IEEE Trans. Circuits and Systems-II, **vol. 45. no. 5, May 1998.**

A Study of Stratified Sampling in Variance Reduction Techniques for Parametric Yield Estimation

Mansour Keramat, *Student Member, IEEE,* and Richard Kielbasa

Ecole Supérieure d'Electricité (SUPELEC)

Service des Mesures

Plateau de Moulon

F-91192 Gif-sur-Yvette Cédex France

Abstract

The Monte Carlo method exhibits generality and insensitivity to the number of stochastic variables, but is expensive for accurate yield estimation of electronic circuits. In the literature, several variance reduction techniques have been described, e.g., *stratified sampling*. In this contribution the theoretical aspects of the partitioning scheme of the tolerance region in stratified sampling is presented. Furthermore, a theorem about the efficiency of this estimator over the Primitive Monte Carlo (PMC) estimator vs. sample size is given. To the best of our knowledge, this problem was not previously studied in parametric yield estimation. In this method we suppose that the components of parameter disturbance space are independent or can be transformed to an independent basis. The application of this approach to a numerical example (Rosenbrock's curved-valley function) and a circuit example (Sallen-Key low pass filter) are given.

Index Terms: Parametric yield estimation, Monte Carlo method, optimal stratified sampling.

I. INTRODUCTION

Traditionally, Computer-Aided Design (CAD) tools have been used to create the nominal design of an integrated circuit (IC), such that the circuit nominal response meets the desired performance specifications. In reality, however, due to the disturbances of the IC manufacturing process, the actual performances of the mass produced chips are different from those of the

nominal design. Process-related performance variations may lead to low manufacturing yield, and unacceptable product quality. For these reasons, statistical circuit design techniques are required to design the circuit parameters, taking the statistical process variations into account. Moreover, the statistical analysis permits finding out the causes of low yield or unacceptable quality, to change the topology and to redesign the circuit.

The Monte Carlo (MC) method [3], [4] is the most reliable technique in yield estimation of electronic circuits. The method is applicable to any type of circuit without requiring simplifying assumptions of the forms of the probability distribution of parameter values or restrictions on the number of parameters. Nevertheless, it requires a large number of circuit simulations to have a valuable estimation (to have a low variance estimator).

To date, several variance reduction techniques [3] which can be applied to the yield estimation, (in particular, *importance sampling* [6]-[8], *stratified sampling* [6], [9], *control variates* [6], [10]-[12], and *Latin hypercube sampling* [13]-[15]), have been studied. Hocevar *et al.* [6] have shown that the importance sampling method is not generally very useful for variance reduction in the MC yield estimation. In addition, these methods require some information about the acceptability region of the circuit. The generality and usefulness of the control variate or *shadow model* technique, in comparison with alternative methods in MC yield estimation, has since been confirmed by Hocevar *et al.* [6], [12] and Soin and Rankin [10], [11]. But an approximate model of the circuit, which involves some additional simulation costs, is required in this method. Latin hypercube sampling (LHS) [13] is a type of stratified MC sampling, and its efficiency in circuit yield and Average Quality Index (AQI) estimation has been shown in [14], [15], [19].

In this paper, we suppose that there is no particular information about the performance behavior or any approximation of acceptability region. Therefore, among the variance reduction techniques, the stratified sampling and LHS can be chosen for variance reduction technique. In what follows, the application of stratified sampling in MC yield estimation will be discussed.

The plan of paper will be as follows. In Section 2, we describe the MC parametric yield estimation and briefly review the variance reduction techniques in MC method. Section 3 describes the stratified sampling and some aspects of the theoretical bases of the partitioning

scheme of the tolerance region. Asymptotic behavior of stratified sampling efficiency with respect to (w.r.t.) the sample size is discussed in Section 4. In Section 5, several examples illustrate good agreement with theoretical aspects. Finally, concluding remarks are made in Section 6.

II. MONTE CARLO PARAMETRIC YIELD ESTIMATION

In order to estimate yield for analysis and design, one first defines a set of performance or response functions, along with constraints for those functions, for each circuit as

$$g_i(\mathbf{p}) \le 0 \qquad i = 1,2,\ldots,m, \tag{1}$$

where $\mathbf{p}$ is the vector of circuit parameters. Also, the acceptability region A_p is defined as

$$A_p = \left\{ \ \mathbf{p} \ \middle| \ g_i(\mathbf{p}) \le 0 \quad i = 1,2,..,m \ \right\}. \tag{2}$$

The constraints functions are usually only known implicitly via simulations, and thus their evaluation can be very costly.

Mathematically, parametric yield of an integrated circuit is defined as the probability of a circuit meeting the design specifications, i.e.,

$$Y = \int_{R^d} I_p(\mathbf{p}) f_{\mathbf{p}}(\mathbf{p}) d\mathbf{p}, \tag{3}$$

where $f_{\mathbf{p}}(\mathbf{p})$ is the joint probability density function (pdf), and $I_p(\mathbf{p})$ is an indicator function defined as

$$I_p(\mathbf{p}) = \begin{cases} 1 & \mathbf{p} \in A_p \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

In VLSI circuits, the probability distribution of integrated circuit parameters is difficult to model. The parameters $\mathbf{p}$ are correlated in a complicated way, and it is impossible to find an analytic expression for their pdf [1].

In the case of integrated circuits, the circuit parameters can be modelled as functions of their deterministic nominal values, $\mathbf{p^0}$, and a set of process disturbances, ξ, i.e., $\mathbf{p} = \mathbf{p}(\mathbf{p^0}, \xi)$. The

components of ξ can be considered mutually independent [2]. One can formulate the yield in disturbance space as

$$Y = E_{\xi}\left[I_{\xi}(\xi)\right] = \int_{R^n} I_{\xi}(\xi)\, f_{\xi}(\xi) d\xi\,, \tag{5}$$

where $I_{\xi}(\xi)$ and $f_{\xi}(\xi)$ are defined in the same way as in (3).

Yield can be evaluated numerically using either the quadrature-based, or MC based [3] methods. The quadrature-based methods have computational costs that grow exponentially with the dimensionality of the disturbance space. The MC method is the most reliable technique for the statistical analysis of electrical circuits. The unbiased PMC based estimator of yield can be expressed as

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

where ξ^i's are independently drawn random samples from $f_{\xi}(\xi)$, and N is the sample size. The variance of $\hat{Y}_{MC}$ is as follows:

$$\sigma_Y^2 = \frac{\mathrm{Var}(I_{\xi})}{N} = \frac{Y(1-Y)}{N} = \frac{\sigma_B^2}{N}\,, \tag{7}$$

where σ_B^2 is the variance of the binomial random variable $I_{\xi}(\xi)$ [5]. The variance of the estimator is independent of the dimension of the parameter space, but depends on the sample size N.

III. STRATIFIED SAMPLING

Stratified sampling [3], [4], [6] is a variance reduction technique in MC estimation. It consists of dividing the sample space to strata and then estimating the yield in each stratum. It was developed by statisticians for use in sample surveys many years ago, and has since been adapted for use in MC estimation.

A. Sample Generation

The sample space is stratified to r disjoint region by choosing a partitioning scheme. Let these partitions be given by $\{Q_i\}_1^r$ with $Q_i \cap Q_j = \varnothing$ for $i \neq j$, and $\cup_i Q_i = R_T$, where R_T denotes the tolerance region [16] of disturbances, which is usually an orthotope. Then in each partition, N_i samples are generated by the following pdf

$$f_i(\xi) = \begin{cases} f_\xi(\xi)/p_i & \xi \in Q_i \\ 0 & \xi \notin Q_i \end{cases},$$

$$p_i = \int_{Q_i} f_\xi(\xi) d\xi, \tag{8}$$

and the N_i are chosen such that

$$N = \sum_{i=1}^{r} N_i . \tag{9}$$

Then the stratified sampling yield estimator is expressed as follows:

$$\hat{Y}_S = \sum_{i=1}^{r} p_i \frac{N_{si}}{N_i}, \tag{10}$$

where N_{si} is the number of circuits which meets all the specifications in the partition Q_i. This is an unbiased estimator for the parametric yield, and its variance can be expressed as

$$\sigma_s^2 = \sum_{i=1}^{r} p_i^2 \frac{\mathrm{Var}\left(I_\xi \mid \xi \in Q_i\right)}{N_i} = \sum_{i=1}^{r} p_i^2 \frac{Y_i(1-Y_i)}{N_i}, \tag{11}$$

where Y_i is the partial yield related to partition Q_i, and is stated as

$$Y_i = \int_{Q_i} I_\xi(\xi) \frac{f_\xi(\xi)}{p_i} d\xi = \int_{R_T} I_\xi(\xi) f_i(\xi) d\xi . \tag{12}$$

B. Optimal Sample Size of Partitions

Suppose that the number of samples N is given, and one wishes to determine the optimal

value N_i such that the variance σ_s^2 is minimized subject to the constraint (9) and $N_i \geq 0$ for all i. This has been done analytically [6], [17] by solving the Kuhn-Tucker equations and verifying the second-order sufficiency condition. The optimum sample size for a partition is

$$N_i^{opt} = \frac{p_i S_i}{\sum_{i=1}^{r} p_i S_i} N \,, \tag{13}$$

where $S_i^2 = Y_i(1-Y_i)$. This is called Neyman or optimal allocation, and the variance of the estimator is

$$\sigma_N^2 = \frac{1}{N}\left[\sum_{i=1}^{r} p_i S_i\right]^2 . \tag{14}$$

From (13), it is seen that the optimal number of sample size for each partition depends on the partial yield Y_i. Since for a practical problem Y_i is unknown, then this method cannot be in practice used for yield estimation.

Another method for choosing the sample size in partitions is called *"Proportional Allocation Stratified Sampling"* (PASS) where $N_i = p_i N$ for all i. By using (11), the corresponding variance is obtained as follows:

$$\sigma_P^2 = \frac{1}{N}\sum_{i=1}^{r} p_i \operatorname{Var}\left(I_\xi \mid \xi \in Q_i\right) = \frac{1}{N}\sum_{i=1}^{r} p_i Y_i (1-Y_i) \,. \tag{15}$$

Furthermore, it can be shown that

$$\sigma_N^2 \leq \sigma_P^2 \leq \sigma_Y^2 = \frac{Y(1-Y)}{N} , \tag{16}$$

because

$$\sigma_P^2 - \sigma_N^2 = \frac{1}{N}\sum_{i=1}^{r} p_i \left[S_i - \sum_{j=1}^{r} p_j S_j\right]^2 \geq 0 \,. \tag{17}$$

The proof of (17) is shown in [17]. Consider the following property for showing $\sigma_P^2 \leq \sigma_Y^2$.

Property 1: *The variance of the Monte Carlo estimator with PASS is less than or equal to the PMC estimator, and the difference between the variances of corresponding estimators is described as [6]*

$$\Delta_{yp} \equiv \sigma_Y^2 - \sigma_P^2 = \frac{1}{N}\sum_{i=1}^{r} p_i (Y_i - Y)^2 \geq 0 . \tag{18}$$

It should be emphasized that the effectiveness of PASS is due to the differences of the partition yields and the overall yield. The gain of Neyman allocation over PASS is due to differences of the partition standard deviations and the overall average standard deviation. The Neyman allocation is difficult to implement because the partition standard deviations must be known. On the contrary, the PASS method can be easily implemented. In what follows the properties of PASS will be studied.

C. Some Theoretical Aspects of Geometric Shapes of Partitions

The efficiency of the PASS over the PMC estimator can be defined as the ratio of their variances

$$\eta = \frac{\sigma_Y^2}{\sigma_P^2} = \frac{Y(1-Y)}{\sum_{i=1}^{r} p_i Y_i (1-Y_i)} . \tag{19}$$

It is seen that this efficiency is independent of sample size N. But it depends on the following:

1- the partitioning scheme of the tolerance region,
2- the number of partitions.

To the best of our knowledge, the problem of optimal partitioning scheme of the tolerance region R_T in stratified sampling for parametric yield estimation has not been studied. In what follows, in order to obtain some insight of efficiency, the effect of the partitioning scheme in PASS is studied in some special cases.

Problem 1: Consider a two dimensional disturbance space such that the tolerance region R_T is a rectangle (Fig. 1(a)) with uniform pdf. Assume that the boundary of acceptability region A_ξ cuts R_T such that R_T is divided into two rectangles, $R_T \cap A_\xi$ and $R_T \cap A_\xi^c$ (Fig. 1(a)). However,

the position of the cut line L is unknown. The normalized and simplified model is given in Fig. 1(b). Moreover, we consider a class of circuits [20] in which the position of the cut line L is a uniform random variable over R_T. The criterion of partitioning is to minimize the expectation of PASS estimator variance (15) for the defined class of circuits. Now the two following partitioning schemes will be considered.

Scheme a: The tolerance region is partitioned into r vertical strips (Fig. 2(a)). Let a_i denote the width of partition Q_i. The problem of optimal partitioning in this case can be formulated as the following optimization problem

$$\text{(P1)} \begin{cases} \underset{\mathbf{a}}{\text{minimize}} & \left\{ F(\mathbf{a}) = E_l\left[\sum_{i=1}^{r} p_i Y_i (1-Y_i) \right] \right\} \\ \text{subject to} & \sum_{i=1}^{r} a_i = 1 \end{cases}, \tag{20}$$

where $\mathbf{a} = \left[a_1, a_2, \ldots, a_r\right]$, and $E_l[.]$ is the expectation with respect to *l*. By using the "*Total Probability Theorem*" [5], the objective function $F(\mathbf{a})$ can be expressed as

$$F(\mathbf{a}) = \sum_{j=1}^{r} E_l\left[\sum_{i=1}^{r} p_i Y_i (1-Y_i) \;\middle|\; L \in Q_j \right] P_r(L \in Q_j). \tag{21}$$

The following results can be shown easily

$$\begin{aligned} P_r(L \in Q_j) &= a_j \\ p_i &= a_i \end{aligned}, \tag{22}$$

$$E_l\left[p_i Y_i (1-Y_i) \;\middle|\; L \in Q_j \right] = \begin{cases} p_j/6 & \text{for } i = j \\ 0 & i \neq j \end{cases}. \tag{23}$$

By substituting (22) and (23) into (21), one obtains

$$F(\mathbf{a}) = \frac{1}{6} \sum_{i=1}^{r} a_i^2 . \tag{24}$$

Using the *Lagrangian* function, problem (P1) can be transformed to an unconstrained optimization. It can be shown that the solution of this problem is

$$a_i = 1/r \quad i = 1,2,\ldots,r\,. \tag{25}$$

Therefore, the optimal partitioning of *Scheme a* is obtained when the width of all strips are equal. The variance of the estimator corresponding to optimal proportional allocation w.r.t. yield value is shown in Fig. 3. It is seen that in order to decrease the maximum value of σ_P^2 for a given value of N, one can increase the number of partitions r. Here, the maximum of r is obtained when one sample per stratum is chosen, i.e., $r = N$.

Furthermore, the solution of the problem for other pdf (e.g., Gaussian) of disturbance space is very complicated. However, it is felt that the solution should be almost in such a way that the tolerance region is divided into equal probability partitions.

Scheme b: Consider that the tolerance region is partitioned into r horizontal strips (Fig. 2(b)). The problem of optimal partitioning can be formulated in the same way as that of *Scheme a.* From Fig. 2(b), one can see that the partial yield Y_i is equal to the overall parametric yield Y. Therefore we have

$$\sigma_P^2 = \frac{1}{N}\sum_{i=1}^{r} p_i Y_i (1-Y_i) = \frac{Y(1-Y)}{N}\sum_{i=1}^{r} p_i = \frac{Y(1-Y)}{N} = \sigma_Y^2\,. \tag{26}$$

Thus the variance is independent of the position of the cut line L and the width of the strips. Consequently, proportional allocation with this partitioning scheme has no advantages over the PMC method.

From the results of the two schemes, one can conclude that in order to have an efficient estimator, the partitioning scheme should be such that the probability of each stratum belonging to A_ξ or its complement A_ξ^c must be maximum. Now consider the following problem which describes a more general case than *Problem 1*.

Problem 2: A two dimensional disturbance space with a uniform pdf over tolerance region R_T is considered. Suppose that the region $R_T \cap A_\xi$ is a rectangle as shown in Fig. 4(a). In order to define a criterion for optimization, we need to define a class of circuits. Suppose that the positions of cut lines L_1 and L_2 have independent uniform pdf's over R_T. Here, the tolerance region is partitioned by vertical and horizontal lines (Fig. 4(b)) such that the overall number of

partitions is r. The partitioning criterion is to minimize the expectation of the PASS estimator variance (15) for the above class of circuits. Thus, the problem of optimal partitioning can now be formulated as the following optimization problem

$$\text{(P2)} \begin{cases} \underset{\mathbf{a},\mathbf{b},m,k}{\text{minimize}} & \left\{ F(\mathbf{a},\mathbf{b},m,k) = E_l\left[\sum_{i,j=1}^{m,k} p_{ij} Y_{ij}(1-Y_{ij})\right]\right\} \\ \text{subject to} & \sum_{i=1}^{m} a_i = 1 \ ; \ \sum_{j=1}^{k} b_j = 1 \ ; \ mk = r \end{cases}, \tag{27}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_m]$, $\mathbf{b} = [b_1, b_2, \ldots, b_k]$, and $\mathbf{l} = [l_1, l_2]$.

In problem (P2), by using the *"Total Probability Theorem"* [5], the objective function can be expressed as a sum of expectations over the disjoint region as follows:

$$F(\mathbf{a},\mathbf{b},m,k) = \sum_{i'=1}^{m}\sum_{j'=1}^{k} E_{\mathbf{l}}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} Y_{ij}(1-Y_{ij}) \,\middle|\, L_1, L_2 \in Q_{i'j'}\right] P_r(L_1, L_2 \in Q_{i'j'}) . \tag{28}$$

From the assumptions of *Problem 2*, the following results are straightforward

$$\begin{aligned} &P_r(L_1, L_2 \in Q_{i'j'}) = a_{i'} b_{j'} \\ &p_{ij} = a_i b_j \end{aligned}, \tag{29}$$

and

$$\begin{aligned} E_{\mathbf{l}}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} Y_{ij}(1-Y_{ij}) \,\middle|\, L_1, L_2 \in Q_{i'j'}\right] &= p_{i'j'} E_{\mathbf{l}}\left[Y_{i'j'}(1-Y_{i'j'}) \middle| L_1, L_2 \in Q_{i'j'}\right] \\ &+ \sum_{j=1}^{j'-1} p_{i'j} E_{l_1}\left[Y_{i'j}(1-Y_{i'j}) \middle| L_1 \in Q_{i'j'}\right] + \sum_{i=1}^{i'-1} p_{ij'} E_{l_2}\left[Y_{ij'}(1-Y_{ij'}) \middle| L_2 \in Q_{i'j'}\right] \end{aligned} . \tag{30}$$

By evaluating the expectations in (30), we have

$$\begin{aligned} &E_{\mathbf{l}}\left[Y_{i'j'}(1-Y_{i'j'}) \middle| L_1, L_2 \in Q_{i'j'}\right] = \frac{5}{36} \\ &E_{l_1}\left[Y_{i'j}(1-Y_{i'j}) \middle| L_1 \in Q_{i'j'}\right] = E_{l_2}\left[Y_{ij'}(1-Y_{ij'}) \middle| L_2 \in Q_{i'j'}\right] = \frac{1}{6} \end{aligned} . \tag{31}$$

By substituting (29)-(31) into (28), the objective function can be expressed as

$$F(\mathbf{a},\mathbf{b},m,k)=\frac{5}{36}\sum_{i'=1}^{m}\sum_{j'=1}^{k}a_{i'}^{2}b_{j'}^{2}+\frac{1}{6}\left(\sum_{i'=1}^{m}a_{i'}^{2}\right)\left(\sum_{j'=2}^{k}\sum_{j=1}^{j'-1}b_{j'}b_{j}\right)+\frac{1}{6}\left(\sum_{j'=1}^{k}b_{i'}^{2}\right)\left(\sum_{i'=2}^{m}\sum_{i=1}^{i'-1}a_{i'}a_{i}\right). \quad (32)$$

Furthermore

$$\begin{aligned}\sum_{j'=2}^{k}\sum_{j=1}^{j'-1}b_{j'}b_{j}&=\frac{1}{2}\left(\sum_{j'=1}^{k}b_{j'}\right)^{2}-\frac{1}{2}\sum_{j'=1}^{k}b_{j'}^{2}=\frac{1}{2}-\frac{1}{2}\sum_{j'=1}^{k}b_{j'}^{2}\\ \sum_{i'=2}^{m}\sum_{i=1}^{i'-1}a_{i'}a_{i}&=\frac{1}{2}\left(\sum_{i'=1}^{m}a_{i'}\right)^{2}-\frac{1}{2}\sum_{i'=1}^{m}a_{i'}^{2}=\frac{1}{2}-\frac{1}{2}\sum_{i'=1}^{m}a_{i'}^{2}\end{aligned}. \quad (33)$$

By substituting (33) into (32), we obtain

$$F(\mathbf{a},\mathbf{b},m,k)=\frac{1}{12}\left(1-\frac{1}{3}\sum_{j'=1}^{k}b_{j'}^{2}\right)\sum_{i'=1}^{m}a_{i'}^{2}+\frac{1}{12}\sum_{j'=1}^{k}b_{j'}^{2}. \quad (34)$$

In order to solve problem (P2), one can first optimize it w.r.t. $\mathbf{a}$, i.e.,

$$\begin{cases}\underset{\mathbf{a}}{\text{minimize}} & \left\{F(\mathbf{a},\mathbf{b},m,k)=K_1\left(\sum_{i'=1}^{m}a_{i'}^{2}\right)+K_2\right\}\\ \text{subject to} & \sum_{i'=1}^{m}a_{i'}=1\end{cases}, \quad (35)$$

where K_1 and K_2 are determined from (34) and are not functions of $\mathbf{a}$. It can also be shown that they are always positive numbers. This problem can be solved by the *Lagrangian* function, and the solution is

$$a_i=1/m \qquad i=1,2,\ldots,m. \quad (36)$$

Now, we substitute solution (36) into the objective function (34), and optimize it with respect to **b**. The solution is obtained in the same way

$$b_j=1/k \qquad j=1,2,\ldots,k. \quad (37)$$

Using (34), (36), and (37) in problem (P2), the following equivalent problem is obtained

$$\begin{cases} \underset{m,k}{\text{minimize}} & \left\{ F(m,k) = \frac{1}{12}\left(1-\frac{1}{3m}\right)\frac{1}{k} + \frac{1}{12m} \right\} \\ \text{subject to} & mk = r \end{cases}. \tag{38}$$

If $\sqrt{r}$ is an integer number, then it can easily be shown that the solution of this problem is $m = k = \sqrt{r}$. Therefore, the overall solution of *Problem (P2)* is given by

$$a_i = b_i = 1/\sqrt{r} \qquad i = 1,2,\ldots,\sqrt{r}\,. \tag{39}$$

If $\sqrt{r}$ is not an integer, one should choose values of m and k such that $mk = r$, and each as near as possible to $\sqrt{r}$.

Moreover, by substituting the optimal solution (39) into (P2), the objective function is obtained as

$$F(r) = \frac{6\sqrt{r}-1}{36r}. \tag{40}$$

It can easily be verified that $F(r)$ is a decreasing function. Thus one can choose the maximum value of r which is N, i.e., one sample per stratum. In addition, the results of *Problem 2* can be extended to higher dimensions.

In fact, the tolerance region can be considered as a weighted hypervolume by the pdf of disturbances, and the performance specifications create a hypersurface in this hypervolume. The hypersurface divides the hypervolume into acceptability region and failure region. The optimal partitioning scheme can be intuitively stated as follows: one should partition the hypervolume in such a way that the probability that each partition belongs to either the acceptability region or the failure region be maximum. For example, in the case of Gaussian pdf , one should first choose the number of divisions related to each disturbance, then the ranges of each of the *n* components of ξ is partitioned into related n_i intervals of probability size $1/n_i$. The Cartesian product of these intervals partitions the tolerance region into $\prod_{i=1}^{n} n_i$ cells each of probability size $\prod_{i=1}^{n} n_i^{-1}$. This is, to some extent, similar to *"Latin Hypercube Sampling"* (LHS) [13], [19].

IV. ASYMPTOTIC EFFICIENCY OF OPTIMAL PASS W.R.T. SAMPLE SIZE

In this section the behavior of optimal PASS efficiency over PMC w.r.t. sample size N will be discussed theoretically. Throughout this section, it is assumed that one sample per stratum is chosen, i.e., $N = r$ and $p_i = 1/N$ (equal probability strata).

The variance of PASS estimator (15) can also be written as

$$\sigma_P^2 = \frac{1}{N} E\left[\mathrm{Var}\left(I_\xi \middle| \xi \in Q_i\right)\right]. \tag{41}$$

Using (7) and (41) in (19), the efficiency of the PASS estimator can be described as

$$\eta = \frac{\mathrm{Var}(I_\xi)}{E\left[\mathrm{Var}\left(I_\xi \middle| \xi \in Q_i\right)\right]} = \frac{Y(1-Y)}{E\left[\mathrm{Var}\left(I_\xi \middle| \xi \in Q_i\right)\right]}. \tag{42}$$

Now consider the following theorem.

Theorem 1: *It is assumed that the set of the non continuity points of the indicator function $I_\xi(\xi)$ over disturbance space is of the zero Lebesgue measure [5]. Then the efficiency of optimal PASS over PMC approaches infinity when the sample size $N \to \infty$ and $Y \notin \{0,1\}$.*

Proof: Suppose that a hyperbox stratum Q_i can be determined by $\Delta\xi^i = \left[\Delta\xi_1^i, \Delta\xi_2^i, \ldots, \Delta\xi_n^i\right]$ and ξ^i is an arbitrary point in this stratum (Fig. 5). For a sufficiently large value of N, we have for each stratum

$$\frac{1}{N} \approx f_\xi(\xi^i) \prod_{j=1}^{n} \Delta\xi_j^i . \tag{43}$$

Therefore, the denominator of (42) can be approximated as

$$\begin{aligned} E\left[\mathrm{Var}\left(I_\xi \middle| \xi \in Q_i\right)\right] &= \frac{1}{N} \sum_{i=1}^{N} \mathrm{Var}\left(I_\xi \middle| \xi \in Q_i\right) \\ &\approx \sum_{i=1}^{N} \mathrm{Var}\left(I_\xi \middle| \xi \in Q_i\right) f_\xi(\xi^i) \delta\xi^i , \\ &= \sum_{i=1}^{N} v_N(\xi^i) f_\xi(\xi^i) \delta\xi^i \end{aligned} \tag{44}$$

where $\delta\xi^i = \prod \Delta\xi_j^i$ and $v_N(\xi^i)$ stands for $\mathrm{Var}\left(I_\xi \mid \xi \in \Delta\xi^i\right)$. It can be concluded that if $I_\xi(.)$ has no discontinuity in $\Delta\xi^i$ then $v_N(\xi^i) = 0$.

Now, if we take the limit on both sides of (44) w.r.t. sample size *N*, the right hand side can be written in integral form as follows:

$$\lim_{N\to\infty} E\left[\mathrm{Var}\left(I_\xi \mid \xi \in Q_i\right)\right] = \int_{R_T} v_\infty(\xi) f_\xi(\xi) d\xi, \tag{45}$$

where $v_\infty(\xi)$ is non zero at the discontinuity points of the indicator function $I_\xi(.)$. It can be demonstrated that the maximum value of $v_\infty(\xi)$ at discontinuity points equals 1/4. Consequently, from the assumption of the theorem about zero Lebesgue measure over discontinuity of $I_\xi(.)$, one can conclude that the right hand side of (45) equals zero. The result follows immediately.☑

Two brief comments can be made about *Theorem 1*. First, it should be emphasized that the assumption of this theorem holds for all realistic problems and is not therefore a restrictive condition. Secondly, it can be shown that for other partitioning schemes where the dimensions of strata do not approach zero, the efficiency of PASS generally converges to a bounded value (e.g., in [6] the strata are chosen to be hyperellipsoid annuli with identical probability).

Furthermore, the result of *Theorem 1* can be extended to AQI estimation of integrated circuits which is a more general definition than parametric yield. Moreover, this theorem confirms some aspects of the optimality of the partitioning scheme that is obtained from the extension of the results of studying some special cases (*Problem 1* and *2*) in the previous section.

V. NUMERICAL AND CIRCUIT EXAMPLES

In this section, we will first present the computational results of a numerical example to show the advantage of the optimal partitioning scheme of PASS over other partitioning procedures. Then a circuit example is given.

Example 1: *Analytic Boundary of Acceptability Region*

The acceptability region A_ξ is defined as

$$A_\xi = \left\{\xi \in R^2 \middle| 100(\xi_2 - \xi_1^2)^2 + (1-\xi_1)^2 \le 8\right\} \tag{46}$$

(Rosenbrock's curved-valley function [18]). The acceptability region and the tolerance region R_T are shown in Fig. 6. It is assumed that the joint pdf of variables are Gaussian with $\bar{\xi} = [-0.5, 0]$ and $\sigma_{\xi_1} = \sigma_{\xi_2} = 0.07$. Figs. 7(a) and 7(b) illustrate the standard deviations of the optimal PASS and PMC yield estimators and the efficiency of PASS over PMC method, respectively. In order to study the behavior of efficiency gain vs. different values of yield, the boundary of acceptability region is moved. The results of efficiency gain are shown in Figs. 7(c) and 7(d). The curve of Fig. 7(b) confirms the result of *Theorem 1*. From the experimental results, it is found that the rate of convergence of efficiency to infinity depends on the dimension of disturbance space. Greater dimension results in slower convergence.

Example 2: *Sallen and Key Lowpass Filter* [6]

One practical example considered is the second-order lowpass Sallen and Key active filter shown in Fig. 8. The nominal frequency response of the filter is illustrated in Fig. 9. The specifications of this filter are given in Table I. The parameters are chosen to be independently distributed Gaussian random variables with standard deviations equal to one-third of 10 percent of the nominal values for the resistors and 5 percent for the capacitors. Figs. 10 and 11 shows the histograms of the optimal PASS and PMC estimators for two different values of sample size. The efficiency and standard deviation of the estimators are illustrated in Fig. 12. It is seen that the efficiency gain increases w.r.t. sample size. In [6], the yield estimation of this circuit by stratified sampling has been studied. There, the strata were chosen to be hyperellipsoid annuli with equal probability. The efficiency gain was 1.2 versus PMC. It can be shown that the efficiency gain by ellipsoidal partitioning converges to a bounded value (in this case ≈ 1.2) when the number of strata approaches infinity. On the contrary, in Fig. 11, it is seen that the efficiency gain η is equal to 4.

VI. CONCLUSIONS

In this paper, the partitioning scheme of *Proportional Allocation Stratified Sampling* (PASS), which is a variance reduction technique in Monte Carlo yield estimation of discrete and integrated circuits was presented. For accurate yield estimation by Primitive Monte Carlo [3], [4] a large number of simulations is required. Several variance reduction techniques have been

developed for reducing the computational cost [3], [6]-[12]. The majority of these methods generally require some knowledge about the circuit responses, which in turn requires further computational procedures. In addition, the efficiency of these methods depend on the quantity of the knowledge. The PASS approach can be realized without any assumptions on circuit responses and *a priori* knowledge of circuit behavior. Here, some optimal aspects of the partitioning scheme of the tolerance region in PASS, which provide some theoretical and practical insights, were discussed. This scheme permits obtaining potentially higher efficiency w.r.t. the PMC method, i.e., to have a more accurate yield estimation than that of PMC.

References

[1] S. W. Director, P. Feldmann, and K. Krishna, "Optimization of parametric yield: a tutorial," in *Proc. IEEE Custom Integrated Circuits Conference*, 1992, pp. 3.1.1-3.1.8.

[2] P. Cox, P. Yang, S. S. Mahant-Shetti, and P. Chatterjee, "Statistical modeling for efficient parametric yield estimation of MOS VLSI circuits," *IEEE Trans. Electron Devices*, vol. ED-32, no. 2, pp. 471-478, February 1985.

[3] R. Y. Rubinstein, *Simulation and the Monte Carlo Method.* John Wiley & Sons, Inc., 1981.

[4] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods.* Norwich, England, Fletcher, 1975.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Process, 3rd edition*. McGraw-Hill, Inc., 1991.

[6] D. E. Hocevar, M. R. Lightner, and T. N. Trick, "A study of variance reduction techniques for estimating circuit yields," *IEEE Trans. Computer-Aided Design*, vol. CAD-2, no. 3, pp. 180-192, July 1983.

[7] K. Singhal and J. F. Pinel, "Statistical design centering and tolerancing using parametric sampling," *IEEE Trans. Circuits Syst.*, vol. CAS-28, no. 7, pp. 692-702, July 1981.

[8] M. A. Styblinski, "Efficiency of yield estimation by the Monte Carlo methods," in *Int. Conf. Computer-Aided Design Manufacture Electronic Components Circuits Syst.*, Sussex, England, July 1979, pp. 150-153.

[9] M. L. Stein, "An efficient method of sampling for statistical circuit design," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, no. 1, pp. 23-29, January 1986.

[10] P. J. Rankin and R. S. Soin, "Efficient Monte Carlo yield prediction using control variates," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1981, pp.143-148.

[11] R. S. Soin and P. J. Rankin, "Efficient tolerance analysis using control variates," *IEE Proc.*, vol. 132, Pt. G, no. 4, pp. 131-142, August 1985.

[12] D. E. Hocevar, M. R. Lightner, and T. N. Trick, "Monte Carlo based yield maximization with a quadratic model," in *Proc. IEEE Int. Symp. Circuits and Systems*, New Port Beach, CA, 1983, pp. 550-553.

[13] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in analysis of output from a computer code," Technometrics, vol. 21, no. 2, pp. 239-245, May 1979.

[14] M. Keramat and R. Kielbasa, "Latin hypercube sampling Monte Carlo estimation of average quality index for integrated circuits," *Analog Integrated Circuits and Signal Processing*, vol. 14, no. 1/2, pp. 131-142, 1997.

[15] M. Keramat and R. Kielbasa, "Worst case efficiency of Latin hypercube sampling Monte Carlo (LHSMC) yield estimator of electrical circuits," in *Proc. IEEE Int. Symp. Circuits Syst.*, Hong Kong, June 1997, pp. 1660-1663.

[16] R. Spence and R. S. Soin, *Tolerance Design of Electronic Circuits.* Addison-Wesley Publishers Ltd., 1988.

[17] W. G. Cochran, *Sampling Techniques, 3rd ed.* New York: John Wiley, 1977.

[18] M. A. Styblinski and A. Ruszczynski, "Stochastic approximation approach to statistical circuit design," *Electronics Letters*, vol. 19, no. 8, pp. 300-302, April 1983.

[19] M. Keramat and R. Kielbasa, "Efficient average quality index estimation of integrated circuits by modified Latin hypercube sampling Monte Carlo," in *Proc. IEEE Int. Symp. Circuits Syst.*, Hong Kong, June 1997, pp. 1648-1651.

[20] M. Keramat and R. Kielbasa, "Optimality aspects of centers of gravity algorithm for statistical circuit design," in *Proc. IEEE 40th Midwest Symposium on Circuits and Systems*, Sacramento, CA, August 3-6 1997.

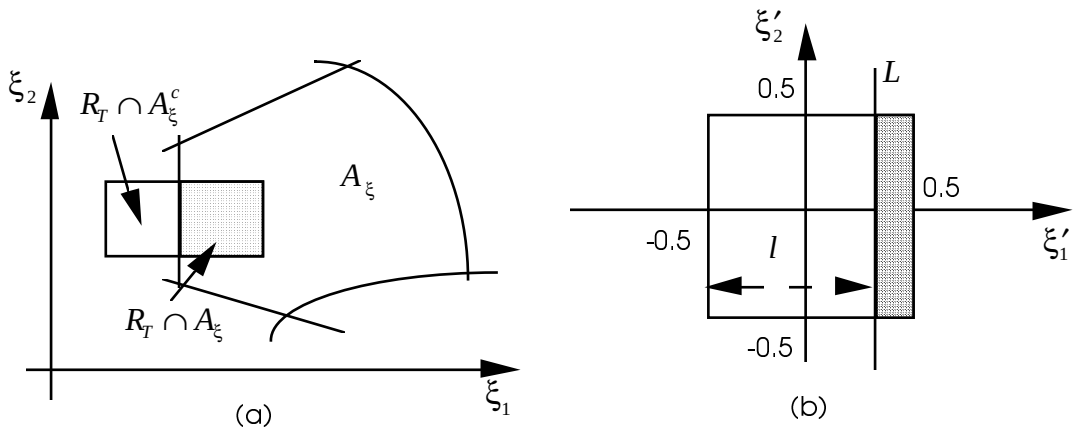

Fig. 1. Description of *Problem 1*.

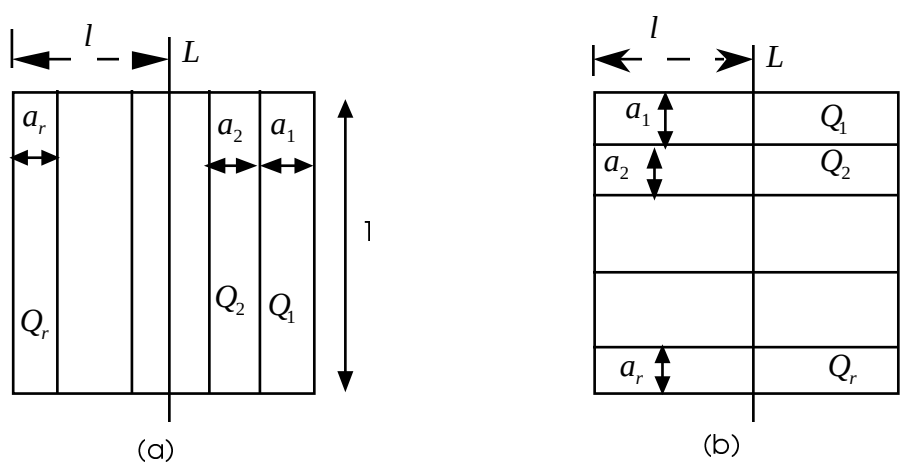

Fig. 2. Partitioning scheme of the tolerance region. (a) Vertical partitioning. (b) Horizontal partitioning.

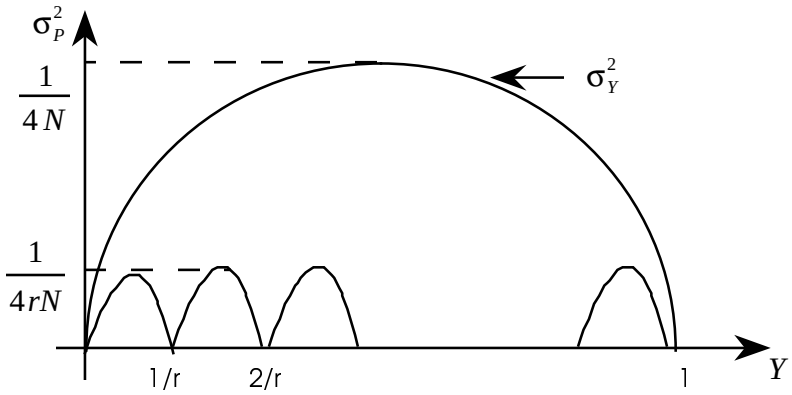

Fig. 3. Variance of estimator for proportional allocation.

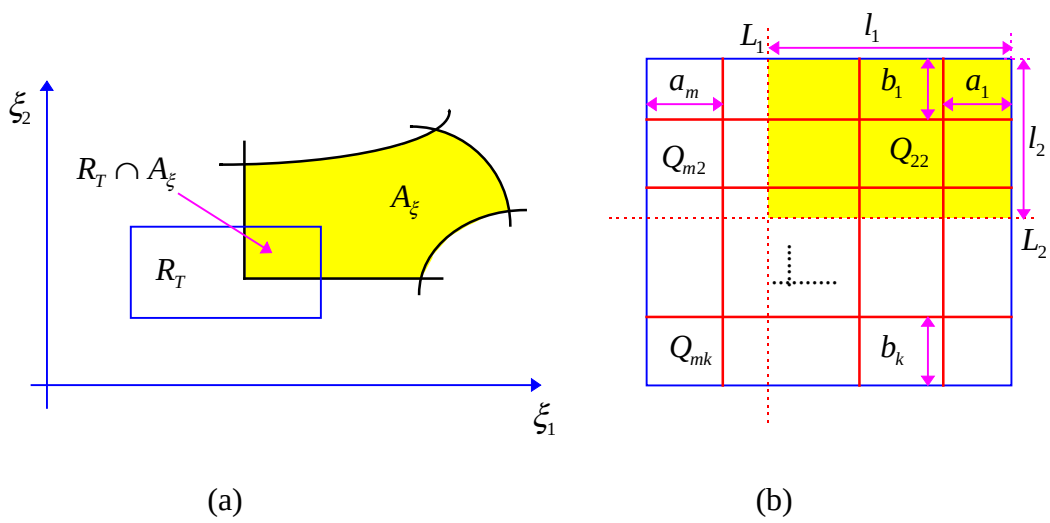

Fig. 4. Description of partitioning for *Problem 2*.

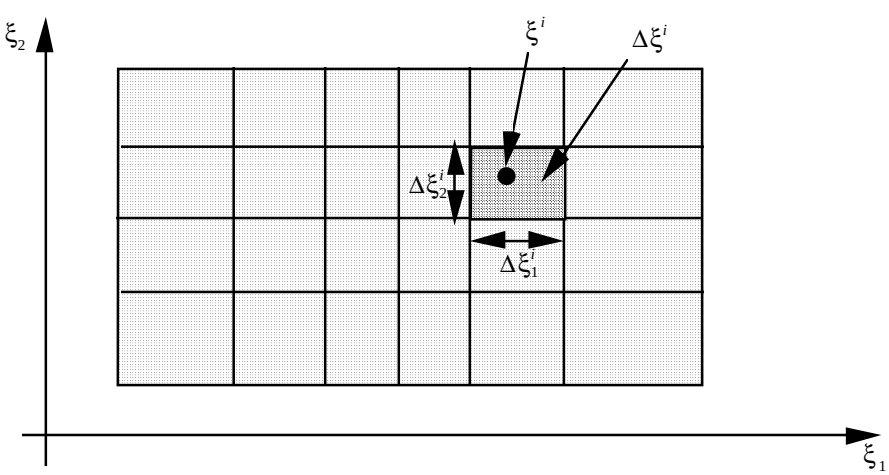

Fig. 5. A 2-dimensional strata scheme of optimal PASS.

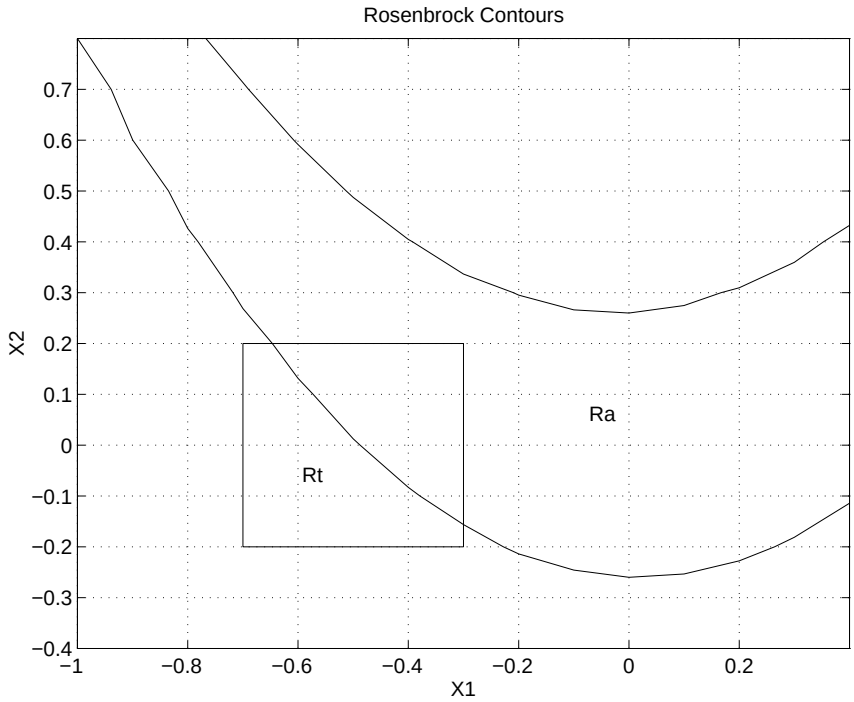

Fig. 6. Rosenbrock's curved-valley acceptability region.

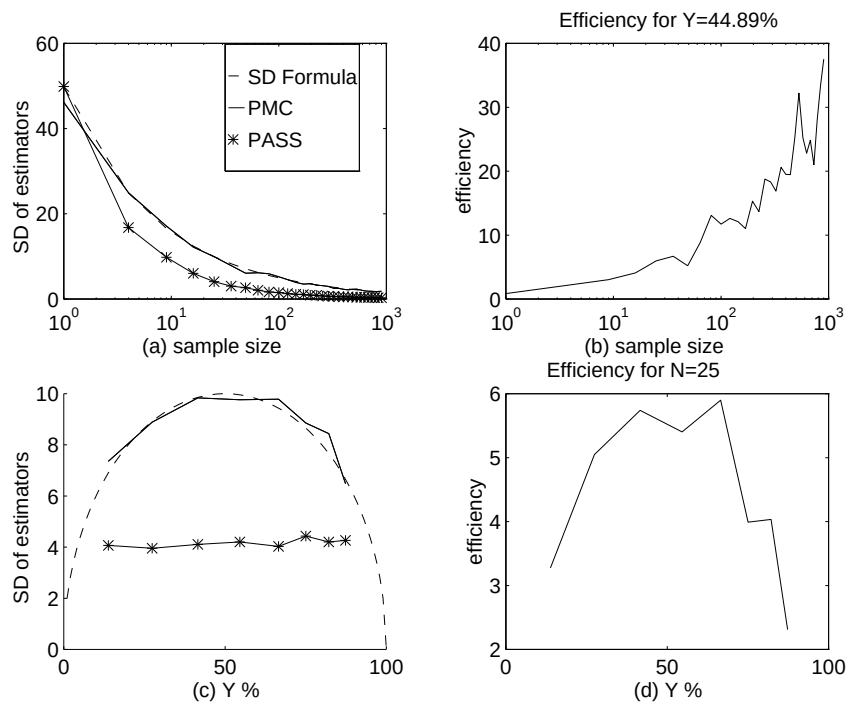

Fig. 7. Efficiency of optimal PASS over PMC estimator for Example 1.

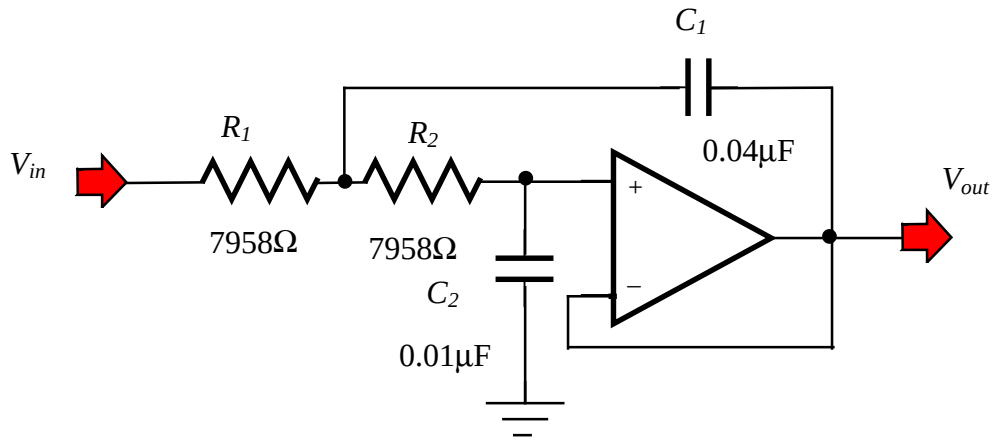

Fig. 8. Schematic circuit of Sallen and Key filter.

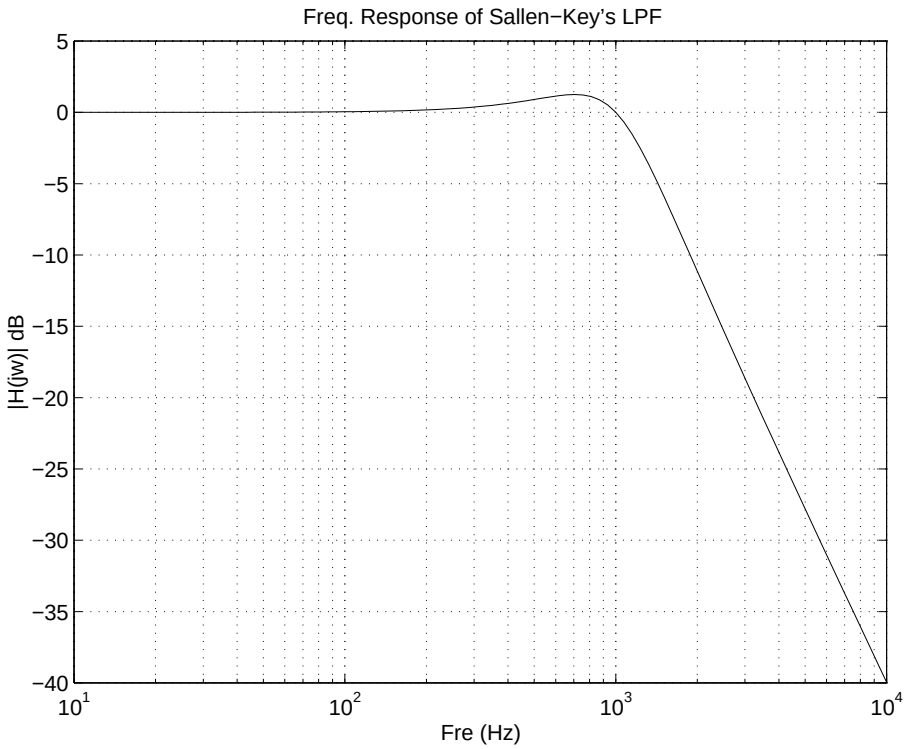

Fig. 9. Frequency response of Sallen and Key filter.

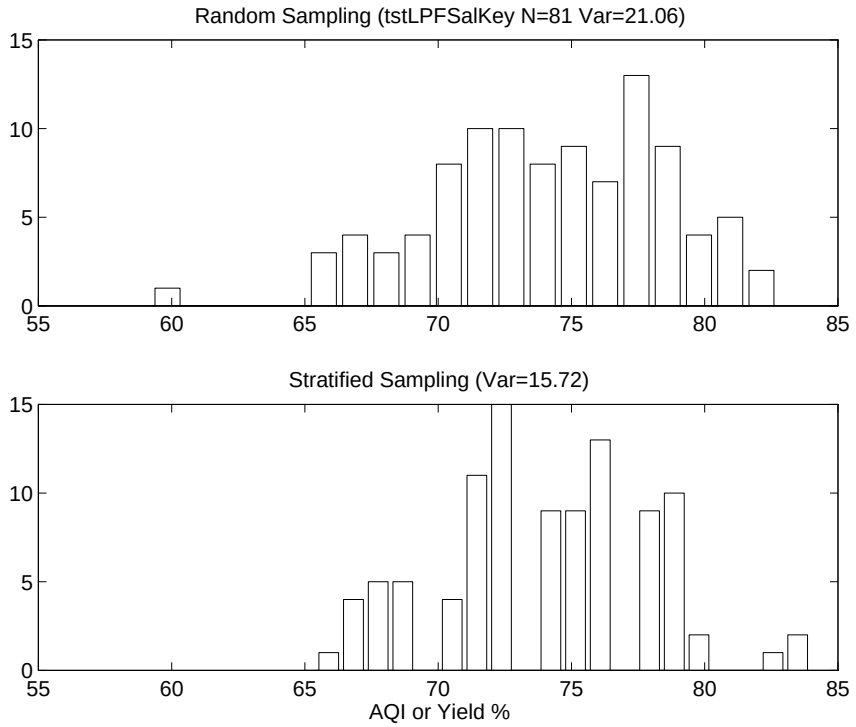

Fig. 10. Histograms of estimators for Sallen and Key filter, *N*=81.

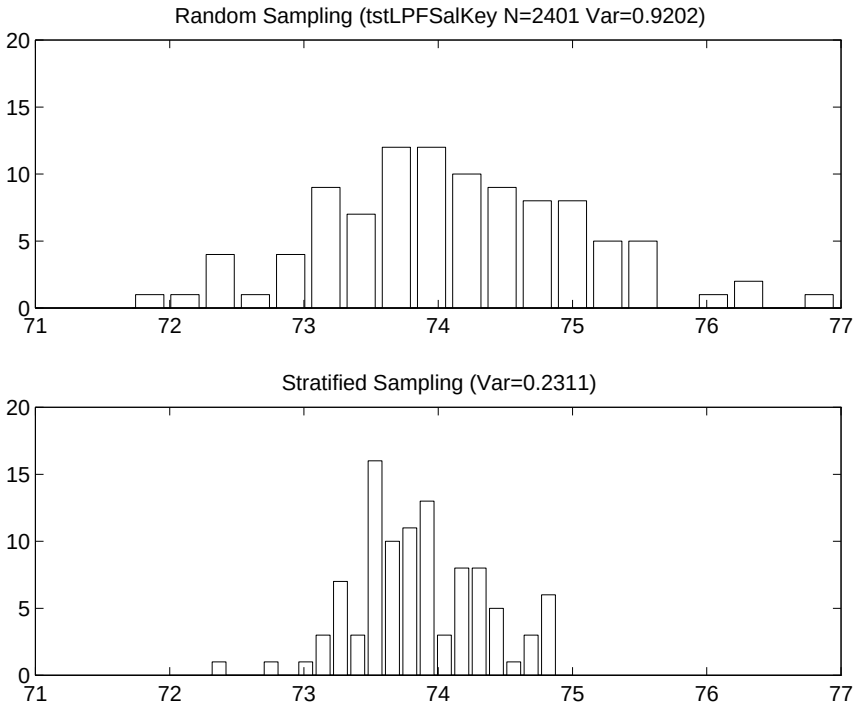

Fig. 11. Histograms of estimators for Sallen and Key filter, *N*=2401.

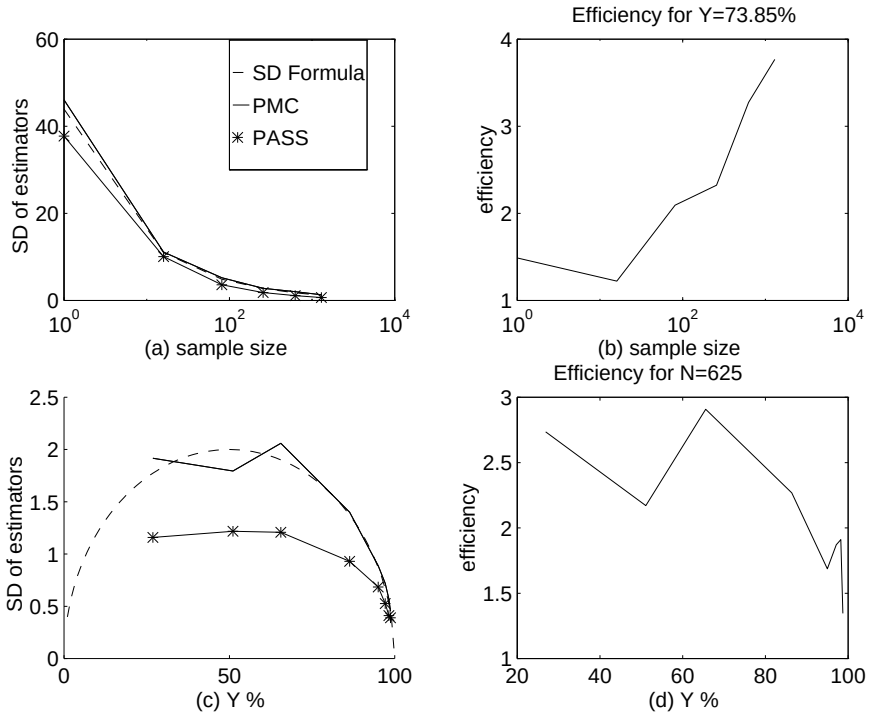

Fig. 12. Efficiency of optimal PASS over PMC estimator for Sallen and Key filter.

Table I PERFORMANCE SPECIFICATIONS FOR SALLEN AND KEY FILTER

Frequency (Hz)	Lower magnitude (dB)	Upper magnitude (dB)
10	-0.25	1.5
100	-0.25	1.5
500	-0.25	1.5
600	-0.25	1.5
700	-0.25	1.5
800	-0.25	1.5
900	-0.25	1.5
1000	-0.25	1.5
3000	$-\infty$	-18
5000	$-\infty$	-27

Figure Captions

Fig. 1. Description of *Problem 1*.

Fig. 2. Partitioning scheme of the tolerance region. (a) Vertical partitioning. (b) Horizontal partitioning.

Fig. 3. Variance of estimator for proportional allocation.

Fig. 4. Description of partitioning for *Problem 2*.

Fig. 5. A 2-dimensional strata scheme of optimal PASS.

Fig. 6. Rosenbrock's curved-valley acceptability region.

Fig. 7. Efficiency of optimal PASS over PMC estimator for Example 1.

Fig. 8. Schematic circuit of Sallen and Key filter.

Fig. 9. Frequency response of Sallen and Key filter.

Fig. 10. Histograms of estimators for Sallen and Key filter, N=81.

Fig. 11. Histograms of estimators for Sallen and Key filter, N=2401.

Fig. 12. Efficiency of optimal PASS over PMC estimator for Sallen and Key filter.

Table Captions

TABLE I. PERFORMANCE SPECIFICATIONS FOR SALLEN AND KEY FILTER