On Dynamic Delay and Repeater Insertion[1]

Hannu Tenhunen and Dinesh Pamunuwa

Royal Institute of Technology (KTH), IMIT, LECS
Electrum 229, SE-164 40 Kista, Sweden
dinesh/hannu@ele.kth.se

ABSTRACT

In deep sub-micron technologies, as the wires are placed ever closer and signal rise and fall times go into the sub-nano second region, increased cross talk has implications on the data throughput and on signal integrity. Depending on the data correlation on the coupled lines, the delay can either decrease or increase. Here we show that in uniform coupled lines, the response for several important switching configurations has a dominant pole characteristic. This allows easy prediction for the average, worst-case and best-case delay of buffered lines. We show that the repeater numbering and sizing can be optimised to deal with cross-talk under different constraints to best match the application. Area and power issues are considered and all equations are checked against a dynamic circuit simulator (SPECTRE).

1. INTRODUCTION

In the future generation of VLSI circuits when the feature size shrinks to a fraction of a micro meter, the aspect ratio (width/height) of interconnect is reduced in order to keep the resistance increase to a minimum. This means the capacitance between wires increases, and cross talk - which couples a noise voltage onto the victim net, and has an effect on the delay- poses a serious challenge in designing VLSI systems. Our interest in this paper is in cross-talk induced delay, and further in a parallel line configuration, where the nets are laid out alongside each other for a relatively long distance as would occur in an intermediate or global level bus. Recently there has been a profusion of research into block oriented architectures [1] with different modules communicating via buses and the parallel net topology in Fig. 1 will occur very often.

Capacitive coupling can result in speeding up of the signal or cause delay- depending on the correlation between the data on the different lines. This input dependent dynamic delay can exactly be captured only by dynamic simulators. However when the line under consideration is reduced to a uniformly coupled two aggressor configuration as shown in Fig. 1, certain simplifications are possible which allow delay predictions depending on the switching of the aggressors. There have been previous works which have distributed the capacitance over ground and coupled components and presented closed form delay equations with various switching configurations. However these use a single T or Π section, which does not represent a distributed line with reasonable accuracy. We shall show that for uniformly coupled parallel nets, when the aggressors switch simultaneously in a variety of ways, the response of the victim line has a dominant pole characteristic. This allows the delay to be modelled by a single time constant, with a changing coefficient giving measures of average, worst-case and best-case delays with over 90% accuracy. This analysis is extended to buffered lines, where we give closed form equations which quantify the effect that repeater sizing and numbering have on the delay for different switching patterns. Finally we show how these expressions facilitate repeater optimisation under different constraints.

2. DELAY MODELLING

From now on, whenever delay is mentioned we are always talking about the 50% delay, since this corresponds to the delay of the output to the switching threshold of an inverter. Also in all cases the victim line is assumed to switch from zero to one, without loss of generality. When a line switches up(down) from zero(one) it is assumed to have been zero(one) for a long time. We consider a line with coupling on two sides as shown in Fig. 1. To build up our delay model for the distributed line, we analyse first the lumped model which consists of a single section. For simultaneously switching lines, six different switching scenarios can be identified.

(a) Both aggressors switch from one to zero
(b) One switches from one to zero, the other is quiet
(c) Both are quiet
(d) One switches from one to zero, the other switches from zero to one
(e) One switches from zero to one, the other is quiet

1. The funding support of Sida and that of Vinnova via the Socware and Exsite Programs are gratefully acknowledged

0-7803-7448-7/02/$17.00 ©2002 IEEE

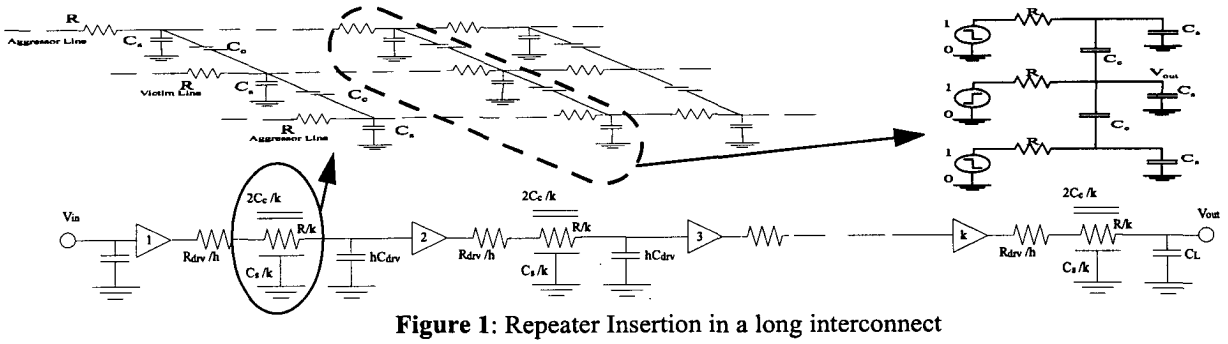


Figure 1: Repeater Insertion in a long interconnect

(f) Both switch from zero to one

Consider (c) above as the reference delay, where the driver of the victim line charges the entire capacitance. Cases (a) and (b) slow down the victim line, (d) is equivalent to (c), and (e) and (f) speed up the victim. Now given in (1) is the complete response of the victim line where the coefficients A_i and B_i take the values given in Table 1 depending on how the aggressor lines switch.

Table 1. Coefficients for different switching configurations

i	Switching Configuration	A_i	B_i	λ_i	k_i
1	(a)	-4/3	1/3	1.51	2.20
2	(b)	1	0	1.13	1.50
3	(c)	-2/3	-1/3	0.57	0.65
4	(d)	-2/3	-1/3	0.57	0.65
5	(e)	-1/3	-2/3	--	--
6	(f)	0	1	0	0

In cases (b) and (f), the response is a single decaying exponential, while in the other cases the slow or dominant time constant is $R(C_s+3C_c)$. In cases (a), (c) and (d), the slower time constant is also associated with the larger coefficient, and hence the faster time constant can be neglected with good accuracy in the delay. This is especially so in case (a). Now to state some well known results, a lumped RC circuit with no aggressors has a single pole response and the delay is as given in (2). Signal propagation along a distributed RC line is governed by the diffusion equation which does not lend itself readily to closed form predictions for the delay at a given threshold. However it turns out that a simple exponential is a very good predictor [2] which leads to (3) as the model for the 50% delay of a distributed RC line to a step input. This is a very good approximation and is reputed to be accurate to within 4% for a very wide range of R and C.

$$V = 1 + A_i e^{-\frac{t}{R(C_s+3C_c)}} + B_i e^{-\frac{t}{RC_s}} \quad (1)$$

$$T_{0.5,\,lumped} = 0.7RC \quad (2) \qquad T_{0.5,\,distr} = 0.4RC \quad (3)$$

$$T_{0.5,\,distr} = 0.4RC_s + \lambda_i RC_c \quad (4)$$

For the kinds of RC lines shown, whenever the response of the lumped model corresponding to a single section of the distributed line is or can be approximated by a waveform containing a single exponential, the response of the distributed line can also be approximated by a waveform with a single exponential. Hence we propose to model the delay of the *distributed* lines corresponding to (a), (b), (c), (d) and (f) with single time constant expressions. (In the case of (e) the accuracy is not high enough to justify such an approach because the lumped model does not have a dominant time constant). Since the time constants in question are linear combinations of R, C_s and C_c, changing coefficients are sufficient to distinguish between the different cases. The delay is as given in (4) where λ_i take the values in Table 1. These constants were obtained by running sweeps with the circuit analyser SPECTRE. For all i, the accuracy is more than 93% for a wide range of R, C_s and C_c values. In the interest of brevity, only a representative subset of the values for i=1, which is of special interest, is shown here in Table 2.

3. REPEATER INSERTION

To reduce delay the long lines in Fig. 1 are broken up into shorter sections, with a repeater (an inverter) driving each section as shown in Fig. 2. The analysis for repeater insertion is carried out by characterizing the non-linear buffers by an output resistance and input capacitance. Let the number of repeaters including the original driver be k, and the size of each repeater be h times a minimum sized inverter (all lines are buffered in a similar fashion). The output impedance of a minimum sized inverter for the particular technology is $R_{drv,m}$ and the output capacitance $C_{drv,m}$. Then the output impedance of an h sized driver is assumed to be $R_{drv,m}/h$, and the output capacitance $h \times C_{drv,m}$.

Table 2. Comparison of simulated and predicted delay for a distributed RC line with worst-case cross talk

R (ohms)	Cs (fF)	Cc (fF)	Td (simulated) (fs)	Td (model) (fs)	Error percentage (%)
10	1	10	153.8	154	-0.2%
10	100	1	403	415	-2.8%
10	100	10	546	550	-0.8%
100	1	1	197	190	3.7%
100	1	10	1537	1540	-0.1%
100	10	10	1984	1900	4.2%
200	10	30	9938	9800	1.4%
300	30	10	8393	8100	3.5%
300	30	20	13222	12600	4.8%
300	30	30	17850	17100	4.2%

Now with reference to Fig. 1 and using superposition with the delay equations (2, 3 and 4) the total delay for a line takes the expression given in (5). This expression follows the Bakoglu model [3] of equalising the repeaters, and can be explained as follows. The distributed and lumped resistances combine with the distributed and lumped capacitances to produce various delay terms. The terms in bold are the result of modelling cross talk in the delay. λ_i and μ_i take the values given in Table 1, where μ_i is a coefficient introduced to take the Miller effect into account.[1] It is assumed that the load C_L is equal to the input capacitance of an h sized inverter. Also the signal rise time has been included here. Because in general the delay per section is much greater than half the rise time, the non-zero rise(fall) time of the input signal is approximated in (5) as a simple addition. Hence the fact that the entire analysis is based on step inputs does not cause grave drops in the accuracy of the final expressions. This is ever more true for future generations of technologies where decreasing feature sizes allow transistors to be gated with faster signals, but also cause wire parasitics to become more pathological. This delay expression was checked against simulated values,

$$t_{0.5} = k\left[0.7\frac{R_{drv_m}}{h}\left(\frac{C_s}{k} + hC_{drv_m} + \mu_i\frac{\mathbf{2C_c}}{k}\right) + \frac{R}{k}\left(0.4\frac{C_s}{k} + \lambda_i\frac{\mathbf{C_c}}{k} + 0.7hC_{drv_m}\right)\right] + \frac{t_r}{2} \quad (5)$$

$$k_{i,opt} = \sqrt{\frac{0.4RC_s + \lambda_i RC_c}{0.7R_{drv_m}C_{drv_m}}} \quad (6) \qquad h_{i,opt} = \sqrt{\frac{0.7R_{drv_m}C_s + 1.4\mu_i R_{drv_m}C_c}{0.7RC_{drv_m}}} \quad (7)$$

1. Because of the approximate models used for the delay, the final accuracy is improved if the Miller coefficients take non-integer values as shown.

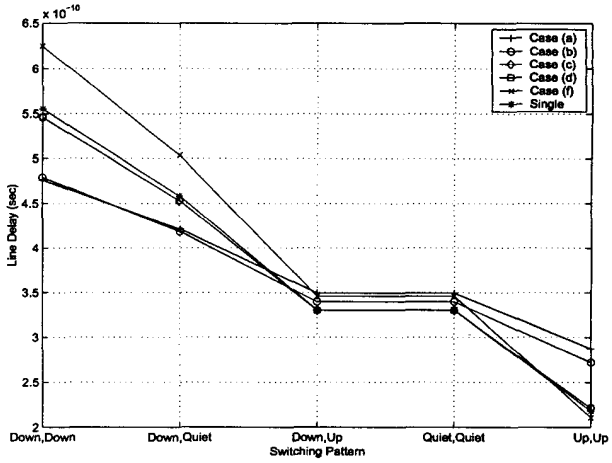


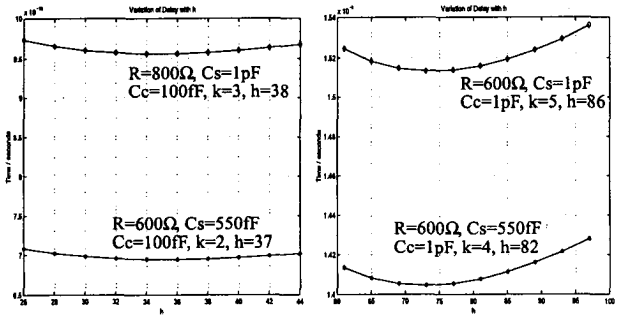
Figure 2: Delays for different repeater insertion strategies

and the accuracy found to be limited only by the accuracy of the initial expression (4).

To find the optimum h and k for minimising delay, the partial derivatives of (5) with respect to k and h are equated to zero, resulting in (6) and (7). Case (a) is of special significance because it represents the worst-case cross-talk of all the cases considered (the delay for this pattern is only 1 or 2% less than the worst-case delay caused by non-simultaneously switching aggressors). A repeater insertion strategy that is optimised for a certain pattern will not be optimal for other patterns, and of interest is exactly how it performs. Given in Fig. 2 are the delays for different patterns, when the repeater insertion strategies are optimised for cases (a) through (f), excepting (e). The net considered here has a resistance of 1k Ω and capacitances of 100f F to ground and to each of the adjacent wires. R_{drv} and C_{drv} are set to 7.7k Ω and 9.5f F to match the 0.35 mm technology we use for testing. The legend termed *single* refers to the conventional optimisation strategy that would be carried out by treating the total capacitance as a single lumped component. Obviously for each switching pattern, the delay is minimum for the h and k that is optimised for that particular pattern. What is interesting here is that pattern (a) always causes the maximum delay (hence defining the maximum bit frequency over the line as the worst-case has to be expected in general), and this can be reduced by a repeater insertion strategy that is more aggressive than would be predicted as the optimal by a conventional analysis. By inspecting the optimal k and h values for the different switching patterns and considering the delay constraints and the available resources for repeater insertion, the k and h values that best suits ones application can be selected.

To check the accuracy of our models we ran simulations for transistor models in an actual 0.35 μm technology

where $R_{drv,m}$ and $C_{drv,m}$ take the values given above. Shown in Fig. 3 are the results of simulations for a range of h situated either side of the value predicted by (6), where the k and h values associated with each graph refer to $k_{l,opt}$ and $h_{l,opt}$. It can be seen that the fidelity of (6) and (7) are quite good.

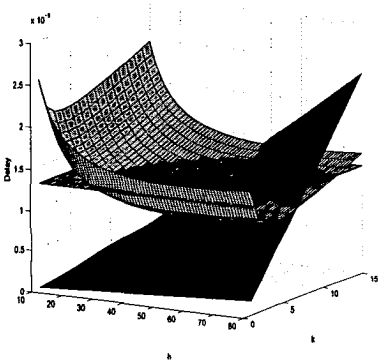


a: Examples with two and three Repeaters

b: Examples with four and five Repeaters

Figure 3: Effect of Repeater Sizing on Delay

4. AREA AND POWER



igure 4: Delay constraint matching for Net in row 1, Tab. 4.

Minimising power consumption is equivalent to minimising area, or the product hk. When the delay is equalised over each line segment, the problem of repeater optimisation can take two forms. Either the maximum acceptable delay for the net is specified, and the objective is to minimise hk subject to the constraint $t \le t_{max}$, or the maximum acceptable area is specified and the objective is to minimise the delay subject to the constraint $A \le A_{max}$. Consider Fig. 4 which shows the variation of delay with h and k where the line parasitics are defined by R=800Ω and C_s=C_c=550fF. The plane shows a delay constraint of 1.3n seconds for that net, and any of the k and h combinations which lie below this and on the curved surface showing the delay is acceptable to meet the delay constraint. Also shown is an appropriately scaled plot of hk. Because hk is quasi concave in the quadrant of positive h and k, it is not possible to find an analytical solution to the first optimisation problem, which has to be solved numerically. The solution to the second optimisation problem is obtained by solving the Kuhn Tucker conditions [4] given in (8) through (12) where L_i refer to the Lagrangian constants. The coefficients corresponding to case (a) have been used as the worst-case needs to be considered.

$$0.7\frac{R_{drv}}{h^2}(C_s + 4.4C_c) - 0.7RC_{drv} + L_1 k + L_2 = 0 \quad (8)$$

$$\frac{R}{k^2}(0.4C_s + 1.5C_c) - 0.7R_{drv}C_{drv} + L_1 h + L_3 = 0 \quad (9)$$

$$L_1(hk - A_{max}) = 0 \quad hk \le A_{max} \quad L_1 \ge 0 \quad (10)$$

$$L_2(h-1) = 0 \quad h \ge 1 \quad L_2 \ge 0 \quad (11)$$

$$L_3(k-1) = 0 \quad k \ge 1 \quad L_3 \ge 0 \quad (12)$$

5. CONCLUSIONS

In this paper we have investigated the issue of dynamic delay in buffered lines and shown that distributing the capacitance into two components as we have proposed allows the effect of switching aggressors in a buffered net to be quantified in simple equations. The optimal k and h values that minimise delay for any given switching pattern were then derived. All these expressions give the designer more information about when and how to insert repeaters in long nets and are proposed as being suitable for static timing tools. The closed form nature of the equations allow iterations to be made much more cheaply than with a dynamic simulator. For all patterns, when the coupling capacitance term C_c is set to zero (i.e. total capacitance is lumped into the term C_s), the equations describing h and k simplify to the Bakoglu equations [3]. Hence we have proposed a simple yet accurate way of distributing the capacitance and including the effect of switching aggressors.

6. REFERENCES

[1] D. Sylvester and K. Keutzer, "Getting to the bottom of deep submicron II: a global wiring paradigm", in *Proc. ISPD*, 1999, pp. 193-200.

[2] J. Rubinstein, P. Penfield and M. Horowitz "Signal delay in RC tree networks", *IEEE Trans. Computer Aided Design*, vol CAD-2, no. 3, pp. 202-211, July 1983.

[3] H. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Reading, MA: Addison Wesley 1990

[4] S. Dar and M. Franklin, "Optimum buffer circuits for driving long uniform lines", *IEEE J. Solid State Circuits*, vol. 26, pp. 32-40, Jan. 1991.

[5] Y. Ismail and E. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits", *IEEE Trans. VLSI Systems*, April 2000, vol. 8, pp. 195-206.