Analysis of On-Chip Inductance Effects using a Novel Performance Optimization Methodology for Distributed *RLC* Interconnects

Kaustav Banerjee
Center for Integrated Systems
Stanford University
Stanford CA 94305
kaustav@stanford.edu

Amit Mehrotra
Computer and Systems Research Laboratory
University of Illinois at Urbana-Champaign
Urbana IL 61801
amehrotr@uiuc.edu

ABSTRACT

This work presents a new and computationally efficient performance optimization technique for distributed *RLC* interconnects based on a rigorous delay computation scheme. The new optimization technique has been employed to analyze the impact of line inductance on the circuit behaviour and to illustrate the implications of technology scaling on wire inductance. It is shown that reduction in the driver capacitance and output resistance with scaling makes deep submicron (DSM) designs increasingly susceptible to inductance effects. Also, the impact of inductance variations on performance has been quantified. Additionally, the impact of the wire inductance on catastrophic logic failures and IC reliability issues have been analyzed.

1. INTRODUCTION

1.1 Inductance Effects in DSM Interconnects

For deep submicron interconnects on-chip inductive effects arising due to increasing clock speeds, increasing interconnect lengths and decreasing signal rise times are a concern for signal integrity and overall interconnect performance [1, 2]. Inductance causes overshoots and undershoots in the signal waveforms, which can adversely affect signal integrity. For global wires inductance effects are more severe due to the lower resistance of these lines, which makes the reactive component of the wire impedance comparable to the resistive component, and also due to the presence of significant mutual inductive coupling between wires resulting from longer current return paths [3]. Furthermore, since the global wires are the farthest away from the substrate, they are most susceptible to large variations in the current return path and therefore large variations in the inductance. With the recent adoption of Copper as the interconnect metal [4, 5], line resistances have decreased further and as a result, inductive effects have become more prominent. Hence the traditional lumped or distributed *RC* model of the interconnects, specially of the global wires, is no longer adequate since it can result in substantial errors in predicting both delay and crosstalk [6].

Line inductance affects the circuit performance in two distinct ways. Firstly, it can affect the rise/fall time and signal delay through an interconnect. Traditional delay models of interconnects are based on Elmore delay [7] which does not take into account the inductance of the interconnect. If the line inductance is "small enough", the step response is very similar to the step response obtained by ignoring the inductance and therefore Elmore delay predictions are accurate. However, as the line inductance increases beyond a certain value, the actual delay and Elmore delay diverge and one needs to compute signal delay by accurately modelling line inductance.

Secondly, a VLSI interconnect can be viewed as a lossy distributed *RLC* transmission line with a characteristic impedance of $Z_0 = \sqrt{(r+sl)/(sc)}$ where r, l and c are the line resistance, inductance and capacitance per unit length respectively and s is the complex frequency ($j\omega$). If the series output impedance of the driver and the input impedance of the receiver are equal to Z_0 then according to the transmission line theory, there are no reflections present in the system. However in a practical VLSI circuit, the load is almost exclusively capacitive. Also, the driver size is typically optimized for delay minimization and it series impedance may not necessarily be equal to Z_0. Therefore in such systems, line inductance can give rise to reflections which result in overshoots and undershoots in voltage waveforms. Voltage overshoot may cause reliability concerns in the circuit whereas undershoot will, in best case, cause glitches and, in worst case, cause false transitions at the output of a gate. Glitches increase the dynamic power dissipation while false transitions can cause logic errors and severe timing violations.

In the past a lot of research effort has been devoted to the areas of inductance computation [8, 9, 2], inductance extraction using both numerical and experimental techniques [2, 10, 11, 12, 13, 14, 15, 16] and modelling of on-chip inductance [2, 17, 18, 19] in integrated circuits. However, accurate inductance modelling still remains a challenging problem. This is due to the fact that magnetic fields have much longer spatial range compared to that of electric fields and therefore, in practical high-performance ICs containing several layers of densely packed interconnects the wire inductances are sensitive to even distant variations in the interconnect topology and switching activity [12]. Moreover, uncertainties in the termination of neighbouring wires can significantly affect the signal return path and also the return current distributions, and therefore the effective loop inductance and resistance. Additionally, the series resistance of the conductor affects the transient behaviour of the on-chip current distribution [20]. Although the effective wire inductances in complex 3D interconnect structures can be obtained by rigorous electromagnetic field solvers [10, 12], the results are at best approximate for real high-performance circuits due to the uncertainties in providing valid models of the local physical and electromagnetic environment formed by the orthogonal and parallel interconnects. Also, accurate estimation of effective inductance values not only requires details of the 3D interconnect geometry and layout, technology information such as metal resistivity, insulator dielectric constant etc., but also accurate model for the current distributions and switching activities of the wires, which are difficult to predict a priori.

As pointed out earlier, line inductance can cause a significant change in the interconnect delay and the effect of line inductance needs to be taken into account in the optimization of repeater insertion. There have been some recent work that suggest an optimum buffer insertion scheme [21, 22] which also takes inductance into account. However, their approach is based on optimizing the interconnect delay which is calculated using an empirical equation which has been curve-fitted using circuit simulation results and has a limited range of validity. It is also not clear how the delay varies as the line inductance changes which depends on input signal pattern and therefore cannot be easily predicted a priori as discussed earlier. Furthermore, it is not clear how does the effect of line inductance change

DAC 2001 June 18-22, 2001, Las Vegas, Nevada, USA

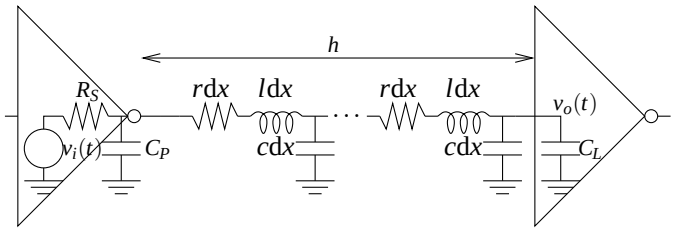

Figure 1: Driver-interconnect-load structure.

which scaling. Given that minimum and maximum line inductances per unit length do not change appreciably with scaling, it is not clear as to what factors are responsible for increased inductance effects.

1.2 Scope of This Study

In this work we first derive the transfer function and the time-domain response of a realistic driver-interconnect-load configuration based on a rigorous analysis of the distributed structure. Using these results we develop a new methodology for optimum repeater insertion for a distributed *RLC* interconnect based on the solution of the general delay equation for a repeater driving an interconnect with a load capacitance. Unlike previous such attempts [21] which are based on curve-fitted delay equations from circuit simulations which have limited validity, our approach, which is described in Section 2, is based on the analytical minimization of interconnect delay per unit length. We show that optimum repeater sizes and interconnect lengths can be efficiently computed for given technology and interconnect parameters. In Section 3 we use this methodology to compute the optimum buffer sizes and interconnect lengths for a wide range of line inductances found in global interconnects for 250 nm and 100 nm technology nodes. We use these results to demonstrate that reduction in minimum-sized driver capacitance and output resistance with device scaling is primarily responsible for increasing susceptibility of VLSI designs to inductance effects. We also show that excessive overshoot and undershoot in voltage waveforms due to inductance can cause catastrophic circuit failures but only circuits at 100 nm technology node can be susceptible to such failures for practical values of line inductances.

2. PERFORMANCE OPTIMIZATION METHODOLOGY

We now derive the step response of a realistic driver-interconnect-load structure from first principles and then apply it to develop our new methodology for optimum repeater insertion for distributed *RLC* interconnects.

2.1 Delay Calculation

Consider a uniform line with resistance, capacitance and inductance per unit length of r, c and l respectively, driven by a repeater of series resistance R_S and output parasitic capacitance C_P and driving an identical repeater with load capacitance C_L (Figure 1). For a given technology, let the output resistance, output parasitic capacitance and input capacitance of a minimum sized repeater be r_s, c_p and c_0 respectively. Therefore if the repeater size is k times the size of a minimum sized repeater, $R_S = r_s/k$, $C_P = c_p k$ and $C_L = c_0 k$. For this analysis it is assumed that the repeater resistance and output parasitic capacitance is linear throughout the output voltage transition range.

The ABCD parameter matrix for a uniform *RLC* transmission line of length h is given by:

$$\begin{bmatrix} \cosh(\theta h) & Z_0 \sinh(\theta h) \\ \frac{1}{Z_0}\sinh(\theta h) & \cosh(\theta h) \end{bmatrix}$$

where $Z_0 = \sqrt{(r+sl)/sc}$ and $\theta = \sqrt{(r+sl)sc}$. Therefore the ABCD parameter matrix of the configuration in Figure 1 is given by

$$\begin{bmatrix} 1 & R_S \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC_P & 1 \end{bmatrix} \begin{bmatrix} \cosh(\theta h) & Z_0 \sinh(\theta h) \\ \frac{1}{Z_0}\sinh(\theta h) & \cosh(\theta h) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC_L & 1 \end{bmatrix}$$

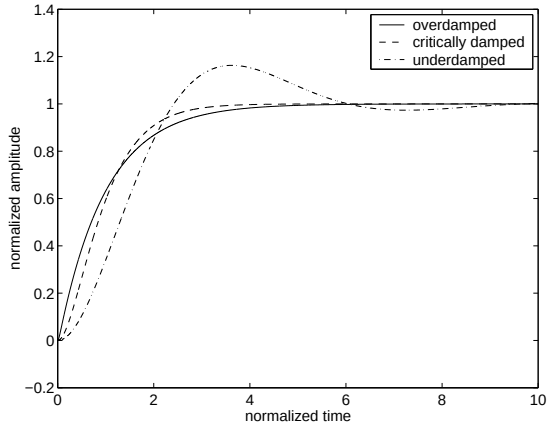

Figure 2: Step response of a second-order (*RLC*) system.

and the input-output transfer function is given by

$$H(s) = \frac{V_o(s)}{V_i(s)} = \frac{1}{[1+sR_S(C_P+C_L)]\cosh(\theta h) + \left[\frac{R_S}{Z_0} + sC_L Z_0 + s^2 R_S C_P C_L Z_0\right]\sinh(\theta h)} \tag{1}$$

The step-response of this system is given by $V_o(s) = \frac{1}{s}H(s)$ in the Laplace domain. However, computing the response in the time-domain is analytically intractable. Kahng and Muddu [23] suggested using a second-order Padé expansion of the transfer function[1], i.e.,

$$H(s) \approx \frac{1}{1+sb_1+s^2b_2} \tag{2}$$

where

$$b_1 = R_S(C_P+C_L) + \frac{rch^2}{2!} + R_S ch + C_L rh$$

$$b_2 = \frac{lch^2}{2!} + \frac{r^2c^2h^4}{4!} + R_S(C_P+C_L)\frac{rch^2}{2!} + (R_S ch + C_L rh)\frac{rch^2}{3!} + (C_L lh + R_S C_P C_L rh)$$

The two poles of the transfer function are

$$s_{1,2} = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_2}}{2b_2}$$

and the step response, which is the inverse Laplace transform of $\frac{1}{s}H(s)$, is given by

$$v(t) = V_0\left[1 - \frac{s_2}{s_2-s_1}\exp(s_1 t) + \frac{s_1}{s_2-s_1}\exp(s_2 t)\right]$$

The two poles s_1 and s_2 could be real or complex conjugate depending on the sign of $(b_1^2 - 4b_2)$. The response of a second order system is underdamped, critically damped or overdamped if $b_1^2 - 4b_2$ is less than zero, equal to zero and greater than zero respectively. The step response of a second order system for overdamped, underdamped and critically damped case is shown in Figure 2. The step response shows overshoot and undershoot for the underdamped case.

Therefore for a step input, the $f \times 100\%$, (where $0 \le f < 1$) delay, τ, (i.e., $v(\tau) = fV_0$) is the solution of the following equation

$$1 - f - \frac{s_2}{s_2-s_1}\exp(s_1\tau) + \frac{s_1}{s_2-s_1}\exp(s_2\tau) = 0 \tag{3}$$

Kahng and Muddu [23] gave an approximate expression for the delay when the poles are real or complex. Their approximation is accurate if the system is highly underdamped or highly overdamped ($|b_1^2 - 4b_2| \gg |b_2|$). When this is not the case, they suggest using the delay expression of the

[1] their driver-interconnect configuration did not include C_P and included a driver inductance L_S in series with R_S

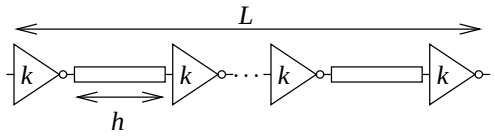

Figure 3: Long interconnect broken up into buffered segments.

critically damped case, i.e., $b_1^2 - 4b_2 = 0$. According to this condition, the value of line inductance l_{crit} required to make the system critically damped is given by

$$l_{crit} = \frac{\frac{b_1^2}{4} - \frac{r^2c^2h^4}{4!} - R_S(C_P+C_L)\frac{rch^2}{2!} - (R_Sch+C_Lrh)\frac{rch^2}{3!} - R_SC_PC_Lrh}{\frac{ch^2}{2!} + C_Lh} \quad (4)$$

Note that l_{crit} depends on the driver size k and the interconnect length h. As shown in Section 3.1, for driver sizes and interconnect lengths optimized for delay, l_{crit} is very similar to the line inductance l. The approximation in [23] predicts that the delay for such values of l is $\frac{1.9}{b_1}$. Note that b_1 is independent of l and therefore this approximation would predict that the delay would not change for values of l which are close to l_{crit}. Therefore, their approximation cannot be used for the delay optimization.

In our optimization, (3) is solved numerically for given values of s_1 and s_2. We solved (3) using Newton-Raphson method and observed that convergence is achieved in less than four iterations in all cases. Therefore obtaining the exact numerical solution of (3) is extremely efficient.

2.2 Delay Minimization

Ismail and Friedman [22, 21] presented empirical formulas for finding the optimum buffer size and interconnect length to minimize the delay in an interconnect of a fixed length. In their work an empirical expression for the 50% delay, which is obtained by curve-fitting with circuit simulation results, was minimized and optimized values of repeater size and interconnect length were plotted. Using those plots, empirical formulas for optimized values of repeater size and interconnect length were obtained using curve-fitting. However, their delay formula is valid only for 50% delay and is applicable only if the ratio of total line capacitance to load capacitance $\left(\frac{ch}{c_0k}\right)$ and the ratio of source resistance to total line resistance $\left(\frac{r_s}{krh}\right)$ are between 0 and 1.

We now present a new approach of optimizing repeater sizes and interconnect lengths which does not suffer from the limitations of the approach of [22, 21]. In our approach the delay is minimized using numerical techniques for any values of s_1, s_2 and f without any curve-fitting with circuit simulation results.

Consider a long interconnect of length L. In order to minimize its delay, the line is broken up into buffered segments of length h, each of which is driven by a buffer of size k and has a delay τ (Figure 3). The overall delay of the line is given by

$$\text{total delay} = \frac{L}{h}\tau$$

We therefore seek to minimize the *delay per unit length* τ/h. Setting the derivative of delay per unit length with respect to h and k to zero we have:

$$\frac{\partial \tau/h}{\partial h} = 0 \Rightarrow \frac{\partial \tau}{\partial h} = \frac{\tau}{h} \quad (5)$$

$$\frac{\partial \tau/h}{\partial k} = 0 \Rightarrow \frac{\partial \tau}{\partial k} = 0 \quad (6)$$

Multiplying (3) with $(s_2 - s_1)$, differentiating it with respect to h and k and using (5) and (6) we get

$$\begin{aligned} 0 = g_1 = (1-f)\left[\frac{\partial s_2}{\partial h} - \frac{\partial s_1}{\partial h}\right] - \frac{\partial s_2}{\partial h}\exp(s_1\tau) + \frac{\partial s_1}{\partial h}\exp(s_2\tau) \\ - s_2\tau\left[\frac{\partial s_1}{\partial h} + \frac{s_1}{h}\right]\exp(s_1\tau) + s_1\tau\left[\frac{\partial s_2}{\partial h} + \frac{s_2}{h}\right]\exp(s_2\tau) \end{aligned} \quad (7)$$

$$\begin{aligned} 0 = g_2 = (1-f)\left[\frac{\partial s_2}{\partial k} - \frac{\partial s_1}{\partial k}\right] - \frac{\partial s_2}{\partial k}\exp(s_1\tau) - s_2\tau\frac{\partial s_1}{\partial k}\exp(s_1\tau) \\ + \frac{\partial s_1}{\partial k}\exp(s_2\tau) + s_1\tau\frac{\partial s_2}{\partial k}\exp(s_2\tau) \end{aligned} \quad (8)$$

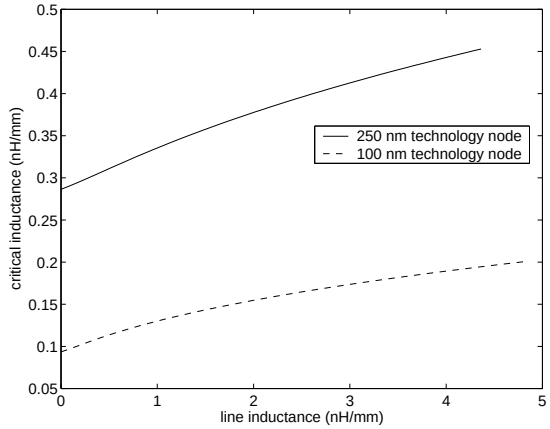

Figure 4: Critical inductance (l_{crit}) as a function of line inductance.

where

$$\frac{\partial s_{1,2}}{\partial h,k} = \frac{1}{2b_2}\left[-\frac{\partial b_1}{\partial h,k} \pm \frac{1}{\sqrt{b_1^2-4b_2}}\left(b_1\frac{\partial b_1}{\partial h,k} - 2\frac{\partial b_2}{\partial h,k}\right) - \frac{1}{b_2}\left(-b_1 \pm \sqrt{b_1^2-4b_2}\right)\frac{\partial b_2}{\partial h,k}\right]$$

Equations (7) and (8) can be numerically solved to obtain values of buffer size $k_{opt_{RLC}}$ and interconnect length $h_{opt_{RLC}}$ which minimize the delay per unit length. We used Newton-Raphson method for this purpose and observed that convergence is achieved in less than six iterations in all cases. The computation steps involved in each Newton iteration are:

1. find $b_{1,2}$, $s_{1,2}$ and their derivatives w. r. t. h and k.
2. compute τ by numerically solving (3).
3. compute $g_{1,2}$ using (7) and (8) and their derivatives w. r. t. h and k.
4. solve

$$\begin{bmatrix} \frac{\partial g_1}{\partial h} & \frac{\partial g_1}{\partial k} \\ \frac{\partial g_2}{\partial h} & \frac{\partial g_2}{\partial k} \end{bmatrix} \begin{bmatrix} \Delta h \\ \Delta k \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

Therefore this entire optimization step is extremely efficient. Note that the only approximations in the above optimization steps are

1. use of (2) instead of (1) for the transfer function
2. linear r_s and c_p for the entire voltage range

3. RESULTS AND DISCUSSION

We now apply our optimization technique to the top level metal interconnect, i.e., metal 6 and metal 8 for 250 nm and 100 nm technology nodes respectively. The relevant interconnect parameters are shown in Table 1 and are based on the NTRS roadmap [24]. Capacitance extraction was performed using a full 3D extractor (FASTCAP) [25].

It should be pointed out that a real VLSI interconnect is not an isolated line as shown in Figure 1 but is embedded in a large multi-level interconnect system. Therefore a significant amount of coupling, both capacitive and inductive, can exist between interconnects. Typically, interconnects on one layer of metal are routed in one direction and on the neighbouring metal layers, interconnects are routed in orthogonal direction. Therefore there is minimal capacitive coupling between interconnects on different metal layers. Additionally, lines on one metal layer couple only to their two nearest neighbouring lines. The effective line capacitance can therefore change due to Miller effect depending on the switching activity on these two nearest neighbouring lines. Since the aspect ratios of interconnects in DSM technologies is typically greater than one, effective line capacitance can vary by as much as 4×. However, as pointed out earlier, since magnetic fields have much longer spatial range compared to electric fields, line inductance values are sensitive to switching activities even in distant lines and therefore experience much larger variation and uncertainty in effective line inductance. In our experiments, for simplicity we have assumed that line capacitance is fixed and concentrated on variations in line inductances. These results can be easily modified to incorporate variations in line capacitance as well.

Tech. (nm)	r (Ω/mm)	c (pF/m)	ε_r	width (μm)	pitch (μm)	height (μm)	t_{ins} (μm)	$h_{opt_{RC}}$ (mm)	$k_{opt_{RC}}$	$\tau_{opt_{RC}}$ (ps)	r_s (kΩ)	c_0 (fF)	c_p (fF)
250	4.4	203.50	3.3	2	4	2.5	13.9	14.4	578	305.17	11.784	1.6314	6.2474
100	4.4	123.33	2	2	4	2.5	15.4	11.1	528	105.94	7.534	0.758	3.68

Table 1: Interconnect technology parameters. Interconnect material is Copper for both technologies. t_{ins} is the distance between the top layer metal and substrate.

3.1 Impact of Inductance on Circuit Behaviour

First consider the case of optimum repeater insertion by considering only the line resistance and capacitance. The total Elmore delay of interconnect of length L (Figure 3) is given by

$$t_{Elmore} = \frac{L}{h}\left[\frac{r_s}{k}(c_p k + c_0 k) + \frac{r_s}{k}ch + rhc_0 k + \frac{1}{2}rch^2\right]$$

Therefore the optimum repeater size $k_{opt_{RC}}$ and interconnect length $h_{opt_{RC}}$ for minimum delay is given by

$$h_{opt_{RC}} = \sqrt{\frac{2r_s(c_0+c_p)}{rc}} \quad k_{opt_{RC}} = \sqrt{\frac{r_s c}{rc_0}}$$

Furthermore, the delay of one segment of length $h_{opt_{RC}}$ driven by a buffer of size $k_{opt_{RC}}$ is given by

$$\tau_{opt_{RC}} = 2r_s(c_0+c_p)\left(1+\sqrt{\frac{2c_0}{c_0+c_p}}\right)$$

Note that $\tau_{opt_{RC}}$ is independent of r and c and therefore the wiring level. Therefore it can be treated as a technology parameter.

In general, for a given technology, r_s, c_p and c_0 cannot be easily determined. Moreover, r_s and c_p are voltage dependent. Therefore for this study, we find $h_{opt_{RC}}$ and $k_{opt_{RC}}$ by SPICE simulations. These simulations also provide $\tau_{opt_{RC}}$. Using the above equations, r_s, c_p and c_0 can be determined for that particular technology. These values are shown in Table 1 for the two technologies.

We now show the effect of including line inductance in the optimization as derived in Section 2.2. As pointed out in Section 1.1 l is not a fixed parameter for a given technology and metal layer but depends on the current return path and varies substantially with input vectors. However, the line inductance would be greatest when the current return path is the farthest from the line, i.e., the substrate. The worst-case line inductance for both these technologies was calculated to be < 5 nH/mm. Therefore we have carried out delay minimization for $0 \leq l < 5$ nH/mm.

As suggested by (4), for a given value of h and k, a critical value of line inductance can be obtained for which the system will be critically damped. It is instructive to compute l_{crit} for the optimized values of h and k, i.e., $h_{opt_{RLC}}$ and $k_{opt_{RLC}}$ and compare it with l. Figure 4 shows the variation of l_{crit} as a function of l for the two technology nodes. Recall that the second order system is overdamped, critically damped or underdamped if $l < l_{crit}$, $l = l_{crit}$ or $l > l_{crit}$. Note that l and l_{crit} are of the same order of magnitude for most practical values of l. Therefore the delay approximation presented in [23] cannot be used for these values of l. Also note that the values of l_{crit} for the 100 nm technology node are smaller than l_{crit} values for the 250 nm technology node. Therefore, compared to the 250 nm technology, top layer interconnects for the 100 nm technology node will be underdamped for a larger range of values of l and therefore using Elmore delay for these lines will lead to larger errors in delay.

Figure 5 shows the ratio of the optimum interconnect length $h_{opt_{RLC}}$ and the interconnect length optimized for Elmore delay, $h_{opt_{RC}}$. Consider the special case when $l = 0$. Since the optimization of Section 3.1 is carried out using a second-order transfer function (2), $h_{opt_{RLC}}$ is slightly smaller than $h_{opt_{RC}}$. This effect cannot be predicted using the curve-fitted equations of [22, 21]. Figure 6 shows ratio of the optimum buffer size $k_{opt_{RLC}}$ and the buffer size optimized for Elmore delay, $k_{opt_{RC}}$. These plots also corroborate the observations made in [21, 2] that with increasing line inductance l, the *RLC* interconnect increasingly resembles an ideal *LC* transmission line and the delay becomes progressively linear with interconnect length. Therefore $h_{opt_{RLC}}$ increases as the line inductance is increased and $k_{opt_{RLC}}$ reduces and asymptotes to a value for which the driver output impedance is equal to the characteristic impedance of the line.

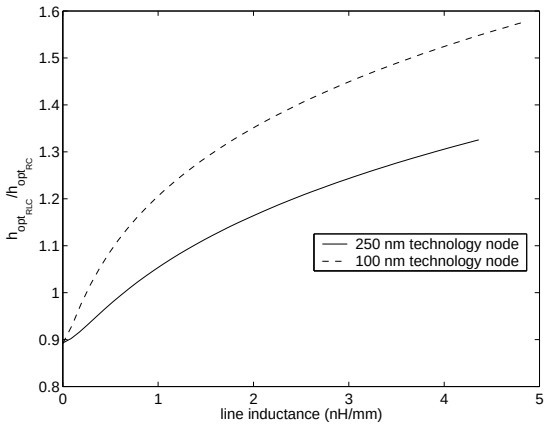

Figure 5: $h_{opt_{RLC}}/h_{opt_{RC}}$ as a function of line inductance.

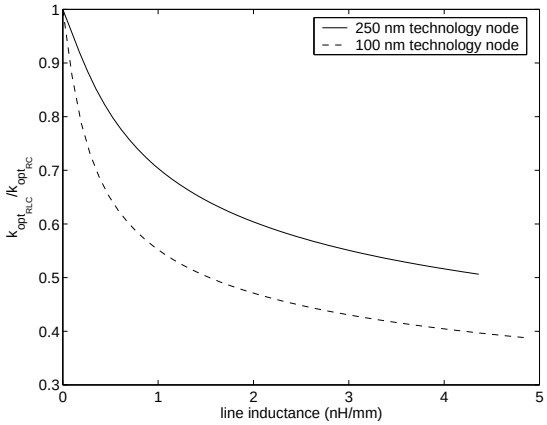

Figure 6: $k_{opt_{RLC}}/k_{opt_{RC}}$ as a function of line inductance.

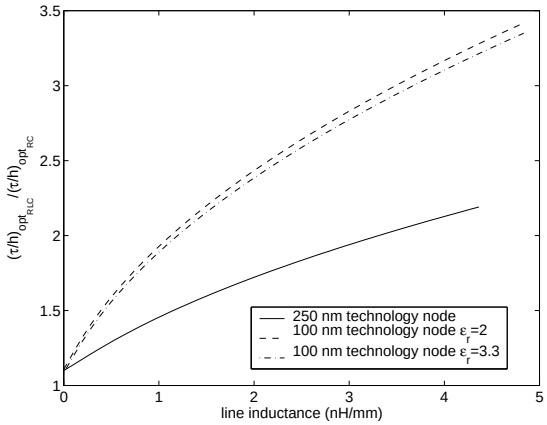

Figure 7: Ratio of optimum delay per unit length, $\frac{\tau}{h}$, for *RLC* and *RC* case as a function of line inductance l.

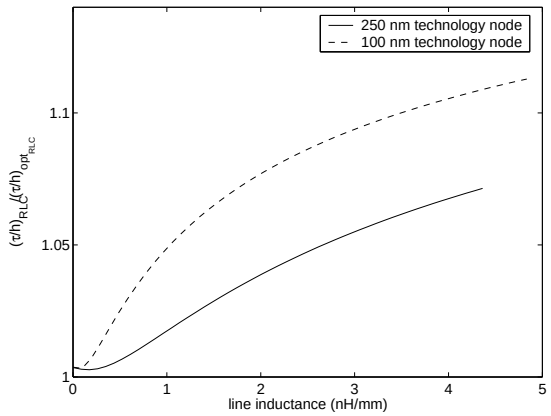

Figure 8: Ratio of *RLC* delays per unit length of an interconnect with $h = h_{opt_{RC}}$ and $k = k_{opt_{RC}}$, and with $h = h_{opt_{RLC}}$ and $k = k_{opt_{RLC}}$ as a function of l.

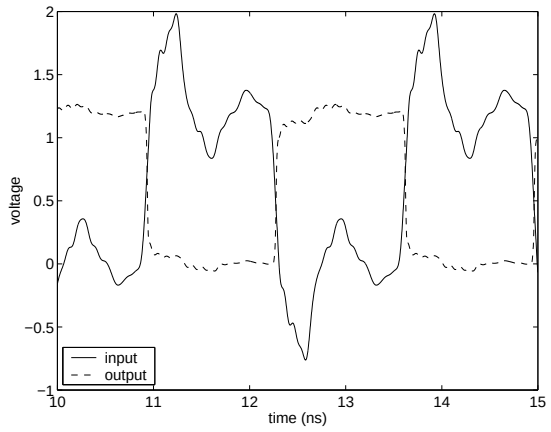

Figure 9: Voltage waveform at the input and output of an inverter in a five stage ring oscillator with $l = 1.8$ nH/mm.

Figure 7 plots the ratio of optimized interconnect delays per unit length with and without considering line inductance as a function of l. For the 250 nm technology node, the delay increases by a factor of 2 for the range of line inductances considered. However, for the 100 nm technology node, this ratio increases rapidly and for large values of l, optimized delay per unit length is 3.5× the corresponding optimized delay per unit length obtained without considering l. Therefore VLSI circuits will progressively become more susceptible to inductance effects as the technology scales. Recall that the top layer metal geometry is identical for both technologies and we have considered similar ranges of l for both technologies. We also carried out this experiment where the dielectric constant for 100 nm was assumed to be same as that of the 250 nm node and therefore the capacitance per unit length was identical. From Figure 7 we observe that even with identical c, the ratio of the optimized delay per unit length increases rapidly with l for the 100 nm technology node. Therefore this increased susceptibility is entirely due to scaling of driver capacitance and output resistance.

3.2 Impact of Inductance Variation on Delay

As discussed in Section 1.1, in a realistic scenario, it is very difficult to predict the effective interconnect inductance because the current return path varies a lot for different input vector patterns. As a result, it is difficult to target a specific value of the line inductance l and optimize the buffer size and interconnect length for that value of l. Therefore it is useful to determine the change in the interconnect delay for a given buffer size and interconnect length as the line inductance l is varied and compare it with the *RLC* based optimum delay for the corresponding values of l. As an example, if the driver size and interconnect length are chosen to be $k_{opt_{RC}}$ and $h_{opt_{RC}}$ respectively, it is useful to determine the increase in the

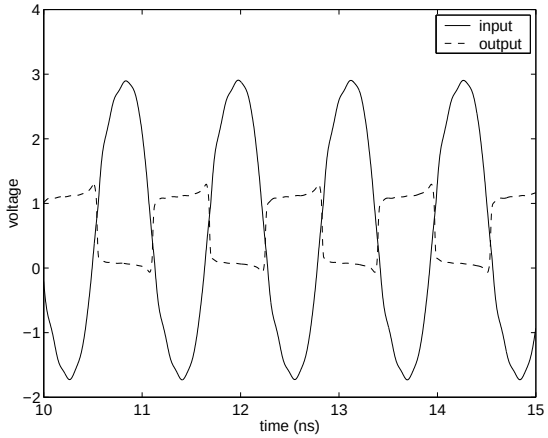

Figure 10: Voltage waveform at the input and output of an inverter in a five stage ring oscillator with $l = 2.2$ nH/mm.

interconnect delay over the *RLC* based optimum delay for the corresponding values of l. Figure 8 plots the ratio of these two delays as a function of l. For the 250 nm technology node, the worst-case increase in the delay over the optimized *RLC* case is 6% whereas for the 100 nm node, the corresponding increase is 12%. This again demonstrates that interconnect delay will progressively become more susceptible to inductance effects as the technology scales.

3.3 Catastrophic Failures due to Inductance

As shown in Figure 2, the step response of an underdamped system exhibits overshoot and undershoot. This overshoot and undershoot can cause catastrophic failures both in terms of device life-time degradation and errors during the operation of the logic circuits.

3.3.1 Logic Failures

As an illustration of this phenomenon, consider a five stage ring oscillator in the 100 nm technology node in which each stage consists of an inverter of size $k_{opt_{RC}}$ driving an interconnect of length $h_{opt_{RC}}$. Figure 9 shows the voltage waveform at the input and output of an inverter in this ring oscillator. The line inductance is assumed to be 1.8 nH/mm. Note that even though the input waveform shows a significant amount of overshoot and undershoot, the inverter output is relatively "clean". However, if l is increased, the undershoot can become large enough to cause the inverter to switch and since this inverter is a part of a ring oscillator, the false switching propagates throughout the chain and the period of oscillation becomes very small. Figure 10 shows the voltage waveform at the input and output of an inverter in this ring oscillator with a line inductance of 2.2 nH/mm. Note that with a small increase in l, the period of oscillation is less than half of the corresponding period for $l = 1.8$ nH/mm. To illustrate this phenomenon further, Figure 11 plots the period of oscillation as a function line inductance. Around $l = 2$ nH/mm, the period drops sharply indicating the onset of false switching in the circuit. A similar behaviour was observed for a five stage buffered *RLC* line which was excited by a square wave at one end with the other end connected to an identical repeater. Therefore this behaviour is not an artifact of the ring oscillator configuration. However, for the 250 nm technology node, this phenomenon is not observed for $0 \leq l < 5$ nH/mm. This again indicates that designs in 100 nm technology will be more susceptible to inductance effects as compared to the 250 nm technology designs.

3.3.2 Reliability Failures

As shown in Figures 9 and 10 the voltage at the repeater input is greater than V_{DD} due to overshoot. Since this voltage is applied at the gate of a MOS transistor, this can cause oxide reliability problems [26, 27]. In current technologies the supply voltage is limited by the electric field that can be reliably sustained in the oxide. In DSM technologies, the supply voltage is also scaling with gate oxide thickness in order for the oxide electric field to stay below some critical value [27]. Hence if the gate voltage is greater than V_{DD} due to overshoot, device reliability can degrade rapidly due to gate oxide wear out.

It was shown in [28] that if interconnect rms and peak current densities

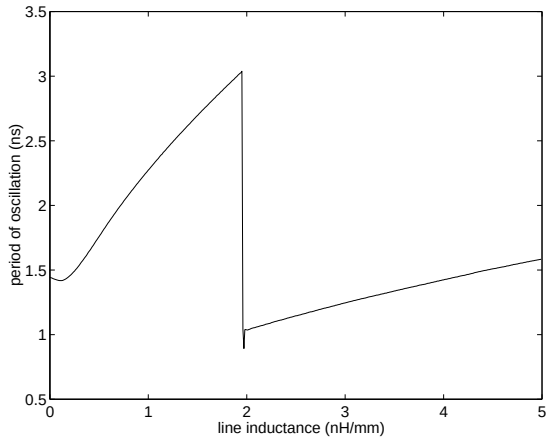

Figure 11: Period of oscillation for the five stage ring oscillator as a function of line inductance.

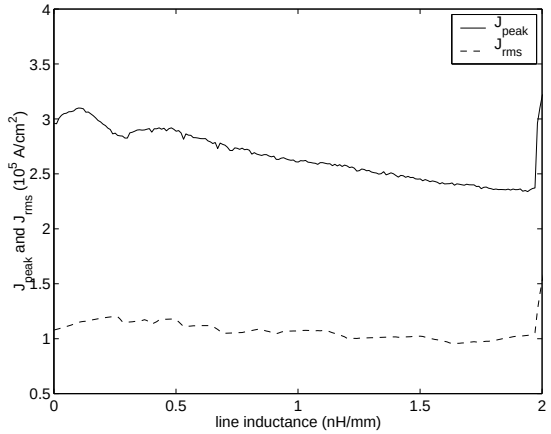

Figure 12: Peak and rms interconnect densities vs line inductance for 100 nm technology top level metal.

increase beyond a certain limit, interconnect reliability can be affected due to increased Joule heating and electromigration of the wire. We therefore investigated the effect of line inductance on interconnect reliability. Figure 12 plots the peak and rms current densities in the interconnect for the five stage ring oscillator as a function of l. It can be observed that the peak and rms current densities do not change appreciably as l increases. Therefore the interconnect reliability does not degrade as the line inductance varies.

4. CONCLUSIONS

In summary, we have derived an accurate expression for the step response of a realistic driver-interconnect-load structure and applied it to develop a novel methodology for optimum repeater insertion for a distributed *RLC* interconnect which is based on analytical minimization of the delay of a repeater driving an interconnect with a load capacitance. The proposed minimization can be performed efficiently and does not require curve-fitting with circuit simulation results. With this methodology optimum repeater sizes and interconnect lengths can be efficiently computed for given technology and interconnect parameters. We have used this methodology to compute the optimum buffer size and interconnect lengths for a wide range of line inductances found in global interconnects for 250 nm and 100 nm technology nodes and compared these with Elmore delay optimization based values. We show that designs at 100 nm technology node are more susceptible to delay variations due to change in inductance as compared to design at 250 nm technology node, even if the interconnect parameters are (artificially) made identical to those of 250 nm technology node. This is due to the reduction of driver capacitance and output resistance due to scaling. We also showed that for practical range of line inductances, 100 nm technology designs are susceptible to logic errors due to excessive voltage undershoot where as the designs at 250 nm node do not suffer from this problem over similar range of values of line inductances. Furthermore, we showed that interconnect reliability will not be affected due to inductance variations.

5. REFERENCES

[1] E. E. Davidson, B. D. McCredie, and W. V. Vilkelis, "Long lossy lines (L^3) and their impact upon large chip performance," *IEEE Transactions on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging*, vol. 20, pp. 361–375, Nov. 1998.

[2] C.-K. Cheng, J. Lillis, S. Lin, and N. Chang, *Interconnect Analysis and Synthesis*. John Wiley & Sons, 1999.

[3] A. Deutsch, H. H. Smith, C. W. Surovic, G. V. Kopcsay, D. A. Webber, P. W. Coteus, G. A. Katopis, W. D. Becker, A. H. Dansky, G. A. Sai-Halasz, and P. J. Restle, "Frequency-dependent crosstalk simulation for on-chip interconnections," *IEEE Transactions on Advanced Packaging*, vol. 22, pp. 292–308, Aug. 1999.

[4] D. Edelstein *et al.*, "Full copper wiring in a sub-0.25 μm CMOS ULSI technology," in *International Electron Devices Meeting. Digest of Technical Papers*, pp. 773–776, 1997.

[5] N. Rohrer *et al.*, "A 480MHz RISC microprocessor in a 0.12 μm L_{eff} CMOS technology with copper interconnects," in *International Solid-State Circuits Conference. Digest of Technical Papers*, pp. 240–241, 1998.

[6] A. Deutsch, G. V. Kopcsay, P. J. Restle, H. H. Smith, G. Katopis, W. D. Becker, P. W. Coteus, C. W. Surovic, B. J. Rubin, R. P. Dunne, Jr., T. Gallo, K. A. Jenkins, L. M. Terman, R. H. Dennard, G. A. Sai-Halasz, B. L. Krauter, and D. R. Knebel, "When are transmission-line effects important for on-chip interconnections?," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, pp. 1836–1846, Oct. 1997.

[7] W. C. Elmore, "The transient response of damped linear networks," *Journal of Applied Physics*, vol. 19, pp. 55–63, Jan. 1948.

[8] A. E. Ruehli, "Inductance calculations in a complex integrated circuit environment," *IBM Journal of Research and Development*, pp. 470–481, Sept. 1972.

[9] G. L. Matthaei, G. C. Chinn, C. H. Plott, and N. Dagli, "A simplified means for computation of interconnect distributed capacitances and inductances," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 11, no. 4, pp. 513–524, 1992.

[10] M. Kamon, M. J. Tsuk, and J. K. White, "FASTHENRY: a multipole-accelerated 3-D inductance extraction program," *IEEE Transactions on Microwave Theory and Techniques*, vol. 42, pp. 1750–1758, Sept. 1994.

[11] B. Krauter and S. Mehrotra, "Layout based frequency dependent inductance and resistance extraction for on-chip interconnect timing analysis," in *Proceedings 1998 Design Automation Conference*, pp. 303–308, 1998.

[12] Y. Massoud, S. Majors, T. Bustami, and J. White, "Layout techniques for minimizing on-chip interconnect self-inductance," in *Proceedings 1998 Design Automation Conference*, pp. 566–571, 1998.

[13] M. W. Beattie and L. T. Pillegi, "IC analyses including extracted inductance models," in *Proceedings 1999 Design Automation Conference*, pp. 915–920, 1999.

[14] S. V. Morton, "On-chip inductance issues in multiconductor systems," in *Proceedings 1999 Design Automation Conference*, pp. 921–926, 1999.

[15] N. Chang, S. Lin, L. He, O. S. Nakagawa, and W. Xie, "Clocktree RLC extraction with efficient inductance modeling," in *Proceedings Design, Automation and Test in Europe Conference and Exhibition 2000*, pp. 522–526, 2000.

[16] B. Kleveland, X. Qi, L. Madden, R. W. Dutton, and S. S. Wong, "Line inductance extraction and modeling in a real chip with power grid," in *International Electron Devices Meeting 1999. Technical Digest*, pp. 901–904, 1999.

[17] X. Qi, B. Kleveland, Z. Yu, S. S. Wong, R. W. Dutton, and T. Young, "On-chip inductance modeling of VLSI interconnects," in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pp. 172–173, 2000.

[18] K. Gala, V. Zolotov, R. Panda, B. Young, J. Wang, and D. Blaauw, "On-chip inductance modeling and analysis," in *Proceedings 2000 Design Automation Conference*, pp. 63–68, 2000.

[19] S. Lin, N. Chang, and S. Nakagawa, "Quick on-chip self- and mutual-inductance screen," in *Proceedings IEEE 2000 First International Symposium on Quality Electronic Design*, pp. 513–520, 2000.

[20] P. Restle, A. Ruehli, and S. G. Walker, "Dealing with inductance in high-speed chip design," in *Proceedings 1999 Design Automation Conference*, pp. 904–909, 1999.

[21] Y. I. Ismail and E. G. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, pp. 195–206, Apr. 2000.

[22] Y. I. Ismail and E. G. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits," in *Proceedings 1999 Design Automation Conference*, pp. 721–724, 1999.

[23] A. B. Kahng and S. Muddu, "An analytical delay model for RLC interconnects," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 16, pp. 1507–1514, Dec. 1997.

[24] "National Technology Roadmap for Semiconductors (NTRS)," 1997.

[25] K. Nabors and J. K. White, "FASTCAP: a multipole-accelerated 3-D capacitance extraction program," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, pp. 1447–1459, Nov. 1991.

[26] C. Hu, "MOSFET scaling in the next decade and beyond," *Semiconductor International*, pp. 105–114, 1994.

[27] C. Hu, "Gate oxide scaling limits and projection," in *International Electron Devices Meeting. Digest of Technical Papers*, pp. 319–322, 1997.

[28] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu, "On thermal effects in deep sub-micron VLSI interconnects," in *Proceedings 1999 Design Automation Conference*, pp. 885–891, 1999.