Gate Electrode *RC* Delay Effects in VLSI's

TAKAYASU SAKURAI, MEMBER IEEE, AND TETSUYA IIZUKA, MEMBER IEEE

***Abstract*—A poly-silicon gate electrode can be considered as a distributed *RC* line. The delay induced by this *RC* time constant can become a limitation in designing high-speed VLSI's. This effect, called the gate electrode *RC* delay effect (GERDE), is studied for short-channel MOSFET's. A simple formula is derived to roughly estimate the GERDE, which can be used as a rule-of-thumb in VLSI design. An approximation of the GERDE by a simple lumped-circuit model is also described. The future trends of the GERDE are investigated and it is concluded that the GERDE gets more severe for shorter channel MOSFET's, but, if the gate width is confined up to 30 μm, the GERDE can be neglected for MOSFET's with a channel length of more than 0.8 μm. For a large conductance, division of the MOSFET width is shown to be effective through experiments.**

I. Introduction

In VLSI's, parasitic resistance and capacitance give serious limitations in designing high-speed devices. This is basically because *RC* delay is not decreased as the design rule is scaled down, although the switching time of MOSFET's is decreased by the shrinkage, as is well known [1]. The effect of the source and drain resistance on the operation speed has been extensively investigated [2]. In addition to the effect, an *RC* time constant associated with a poly-silicon gate electrode may cause a significant parasitic delay. Sheet resistivity of a poly-silicon gate is several tens of ohms per square (about 30 Ω per square for the 1.2-μm rule). This resistance, together with the gate input capacitance of the order of several femto-farads per square micrometer, makes the poly-silicon gate electrode a distributed *RC* line. Very wide MOSFET's are frequently used for buffer outputs, bus drivers, and some logic stages. For these MOSFET's, the *RC* time constant amounts to more than five hundred ps, which is larger than the ideal switching time of the MOS gates.

This gate electrode *RC* delay effect (GERDE) has not been studied very extensively, yet. In 1975, Lin *et al.* first reported the effect [3]. Their interest is mainly centered on the frequency domain behavior of the GERDE, and no systematic treatment on the the transient characteristics is done except four simulations with L being around 10 μm and W being around 300 μm. The effect should be studied further in the following points. First, they are studying MOSFET's of 10-μm channel length, whereas current VLSI's/ULSI's use 1.2-μm MOSFET's [4]. The GERDE becomes more significant in shorter channel regions where the intrinsic gate delay is small, so that a reexamination of the effect is required. Secondly, Lin *et al.* are employing L ladder circuits for approximating a distributed *RC* line in computing transient response by a circuit simulator, which has been pointed out to be a poor approximation [5], [6]. Either T or π ladder circuits should be used instead of an L ladder. Thirdly, a simple formula for the GERDE is desirable for a VLSI designer to estimate the speed degradation induced by the GERDE. Approximation of the GERDE for circuit simulators is also important. Lastly, future trends of the GERDE are to be investigated. In order to answer these requests, this work is carried out by putting more stress on short-channel MOSFET's and easy-to-use results.

The basic model and calculation based on the model are described in Sections II and III, respectively. Measurements to assure the validity of the results are given in Section IV. In Section V, the GERDE in future ULSI's is considered, followed by conclusions in Section VI.

II. Basic Model

Since a gate electrode is considered as a distributed *RC* line, it can be approximated by n-step π or T ladder circuits [5], [6]. More concretely, a MOSFET with a large gate width is approximated as ladder circuits shown in Fig. 1(a) and (b). In the figure R_g and C_g designate total gate resistance and capacitance, respectively. The T ladder circuit consists of a T-shaped unit, each of which is made of two $R_g/2n$ resistors and one MOSFET with W/n channel width, being concatenated n times. In order to estimate the extra switching delay caused by the GERDE, the step response of the output node connected to a load capacitance C_0, is calculated. As is also shown in Fig. 1(a), the propagation delay t'_{pd} is defined as the discharging time from 0.9 to 0.1 V_{DD}, whereas t_{pd} is defined as an intrinsic or ideal propagation delay which can be calculated by setting gate resistance R_g equal to zero. A propagation delay in real devices is not exactly the same as t_{pd} defined like this, because input waveform of the gate is not necessarily a step voltage but has finite rising and falling time. VLSI designers, however, usually use the t_{pd} for the real propagation time delay in the early design stage and this approximation is, in most cases, tolerable. For this reason, this t_{pd} is used throughout the work to obtain a general understanding on the gate *RC* delay problem.

Manuscript received June 22, 1984; revised August 21, 1984.
The authors are with the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, 210 Japan.

0018-9200/85/0200-0290$01.00 ©1985 IEEE

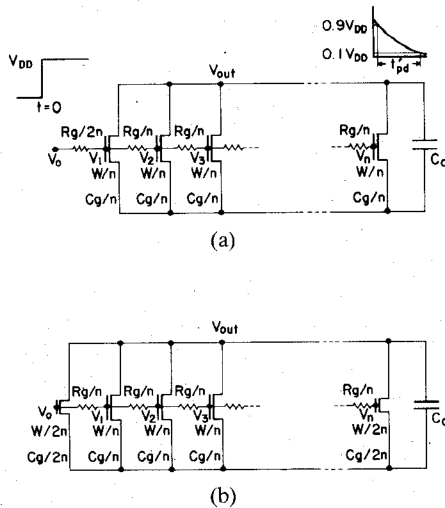

Fig. 1. (a) n-step T ladder model for very wide gate. Step voltage is incident at a gate contact. t'_{pd} is the discharging time of the load capacitor with gate electrode RC delay effects. (b) n-step π ladder model. W is a gate width, R_g total gate resistance, C_g total gate capacitance, V_{out} output voltage, and C_0 load capacitance.

The circuit equations to be solved for the π ladder model are as follows.

$$\frac{C_g}{n}\frac{dV_1}{dt}=\frac{V_0-V_1}{R_g/n}-\frac{V_1-V_2}{R_g/n}$$

$$\vdots$$

$$\frac{C_g}{n}\frac{dV_i}{dt}=\frac{V_{i-1}-V_i}{R_g/n}-\frac{V_i-V_{i+1}}{R_g/n}$$

$$\vdots$$

$$\frac{C_g}{2n}\frac{dV_n}{dt}=\frac{V_{n-1}-V_n}{R_g/n}$$

$$\vdots$$

$$C_0\frac{dV_{out}}{dt}=-\frac{1}{2}I_d(V_0)-\sum_{i=1}^{n-1}I_d(V_i)-\frac{1}{2}I_d(V_n)$$

where $I_d(V_i)$ is a drain current for gate input voltage V_i. Shockley's simple expression for the drain current [7] is employed here to estimate the general behavior of the GERDE. The explicit integration scheme failed to solve the resultant differential equations because of nonconvergence problems, whereas the trapezoidal rule is successfully applied with discretization error of less than 0.1 percent.

III. Results

Fig. 2 shows relative error as a function of ladder steps. The exact delay can be obtained only when the number of the ladder steps goes infinity. Here, however, due to practical limitations, the value for t'_{pd} is approximated as an average of π and T ladder models with nine ladder steps. The relative error of this approximation is considered less than 1 percent, as is shown in Fig. 2.

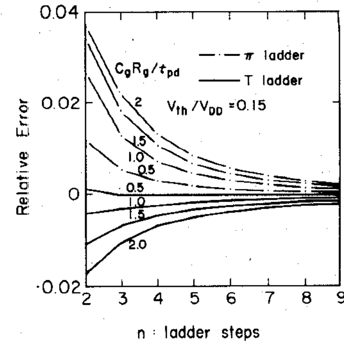

Fig. 2. Calculation error of n-step ladder models. t'_{pd} is approximated as an average of nine-step T and π ladder models.

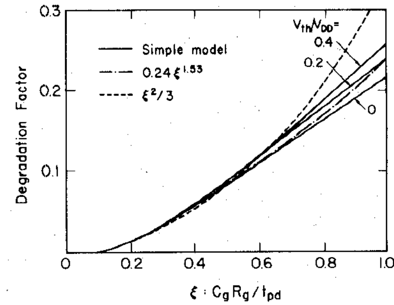

Fig. 3. Degradation factor (DF) versus normalized RC time constant of a gate. This graph is for $\xi<1$, with ξ being R_gC_g/t_{pd}. DF of 20 percent means that the switching time is delayed by 20 percent compared with an ideal gate structure where R_g is equal to zero. The broken line is a formula $DF=\xi^2/3$ which can be used to roughly estimate the DF. A more concise but complicated formula is $DF=0.24\xi^{1.53}$, a curve of which is also shown in the figure in a dash-dot line.

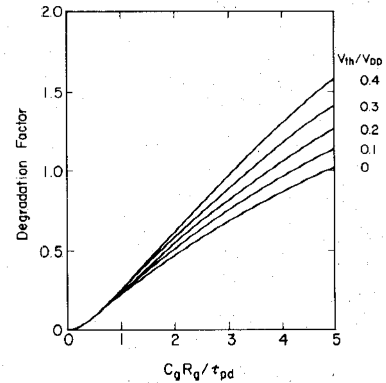

Fig. 4. Degradation factor (DF) versus normalized RC time constant of a gate. This graph is for $\xi<5$.

Figs. 3 and 4 are calculated results of the GERDE. The degradation factor (DF) is a delay increase by the GERDE compared with an ideal gate structure and is defined as

$$DF\equiv\frac{t'_{pd}-t_{pd}}{t_{pd}}.$$

Effects of the MOSFET conductance parameters and load capacitance can all be included in t_{pd} due to the structure of the basic differential equations. When the gate RC time

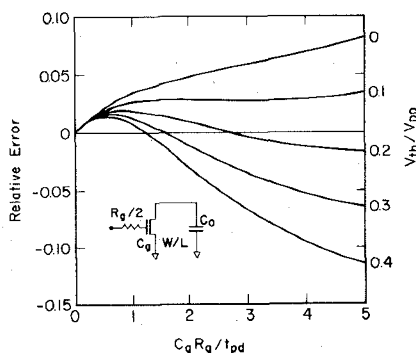

Fig. 5. Relative error of the one-step T ladder model. The model is shown in the figure and is the most simple circuit to introduce the GERDE in a circuit simulator.

constant, i.e., the product of R_g and C_g, is equal to the propagation delay of an ideal MOSFET, the DF amounts to about 25 percent, regardless of a threshold voltage. The degradation is rather small because the part of the MOSFET near a gate contact operates without any gate signal delay, even if the part far from the gate contact suffers from the RC delay effects. In this sense, the GERDE can be said to be a secondary effect. The DF depends weakly on a threshold voltage for usual situations of R_gC_g being less than t_{pd}. This result indicates that the exact $I_d(V_i)$ curve is not of importance and that the choice of the normalization factor of t_{pd} is effective to estimate the DF accurately.

The DF can be approximated by a simple formula

$$DF \sim \frac{1}{3}\left(\frac{R_gC_g}{t_{pd}}\right)^2, \quad \text{for } R_gC_g < t_{pd}. \tag{1}$$

It is convenient to use this formula to estimate the GERDE in the initial stage of VLSI design. It should be noted that, although the DF depends quadrically on the ratio R_gC_g/t_{pd} when $R_gC_g < t_{pd}$, the DF depends linearly on the ratio when $R_gC_g \gg t_{pd}$ as is shown in Fig. 4.

In the simulation stage, it is not cost effective to use nine-step ladder circuits to include the GERDE in a timing design. For this purpose, a one-step T ladder model, as illustrated in Fig. 5, is the most simple. It should be noted that the resistor is $R_g/2$ because of the T ladder model nature. The calculation error of this model is shown in the same figure. For poly-silicon gates with R_gC_g being less than t_{pd}, the relative error is within 4 percent, which is usually tolerable.

IV. Measurement

Two types of 21-stage ring oscillators, whose photomicrograph is shown in Fig. 6, are fabricated to observe the GERDE. Both of the test devices consist of inverters with W/L being 100 μm/1.2 μm for n-channel and 100 μm/1.6 μm for p-channel. One has a simple gate structure of straight 100 μm and the other has four divided gates of 25 μm $\times$ 4. Junction areas are designed to be the same for the two oscillators. Therefore, the latter has a gate RC time-

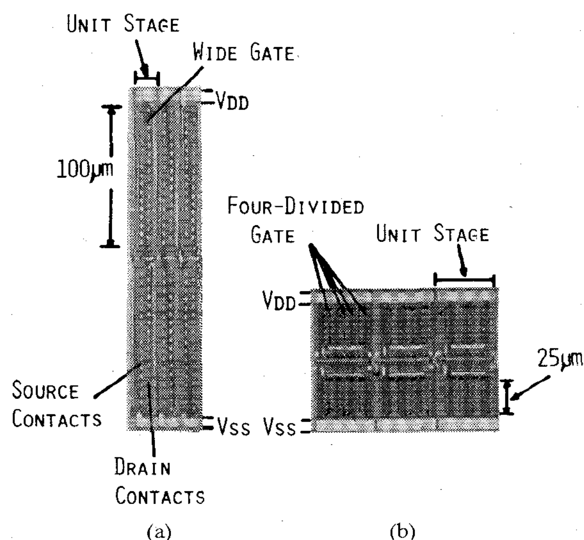

Fig. 6. Photomicrograph of test ring oscillator of 21-stage. Only three stages are shown in the photo. (a) Straightly spread gate structure of 100-μm width, and (b) four divided-gate structure to decrease R_gC_g time constant.

constant of 1/16 that of the former. According to (1), the DF decreases by a factor of 1/256 so that the DF is negligible for the latter.

An extra delay of 45 percent is observed for the straight wide gate at 5 V as, is shown in Fig. 7. In the same figure, the calculated results from the proposed model calculation and SPICE2 [8] simulation are also plotted. For n-channel and p-channel transistors, R_gC_g is equal to 450 and 440 ps, respectively, including two-dimensional effects [9]. The error amounts to 25 percent, if the two-dimensional effects are not taken into account in estimating the gate capacitance. In the simple model calculation, values for t_{pd} and R_gC_g are approximated as averages of p-channel and n-channel transistors. As seen from the figure, overall agreement is good and the validity to use the calculated DF is assured in roughly estimating the GERDE. The Miller capacitance effect is not taken into account in the simple model calculation, but this effect is not seen to give significant error to the results of the DF since a ratio of C_g and t_{pd} is of interest, and the Miller effect increases both quantities by the same factor. Besides, a lightly doped drain (LDD) structure widely used in recent miniaturized MOSFET's minimizes the Miller capacitance.

The effectiveness of the gate width division to reduce the GERDE is also seen from the figure. Measured parasitic drain and source resistance effect is also plotted. The source resistance gives the strongest degradation effect on the switching speed, but it is usually possible to minimize these resistances by carefully putting contacts near an intrinsic transistor.

V. Discussions

Fig. 8 shows the predicted degradation factor for shorter channel MOSFET's when straight gate width is kept constant. Fanout is assumed to be unity and an n-type channel

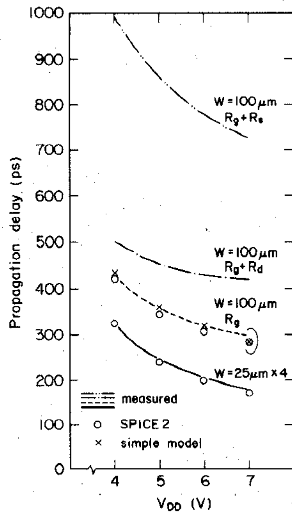

Fig. 7. Comparison among measurements and model calculation[1] for the GERDE. SPICE2 simulation is carried out by using nine-step π and T ladder models. The measured drain and source resistance effects by means of 21-stage ring oscillators are also shown in the figure. Lines designated as $R_d + R_g$ and $R_s + R_g$ are for drain and source resistance effect, respectively. For drain (source) resistance effect measurement, only one contact is patterned in the 100-μm drain (source) area, so that the total drain (source) resistance of 2.0 and 2.3KΩ is inserted for n-channel and p-channel MOSFET, respectively. Since these transistors also include the gate electrode RC delay effect, the notations $R_d + R_g$ and $R_s + R_g$ are used.

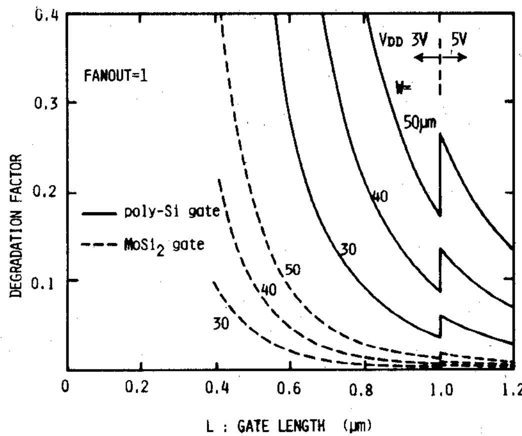

Fig. 8. Predicted degradation factor for shorter channel MOSFET's when the straight gate width is kept constant. The value of the electron mobility is calculated from the measured saturation region characteristics of MOSFET's with down to a 0.7-μm channel length. For the shorter channel length, the measured value is extrapolated.

is assumed, which is the worst condition for the GERDE. As seen from the figure, the GERDE puts more severe limitations on the choice of straight gate width for shorter channel transistors. This trend is reinforced by Fig. 9,

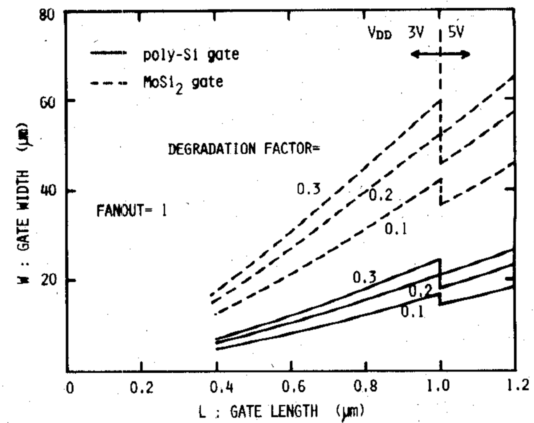

Fig. 9. The predicted maximum straight gate width for shorter channel MOSFET's when a certain degradation factor is admitted.

where maximum straight gate width is plotted, assuming that a certain degradation factor is admitted. If straight gate width, however is limited within 30 μm, the GERDE can be neglected except for submicrometer devices. It is to be noted that this criterion is for the most severe case. Therefore, if a VLSI is designed within this criterion, the timing error due to the GERDE will not occur. Generally speaking, the limit width is a function of a load capacitance and should be calculated using (2) for individual cases. Results for silicide gates are also shown in the figure. The sheet resistivity of the silicide gate is assumed to be 5 Ω/square, which is one sixth of that of the poly-silicon gate. The situation is better for the low-resistive gate material, but the GERDE still becomes a problem when a very short-channel MOSFET of less than a half-micrometer channel length is in use.

Here, let us consider the case where channel width W and length L and oxide thickness t_{ox} are scaled down simultaneously. If electron mobility is kept constant for a scaled-down transistor, t_{pd} is decreased proportionally to the channel length, whereas R_gC_g is essentially unchanged. As a result, the DF increases quadrically in small DF region according to the (1). Actually, the electron mobility is degraded and the DF is not so rapidly increasing but the DF still increases because the intrinsic switching delay decreases as device dimensions are scaled down.

VI. Conclusion

An RC delay effect caused by a poly-silicon gate electrode (GERDE) is studied for MOSFET's of channel length around 1 μm. As a result, the following points are made. The degradation factor (DF) by the GERDE can be approximated as

$$DF \sim \frac{1}{3}\left(\frac{R_gC_g}{t_{pd}}\right)^2, \quad \text{for} \quad R_gC_g < t_{pd}.$$

This simple formula can be used as a rule-of-thumb in the initial design stage. If the gate RC delay turns out to be

[1] In calculating the DF for the experimental ring oscillator with straight 100-μm gate width, t_{pd} is taken from the measured value for a 25 μm$\times$4 ring oscillator. After obtaining the DF, t'_{pd} for a 100-μm ring oscillator is calculated using again the measured value for a 25 μm$\times$4 ring oscillator. Therefore, there was no need for estimating t_{pd} of the 25 μm$\times$4 using the simple model. The simple model is used only for estimating the DF, but not used for estimating t_{pd} itself. But if we estimate the t_{pd} for the 25 μm$\times$4 using the simple model, we will get 165, 193, 239, and 292 ns for 7, 6, 5, and 4 V V_{dd}, respectively. In obtaining these values, the mobility value for saturation region is used. The errors between these values and the measured values are within 10 percent.

significant, division of the gate width is effective in reducing the delay. This is assured by the experiments. In the simulation stage, a one-step T ladder lumped circuit model, where only one resistor of $R_g/2$ is added to a gate, can be used to approximate the resistive gate electrode which is exactly a distributed RC line. The relative error of this approximation is less than 4 percent for $R_g/C_g < t_{pd}$.

The gate RC delay effect causes more severe influences on scaled-down MOSFET's. This is basically because RC delay is not decreased as the design rule is scaled down, although the switching time of MOSFET's is decreased by the shrinkage. If the straight gate width is limited within 30 μm, the gate electrode RC delay effect can be neglected, except for submicron devices. In order to mitigate the GERDE, low-resistive gate material such as $MoSi_2$ is effective, but the GERDE still remains a problem when short- and wide-channel MOSFET is in use.

Acknowledgment

The authors would like to express their thanks to S. Tanaka for useful suggestions, to Dr. K. Tamaru for helps in computer calculation, and to M. Isobe, T. Ohtani, K. Sawada, and A. Aono for their help throughout the work.

References

[1] T. Sakurai and T. Iizuka, "Double word line and bit line structure for VLSI RAM's," in *Proc. 15th Conf. Solid State Devices and Materials*, pp 269–272, 1983.
[2] K. Anami, M. Yoshimoto, H Shinohara, O Tomisawa, and T. Nakano, "Analysis of parasitic resistance effects in MOS LSI," *IECE of Japan*, vol J66-C, no. 9, pp 646–652, Sept. 1983.
[3] H. C. Lin, Y. F. Arzoumanian, J. L Halsor, M. N Giuliano, and H F. Benz, "Effect of silicon-gate resistance on the frequency response of MOS transistors," *IEEE Trans. Electron Devices*, vol. ED-22, pp 255–264, May 1975.
[4] T. Sakurai, J. Matsunaga, M. Isobe, T. Ohtani, K. Sawada, A. Aono, H. Nozawa, T. Iizuka, and S. Kohyama, "A low power 46ns 256Kbit CMOS static RAM with dynamic double word line," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 5, Oct. 1984.
[5] T. Sakurai, "Approximation of wiring delay in MOSFET LSI," *IEEE J. Solid-State Circuits*, vol. SC-18, no. 4, pp. 418–426, Aug. 1983.
[6] T Sakurai, T. Furuyama, and T Iizuka, "Methods for analyzing a word line delay and their applications," *The Community of Solid State Device of IECE of Japan*, vol. SSD82, p. 15, Oct. 1982.
[7] C. T Sah, "Characteristics of metal-oxide-semiconductor transistors," *IEEE Trans. Electron Devices*, vol. ED-11, pp 324–345, July 1964.
[8] L. W. Nagel, "A computer program to simulate semiconductor circuits," U. C. Berkeley Memo ERL-M520, May 1975.
[9] T Sakurai and K. Tamaru, "Simple formulas for two- and three-dimensional capacitances," *IEEE Trans. Electron Devices*, vol. ED-30, pp. 183–185, Feb. 1983.