Modeling Process Variations Using a Compact Model

Raghunath Murali, James D. Meindl

Microelectronics Research Center, Georgia Tech, Atlanta GA. Email: ragu@gatech.edu

ABSTRACT

Inclusion of manufacturing variations has become an important part of static timing analysis. Existing statistical timing analysis methods involve the use of time-consuming circuit simulations or use fit polynomials. In this work a compact model is derived to predict the effect of manufacturing variations on the delay distribution of circuits. The model is physics-based and exhibits excellent match with simulations. The model can be used to rapidly assess the impact of various circuit and device parameters on delay variation and thus is a useful tool for design-for-manufacturing. Use of the model for design space exploration is shown.

Keywords: process variations, threshold voltage, compact model

1 INTRODUCTION

The tremendous increase in the number of transistors on a chip along with sub-wavelength lithography has amplified the effect of manufacturing variations. The impact of manufacturing variations is to increase the mean and variance of the delay distribution, thereby reducing yield. The main sources of manufacturing variations are due to variations in gate-length, threshold voltage, and interconnect [1]. Gate-length fluctuations occur because of lithography and etch process steps, and is currently the major contributor to the chip delay distribution in logic circuits. Threshold voltage variations are caused due to random dopant fluctuations and poly line-edge-roughness. Interconnect RC variations are caused due to the CMP process.

Manufacturing variations can be classified into inter and intra-die variations. Inter-die variations are variations in device characteristics from one die to another, from wafer to wafer or from lot to lot. Intra-die variations are variations in device characteristics between devices on the same die. Intra-die variations have two components - a systematic and a random component. The systematic component exhibits spatial correlation between devices; this correlation falls off with increasing distance between a pair of devices. Inter-die variations result in an increase in delay variance whereas intra-die variations result in an increase in the delay mean [2]. With continued device scaling, within-die variations are becoming as important as die-to-die variations.

Designers are tackling manufacturing variations at different levels of the IC-design hierarchy. At the manufacturing level, process engineers optimize the process for tighter control of device parameters. At the logic-design and circuit-design level, statistical analysis is used to guard-band the delay variations [3]. Timing yield prediction has become more difficult because of process variations. Worst-case analysis tends to overestimate the effect of variations and results in design-effort wastage. Thus accurate gate-level timing models are needed to correctly estimate the impact of manufacturing variations.

Statistical methods for timing analysis mostly use circuit-level Monte Carlo simulations [4], [5]; such simulations become time-consuming for complex circuits. Some timing methods might use fit polynomials for synthesis/optimization [6]. Such a curve-fit method has a limited range of validity and does not allow full exploration of the design space. Thus, there is a need for compact, physics-based models for delay distribution so as to achieve fast circuit optimization. An additional benefit of compact models is that they present a physical insight into the design process. In [7], an analytical model is derived for analyzing the impact of process variations. But the model does not explicitly relate the delay variance to process and circuit parameters. Also, many fitting parameters are used that make the model less physical.

In this work, a compact, physics-based delay variance model is derived that can be used to provide insights into the effect of process/device parameters on the delay variance. The model can be used in a standard-cell based approach to greatly speed-up statistical timing analysis for estimating delay variance and for early circuit design including process variations.

2 COMPACT MODEL

The compact model derived in this section primarily models the effect of gate-length variation and power-supply variation on the delay distribution. This is because the other sources of variation - oxide thickness, threshold voltage variation due to dopant fluctuation,
and interconnect - are much smaller for logic circuits [1]. A decrease in the gate-length causes an increased short-channel effect; this results in a smaller threshold voltage, a higher drain induced barrier lowering (DIBL) and a higher subthreshold slope. Previous studies have shown gate-length variation to be the primary contributor to delay variance [8].

A chain of NAND gates is considered to model the effect of manufacturing variations on the delay distribution. A chain of NAND gates is considered since the delay of a critical path can be simulated by replacing complex CMOS gates with equivalent NAND gates [9]. The delay of such a chain is given by [10]

$$ T_{PD-chain} = \frac{1}{b} n_{CP} T_{PD} $$

(1)

where $b$ is the clock-skew factor, $n_{CP}$ is the number of gates in the critical-path chain, and $T_{PD}$ is the delay of a single gate in the chain. To be able to derive a compact expression for variance of $T_{PD}$, we assume $T_{PD}$ to be

$$ T_{PD} = f_{\text{inff}} \frac{C_L V_{DD}}{I_{\text{dat}}} $$

(2)

where $f_{\text{inff}}$ is the effective fan-in factor to account for the gate stacking effect, $C_L$ is the load capacitance, $V_{DD}$ is the supply-voltage, and $I_{\text{dat}}$ is the saturation current. To model $I_{\text{dat}}$, the alpha-power law model of Sakurai [11] is used:

$$ I_{\text{dat}} = K \mu C_{ox} \frac{W}{L} (V_{DD} - V_t)^\alpha $$

(3)

where $V_t$ is the threshold voltage. Here, $K$ and $\alpha$ would need to be extracted from the process data. To make the $I_{\text{dat}}$ model more physics-based, $\alpha$ could be evaluated from device parameters as in [12]. A threshold voltage model that includes drain-induced barrier lowering (DIBL) and short-channel effect (SCE) has been derived in [13]; this model is physics-based and can be used to project performance of future technology generation. The threshold voltage is modeled as $V_t = V_{IL} + \Delta V_t$ where $V_{IL}$ is the long-channel component of the threshold voltage and $\Delta V_t$ is the short-channel component that includes DIBL and SCE.

Say the pdf of gate-length distribution is $f(L)$, the mean gate-length is $\mu_L$ (= nominal gate-length), and the standard-deviation is $\sigma_L$. Let the pdf of the delay distribution be $f(T_{PD})$, the mean be $\mu_{T_{PD}}$ (evaluated from (2) with $L = \mu_L$), and the standard-deviation be $\sigma_{T_{PD}}$. The pdf of gate-length, $f(L)$, is taken to be a Gaussian [2]. Then, $\sigma_{T_{PD}}$ can be evaluated from

$$ \sigma_{T_{PD}} = \sigma_L \frac{d T_{PD}}{dL} $$

(4)

d$T_{PD}/dL$ can be evaluated from (2). Evaluation of $d T_{PD}/dL$ involves the evaluation of $dI_{\text{dat}}/dL$ which in turn requires the evaluation of $dV_t/dL$. This series of dependencies accounts for the inclusion of SCE and DIBL effects into the final equation. Thus,

$$ \sigma_{T_{PD}} = \sigma_L \mu_{T_{PD}} \left[ \frac{\alpha \Delta V_t}{V_{DD} - V_t} \right] \left[ 1 + \left( \frac{\theta}{\Delta L_{eff}} \right) \right] $$

(5)

where $L_{\text{eff}} = L - \Delta L$ is the effective channel length, $\theta = \frac{1}{\sqrt{1 + \frac{\Delta L_{eff}}{\Delta L}}}$ and $\kappa$ is a parameter that depends on the device technology, For the 130nm technology considered in this work, $\kappa$ is 0.54 for NMOS and 1.50 for PMOS transistors.

In the analysis of process variations, both die-to-die (D2D) and within-die (WID) fluctuations need to be considered [2]. The D2D standard deviation for a gate-chain is given by the sum of the gate $\sigma_{T_{PD}}$. For WID variations that are un-correlated, the total variance is given by the RMS sum of the individual variances. For WID variations that have spatial correlations, the delay standard deviation can be calculated from the covariance matrix once $\sigma_{T_{PD}}$ for the individual gates are calculated.

3 MODEL VERIFICATION

Monte Carlo simulations using HSPICE are used to verify the accuracy of the model. A 130nm generation technology with 70nm gate-length transistors is selected and the BSIM parameters are obtained from [14]. The model parameters $K$ and $\alpha$ are extracted from I-V data, and $\kappa$ is extracted from transient simulations. Since there are only these parameters to extract and since only...
a small data set is needed for extraction, the extraction is simple and fast. A NAND3 gate with a step-input is simulated, and the mean and variance information is extracted from Monte Carlo simulations. An excellent match is obtained between the model and simulations for the mean delay. Fig. 1 shows the results for the delay standard-deviation, for a rising input. In the inset of Fig. 1, results are shown for a NAND3 gate with a falling input. Since the technology under consideration has worse SCE and DIBL for PMOS transistors, $\sigma_{TPD}$ is higher (while $\mu_{TPD}$ remains the same) for a falling input. The model is seen to work well over a wide range of $V_{DD}$.

Next, D2D variations on a chain of NAND3 gates are considered. The number of gates in the chain is varied so as to test the model over a range of conditions. Fig. 2 shows the match for the variance of the delay; it can be seen that as $n_{CP}$ increases, $\sigma_{TPD}$ almost remains constant. This can be explained from the following equation.

$$\frac{\sigma_{TPD-chain}}{T_{PD-chain}} = \frac{b}{2T_{PD}} (\sigma_{TPD_n} + \sigma_{TPD_p})$$

where $T_{PD-chain} = n_{CP} T_{PD}/b$. Thus $(\sigma/\mu)$ of the gate-chain is independent of $n_{CP}$ when D2D variations are considered.

Fig. 3 shows the effect of gate-length variation. It can be seen that the model is able to predict the delay mean and standard deviation with good accuracy. It can be seen that by increasing gate-length by 10nm, the delay standard-deviation can be reduced by 13%. This will be further explored in the next section.

4 MODEL APPLICATIONS

The compact model for $\sigma_{TPD}$ can be used to rapidly explore a design space so as to achieve the minimum delay variation for a given set of process and circuit parameters. A circuit designer has to meet constraints such as power, frequency, and area. The maximum allowable $V_{DD}$ is set by the power limit and the minimum allowable $V_{DD}$ is set by the frequency limit. Also, the foundry places a limit on the best achievable $\sigma_L$, which is typically 4% of the nominal gate-length [15]. For a given $\sigma_{TPD}$, the design window is composed of the power, frequency, manufacturing, and the delay variation limits. Such a design window is shown in Fig. 4 for various contours of $\sigma_{TPD}$ on a $V_{DD} - \sigma_L$ plane. As $\sigma_L$ increases, a higher $V_{DD}$ is needed to achieve the same delay variation as before; this explains the shape of the contours.

The design window is seen not to exist for $(\sigma/\mu)_{TPD} = 4\%$. For $(\sigma/\mu)_{TPD} = 6\%$, we obtain a small design window (shaded area). Gate-length bias reduces the effect of threshold-voltage variations and improves the delay variance. When gate-length is increased from 70nm to 80nm, it can be seen that the new design window (shaded area) is now bigger than before. This relaxes the manufacturing constraint and increases yield.

5 CONCLUSIONS

Process variations are becoming a dominant factor in the design of circuits. This work has presented a circuit optimization methodology using a compact model. The model is physics-based, uses few fitting parameters and matches well with Monte Carlo simulations, for both
Figure 4: Design window with gate-length bias. The solid lines are contours for a gate-length of 80nm whereas the dashed lines are contours for a gate-length of 70nm.

single gates and chains of gates. Spatial correlations can be included in circuit analysis using the model and results in huge time savings compared to SPICE simulations. Thus the model is ideal for analyzing critical paths in early-stage circuit design. The model allows for optimization of circuit parameters to obtain increased yield. A design plane with $\sigma_{V_{PD}}$ contours was shown.

REFERENCES


