Power Dissipation of Tapered Serially Connected MOSFETs*

Brian S. Cherkauer and Eby G. Friedman

Department of Electrical Engineering
University of Rochester
Rochester, NY 14627

Introduction

In order to increase the performance of a CMOS circuit, integrated circuit (IC) designers apply many specialized techniques to decrease the time and power required for signals to propagate through combinatorial networks. One technique is the sizing of individual transistors for minimal delay and/or power dissipation.

Many CMOS logic structures are composed of chains of MOSFETs serially connected between a power supply rail and the output of the subcircuit. These serially connected MOSFETs are often a major source of delay and power dissipation, and therefore optimal sizing of these transistors is important in reducing their delay and power dissipation.

Typically, IC designers constrain the size of the transistors in the discharge chain to have the same channel dimensions and choose the channel width such that it fits their design criteria for speed and area. Shoji [1] first pointed out that under certain circumstances (specifically, the load capacitance must be of the same order of magnitude as the parasitic drain/source capacitances between the serially connected transistors), the constant width approach to transistor sizing may not be optimal. He proposed both a linear and an exponential tapering of transistor aspect ratios. He further demonstrated that it was often possible under these circumstances, with the proper choice of tapering factor, to produce a circuit which would discharge more quickly than an untapered chain.

Experimental Results of Tapering

In gates where the output capacitance is approximately equal to or smaller than the capacitance at the drain/source nodes of the serial chain, a tapered chain may be faster than an untapered chain. However, in many circuits where the output capacitance is a small multiple of the drain/source capacitance, tapering was found to be detrimental to the 50% delay time, which is the time from the input transition until the output reaches $V_{DD}/2$, if only the speed of the tapered gate is considered. Alternatively, if the effects of lowering the input capacitance are not considered, the delay of a tapered gate driving a capacitive load somewhat larger than the internodal capacitance of the serial chain will be slightly slower at the 50% output voltage level than an untapered gate with transistors the same size as the largest transistor of the tapered chain. It is only after the output voltage reaches $V_{DD}/2$ that the tapered structure responds more quickly than the untapered structure and reaches the 10% output voltage level earlier, as shown in Figure 1.

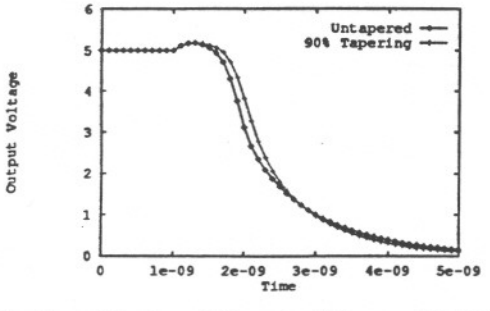

Figure 1: Output Waveform of Tapered and Untapered Serial Chain

It is this characteristic of tapering which is most useful in reducing CMOS short-circuit power dissipation [2]. The time during which the output voltage of the circuit operates between $V_{DD} - |V_{Tp}|$ and $V_{ss} + V_{Tn}$ is reduced, which translates into a reduction in short-circuit power dissipation in the following stage. This is the portion of the input waveform in which both the NMOS and PMOS devices in the following stage conduct short-circuit current, allowing DC current to flow from V_{DD} to V_{SS}. This is illustrated in Figure 2. Coupled with the reduction in dynamic power dissipation due to the decreased parasitic drain and source capacitances and input gate capacitance, tapering becomes advantageous in those designs for which power dissipation is often a major concern. This reduction in dynamic and short-circuit power dissipation comes with only a minimal increase or possibly a reduction in 50% delay and a reduction in 90–10% delay [1,3].

The short-circuit current depicted in Figure 2 is derived from an example circuit consisting of a six input domino Nand gate with a minimum sized inverter at the output. Tapering the serial chain with α=0.8 results in a 22% reduction in energy expended through short-circuit current in the load inverter. The maximum instantaneous short-circuit power was reduced by 10% as compared to an untapered serial chain.

The ratio of dynamic (CV^2f) power dissipation, P_d, originating from the input gate capacitance of an n transistor tapered serial chain, where n is the number of transistors in the chain, as compared to an untapered chain is given by

$$\frac{P_{d\ Tapered}}{P_{d\ Untapered}} = \frac{1}{n}\sum_{i=0}^{n-1}\alpha^i \qquad (1)$$

Since the largest transistor of the serial chain remains the same width with tapering, the impact on the clock distribution network is minimal in domino logic designs. This permits existing designs to incorporate tapering, thereby minimizing power dissipation without requiring the clock distribution network to be redesigned.

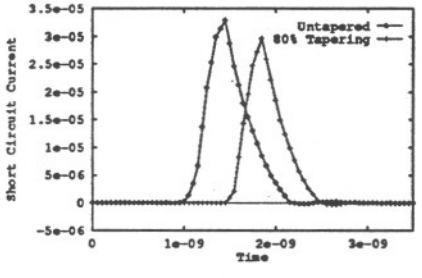

Figure 2: Short-Circuit Current in an Inverter Following a Serial MOSFET Chain

Conclusions

This paper empirically demonstrates that channel width tapering of serially connected MOSFETs affects both the speed and the shape of the output waveform. It is shown that in many cases where tapering actually increases delay, the change in the shape of the output waveform is useful in reducing short-circuit power dissipation in the stage following the tapered gate. This occurs in systems where the output capacitance is somewhat larger, but of approximately the same order of magnitude, as the internodal capacitance in the serial chain. Coupled with the reduction in dynamic power dissipation shown in (1), tapering has been shown to be a viable option in many designs where power dissipation is of fundamental concern.

References

[1] M. Shoji, "FET Scaling in Domino CMOS Gates," *IEEE Journal of Solid-State Circuits*, vol. SC-20, pp. 1067–1071, Oct. 1985.

[2] H. J. M. Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and Its Impact on the Design of Buffer Circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-19, pp. 468–473, Aug. 1984.

[3] S. S. Bizzan, G. A. Jullien, and W. Miller, "Analytical Approach to Sizing nFET Chains," *Electronics Letters*, vol. 28, pp. 1334–1335, July 1992.

* This material is based upon work supported by the National Science Foundation under Grant No. MIP-9208165