A Probabilistic Framework for Power-Optimal Repeater Insertion in Global Interconnects under Parameter Variations

Vineet Wason and Kaustav Banerjee
Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106
{vwason, kaustav}@ece.ucsb.edu

ABSTRACT

This paper addresses the problem of power dissipation during the buffer insertion phase of interconnect performance optimization in nanometer scale designs taking all significant parameter variations into account. The relative effect of different device, interconnect and environmental variations on delay and different components of power has been studied. A probabilistic framework to optimize buffer-interconnect designs under variations has been presented and results are compared with those obtained through simple deterministic optimization. Also, statistical models for delay and power under parameter variations have been developed using linear regression techniques. Under statistical analysis, both power and performance of buffer-interconnect designs are shown to degrade with increasing amount of variations. Also, % error in power estimation for power-optimal repeater designs is shown to be significant if variations are not taken into account. Furthermore, it has been shown that due to variations, significantly higher penalties in delay are needed to operate at power levels similar to those under no variations. Finally, the percentage savings in total power for a given penalty in delay are shown to improve with increasing amount of parameter variations.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles – *VLSI.*

General Terms: Performance, Design.

Keywords: Parameter variations, buffer-interconnect system, sensitivity analysis, statistical delay and power models, statistical optimization.

1. INTRODUCTION

As very large scale integrated (VLSI) circuits continue to be scaled beyond 100 *nm* technology node, the performance of these ICs is being increasingly dominated by global interconnects since their delay is a quadratic function of their length [1]. In order to optimize interconnect delay, buffers are inserted at regular intervals as a result of which delay becomes a linear function of length [2]. For large high-performance designs, the number of such buffers can be prohibitively high and can take up large fraction of silicon and routing area. Also, it is known that a significant fraction of the total chip power dissipation arises due to the loading caused by long global and semi-global-tier interconnect networks, especially in high performance designs. Since, increasing power dissipation in integrated circuits has been highlighted as a real limiter to future scaling of CMOS circuits [3]-[4], optimizing both power as well as delay is more meaningful, especially for sub-100 *nm* CMOS circuits, where leakage power forms a major component of total power. Previous works [5] on simultaneous power-delay optimization for buffer-interconnect system have already shown the existence of huge potentials for power savings at the expense of a small penalty in delay. This is mainly because of the fact that the interconnect delay is very shallow with respect to both repeater size and inter-repeater separation close to the optimum point as illustrated in Figure 1. Hence, by operating *'slightly away'* from the optimal delay point, large power savings can be achieved.

ISLPED'05, August 8–10, 2005, San Diego, California, USA.

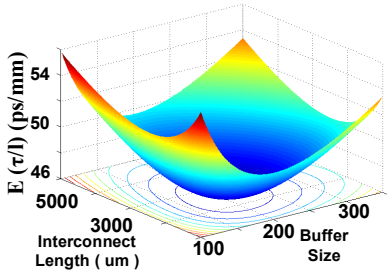

Figure 1. Expected value of delay per unit length as a function of buffer size (*s)* and interconnect length (*l*) for 100 *nm* tech. node under parameter variations.

Although a comprehensive power-optimal repeater insertion methodology was presented in [5], it neglected the impact of parameter variations on delay and power. Path delays and total power in a buffer-interconnect system depends upon device parameters such as channel length (L), oxide thickness (t_{ox}) and doping concentration (N_{ch}), interconnect parameters such as interconnect width (w), thickness (t), spacing (s) and inter layer dielectric (ILD) height (h) and environmental parameters such as temperature (T), supply voltage (V_{dd}) and crosstalk. Numerous works [6]-[13] in the past have shown that variations in these parameters can have a significant impact on delay, and in worst case, lead to timing violations. Also, power dissipation, especially, leakage power has been shown to be significantly affected due to parameter variations [3]-[4], [14]-[15].

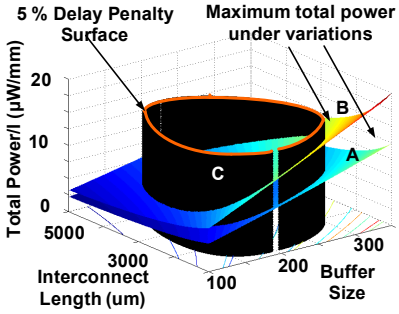

Figure 2. Maximum value of total power per unit length as a function of buffer size (s) and interconnect length (l) under two different cases. Also drawn is a delay penalty surface, along which $(\tau/l) = 1.05\ (\tau/l)_{opt}$.

We qualitatively illustrate the importance of considering variations in power-optimal repeater insertion by considering two cases A and B. The amount of parameter variations is much higher for case B as compared to case A. Figure 2 shows the maximum value of total power dissipation per unit length for buffer-interconnect system under these two cases. Due to variations, the maximum power plane shifts upwards for case B in comparison to case A. Also drawn are the set of values of buffer sizes and interconnect lengths which would give a 5% penalty in delay compared to the optimum point. It should be noted that the 2-D contour which gives a 5% delay penalty has been extended in 3D plane (surface C) for better visualization purpose. Thus, the goal of the statistical power-optimal repeater insertion methodology is to find the optimum point on the contour formed by the intersection of C with A (or B) which yields the minimum total power per unit length. Clearly, the optimum for case A and case B can be significantly different in s_{opt}, l_{opt} and P_{opt}/l, thus

emphasizing the fact that parameter variations can significantly impact the power optimal repeater insertion, which is the focus of this work.

In order to understand and account for variation effects in devices and interconnects, it is important that electrical behavior (such as delay, power dissipation) of interconnect and devices is modeled accurately. There can be multiple ways of obtaining these models. One way could be to start from a deterministic model for power and delay and perform a statistical analysis to derive analytical models that are valid under variations. However, the complexity of such an analysis would increase significantly with the increase in the number of parameters that contribute to uncertainty. Furthermore, the deterministic models that will be used for statistical analysis might not be accurate enough to capture the variations of device and interconnect parameters. Hence, for 'exact' computations of delay and power, SPICE simulations are mandatory which can be later used to develop accurate models using linear regression techniques. Accurate statistical models for delay and different components of power thus developed could then be used to carry out statistical optimization of buffer-interconnect designs under parameter variations.

This paper is organized as follows. The preliminary results for deterministic delay optimization for a buffer-interconnect system is presented in Section 2. Sensitivity analysis of various parameters on delay and power is presented in Section 3. The results for statistical delay optimization are presented in Section 4. Power-optimal repeater insertion methodology for a given delay penalty under the impact of parameter variations is presented in Section 5. Results for statistical power optimal repeater insertion are presented in Section 6. Concluding remarks are presented in Section 7.

2. Deterministic Delay Modeling and Optimization

We consider a uniform interconnect buffered by identical repeaters as shown in Figure 3.

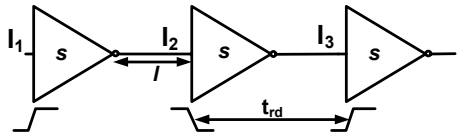

Figure 3. Schematic representation of a buffer-interconnect system with s and l representing buffer size and inter-buffer interconnect length respectively while t_{rd} represents the rise delay.

Since we are dealing with global interconnects, we assume the following interconnect structure as shown in Figure 4 (a)-(b). Here, *w, t, s* and *h* represent interconnect width, thickness, spacing and ILD height while C_c is the coupling capacitance and C_g is capacitance to ground.

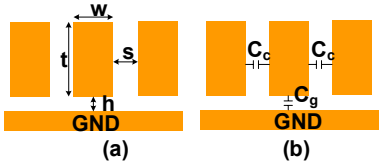

Figure 4. (a) Interconnect structure for top global layer
(b) Interconnect capacitances for top global layer

In order to carry out the deterministic optimization, we first develop a deterministic model to calculate the average (of rise and fall) delay from I_2 to I_3 (Figure 3). This can be done by carrying out SPICE simulations for different sets of s and l and then by using linear regression techniques, a model can be developed. While carrying out the SPICE simulations, following distributed model (Figure 5) for interconnect was used. The values of interconnect capacitance (crosstalk and ground) and resistance were derived using interconnect BPTM model [16]. The signals $xtalk_i$ are set to ground for simulations without variations. These signals will be used later while simulating under parameter variations. It should be noted that here interconnect segment is considered as an RC element and not an RLC element. This is done because of the fact that the effect of line inductance reduces with technology scaling for minimum sized global interconnects [17].

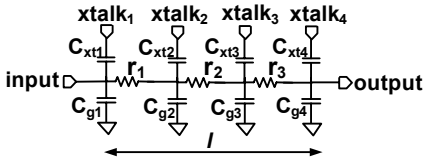

Figure 5. Interconnect model used for SPICE simulations

With the help of linear regression techniques and data generated for τ/l from SPICE simulation, we develop the following model:

$$\tau / l = \beta_1 s + \frac{\beta_2}{s} + \beta_3 l + \frac{\beta_4}{l} \quad (1)$$

Here, for an interconnect segment, τ is the time difference between input and output waveforms crossing 50% of their full swing value and is averaged over rise and fall cycles. β_1, β_2, β_3, β_4 are empirical constants and are listed in Table 1 for 100 *nm* technology node. Note that here τ/l is measured in *ps/mm*. From now onwards, if not mentioned otherwise, all results are obtained for 100 *nm* technology node. BPTM parameters [16] have been used throughout the simulations. The following values for interconnect geometry parameters have been used: w=0.5μm, s=0.5μm, t=1.2μm and h=0.3μm [16].

Table 1. Summary of linear model used to estimate parameters in (1). The t value is the ratio of estimate and std. error. Pr (>|t|) is measure of significance of the corresponding estimate, evaluated using t-statistics. R-squared values are a measure of how well the model fits the data. F-statistics is used to test if any of the predictors (here, s, 1/s, l, 1/l) have any significance in the model. A high p-value in F-statistics indicates low significance. Here, DF represents the degrees of freedom for F-test [18].

	Estimate	Std. Error	t value	Pr(>|t|)
β_1	6.096e-02	7.43E-04	82.09	<2e-16
β_2	2.476e+03	1.56E+01	158.71	<2e-16
β_3	3.083e-03	4.22E-05	73.03	<2e-16
β_4	3.740e+04	4.11E+02	91.08	<2e-16
Residual standard error of response: 0.4526				
Multiple R-Squared: 0.9999, Adjusted R-squared: 0.9999				
F-statistic: 2.934e+05 on 4 and 76 DF, p-value: < 2.2e-16				

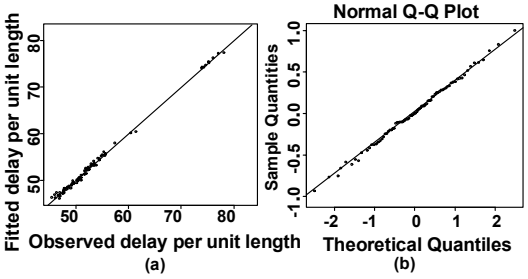

Figure 6. (a) Fitted vs. observed values for delay per unit length
(b) Normal plots for residuals for model given by (1)

The estimates of above parameters (β_1, β_2, β_3, β_4) were obtained using linear regression techniques. A small p-value and multiple R-squared values close to 1 show that the model given by (1) is a reasonable model for the given data. The plots for fitted vs. observed values for delay per unit length and residuals are shown in Figure 6(a)-(b). Figure 6(a) shows a good fit, while residuals (Figure 6(b)) roughly show normal distribution. The values of s and l for which (1) is optimized is given by (2):

$$s_{opt} = \sqrt{\frac{\beta_2}{\beta_1}} \qquad l_{opt} = \sqrt{\frac{\beta_4}{\beta_3}} \quad (2)$$

Using values listed in Table 1, we get s_{opt} = 201.53 and l_{opt} = 3482.96 μm. In the next section, we present a sensitivity analysis to screen out the parameters that do not contribute to spread in delay and power. Later in the section, closed form statistical models for delay and power per unit length are also developed.

3. Sensitivity Analysis for Delay and Power

3.1. Sensitivity study for delay

The uncertainties in device, interconnect and device parameters can result in a spread in the delay of an interconnect-buffer system. As a result, delay optimization presented in Section 2 needs to be solved statistically. However, the complexity of any statistical optimization algorithm will be directly proportional to the number of varying parameters. Hence, before carrying out any statistical optimization, we perform screening experiment to screen out the parameters that do not contribute much to spread in delay and power using Monte Carlo (MC) simulations.

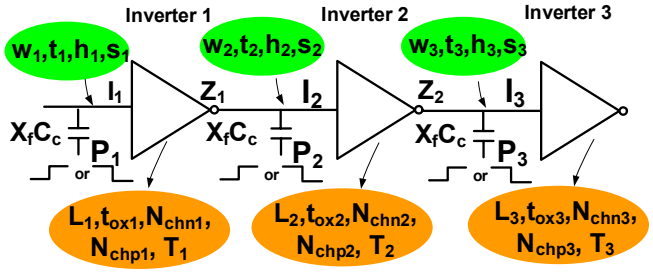

Figure 7. Buffer-interconnect system used for Monte Carlo simulations in sensitivity analysis experiment.

The circuit shown in Figure 7 shows the set up used for MC simulations. Here, N_{chn} and N_{chp} represent the doping density for NMOS and PMOS substrates respectively while X_f is the percentage of the coupling capacitance that contributes to crosstalk ($0 < X_f < 1$). The crosstalk signals P_1, P_2, P_3 are assumed to switch either from 0 to 1 or from 1 to 0 with equal probabilities. For carrying out the sensitivity analysis, Gaussian variations have been assumed for parameters such as L_i, t_{oxi}, N_{chi}, w_i, t_i, h_i, s_i, and T_i (i=1, 2, 3). The mean (μ) and sigma (σ) values for these parameters are listed in Table 2. It should be noted that only fluctuations in the number of dopant atoms has been taken into account here, since fluctuations in dopant position would require atomistic simulations which is beyond the scope of this work. Also, it has been assumed for simplification that at all times only X_f fraction of the total coupling capacitance (C_c) contributes to crosstalk while the rest of the capacitance ($1-X_f$).C_c is assumed to add to the ground capacitance (C_g). This might be a valid simplification since the aim of this experiment is to screen out the insignificant parameters. Also, X_f is varied from 0.1 to 0.3 later in the simulations to account for varying percentages of coupling capacitance contributing to crosstalk.

Table 2. Mean (μ) and sigma (σ) values for different parameters used during Monte Carlo simulations. The following set of values will now be referred to as Case 1 in rest of the paper.

Parameters	Mean	3σ
Supply Voltage (V)	1.2	0.036
Channel Length (*nm*)	100	10
Oxide Thickness (*nm*)	2.5	0.25
Doping Concentration for NMOS (cm^{-3})	9.7E17	0.97E17
Doping Concentration for PMOS (cm^{-3})	1.04E18	0.104E18
Interconnect Width (μm)	0.5	0.05
Interconnect Thickness (μm)	1.2	0.12
ILD Height (μm)	0.5	0.05
Interconnect Spacing (μm)	0.5	0.05
Temperature (°C)	100	50
Crosstalk Factor	0.1	

Since the actual numbers for variances listed in Table 2 were not known accurately, we consider 3 cases here. The 3σ percentage variation for parameters namely L, t_{ox}, N_{chn}, N_{chp}, w, t, h, s, were kept at 10%, 20%, and 30% of their mean value for case 1, 2, 3 respectively, the crosstalk factor was kept at 0.1, 0.2 and 0.3, while 3σ variations for temperature and V_{dd} were kept at 50, 70, 100°C and 3%, 6%, 10% respecively for the three cases. Note that Table 2 corresponds to case 1. The spread of delay obtained under 3 different cases has been shown in Figure 8, while the normalized maximum delay per unit length as function of different %age variations is shown in Figure 9. Here delay refers to the average of rise and fall delay from I2 to I3 (Figure 7). From figure 8 and 9, it can be clearly inferred that out of all device parameters, L has the most significant impact on delay spread, while among interconnect parameters, w variations result in most significant spread. Temperature (T) variations result in most dominant spread among environmental variations. The result that w variations result in most significant spread (among interconnect variations) in delay has been derived assuming equal percentages of variations in interconnect parameters. However, this may not be generally true and percentage variations for different interconnect parameters in a CMP (Chemical Mechanical Planarization) process might be different. Hence, once the actual numbers for variations are known, the corresponding values for maximum delay can be found from Figure 9. Similarly for other parameters such as temperature and supply voltage, once actual numbers are known for a particular design, more reasonable estimates for spread in delay can be found using Figure 9.

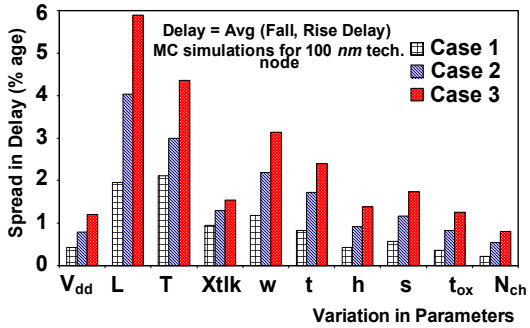

Figure 8. Percentage spread in delay for different percentages of variations in different parameters. Here Xtlk represents crosstalk. Rest of the parameters have their usual meanings as defined in the text.

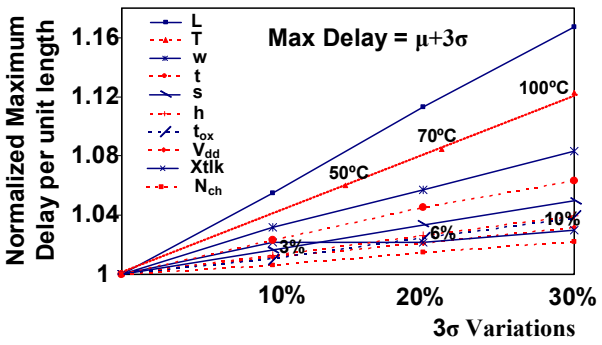

Figure 9. Normalized maximum delay for different percentages of variations in different parameters. Note that for temperature (T) and supply voltage (V_{dd}) the 3σ values are shown separately next to the corresponding temperature and Vdd curves.

However, it should be noted that the result of the screening experiment as performed above is circuit and technology dependent and also dependent on layout to some extent. Hence, the results obtained above cannot be simply generalized for even the same style of circuits at different technology nodes without further analysis. However, similar experiments as performed above can be carried out for different circuit styles and technology nodes, once exact variance values for each of the parameters are known. Based on the MC simulation, we obtained the following

models for μ and σ of delay per unit length (τ/l). The empirical constants are listed in Table 3 for three different cases.

$$E(\tau/l) = \beta_1 s + \frac{\beta_2}{s} + \beta_3 l + \frac{\beta_4}{l} \quad (3)$$

$$\sigma(\tau/l) = \beta_5 s + \frac{\beta_6}{s} + \beta_7 l + \frac{\beta_8}{l} \quad (4)$$

The above set of models showed a good fit with experimental data. However, the detailed summary of the linear models has been omitted here due to lack of space.

Table 3. Empirical constants for E(τ/l) and σ(τ/l) obtained using linear regression.

	Case 1	Case 2	Case 3
β_1	6.078e-02	6.099e-02	6.182e-02
β_2	2.482e+03	2.497e+03	2.521e+03
β_3	3.124e-03	3.163e-03	3.224e-03
β_4	3.767e+04	3.803e+04	3.837e+04
β_5	3.995e-03	6.228e-03	8.925e-03
β_6	1.091e+02	2.275e+02	3.482e+02
β_7	8.121e-05	9.591e-05	1.617e-04
β_8	3.927e+02	1.057e+03	1.678e+03

3.2. Sensitivity study for power

The power dissipation of a repeater shown in Figure 7 is given by:

$$P_{rep} = P_{sw} + P_{sc} + P_{le} \quad (5)$$

where P_{sw}, P_{sc} and P_{le} represent switching, short-circuit and leakage power. SPICE simulations were carried out to get an accurate estimate of these components. The results for spread *(s)* and delta *(δ)* for different components of power are summarized in the figures below. Here, *s* and *δ* are given by (6), where σ represents the standard deviation, μ represents the mean, while η represents the nominal value.

$$s = \frac{\sigma}{\eta} \qquad \delta = \frac{\mu - \eta}{\eta} \quad (6)$$

As mentioned earlier, since exact values of variances were not known, we again consider 3 different cases similar to those considered in Section 3.1. As shown in Figures 10-12, leakage power is most dominantly impacted due to variations since leakage power is exponentially dependent on channel length (L) (through V_{th}), temperature (T) and supply voltage (V_{dd}). Leakage power is only affected by device and environmental variations (not by interconnect variations). On the other hand, short circuit is sensitive to almost all the parameters as shown in Figure 11.

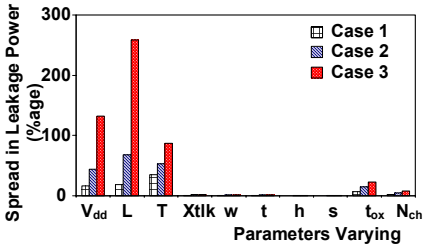

Figure 10. Spread in leakage power due to parameter variations under cases 1-3.

Also, the spread in dynamic power due to variations is almost negligible as compared to that of the leakage power. Note that the scale for y-axis in Figure 12 is not the same as that in Figures 10-11. Moreover, it can be seen that variations in short circuit power are significantly higher than those of dynamic power. This is because, the short circuit power is critically dependent on input and output waveforms, which can significantly vary due to variations. Apart from a spread in the distribution, variations also result in an average shift in the mean value of the power. However, only leakage power resulted in a significant shift from its nominal value. Hence, shift in only leakage power is shown in Figure 13 and the shift in rest of the power components could be safely neglected.

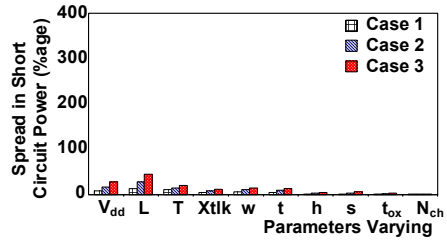

Figure 11. Spread in short-circuit power due to parameter variations under cases 1-3.

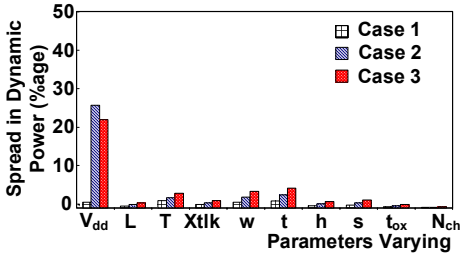

Figure 12. Spread in dynamic power due to parameter variations under cases 1-3.

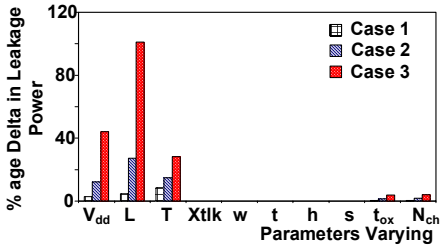

Figure 13. Delta for leakage power due to parameter variations under cases 1-3. Delta for other power components can be safely neglected.

Based on the data generated from MC simulations, the following statistical models (7)-(8) were developed for the mean and standard deviation of power. The empirical constants β_9- β_{22} are tabulated in Table 4.

$$E(P_{le}/l) = \beta_9 + \frac{\beta_{10}}{l} s \quad (7a)$$

$$E(P_{sw}/l) = \beta_{11} + \beta_{12} s + \frac{\beta_{13}}{l} + \beta_{14} \frac{s}{l} \quad (7b)$$

$$E(P_{sc}/l) = \beta_{15} + \beta_{16} s + \beta_{17} s^2 + \beta_{18} l + \beta_{19} sl \quad (7c)$$

$$\sigma(P_{rep}/l) = \beta_{20} + \beta_{21} s + \beta_{22} \frac{s}{l} \quad (8)$$

Table 4. Empirical constants in (7)-(8) under different conditions.

	No Variations	Case 1	Case 2	Case 3
β_9	0.01827	0.020228	0.01211	-0.015917
β_{10}	14.49846	16.398218	21.57619	37.083546
β_{11}	7.178e-01	7.110e-01	5.724e-01	2.124e-01
β_{12}	-1.273e-03	-1.264e-03	-8.553e-04	3.029e-04
β_{13}	3.201e+02	3.254e+02	4.786e+02	1.197e+03
β_{14}	1.261e+01	1.260e+01	1.213e+01	9.606e+00
β_{15}	2.825e-01	2.848e-01	3.513e-01	2.225e-01
β_{16}	-3.016e-03	-3.058e-03	-3.373e-03	-3.045e-03
β_{17}	7.168e-06	7.371e-06	8.087e-06	9.365e-06
β_{18}	-5.519e-05	-5.283e-05	-5.085e-05	1.566e-05
β_{19}	7.277e-07	7.199e-07	7.216e-07	4.987e-07
β_{20}	0	1.969e-02	1.117e-01	1.200e-01
β_{21}	0	1.244e-06	2.077e-06	3.474e-06
β_{22}	0	6.325e+00	1.512e+01	5.322e+01

Again, the detailed results of model-fitting using linear regression techniques have been omitted from here due to lack of space. From the above analysis, we can conclude that device variations impact both power and delay distribution while interconnect variations impact mainly the delay distribution.

4. STATISTICAL DELAY OPTIMIZATION

Due to parameter variations, the delay no longer remains a deterministic parameter and shows a spread as shown in Figure 14. It is obtained using MC simulations under parameter variations for case 1. The spread in the distribution for case 2 and 3 was more than that in case 1.

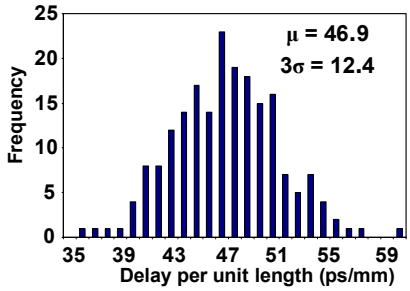

Figure 14. Delay distribution for delay per unit length under parameter variations for Case 1 (see Table 2).

In order to carry out statistical delay optimization, the following quantity (given by (9)) is optimized.

$$E(\tau/l)+3\sigma(\tau/l) \tag{9}$$

Using (9) as an optimization function, the following results were obtained (Table 5). Because of a delay spread introduced due to variations, there is an increase in the maximum delay at optimum point.

Table 5. Results for s_{opt} and l_{opt} obtained by statistical delay optimization under 3 different cases.

	s_{opt}	l_{opt} (μm)	$(\tau/l)^{Max}$ ps/mm
No Var	201.53	3483.0	48.35
Case 1	196.49	3396.4	51.47
Case 2	204.42	3555.4	56.42
Case 3	200.61	3420.8	60.92

5. STATISTICAL POWER OPTIMIZATION

As discussed in Section 1, there exist huge potentials for power savings if one operates slightly away from the optimum point governed by simple delay optimization. In this section, power optimization is carried out for a 5% penalty in delay. Later in the paper, this delay penalty is varied from 5% to 30% and corresponding power savings are shown. Again, here we try to minimize the following quantity based on similar argument as delay.

$$(P_{rep}/l)^{\max} = E(P_{rep}/l)+3\sigma(P_{rep}/l) \tag{10}$$

where P_{rep} is the power dissipated by the repeater and $(P_{rep}/l)^{max}$ is the maximum power dissipation. If the fractional delay penalty to be tolerated is f, then

$$E(\tau/l) = (1+f)(E(\tau/l))_{opt} \tag{11a}$$

$$\beta_1 s+\frac{\beta_2}{s}+\beta_3 l+\frac{\beta_4}{l} = 2(1+f)(\sqrt{\beta_1\beta_2}+\sqrt{\beta_3\beta_4}) \tag{11b}$$

Now using (7)-(8),

$$(P_{rep}/l)^{\max} = (\beta_9+\beta_{11}+\beta_{15}+3\beta_{20})+(\beta_{12}+\beta_{16}+3\beta_{21})s+\beta_{17}s^2 + \beta_{18}l+\beta_{19}sl+\beta_{13}/l+(\beta_{10}+\beta_{14}+3\beta_{22})s/l \tag{12}$$

Setting the derivative of (12) with respect to s to zero, we have

$$(\beta_{12}+\beta_{16}+3\beta_{21})+2\beta_{17}s+\beta_{19}l+\frac{(\beta_{10}+\beta_{14}+3\beta_{22})}{l}+\frac{dl}{ds}[\beta_{18}+\beta_{19}s - \frac{\beta_{13}}{l^2}-\frac{(\beta_{10}+\beta_{14}+3\beta_{22})}{l^2}s] = 0 \tag{13}$$

and dl/ds can be obtained by differentiating (11b) with respect to s, which gives,

$$\frac{dl}{ds} = -(\beta_1-\frac{\beta_2}{s^2})/(\beta_3-\frac{\beta_4}{l^2}) \tag{14}$$

The equations (11b), (13) and (14) can be solved numerically by Newton-Raphson method to obtain new s and l.

6. RESULTS

The results for power-optimal buffer insertion are tabulated in Table 6. Here, s is the new repeater size, s/s_{opt} is ratio of the new repeater size and delay optimal repeater size, l is the interconnect length, l/l_{opt} is the ratio of new interconnect length between successive repeaters and delay optimal interconnect length, P/l is the power dissipation per unit length, $(P/l)/(P/l)_{opt}$ is the power dissipation per unit length as a ratio of power unit length of delay optimal repeater, P/P_{opt} is the power dissipation of a single repeater as a ratio of the power dissipation of the delay optimal repeater. From the table, it is clear that for optimal power dissipation at a given delay penalty, the repeater size needs to be reduced ($s/s_{opt} < 1$) and interconnect lengths between successive repeaters needs to be increased ($l/l_{opt} > 1$) compared to delay optimal case, in agreement with [5]. Also, as parameter variations increase from Case 1 to Case 3, both buffer size and interconnect length between two repeaters needs to be increased as compared to the case where no variations are considered. This is obvious since leakage power increases with increase in parameter variations; hence power optimal buffer insertion methodology should insert less power-consuming repeaters. However, to maintain the same delay as for the case of no variations, buffer size needs to be increased to compensate for the increase in delay due to increase in inter-repeater separation. Also, note that as parameter variations increases, the power savings ($\propto (P/l)_{opt}/(P/l)$) for a given delay penalty increases. This is due to the fact that leakage current increases substantially with parameter variations and therefore reducing the number of buffers for a given interconnect length results in larger savings in power dissipation as also shown in [5]. Also, note that $(P/l)^{max}$ and $(\tau/l)^{max}$ increases as variations increases. The increase in power is mainly contributed by the device variations while the increase in delay is contributed by both device as well as interconnect variations as shown in Figures 9 and 13. Figure 15 shows the error in estimation of mean and maximum power for power optimal case, if one were to ignore the effect of parameter variations. The optimum s and l for power optimal case were calculated ignoring variations, while errors in mean and maximum power were calculated considering variations in account. It is obvious that error in power estimation increases with parameter variations, mainly because both mean as well as maximum leakage power increases substantially due to variations. Since, with scaling, variations are expected to increase, hence the significance of applying above methodology is also expected to increase with scaling. As observed in Table 6, the total power dissipated in cases 1-3 could be much higher than that for the case where no variations are considered. Thus, if under variations, one wishes to operate at the same power levels as dissipated under no variations, much higher delay penalties are required. This effect is quantified in Table 7. Power dissipated under different penalties in delay (for no variation case) is considered, and the delay penalties required to dissipate exactly the same power under three different cases are calculated.

Table 6. Results for power-optimal repeater insertion under different cases.

	s	s/s_{opt}	l (μm)	l/l_{opt}	E(P/l) (uW/mm)	Max(P/l) (uW/mm)	$(P/l)/(P/l)_{opt}$	P/P_{opt}	E(τ/l) (ps/mm)	Max(τ/l) (ps/mm)
No Variations	142.618	0.7076	4586.5	1.3168	1.6883	1.6883	0.6964	0.9170	48.35	48.35
Case 1	145.298	0.7190	4673.7	1.3459	1.7565	2.4063	0.6485	0.8728	48.57	53.96
Case 2	146.882	0.7259	4721.7	1.3617	1.8995	3.6465	0.6255	0.8518	48.95	59.39
Case 3	148.114	0.7335	4752.6	1.3776	2.3507	7.6880	0.5722	0.7882	49.57	63.96

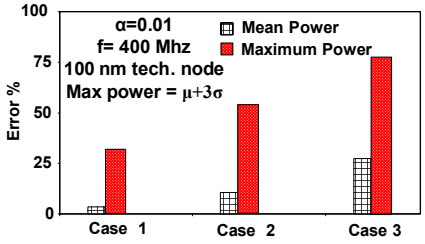

Figure 15. Percentage error in mean and maximum power when operating at deterministic optimum. The percentage error increases with variations.

Table 7. Optimization results for repeater insertion with a desired level of power dissipation. The numbers in the parenthesis in the first column represent the penalties incurred under no variations for operating at desired power level.

$(P_{rep}/l)^{des}$ (μW/mm)	Case 1		Case 2		Case 3	
	τ/l (ps/mm)	% Delay Penalty	τ/l (ps/mm)	% Delay Penalty	τ/l (ps/mm)	% Delay Penalty
2.42 (0%)	48.4	4.8	56.7	21.5	82.8	75.3
1.86 (2.5 %)	53.1	14.9	67.0	43.7	110.2	133.4
1.69 (5 %)	55.8	20.7	72.3	55.1	127.8	170.6
1.58 (7.5%)	58.2	25.7	76.8	64.7	145.6	208.5
1.49 (10%)	60.3	30.3	80.7	73.0	165.1	249.8

This is equivalent to solving (13), (14) and (15) simultaneously.

$$(P_{rep}/l)^{des} = (\beta_9 + \beta_{11} + \beta_{15} + 3\beta_{20}) + (\beta_{12} + \beta_{16} + 3\beta_{21})s + \beta_{17}s^2 + \beta_{18}l + \beta_{19}sl + \beta_{13}/l + (\beta_{10} + \beta_{14} + 3\beta_{22})s/l \quad (15)$$

Here, $(P_{rep}/l)^{des}$ is the desired power dissipation. It is clear from Table 7, that as variations increase, higher penalties in delay are required to dissipate same amount of power. Note that along a row, the power dissipated per unit length remains constant while delay per unit length increases from left to right. Thus, from Table 7, we can conclude that with respect to maximum power dissipation, 0% penalty in delay under no variations is equivalent to delay penalty of 4.8% under case 1, 21.5% under case 2 and 75% under case 3. Furthermore, As the percentage penalty in delay increases, the power savings also increase. For different cases, delay penalty is tabulated in Figure 16. As expected, $(P/l)/(P/l)_{opt}$ decreases with increase in delay penalty. Also, for a given delay penalty power savings are greater under higher percentage of variations, mainly because of the increase in leakage power under variations. Also, plotted are maximum delay curves under different operating conditions. The maximum delay here is normalized to the delay for no variations case. The normalized delay increases because of variations.

7. CONCLUSIONS

This paper addressed the problem of power dissipation during the buffer insertion phase of interconnect performance optimization taking all significant parameter variations into account. Delay was shown to be sensitive to both device and interconnect variations while power, especially leakage power was pre-dominantly sensitive to device variations. Also, variations were shown to cause a significant spread as well as a shift in power dissipation unlike buffer-interconnect delay which shows only a spread around the nominal value. Moreover, accurate statistical empirical models were developed using SPICE and linear regression techniques. The results of statistical power optimal repeater insertion were compared to those of deterministic optimization and it was concluded that a higher repeater size (s) and greater inter-repeater interconnect lengths (l) have to be used for variations aware power-optimal designs. Also, power dissipation under variations was found to be much greater (upto 4.5X) which could result in upto 75% error, if variations are ignored. Furthermore, it was shown that under variations, higher penalties in delay are needed to operate at power levels similar to no variation conditions. Finally, the significance of applying power-optimal repeater insertion methodology was shown to increase with increasing amounts of variation as technology scales.

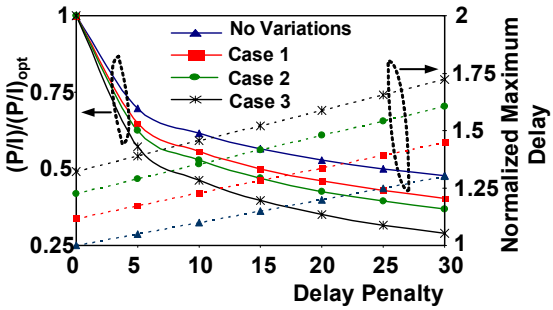

Figure 16. $(P/l)/(P/l)_{opt}$ for different delay penalties under different conditions. Also, normalized maximum delay curves are also drawn.

ACKNOWLEDGEMENTS: This work was supported by Intel Corp. and the University of California- MICRO program.

REFERENCES

[1] International Technology Roadmap for Semiconductors (ITRS), SIA, 2003.
[2] H. B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Addision-Wesley, 1990.
[3] S. Borkar et al., "Parameter variations and impact on circuits and microarchitecture," *DAC*, 2003, pp. 338-342.
[4] P. Gelsinger, *41st DAC Keynote*, DAC, 2004, (www.dac.com)
[5] K. Banerjee and A. Mehrotra, "A power-optimal repeater insertion methodology for global interconnects in nanometer designs," *IEEE Trans. on Elec Dev.*, vol. 49, pp. 2001-2007, Nov. 2002.
[6] M. Eisele et al., "The impact of intra-die device parameter variations on path delays and on the design for yield of low voltage digital circuits," *IEEE Trans. on VLSI Sys.*, Vol. 5, pp. 360-368, 1997.
[7] Z. Lin et al., "Circuit sensitivity to interconnect variations," *IEEE Trans. on Semiconductor Manufacturing*, vol.11, no. 4, pp. 557-568, 1998.
[8] D. Boning and S. Nassif, "Models of process variations in device and interconnect," *Design of high performance microprocessor circuits*, A. Chandrakasan et al., Wiley-IEEE Press, 2000.
[9] K. A. Bowman et al., "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE JSSC*, vol. 37, pp. 183-190, Feb. 2002.
[10] C. Visweswariah et al., "First-order incremental block based statistical timing analysis," *DAC*, 2004, pp. 331-336.
[11] J.A.G. Jess et al., "Statistical timing for parametric yield prediction for digital integrated circuits," *DAC*, 2003, pp. 932-937.
[12] S. Nassif, "Techniques for modeling delay variability," *ICCAD*, 2004, pp. 138-139.
[13] P. S. Zuchowski et al., "Process and environmental variations impacts on ASIC timing", *ICCAD*, 2004, pp. 336-342.
[14] S. Borkar, T. Karnik and V. De, "Design and reliability challenges in nanometer technologies," *DAC*, 2004, pp. 75.
[15] S. Zhang, V. Wason and K. Banerjee, "A probabilistic framework to estimate full-chip subthreshold leakage power distribution considering within-die and die-to-die P-T-V variations," *ISLPED*, 2004, pp. 156-161.
[16] http://www-device.eecs.berkeley.edu/~ptm/cgi-bin/
[17] K. Banerjee and A. Mehrotra, "Analysis of on-chip inductance effects for distributed RLC interconnects," *IEEE Trans. on CAD*, vol. 21, no. 8, pp. 904-915, 2002.
[18] J. Faraway, *Linear Models with R*, CRC Press, 2004.