A numerical integration-based yield estimation method for integrated circuits

Liang Tao(梁涛)† and Jia Xinzhang(贾新章)

Key Laboratory of Ministry of Education for Wide Bandgap Semiconductor Materials and Devices, School of Microelectronics, Xidian University, Xi'an 710071, China

Abstract: A novel integration-based yield estimation method is developed for yield optimization of integrated circuits. This method tries to integrate the joint probability density function on the acceptability region directly. To achieve this goal, the simulated performance data of unknown distribution should be converted to follow a multivariate normal distribution by using Box–Cox transformation (BCT). In order to reduce the estimation variances of the model parameters of the density function, orthogonal array-based modified Latin hypercube sampling (OA-MLHS) is presented to generate samples in the disturbance space during simulations. The principle of variance reduction of model parameters estimation through OA-MLHS together with BCT is also discussed. Two yield estimation examples, a fourth-order OTA-C filter and a three-dimensional (3D) quadratic function are used for comparison of our method with Monte Carlo based methods including Latin hypercube sampling and importance sampling under several combinations of sample sizes and yield values. Extensive simulations show that our method is superior to other methods with respect to accuracy and efficiency under all of the given cases. Therefore, our method is more suitable for parametric yield optimization.

Key words: integrated circuits; yield estimation; Box–Cox transformation; orthogonal array; Latin hypercube

DOI: 10.1088/1674-4926/32/4/045012 **EEACC:** 2570; 0240; 1130B

1. Introduction

Due to the fluctuations of process parameters, mass-produced chips suffer from yield loss and low product quality. Traditionally, worst-case based analysis methods are adopted to ensure conservative design margins under process variability. As manufacturing geometries continue to shrink, those margins are becoming unnecessarily large and with them the risks of over-design and profit loss increase. For these reasons, parametric yield optimization techniques have been proposed to tackle this problem[1−3].

During parametric yield optimization (PYO), a maximization algorithm is utilized to find the circuit designs that improve the yield iteratively. Mathematically, parametric yield, which is the probability of a fabricated circuit meeting the performance specifications under process variability, can be formulated as a multivariate integral. Thus it is significant for the algorithm to evaluate the integral accurately and efficiently. Yield can be evaluated numerically using either the quadrature-based or Monte Carlo (MC) based methods. Both methods are put forward in the light of the fact that the integral cannot be calculated directly.

The quadrature-based method can approximate the integral in a convex acceptability region, and its computational costs explode exponentially with the dimensionality of the disturbance space[1, 3]. The MC method, by contrast, is more widely used free from the restriction on the dimensionality[4]. However, in order to gain a reasonably accurate estimation of the yield, the number of circuit simulations can grow prohibitively high. As a result, variance reduction techniques have long been a concern of studies[5−10].

In this paper, a novel integration-based yield estimation method is developed. Unlike the above two methods, the yield is estimated by integrating the joint probability density function (JPDF) directly through orthogonal array-based modified Latin hypercube sampling (OA-MLHS) and Box–Cox transformation (BCT). In Section 2, the necessary background on parametric yield estimation and the MC method are briefly reviewed, which shows that the MC method is inadequate for PYO for its big misjudgment probability and minimum sample size required, in addition to the well-known low accuracy. Next, two widely-used MC-based variance reduction techniques are discussed in Section 3. The yield estimation procedure of our integration-based method is depicted in detail in Section 4. The principle of variance reduction for model parameter estimation through OA-MLHS is also given. In Section 5, successful results are obtained for our method when compared with different methods under several combinations of sample sizes and yield values. Finally, concluding remarks are made in Section 6.

2. Monte Carlo method and its drawbacks

2.1. Monte Carlo method based parametric yield estimation

Suppose $y = [y_1, y_2, \cdots, y_m]$ is a set of circuit performances and $d = [d_1, d_2, \cdots, d_p]$ is the design parameters involved. $\xi = [\xi_1, \xi_2, \cdots, \xi_n]$ denotes those transistor parameters of the circuit that are subject to process variations and the components of ξ can be considered mutually independent with normal distribution[6, 7]. For a fixed circuit design d, y is given by a deterministic function $y(\xi)$. Since ξ is known to

† Corresponding author. Email: t_liang@yahoo.cn
Received 17 June 2010, revised manuscript received 8 December 2010
© 2011 Chinese Institute of Electronics

obey a multivariate normal distribution with its JPDF $f_\xi = \prod_{j=1}^{n} f_{\xi_j}$, $y(\xi)$ becomes statistical too. Now, the circuit performances can be characterized by a JPDF f_y and the yield can be expressed as

$$Y = \int_{R_y^A} f_y(y)\mathrm{d}y, \quad (1)$$

where R_y^A is the acceptability region defined by the specifications and usually a convex subspace of R^m. Normally, f_y is unknown because $y(\xi)$ usually does not have an explicit expression, like a simulation code. In order to compute the integral in Eq. (1), define

$$I_y(y) = \begin{cases} 0, & y \notin R_y^A, \\ 1, & y \in R_y^A, \end{cases} \quad (2)$$

as an indicator function and the yield can be re-expressed as

$$Y = \int_{R^m} I_y(y) f_y(y)\mathrm{d}y = E_{f_y}(I_y(y)). \quad (3)$$

Use $\hat{Y} = \left(\sum_{i=1}^{N} I_y(y^i)\right)/N$ as an unbiased estimator of the yield. Since f_y is unknown, y^i s cannot be drawn directly. Alternatively, a random sample $\xi^1, \xi^2, \cdots, \xi^N$ is independently drawn in the disturbance space according to f_ξ. Thus, the yield can be redefined due to the sources of variability as

$$Y = \int_{R^n} I_y(y(\xi)) f_\xi(\xi)\mathrm{d}\xi = \int_{R^n} I_\xi(\xi) f_\xi(\xi)\mathrm{d}\xi = E_{f_\xi}(I_\xi(\xi)), \quad (4)$$

where

$$I_y(y(\xi)) = I_\xi(\xi) = \begin{cases} 0, & \xi \notin R_\xi^A, \\ 1, & \xi \in R_\xi^A. \end{cases} \quad (5)$$

Seen from Eq. (5), $I_y(y(\xi))$ defines a mapping of the acceptability region from R_y^A to R_ξ^A. In most cases, R_ξ^A is only known implicitly and unnecessarily convex subspace of R^n. It seems that the Monte Carlo method is the only general way to numerically solve the integral in Eq. (4). The unbiased MC-based estimator of yield is then given by

$$\hat{Y}_{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} I_\xi(\xi^i). \quad (6)$$

Note that $B_\mathrm{I} = \sum_{i=1}^{N} I_\xi(\xi^i)$ is binomially distributed as $B_I \sim B(N, Y)$ and $\hat{Y}_{\mathrm{MC}}$ is an unbiased estimator of the unknown parameter Y. The variance of $\hat{Y}_{\mathrm{MC}}$ can be easily derived as

$$\mathrm{var}(\hat{Y}_{\mathrm{MC}}) = \sigma_{\mathrm{MC}}^2 = \frac{\sigma_{B_\mathrm{I}}^2}{N^2} = \frac{Y(1-Y)}{N}. \quad (7)$$

Since σ_{MC} is inversely proportional to the square root of N, the sample size must be quadrupled to estimate the unknown Y twice accurately.

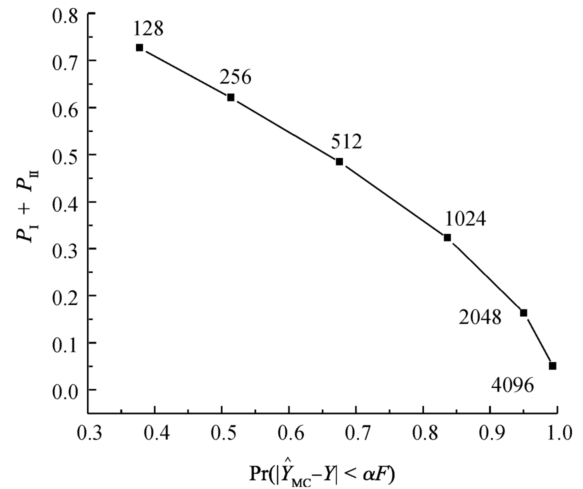


Fig. 1. Relationship between $\Pr(|\hat{Y}_{\mathrm{MC}} - Y| < \alpha F)$ and $P_\mathrm{I} + P_\mathrm{II}$ when Y = 95.45% and α = 0.2.

2.2. The misjudgment probability and minimum sample size requirement

If N is large enough, according to the de Moivre–Laplace theorem, $B(N, Y)$ can be approximated to a normal distribution $N(NY, NY(1-Y))$[11]. Accordingly, we have $\hat{Y}_{\mathrm{MC}} \sim N(Y, \sigma_{\mathrm{MC}}^2)$. Suppose there are three independent candidates of circuit designs d^A, d^B, d^C involved in PYO. The yield values of them are known to be $Y^\mathrm{B} = Y^\mathrm{A} - \alpha F^\mathrm{A}$, and $Y^\mathrm{C} = Y^\mathrm{A} + \alpha F^\mathrm{A}$, $0 < \alpha < 1$, where $F^\mathrm{A} = 1 - Y^\mathrm{A}$ is the reject rate. A wrong step will happen in iterations if $\hat{Y}_{\mathrm{MC}}^\mathrm{A} < \hat{Y}_{\mathrm{MC}}^\mathrm{B}$ or $< \hat{Y}_{\mathrm{MC}}^\mathrm{C} < \hat{Y}_{\mathrm{MC}}^\mathrm{A}$ when the yield is estimated by the MC method. Since $\hat{Y}_{\mathrm{MC}}^\mathrm{A} - \hat{Y}_{\mathrm{MC}}^\mathrm{B} \sim N(\alpha F^\mathrm{A}, \mathrm{var}(\hat{Y}_{\mathrm{MC}}^\mathrm{A}) + \mathrm{var}(\hat{Y}_{\mathrm{MC}}^\mathrm{B}))$, the probability $P_\mathrm{I} = \Pr(\hat{Y}_{\mathrm{MC}}^\mathrm{A} - \hat{Y}_{\mathrm{MC}}^\mathrm{B} < 0)$ can be easily computed, likewise for $P_\mathrm{II} = \Pr(\hat{Y}_{\mathrm{MC}}^\mathrm{C} - \hat{Y}_{\mathrm{MC}}^\mathrm{A} < 0)$; hence, the probability of misjudgment is $P_\mathrm{I} + P_\mathrm{II}$, which can also be viewed as the probability of the iterations going in the wrong direction. In order to lower this probability, $\hat{Y}$ should be made as accurate as possible.

A precision index of a general yield estimator $\hat{Y}$ is defined as $\Pr(|\hat{Y} - Y| < \alpha F)$. Note that the reject rate F is introduced because the relative accuracy of F is worse than that of Y when the circuit yield is high, and αF is more suitable to mimic the step size in PYO. Intuitively, if $\hat{Y}$ is more likely to locate near Y, the estimated yield is unlikely to be confronted with the adjacent values of Y. Figure 1 shows this relationship between $\Pr(|\hat{Y}_{\mathrm{MC}}^\mathrm{A} - Y^\mathrm{A}| < \alpha F^\mathrm{A})$ and $P_\mathrm{I} + P_\mathrm{II}$ under different sample sizes when Y^A = 95.45% and α = 0.2. Clearly, the misjudgment probability of the MC method cannot be neglected until the sample size is big enough. As the number of iterations increases, the step size decreases and the yield values are raised; meanwhile, the misjudgment probabilities also increase. The low accuracy of $\hat{Y}_{\mathrm{MC}}$ due to unaffordable simulation costs may reduce the efficiency and effectiveness of PYO.

Another major limitation of the MC method is the minimum sample size required when estimating high yield. It is obvious that using 100 random samples, MC cannot reliably approximate the yield values as high as 99.73%. The reason is that according to Poisson approximation[11], the probability of at least 1 sample being rejected is only $1 - \mathrm{e}^{-(100\times0.0027)} =$

0.2366. Even if the rejected samples do exist, the highest yield one can get is just 99%. One rule to decide the minimum sample size is that $\lambda = NF$ must be greater than 5, in which case the probability of rejection occurrence is higher than $1 - e^{-\lambda} = 0.9933$. This rule also gives the minimum sample size needed for the adequacy of normal approximation[11]. Therefore, the minimum sample size required is 1852 for $Y = 99.73\%$ and the misjudgment probability in PYO is 0.75 for $\alpha = 0.2$ and 0.42 for $\alpha = 0.5$. Both the sample size and the misjudgment probability are disasters for PYO using the MC method.

3. Variance reduction techniques

Variance reduction techniques are MC-based methods and can be used to improve the accuracy of yield estimation. Two of them, Latin hypercube sampling (LHS)[6, 7] and importance sampling (IS)[5, 8–10] are extensively studied and thus are introduced here for comparison with our method.

3.1. Latin hypercube sampling

LHS is a type of stratified MC sampling widely used in computer experiments. LHS generates an N-sized sample in disturbance space $\xi = [\xi_1, \xi_2, \cdots, \xi_n]$ as follows. The interval [0, 1] is divided into N equal intervals. Then in each interval a sample is drawn randomly with uniform distribution. This procedure is repeated independently for each dimension. Thus for each dimension j, we have $\tilde{x}_j^k \sim U((k-1)/N, k/N)$, $k = 1, \cdots, N$. Next, $[\tilde{x}_j^1, \tilde{x}_j^2, \cdots, \tilde{x}_j^N]^T$ is randomly ordered as $[x_j^1, x_j^2, \cdots, x_j^N]^T$. Let k_j be a random permutation of the integers $[1, 2, \cdots, N]$, then we have $[\tilde{x}_j^{k_j(1)}, \tilde{x}_j^{k_j(2)}, \cdots, \tilde{x}_j^{k_j(N)}]^T = [x_j^1, x_j^2, \cdots, x_j^N]^T$. The samples of each component of ξ, ξ_j^i, $i = 1, \cdots, N$ are obtained by mapping x_j^i, $i = 1, \cdots, N$ to the ξ_j axis by the corresponding univariate normal inverse cumulative distribution function (ICDF) $F_{\xi_j}^{-1}(x_j^i)$. By juxtaposing these n vectors, the LHS sample is achieved. This sampling approach ensures that each component of ξ has all portions of its range represented.

After a random sample $\xi^1, \xi^2, \cdots, \xi^N$ is generated by LHS, the unbiased estimator of the yield is given as

$$\hat{Y}_{\text{LHS}} = \frac{1}{N}\sum_{i=1}^{N} I_\xi(\xi^i) = \frac{1}{N}\sum_{i=1}^{N} I_\xi([\xi_1^i, \xi_2^i, \cdots, \xi_n^i]) = \frac{1}{N}\sum_{i=1}^{N} I_\xi([F_{\xi_1}^{-1}(x_1^i), F_{\xi_2}^{-1}(x_2^i), \cdots, F_{\xi_n}^{-1}(x_n^i)]), \quad (8)$$

where $x_j^i \in [0, 1]$, $i = 1, \cdots, N$, $j = 1, \cdots, n$. Let $x^i = [x_1^i, x_2^i, \cdots, x_n^i] \in [0, 1]^n$, then $\hat{Y}_{\text{LHS}}$ is simplified as

$$\hat{Y}_{\text{LHS}} = \frac{1}{N}\sum_{i=1}^{N} I_x(x^i). \quad (9)$$

Therefore the yield can be viewed as the mean of the indicator function $I_x(x)$, $Y = E(I_x(x))$. Note that if all x_j^i are randomly sampled uniformly on [0,1] like $x_j^i \sim U(0, 1)$, $\left(\sum_{i=1}^{N} I_x(x^i)\right)/N$ is then a MC estimator. Stein pointed out that the variance of $\left(\sum_{i=1}^{N} I_x(x^i)\right)/N$ under LHS is asymptotically smaller than that of the MC method [12].

LHS is widely used for its simple implementation and low computational costs. However, the efficiency provided by LHS over MC is modest especially when yield is high [6, 7]. In addition, LHS also suffers from the minimum sample size requirement as the MC method does.

3.2. Importance sampling

When the IS method is used in yield estimation, the original sampling function f_ξ is replaced by a new one $f_\xi^{\text{IS}} = \prod_{j=1}^{n} f_{\xi_j}^{\text{IS}}$ in order to produce more samples in the importance regions. After a random sample $\xi^1, \xi^2, \cdots, \xi^N$ is generated according to f_ξ^{IS}, the yield can be estimated without bias by

$$\hat{Y}_{\text{IS}} = \frac{\sum_{i=1}^{N} w(\xi^i) I_\xi(\xi^i)}{\sum_{i=1}^{N} w(\xi^i)}, \quad (10)$$

where $w(\xi^i) = f_\xi(\xi^i)/f_\xi^{\text{IS}}(\xi^i)$ is called the weight function. When the yield is high, samples are more likely to fall into the acceptability region R_ξ^{A}. The importance regions can be defined as $\{\xi^i | \xi^i \in R^n, \xi^i \notin R_\xi^{\text{A}}\}$. Properly choosing f_ξ^{IS}, IS can produce a lower variance estimator of the yield by selecting less likely samples in the importance regions. The required minimum sample size also decreases accordingly.

Traditionally, IS is hard to implement for the implicity of R_ξ^{A}[5]. Recently, several techniques have been proposed to use IS in yield estimation[8–10]. In Ref. [10], $f_{\xi_j}^{\text{IS}}$ is fulfilled by enlarging the standard deviation of f_{ξ_j}, which results in more samples being drawn from the extremes of ξ_j. This method requires much less implementation effort than other IS methods and is adopted in comparison with other yield estimation methods in section 5.

IS is efficient only for very high yield because more samples within the importance regions are not adequate for estimating a moderate yield[5]. Thus, IS cannot be used as a general estimation method in PYO since the yield is unknown beforehand. Even when the yield is high, the performance of IS is sensitive to the choice of f_ξ^{IS} and its parameters [5, 8–10].

4. Numerical integration based yield estimation via BCT and OA-MLHS

The low accuracy of MC-based estimators is due to the lost information of the simulated data. If the statistical laws behind the data may be recognized, the yield can be estimated more accurately. Traditionally, $\int_{R_y^{\text{A}}} f_y(y)\text{d}y$ cannot be computed directly because f_y is unknown. Things will change when y can be transformed to obey a multivariate normal distribution. Then the yield can be obtained by integrating f_y on the convex region R_y^{A} as

$$\hat{Y} = \int_{R_y^{\wedge}} f_y(y)\mathrm{d}y = \int_{R_y^{\wedge}} \frac{1}{(2\pi)^{m/2}|\sum_y|^{1/2}} \times \exp\left\{-\frac{1}{2}(y-\mu_y)^T \sum\nolimits_y^{-1}(y-\mu_y)\right\}\mathrm{d}y, \quad (11)$$

where μ_y and $\sum_y$ are the mean vector and covariance matrix of y, respectively, and can be estimated by a random sample $y^1, y^2, \cdots, y^N$ as $\hat{\mu}_y = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y^i$, and $\hat{\sum}_y = \frac{1}{N-1}\sum_{i=1}^{N}(y^i - \bar{y})(y^i - \bar{y})^T$.

Note that every value of the sample participates into the estimation of Y, whereas a lot of information is lost by the two-valued indicator function in the MC-based method. In order to obtain a precise estimator of Y using the integration method, y^1, y^2, Λ, y^N should first be transformed into multivariate normal data and the parameters, μ_y and $\sum_y$, should then be estimated as accurately as possible. These two steps can be fulfilled by BCT and OA-MLHS, respectively. Therefore, our numerical integration-based yield estimation method involves the combination of BCT and OA-MLHS. To the best of our knowledge, it is the first time that this method is introduced into PYO. Next, the procedures needed for yield estimation, including BCT and OA-MLHS, are depicted in Sections 4.1 and 4.3. The reason why the model parameters can be estimated accurately through the combination of BCT and OA-MLHS is explained in Sections 4.2. In Section 4.4, general accuracy measures are defined.

4.1. Box–Cox power transformation

Box–Cox transformation is the most commonly used univariate power transformation family[13] and has long been adopted in process capability evaluation during semiconductor manufacturing[14]. The jth column of $y^1, y^2, \cdots, y^N$ can be transformed by BCT as

$$y_j^{(\lambda_j)} = \mathrm{BCT}(y_j) = \begin{cases} \dfrac{y_j^{\lambda_j} - 1}{\lambda_j}, & \lambda_j \neq 0, \\ \ln y_j, & \lambda_j = 0. \end{cases} \quad (12)$$

This family is only appropriate for positive data. For data that are possibly negative, one can first transform them to positive values by adding a constant and then apply a BCT. This (continuous) family depends on a single parameter λ_j, which can easily be estimated by the method of maximum likelihood. First, a value of λ_j from a pre-assigned range, say [–5, 5], is collected. Next, we evaluate

$$L(\lambda_j) = -\frac{1}{2}\ln(\hat{\sigma}^2(\lambda_j)) + (\lambda_j - 1)\sum_{i=1}^{N}\ln(y_j^i). \quad (13)$$

The estimate of $\hat{\sigma}^2(\lambda_j)$ for a fixed λ_j is $\hat{\sigma}^2(\lambda_j) = S(\lambda_j)/N$, where $S(\lambda_\mathrm{j})$ is the residual sum of squares in the analysis of variance of $y_j^{(\lambda_j)}$. Box and Cox show that the maximum likelihood estimate of λ_j is the value of λ_j that maximizes $L(\lambda_j)$. The above process can be easily implemented by a 1D searching algorithm like the golden section method. With the optimal λ_j^*, the specification limits are also transformed using Eq.(12). This process is repeated for data samples of each performance, and$y^1, y^2, \cdots, y^N$ can be transformed into multivariate normal data.

Generally, the normality test should be used to decide how likely the transformed data is to be normally distributed. The more likely it is, the more accurate the estimate is. Since most performance distributions encountered in IC designs can be thought as moderately deviate from normal distribution, Box-Cox transformation is supposed to work well for yield estimation.

4.2. The principle of variance reduction of model parameter estimation through OA-MLHS together with BCT

As seen in Sections 2 and 3, the yield is estimated by the sample mean of the indicator function. LHS can reduce the variance of the estimator. Similarly, each element of $\hat{\mu}_y$ and $\hat{\sum}_y$ can be expressed as a sample mean of the transformed data. The variance of the estimator for each element of μ_y and $\sum_y$ can also be reduced as long as $y^1, y^2, \cdots, y^N$ is obtained through a proper sampling plan. Consider the first element of $\hat{\mu}_y$, $\bar{y}_1^{(\lambda_1)}$ is the sample mean of the dataset transformed from $[y_1^1, y_1^2, \cdots, y_1^N]^T$ by BCT, then we have

$$\bar{y}_1^{(\lambda_1)} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{BCT}(y_1^i) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{BCT}(y_1(\xi^i)), \quad (14)$$

where $y_1(\xi^i)$ stands for the deterministic function of the first performance defined on ξ and is implicit in essence.

If F_ξ^{-1} denotes the joint inverse cumulative distribution function, $[\xi_1^i, \xi_2^i, \ldots, \xi_n^i] = [F_{\xi_1}^{-1}(x_1^i), F_{\xi_2}^{-1}(x_2^i), \ldots, F_{\xi_n}^{-1}(x_n^i)]$ can be reduced as $\xi^i = F_\xi^{-1}(x^i)$. Thus Equation (14) can be re-expressed as

$$\bar{y}_1^{(\lambda_1)} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{BCT}(y_1(F_\xi^{-1}(x^i))) = \frac{1}{N}\sum_{i=1}^{N} g(x^i), \quad (15)$$

where $x^i = [x_1^i, x_2^i, \ldots, x_n^i] \in [0,1]^n$, $[0,1]^n$ is a unit cube of R^n. The only parameter of $g(x^i)$, λ_1^* can be estimated by the maximum likelihood method. If we define

$$g_j(x_j) = E(g(x)|x_j) - E(g(x)), \quad (16)$$

as the main effect of x_j, and

$$g_{ij}(x_i, x_j) = E(g(x)|x_i, x_j) - g_i(x_i) - g_j(x_j) - E(g(x)), \quad (17)$$

is the second-order interaction between x_i and x_j, Thus $g(x)$ can be formulated as

$$g(x) = E(g(x)) + \sum_j g_j(x_j) + \sum_{i<j} g_{ij}(x_i, x_j) + r(x), \quad (18)$$

where $r(x)$ is the remainder, which can be expressed by the other terms. According to Stein's result[12], the variance of $\left(\sum_{i=1}^{N} g(x^i)\right)\Big/N$ under MC method is

$$\sigma_{\rm MC}^2 = \frac{1}{N}\left(\sum_j \mathrm{var}(g_j(x_j)) + \sum_{i<j} \mathrm{var}(g_{ij}(x_i, x_j)) + \mathrm{var}(r(x))\right), \quad (19)$$

and the variance under the LHS method is

$$\sigma_{\rm LHS}^2 = \frac{1}{N}\left(\sum_{i<j} \mathrm{var}(g_{ij}(x_i, x_j)) + \mathrm{var}(r(x))\right) + o\left(\frac{1}{N}\right). \quad (20)$$

It is clear that by using LHS the main effects are filtered out even without knowing the expression of $g(x)$. The variance can be further reduced by filtering out all of the second-order interactions and the main effects through Orthogonal Array-based Latin Hypercube Sampling (OA-LHS)[15] as

$$\sigma_{\rm OA\text{-}LHS}^2 = \frac{1}{N}\mathrm{var}(r(x)) + o\left(\frac{1}{N}\right). \quad (21)$$

On the other hand, if $\tilde{x}_j^k = (2k-1)/2N$, $k = 1, \cdots, N$ during LHS generation, one gets a modified LHS (MLHS). According to the statistical properties of MLHS discussed in Ref. [7], the variance of $\left(\sum_{i=1}^{N} g(x^{\rm i})\right)/N$ is expected to be smaller than that of standard LHS. Therefore, a new sampling plan called OA-MLHS is presented, originating from OA-LHS. The above reasoning of variance reduction is certainly suitable to any other element of $\hat{\mu}_y$ and $\hat{\sum}_y$. Since all of the model parameters are expected to be more accurate by using OA-MLHS after the simulated performance data are transformed through BCT, the yield estimation is supposed to be more accurate than by using other data sampling plans, such as MC, LHS and OA-LHS.

4.3. Procedures of generating OA-MLHS

The procedures of generating OA-MLHS are basically the same as those of LHS in Section 3.1 except that (1) $\tilde{x}_j^k$ is taken to be $(2k-1)/2N$ other than $\tilde{x}_j^k \sim U((k-1)/N, k/N)$, $k = 1, \cdots, N$. (2) $[\tilde{x}_j^1, \tilde{x}_j^2, \cdots, \tilde{x}_j^N]^{\rm T}$ is not simply ordered as $[x_j^1, x_j^2, \cdots, x_j^N]^{\rm T}$ according to a random permutation k_j but ordered according to the jth column of an orthogonal sequence of permutations[15]. The typical steps for producing this sequence are:

Step 1: Select an orthogonal array OA(N, n_0, s, r) with strength $r = 2$, N rows, n_0 ($n_0 > n$) columns and s levels (levels are often denoted as 1, 2,..., s).

Step 2: Draw n columns from OA(N, n_0, s, r) by simple random sampling without replacement.

Step 3: Let P_i denote a random permutation of [1, 2,..., N/s] plus a constant $N(i-1)/s$ for $i = 1, 2,..., s$. For each column, replace the N/s elements with the same level by the corresponding P_i. For example, if $N = 32$, $s = 4$, then replace eight 1s by P_1, eight 2s by P_2 and so on. Denote each column of the final orthogonal sequence by O_1, O_2,..., O_n.

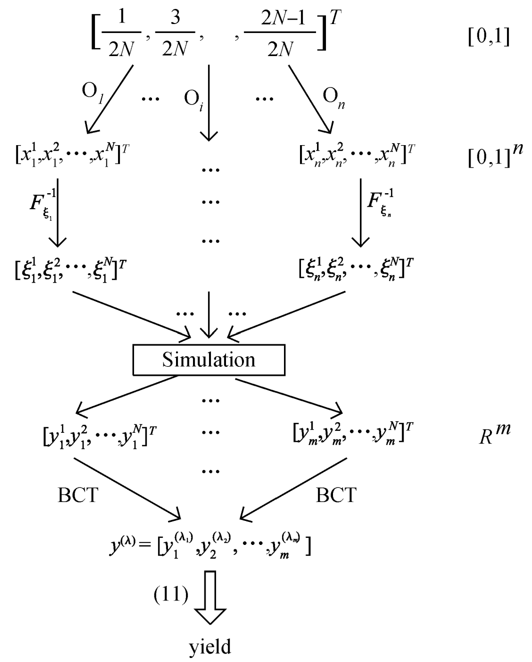


Fig. 2. Integration-based yield estimation via BCT and OA-MLHS.

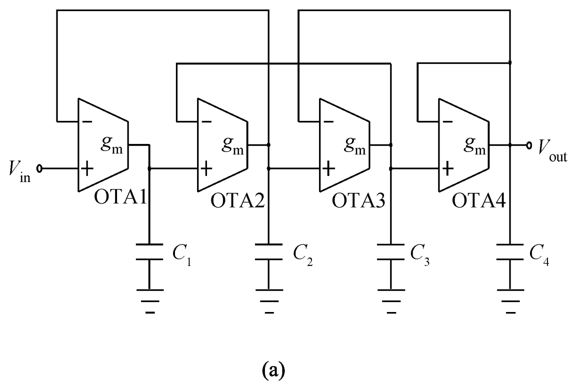
All of the procedures needed for yield estimation by BCT and OA-MLHS are depicted in Fig. 2. The integration in Eq. (11) is calculated efficiently by the sophisticated algorithm of Genz[16]. The majority of extra computation time is occupied by the 1D searching of BCT and the integration; both are very computationally cheap. Thus, the extra computation costs are assumed to increase linearly with the dimension of performance space and very little compared to circuit simulation. In addition, the integration-based methods are free from the restriction on minimum sample size.

4.4. Measures of accuracy for a general estimator of yield

The MC, LHS and IS methods all produce unbiased estimators of the yield. Thus, the accuracy of these estimators can be measured by their variances, $\sigma^2 = E((\hat{Y} - E(\hat{Y}))^2)$. Since the integration-based estimator is not generally unbiased due to the introduction of BCT and MLHS[7], it is incorrect to measure the accuracy of our method still by σ^2. For a general estimator of the yield, with its bias $B = E(\hat{Y}) - Y$, one can define the mean square error (MSE) as

$$\mathrm{MSE} = E((\hat{Y} - Y)^2) = B^2 + \sigma^2. \quad (22)$$

Clearly, for MC-based methods, MSE = σ^2. From the perspective of MSE, an unbiased estimator with large σ^2 is less accurate than an estimator with a small sum of B^2 and σ^2. Another measure of accuracy is $\Pr(|\hat{Y} - Y| < \alpha F)$ given before. $\Pr(|\hat{Y} - Y| < \alpha F)$ defines the accuracy using probabilities and can also be used for measuring the misjudgment probability. In order to obtain these two accuracy measures, the whole process of Fig. 2 is repeated $N_{\rm r}$ times independently. It is eas-



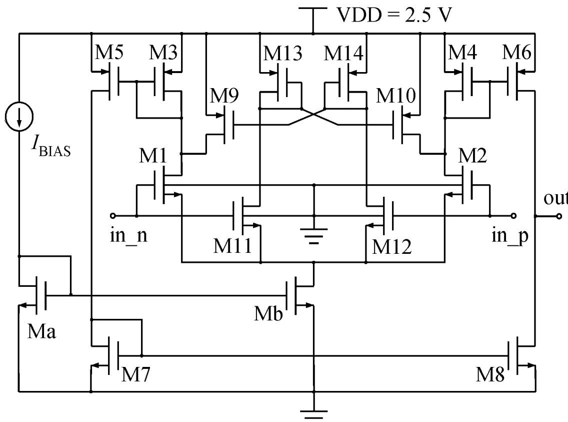


Fig. 3. (a) Schematic of the fourth-order low-pass OTA-C filter. (b) Schematic of the OTA.

Table 1. Circuit parameters and bias conditions.

Transistor	W/L (μm/μm)	Capacitor	C (pF)
M1, M2	51.3/0.5	C_1	13.12
M3, M4	6.2/1	C_2	13.52
M5, M6	6.3/1	C_3	9.28
M7, M8	5.2/1.7	C_4	3.28
M9, M10	0.85/0.5	Bias condition	
M11, M12	42.2/0.5	V_{dd}	2.5 V
M13, M14	2.2/0.5	I_{BIAS}	86 μA

ily found that $\hat{\text{MSE}} = \frac{1}{N_r}\sum_{i=1}^{N_r}(\hat{Y}^i - Y)^2$ and N_α/N_r (N_α is the frequency number for $|\hat{Y}^i - Y| < \alpha F$) are unbiased estimators of MSE and $\Pr(|\hat{Y} - Y| < \alpha F)$, respectively. Thus, $\hat{\text{MSE}}$ and N_α/N_r or N_α can serve as the accuracy measures for the comparison of different yield estimation methods.

5. Experimental results

In this section, the effectiveness of the proposed method is illustrated in two yield estimation examples: a fourth-order OTA-C filter[17] and a quadratic performance function[6, 7]. Six different yield estimation methods are compared through the accuracy measures proposed above. These methods are three MC-based methods (1) MC, (2) LHS, (3) IS and three integration-based methods which are categorized by their respective sampling plans on the disturbance space as (4) MC based integration (MCI), (5) LHS based integration (LHSI) and (6) OA-MLHS based integration (OA-MLHSI).

Table 2. Process noise factors.

ξ_i	$E(\xi_i)$	$3\sigma_{\xi_i}$	Description
ξ_1	12.8	0.7	Oxide thickness (nm)
ξ_2	–14.2	90	NMOS zero-bias threshold voltage shift (mV)
ξ_3	–13.0	100	PMOS zero-bias threshold voltage shift (mV)
ξ_4	60	70	MOSFET length reduction (nm)
ξ_5	55	70	MOSFET width reduction (nm)
ξ_6	1.57 $\times10^{-3}$	1.57 $\times10^{-4}$	Sheet capacitance for filter capacitors (F/m^2)

Table 3. Filter specifications (Y = 99.73%).

y_j	Specification	Description
y_1	8.967–11.277	–3dB frequency f_c range (MHz)
y_2	< 104.414	Stop band (MHz) @ –80 dB attenuation
y_3	< 1.695	Pass band ripple (dB)
y_4	< 11.562	Group delay variation (ns) @ 0.01f_c–0.5f_c

5.1. Fourth-order all-pole canonical OTA-C filter

Schematics of the OTA and the low-pass filter are given in Fig. 3 together with the circuit parameters shown in Table 1. The filter is designed in a 0.5 μm CMOS process. The parameters of six process noise factors are listed in Table 2; they are considered to be the principal independent factors in disturbance space of an MOS circuit[6, 7]. According to our simulation, the statistical properties of the filter using all of the noise factors in foundry variation models are nearly identical to those using only ξ. Several combinations of sample sizes and yield values are used in comparison. The yield requirements are typically set to be 99.73% (corresponds to the so-called three-sigma design), 95.45% (two-sigma), 86.64% and 68.27% (one-sigma) by 800,000 MC runs. For the purpose of studying different yield estimation methods, fixing the specifications and varying the circuit parameters is basically equivalent to fixing the parameters and varying the specifications. However, the implementation effort of the latter is far simpler. That is why it is widely used in Refs. [6, 7, 9, 10]. The performance specifications when the yield is 99.73% are given in Table 3. To compare the accuracy measures of different methods, repeat 200 times independently for each combination of N and Y. Since the MC and MCI methods are only different in how they deal with the performance data, they can share the same simulation data; so do LHS and LHSI. The standard deviation of f_{ξ_j} becomes 1.5 times its original value during IS, which proves to be reasonable to estimate high yield values in our example.

When N is big enough, the skewness test statistic should not be significantly different from zero if the data sample is approximately normally distributed[11]. The average of the skewness test statistics of the simulated cut-off frequency f_c by using the MC and three integration-based methods are compared

Table 4. Average of the skewness test statistics by 200 times repetitions.

Sample size	Average of the skewness test statistics				$OA(N, n_0, s, r)$
	MC	MCI	LHSI	OA-MLHSI	
1024	3.497	0.011	0.008	0.002	(1024, 33, 32, 2)
512	2.464	0.016	0.012	0.004	(512, 33, 16, 2)
256	1.781	0.020	0.017	0.006	(256, 17, 16, 2)
128	1.323	0.026	0.028	0.015	(128, 17, 8, 2)
64	–	–	–	–	(64, 9, 8, 2)
32	–	–	–	–	(32, 9, 4, 2)

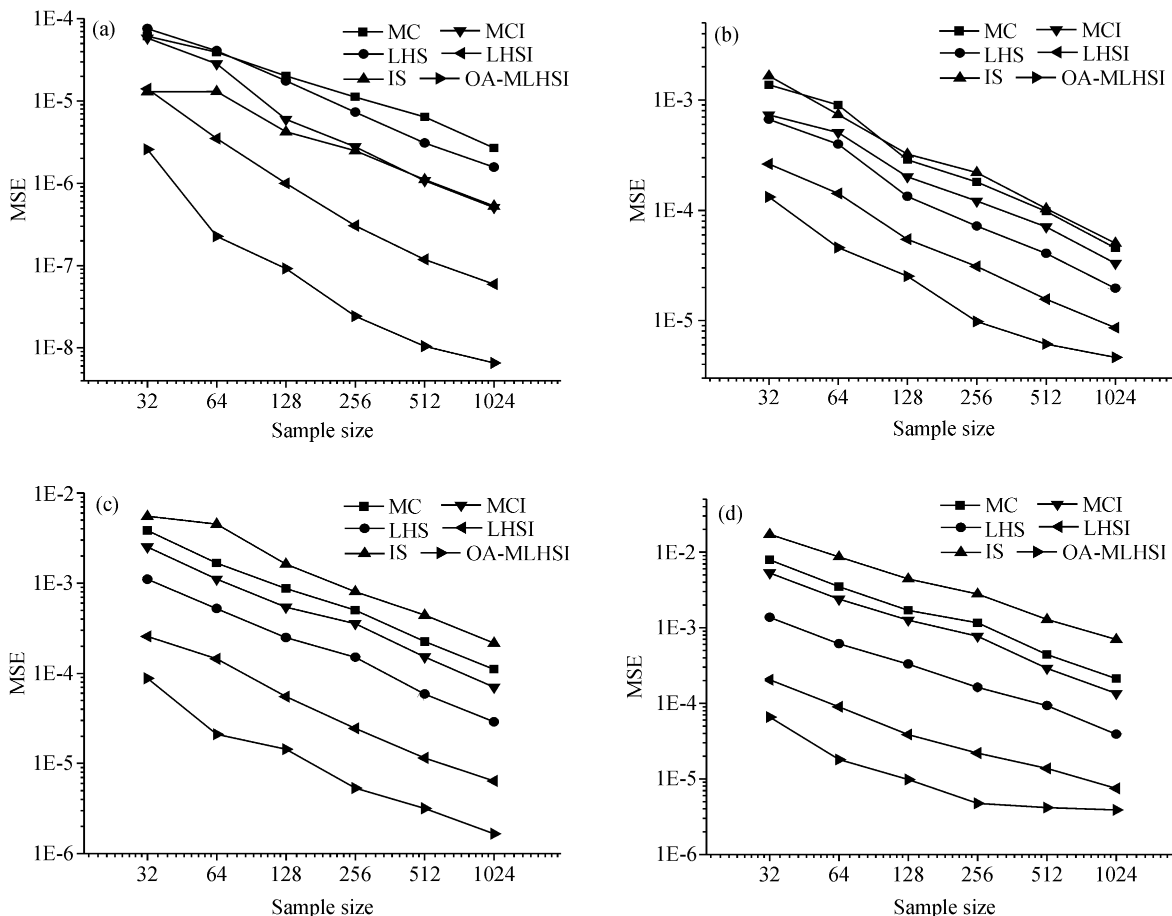


Fig. 4. MSE of different methods under different sample sizes when estimating (a) Y = 99.73%, (b) Y = 95.45%, (c) Y = 86.64% and (d) Y = 68.27%.

in Table 4 after 200 times repetitions (the test statistics when N = 64 and 32 are not given because the skewness test is qualified for N >100). Obviously, the simulated performance data do not obey normal distribution and thus are successfully transformed to normal through BCT and various sampling methods. The orthogonal arrays OA(N, n_0, s, r) chosen for different sample sizes are also given in Table 4.

In Fig. 4, the MSE of each method decreases with the sample size for different yield values. Consequently, the yield estimation accuracy can be controlled via the sample size. In Fig. 5, the frequency number, N_α is compared for different methods under small sample sizes. As seen in Figs. 4 and 5, our method is superior to all of the other methods as for both MSE and $\Pr(|\hat{Y} - Y| < \alpha F)$ under all given combinations of N and Y. For the three MC-based methods, LHS performs best on the whole and is always more accurate than MC. When Y = 99.73%, the MSE of LHS is very close to that of MC, especially for small sample sizes, and IS performs better than the others. However, the accuracy of IS deteriorates rapidly as Y decreases. For the three integration-based methods, MCI performs worst. The accuracy of MCI is close to that of MC and less than that of LHS in many cases. Although LHS and LHSI share the same data, the accuracy measures of them differ a lot since more information contained in the data is exploited by LHSI. However, the superiority of MCI over MC does not seem quite apparent. Therefore, with only the BCT technique, integration-based methods are not at all attractive compared with MC-based methods. That explains why integration-based methods never cause any concern to people.

It is worth noting that N_α of MC and LHS are not given in Fig. 5(a) since $N_\alpha = 0$ for them when Y = 99.73%. In Fig. 6, the histograms of yield estimates for MC and OA-MLHSI when Y

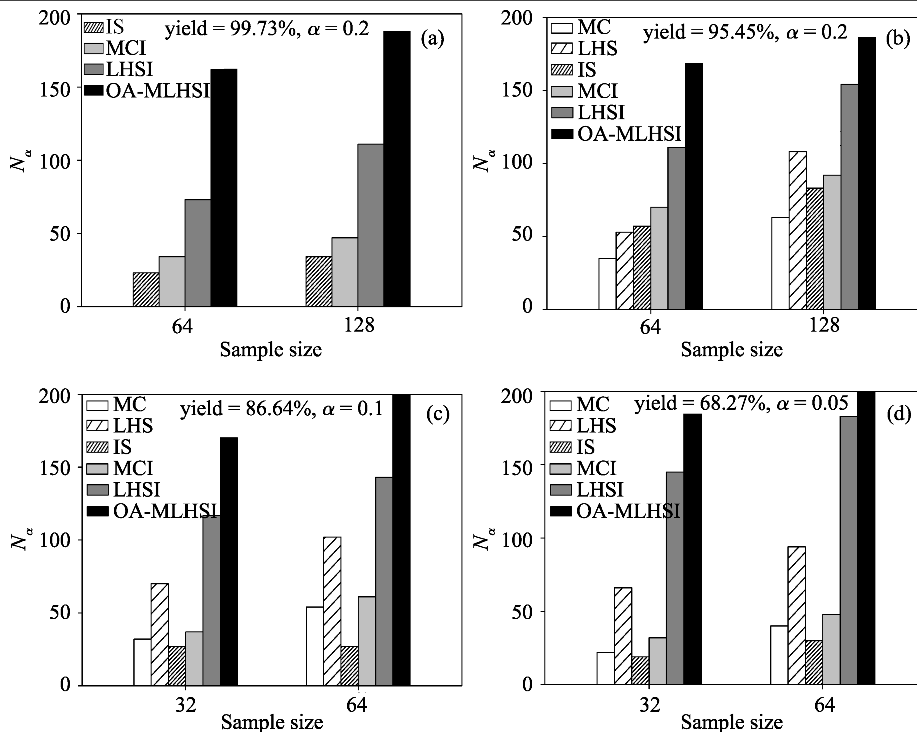


Fig. 5. N_α of different methods under different combinations of sample size, α and yield value.

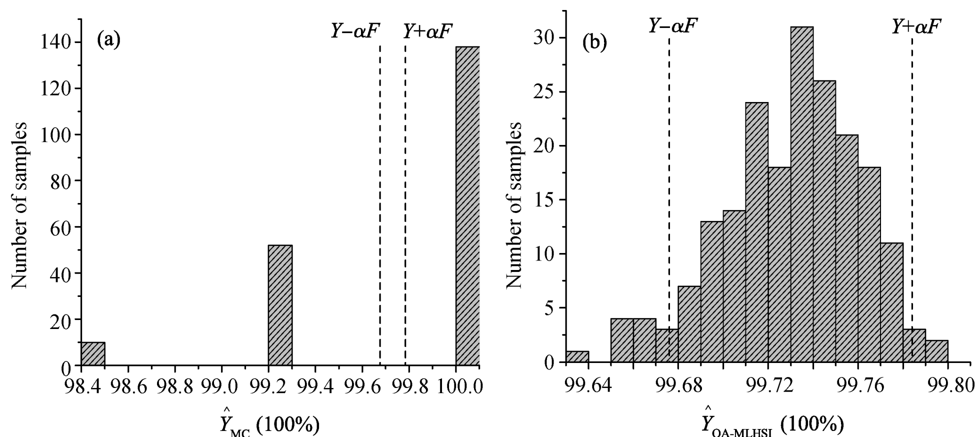


Fig. 6. Histograms of $\hat{Y}$ for (a) MC and (b) OA-MLHSI when Y = 99.73% and N = 128.

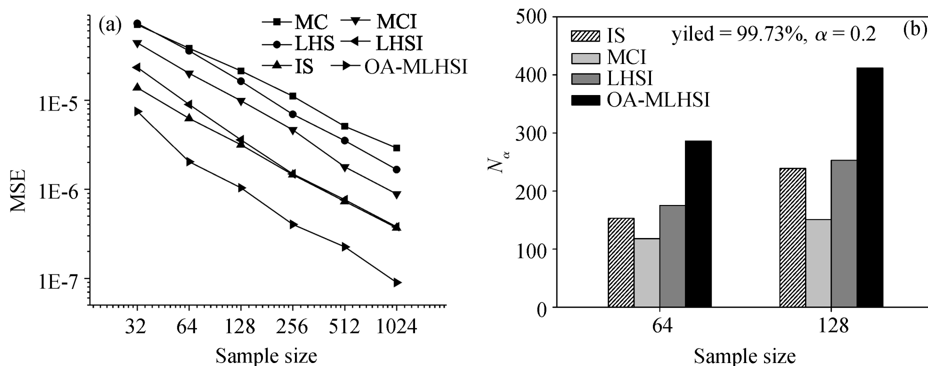


Fig. 7. Accuracy measures of different methods under different sample sizes when estimating Y = 99.73%. (a) MSE. (b) N_α.

= 99.73% and N = 128 are compared. The distribution of yield estimates by using the MC method resembles the shape of a binomial distributed variable, and $|\hat{Y}_{MC} - Y| < \alpha F$ never happens because $\hat{Y}_{MC}$ can only take discrete values. The origin of this phenomenon is that the sample size is far smaller compared with the minimum requirement for the MC method to estimate

$Y = 99.73\%$. Since the performance of LHS is close to that of MC for very high yield values[6, 7], LHS also produces bad results. By contrast, the distribution of yield estimates by using OA-MLHSI is close to a normal distribution via such small samples. The misjudgment probability is expected to be very small because most yield estimates stay around Y. Therefore, the integration-based method using BCT and OA-MLHSI can efficiently produce accurate yield estimation even for very high yield values.

5.2. Quadratic performance function

Let's consider a more general case. Suppose that the behavior of a circuit performance can be expressed as a 3D quadratic function. We adopt the function used in Refs. [6, 7] in this example,

$$y(\xi) = a_0 + \boldsymbol{J}\xi + \frac{1}{2}\xi^T \boldsymbol{H}\xi, \tag{23}$$

where $a_0 = 3$, and the matrices $\boldsymbol{J}$ and $\boldsymbol{H}$ are

$$\boldsymbol{J} = \begin{bmatrix} -1 \\ 10 \\ 2 \end{bmatrix}^T, \quad \boldsymbol{H} = \begin{bmatrix} 1 & -4 & 5 \\ -4 & 3 & 4 \\ 5 & 4 & -2 \end{bmatrix}. \tag{24}$$

Each component of ξ is considered mutually independent and identically distributed normal variable with $E(\xi_i) = 0.5$ and $3\sigma_{\xi i} = 1$, $i = 1, 2, 3$. y is verified to obey three parameters lognormal distribution and the yield value of y is 99.73% when $y \geqslant -0.23$. The MSE and N_α of each method are compared in Fig. 7 after repeating $N_r = 1000$ times. Again, our method is superior to all other methods. Similar results will be obtained for the six methods under different combinations of sample sizes and yield values. It is noteworthy that the IS method performs better in this example than in Section 5.1. In fact, the accuracy of the IS method is very dependent upon the dimensionality of the performance space, the higher the dimensionality the lower the accuracy[5]. People can expect IS to perform well only in the extreme situations where the performances involved are very few and the yield is know to be fairly high.

6. Conclusion

In this paper, an integration-based method for estimating circuit yield is presented. This method tries to integrate the joint probability density function on the acceptability region directly. To do that, the simulated performance data of unknown distribution should be converted to follow a multivariate normal distribution by using Box-Cox transformation. According to our analysis, if the performances are evaluated using the noise dataset sampled on the disturbance space by OA-MLHS during simulations, the estimation variances of model parameters of the multivariate normal JPDF can be reduced greatly. Correspondingly, the accuracy of yield estimation is improved drastically even under small sample sizes. The extra computation costs of our method are linearly proportional to the dimension of performance space. The yield estimations of a fourth-order OTA-C filter and a quadratic performance function are used for comparison of different methods under several combinations of sample sizes and yield values. A vast number of simulations show that our method is superior to other methods with respect to accuracy and efficiency under all of the given cases. Hence, our method is more suitable for parametric yield optimization.

References

[1] Director S W, Feldmann P, Krishna K. Optimization of parametric yield: a tutorial. Proceedings of IEEE Custom Integrated Circuits Conference, 1992: 3.1.1

[2] Kermat M, Kielbasa R. Parametric yield optimization of electronic circuits via improved centers of gravity algorithm. Proceedings of the 40th Midwest Symposium on Circuits and Systems, 1997: 1415

[3] Jing M E, Hao Y, Zhou D, et al. A novel optimization method for parametric yield: uniform design mapping distance algorithm. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2007, 26(6): 1149

[4] Zhang Youtao, Wang Yang, Xia Guanqun, et al. Monte Carlo analysis of yield and performance of a GaAs flash ADC. Chinese Journal of Semiconductors, 2005, 26(8): 1509

[5] Hocevar D E, Lightner M R, Trick T N. A study of variance reduction techniques for estimating circuit yields. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1983, 2(3): 180

[6] Keramat M, Kielbasa R. Latin hypercube sampling Monte Carlo estimation of average quality index for integrated circuits. Analog Integrated Circuits and Signal Processing, 1997, 14(1/2): 131

[7] Keramat M, Kielbasa R. Modified Latin hypercube sampling Monte Carlo (MLHSMC) estimation for average quality index. Analog Integrated Circuits and Signal Processing, 1999, 19(1): 87

[8] Jess J A G, Kalafala K, Naidu S R, et al. Statistical timing for parametric yield prediction of digital integrated circuits. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2006, 25(11): 2376

[9] Kanj R, Joshi R, Nassif S. Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events. Proceedings ACM/IEEE Design Automation Conference, 2006: 69

[10] Doorn T S, Matenter E J W, Croon J A, et al. Importance sampling Monte Carlo simulations for accurate estimation of SRAM yield. Proceedings of the 34th European Solid-State Circuits Conference, 2008: 230

[11] Sheng Zhou, Xie Shiqian, Pan Chengyi. Probability theory and mathematical statistics. Beijing: Higher Education Press, 2003 (in Chinese)

[12] Stein M. Large sample properties of simulations using Latin hypercube sampling. Technometrics, 1987, 29(2): 143

[13] Box G E P, Cox D R. An analysis of transformations. Journal of the Royal Statistical Society B, 1964, 26: 221

[14] Wang Shaoxi, Jia Xinzhang. Model of non-normal process capability indices to semiconductor quality control. Chinese Journal of Semiconductors, 2007, 28(2): 227

[15] Tang B. Orthogonal array-based Latin hypercubes. Journal of the American Statistical Association, 1993, 88(424): 1392

[16] Genz A. Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics, 1992, 1(2): 141

[17] Yichuang Sun, Fidler J K. Structure generation and design of multiple loop feedback OTA-grounded capacitor filters. IEEE Trans Circuits Syst I: Fundamental Theory and Applications, 1997, 44(1): 1