Tapered Transmission Gate Chains for Improved Carry Propagation

Boris D. Andreev, Edward Titlebaum, and Eby G. Friedman

Department of Electrical and Computer Engineering
University of Rochester
Rochester, New York 14627

{bandreev, tbaum, friedman} @ece.rochester.edu

ABSTRACT

Carry propagation chains are commonly found along the critical paths of many digital VLSI systems and, in particular, arithmetic circuits. The carry propagation delay, therefore, has a significant effect on system performance. Innovative design approaches, such as carry-lookahead adders and redundant arithmetic, trade off area, delay, and power through shorter carry-propagation paths. The focus of this paper is an alternative and complementary solution for decreasing the carry-propagation delay, particularly for those cases where chains of transmission gates are used. The behavior of transmission gate chain tapering, supported by simulation results, is presented. With this proposed circuit technique, the area, delay, and power of the carry propagation chain may be significantly improved.

I. INTRODUCTION

The speed of arithmetic logic circuits is a primary criterion in many digital VLSI systems, and is often achieved at the expense of increased area or power dissipation. The propagation delay of the carry signals through long chains of logic, a common structure in a conventional ripple-carry adder, is a major source of performance degradation in arithmetic circuits. Innovative design approaches, such as carry-lookahead adders (CLA) and redundant arithmetic, trade off area, delay, and power in order to produce 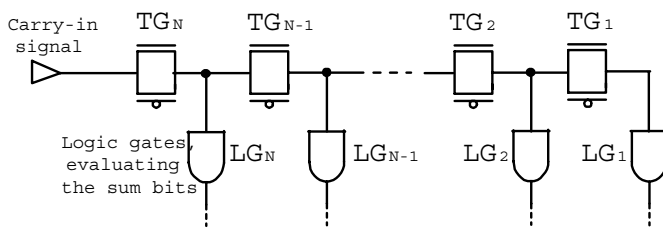shorter carry-propagation paths. In this paper, an alternative and complementary solution for increasing the speed of the carry-propagation chain is proposed, particularly for the case where transmission gate (TG) chains are used as shown in Fig. 1. TG chains are used in a large variety of practical VLSI-based circuits in both linear and binary tree carry propagation [1]-[4]. Ripple-carry adders based on TG chains are among the most area-efficient binary adders. High performance wide binary adders, such as carry-select and carry-lookahead, also include TG carry propagation sections typically ranging from four to six bits.

The transmission gate is one of the most important structures in CMOS integrated circuits, supporting a switch function, logic reduction, and an efficient layout [1]. The resistance between the input and output of an "on" TG is dependent on the voltage potentials at all four terminals of each of the two MOSFET transistors. The equivalent resistance of a TG can be approximated by a constant resistance, with a magnitude inversely proportional to the width of the transistor channels [1,5].

This paper is organized as follows: In section II, the tapered TG chain approach is motivated. An *RC* model of the circuit is briefly described in section III. The application of tapering to several circuits is presented in section IV. Some final remarks are provided in section V.

II. TAPERED TRANSMISSION GATE CHAINS

The feasibility of a tapered transmission gate chain is motivated by tapered CMOS buffering [6] and the constant capacitance-to-current ratio tapering (C^3RT) method [7]. Tapered CMOS buffers have been widely used to provide significant area/delay benefits when driving large capacitive loads. The approach used in these circuits is to provide a large output current without degrading the performance of the signal path by placing an excessively large capacitive load on the previous stage.
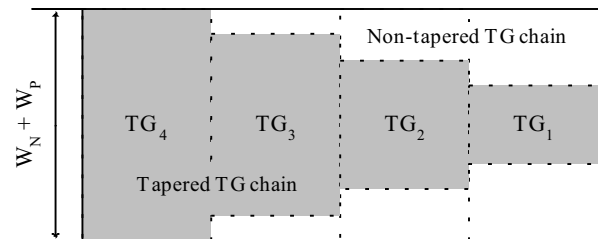

Fig. 1: A carry propagation chain

Fig. 2: Layout of transmission gate chains
Shaded area represents a tapered chain and
the non-shaded area represents a non-tapered chain

Tapering the geometric transistor width of the TGs in arithmetic circuits is the primary focus of this paper. This process is different from tapering standard CMOS buffers because the number of stages is fixed by the wordlength and there are significant local node capacitances, rather than a single large capacitive load. TG tapering is similar to a non-uniform distributed *RC* line because the input TG passes all of the current through the whole chain while in a tapered buffer every stage inverter drives an isolated local node capacitance. The objective presented here is the design of a TG chain that minimizes a specified cost function, or equivalently, requires the least amount of resources, area, power, and delay, for a specified constraint on one of these resources.

Non-tapered TG chains are less efficient because equally sized TGs pass different amounts of charge to charge/discharge different capacitances. The TG closest to the carry-in signal constrains the delay because all of the currents charging/discharging the individual nodes must pass through this circuit element.

Another way to describe this approach is as an additional degree of freedom in the design of carry propagation chains to achieve more efficient circuits. Several design variables exist for a tapered chain of N transmission gates: the geometric size (or channel width) of the TG, the tapering coefficient(s), and the size and number of repeaters inserted along the chain. The focus of this paper is on optimizing the basic TG size and the tapering coefficient(s). Once these variables are fixed, the repeater insertion problem is similar to delay reduction in resistive interconnect [8, 9].

The tapered TG chain is a convenient structure for the carry-lookahead adder architecture, where the carry-generate and carry-propagate signals for the least significant bit are available first, providing additional time to charge the larger input gates. The carry signal is also propagated through the more significant bits but the TGs for these bits are smaller, requiring less time to charge the gate capacitances.

III. CIRCUIT MODEL OF A TG CHAIN

Based on [10] - [14], a transmission gate chain can be approximated by an *RC* model as shown in Fig. 3. The model is used to analyze the delay, area, and power characteristics of the TG chain. An important assumption in this model is that the transmission gate maintains an approximately constant resistance through all regions of
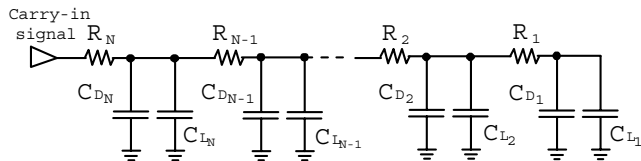

Fig. 3: An *RC* model for a chain of transmission gates

operation. Different node capacitances C_{Li} model both the input capacitance of the logic gate connected to node i and the interconnect capacitance between the two TGs. The basic parameters of the model are

k_i Tapering coefficient of the ith TG

$R_i = \dfrac{R_1}{k_i} = \dfrac{R}{k_i}$ Equivalent resistance of the ith TG

$C_{Di} = (k_i + k_{i-1})C_{D1}$ Scaled drain capacitance of ith TG and source capacitance of $(i-1)$[st] TG

C_{Li} Load and interconnect node capacitance

Due to the degradation of the signal waveform, the optimal number of TGs within a section is in the range $N = \{2, 3, 4\}$. For $N > 4$, the transition time of the signal waveform typically becomes excessively long, requiring a repeater to amplify the signal along the TG path.

Only uniform tapering is considered in this paper, where the TGs in all of the sections are scaled with the same tapering factor k, such that $k_i = k^{i-1}$. When the node capacitances C_{Li} differ significantly, the chain becomes irregular and different scaling coefficients should be used for each stage to maintain the same current to capacitance ratio. Such an approach is referred to as non-uniform tapering and is not addressed in this paper. In the following discussion, all of the local node capacitances C_{Li} are assumed to be equal to C_L; therefore, a single tapering variable k is applied.

Using the circuit model shown in Fig. 3, a conventional analysis of the Elmore propagation delay shows that the delay of a non-tapered circuit reaches a minimum for a specific ratio C_D/C_L. Increasing the TG size and the corresponding C_D beyond this ratio leads to a degradation in performance. Tapering with $k > 1$ can be used to achieve a higher speed. Spice simulations confirm that this crossover point is close to the ratio predicted by the analysis. Above this value, the delay of the non-tapered chain converges to a limit.

IV. OPTIMUM TAPERING

Transmission gate chains with three sections have been analyzed with the Spectre simulator within the Cadence IC design environment. The models used are for a TSMC 0.25 μm CMOS technology with a supply voltage of $V_{DD} = 2.5$ volts. The application of tapering to high speed carry propagation circuits is presented in section IV-A. Area-delay and energy-delay tradeoffs are discussed in sections IV-B and IV-C, respectively.

A. Optimum tapering for speed

Speed is the primary design criterion in high performance arithmetic circuits. The design objective in such circuits is to achieve the highest speed possible without greatly affecting area and power resources. A non-tapered transmission gate chain imposes a limitation on the minimum achievable delay, as illustrated in Fig. 4. Without tapering, the primary circuit design variable for
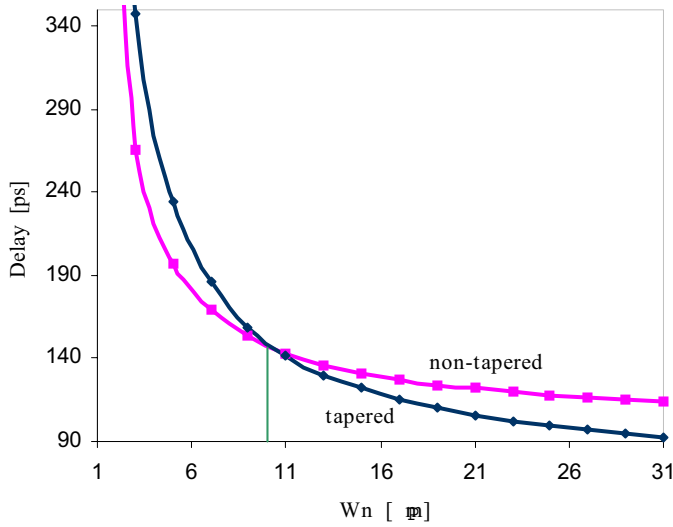
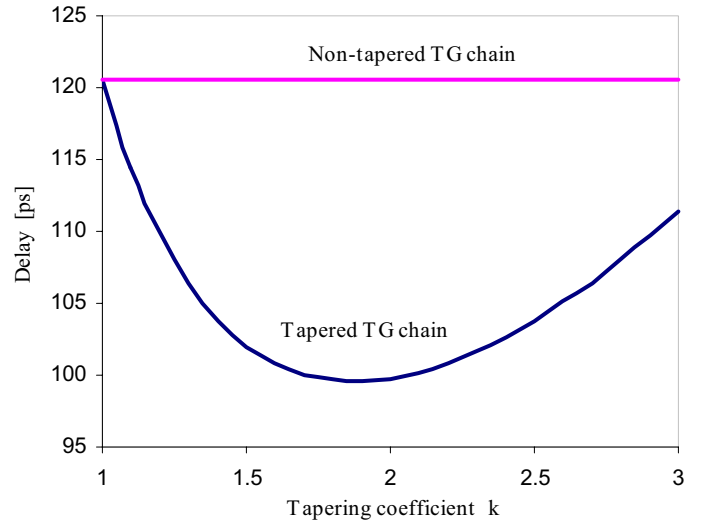Fig. 4: Delay for tapered and non-tapered TG chains,
C_L = 50 fF and k = 1.8

Fig. 5: Effect of variable tapering on the delay of a TG chain,
C_L = 50 fF and W_n = 25 µm

enhancing circuit speed is the width of the TGs along the chain. Increasing the width of the pass devices produces a corresponding increase in the drain, source, and gate capacitances. When the TG capacitances are on the order of the node capacitance C_L (from the logic gates and interconnect), any decrease in the signal propagation delay is limited. This behavior is similar to tapered CMOS buffers [6, 7]. Using the additional degree of freedom provided by tapering a chain of TGs, the delay may be reduced by more than 20% for the same base transistor width, shown as W_n in Fig. 2.

There are two key issues in the design of high performance TG chains: 1) for a fixed W_n, a lower delay can be achieved by tapered TGs, and 2) the limit on W_n can be extended such that the delay can be further reduced by increasing both W_n and k. As an example, the delay achieved by a non-tapered circuit for C_L = 50 fF is about 105 ps, while for a tapered circuit with k = 1.8, a propagation delay of 70 ps is achieved. For increasing node capacitance C_L, a higher speed is achieved by increasing both W_n and k.

B. Optimum tapering with area constraint

Although delay reduction is important in many arithmetic circuits, typically, the power and area are also constrained. It is, therefore, necessary to analyze the efficiency of tapered and non-tapered circuits under these practical constraints.

For both tapered and non-tapered TG chains, the size of the largest TG, determined by the width W_n of the NMOS transistor, may be used as a measure of the effective area occupied by the circuit. Above a certain value of transistor width W_n, a uniform TG chain is slower than a tapered circuit, as shown in Fig. 4. This width is denoted as W_{n0}. For smaller and slower circuits, a non-tapered TG chain is more efficient, while for faster and larger circuits, the tapered approach is preferable. As illustrated in Fig. 5, an optimal tapering factor k_{opt} exists for each set of design
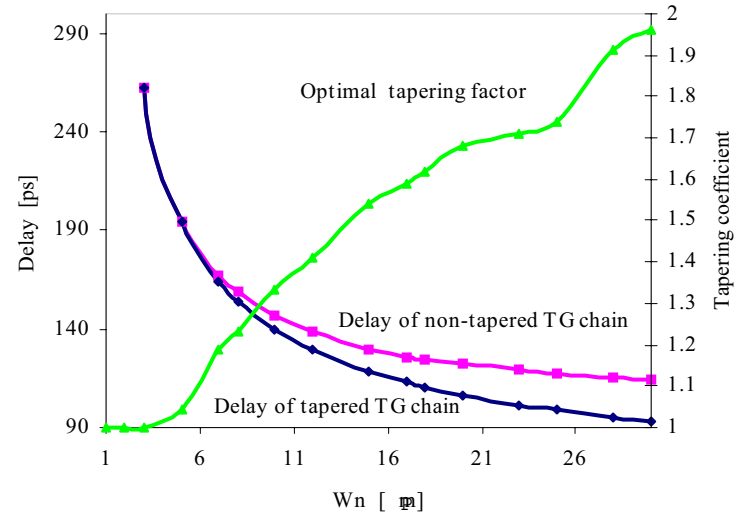

Fig. 6: TG chain delay and optimal tapering coefficient
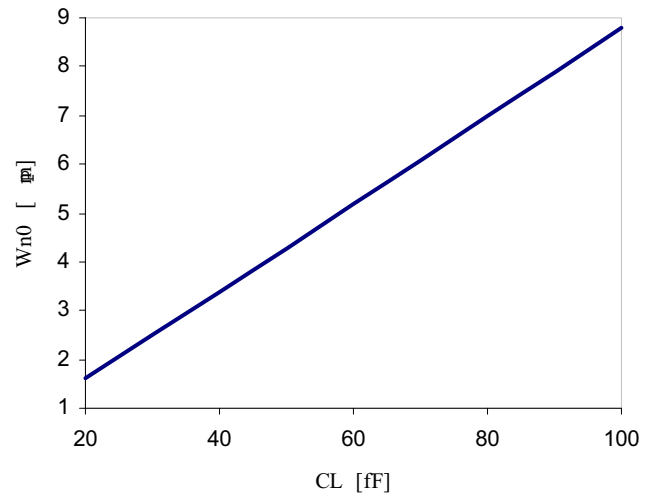for node capacitance C_L = 50 fF for variable area constraint

Fig. 7: TG width W_{n0} for different node capacitances
(for $W_n > W_{n0}$, the tapered circuit is faster than a non-tapered circuit)

constraints, W_n and C_L. The optimal tapering coefficient k_{opt} and width W_{n0} depend on the local node capacitance C_L and the maximum width W_n, as shown in Figs. 5 and 6, respectively. For $C_L = 50$ fF and $W_n = 25$ μm, a tapering factor of $k_{opt} = 1.8$ achieves a minimum delay. The relation between the node capacitance C_L and the width W_{n0} is linear as illustrated in Fig. 7. As noted in the previous section, as the node capacitance C_L grows, k should be increased to achieve the minimum delay for a given area constraint.

C. Optimum tapering to minimize the energy-delay product

Power dissipation is a growing concern in modern high performance circuits. The proposed tapered TG chain achieves a significant reduction in energy, occasionally exceeding 50%. The dynamic power dissipation is assumed to be directly proportional to the active area
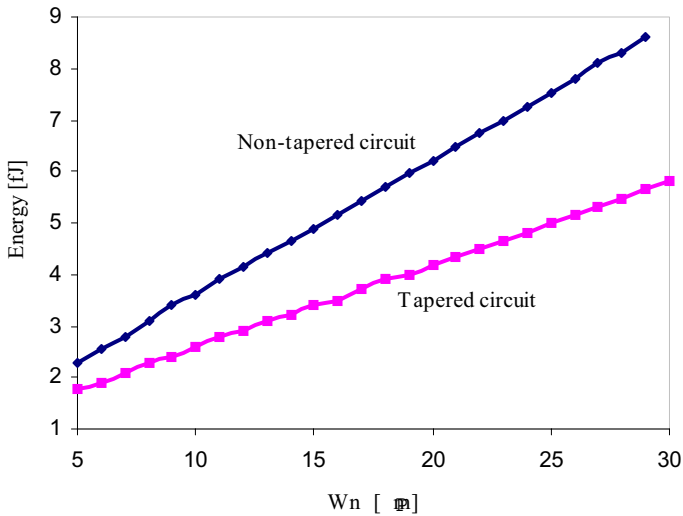

Fig. 8: Energy dissipation versus TG channel width, k = 1.8 and C_L = 50 fF

occupied by all of the TGs in a circuit, which corresponds to the total capacitance being charged. The savings in energy depends upon C_L and the target delay. An example is shown in Fig. 8 for $C_L = 50$ fF, where a non-tapered TG chain consumes significantly more power at the same or lower speed as compared to a tapered chain. This difference grows linearly as the target speed is increased.

V. CONCLUSIONS

A design technique for enhancing the carry-propagation delay of transmission gate chains has been proposed. This method is complementary to existing architectural and circuit techniques for designing fast arithmetic circuits. An additional degree of freedom, the tapering coefficient(s) of the TGs along the chain, is introduced to enhance the delay and energy characteristics of a TG chain. The area, delay, and power dissipation of a carry propagation chain is significantly enhanced.

The tapered approach in high performance TG chains has two primary beneficial factors: for a given area (fixed

W_n), tapered TGs exhibit a lower delay. Also with tapering, the minimum delay for wider TGs is extended to much higher values of W_n, achieving significantly enhanced overall performance.

As W_n is increased, the drain, source, and gate capacitances become a larger share of the total node capacitance, requiring a higher tapering factor to achieve a lower delay. The smaller the C_D/C_L ratio, the less tapering is required to compensate for the increase in the drain/source capacitances. When W_n is small, C_D is also small in comparison to C_L; therefore, tapering is not effective in this case.

REFERENCES

[1] G. Bouhasin, "The Transmission Gate: An Advantage of CMOS Gate-Arrays," *Journal of Semicustom ICs*, Vol. 2, No. 3, pp. 16-22, May 1985.
[2] B. D. Andreev, E. G. Friedman, and E. Titlebaum, "Efficient Implementation of a Complex ±1 Multiplier," *Proceedings of the ACM/SIGDA Great Lakes Symposium on VLSI*, pp. 83-88, April 2002.
[3] T. Sato *et al.*, "An 8.5-ns 112-b Transmission Gate Adder with a Conflict-Free Bypass Circuit," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 4, pp. 657-659, November 1993.
[4] H. Srinivas and K. Parhi, "A Fast VLSI Adder Architecture," *IEEE Journal on Solid-State Circuits*, Vol. 27, No. 5, pp. 761-767, May 1992.
[5] S. M. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits: Analysis and Design*, McGraw Hill, 1999.
[6] B. Cherkauer and E. G. Friedman, "A Unified Design Methodology for CMOS Tapered Buffers," *IEEE Transactions on VLSI*, Vol. 3, No. 1, pp. 99-111, March 1995.
[7] B. Cherkauer and E. G. Friedman, "Design of Tapered Buffers with Local Interconnect Capacitance," *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 2, pp. 151-155, February 1995.
[8] V. Adler and E. G. Friedman, "Repeater Design to Reduce Delay and Power in Resistive Interconnect," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 45, No. 5, pp. 607-616, May 1998.
[9] V. Adler and E. G. Friedman, "Uniform Repeater Insertion in RC Trees," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, Vol. 47, No. 10, pp. 1515-1523, October 2000.
[10] D. Deschacht, M. Robert, and D. Auvergne, "Explicit Formulation of Delays in CMOS Data Paths," *IEEE Journal of Solid-State Circuits*, Vol. 23, No. 5, pp. 1257-1264, October 1988.
[11] L. Brocco *et al.*, Macromodeling CMOS Circuits for Timing Simulation," *IEEE Transactions on Computer-Aided Design*, Vol. 7, No. 12, pp. 1237-1249, December 1988.
[12] S. Vemuru, "Delay Macromodelling of CMOS Transmission-Gate-Based Circuits," *International Journal of Modelling and Simulation*, Vol. 15, No. 3, pp. 90-97, March 1995.
[13] E. G. Friedman and J. H. Mulligan, Jr., "Ramp Input Response of RC Tree Networks," *Analog Integrated Circuits and Signal Processing*, Volume 14, No. 1/2, pp. 53-58, September 1997.
[14] J. Rubinstein, P. Penfield, and M. Horowitz, "Signal Delay in RC Tree Networks," *IEEE Transactions on Computer-Aided Design*, Vol. 2, No. 3, pp. 202-211, July 1983.