An Optimized Output Stage for MOS Integrated Circuits

HUNG CHANG LIN, FELLOW, IEEE, AND LOREN W. LINHOLM, MEMBER, IEEE

Abstract—**An output device for optimizing propagation delay and minimizing chip area is described. An optimum means of tapering the output stages to minimize propagation delay is determined. The minimum delay is a function of the capacitive load to node ratio, the number of output stages, and the interstage propagation delay. The effects on area are also presented. A figure of merit which is a function of area and propagation time is defined which is of use in designing output stages. An optimum exists which can be considered the best compromise between further decreasing propagation delay and increasing chip area. Data are also presented which allow a designer to determine the minimum chip area once the capacitive load and the maximum allowable delay are known.**

Introduction

In iterative MOS integrated circuits, the output transistors should be large enough to drive the required load capacitance. The last output transistor will load the previous stage and slow it down. To improve the propagation time, the driver for the output stage should also be enlarged. Obviously, if every stage is enlarged, excessive area will be consumed. However, if the drivers are made progressively smaller than the succeeding stage, the total area can converge to a reasonable value. This paper is an analysis of such a tapered output stage for optimizing propagation time and chip area. This technique is applicable, in general, to MOS integrated circuits where a high output drive capability is required.

Output Stage Design Principles

In order to drive typical load capacitances required in an integrated electronic system, the output transistors in an integrated circuit should be of suitable size to provide large charging and discharging currents. Unfortunately, large output transistors load the previous stage, which decreases its operating speed. To improve this situation, the driver for the output stage should also be enlarged.

Consider the output circuit such as the shift register shown in Fig. 1. If the load capacitance, C_L, is equal to M times the interstage node capacitance, C_0, and the MOS devices are of minimum sizes, the propagation delay (which is proportional to load capacitance) through the last two stages is:

$$t_{pL} = (M+1)\, t_{p0} \tag{1}$$

where

Manuscript received August 21, 1974; revised November 22, 1974. The technique described here was used in the design of a DMOS shift register for the National Aeronautics and Space Administration under Contract NAS-1-12533 awarded to the Westinghouse Electric Corporation.

H. C. Lin is with the Department of Electrical Engineering, University of Maryland, College Park, Md. 20742.

L. W. Linholm is with the Department of Defense, Ft. Meade, Md. 20755.

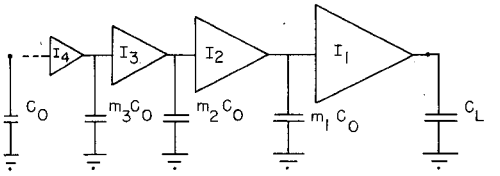


Fig. 1. Cascade MOS stages.

t_{pL} = propagation delay through the last two stages
t_{p0} = propagation delay at the low level interstage
M = C_L/C_0
C_L = load capacitance
C_0 = interstage node capacitance.

The propagation time can be reduced by enlarging the sizes of the MOS devices of the last stage, I_1, by m times. The driver stage, I_2, is slowed down because the output transistors now see an m-fold increase in capacitive load. The propagation delay of the driver and output stages becomes

$$t_{pL} = \frac{M}{m}(t_{p0}) + m\,(t_{p0}). \tag{2}$$

To find the minimum t_{pL}, we differentiate the above with respect to m and set the differentiated equation equal to zero.

$$\frac{d(t_{pL})}{dm} = \left(\frac{-M}{m^2} + 1\right) = 0 \tag{3}$$

or

$$m = M^{1/2}$$

and

$$t_p\,(\min) = 2M^{1/2} t_{p0}. \tag{4}$$

For shorter propagation time, it is better to enlarge the sizes of the last few stages. For instance, one may increase stages I_1 by m_1 times, increase stage I_2 by m_2 times, I_3 by m_3 times, etc. Then the propagation delay for the last four stages becomes

$$t_{p4} = \left(\frac{M}{m_1} + \frac{m_1}{m_2} + \frac{m_2}{m_3} + m_3\right) t_{p0}. \tag{5}$$

By differentiating t_{p4}, we find

$$m_1 = M^{3/4},\quad m_2 = M^{2/4},\quad m_3 = M^{1/4} \tag{6}$$

and

$$t_{p4} = 4M^{1/4} t_{p0}. \tag{7}$$

The propagation delay of every stage is

$$t_{pL} = M^{1/4} t_{p0}. \tag{8}$$

Note that the propagation time per stage resulting from enlarging the last three stages given in (8) is shorter than that of enlarging just the last stage. Also, the optimum sizes of the device decrease monotonically as $M^{3/4}$, $M^{2/4}$, and $M^{1/4}$ times the minimum interstage sizes.

From the last two equations, one may generalize that for an output device of N stages, the propagation time per stage is

$$t_{pS} = M^{1/N} t_{p0}. \tag{9}$$

A computer analysis of a ten stage output device utilizing a two-phase ratioless logic interstage was made. The results of the analysis showed good agreement of the predicted propagation delay to the computed delay to within the accuracy of the program used.

Area Optimization

From the above derivation, one can see that the optimum device size should decrease on the order

$$m_1 = M^{(N-1)/N}, \quad m_2 = M^{(N-2)/N}$$
$$m_3 = M^{(N-3)/N}, \quad \cdots M_{N-1} = M^{1/N}$$
$$m_N = 1.$$

When a number of stages are connected in cascade, the area is the sum of all the stages. When the stages are tapered to optimize the speed, each preceding stage is reduced by a factor $1/m$. If there are N-1 enlarged stages, the area of the last N stage is

$$A_N = A_0 (1 + m + m^2 + m^3 + \cdots m^{N-1}) \tag{10}$$

where A_0 is the area of the standard stage (not enlarged). This summation is equal to

$$A_N = \frac{A_0 (m^N - 1)}{m - 1}. \tag{11}$$

But since

$$M = m^N \tag{12}$$

$$A_N = \frac{(M - 1)}{M^{1/N} - 1} A_0. \tag{13}$$

This is the ratio of the enlarged area to the standard cell area. Equation (13) is plotted in Fig. 2 for $M = 100$. Also plotted is the propagation time of stage t_{pS}, which is normalized.

$$t_{pS}/t_{p0} = M^{1/N}. \tag{14}$$

From Fig. 2 we find that the increased area ratio is of the order of the load capacitance to node capacitance ratio, M; the area increases as the number of enlarged stages N increases; and the propagation time per stage decreases as the numbers of enlarged stages decreases.

Optimum Stage Design

The optimum design is a compromise between speed and area. The choice is based on economics. We can choose a figure of merit, F, defined as

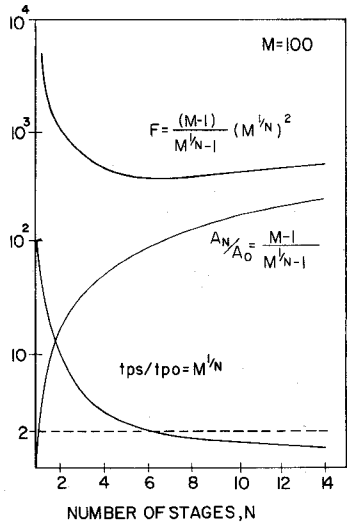


Fig. 2. Normalized propagation delay, normalized area, and figure of merit F versus number of stages for $M = 100$, $K = 2$.

$$F = \left(\frac{A_N}{A_0}\right)\left(\frac{t_{pS}}{t_{p0}}\right)^K \tag{15}$$

where K is a weighting exponent. If K is greater than unity, it means more weight is placed on speed than area. Substituting (13) and (14) into (15)

$$F = \frac{(M - 1)}{M^{1/N} - 1} (M^{1/N})^K. \tag{16}$$

The minimum F can be found by differentiating F with respect to N.

$$M^{1/N} = \frac{K}{K - 1}. \tag{17}$$

For instance, if $K = 2$

$$M^{1/N} = 2.$$

This has several implications. From (14)

$$t_{pS}/t_{p0} = M^{1/N}.$$

The minimum area and propagation delay is achieved when

$$t_{pS}/t_{p0} = 2. \tag{18}$$

The optimum area is

$$A_N \text{ (opt)} = (M - 1) A_0. \tag{19}$$

Fig. 3 is representative of how the figure of merit varies for several load to interstage capacitance values. Below the optimum value, t_{pS} decreases more slowly than the area increases; above this value t_{pS} increases more quickly than the area decreases.

At some point, however, a large increase in propagation delay occurs. As seen in Fig. 2, the area-propagation delay square product increases drastically for small values of N. This is not the case for larger values of N. Within the design constraints of a particular circuit, the designer pays a smaller penalty in the area propagation delay square product by in-

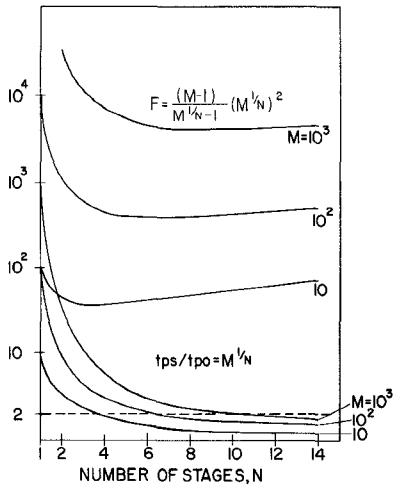


Fig. 3. Normalized propagation delay and area-propagation delay square product F versus number of stages.

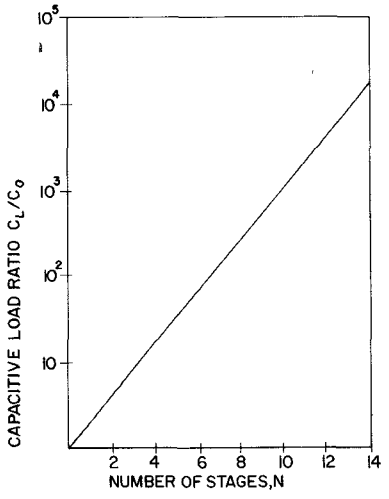


Fig. 4. Optimum number of enlarged output stages for different load to node capacitance ratios.

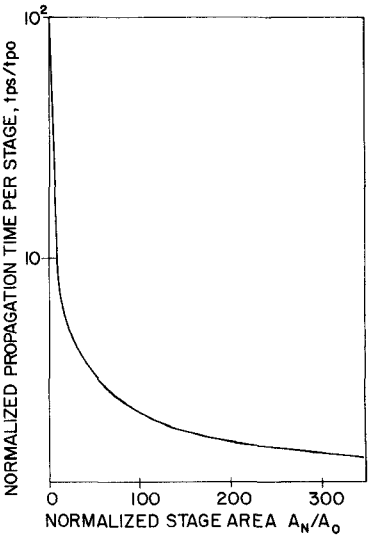


Fig. 5. Minimum propagation delay obtainable for a given area and a fixed load to node ratio of 100.

creasing the number of stages beyond the optimum point than decreasing the number of stages. Also, since

$$M^{1/N} = 2 \tag{20}$$

$$M = m^N.$$

An optimum condition exists when

$$m = 2 \tag{21}$$

or when the capacitance ratio per stage doubles. Simply stated, the best compromise between area increase and the square of propagation delay decrease exists when the output device is designed such that each required stage in the output circuit doubles in size or capacitance between the last interstage circuit and the load. Fig. 4 shows the number of stages required for a given load to interstage capacitance ratio for optimum performance.

From a design standpoint, it is most economical to minimize the area. If the entire integrated circuit consists of B number of cells, of which N stages are enlarged, the total area is

$$A = (B - N)A_0 + A_N \tag{22}$$

assuming each cell occupies A_0 area. In most instances, B is large. The enlarged area A_N does not add any more bits but merely improves the loading and speed capability. Thus, a compromise exists between speed and area. If $(B - N) \gg A_N/A_0$, the addition of enlarged area is insignificant. If A_N becomes comparable or larger than $(B - N)A_0$, the enlarged buffer can add substantially to the area and cost of the integrated circuit. In this case, the design engineer may choose to commit some fraction of the total chip area to an output device. Fig. 5 represents the minimum propagation delay obtainable for a given area and a fixed capacitive load.

Conclusions

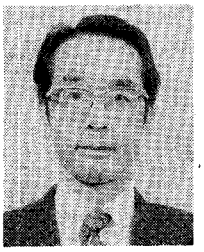
A means of optimizing propagation delay for an output inverter stage has been described. It is applicable in general to

MOS integrated circuits where a high output drive capability is required. Depending on the capacitive load to node ratio, an optimum point exists for designing an output device with a minimum propagation delay in a minimum area.

The data presented also give the design engineer a means of determining the minimum propagation delay for a given area and what tradeoffs are available when optimum conditions cannot be achieved. Also, if the capacitive load to node ratio and the maximum allowable delay are known for a given circuit, the designer can now determine the minimum chip area required to meet these constraints.

ACKNOWLEDGMENT

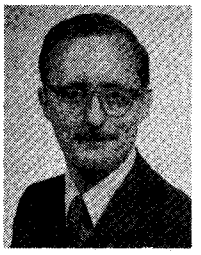
The authors wish to thank R. C. Jaeger for his perceptive review and suggestions.



Hung Chang Lin (S'47-A'52-M'56-SM'65-F'69) received the B.S. degree in electrical engineering from Chiaotung University, Shanghai, China, in 1941, the M.S. degree from the University of Michigan, Ann Arbor, in 1948, and the Doctor of Electrical Engineering from the Polytechnic Institute of Brooklyn, Brooklyn, N.Y., in 1956.

From 1941 to 1946, he worked in China as an Engineer at the Central Radio Works and Central Broadcasting Administration. From 1948 to 1956, he was a Research Engineer at RCA Laboratories engaged mostly in early transistor work. In 1956 he joined the Hytron Division of Columbia Broadcasting System where he worked until 1959 as Manager of the Semiconductor Applications Laboratory. Since July 1, 1959, he has been with Westinghouse Electric Corporation Baltimore, Md., working on integrated circuits in various capacities as Advisory Engineer at the Research and Development Laboratory, Manager of Advanced Development at the Molecular Electronics Division, and Consultant to the System Development Division. In the academic field, he has been a part-time Adjunct Professor at the University of Pittsburgh, Pittsburgh, Pa., Visiting Lecturer at the University of California, Berkeley, and is presently a Professor at the University of Maryland, College Park. He is the holder of 35 U.S. patents. Among his important inventions are the lateral transistor, widely used in linear integrated circuits, and the quasi-complementary output amplifier used in most high fidelity amplifiers. For his inventions he received three Westinghouse patent awards and the Technology utilization awards from the National Aeronautical and Space Administration. He has published some 60 technical articles, and is author of *Integrated Electronics*, and co-author of two other books.

Dr. Lin is a member of Sigma Xi and Phi Tau Phi. He has served as the Chairman of the IEEE Linear Integrated Circuits Task Group, as an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS, and as Technical Program Committee Member and/or Session Chairman in numerous International Solid-State Circuit Conferences and International Electron Devices Meetings.



Loren W. Linholm (S'67-M'68) was born in Calexico, Calif., on November 14, 1945. He received the B.S. degree in electronics engineering from the University of California, Berkeley, in 1968, and the M.S. degree in electrical engineering from the University of Maryland, College Park, in 1973.

From 1963 to 1968 he was employed by the Naval Missile Center, Ft. Mugu, Calif. Since 1968, he has been with the Department of Defense, Ft. Meade, Md. He has worked on input protection devices and MOS device prototyping. He is currently responsible for setting up an integrated circuit pilot line.

Mr. Linholm is a member of Tau Beta Pi and Eta Kappa Nu.