g_j^l$ $\quad s_i = p_i^0 \oplus c_i$

$p_i^0 = a_i \oplus b_i$ $\quad p_i^{l+1} = p_i^l p_j^l$

Figure 4. Black computation nodes

For simplicity, we assume equal input bit arrival times in the following analysis. Under unit gate-delay model with no consideration of wiring and fanout, these structures would have the same minimum logic depth and hence have the same delay. Such a delay estimation model, however, can only be useful for coarse evaluation. Nowadays, the impact of large fanout and more importantly, the impact of wire delay cannot be neglected. To get accurate wire delay estimation, the detailed layout is usually required. Fortunately, the graph representations of the above prefix structures can be directly mapped into a layout topology under structured custom design style [8][14], which makes wire delay estimation with relatively high accuracy possible. From the dot graphs, it can be seen that each critical path goes from the top-right corner to the bottom-left corner. The critical path can be divided into the several logic/wire stages. Each logic/wire stage can be abstracted as some logic driving a distributed RC chain modeling wiring effect with *N* spaced capacitive loads, as shown in Figure 5:

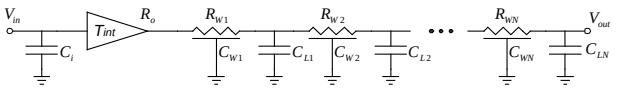


Figure 5. One logic/wire stage

The logic has input capacitance C_i, output resistance R_o and intrinsic delay T_{int}. The effect of diffusion capacitance in the logic output, C_d, has been included in T_{int}. $(C_W)_i$ and $(R_W)_i$ are the capacitance and resistance of a wire segment which should be treated as a distributed RC line. $(C_L)_i$ is the load capacitance at sink node *i*.

The total delay is the delay sum of each logic/wire stage. To clarify, we focus on the delay calculation of one logic/wire stage in Figure 5. The delay has two parts: logic delay and wire delay, i.e., $T_{total} = T_{logic} + T_{wire}$. The logic delay is defined as the delay with zero wire effects. By assuming all $(C_W)_i$ and $(R_W)_i$ to be zero, the logic delay can be easily calculated through a lumped RC model:

$$T_{logic} = T_{int} + 0.7 R_o \sum_{i=1}^{N} (C_L)_i \qquad (1)$$

Traditionally, the propagation delay is defined as the time period from 50% point of the input to 50% point of the output, which generates a 0.7 factor [1].

The wire delay includes the effects of distributed $(C_W)_i$ and $(R_W)_i$. The wire capacitance problem is further complicated if the wire fringing capacitance and coupling capacitance are considered. Before it is justified to make some simplification, all possible effects should be taken into account.

3. Wire delay model

Elmore [7] showed that the following closed-form time-constant expression as a first-order approximation is valid for a non-branched N-stage RC chain in Figure 6:

$$\tau_N = \sum_{i=1}^{N} \left(R_i \sum_{j=i}^{N} C_j \right) = \sum_{i=1}^{N} \left(C_i \sum_{j=1}^{i} R_j \right) \qquad (2)$$

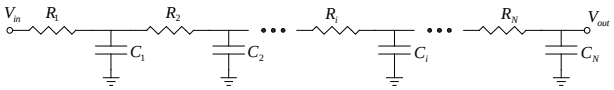


Figure 6. RC chain

To some extent, the RC chain in Figure 6 is similar to the distributed RC chain in our study. According to [1], a good approximation can still be made by combining weighted resistive and capacitive terms in the same way as in Elmore delay model. The weight is decided as follows. Under step voltage excitation, the propagation delay in a distributed RC network and the delay in a lumped RC network, are 0.4RC and 0.7RC, respectively. Thus, the resistive and capacitive terms are weighted by 0.4 when they are distributed and by 0.7 when they are lumped. Now a wire delay estimation equation can be formed for the wire part in Figure 5:

$$T_{wire} = 0.7R_o \sum_{i=1}^{N} (C_W)_i + \sum_{i=1}^{N} (R_W)_i \left(0.4(C_W)_i + 0.7(C_L)_i + 0.7 \sum_{j=i+1}^{N} \left((C_L)_j + (C_W)_j \right) \right) \quad (3)$$

The first item in Equation (3) is the pure effect of wire capacitance, which is a lumped RC model. The second item represents the effect of wire resistance, which is treated as a distributed RC model. The first item grows linearly with the wire length while the second item grows quadratically with the wire length. Note that the wire length is represented by the number of wire segments in Equation (3). In many cases of prefix adder structures, each load in one logic/wire stage has the same value and is uniformly spaced, which can help simplify Equation (3).

4. Simulation results

Prefix adders are organized in three parts: preprocessing part generating g_i^0 and p_i^0, prefix computation part generating c_i and post-processing part generating s_i. Different prefix structures only differ in the computation parts. Thus, our simulation only calculates the delay of the prefix computation part. To make the simulation data more practical and meaningful, we do not use overly simplified scaling models, such as uniform scaling and general scaling [12]. Instead, our simulation uses wire and device parameters that are derived from NTRS'97 [13]. The derivation has been conducted by Cong and Pan [5][6]. We directly borrowed their results for our purpose. The related parameters are listed in Table 1. The device here is a buffer, made up of two cascaded inverters with stage ratio of 1:5. The device parameters are obtained through HSPICE simulation. The wire capacitances are obtained using a 3D capacitance solver FASTCAP [11] for a wire of 2× minimum width and with two parallel neighboring wires of 2× minimum spacing.

Table 1: Device/Wire parameters [5]

Tech	0.25	0.18	0.15	0.13	0.10	0.07
Tint	86.6	66.4	65.5	54.4	50.1	29.8
Ro	16.2	17.1	17.3	22.1	23.4	22.1
Cg	0.282	0.234	0.220	0.135	0.072	0.066
Rs	0.0733	0.0679	0.0733	0.0806	0.0917	0.0952
Ca	0.0580	0.0589	0.0540	0.0462	0.0720	0.0611
Cf	0.0418	0.0302	0.0248	0.0183	0.0141	0.0148
Cx	0.0710	0.0583	0.0494	0.0428	0.0453	0.0416

Tech(μm): technology feature size;
Tint(ps): intrinsic device delay;
Ro(kΩ): output resistance of a minimum device;
Cg(fF): input capacitance of a minimum device;
Rs(Ω/□): sheet resistance of wire;
Ca(fF/μm^2): unit area capacitance of wire;
Cf(fF/μm): unit fringing capacitance of wire;
Cx(fF/μm): unit coupling capacitance of wire

In our study, each logic device on the critical path implements $g_i^{l+1} = g_i^l + p_i^l g_j^l$ (Figure 4). We assume each device is physically composed of one AND-NOR gate and one inverter of 5× minimum size. The intrinsic delay from g_j^l to g_i^{l+1} in such a device is approximated as $7/6 T_{int}$. The load capacitance is the AND-input capacitance in an AND-NOR gate, which is about 1.5Cg. The total wire capacitance of a wire of width W_w and of length L_w can be estimated as [4]:

$$C_w = C_a W_w L_w + C_f L_w + C_x L_w \quad (4)$$

The custom designer can change the wire spacing to reduce the coupling effect. For simplicity, we assume each channel wire to be of 2× minimum width and of 2× minimum spacing with neighboring wires if exist. Among the three structures considered, Ladner-Fisher structure has only one wire each channel and hence no coupling effect while both Knowles and Kogge-Stone structures exhibit coupling effects.

Because of the array topology of prefix adder structures, wire units can be defined to simplify the simulation. A minimum wire unit is defined as a transverse channel wire segment between two adjacent bits. In our study, the wire unit is of length 160λ and of width 4λ, where λ is traditionally defined as half of the technology feature size.

Based on the above analytical model and NTRS'97 technology parameters, the effects of wire resistance and capacitance on the prefix adders have been simulated and the results are discussed in the following.

4.1 Wire delay contribution

In DSM domain, wire delay plays an important role in circuits' performance. Figure 7 shows the simulation

results on wire delay contribution to the total critical path delay. In many case, wire delay exceeds logic delay and dominates critical path delay. As the feature size decreases, however, the wire delay contribution decreases slowly! When the feature size becomes less than 0.15um, the logic delay and wire delay of the critical path in prefix adders scale fairly well with the technology advancement. There are two major reasons for such good scaling. First, the technology roadmap has made great effect to reduce the wiring effect when the technology scales down. Second, the largest wire length in these adders (e.g., 1.28mm in 128-bit prefix adders) is still much less than a typical global wire length (e.g., 2cm) and the quadratic part in wire delay Equation (3) is very small.

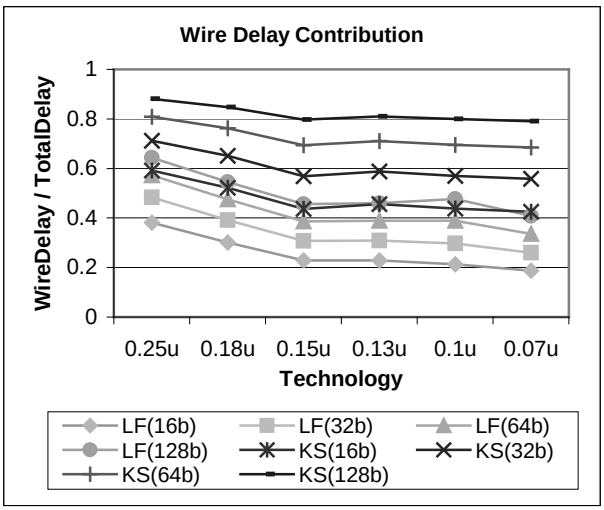


Figure 7. Wire delay contribution

For a given technology, the wire delay contribution increases steadily as the adder width increases. Such a steady increase implies that more effort should be put on wire design as datapath becomes wider. A typical effort example is to apply sophisticated buffering technique. The wire delay impact on Ladner-Fisher adder is much less than the impact on Kogge-Stone adder. This is because Ladner-Fisher adder has the least parallel wires in the channels and hence the least wiring effect. In fact, Ladner-Fisher adder and Kogge-Stone adder are two extrema in respect to wire delay contribution.

4.2 The impact of wire resistance

As shown in wire delay Equation (3), the second part that represents wire resistance impact increases quadratically with the wire length. In the design of global wires, such as clock and bus, this quadratic part has posed a very difficult problem to system designers. In prefix adder design, however, our simulation shows that the wire resistance delay is less than 1% of the total wire delay under any situation. Table 2 lists the simulation result of such wire resistance delay percentages in Ladner-Fisher adder, which has the largest percentages among the three prefix structures under study.

Table 2: wire-resistance delay / wire delay (%)

	0.25µ	0.18µ	0.15µ	0.13µ	0.10µ	0.07µ
16bit	0.07	0.08	0.09	0.08	0.07	0.10
32bit	0.14	0.14	0.17	0.14	0.14	0.18
64bit	0.26	0.26	0.32	0.27	0.26	0.34
128bit	0.51	0.51	0.62	0.52	0.50	0.66

Therefore, there is little need to consider the effect of wire resistance in prefix adder design and a lumped delay model is sufficiently accurate. By neglecting the quadratic part in the wire delay equation (3), the total delay can be expressed in a simple form:

$$T_{total} = T_{int} + 0.7R_o \sum_{i=1}^{N} \left((C_L)_i + (C_W)_i \right) \quad (5)$$

4.3 Comparison of three prefix adder structures

The comparisons of the overall delay of three prefix structures are shown in Figure 8 and 9. In ideal cases where there is no wire coupling capacitance, Ladner-Fisher structure would not be a good choice in terms of delay. In this case with no coupling effect, the wire delays of three structures are about the same and the difference in total delays comes from the logic delays. Because the logic gates in Ladner-Fisher structure have the largest fanout loads, Ladner-Fisher adder exhibits the largest logic delay and also the largest total delay.

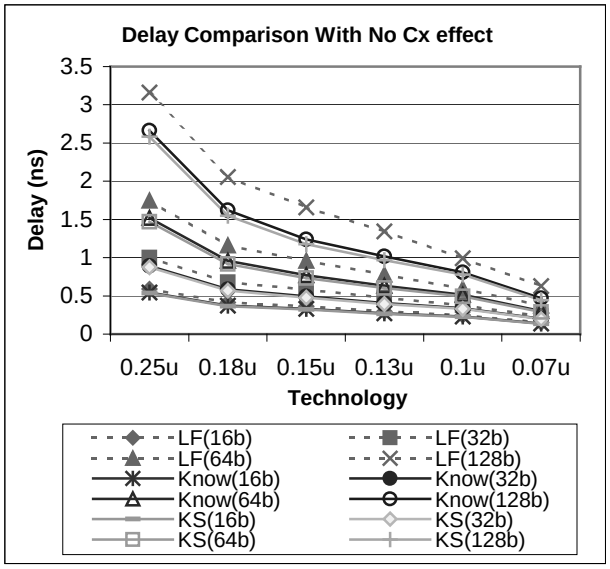


Figure 8. Delay comparison without coupling capacitance effect (ideal case)

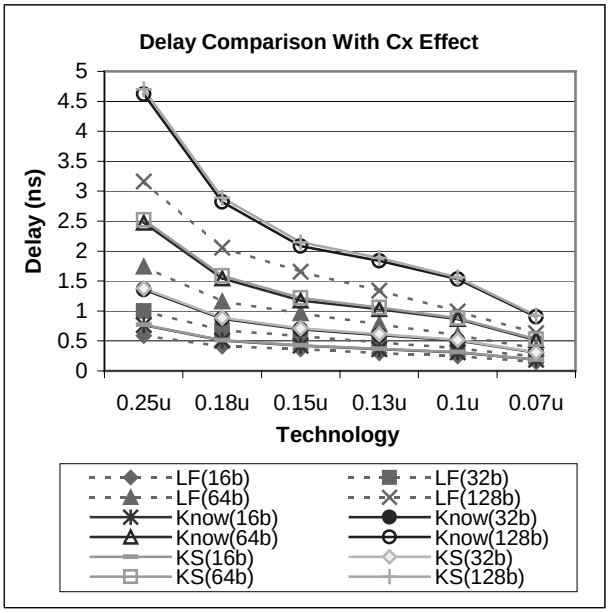


Figure 9. Delay comparison with coupling capacitance effect

However, the ideal case is seldom acceptable if the chip area and cost are considered. When the coupling effect is taken into account, the conclusion is reversed: Ladner-Fisher structure is the best choice under all technologies and datapath width cases. It has about 20% to 30% less delay. Thus, the effect of wire coupling capacitance plays a critical role in the design choice. To achieve the best performance, designers could keep wire spacing large enough to neglect coupling effect.

Figure 8 and 9 also show that the delay of Knowles structure and that of Kogge-Stone structure are almost the same in all circumstances. Because Knowles structure has much lower wire complexity, it is preferable when designers need to make choice between Knowles structure and Kogge-Stone structure.

5. Conclusions

The wire delay effect on the design of prefix adders in deep-submicron technology has been studied. The simulation is based on practical technology parameters from NTRS'97 and uses an analytical wire delay model that considers the fanout effect and the distributive nature of wire capacitance and resistance.

Simulation results show that wire delay exceeds logic delay and dominate the critical path delay of prefix adders in many cases. For a given technology, the wire delay contribution increases steadily as the adder width increases. As the feature size decreases, however, the wire delay contribution decreases slowly. The simulation data also imply that there is no need to consider wire resistance in prefix adder design and hence a lumped delay model is sufficiently accurate as a first-order estimation. On the other hand, the effect of wire coupling capacitance plays a critical role in prefix adder performance. Without coupling effect, Knowles and Kogge-Stone structures are better. When the coupling effect is taken into account, Ladner-Fisher structure is the best choice under all technologies and datapath widths.

Acknowledgment: This research has been supported in part by the NSF Grant MIP-9725771.

References

[1] H.B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley Publishing Company, 1990.

[2] R.P.Brent and H.T. Kung, "A regular layout for parallel adders", IEEE Trans. Computers, Vol. C-31, No. 3, pp.260-264, Mar. 1982.

[3] G. Choe, E.E. Swartzlander, "Interconnection effects in fast multipliers", Proc. 33rd Asilomar Conf. on Signals, Systems, and Computers, pp.1224-1227, 1999.

[4] J. Cong, L. He, C.-K. Koh and Z. Pan, "Global interconnect size and spacing with consideration of coupling capacitance", Proc. Int. Conf. on Computer Aided Design, pp.628-633, 1997.

[5] J. Cong, "Challenges and opportunities for design innovations in nanometer technologies", in SRC Working Paper, http://www.src.org/prg_mgmt/frontier.dgw, Dec. 1997.

[6] J. Cong and D.Z. Pan, "Interconnect delay estimation models for synthesis and design planning", Proc. Asia and South Pacific Design Automation Conf., pp.97-100, 1999.

[7] E. Elmore, "The transient response of damped linear networks with particular regard to wideband amplifiers", Journal of Applied Physics, pp.55-63, Jan. 1948.

[8] S. Knowles, "A family of adders", Proc. 14th IEEE Symp. on Computer Arithmetic, pp.30-34, 1999.

[9] P.M Kogge and H.S. Stone, "A parallel algorithm for the efficient solution of a general class of recurrence equations", IEEE Trans. Computers, Vol. C-22, No. 8, pp.786-793, Aug. 1973.

[10] R.E. Ladner and M.J. Fischer, "Parallel prefix computation", Journal of ACM, Vol. 27, No. 4, pp.831-838, Oct. 1980.

[11] K. Nabors and J. White, "Fastcap: A multipole accelerated 3-D capacitance extraction program", IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, pp.1447-1459, Nov. 1991.

[12] J. Rabaey, *Digital Integrated Circuits: A Design Perspective*, Prentice-Hall, 1996.

[13] Semiconductor Industry Association, *National Technology Roadmap for Semiconductors*, 1997.

[14] R. Zimmermann, *Binary Adder Architectures for Cell-Based VLSI and their Synthesis*, Ph.D. thesis, Hartung-Gorre Verlag Konstanz, 1998.