8

Modeling Process Variability in Scaled MOSFETs

8.1 Introduction

This chapter presents compact MOSFET (metal-oxide-semiconductor field-effect transistor) modeling approaches for process variability-aware VLSI (very-large-scale-integrated) circuit CAD. The circuit design for advanced VLSI technology is severely constrained by random and systematic process variability [1]. With continued miniaturization of MOSFET devices [2–8], performance variability induced by process variability has become a critical issue in the design of VLSI circuits using advanced CMOS (complementary metal-oxide-semiconductor) technologies. Process variability in scaled CMOS technologies severely impacts the delay and power variability in VLSI devices, circuits, and chips, and this impact keeps increasing as MOSFET devices and CMOS technologies continue to scale down [1,9–13]. The increasing amount of within-die process variability on the yield of VLSI circuits, such as static random access memory (SRAM), has imposed an enormous challenge in the conventional VLSI design methodologies. Similarly, the chip mean variation due to across-the-chip systematic process variability also imposes serious challenge in the conventional VLSI circuit design methodologies. Because of process variability constraints, an advanced VLSI circuit, optimized using the conventional design methodology, is more susceptible to random performance fluctuations. Thus, new circuit design techniques to account for the impact of process variability in VLSI circuits have become essential [1,9]. And, compact model addressing the impact of random and systematic process variability in scaled MOSFET devices is crucial for the simulation and analysis of advanced VLSI circuits. Process variability in manufacturing technology includes *front-end* or intrinsic process variability due to various dopant implant and thermal processing steps and *back-end* variability of metal lines for interconnecting the devices in the VLSI circuits. Both the front-end and interconnection process variabilities are important for circuit analysis. Over the years, different approaches have been used to develop statistical models for circuit analysis to account for intrinsic process

variability [9,14–21]. In this chapter, we will present different approaches to develop statistical compact MOSFET device models to account for the front-end process variability in VLSI circuit design.

8.2 Sources of Front-End Process Variability

The intrinsic sources of variability in VLSI device performance arise from the random variability of fabrication-processing steps [1,9–11]. Typically, the intrinsic process variability is grouped as *systematic* and *stochastic* or random as shown in Figure 8.1.

8.2.1 Systematic or Global Process Variability

Systematic variation in IC (integrated circuit) device and chip performance is caused by inherent random character of IC-processing steps. The systematic component is defined as the global or inter-die process variability [1,9–13]. The global process variability causes die-to-die, wafer-to-wafer, or lot-to-lot systematic parametric fluctuations between identical devices [1,9–13]. Global variability causes a shift in the mean value of the sensitive design parameters, including the channel length (L), channel width (W), gate oxide thickness (T_{ox}), resistivity, doping concentration, and body effect as shown in Figure 8.2. Systematic differences may lead to longer channel length transistors than the nominal devices, causing them to switch more slowly due to reduced drive current, resulting in slower ICs with lower leakage current. On the other hand, the shorter (than the nominal) channel length devices would lead to faster die easily meeting clock-frequency specifications; however, these devices may exhibit excessive leakage current and fail leakage

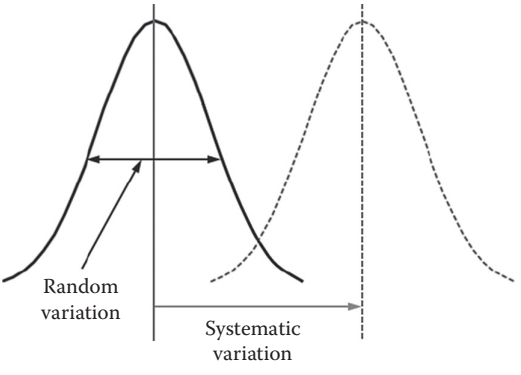

FIGURE 8.1
Types of process variability: random variation of a parameter around its mean value and systematic variation of the mean value of a parameter.

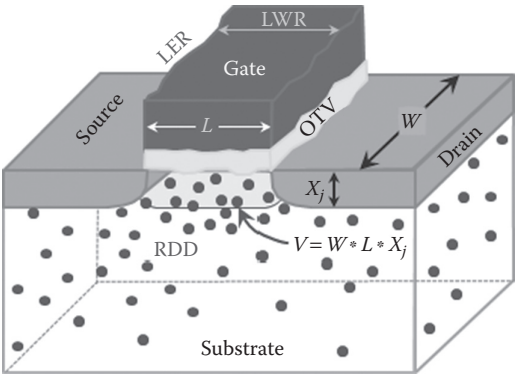

FIGURE 8.2
The critical sources of variability in CMOS devices; here L, W, and X_j are the channel length, channel width, and S/D junction depth of MOSFET devices, respectively; and V is the volume of charge in the channel region. (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.)

current specifications. In the semiconductor industry, the systematic process variation has been the major interest to IC chip manufacturers in maintaining competitive yield over multiple technology nodes [4]. The systematic process variability in manufacturing technology has been accounted in compact modeling by defining process corners, that is, boundaries in parameter variation that account for process tolerances [1,9].

8.2.2 Random or Local Process Variability

The random variation in IC device and chip performance arises from stochastic variations inherent to the discrete nature of dopant impurities and point defects as well as random variations in the complex processing steps of nanometer scale CMOS technology. Random variability is defined as the *local* or *intra-die* process variability [1,9]. Local process variability causes parametric fluctuations or mismatch between identically designed devices within a die as shown in Figure 8.1. The major sources of intrinsic process variability in advanced CMOS technologies include random discrete doping (RDD), line-edge roughness (LER), line-width roughness (LWR), and oxide thickness variation (OTV) as shown in Figure 8.2 [1,9–13].

The front-end process variability in CMOS technology has been extensively studied and the major sources of process variability along with their impact on device and VLSI circuit performance have been reported [1,9–11]. In the following section, a brief overview of the sources of front-end process variability is presented.

8.2.2.1 Random Discrete Doping

In the channel region of a MOSFET device, RDD results from the discreteness of dopant atoms as shown in Figure 8.2. In a MOSFET device, the channel

region is doped with dopant atoms to control its threshold voltage (V_{th}) [5–8]. For a device with channel doping concentration N_{CH} and source/drain (S/D) junction depth X_j, the total number of dopant atoms in the channel region is given by [1,9]:

$$N_{CHtotal} \cong N_{CH} \cdot W \cdot L \cdot X_j \tag{8.1}$$

Equation 8.1 shows that the continuous scaling down of L, W, and X_j causes the total number of dopants in the channel to decrease, despite the corresponding increase in the channel-doping concentration according to the CMOS scaling rule [2,3]. Using Equation 8.1 and the target specifications for advanced CMOS technology scaling by *International Technology Roadmap for Semiconductors* [22], the estimated decrease in $N_{CHtotal}$ over the scaled technology nodes is shown in Figure 8.3. Figure 8.3 implies that the number of dopants in a transistor channel is a discrete statistical quantity with probability to occupy any random location. Therefore, in an advanced CMOS technology, two identical transistors next to each other have different electrical characteristics because of the randomness in a few dopant atoms, resulting in intra-die device and circuit performance variability.

The major effects of RDD include significant variability in V_{th}, variability in the overlap capacitance (C_{ov}) due to the uncertainty in the position of S/D dopants under the gate, and variability in the effective S/D series resistance (R_{DS}). The impact of RDD-induced process variability on V_{th} mismatch between two identically designed within-die devices is given by [9]

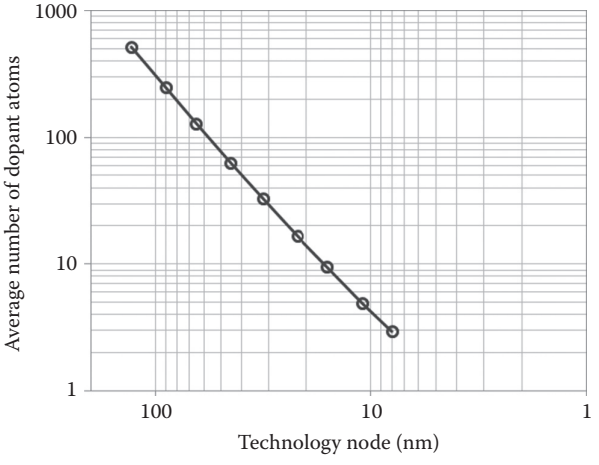

FIGURE 8.3
Estimated average channel doping concentration with scaling bulk CMOS devices in the nanoscale regime; the calculation is performed following ITRS. (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.)

$$\sigma V_{th,RDD} \cong C \cdot \left(\sqrt[4]{q^3 \varepsilon_{si} \phi_B} \right) \frac{T_{ox}}{\varepsilon_{ox}} \left(\frac{\sqrt[4]{N_{CH}}}{\sqrt{W_{eff} L_{eff}}} \right) \tag{8.2}$$

where:

C is a number and is given by 0.8165 [23] or 0.7071 [24] with or without the dopant variation along the depth of the channel region, respectively

q is the electronic charge

ε_{si} and ε_{ox} are the permittivity of silicon and silicon-dioxide (SiO_2), respectively

$\phi_B = v_{kT} \ln\left(N_{CH}/n_i\right)$ is the bulk potential of the channel region of MOSFETs and v_{kT} and n_i are the thermal voltage and intrinsic carrier concentration, respectively

W_{eff} and L_{eff} represent the effective dimension of W and L, respectively

Since the device area ($W_{eff} \times L_{eff}$) decreases with each new technology generation, it is obvious from Equation 8.2 that the net result of RDD is a significant increase in process variability for scaled CMOS technology as shown in Figure 8.4. In fact, RDD is a major contributor to mismatch (σV_{th}) in advanced MOSFETs [25]. As the device size scales down, the total number of channel dopants decreases [1,9], resulting in a larger variation of dopant numbers, and significantly impacting V_{th} as shown in Figure 8.4.

Equation 8.2 is the generalized analytical expressions for σV_{th} in planar devices due to RDD that represents σV_{th} equations derived by Stolk et al. [23] and Mizuno et al. [24] with appropriate value of the parameter C [9]. For devices of a particular process technology, Equation 8.2 can be expressed as

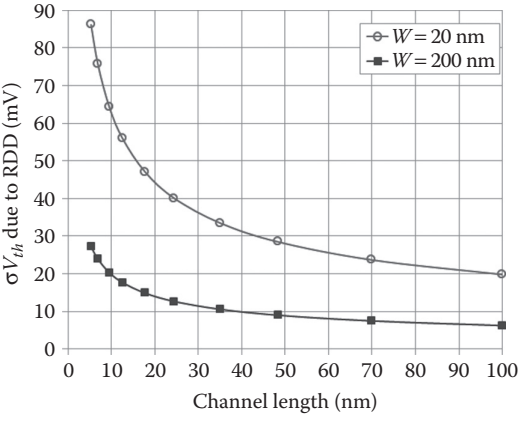

FIGURE 8.4
Estimated threshold voltage variation for a typical 20-nm bulk CMOS technology as a function of device channel length for different channel width following ITRS (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.); parameters used in Equation 8.2 are $N_{CH} = 6 \times 10^{18}$ cm^{-3}; SiO_2 equivalent oxide thickness (EOT) = 1.1 nm; and $C = 0.8165$.

$$\sigma V_{th,RDD} = \frac{C_{vt}}{\sqrt{W_{eff} L_{eff}}} \tag{8.3}$$

where C_{vt} is a process technology dependent constant given by

$$C_{vt} \cong C \cdot \left(\sqrt[4]{q^3 \varepsilon_{si} \phi_B N_{CH}} \right) \frac{T_{ox}}{\varepsilon_{ox}} \tag{8.4}$$

Equation 8.3 shows that V_{th} variability due to RDD is inversely proportional to the square root of device area. Thus, Equation 8.3 can be used to model the impact of RDD-induced V_{th} variability and the parameter C_{vt} is the slope of $\sigma V_{th,RDD}$ versus $1/\sqrt{W_{eff} L_{eff}}$ plots obtained from a large set of measurement data on a set of paired devices with varying W and L.

8.2.2.2 Line-Edge Roughness

In CMOS technology, LER results from sub-wavelength lithography and etching processes that cause variation in the critical dimension of the transistor feature size, as shown in Figure 8.2 [26]. The impact of LER includes variation in V_{th} and higher subthreshold current. LER-induced V_{th} mismatch depends on the variability in W_{eff} of MOSFETs and is given by [1,9,26,27]

$$\sigma V_{th,LER} \propto \frac{1}{\sqrt{W_{eff}}} < \sigma V_{th,RDD} \tag{8.5}$$

Thus, LER increases as VLSI technology scales down. In scaled MOSFET devices, LER has become a larger fraction of L and a major source of intrinsic statistical variation, causing significant variability in VLSI device and circuit performance. The mismatch due to LER and RDD is statistically independent and can be modeled independently [1,9,26].

8.2.2.3 Oxide Thickness Variation

In CMOS technologies, OTV, shown in Figure 8.2, is caused by atomic level interface roughness between silicon and gate dielectric and remote interface roughness between gate material and gate dielectric, hereafter referred to as the *surface roughness (SR)*. This SR causes fluctuations of the voltage drop across the oxide layer, resulting in V_{th} variation [1,9,28,29]. In nanoscale MOSFETs, OTV is becoming more dominant as T_{ox} approaches the length of a few silicon atoms and is comparable to the thickness of interface roughness.

In nano-MOSFET devices, OTV causes significant device parameter variability. In polysilicon gate MOSFETs, OTV introduces a gate current (I_g)

variation. This I_g variation induces a voltage drop in the polysilicon gate and significantly changes V_{th}. In addition, the device transconductance g_m changes significantly because of the reduction in the gate voltage V_{gs} due to the voltage drop in the polysilicon gate. In high-k gate dielectric and metal gate devices, OTV introduces significant mobility degradation [1,9].

8.2.2.4 Other Sources Process Variability

Other sources of process variability include variation associated with polysilicon as well as metal gates granularity [30,31]; variation in fixed charge [32] and defects and traps in gate dielectric [33]; variation associated with patterning proximity effects such as optical proximity correction [34]; variation associated with polish such as shallow trench isolation [35] and gate [36]; variation associated with the strain such as in wafer-level biaxial strain [37], high-stress capping layers [38], and embedded silicon germanium (SiGe) [39]; and variation associated with implants and anneals due to implant tools, the implant profile, and millisecond annealing [40,41].

Thus, from the above discussions, it is clear that the advanced CMOS process technologies introduce within-die random performance variability, which causes severe variability in the performance of advanced VLSI circuits and systems. Therefore, it is critical to accurately model process variability when predicting the performance of advanced VLSI circuits and systems.

8.3 Characterization of Parametric Variability in MOSFETs

The random parametric variation such as threshold voltage variation (σV_{th}) is a key factor in determining the yield of memory elements such as SRAM and register file cells. Equation 8.3 can be used to characterize random V_{th} variation in devices.

8.3.1 Random Variability

In Figure 8.1, the random variability of a parameter is defined as the variation around its mean value. Therefore, random variability can be characterized by monitoring the differences in the value of a parameter of two closely spaced identical transistors, that is, paired transistor. Thus, the random V_{th} variation of identical transistor pairs can be determined by measuring the difference in V_{th} (i.e., ΔV_{th}) between a number of sets of closely spaced paired transistors (e.g., all the transistor pairs on a wafer) and computing the standard deviation of the difference ΔV_{th} (i.e., $\sigma\Delta V_{th}$). Thus

$$\sigma_{random\text{-}pair} = \sigma(V_{th1} - V_{th2}) = \sigma\left(\Delta V_{th}\right) \equiv \sigma\left(\sigma V_{th}\right) \tag{8.6}$$

In order to determine the local V_{th} variability (also referred to as the *mismatch*) in devices, V_{th} shifts (ΔV_{th}) between the closely spaced identical paired transistors are measured for a large number of pairs. Any standard procedure can be used to extract V_{th} for individual devices. Typically, the V_{th} extraction procedure is used on a set of device geometries (L, W) for both *n*-channel and *p*-channel MOSFETs. Then, $\sigma\Delta V_{th}$ is plotted as a function of $1/\sqrt{W_{eff}L_{eff}}$ as shown in Figure 8.5, which is known as Pelgrom plot [42]. Pelgrom plot presents the V_{th} variability ($\sigma\Delta V_{th}$) extracted for various (L, W) gate dimensions.

The slope A_{vt} of Pelgrom plot as shown in Figure 8.5 is called the *mismatch coefficient* and describes the mismatch between closely spaced identical transistor pairs [42,43]. Thus, we can write

$$A_{vt} = \sigma\left(\Delta V_{th}\right) \cdot \sqrt{W_{eff} \cdot L_{eff}} \tag{8.7}$$

In order to determine the local V_{th} variation of an individual transistor in the pair, we consider

$$\sigma^2_{\Delta V_{th}} = \sigma^2_{V_{th1}} + \sigma^2_{V_{th2}} - 2 \cdot \rho \cdot \sigma_{V_{th1}} \cdot \sigma_{V_{th2}} \tag{8.8}$$

where:

V_{th1} and V_{th2} are the threshold voltages of the transistors 1 and 2 in the pair

ρ is the correlation coefficient between V_{th1} and V_{th2} fluctuations

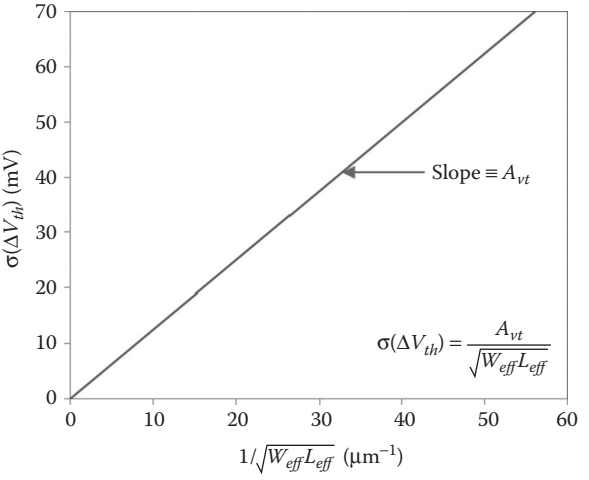

FIGURE 8.5
Plot of $\sigma(\Delta V_{th})$ versus $\left(\sqrt{W_{eff}L_{eff}}\right)^{-1}$ of identically designed paired transistors; A_{vt} is the slope of the plot.

Note that the fluctuations on one transistor of the pair cannot induce fluctuations on the second one (i.e., σV_{th1} and σV_{th2} are independent); thus, $\rho = 0$. In addition, $\sigma V_{th1} = \sigma V_{th2} \equiv \sigma V_{th}$ (i.e., $V_{th1} - V_{th2} = V_{th2} - V_{th1}$). Therefore, defining $\sigma V_{th1} = \sigma V_{th2} \equiv \sigma V_{th}$ from Equation 8.8, we can show that

$$\sigma V_{th} = \frac{\sigma \Delta V_{th}}{\sqrt{2}} \tag{8.9}$$

Equation 8.9 describes σV_{th} of individual transistors of the closely spaced pair. Equation 8.9 can be experimentally verified by comparing the local V_{th} variability obtained either in paired transistors (which give a value of $\sigma(\Delta V_{th})$) or in dense transistor arrays (which give a value of (σV_{th})) [44]. Note that the A_{vt} factor is defined historically from $\sigma(\Delta V_{th})$. Therefore, in order to develop compact variability model to simulate mismatch between identical devices, we get σV_{th} from Equations 8.7 and 8.9, as

$$\sigma V_{th} = \frac{A_{vt}}{\sqrt{2}} \cdot \frac{1}{\sqrt{W_{eff} L_{eff}}} \tag{8.10}$$

Comparing Equations 8.3 and 8.10 we get, $C_{vt} = A_{vt}/\sqrt{2}$; Thus, we can estimate the mismatch coefficient A_{vt} for any technology from Equation 8.4 using the technology parameters T_{ox} and N_{CH}. However, A_{vt} is extracted from the measured data from a set of closely spaced identical paired transistors.

The same procedure is used to determine the mismatch σP of any parameter P between closely spaced identical devices with mismatch coefficient A_p such that

$$\sigma P = \frac{A_p}{\sqrt{2}} \frac{1}{\sqrt{W_{eff} L_{eff}}} \tag{8.11}$$

8.3.2 Systematic Variability

As shown in Figure 8.1, the systematic or global variability is the shift of the mean value of a parameter. Therefore, global variability is obtained simply by calculating the standard deviation (σ) of any parameter P causing systematic variability. Thus, the systematic variability of V_{th} is characterized by calculating σV_{th} of the total V_{th} population, that is, V_{th} data from the target MOSFET test structures distributed across the wafer. The total V_{th} population could include devices from several wafers of a lot or from several lots collected over a period of time. The same procedure is used to determine the systematic variation σP of any parameter causing global device performance variability.

8.4 Conventional Process Variability Modeling for Circuit CAD

In order to account for process variability in circuit performance, typically corner models are used to set the lower and upper limits of process variation. These models are implemented in the process design kit to support process variability-aware VLSI circuit design.

8.4.1 Worst-Case Fixed Corner Models

In conventional circuit design technique, process variability is modeled by *four* worst-case corners: two for analog applications and two for digital [1,9]. The corners for analog applications are generated from slow NMOS (*p*-type body with *n*+ source-drain) and slow PMOS (*n*-type body with *p*+ source-drain; SS) to model the worst-case speed and from fast NMOS and fast PMOS (FF) to model the worst-case power. The corners for digital applications are generated from fast NMOS and slow PMOS (FS) to model the worst-case "1" and from slow NMOS and fast PMOS (SF) to model the worst-case "0". A standard set of model parameters (e.g., V_{th}) is used to account for process variability and model worst-case corner performance of the devices and circuits for the target CMOS technology [1,9].

In this modeling approach, the standard deviation (σ) limits are preset pessimistically to include any potential process variability over a wide range. The worst-case corner models are generated by offsetting the selected compact-model parameter, *P*, of the typical (TT) compact model by $\pm dP = n\sigma$ to account for the window of process variability, where *n* is the number of σ for *P*. Typically, $3 \le n \le 6$ is selected to set the fixed *lower limit* (LL) and *upper limit* (UL) of the worst-case models; and TT is the typical compact model extracted from the *golden die* of *golden wafer*, representing the centerline process technology [9]. For example, the TT model parameter V_{TH0} of BSIM4 [45] corner models is defined as $V_{TH} = V_{TH0} \pm dvth$, where *dvth* is used to set the target LL and UL of the worst-case models.

To obtain the worst-case corner of drain current I_{ds}, let us consider the basic I_{ds} expression in the ON state (saturation regime) of a large MOSFET device [46] (Equation 4.87)

$$I_{ds} \cong \left(\frac{W}{2L}\right)\mu_{eff}C_{ox}\left(V_{gs}-V_{th}\right)^2; \quad 0 < \left(V_{gs}-V_{th}\right) < V_{ds} \tag{8.12}$$

where:

μ_{eff}, C_{ox}, and V_{ds} are the inversion carrier mobility, gate oxide capacitance, and drain-to-source voltage, respectively

$(V_{gs} - V_{th}) \equiv V_{dsat}$

the remaining parameters have their usual meanings as defined in Chapters 4 and 5

Then, the UL is set by taking the appropriate maximum or minimum offset of model parameters to maximize the value of I_{ds}. Thus, the UL of ION, defined at $V_{ds} = V_{dsat}$ for nMOSFETs is given by:

$$IONN(UL) \cong \mu_{eff}\left(\frac{W+dW}{2(L-dL)}\right)\left(\frac{\varepsilon_{ox}}{T_{ox}-dT_{ox}}\right)\left[V_{gs}-\left(V_{th}-dV_{th}\right)\right]^2 \tag{8.13}$$

In Equation 8.13, W is increased by dW, L is reduced by dL, T_{ox} is reduced by dT_{ox}, and V_{th} is reduced by dV_{th} to achieve the UL of ION specification. Similarly, the LL for ION is set by

$$IONN(LL) \cong \mu_{eff}\left(\frac{W-dW}{2(L+dL)}\right)\left(\frac{\varepsilon_{ox}}{T_{ox}+dT_{ox}}\right)\left[V_{gs}-\left(V_{th}+dV_{th}\right)\right]^2 \tag{8.14}$$

The FF corner is obtained using the UL values of the selected model parameters for both NMOS and PMOS devices whereas SS corner is obtained considering the LL values of the selected model parameters for both NMOS and PMOS devices. The SF corner is derived using LL values of NMOS and UL values of PMOS model parameters. Similarly, The FS corner is derived using UL values of NMOS and LL values of PMOS model parameters.

Figure 8.6 shows ION plots for both nMOSFET and pMOSFET devices obtained by fixed corner models along with the distribution of electrical test (ET) data. It is observed from Figure 8.6 that the simulation results obtained by fixed corner models are too wide, so they could end up rejecting a valid

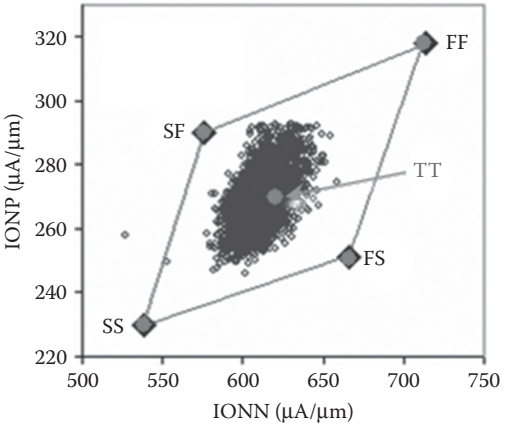

FIGURE 8.6
Distribution of measurement and simulation data generated using fixed corner models: NMOS ON current (IONN) versus PMOS ON current (IONP). (FF: fast NMOS and fast PMOS; FS: fast NMOS and slow PMOS; SF: slow NMOS and fast PMOS; SS: slow NMOS and slow PMOS). (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.)

design, causing yield loss. The major problems with the worst-case corner models are that in most cases the existing correlations between the device parameters are ignored and the models include pessimistic corner values. As a result, the models generate a large spread of data during analog circuit simulation.

The worst-case corner models offer designers capability to simulate the pass/fail results of a typical design and are usually pessimistic.

8.4.2 Statistical Corner Models

During IC chip manufacturing, a large set of ET data on critical device and process parameters are collected for process monitoring. Therefore, unlike fixed corner models, statistical corner models can be generated using ET data from different die, wafers, and wafer lots collected over a certain period of time to represent realistic process variability of a target technology [14–21].

In one approach, ET data are collected from a large number of sites of the target technology. And, for each site of ET data a compact model file is generated. Thus, a large number of compact model files, referred to as the *performance-aware model* (PAM) cards, are generated for the target technology [17,18]. In this approach about 1000 PAM cards or model files are generated for realistic statistical analysis of circuit performance.

In another approach, ET data are used to determine the depth of the location of device parameters in the distribution to generate corner models, referred to as the *location depth corner modeling* (LDCM) [19]. In LDCM, the wafers corresponding to the extreme data points in the distribution are used to extract separate compact models. Thus, using LDCM, the number of model cards is reduced significantly (<20) in contrast to PAM. An enhanced LDCM is used with proper guard banding to ensure design validation against future process shift from the baseline specifications [19].

8.4.3 Process Parameters–Based Compact Variability Modeling

The statistical modeling approach, referred to as the *backward propagation of variance* (BPV) [20], formulates statistical models as a set of independent, normally distributed process parameters. These parameters control the variations seen in the device electrical performances through the behavior described in the TT compact models. With recent extensions [21], BPV is used to characterize physical process–related compact model parameters. For an accurate analysis of process variability–induced circuit performance variability using BPV, the TT model file must be physical, the sensitivity matrix must be well-conditioned, and the variances of parameters must be physically consistent.

8.5 Statistical Compact Modeling

In the conventional variability modeling approaches, a standard set of model parameters are used for fixed corner modeling or a large number of model files are generated from ET data. The fixed corner models are inadequate, whereas, ET data–based modeling is resource-intensive. Therefore, an analytical technique to obtain the process-sensitive compact model parameters of any industry standard compact model to generate compact variability model library for circuit analysis is crucial for variability-aware circuit design as described in the following section.

A generalized approach for process variability modeling is shown in Figure 8.7. The method includes selection of target compact model, consideration of basic I_{ds} expression, derivation of a generalized expression of I_{ds} variance, selection of device parameters causing process-induced I_{ds} variation,

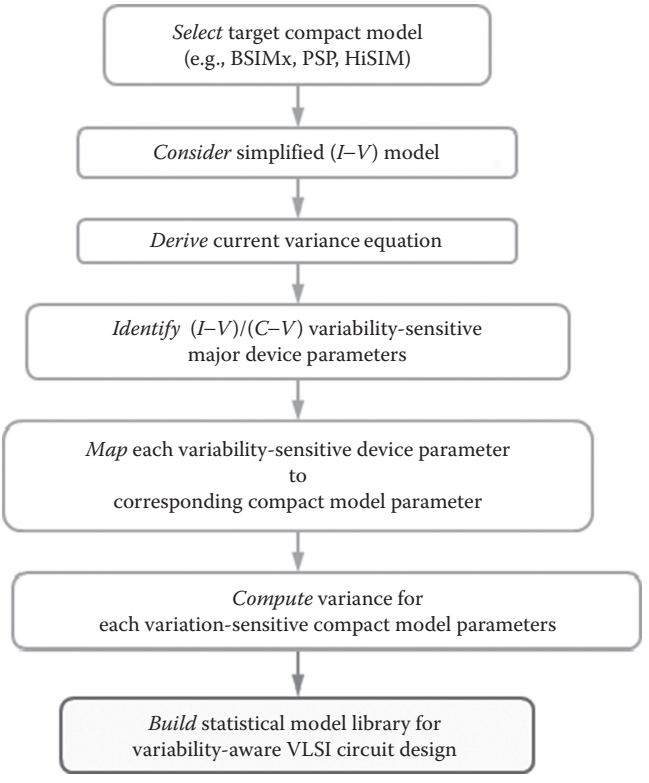

FIGURE 8.7
Generalized modeling approach for process variability-aware VLSI circuit design; here, BSIM*x*, PSP, and HiSIM represent industry standard MOSFET compact models; where $x = 3$, 4, and 6. (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.)

mapping process-sensitive device parameters to corresponding compact model parameters, determination of variances for mismatch modeling and global variability modeling, and finally, building compact variability model.

The modeling methodology outlined in Figure 8.7 is described in the following section.

8.5.1 Determination of Process Variability-Sensitive MOSFET Device Parameters

It is clear from our discussions in Section 8.2 that process variability causes variability in MOSFET device performance, which in turn causes variability in VLSI circuit performance. Since, the MOSFET device performance is determined by I_{ds}, in order to determine the impact of process variability on circuit performance, we determine the process variability-sensitive device parameters causing I_{ds} variability. For the selection of major process variability-sensitive device parameters, we consider the basic I_{ds} model in the subthreshold, linear, and saturation regions of MOSFETs (Equations 4.122 and 4.135)

$$I_{ds} \cong \begin{cases} \left(\frac{W}{L}\right)\mu_{eff}C_{ox}(n-1)v_{kT}^2 e^{(V_{gs}-V_{th})/nv_{kT}}\left(1-e^{-(V_{ds}/v_{kT})}\right); & (V_{gs}-V_{th})<0 \\ \left(\frac{W}{L}\right)\mu_{eff}C_{ox}\left(V_{gs}-V_{th}-\frac{V_{ds}}{2}\right)V_{ds}; & 0<(V_{gs}-V_{th})>V_{ds} \\ \left(\frac{W}{2L}\right)\mu_{eff}C_{ox}(V_{gs}-V_{th})^2; & 0<(V_{gs}-V_{th})\le V_{ds} \end{cases} \quad (8.15)$$

where the parameters have their usual meanings as defined in Chapter 4. From Equation 8.15, we can determine the major device parameters most sensitive to process variability in each region of MOSFET device operation.

8.5.1.1 Selection of Local Process Variability-Sensitive Device Parameters

The local process variability or mismatch between identically designed transistors is caused by microscopic process that makes every transistor different from its neighbors [1,9–13]. As a result, a device parameter P can be considered as consisting of a fixed component P_0 and a randomly varying component p resulting in different values of P for closely spaced identical paired transistors. Then the difference ΔP between two identical transistors within a die is a randomly varying parameter and is defined as the "mismatch" in P between two identical paired transistors. For a large number of samples, ΔP converges to a Gaussian distribution with zero mean. Then the mismatch in relative drain current, $\Delta I_{ds}/I_{ds}$, between paired transistors due to P is given by [47]:

$$\sigma^2_{\Delta I_{ds}/I_{ds}} = \sum_{i=1}^{l}\left(\frac{1}{I_{ds}}\frac{\partial I_{ds}}{\partial P_i}\right)^2 \sigma^2_{\Delta P_i} + \frac{2}{I^2_{ds}}\sum_{i=1}^{l}\frac{\partial I_{ds}}{\partial P_i}\frac{\partial I_{ds}}{\partial P_{i+1}}\rho\left(\Delta P_i, \Delta P_{i+1}\right) \tag{8.16}$$

where l is the total count of ΔP contributing to I_{ds} mismatch; ΔP_i is the ith count of ΔP with standard deviation $\sigma_{\Delta P_i}$; and $\rho\left(\Delta P_i, \Delta P_{i+1}\right)$ is the correlation between ΔP_i and ΔP_{i+1}. Since ΔP_i is random and independent, the correlation $\rho\left(\Delta P_i, \Delta P_{i+1}\right) = 0$ as discussed in Section 8.3.1. In order to model I_{ds} mismatch between paired transistors, we determine the major local process variability-sensitive device parameters P.

From Equation 8.15, we find that for all regions of MOSFET device operation, the value of I_{ds} depends on a common set of parameters $\{V_{th}, W, L, C_{ox}, \mu_{eff}, V_{gs}, V_{ds}\}$. We know $C_{ox} = f(T_{ox})$; then considering only parametric variation in Equation 8.16, ΔP represents any of the mismatch parameters of the set $\{\Delta V_{th}, \Delta W, \Delta L, \Delta T_{ox}, \Delta\mu_{eff}\}$. It is to be noted that the parameter set $\{\Delta W, \Delta L, \Delta T_{ox}, \Delta\mu_{eff}\}$ describes the mismatch in current gain, $\beta = [(W/L)C_{ox}\mu_{eff}]$, defined in Equation 4.74.

Again, V_{th} can be expressed as $V_{th} = f\left(V_{th0}, \gamma, \phi_s, V_{bs}\right)$, where V_{bs} is the applied body bias and $V_{th0} = V_{th}$ at $V_{bs} = 0$ whereas γ and ϕ_s are the body effect coefficient and channel surface potential, respectively. Here, ΔV_{th0} describes the mismatch $\Delta I_{ds}(V_{bs} = 0)$ due to RDD of the channel doping concentration N_{CH} of MOSFETs whereas, $\Delta\gamma$ describes the mismatch in $\Delta I_{ds}(V_{bs})$ due to the variation in N_{CH} in the depletion region under the gate. We know that $\gamma = f(N_{CH})$ (Equation 4.11) and with the change in the value of V_{bs}, the depth of the depletion layer under the gate changes due to nonuniform channel doping profile [1,9,48–51]. As a result, the amount of bulk charge qN_{CH} changes with the change in V_{bs} as shown in Figure 8.8 for the graded retrograde channel doping profile [49]. Thus, RDD of the vertical channel doping profile under the gate contributes to the mismatch in $I_{ds}(V_{bs})$. Hence, $I_{ds}(V_{bs})$ mismatch between the identical paired transistors due to variation in the vertical channel doping concentration must be modeled by γ.

Thus, the set of major local process variability-sensitive device parameters contributing to the mismatch between identically designed paired transistors within a die is $\{V_{th0}, W, L, T_{ox}, \mu_{eff}, \gamma\}$ as shown in Table 8.1. Here, ΔV_{th0} describes the variation in ΔI_{ds} due to RDD; ΔW and ΔL describe ΔI_{ds} due to LER and LWR; ΔT_{ox} defines ΔI_{ds} due to OTV; $\Delta\mu_{eff}$ defines ΔI_{ds} due to mobility variation caused by SR scattering; and γ models $\Delta I_{ds}(V_{bs})$ due to RDD in the vertical channel doping profile. Therefore, we have used the basic I–V relation to determine the major process variability-sensitive device parameters for modeling mismatch in VLSI circuit performance.

8.5.1.2 Selection of Global Process Variability-Sensitive Device Parameters

The global process variability is caused by nonuniform processing temperature as well as by the variation of implant doses across wafers and relative

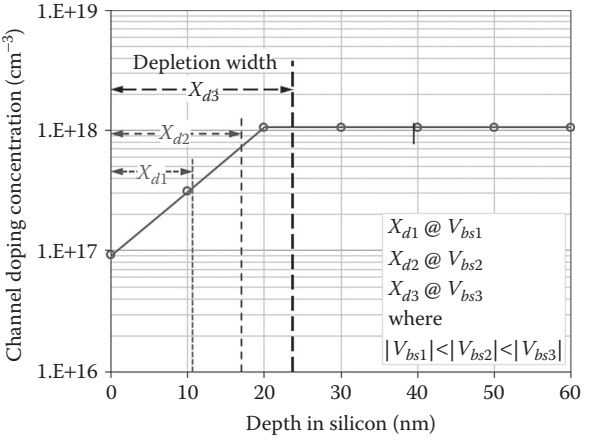

FIGURE 8.8
A piecewise graded-retrograde MOSFET channel doping profile from the silicon/SiO_2 interface at depth $x = 0$ into the substrate; here X_{d1}, X_{d2}, and X_{d3} are the depletion width due to the applied body bias V_{bs1}, V_{bs2}, and V_{bs3}, respectively, causing $V_{th}(V_{bs})$ variability due to RDD along the depth of the channel. (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.)

TABLE 8.1

Process Variability-Sensitive Local Device Parameters Mapped to the Corresponding BSIM4 Compact Model Parameters

Device Parameter		Compact Model Parameter	
Symbol	**Definition**	**Symbol**	**Definition**
V_{th0}	Threshold voltage	VTH0	V_{th} at $V_{bs} = 0$
W	Channel width	XW	W offset due to masking and lithography
L	Channel length	XL	L offset due to masking and lithography
T_{ox}	Gate oxide thickness	TOXE/TOXM	Equivalent T_{ox}
μ_{eff}	Inversion carrier mobility	U0	Low field mobility
γ	Body bias coefficient	K1	1st order body bias coefficient

Source: S.K. Saha, *IEEE Access*, 2, 104–115, 2014.

location of devices [7,8]. The global variation shifts the average or mean value of device performance. As a result, a device parameter within a chip varies for two identically designed devices. For a large count of P from a large number of on-chip measurement data, P converges to a Gaussian distribution with mean value P_0 and standard deviation $\sigma = \Delta P$. Then the chip mean variation in I_{ds} due to global process variability-sensitive parameter P is given by [9]

$$\sigma_{I_{ds}}^2 = \sum_{i=1}^{l}\left(\frac{\partial I_{ds}}{\partial P_i}\right)^2 \sigma_{P_i}^2 + 2\sum_{i=1}^{l}\frac{\partial I_{ds}}{\partial P_i}\frac{\partial I_{ds}}{\partial P_{i+1}}\rho\left(P_i, P_{i+1}\right) \tag{8.17}$$

where l is the total number of occurrence (total count of data) of the device parameter P contributing to global I_{ds} variation; P_i is the ith count of P with standard deviation σP_i from its mean value P_0; and $\rho\left(P_i, P_{i+1}\right)$ is the correlation between the occurrence P_i and P_{i+1}. In order to model the variation of I_{ds} around its mean value, we determine the major global process variability-sensitive parameters P.

Again, from Equation 8.15, the chip mean variation in I_{ds} due to global process variability can be described by the parameter set $\left\{V_{th0}, W, L, C_{ox}, \mu_{eff}, \gamma\right\}$. In addition, the I_{ds} variability due to the variation in the S/D dopant implantation dose and processing temperature across wafers are described by the variation in the S/D series resistance R_{DS} of MOSFET devices. Furthermore, the gate delay, $\tau_{pd} \propto C_{load}$, where C_{load} is the load capacitance of the inverter circuit. Therefore, for an accurate simulation of digital circuits, the across-the-chip variation in MOSFET gate capacitance (C_g) along with the S/D junction capacitance (C_J) must be modeled. Now, the variability in the mean value of C_g is described by the gate overlap capacitance (C_{ov}) whereas that in C_J is described by S/D area as well as S/D sidewall and isolation-edge sidewall capacitances. Thus, the variation in the AC and transient performance of VLSI digital circuits are also described by an additional parameter set $\left\{C_{ov}, C_J\right\}$. Therefore, the set of major MOSFET device parameters sensitive to global process variability can be represented by $\left\{V_{th0}, W, L, T_{ox}, \mu_{eff}, \gamma, R_{DS}, C_{ov}, C_j\right\}$ as shown in Table 8.2.

8.5.2 Mapping Process Variability-Sensitive Device Parameters to Compact Model Parameters

In order to develop compact MOSFET model to analyze the impact of process variability in advanced VLSI circuits, the process variability-sensitive device parameters {P} selected in Section 8.5.1 are mapped to the corresponding compact model parameter {M} of the selected compact model. In this study, we select BSIM4 [45] compact model to describe the methodology of generating compact MOSFET variability model library for VLSI circuit CAD.

8.5.2.1 Mapping Local Process Variability-Sensitive Device Parameters to Compact Model Parameters

In Section 8.5.1.1, we have described an analytical approach to select the randomly variable set of device parameters, $\{V_{th0}, W, L, T_{ox}, \mu_{eff}, \gamma\}$, causing mismatch between identically designed paired transistors. The corresponding set of BSIM4 MOS model parameters, shown in Table 8.1, is

TABLE 8.2

Process Variability-Sensitive Global Device Parameters Mapped to the Corresponding Compact Model Parameters

Device Parameter		Compact Model Parameter	
Symbol	**Definition**	**Symbol**	**Definition**
V_{th}	Threshold voltage	VTH0	V_{th} at $V_{bs} = 0$
W	Channel width	XW	W offset due to masking and lithography
L	Channel length	XL	L offset due to masking and lithography
T_{ox}	Gate oxide thickness	TOXE/TOXM	Equivalent T_{ox}
μ_{eff}	Inversion carrier mobility	U0	Low field mobility
γ	Body bias coefficient	K1	1st order body bias coefficient
R_{DS}	SDE resistance	RDSW	Zero bias R_{DS}
C_{ov}	Gate overlap capacitance	CGSL/CGDL	SDE C_{ov}
		CGSO/CGDO	Non-SDE region C_{ov}
C_J	S/D junction capacitance	CJS/CJD	Area component of C_J
		CJSWS/CJSWD	Isolation-edge sidewall C_J
		CJSWGS/CJSWGD	Gate-edge sidewall C_J

Source: S.K. Saha, *IEEE Access*, 2, 104–115, 2014.

$\{V_{TH0}, XW, XL, T_{ox}, U0, K1\}$; where, XW and XL are the channel width and length offset parameters due to masking and photolithography, respectively, and account for the mismatch due to LER and LWR, whereas $U0$ and $K1$ account for the variation in μ_{eff} and N_{CH} under V_{bs}, respectively. In order to build the compact model, the variance $\sigma_{\Delta M_{mismatch}}$ is computed for each M from a large set of data to account for the mismatch in identical paired transistors.

8.5.2.2 Mapping Global Process Variability-Sensitive Device Parameters to Compact Model Parameters

In Section 8.5.1.2, we have shown an analytical approach to determine the critical set of device parameters, $\{V_{th0}, W, L, T_{ox}, \mu_{eff}, \gamma, R_{DS}, C_{ov}, C_j\}$, impacting MOSFET device performance due to global process variability. The corresponding set of BSIM4 compact model parameters is {VTH0, XW, XL, TOX, U0, K1, RDSW, CGSO, CGDO, CGSL, CGDL, CJS, CJD, CJSWS, CJSWD, CJSWGS, CJSWGD}, where the parameter set {CGSO, CGDO, CGSL, CGDL} defines C_{ov}; {CJS, CJD} defines S/D junction area capacitances and {CJSWS, CJSWD, CJSWGS, CJSWGD} defines S/D *pn*-junction sidewall capacitances as shown in Table 8.2. For each M, the variance σM_{global} is obtained from a large set of ET data and added to M_0 to analyze the impact of chip mean variation in VLSI circuits.

8.5.3 Determination of Variance for Process Variability-Sensitive Compact Model Parameters

The variance σM of the compact model parameter M due to process variability is included to the mean (TT) value M_0 to model the impact of process variability on VLSI circuit performance.

8.5.3.1 Variance of Local Process Variability-Sensitive Compact Model Parameters

For a large number of samples $\Delta M_{mismatch}$ between paired transistors is described by standard normal distribution, $N\left(0, \sigma_{\Delta M_{mismatch}}\right)$, where the variance $\sigma_{\Delta M_{mismatch}}$ is given by: $\sigma_{\Delta M_{mismatch}}\big|_{pair} \cong A_M / \sqrt{WL}$ as described in Section 8.3.1 [43,44], where the parameter A_M is a technology-dependent constant of ΔM and is extracted from ΔM_i versus $\left(1/\sqrt{WL}\right)$ plot for a large number ($i = 1, 2, 3, \ldots l$) of sample ET data [1,43,44]. Thus, for the compact model parameter V_{TH0}, the variance of ΔV_{TH0} between two paired transistors is given by:

$$\sigma_{\Delta V_{TH0}}\big|_{pair} \cong \frac{A_{vt}}{\sqrt{WL}} \tag{8.18}$$

where:

A_{vt} is the area dependent constant of ΔV_{TH0}

Typically, each mismatch parameter $\Delta VTH0$, ΔXW, ΔXL, ΔT_{ox}, $\Delta U0$, and $\Delta K1$ can be represented by an expression similar to Equation 8.16. Again, since ΔM_i is random and independent, the correlation $\rho\left(\Delta M_i, \Delta M_{i+1}\right) = 0$ [43]. Then, for a single device we get

$$\sigma M_{mismatch} = \frac{1}{\sqrt{2}}\sigma_{\Delta M_i} = \frac{1}{\sqrt{2}}\frac{A_M}{\sqrt{WL}} \tag{8.19}$$

In Equation 8.19, $\sigma M_{mismatch}$ represents the variance of ΔM due to within-die stochastic process variability. Thus, the variance of ΔV_{TH0} is given by

$$\sigma V_{TH0,mismatch} = \frac{1}{\sqrt{2}}\sigma_{\Delta V_{TH0}} = \frac{1}{\sqrt{2}}\frac{A_{vt}}{\sqrt{WL}} \tag{8.20}$$

For statistical compact modeling, $\sigma M_{mismatch}$ for each variability-sensitive parameter is added to the corresponding M_0 to compute mismatch between paired transistors. Typically, for each M, A_M is extracted from Pelgrom's plot from a large set of measurement data. For next generation technology development, a large set of data can be obtained by numerical process and device CAD to compute $\sigma M_{mismatch}$ for each variability-sensitive compact model parameters [51–55].

8.5.3.2 Variance of the Global Process Variability-Sensitive Compact Model Parameters

For Monte Carlo (MC) statistical modeling, M_{global} is described by normal distribution $N\left(M_0, \sigma M_{global}\right)$, around its mean (TT) value M_0. The global variance σM_{global} is obtained from the statistical distribution of ET data for each M measured from multiple die, wafers, and lots over a period of time [1,9]. However, for the next-generation technology, the ET data are scarcely available for statistical analysis. In this case, the numerical simulation data can be used for the computation of σM_{global} and generate rev0 compact model for circuit analysis of the target technology [51–55]. Typically, $n\sigma M_{global}$ is used to model global process variability with $3 \le n \le 6$.

8.5.4 Formulation of Compact Model for Process Variability-Aware Circuit Design

As described in Section 8.4.1, the TT model for circuit CAD consists of a set of parameters $\{M_0\}$ that models the device and circuit performance of centerline process of the target technology node. The set $\{M_0\}$ represents the nominal device specifications of the target technology. The local and global components of the variability-sensitive compact model parameter are included in the nominal set $\{M_0\}$ to generate compact variability model library for circuit CAD. The final model library includes the nominal parameters with the components of process variability. Thus, a process variability-sensitive model parameter M including both local and global process variability components is given by

$$M = M_0 + \sigma M_{mismatch} + n\sigma M_{global} \tag{8.21}$$

Equation 8.21 is used to build the compact model of the target technology for process variability-aware circuit analysis. Thus, for the compact model parameter V_{TH}, Equation 8.21 yields

$$V_{TH} = V_{TH0} + \sigma V_{TH0,mismatch} + n\sigma V_{TH0,global} \tag{8.22}$$

Equation 8.22 is used to build statistical corner model for realistic analysis of process variability in scaled MOSFETs. Table 8.3 shows FF and SS corner limit of a set of process variability-sensitive model parameters obtained by analytical approach discussed in Section 8.5.2.2.

For MC statistical compact modeling, the *probability distribution function* (PDF) of the mismatch component of M for HSPICE (see Section 1.2.2.1) [56] circuit CAD is obtained using 1-σ variation between paired transistors

$$PDF\left(\sigma M_{mismatch}\right) = \left(\sigma M_{mismatch}\right) agauss(0,1,1) \tag{8.23}$$

TABLE 8.3

Typical Parameter Limits for Worst-Case BSIM4 Fixed Corner Model Generation

Compact Model Parameter	FF	SS
TOXE	Minimum	Maximum
TOXM		
XL	Minimum	Maximum
XW	Maximum	Minimum
VTH0	Minimum	Maximum
U0	Maximum	Minimum
K1	Minimum	Maximum
RDSW	Minimum	Maximum
CGSL	Maximum	Minimum
CGDL		
CGSO	Maximum	Minimum
CGDO		
CJS	Minimum	Maximum
CJD		
CJSWS	Minimum	Maximum
CJSWD		
CJSWGS	Minimum	Maximum
CJSWGD		

Source: S.K. Saha, *IEEE Access*, 2, 104–115, 2014.
FF: fast NMOS and fast PMOS; SS: slow NMOS and slow PMOS.

Similarly, the PDF for the global component of M is expressed as

$$PDF\left(\sigma_{global}\right) = \left(\sigma_{global}\right) agauss(0,1,3) \tag{8.24}$$

Equations 8.22 through 8.24 are used to formulate the variability-sensitive compact model parameters to develop the final model library for HSPICE circuit CAD. Table 8.4 shows the formulation of variability-sensitive BSIM4 model parameters determined in Section 8.5.2 in the model library. Thus, for the variability-sensitive V_{TH}, we have

$$V_{TH} = V_{TH0} + \frac{1}{\sqrt{2}} \frac{Avt}{\sqrt{WL}} agauss(0,1,1) + \sigma V_{TH0} agauss(0,1,3) \tag{8.25}$$

The above procedure is used to build a BSIM4 MOSFET compact model library for the advanced CMOS technology [5–7]. In order to show the basic functionality of the present modeling approach, all mismatches are lumped into V_{th} mismatch and the correlation between global model parameters is ignored.

TABLE 8.4

Process Variability-Sensitive BSIM4 Model Parameters

Compact Model Parameter	Typical Value (M_0)	Local Component of Model Parameter		Global Component of Model Parameter		Compact Model Parameter for MC Statistical Analysis
		$\sigma M_{mismacth\_z}$	PDF for MC Analysis	σM_{global_z}	PDF for MC Analysis	
TOXE	$toxe0$	$\sigma_{toxm\_z} = \left(1/\sqrt{2}\right)\left(A_{tox}/\sqrt{WL}\right)$	$\sigma_{toxm\_z} agauss(0,1,1)$	σ_{tox_z}	$\sigma_{tox\_z} agauss(0,1,3)$	$toxe0 + \sigma_{toxm\_z} agauss(0,1,1) + \sigma_{tox\_z} agauss(0,1,3)$
TOXM	$toxm0$	$\sigma_{toxm\_z} = \left(1/\sqrt{2}\right)\left(A_{tox}/\sqrt{WL}\right)$	$\sigma_{toxm\_z} agauss(0,1,1)$	σ_{tox_z}	$\sigma_{tox\_z} agauss(0,1,3)$	$toxm0 + \sigma_{toxm\_z} agauss(01,1) + \sigma_{tox\_z} agauss(0,1,3)$
XL	$xl0$	$\sigma_{xlm\_z} = \left(1/\sqrt{2}\right)\left(A_{xl}/\sqrt{WL}\right)$	$\sigma_{xlm\_z} agauss(0,1,1)$	σ_{xl_z}	$\sigma_{xl\_z} agauss(0,1,3)$	$xl0 + \sigma_{xlm\_z} agauss(0,1,1) + \sigma_{xl\_z} agauss(0,1,3)$
XW	$xw0$	$\sigma_{xwm\_z} = \left(1/\sqrt{2}\right)\left(A_{xw}/\sqrt{WL}\right)$	$\sigma_{xwm\_z} agauss(0,1,1)$	σ_{xw_z}	$\sigma_{xw\_z} agauss(0,1,3)$	$xw0 + \sigma_{xwm\_z} agauss(0,1,1) + \sigma_{xw\_z} agauss(0,1,3)$
VTH0	$vth0$	$\sigma_{vthm\_z} = \left(1/\sqrt{2}\right)\left(A_{vt}/\sqrt{WL}\right)$	$\sigma_{vthm\_z} agauss(0,1,1)$	σ_{vth_z}	$\sigma_{vth\_z} agauss(0,1,3)$	$vth0 + \sigma_{vthm\_z} agauss(0,1,1) + \sigma_{vth\_z} agauss(0,1,3)$
U0	$u00$	$\sigma_{u0m\_z} = \left(1/\sqrt{2}\right)\left(A_{u0}/\sqrt{WL}\right)$	$\sigma_{u0m\_z} agauss(0,1,1)$	σ_{u0_z}	$\sigma_{u0\_z} agauss(0,1,3)$	$u00 + \sigma_{u0m\_z} agauss(0,1,1) + \sigma_{u0\_z} agauss(0,1,3)$
K1	$k10$	$\sigma_{k1m\_z} = \left(1/\sqrt{2}\right)\left(A_{k1}/\sqrt{WL}\right)$	$\sigma_{k1m\_z} agauss(0,1,1)$	σ_{k1_z}	$\sigma_{k1\_z} agauss(0,1,3)$	$k10 + \sigma_{k1m\_z} agauss(0,1,1) + \sigma_{k1\_z} agauss(0,1,3)$

(*Continued*)

TABLE 8.4 (*Continued*)

Process Variability-Sensitive BSIM4 Model Parameters

Compact Model Parameter	Typical Value (M_0)	Local Component of Model Parameter		Global Component of Model Parameter		Compact Model Parameter for MC Statistical Analysis
		$\sigma M_{mismacth\_z}$	PDF for MC Analysis	σM_{global_z}	PDF for MC Analysis	
RDSW	*rdsw*0			σ_{rdsw_z}	$\sigma_{rdsw\_z} agauss(0,1,3)$	$rdsw0 + \sigma_{rdsw\_z} agauss(0,1,3)$
CGSL/CGDL	*cgl*0	*	*	σ_{cgl_z}	$\sigma_{cgl\_z} agauss(0,1,3)$	$cgl0 + \sigma_{cgl\_z} agauss(0,1,3)$
CGSO/CGDO	*cgo*0	*	*	σ_{cgo_z}	$\sigma_{cgo\_z} agauss(0,1,3)$	$cgo0 + \sigma_{cgo\_z} agauss(0,1,3)$
CJS/CJD	*cj*0	*	*	σ_{cj_z}	$\sigma_{cj\_z} agauss(0,1,3)$	$cj0 + \sigma_{cj\_z} agauss(0,1,3)$
CJSWS/CJSWD	*cjsw*0	*	*	σ_{cjsw_z}	$\sigma_{cjsw\_z} agauss(0,1,3)$	$cjsw0 + \sigma_{cjsw\_z} agauss(0,1,3)$
CJSWGS/CJSWGS	*cjswg*0	*	*	σ_{cjswg_z}	$\sigma_{cjswg\_z} agauss(0,1,3)$	$cjswg0 + \sigma_{cjswg\_z} agauss(0,1,3)$

Source: Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.

8.5.5 Simulation Results and Discussions

The model library developed in Section 8.5.4 is used for MC statistical analysis of advanced MOSFET devices [5–7]. Since RDD is the dominant contributor in mismatch, Figure 8.9 is obtained using only RDD in mismatch model. Figure 8.9 shows the distribution of IONN and IONP obtained by HSPICE circuit simulation. Here, ION is defined as $|V_{gs}| = |V_{ds}| = 1$ V. The IONN versus IONP distribution in Figure 8.9 clearly shows the impact of local process variability or mismatch, global process variability or chip mean variation, and the local and global process variability combined. In Figure 8.9, the simulation data from statistical corner values of IONN and IONP are also superimposed on the plot for reference. In Figure 8.9, FF and SS corners enclose the MC distribution of ON currents. Thus, in contrast to fixed pessimistic corners, shown in Figure 8.6, the statistical corners offer realistic analysis of process variability similar to MC analysis as shown in Figure 8.9.

Figure 8.9 shows that global variability is dominated mainly by local fluctuations, as observed for advanced bulk technologies [57]. It would indicate that global variability is dominated by local random fluctuations or that most of the systematic process variations are present already within the distance between two mismatch transistors [57].

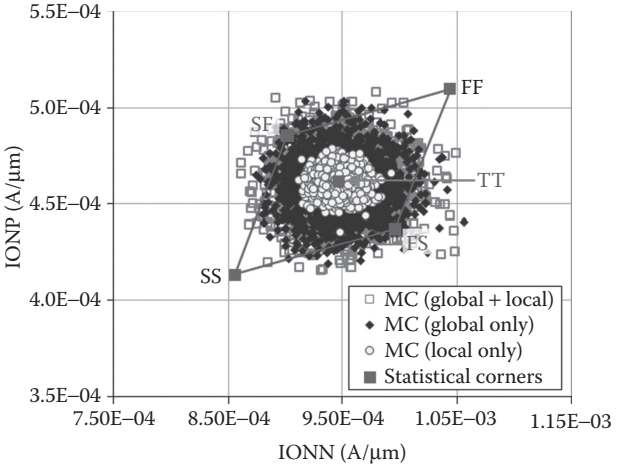

FIGURE 8.9
MC simulation data obtained by HSPICE circuit CAD for an advanced CMOS technology; simulation data show the distribution of ON currents for pMOSFETs (IONP) and nMOSFETs (IONN) for local only, global only, and both local and global process variability. The simulated statistical corners (SS, FF, SF, and FS) along with the nominal (TT) values of drain currents are also superimposed on the plot using solid rectangular symbols. (Data from S.K. Saha, *IEEE Access*, 2, 104–115, 2014.)

8.6 Mitigation of the Risk of Process Variability in VLSI Circuit Performance

Techniques to mitigate the risk of process variability include (1) pure process optimization such as targeting key transistor properties to reduce RDD, improve patterning techniques to reduce LER, and improve polishing techniques to reduce systematic cross-wafer variation; (2) combination of process and design techniques such as optimization of topology, use of OPC to reduce random and systematic variations, and adding dummy features to reduce systematic variations; and (3) pure design techniques such as common-centroid layout to compensate for systematic variation.

As we discussed in Section 8.2.2.1, RDD is a major contributor to random variation and is modeled by Equation 8.2. From Equation 8.2, it is found that we can reduce the impact of RDD by reducing channel doping, N, and gate oxide thickness, T_{ox}. In advanced CMOS technologies, T_{ox} is scaled appropriately using Hi-K dielectric with metal gate to mitigate the risk of process variability due to OTV. However, due to the scaling constraint of N_{CH}, RDD cannot be controlled in nanoscale planar CMOS technology.

Recently, advanced channel engineering has been used to design nanoscale MOSFET devices with undoped or lightly doped channel to mitigate the risk of RDD [48]. The channel is formed on undoped epitaxial layer grown on silicon substrate followed by standard CMOS processing steps [58]. Also, it has been shown that the double-halo MOSFET device architecture [5–8] controls the V_{th} variation in nanoscale devices. Recently, an enhanced double-halo MOSFET [7] device architecture is proposed to design undoped or lightly doped channel MOSFETs and mitigate the risk of process variability in planar CMOS technology [59]. This enhanced double-halo structure is referred to as the buried-halo MOSFET (BH-MOSFET), which is shown in Figure 8.10. The simulation results shown in Figure 8.11 show a significant reduction of threshold voltage variation due to RDD in nanoscale BH-MOSFETs compared to the conventional MOSFET devices.

In order to further mitigate the risk of process variability in nonplanar devices and technologies including Fin field-effect transistors (FinFETs) and ultrathin body (UTB) silicon-on-insulator field-effect transistors referred to as the UTB-SOI MOSFETs [60] have emerged as the most promising alternatives to MOSFET devices and CMOS technology. An overview of the compact models for these devices is presented in Chapter 9.

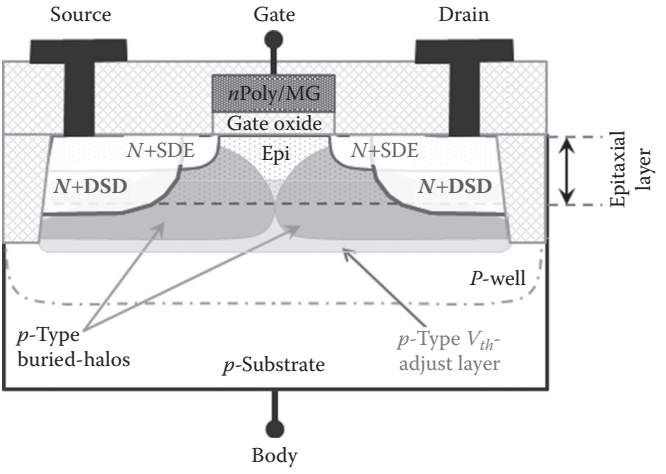

FIGURE 8.10
A variability-tolerant buried-halo MOSFET (BH-MOSFET) device structure; multiple halo implants are buried under the epitaxial layer to obtain undoped or lightly doped channel region.

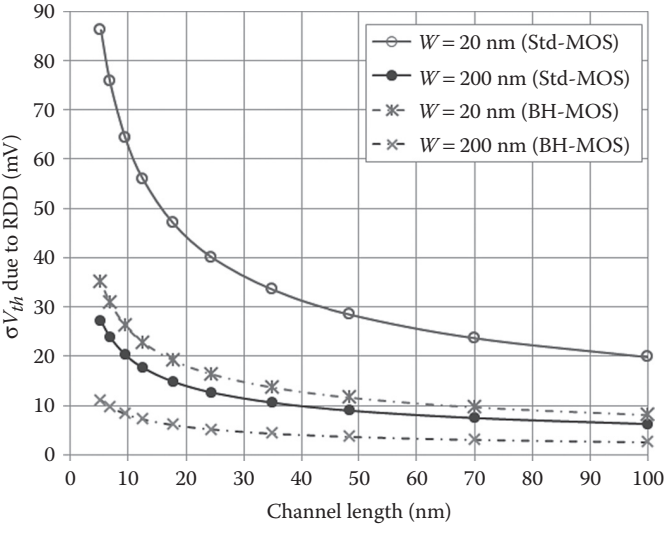

FIGURE 8.11
Comparison of the simulated threshold voltage variation of the conventional (Std-MOS) and BH-MOSFET (BH-MOS) for a typical 20-nm bulk CMOS technology as a function of device channel length for channel width 20 and 200 nm following ITRS (Data from International Technology Roadmap for Semiconductors. http://www.itrs.net/) and using Equation 8.2; parameters used are $N_{CH} = 6 \times 10^{18}$ cm^{-3}; SiO_2 equivalent oxide thickness (EOT) = 1.1 nm; and $C = 0.8165$.

8.7 Summary

This chapter presented the intrinsic process variability in CMOS technology and different approaches to model process variability in VLSI circuit CAD. A brief overview of the systematic and stochastic front-end process variability and sources of process variability is described. A methodology to characterize the random process variability that causes mismatch in the performance of identical MOSFETs in a die is discussed. Conventional approaches to generate compact MOSFET variability models are overviewed and a detailed statistical MOSFET compact modeling approach is discussed. The basic steps to generate statistical compact MOSFET models including selection of expressions defining device performance, selection of device parameters sensitive to process variability, mapping process-variability sensitive device parameters to corresponding compact model parameters, and model formulation are described. The results obtained by MC statistical model and statistical corner model are presented. The basic statistical modeling methodology can be used to generate statistical compact MOSFET models using any compact models considering the basic equations for device performance. Finally, different approaches to mitigate the risk of process variability in VLSI circuits are briefly discussed.

Exercises

8.1 Write an expression for variance $\left(\sigma^2_{\Delta\beta}/\beta^2\right)$ for mismatch in current factor β (a) in terms of the variance of its mutually independent components described in Section 8.5.1.1 and (b) in terms of the mismatch coefficient of each component.

8.2 Consider an nMOSFET device with channel length $L = 100$ nm, channel width $W = 200$ nm, channel doping concentration $N_a = 5 \times 10^{17}$ cm^{-3}, $T_{ox} = 1$ nm, and S/D junction depth $X_j = 50$ nm of the 100 nm CMOS technology node; use $C = 0.8165$ to solve the following problems:

a. Scale down the above technology by 70% up to five times and calculate the total number of dopants (N_{total}) in the channel for all the technology nodes and plot N_{total} versus L. (Scaling: multiply all geometry parameters by 0.7 and divide doping by 0.7.)

b. Considering the device with $W = 200$ nm, calculate and plot $\sigma V_{th,RDD}$ as a function of L calculated in part (a); assume $L = L_{eff}$ and $W = W_{eff}$.

c. Repeat part (b) for $W = 30$ nm to calculate and plot $\sigma V_{th,RDD}$ as a function of L on the same graph (b).

d. Compare your results in parts (b) and (c) and explain.

e. Repeat parts (b) and (c) for C = 0.7071; explain the difference, if any.

8.3 Use the given technology parameters in exercise 8.2 to solve the following problems (consider only RDD):

a. Estimate mismatch coefficient A_{vt} for the nMOSFET devices of the technology using $C = 0.8165$.

b. Use the estimated A_{vt} number from part (a) to calculate $\sigma(\Delta V_{th})$ for a set of devices with varying W and L and plot $\sigma(\Delta V_{th})$ *versus* $1/\sqrt{W \cdot L}$. Explain your plot.

c. Repeat part (a) to calculate σV_{th} for a set of devices with varying W and L and plot σV_{th} versus $1/\sqrt{W \cdot L}$. Explain your results.

d. Compare results from part (b) and part (c) and explain the significance of each plot in compact MOSFET modeling.

First of all, select a wide W (~2 μm) and keeping W constant vary L from the nominal geometry to a long (~250 nm) device and calculate the area *W.L;* then select a long L (=200 nm) and keeping L constant vary W from the nominal geometry to a wide device (~1 μm) and calculate *W.L*.

8.4 If the distance between the identical paired transistors in the x direction is D_x, write an expression for the variance $\sigma^2_{\Delta P}$ of the stochastic parameter P showing the correction factor due to separation between the transistors of the pair.

8.5 Following references [23,24] derive Equation 8.2. Clearly state any assumptions you make.