

# Low Power Digital CMOS Buffer Systems for Driving Highly Capacitive Interconnect Lines

Radu M. Secareanu<sup>1</sup> and Eby G. Friedman

Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627-0231

**Abstract**— A CMOS buffer system to drive highly capacitive interconnect lines with reduced power dissipation is presented. The speed of the proposed buffer system is equivalent to the speed of an optimal tapered buffer that drives an equivalent capacitive load, while the power dissipation and area are significantly reduced. The proposed buffer system also provides low switching noise and improved chip reliability by decreasing the likelihood of electromigration and on-chip hot spots.

## I. INTRODUCTION

In modern high performance digital circuits, high interconnect resistance and capacitance have become an important factor in limiting performance. Multiple solutions for driving highly capacitive loads such as a High Drive (HD) buffer [1] or cascaded tapered buffers [2–6] have been proposed. Approaches for driving highly resistive  $RC$  lines have also been described in [7–11].

Power dissipation is an important limiting factor in integration density. Deep submicrometer (DSM) technologies and high yields permit the integration of many millions of transistors. One primary limitation to increasing integration density, however, is the on-chip power dissipation.

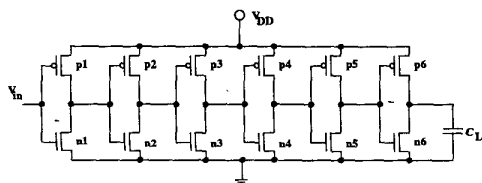


Fig. 1. Transistor-level schematic of a six stage CMOS tapered buffer

The on-chip interconnect also represents a major source of power dissipation. Typically, the highly capacitive interconnect lines are driven by tapered buffer systems, as shown in Fig. 1. An optimally sized tapered buffer to minimize delay has a tapering factor of  $e = 2.73$  [3–5]. Practically, a useful tapering factor is between three and four [5].

There are two primary capacitive elements that induce on-chip power dissipation, the capacitive load of the buffer  $C_L$ , and the internal capacitance of the buffer  $C_{int}$  which is the gate capacitances plus the junction capacitances of each transistor. The junction capacitances are a small fraction of the total gate capacitances, particularly for the transistor sizes considered in this paper, and are

therefore neglected. Accordingly, a tapered buffer driving a capacitive line dissipates a total power  $P_{tb}$  given by

$$P_{tb} = (C_L + C_{int})V^2f, \quad (1)$$

where  $V$  is the voltage swing at the output of each stage of the buffer (typically  $V_{DD}$ ) and  $f$  is the switching frequency. The proposed buffer system reduces the total power dissipation required to drive a capacitive load by reducing the internal capacitance of the buffer,  $C_{int}$ .

A detailed description of the operation of the proposed buffer circuit system with an emphasis on the process in which the power dissipation is reduced and distributed in time is presented in Section II. Simulation results characterizing the performance of the proposed buffer system are described in Section III. Finally, some conclusions are presented in Section IV.

## II. OPERATION OF THE BUFFER SYSTEM

The basic objective of this paper is to present a buffer system that provides the same speed as a tapered buffer (optimized for speed) while providing a significant savings in dissipated power. To achieve this goal, an HD buffer [1] optimized for power is used to drive the same capacitive load as an (optimal speed) tapered buffer. The output signal of the HD buffer optimized for power is restored by an HDR buffer [11] which drives the logic gates. A block level schematic of the proposed buffer system is shown in Fig. 2.

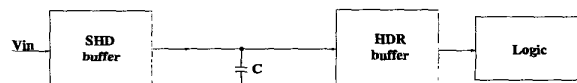


Fig. 2. A block level schematic of the proposed buffer system

A buffer circuit to drive large capacitive loads with improved speed, area, and power dissipation, the HD buffer, has been introduced in [1]. The performance of the HD buffer optimized for power [called a same speed HD buffer (SHD)] has also been presented in [1]. The circuit schematic of the SHD is shown in Fig. 3.

The internal power dissipation of the SHD is reduced as compared to a tapered buffer and an HD buffer according to the following factors:

- For each stage of an optimal tapered buffer (other than the last stage) and a low-to-high input transition, the output of that stage must drive the parasitic gate capacitance of the N-channel transistor of the following stage, representing 25% of the total capacitance at the output of a stage. For the high-to-low input transition of each stage, the output drives the parasitic gate capacitance of the P-channel transistor of the following stage, representing 75% of the total capacitance at the output of a stage. The HD buffer utilizes the driving (p and n), fast nulling (q), and maintenance nulling (m) transistors [1]. By using two separate paths, one path for the low-to-high input transition and another path for the high-to-low input transition, the HD buffer eliminates the parasitic capacitances for both input transitions. This behavior occurs since the load of each driving transistor except for the

This research was supported in part by the National Science Foundation under Grant No. MIP-9610108, the Semiconductor Research Corporation under Contract No. 99-TJ-687, a grant from the New York State Science and Technology Foundation to the Center for Advanced Technology—Electronic Imaging Systems, and by grants from the Xerox Corporation, IBM Corporation, Intel Corporation, Lucent Technologies Corporation, and Eastman Kodak Company.

<sup>1</sup>Now with Motorola, Inc., Semiconductor Product Sector, Digital DNA Laboratories - ASRL Advanced Circuits Research Group, Tempe, AZ 85284.

last stage of each path consists of only the useful gate capacitance of the following driving transistor along the path. This technique results in a significant reduction in both delay and power dissipation.

- The q and m transistors do not contribute to the power dissipation of the data path since the gates are separated from the data path. Any short-circuit power dissipation of the buffer is also eliminated by the q and m transistors by employing a synchronous driving technique. To explain this technique, consider the following example. For a fast output transition, only p6 or n6 should be on at one time. In this example, a fast 1-to-0 output transition is desired, therefore, only the n6 transistor must be on during transition. The p6 transistor must therefore be turned off before n6 is turned on. The p6 transistor is therefore turned off (or nulled) in between transitions when no path is driven by slowly charging the gate of the p6 transistor through the corresponding nulling transistors. To efficiently turn off the p6 transistor, the n5, p4, n3, p2, and n1 transistors are each turned off by the corresponding nulling transistors of each stage. The nulling transistors slowly turn off the driving transistors by charging or discharging the gates of the driving P-channel or N-channel transistors. This process prepares the driver for the impending fast transition. The synchronous driving technique refers to the technique of driving the nulling transistors of each stage so that the nulling transistors of one stage are on immediately after the driving transistor of the corresponding stage is turned off (the nulling process of that particular driving transistor is completed).

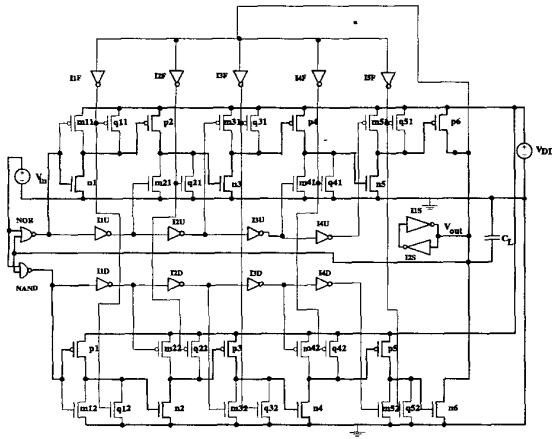


Fig. 3. An example of a six stage HD buffer optimized for power dissipation

- Consider  $D$  as the delay of an HD buffer and  $T$  as the time between two successive input transitions. The HD buffer introduced in [1] is shown in Fig. 4. For a six stage HD buffer with synchronous driving, the time allocated to null one stage is approximately  $T/6$ . Note that the size of the nulling transistors depends upon the  $T/D$  ratio. The transconductance of the nulling transistors is negligible if  $T/D$  is large (greater than five) or is similar to the driving transistors if  $T/D$  is small (less than two). For the extreme case of  $T/D = 1$ , the q transistors have the same transconductance as the corresponding driving transistors. Since the inverters that drive the nulling transistors are connected in a chain, each inverter in the chain is required to drive the parasitic gate capacitance of the following inverter in the chain in addition to

the useful nulling transistors. The total capacitive load at the output of each inverter must be driven within  $T/6$ . In comparing this nulling circuit to the inverters driving the nulling transistors for synchronous driving in the SHD buffer shown in Fig. 3, the followings issues are noted:

- Each inverter only drives the gates of the nulling transistors, eliminating any parasitic capacitance.
- The inverters must drive the nulling transistors of the first stage in  $T/6$ , the nulling transistors of stage 2 in  $2T/6$ , and, for example, the nulling transistors of stage 5 in  $5T/6$ . Note that for the SHD configuration, the larger the capacitance of the node being nulled, the greater the time offered by the circuit to complete the nulling process. This situation permits the size of the five inverters to be considerably reduced, saving power and area. Summarizing, by eliminating any parasitic capacitance from the nulling process and permitting smaller sizes for the inverters driving the q transistors while maintaining the synchronous driving, the SHD buffer provides improved power and area savings as compared to the HD buffer shown in Fig. 4.

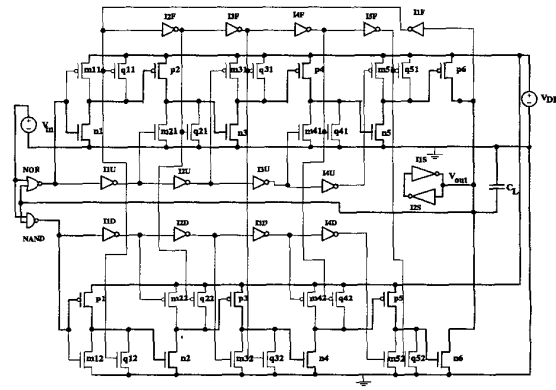


Fig. 4. An example of a six (even) stage HD buffer

- The HD buffer generally has a reduced number of stages as compared to a tapered buffer due to the elimination of the internal parasitic capacitances, thereby permitting a higher tapering factor [1]. The speed and power dissipation of an HD buffer is significantly improved as compared to a tapered buffer [1]. The SHD buffer dissipates even less power than an HD buffer. Also note that an SHD buffer supports a low duty factor  $T$  for the input signal when  $T/D = 1$ . This performance characteristic is not achievable with a tapered buffer, since  $D$  is much smaller than the delay of an optimal speed tapered buffer system driving an equivalent capacitive load. The same duty factor can be achieved by an HD buffer dissipating more power.
- In addition to the aforementioned factors that reduce the internal power dissipation of the SHD buffer, the size of the driving transistors of the buffer provides additional savings in power and area. The SHD buffer may be sized to achieve the same speed as an equivalent optimal speed tapered buffer with a significant reduction in power and area. To achieve this performance, the final stage of the SHD buffer is scaled down by a factor of three to four as compared to the size of an equivalent tapered buffer. The number and size of the remaining stages of the SHD buffer which drive the final stage are chosen to produce the minimum delay. The resulting SHD buffer has far less stages and is geometrically much smaller than a standard

high speed HD buffer, resulting in a significant reduction in total power and area. However, due to the size of the final stage, the output transitions are slower [1].

To improve the slow output signal transitions with a minimal delay penalty while preserving the power and area savings, an HDR buffer [11] is used at the end of the line, as shown in Fig. 2. The HDR buffer circuit is capable of restoring slow transition times and distorted input signals with a minimum delay penalty by featuring a voltage transfer characteristic (VTC) with low threshold voltages and hysteresis [11]. The drawback of the HDR buffer is a higher sensitivity to noise. The VTC with hysteresis of the HDR circuit is shown in Fig. 5. For comparison, the VTC with hysteresis of a Schmitt trigger is also shown. The proposed circuit offers threshold voltages of  $V_{M+}$  ( $V_{IH}$ ) up to 1 V and  $V_{M-}$  ( $V_{IL}$ ) up to 4 V for a process in which  $V_T = 0.8$  V. Note that a Schmitt trigger has  $V_{M+} = 4$  V and  $V_{M-} = 1$  V. Accordingly, the HDR circuit switches when the low-to-high input transition reaches 1 V, while a Schmitt trigger switches at 4 V. Similar behavior occurs for the high-to-low transition (4 V as compared to 1 V). Therefore, this circuit can respond quickly to slow input transitions, providing a significant gain in circuit speed.

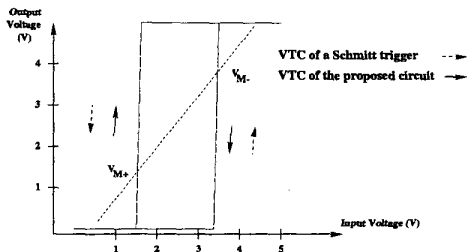


Fig. 5. The VTC of an HDR circuit

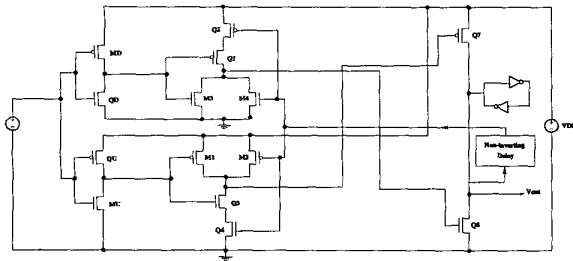


Fig. 6. An HDR circuit to minimize delay

Different versions of the HDR buffer have been discussed in [11]. Since a minimal delay with low threshold voltages and reduced internal capacitances [11] is desired, the circuit shown in Fig. 6 is used.

Note that the low switching threshold voltages of the HDR circuit are achieved through asymmetric sizing of the input inverters [11]. A typical HDR response is depicted in Fig. 7. Note the slow input signal transition, the HDR response at the  $V_{M+}$  and  $V_{M-}$  threshold voltages, and the sharp output response.

Since an optimal speed tapered buffer provides fast output transitions, the tapered buffer generates large noise transients on the power lines particularly when driving large capacitive loads. The tapered buffer may also induce noise through capacitive coupling and dissipate a large amount of power in a short time. This behavior

creates noise and reliability problems such as electromigration within the power lines and hot spots due to large localized on-chip power. Since the SHD buffer has slow output transitions, the power required to drive  $C_L$ , even if the same as for a tapered buffer, is distributed over the output transition. This characteristic minimizes electromigration and noise problems (since the instantaneous current pulses are much smaller), permitting the power to be dissipated gradually.

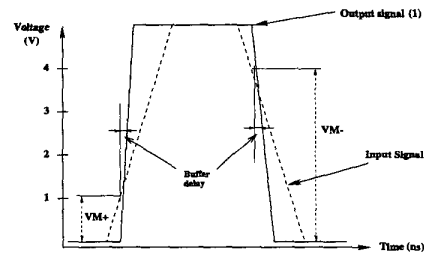


Fig. 7. A typical response of the HDR circuit

### III. SIMULATION RESULTS AND PERFORMANCE COMPARISON

Circuit simulations based on Cadence-Spectre and a  $1.2\ \mu\text{m}$  CMOS technology are described in this section. The results do not include the power dissipation due to  $C_L$ , since this component is similar for both the tapered buffer and the proposed buffer system. By excluding this component of the power dissipation, focus is maintained on the power dissipated by the internal capacitances of the buffers. The internal power dissipation depends on the total gate capacitances of the buffer which is proportional to the total channel width. The same proportionality is maintained for the total area. While the results are based on a  $1.2\ \mu\text{m}$  technology, the ratio of the total width of the tapered buffer system versus the total width of the proposed buffer system is approximately the same for any CMOS technology. Therefore, generalizing, these results are expressed in terms of this ratio.

TABLE I  
THE SIZE OF A SIX STAGE TAPERED BUFFER DRIVING A 20 PF LOAD

Stage	1	2	3	4	5	6
$W_n$ ( $\mu\text{m}$ )	1.8	6.3	22	77	270	949
$W_p$ ( $\mu\text{m}$ )	4.3	15	53	185	650	2278

The conditions for which this analysis is performed are: 1) a tapered buffer is optimally sized for speed to drive a 20 pF capacitance, 2) a buffer system, as shown in Fig. 2, is sized to drive the same 20 pF capacitive load and provide a total delay similar to the tapered buffer system, and 3) the power supply is 5 volts, the frequency of the input signal is 100 MHz, and the input transition times are 0.1 ns.

The optimal size of a tapered buffer system produces a minimal delay of 1.6 ns for seven stages with a tapering factor at 2.8. A six stage tapered buffer system produces a delay of 1.68 ns with a tapering factor at 3.5. The size of the six stage tapered buffer is listed in Table I. The output transition times are 0.44 ns.

The proposed buffer system is sized to obtain a similar delay as the tapered buffer system. The size of the SHD buffer is shown in Table II. Note that a three stage SHD

buffer is required. The total delay required for the signal to reach 1.4 volts (approximately the switching threshold voltage of the HDR buffer) is 1.22 ns. The transition times at the output of the SHD buffer are approximately 1.5 ns. The total width of the input gates, the output latch, the nulling transistors, and the inverters driving the nulling transistors is approximately 100  $\mu\text{m}$ .

TABLE II  
THE SIZE OF A THREE STAGE SHD BUFFER DRIVING A 20 pF LOAD

Stage	1	2	3
$W_n$ ( $\mu\text{m}$ )	1.8	34	215
$W_p$ ( $\mu\text{m}$ )	5	34	640

The HDR buffer size and performance are shown in Table III. The output load is 1 pF, which represents a significant number of minimally sized logic gates. The non-inverting delay element is implemented by inverters [11]. The total width of the NAND and NOR gates, output latch, and delay element is approximately 200  $\mu\text{m}$ . The transition times at the output of the HDR buffer are 0.31 ns. The HDR buffer delay is defined from the time when the HDR input threshold voltages are reached to the time when 50% of the output waveform [11] is reached.

TABLE III  
THE CHARACTERISTICS OF AN HDR BUFFER DRIVING A 1 pF LOAD

No.	QU/ QD ( $\mu\text{m}$ )	$V_{M+}/$ $V_{M-}$ (V)	Final stage ( $\mu\text{m}$ )	Buffer delay (ns)
1	54/ 162	1.3/ 3.7	110/ 320	0.30

To estimate the total internal power and area savings, the total width of the transistors of the two buffer systems is computed. The total width of the transistors for the six stage tapered buffer system is approximately 4400  $\mu\text{m}$ , while for the optimal seven stage tapered buffer system is approximately 6100  $\mu\text{m}$ . The total width of the transistors of the proposed buffer system, in comparison, is approximately 1900  $\mu\text{m}$ . Accordingly, the internal power dissipated by the six stage tapered buffer as compared to the proposed buffer system is, for this example, 2.3 times higher, while for the seven stage tapered buffer is 3.2 times higher. The total area of the two tapered buffers as compared to the proposed buffer system, if the total width of the transistors is considered, is approximately the same ratio. However, due to the more complex routing of the proposed buffer system, the area ratio is expected to be less than 2.3 and 3.2, respectively.

These significant power and area savings are obtained without any degradation of either the speed or output waveforms. The delay of the six stage tapered buffer system is 1.68 ns, as compared to the total delay of the proposed buffer system which is 1.54 ns (approximately 10% smaller). Note that the delay of the proposed buffer system is also smaller than the delay of the seven stage buffer system by approximately 4%. The output signal has a transition time for the six stage tapered buffer of 0.44 ns, as compared to the proposed buffer system of 0.31 ns, approximately 40% smaller. The proposed buffer system also produces less switching noise, thereby improving circuit reliability such as electromigration and minimizing the number of hot spots.

Note that different tapering factors for the tapered buffer system generate different ratios for the transistor

width of the two buffer systems, producing different savings in power and area. As previously discussed, the objective described here is to obtain a buffer system that provides the same speed as a tapered buffer system while offering significant power and area savings. Accordingly, for a 20 pF load, the correct transistor width ratio is 3.2, corresponding to the described seven stage tapered buffer. This ratio can be further improved by equalizing the speed and the output waveforms of the two buffer systems. To equalize the speed, thereby improving the power dissipation, the proposed buffer system can be further improved, by either scaling down the SHD buffer or the HDR buffer. Several methods may be applied to scale down the HDR buffer and reduce the power dissipation: 1) increase the switching threshold voltages, 2) decrease the size of the final and intermediate stages, and 3) use less stages. A buffer system can also be sized to achieve the same speed as a six stage tapered buffer with the resulting width ratio, for this example, in the vicinity of 3.2.

The power and area savings of the proposed buffer system as compared to a tapered buffer system are proportional to the capacitive load. Accordingly, the proposed buffer system provides substantial power and area savings, particularly when driving a large capacitive load.

#### IV. CONCLUSIONS

A buffer system is presented that provides a substantial savings in power and area as compared to a classical tapered buffer system. These savings are achieved when the buffer system is sized to provide approximately the same delay and output waveforms as the tapered buffer system. The proposed buffer system is particularly efficient for driving large capacitive loads. The buffer system also reduces switching noise, provides improved chip reliability by decreasing the likelihood of electromigration, and reduces the probability of on-chip hot spots due to lower instantaneous current levels.

#### REFERENCES

- [1] R. M. Secareanu and E. G. Friedman, "A High Speed CMOS Buffer for Driving Large Capacitive Loads in Digital ASICs," *Proceedings of the IEEE ASIC Conference*, pp. 365-368, September 1998.
- [2] H. C. Lin and L. W. Linholm, "An Optimized Output Stage for MOS Integrated Circuits," *IEEE Journal of Solid-State Circuits*, Vol. SC-10, No. 2, pp. 106-109, April 1975.
- [3] N. Hedenstierna and K. O. Jeppson, "CMOS Circuit Speed and Buffer Optimization," *IEEE Transactions on Computer-Aided Design*, Vol. CAD-6, No. 2, pp. 270-281, March 1987.
- [4] N. C. Li, G. L. Haviland, and A. A. Tuszynski, "CMOS Tapered Buffer," *IEEE Journal of Solid-State Circuits*, Vol. SC-25, pp. 1005-1008, August 1990.
- [5] B. S. Cherkauer and E. G. Friedman, "A Unified Design Methodology for CMOS Tapered Buffers," *IEEE Transactions on VLSI Systems*, Vol. VLSI-3, No. 1, pp. 99-111, March 1995.
- [6] B. S. Cherkauer and E. G. Friedman, "Design of Tapered Buffers with Local Interconnect Capacitance," *IEEE Journal of Solid-State Circuits*, Vol. SC-30, No. 2, pp. 151-155, February 1995.
- [7] 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. CAS II-45, No. 5, May 1998.
- [8] V. Adler and E. G. Friedman, "Delay and Power Expressions for a CMOS Inverter Driving a Resistive-Capacitive Load," *Analog Integrated Circuits and Signal Processing*, Vol. 14, No. 1/2, pp. 29-39, September 1997.
- [9] V. Adler and E. G. Friedman, "Repeater Insertion to Reduce Delay and Power in RC Tree Structures," *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, pp. 749-752, November 1997.
- [10] S. Dhar and M. A. Franklin, "Optimum Buffer Circuits for Driving Long Uniform Lines," *IEEE Journal of Solid-State Circuits*, Vol. SC-26, pp. 151-155, January 1991.
- [11] R. M. Secareanu, V. Adler, and E. G. Friedman, "Exploiting Hysteresis in a CMOS Buffer," *Proceedings of the IEEE International Conference on Electronics, Circuits, and Systems*, pp. 205-208, September 1999.