

Design of Pre-Emphasis Pulses for Large Memory Arrays with Minimal Word-Line Delay Time

メタデータ	言語: eng 出版者: 公開日: 2019-05-09 キーワード (Ja): キーワード (En): 作成者: Matsuyama, Kazuki, Tanzawa, Toru メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00026465

Design of Pre-Emphasis Pulses for Large Memory Arrays with Minimal Word-Line Delay Time

Kazuki Matsuyama, and Toru Tanzawa, *Fellow, IEEE*
Shizuoka University, Hamamatsu 432-8011, Japan

Abstract— This paper formulates minimal word-line (WL) delay time with pre-emphasis pulses to design the pulse width as a function of the overdrive voltage for large memory arrays such as 3D NAND. The theory is validated with a nominal error of 5% in comparison with SPICE simulation for single WL line and three WL line models. The theory can take a finite series resistance of WL driver and decoding transistors into consideration as well. The impact of RC variation in WL and its compensation method are also discussed.

Index Terms—RC lines, delay time, pre-emphasis, NAND Flash, Flat Panel display, Word-line

NOMENCLATURE

T_{pre}	Pre-emphasis time
T_{opt}	Optimum T_{pre} to minimize the delay time
E	Target voltage
α	Ratio of the pre-emphasis voltage to E
β	Error rate to E
γ	Model dependent parameter
x	Delay line position ($x=0$ for the nearest, $x=l$ for the farthest)
r	Resistance per unit length
c	Capacitance per unit length
$e(x, t)$	Voltage at a position x and a time t
$i(x, t)$	Current at a position x and a time t
$E(x, s)$	Laplace transform of $e(x, t)$ with respect to t
$I(x, s)$	Laplace transform of $i(x, t)$ with respect to t
$t_{delay(\beta)}$	Time for $e(x, t)$ to reach βE
$t_{delay_{min}}$	Minimal time for the slowest node voltage to reach βE
R_d	Driver resistance or source resistance
N	Number of divisions of RC delay lines for circuit analysis

I. INTRODUCTION

Pre-emphasis pulses are widely used to reduce wire delay in integrated circuits (IC) design as illustrated in Fig. 1(a). In [1], the technique is applied to transmission line where the signal is attenuated at high speed. Programmable signal pre-emphasis reduces inter-symbol interference to achieve >1Gbps speed with

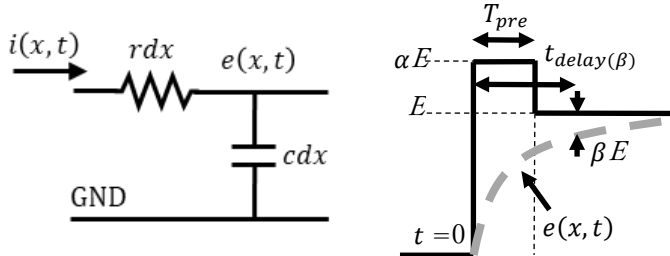


Fig. 2: Distributed element model

0.5um CMOS. The driving current of output buffers is controlled depending on previous output data. In [2], the pre-emphasis pulse is used for column drivers in flat-panel display with compensation against process variation. Prior to user mode, various pre-emphasis pulses are tested to find the best design parameters in the pulse for minimal column delay time. Thus, the optimal pre-emphasis pulse is used in user mode. In [3], pre-emphasis is used to reduce word-line (WL) set-up time for 3D NAND. One of the design challenges in 3D NAND is larger WL loading compared to that of planar devices [3], [4]. WL resistance is measured by using monitoring blocks. By applying the proper voltage and set-up time of a pre-emphasis pulse based on the measured average WL resistance, WL setup time can be minimized even with large process variation. Thus, pre-emphasis is a key design technique to minimize RC delay lines. However, wire delay time has been theoretically analyzed only with step pulses [5-7]. A general optimization method for the pre-emphasis pulses has not been formulated in literature to the best knowledge of authors.

In this study, we report an analytical expression for the pre-emphasis pulse shape to minimize the delay time and the energy-delay product, which can be applied to WL driver design in 3D NAND. In addition, the formulation expands to the case where RC time constant varies and a compensation method without calibration is proposed. More realistic circuit model is also discussed where the series resistance in WL driver and decoding transistors are not negligibly small.

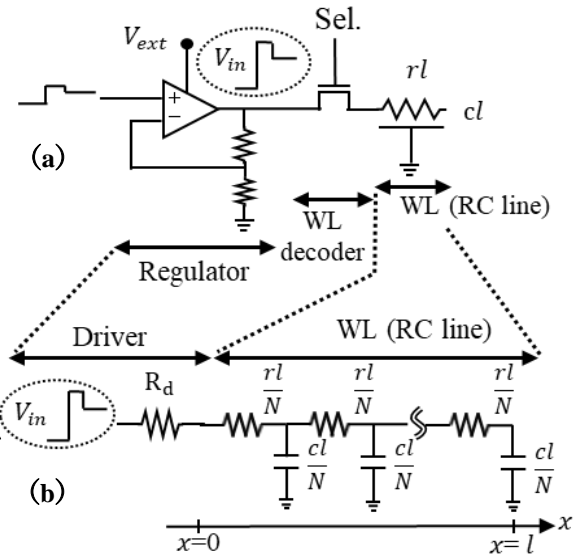


Fig. 1: Equivalent circuit of RC delay line

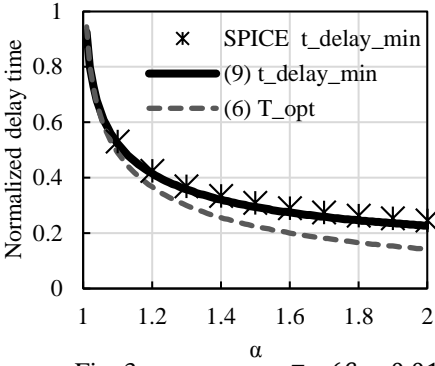


Fig. 3. $\alpha-t_{delay_min}, T_{opt}(\beta = 0.01)$

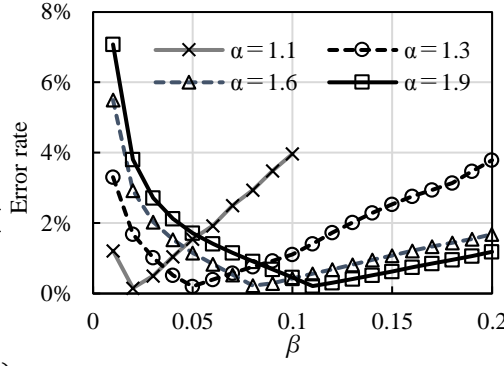


Fig. 4. Accuracy of (9) for various α and β

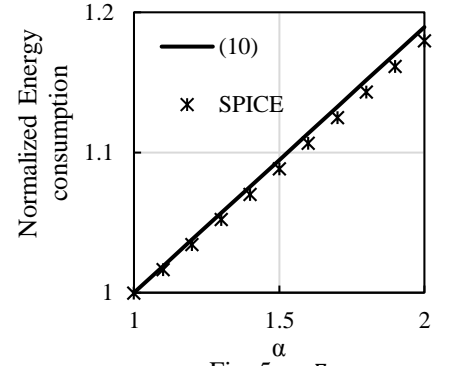


Fig. 5. $\alpha-E_{en}$

II. FORMULATION OF MINIMAL DELAY TIME

A distributed element model (1)(2) as shown by Fig. 2 can be exactly solved to be (3) for $t \leq T_{pre}$ when $e(x, 0) = i(x, 0) = 0, E(0, s) = \alpha E/s, I(l, s) = 0$. These initial and boundary conditions indicate the RC line is fully discharged at $t=0$, the input terminal is driven by a step pulse with $e(x, -0) = 0, e(x, +0) = \alpha E$, and the current at the farthest point ($x = l$) is 0 at any time.

$$-\frac{\partial e(x, t)}{\partial x} = ri(x, t) \quad (1)$$

$$-\frac{\partial i(x, t)}{\partial x} = c \frac{\partial e(x, t)}{\partial t} \quad (2)$$

$$e(x, t) = \alpha E - \frac{4\alpha E}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} e^{-\frac{(2k+1)^2}{\tau} t} \sin \frac{(2k+1)\pi x}{2l} \equiv e_{pre}(x, t) \quad (3)$$

where τ is a time constant given by $4rcl^2/\pi^2$. (1), (2) can be exactly solved to be (4) for $t > T_{pre}$ with $e(x, T_{pre}) = e_{pre}(x, T_{pre}), i(x, T_{pre}) = i_{pre}(x, T_{pre})$. ($i_{pre}(x, t)$ is not shown here.)

$$e(x, t) = E + \frac{4E}{\pi} \sum_{k=0}^{\infty} \left\{ (\alpha - 1) e^{\frac{(2k+1)^2}{\tau} T_{pre}} - \alpha \right\} \frac{1}{2k+1} e^{-\frac{(2k+1)^2}{\tau} t} \sin \frac{(2k+1)\pi x}{2l} \quad (4)$$

Let's determine the optimum T_{pre} (T_{opt}) which minimizes the delay time. For $t \gg \tau$ (4) can be approximated as (5)

$$e(x, t) \approx E + A_0(x, T_{pre})e^{-t/\tau} + A_1(x, T_{pre})e^{-9t/\tau} \quad (5)$$

where A_0 and A_1 are the proportional coefficients for the first two dominant factors. Because $e^{-9t/\tau} \ll e^{-t/\tau}$, one can approximately minimize the delay time with $A_0(x, T_{opt}) = 0$. Thus

$$T_{opt} \approx \tau \ln \frac{\alpha}{\alpha - 1} \quad (6)$$

(5) turns to be (7) with (6).

$$e(x, t) \approx E + A_1(x, T_{opt})e^{-9t/\tau} \quad (7)$$

An equation $E - e(x, t_{delay_beta}) = \beta E$ with (7) is solved to be (8)

$$t_{delay(\beta)} \approx \frac{\tau}{9} \ln \left| \frac{4}{3\pi} \left\{ \alpha \left(\frac{\alpha}{\alpha - 1} \right)^8 - \alpha \right\} \sin \frac{3\pi x}{2l} \right| \quad (8)$$

(8) is maximized at $x = l, l/3$. When $[\alpha/(\alpha - 1)]^8 \gg 1$, (8) turns to be (9), which provides the minimal delay time as a function of α and β .

$$t_{delay_min} \approx \frac{\tau}{9} \ln \left| \frac{4\alpha}{3\pi\beta} \left(\frac{\alpha}{\alpha - 1} \right)^8 \right| \quad (9)$$

As a result, one can easily determine T_{opt} and t_{delay_min} by using (6) and (9). Fig. 3 shows T_{opt} and t_{delay_min} as a function of α for $\beta=0.01$, which is normalized by t_{delay_min} with $\alpha = 1$ in case of a step pulse. By setting $\alpha = 2$, the delay time can be reduced to 1/4. Fig. 4 shows the error of (9) to SPICE simulation. (9) is in good agreement with SPICE simulation results within an error of 8% at most for $0.01 \leq \beta \leq 0.2$ and $1.1 \leq \alpha \leq 1.9$.

III. FORMULATION OF ENERGY CONSUMPTION AND ENERGY DELAY PRODUCT

When the pre-emphasis pulse is generated by a linear regulator as show in Fig. 1(a), the extra charge accumulated in WL is discharged to ground. Therefore, the energy consumption E_{en} is expressed by the product of the power supply voltage V_{ext} and the stored charge in WL at $t = T_{pre}$.

$$E_{en} = V_{ext} \times cl \times \overline{e(x, T_{opt})} \approx V_{ext} \times Ecl \left[\alpha - \frac{8}{\pi^2} (\alpha - 1) \right] \quad (10)$$

where $\overline{e(x, T_{opt})}$ is the average voltage across x . Fig. 5 shows energy consumption as a function of α . (10) agrees with SPICE simulation result within an error rate of 1% for $\alpha \leq 2$. A pre-emphasis pulse with $\alpha=2$ only increases the energy consumption by 20% in consumption with a step pulse. Energy delay product (ED) is used as measure to show a tradeoff between energy consumption and delay time. ED is expressed by

$$ED = t_{delay_min} \times E_{en} \quad (11)$$

using (9) and (10). Fig. 6 shows ED as a function of α . (11) agrees with SPICE simulation result within an error of 10% for $\alpha \leq 3$ Fig. 6 suggests that the energy delay product becomes the minimum of 0.25 when $\alpha = 2.86$. Even with $\alpha = 1.3$, one can reduce ED by 60%.

IV. IMPACT OF PROCESS VARIATION AND COMPENSATION METHOD

Fig. 7 shows t_{delay_min} normalized by that with $\alpha = 1$ and no RC variation. The delay time with RC variation of +20% is much longer than that of -20%. As described in section II,

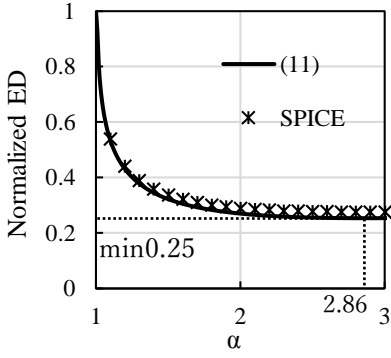


Fig. 6. $\alpha - ED$ ($\beta = 0.01$)

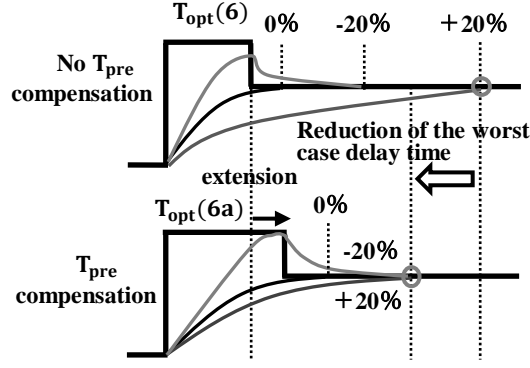


Fig. 8. compensation under RC variation

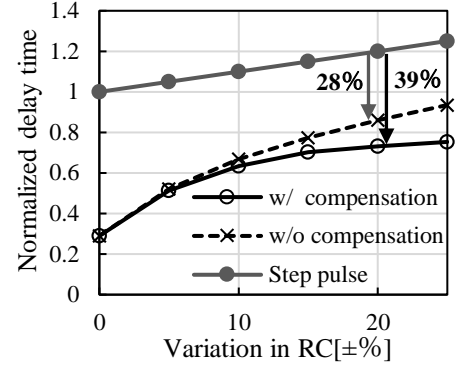


Fig. 10. The effect of compensation ($\alpha = 1.6, \beta = 0.01$)

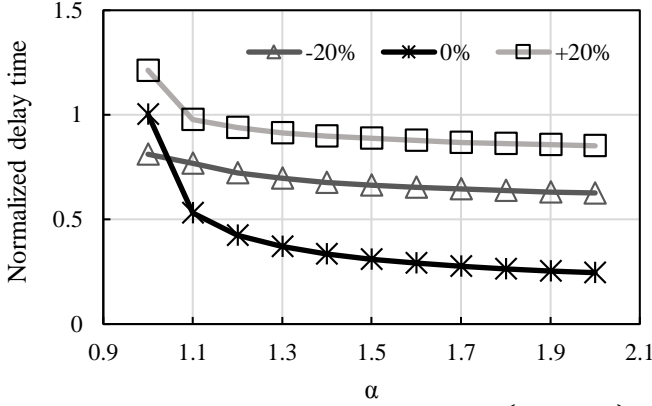


Fig. 7. Impact of process variation ($\beta = 0.01$)

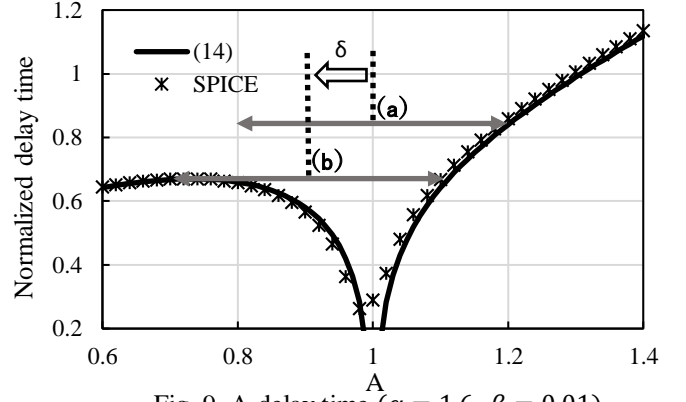


Fig. 9. A-delay time ($\alpha = 1.6, \beta = 0.01$)

t_{delay_min} can be reduced by 75% with a pre-emphasis pulse with T_{opt} in case of no RC variation. However, the improvement in reduction with pre-emphasis pulses becomes much less significant with RC variation of +20%. To compensate the impact of RC variation, we propose a method of RC variation aware pre-emphasis pulses, as illustrated in Fig. 8. The idea is that the delay time with RC variation of 20% can be matched with that of -20% by extending the pre-emphasis time a little. Even though the delay time with no RC variation becomes longer than before, the worst-case delay time can be reduced significantly. T_{opt} under RC variations can be obtained as follows. Let's take the first two terms in (4) for $x = l$.

$$e(l, t) \cong E + \frac{4E}{\pi} \left\{ (\alpha - 1) e^{\frac{T_{opt}}{\tau}} - \alpha \right\} e^{-\frac{t}{\tau}} \quad (12)$$

When RC is increased by a factor of A, (12) becomes (13)

$$e(l, t) \cong E + \frac{4E}{\pi} \left\{ (\alpha - 1) e^{\frac{T_{opt}}{A\tau}} - \alpha \right\} e^{-\frac{t}{A\tau}} \quad (13)$$

The solution for $E - e(x, t_{delay_min}) = \beta E$ with (6) and (13) is (14).

$$t_{delay_min} = A\tau \ln \left\{ \frac{4}{\pi\beta} \left[(\alpha - 1) \left(\frac{\alpha}{\alpha - 1} \right)^{\frac{1}{A}} - \alpha \right] \right\} \quad (14)$$

Fig. 9 compares (14) with SPICE simulation result. The error is very large around 1 due to the approximation, but sufficiently small for $A \leq 0.95$ and $A \geq 1.05$. The arrow δ indicates the method of compensation. The arrow (a) shows the case where the compensation is not performed with an error of $\pm 20\%$, whereas the arrow (b) shows the case where the position is shifted by δ so as to minimize the worst case delay time while

maintaining the length of the arrow (a). Once one identifies a value of δ by using Fig. 9, T_{opt} under RC variation can be determined by (6a)

$$T_{opt} = (1 + \delta)\tau \ln \frac{\alpha}{\alpha - 1} \quad (6a)$$

where R and C in τ are the values in case of no variation. One can figure out δ of 0.1 in the case of process variation in RC of $\pm 20\%$ as shown in Fig. 9. One only needs to increase T_{opt} by 10%. Fig. 10 shows the verification results. When the error is $\pm 20\%$, the maximum delay time can be reduced by 28% without compensation, whereas by 39% with compensation.

V. DESIGN FOR THREE RC LINES

In 3D NAND, the capacitance between adjacent WLs is dominant [3], [4], which can be modeled as three RC delay lines as shown in Fig. 11(a). Because of its symmetry the potential at P is the same as that of Q at any time when all three lines are fully discharged at $t = 0$. As a result, Fig. 11(a) can be reduced to Fig. 11(b). One can introduce an equi-potential line as shown in Fig. 11(c) with $C_1 = 1.5C$, $C_2 = 3C$ which makes the time constant for the target line equal to that for the adjacent one. Thus, the original circuit model of Fig. 11(a) is simplified as an equivalent circuit model as shown in Fig. 11(d). The minimal delay condition and minimal delay time based on Fig. 11(d) are calculated to be (15) and (16), respectively, with the same procedure as section II

$$T_{opt} = \gamma\tau \ln \frac{\alpha}{\alpha - 1} \quad (15)$$

$$t_{delay_min} \cong \frac{\gamma\tau}{9} \ln \left[\frac{4\alpha}{3\pi\beta} \left(\frac{\alpha}{\alpha - 1} \right)^8 \right] \quad (16)$$

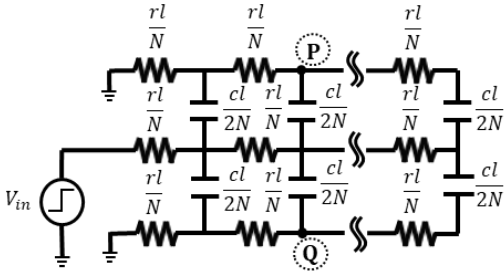


Fig. 11(a): Three RC delay lines

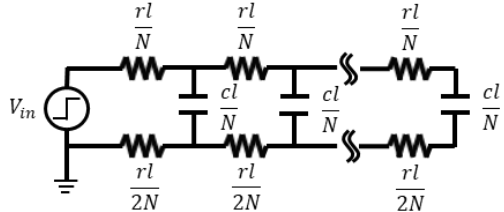


Fig. 11(b): Equivalent circuit of Fig. 11(a)

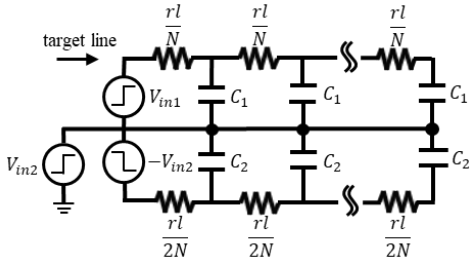


Fig. 11(c): Circuit converted from Fig. 11(b)

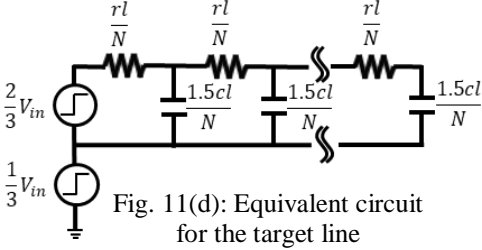


Fig. 11(d): Equivalent circuit for the target line

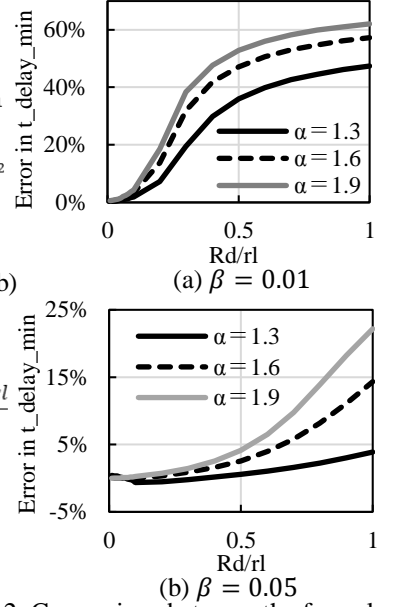


Fig. 12. Comparison between the formula (16) with (15), (18) and SPICE

γ is a model dependent coefficient which is 1.5 for the three RC line model and 1 for the single RC line model.

VI. DESIGN WITH DRIVER RESISTANCE

In NAND, WL is selected by a decoding transistor. The output resistance of the pre-emphasis driver can be designed to be sufficiently smaller than WL resistance. On the other hand, the on resistance R_d may not be sufficiently smaller than WL resistance because the size of the decoding transistor is limited according to the WL pitch. Therefore, in this section, we consider the path resistance for pre-emphasis pulse design. The idea of Elmore delay [5] can be adapted. Elmore delay is obtained by adding the products of the capacitance and the resistance from the power supply at each node. When the time constant τ_d at the farthest node of Fig. 1(b) is calculated with $N = \infty$ to be (17).

$$\tau_d = R_d cl + \frac{1}{2} rl cl \quad (17)$$

Since the worst case delay time is at the far end, γ can be estimated by taking the ratio with the time constant τ_0 at the farthest node of a single WL delay line with $R_d = 0$, i.e.,

$$\gamma = \frac{\tau_d}{\tau_0} = 1 + \frac{2R_d}{rl} \quad (18)$$

Fig. 12 shows an error of the formula (16) with (15) and (18) and the SPICE simulation results. The delay time can be calculated within 5% error for $R_d/rl < 0.1$, $\beta = 0.01$ and $R_d/rl < 0.5$, $\beta = 0.05$. Table 1 shows the main results of this work on the optimum pre-emphasis pulse width T_{opt} and the minimal WL delay time t_{delay_min} with a model-dependent parameter γ .

VII. CONCLUSION

In this research, we formulated a pre-emphasis technique to reduce WL delay time for memories with large arrays such as 3D NAND. Circuit designers can easily estimate the minimum delay condition (15), delay time (16) and energy consumption (11) at initial design phases. The impact of process variation on the delay time was also analyzed with (14). We proposed a method to reduce the delay time under process variation in WL

Table 1: Main results of this work

	Single WL model w/ $R_d=0$	Three WL model w/ $R_d=0$	Single WL model w/ $R_d > 0$
γ	1	1.5	$1 + \frac{2R_d}{rl}$ (18)
T_{opt}	$\gamma \tau \ln \frac{\alpha}{\alpha - 1}$ (15)		
t_{delay_min}	$\frac{\gamma \tau}{9} \ln \left[\frac{4\alpha}{3\pi\beta} \left(\frac{\alpha}{\alpha - 1} \right)^8 \right]$ (16)		

RC. We further derived the calculation method for three WL line model.

ACKNOWLEDGMENT

The authors would like to thank VDEC, Synopsys, Inc., Cadence Design Systems, Inc., Rohm Corp. and Micron Foundation for their support.

REFERENCES

- [1] A. Fiedler et al., "A 1.0625-Gb/s transceiver with 2 oversampling and transmit signal pre-emphasis," ISSCC, pp. 238–239, Feb. 1997.
- [2] J. Bang, et al., "A Load-Aware Pre-Emphasis Column Driver with 27% Settling-Time Reduction in $\pm 18\%$ Panel-Load RC Delay Variation for 240Hz UHD Flat-Panel Displays" ISSCC, 11.7, 2016.
- [3] W. Jeong et al., "A 128 Gb 3b/cell V-NAND Flash Memory With 1 Gb/s I/O Rate" *IEEE Journal of Solid-State Circuits*, Vol. 51, No. 1, pp. 204 – 212, Jan. 2016.
- [4] T. Tanzawa et al., "Design Challenge in 3D NAND Technology: a 4.8X Area- and 1.3X Power-Efficient 20V Charge Pump Using Tier Capacitors," *IEEE Asian Solid-State Circuits Conference*, Nov. 2016.
- [5] W.C. Elmore, "The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers", *J. Applied Physics*, Vol. 19, No. 1, pp. 55-63, 1948.

This is the final version of the paper submitted to IEEE ISCAS 2019.

[6] T. Sakurai, "Approximation of wiring delay in MOSFET LSI," *IEEE J. of Solid-State Circuits*, Vol. SC-18, No. 4, pp. 418–426, Aug. 1983.

[7] T. Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSI's," *IEEE Trans. on Electron Devices*, Vol. 40, No. 1, pp. 118–124, Jan. 1993.