Stratified Random Sampling for Power Estimation

Chih-Shun Ding Qing Wu Cheng-Ta Hsieh Massoud Pedram

Abstract— **In this paper, we present new statistical sampling techniques for performing power estimation at the circuit level. These techniques first transform the power estimation problem to a survey sampling problem, then apply stratified random sampling to improve the efficiency of sampling. The stratification is based on a low-cost predictor, such as zero delay power estimates. We also propose a two-stage stratified sampling technique to handle very long initial sequences. Experimental results show that the efficiency of stratified random sampling and two-stage stratified sampling techniques are 3-10X higher than that of simple random sampling and the Markov-based Monte Carlo simulation techniques.**

I. Introduction

With the continuing reduction in the minimum feature size, chip density and operating frequency of today's ICs are increasing. As a result, power dissipation has become an important concern in IC design. To minimize power, one needs to estimate it first. As a result, there is an increasing need for accurate and efficient power estimation tools.

Existing simulative power estimation techniques [1], [2], [3], [4] explicitly simulate the circuit under a "typical" input vector Their main shortcoming is however that they are very slow. To address this problem, a Monte Carlo simulation technique was proposed in [5]. This technique uses an input model based on a Markov process to generate the input stream for simulation. The simulation is performed in an iterative fashion. In each iteration, a vector sequence of fixed length (called sample) is simulated. The simulation results are monitored to calculate the mean value and variance of the samples. The iteration terminates when some stopping criterion is met. This approach suffers from four major shortcomings. First, since the simulation vectors are generated internally based on statistics of the input stream, a large number of vectors needs to be examined to extract reliable statistics. Second, when the vectors are regenerated for simulation, the spatial correlations among various inputs cannot be adequately captured, which may lead to inaccuracy in the power estimates. Third, the required number of samples, which directly impacts the simulation run time, is approximately proportional to the ratio between the sample variance and square of sample mean value. For certain input sequences, this ratio becomes large, thus significantly increasing the simulation run time. Finally there is a concern about the normality assumption on the sample distribution. Since the stopping criterion is derived based on the normality assumption, if the sample distribution significantly deviate from normal distribution, the simulation may terminate prematurely. Difficult distributions that cause premature termination include bi-modal, multi-modal and distribution with long or asymmetric tails.

In this paper we address the power estimation problem from a survey sampling perspective. We assume a sequence of vectors are provided to estimate the power consumption of a given combinational circuit with certain statistical constraints, such as error and confidence levels. We transform the power estimation problem to a survey sampling problem by dividing the vector sequence into small units, e. g. consecutive vectors, to constitute the population for the survey. Power consumption is the characteristic under study. The average power consumption is estimated by simulating the circuit by a number of samples drawn from the population, a procedure referred to as sampling, using a simulator such as PowerMill [4] or Verilog-XL. Our objective is to design a sampling procedure that will significantly reduce the number of simulated vectors while satisfying the given error and confidence levels.

Stratified sampling techniques have been widely used for surveys because of their efficiency. The purpose of stratification is to partition the population into disjoint subpopulations so that the power consumption characteristic within each subpopulation is more homogeneous than in the original population. The partitioning is based on a low-cost predictor that needs to be efficiently calculated for each member in the population. In this paper, we use the zero delay power estimate as the predictor. Compared to the technique proposed in [5], the proposed technique offers the following advantages: 1) It performs sampling directly on the population and the estimation results are unbiased, 2) It is more efficient, and 3) The sample distributions are more likely to be a normal distribution. When the population size is large, we propose a two-stage stratified sampling procedure to reduce the overhead of predictor calculation and stratification.

The organization of the paper is as follows. In Section 2, we describe the basic principles of survey sampling and its connection with power estimation. In Section 3 and 4, we present a stratified sampling technique for power estimation and discuss its design issues. A two-stage stratified sampling technique is presented in Section 5. Experimental results are presented in Section 6 followed by concluding remarks which are given in Section 7.

II. Background

We first give some useful notation and definitions:

U	the population
N	number of units in the population
u_i	the ith unit in the population
y_i	value of the characteristic under study for u_i
$\overline{Y}$	mean value of y_i in the population
n	sample size

C.-S. Ding is now with Rockwell Semiconductor Systems, Newport Beach, CA.

Q. Wu, C.-T. Hsieh, and M. Pedram are now with Department of Electrical Engineering - Systems, University of Southern California, Los Angeles, CA 90089

We are given a collection (called *population*), $U = \{u_1, u_2, \ldots, u_N\}$ of objects (called *elements* or *units*), of which some property (called *characteristic*) y_i is defined for each u_i. The survey sampling problem deals with ways of selecting samples, i.e., sequences (or collections) of units from the population, to estimate the mean value of the characteristic under study in the population, denoted by $\overline{Y}$, where

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i.$$

For the sake of the simplicity, we will refer to $\overline{Y}$ as *population mean.* The variance of the characteristic under study in the population is simply referred to as *population variance* and is denoted by $V(y)$. The *relative variance* is defined as the ratio between the variance and square of the mean value of a statistic. Number of units included in a sample is referred to as the *sample size* and is denoted by n. An *estimator* (of population mean) is defined as a function of sample characteristic values that estimates the population mean. An estimator is a random variable and may take different values from sample to sample. The difference between the estimator t and $\overline{Y}$, is called *error.* t is said to be an *unbiased* estimator for $\overline{Y}$ if $E(t) = \overline{Y}$, otherwise biased. If an estimator is biased, the bias is given by:

$$B(t) = E(t - \overline{Y}).$$

The *estimator variance* of t is defined as:

$$V(t) = E[(t - E(t))^2].$$

The *mean-square error* (MSE) is defined as:

$$MSE(t) = E[(t - \overline{Y})^2].$$

The relation among $MSE(t)$, $V(t)$, and $B(t)$ is:

$$MSE(t) = V(t) + B^2(t).$$

Given two estimators t_1 and t_2, t_1 is said to be more *efficient* than t_2 if the mean-square error of t_1 is less than that of t_2. The *relative efficiency* of t_1 as compared to t_2 is defined as the reciprocal of the ratio of their estimator variance when the same number of samples are taken in both estimators. Therefore the ratio between the number of required samples is roughly equal to the reciprocal of their relative efficiency.

A. Power Estimation Problem

We are given a vector trace $(v_1, v_2, \ldots, v_M)$ to estimate the average power consumption of a combinational circuit using an accurate simulator. This problem can be easily transformed into a survey sampling problem by grouping a fixed number l of consecutive vectors with overlap occurring only at the group boundaries, and each group becomes an element (unit) of the population in the survey. For example, if $l = 2$ and $M = N + 1$, the grouping will be:

$$(v_1, v_2), (v_2, v_3), \ldots, (v_N, v_{N+1})$$

The power consumption estimated for the vector sequence in each group becomes the characteristic under study. The mean value of the characteristic in the population gives the average power consumption.

B. Simple Random Sampling(SRS)

Simple random sampling is a method of selecting n units out of a population by giving equal probability to all units. Given a sample of n units, $u_1, u_2, \ldots, u_n$ with characteristic values $y_1, y_2, \ldots, y_n$, $\overline{Y}$ is estimated by

$$\overline{y}_{sr} = \sum_{i}^{n} y_i / n,$$

where the subscript sr denotes simple random sampling and $\overline{y}_{sr}$ is referred to as the *sample value.*

Theorem 1: In simple random sampling (wr), the sample value $\overline{y}_{sr}$ is an unbiased estimator of $\overline{Y}$ and its sampling variance is given by:

$$V(\overline{y}_{sr}) = \frac{V(Y)}{n}. \tag{1}$$

Equation (1) shows that the sampling variance is inversely proportional to sample size. In addition, if we select k samples and use the mean of the sample values as the estimator t, the variance is further reduced by k as each sample selection is independent of the other. Thus, the sampling variance of this estimator is inversely proportional to nk, the total number of units drawn from the population. In the following discussion, for the sake of generality, we assume that our estimator t is the mean value of k samples, each with sample size n. When the population variance is not known, one may use t-distribution with degree $(k-1)$ to derive the number of required sample k that achieves a given confidence level, $(1-\alpha)$, and a relative error level, ϵ, provided that the samples follow normality, as given below:

$$k > \left(\frac{t_{\alpha/2} s}{\epsilon\ \sigma}\right)^2 \tag{2}$$

where $t_{\alpha/2}$ is defined so that the area to its right under a t-distribution with degree $(k-1)$ is equal to $\alpha/2$, and σ and s are the mean and standard deviation of the simulated samples, respectively. When the population size is infinite, this procedure is also commonly known as the Monte Carlo simulation approach [5].

In general, the efficiency of simple random sampling is not very high. This can be explained in the context of power estimation as follows. Consider the case where the distribution of the population characteristic (i.e., power consumption) is bi-modal – that is, the characteristic of half of the population is distributed around a right peak, and the characteristic of the other half is distributed around a left peak. Since the selected unit could either come from the right peak or the left peak, the sampling variance is very high.

If, on the other hand, we divide the population into two halves, those units whose characteristic values are around the right peak are put into one subpopulation while the remaining units are put into the other subpopulation, and

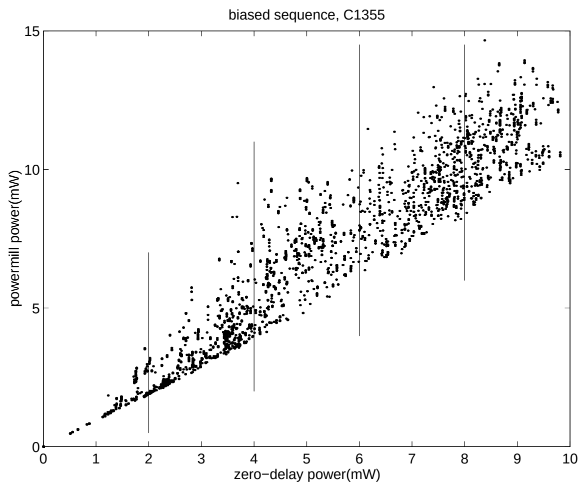

Fig. 1. The plot that shows the power variation after stratification using zero-delay power estimate as the predictor.

select the samples in such a way that half of the units in a sample are selected from each of the subpopulations, then the sampling variance will be significantly reduced. In order to divide the population into subpopulations, a predictor is often used. This predictor need not have a linear relationship with the characteristic under study, as it is only used to divide the population into subpopulations and is not directly used to calculate the power estimates. In the following sections, we will describe a more efficient sampling procedure based on *stratified sampling*.

III. Stratified Random Sampling(STS)

In the stratified random sampling, the population U is partitioned into k disjoint subpopulations, called *strata*. The main objective of stratification is to give a better cross-section of the population so as to gain a higher degree of relative efficiency. The stratification should be done in such a way that strata are homogeneous within themselves w. r. t. the characteristic under study.

For power estimation, we use the following procedure to construct the strata. The zero delay power estimate is used as the predictor. Let the predictor value of unit u_i be x_i. Population U is first sorted according to the x_i value of each unit. Let the new order be $u_1, u_2, \ldots, u_n$. Then $K-1$ separators, $p_1, p_2, \ldots, p_{K-1}$, are selected such that

$$x_1 < p_1 < p_2 < \ldots < p_{K-1} < x_N$$

All the units whose x_i values are between two consecutive separators are put into the same stratum, and the strata disjointly cover the whole population. Let the size of each stratum be N_i, then

$$N_1 + N_2 + \ldots + N_K = N$$

Units in a sample are drawn from each stratum independently so that the sample size within the ith stratum become n_i $(i = 1, 2, \ldots, K)$ and

$$n_1 + n_2 + \ldots + n_K = n$$

If the sample is taken randomly from each stratum, the procedure is known as *stratified random sampling*.

One way to tell how well the x_i behaves in relation to y_i is from the scatter plot of y_i and x_i. In Figure 1, we show the scatter plot for each consecutive vector pair in a biased sequence (i.e. non-random) for $C1355$ circuits (ISCAS85), where x_i and y_i are the zero delay and the PowerMill estimates of each consecutive vector, respectively. As we can see, the zero delay power estimation produces poor accuracy on a vector-by-vector basis. However, in the stratified sampling, x_i is only used for forming the strata so that units of close y_i values can be put in the same stratum, slight inaccuracy of the zero delay power estimates is acceptable.

An intuitive account for how stratified random sampling improves sampling efficiency can be best explained in the context of the scatter plot for biased sequence in Figure 1. In Figure 1, the vertical lines represent the separators. If we apply simple random sampling to this population, the range of variation that we observe from each drawn unit is from 0mW to 15mW. However, when using stratified random sampling, the units drawn from each stratum have a much smaller range of variation due to stratification. For instance, units drawn from the first stratum in Figure 1 now only vary from 0mW to 4mW. Similarly for units from other strata. This result leads to a much smaller sampling variance.

A. Population Parameter Estimation

In this subsection, we mainly focus on the derivation of an estimators for population mean and study of its properties. We first give some useful notations and definitions:

K	number of strata
N_i	size of the ith stratum
W_i	N_i/N, stratum weight of ith stratum
n_i	number of units in a sample falling in ith stratum
y_{ij}	characteristic y of jth unit in ith strata in a sample
$\overline{Y_i}$	stratum mean of the ith stratum, $\overline{Y_i} = \sum_j^{N_i} y_{ij}/N_i$
S_i^2	stratum variance of the ith stratum, $S_i^2 = \sum_j^{N_i}(y_{ij} - \overline{Y_i})^2/N_i$
$\overline{y}_i$	$\sum_j^{n_i} y_{ij}/n_i$, sample mean within strata i

The stratification estimator $\overline{y}_{st}$ is formulated as:

$$\overline{y}_{st} = \sum_i^K W_i \overline{y}_i$$

where the subscript st denotes the stratified sampling method.

Theorem 2: If units in a sample are drawn independently in different strata, then $\overline{y}_{st}$ is an unbiased estimator of the population mean and its sampling variance is given by

$$V(\overline{y}_{st}) = \sum_i^K W_i^2 \frac{S_i^2}{n_i} \tag{3}$$

IV. Stratification Scheme Design

From Equation (3), $V(\overline{y}_{st})$ depends on n_i, W_i and S_i. Our objective is to minimize $V(\overline{y}_{st})$ for a given value of n. The problem of selecting n_i for the ith stratum is referred to as the *sample size allocation* problem. W_i and $V(\overline{y}_i)$ are related and they are determined by strata separators p_i, $i = 1, 2 \ldots, K-1$, when the strata are formed. The problem of finding the optimal values of the separators p_i is referred to as the *stratum selection* problem. These two problems collectively will be referred to as the *stratification scheme design* problem.

A. Sample Size Allocation in Strata

We adopted a variation of the minimum variance allocation scheme proposed by [6]. It was shown that allocating $n_i = n \frac{W_i S_i}{\sum_i^K W_i S_i}$ will give the minimum variance among all possible allocation methods. The difficulty in using the above criterion is that S_i is not known before sampling. One solution is to use the variance of x_i as an approximation of S_i^2. However, this places a condition on the choice x_i since it implies that x_i must be selected judiciously so as to exhibit a variance proportional to that of y_i.

B. Stratum Selection Problem

In general when we increase K, $V(\overline{y}_{st})$ decreases. However, in practice, after K reaches a certain value, the reduction on $V(\overline{y}_{st})$ becomes less significant and sometimes could even increase slightly. Some statisticians [7] have suggested that an increase in K beyond 6 would seldom be profitable.

After determining the value of K, we need to find the strata separator values p_i's. Several selection criteria have been proposed in the literature. Among them, Danlenius and Gurney [8] suggested that the construction of strata on the basis of equalization of $W_i S_i$ and equal sample size allocation to the strata would lead to optimum stratification. Again, this method is not convenient as it requires the knowledge of S_i. Our method is based on a variation of this scheme as described next. The stratum variances S_i^2, where $i = 1, \ldots, K$, is approximated by the variance zero delay power estimate in each stratum. The units of the population are first sorted according to their zero delay power estimates and put in a large number of bins. Adjacent bins are next merged iteratively until K strata are formed and $W_i S_i$ are within 25% of each other.

C. Normality of Sample Distribution

As we have mentioned earlier, the objective of stratification is to make the strata homogeneous within themselves. Thus, it is easier to make the sample distribution follow a normal distribution in stratified random sampling compared to simple random sampling. For example, in *bi-modal or multi-modal population distribution*, if the population modes are widely separated and if we assume that the predictor behaves relatively well, stratification can break up the modes by using a number of strata to cover each mode. Even when one of the strata happens to remain bi-modal, the modes are less likely to be as widely separated as in the original population. In any case, the bimodality behavior only effects the sample units drawn from this stratum and not the whole. Similar arguments hold for population distributions with long or asymmetric tails.

D. Cost Comparison

The cost of simple random sampling is due to the calculation of y_i for all units in the samples, that is, nk times the average simulation time of PowerMill for one vector pair. If we assume the latter to be C_{pwm}, then the cost of simple random sampling can be written as:

$$C_{sr} = a_{sr} + nkC_{pwm}$$

where a_{sr} is a constant overhead for simple random sampling.

For stratified random sampling, the cost comprises of two parts: 1) the calculation of x_i (zero delay power estimate) for all units in the population and stratification (such as sorting and strata selection), and 2) the simulation time of PowerMill for all units in the samples. If we assume the amortized cost of x_i calculation is C_1 and the number of samples is selected so as to achieve the same variance as that of simple random sampling, then the cost of stratified random sampling can be written as:

$$C_{st} = a_{st} + NC_1 + \frac{1}{\eta} nkC_{pwm} \quad (4)$$

where η is the relative efficiency of stratified random sampling vs. simple random sampling and a_{st} is a constant overhead for stratified random sampling.

The constant overheads of both sampling methods are small. Therefore the stratified sampling is more cost-effective if

$$\eta > \frac{1}{1 - \frac{NC_1}{nkC_{pwm}}}$$

In practice, we find that $\frac{C_1}{C_{pwm}} \cong \frac{1}{4000}$. If we assume $N = 4000$ and $nk = 200$, we conclude that when $\eta > 1.005$, stratified random sampling is more cost-effective than simple random sampling.

When the population size is very large, the overhead of calculating the predictor may become significant. To reduce this overhead, a two-stage stratified sampling is proposed in the next section.

V. Two-Stage Stratified Sampling

In the first stage, a sub-population of size $M \ll N$ is first randomly sampled from the original population U. In the second stage, stratified sampling is applied to this sub-population to select a sample of size n. Since stratified sampling is applied to the second stage only, the overhead of calculating x is restricted to the sub-population size. If k samples need to be selected, k subpopulations are randomly selected, and a sample of size n is drawn from each of the selected subpopulation.

The selection of a sample in two-stage sampling consists of two steps. The first step is to select a subpopulation. Once a subpopulation is selected, the selected sample from this subpopulation is an unbiased estimator of the subpopulation mean. Therefore, to calculate the variance of a two-stage sampling, one needs to consider the variance introduced in both steps. The following theorem can be applied to derive the variance of a two-stage estimator.

Theorem 3: [9, p.109] The variance of a random variable X is the sum of the variance of the conditional expected value and the expected value of the conditional variance. Symbolically,

$$V(X) = V_1(E_2(X|Z)) + E_1(V_2(X|Z)) \tag{5}$$

where E_1 stands for expectation of X over the space Z, E_2 stands for the conditional expectation of X for a given Z, V_1 stands for the variance of X over the space Z, and V_2 stands for the conditional variance of X for a given Z.

Using the above theorem and conditioning on the subpopulations, the sampling variance of a two-stage sample can be written as the sum of 1) the variance of the subpopulation means between all possible sub-populations (the first term on the right hand side of (5)), and 2) the mean value of the sampling variance within each subpopulation. The former is referred to as the first stage variance, denoted as S_b^2 and the latter as the second stage variance, denoted as S_w^2, where 'b' stands for "between" subpopulations and 'w' stands for "within" subpopulations.

The difficulty in deriving the sampling variance in a two-stage stratified sampling is that there is a dependency between the stratification and the subpopulation selection. As a result, there is no closed-form expression for the sampling variance in a two-stage stratified sampling. However, the dependency decreases when M increases. When M is large, the probability distribution within a subpopulation will be very close to the original population. Therefore we can assume the relative efficiency η of stratified sampling over simple random sampling is nearly the same in all subpopulations, and can thus approximate the sampling variance using η and the sampling variance in two-stage simple random sampling, in which simple random sampling is employed at both stages.

The S_b^2 in two-stage simple random sampling is exactly $\frac{1}{M}V(Y)$. If the sampling size is 1, the two-stage simple random sampling is equivalent to a single-stage simple random sampling with sample size 1, in which the sampling variance will be $V(Y)$. Therefore

$$S_w^2 = (1 - \frac{1}{M})V(Y)$$

If the sample size is n and k samples are drawn, the estimator variance can be written as:

$$V(t) = \frac{1}{Mk}V(Y) + \frac{1-\frac{1}{M}}{nk}V(Y)$$

Therefore the estimator variance of a two-stage stratified sampling can be approximated by:

$$\begin{aligned} V(t_{2st}) &\cong \frac{1}{Mk}V(Y) + \frac{1-\frac{1}{M}}{nk\eta}V(Y) \\ &\cong (\frac{1}{Mk} + \frac{1}{nk\eta})V(Y) \quad \text{for large M} \end{aligned} \tag{6}$$

TABLE I

Results of 100,000 simulation runs on the random sequence.

	exp.	simulated vec. pairs				improvement of
	rel.	SRS		STS	STS vs. SRS	
ckts	eff.	max	avg	max	avg	
C432	1.76	960	408	690	271	1.50
C880	2.06	720	305	450	201	1.51
C1355	1.57	360	155	270	135	1.15
C1908	1.32	540	226	450	196	1.15
C2670	1.67	420	164	300	139	1.18
C3540	2.14	600	244	450	171	1.43
C5315	1.57	360	150	270	131	1.15
C6288	1.49	360	160	300	141	1.13
C7552	3.53	900	367	450	180	2.04
apex6	4.92	930	422	420	170	2.48
dalu	4.71	990	421	420	173	2.42
des	6.27	480	203	240	114	1.78
i10	2.02	570	225	390	165	1.36
i8	7.32	1080	568	240	118	4.81
pair	3.82	450	180	240	119	1.51
t481	3.51	1050	468	360	146	3.20
avg						1.86

where $2st$ denotes the two-stage stratified sampling.

In the next subsection, we describe how to select M values to maximize the efficiency of the two-stage stratified sampling techniques.

A. Selection of Subpopulation Size

A simple formulation of the sampling cost of the proposed two-stage stratified sampling technique can be written as :

$$C = a_{2st} + kMC_1 + knC_{pwm}$$

where a_{2st} is the constant overhead cost, C_1 and C_{pwm} is the same as in (4). Compared with (6), one may notice that $V(y_{2st})$ decreases with an increase in k, M, and n, while the cost increases. To maximize the sampling efficiency under a given cost constraint, we need to find the optimal values for M, and n. If we apply the Lagrange multipliers method, the optimal ratio of M and n is:

$$\frac{M}{n} = \sqrt{\frac{\eta C_{pwm}}{C_1}}$$

The ratio of the costs in the first stage (kMC_1) and the second stage (knC_{pwm}) is given by:

$$\frac{kMC_1}{knC_{pwm}} = \sqrt{\frac{\eta C_1}{C_{pwm}}}$$

If we assume $\eta = 2$, then $\frac{M}{n} \cong 90$, $\frac{kMC_1}{knC_{pwm}} \cong \frac{1}{45}$. Therefore the impact of M on the total cost is insignificant when the ratio of $\frac{C_{pwm}}{C_1}$ is large. So is its impact on the estimator variance.

VI. Experimental Results

The proposed techniques have been implemented in C and tested on ISCAS85 benchmarks and a set of mid-size to large-size circuits from MCNC91 benchmarks. We report results of two experiments. In the first experiment, we compare the efficiency of the stratified and simple random sampling on a random and a biased sequence. In the second experiment, we compare the efficiency of two-stage stratified sampling with the technique proposed in [5] for

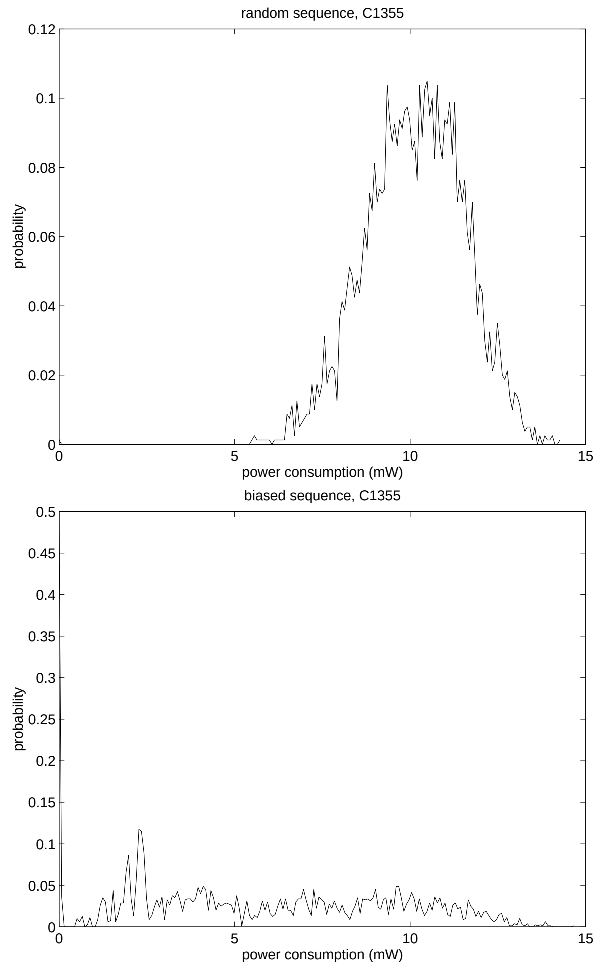

Fig. 2. Typical histograms of power consumption of biased and random sequences.

TABLE II

Results of 100,000 simulation runs on the biased sequence.

ckts	exp. rel. eff.	simulated vec. pairs SRS max	SRS avg	STS max	STS avg	improvement of STS vs. SRS
C432	4.9	4530	2765	1770	1169	2.3
C880	22.2	4650	2955	540	221	13.2
C1355	16.2	2580	1293	450	177	7.3
C1908	7.6	2370	1241	570	252	4.9
C2670	16.0	3900	2328	570	234	10.0
C3540	23.1	3450	1989	450	182	10.9
C5315	39.9	3690	2094	330	152	13.8
C6288	23.8	8730	6141	780	333	18.4
C7552	7.7	4140	2418	1230	755	3.2
apex6	20.9	4320	2554	540	211	12.1
dalu	12.0	4290	2467	720	287	8.6
des	19.1	3240	1648	450	182	9.0
i10	18.4	3930	2420	540	220	11.0
i8	25.3	4710	2932	510	206	14.2
pair	51.0	4080	2441	330	147	16.6
t481	38.2	3630	2103	360	153	13.7
avg						10.5

TABLE III

Error violation percentages.

ckts	error violation ($> x\%$) percentage: biased seq. SRS $> 5\%$	biased seq. SRS $> 20\%$	biased seq. STS $> 5\%$	biased seq. STS $> 20\%$	random seq. SRS $> 5\%$	random seq. STS $> 5\%$
C432	1.46	0.07	0.95	0	1.55	0.75
C880	1.35	0.09	0.39	0	1.05	0.22
C1355	1.96	0.02	0.12	0	0.02	0.00
C1908	2.02	0.01	0.64	0	0.44	0.23
C2670	1.48	0.06	0.50	0	0.03	0.01
C3540	1.70	0.05	0.14	0	0.58	0.07
C5315	2.70	0.07	0.06	0	0.06	0.01
C6288	1.03	0.10	0.86	0	0.03	0.01
C7552	2.16	0.06	0.85	0	1.89	0.18
apex6	1.48	0.07	0.31	0	1.73	0.07
dalu	1.41	0.05	0.90	0	1.80	0.08
des	1.81	0.04	0.13	0	0.25	0.01
i10	1.51	0.06	0.37	0	0.45	0.07
i8	1.36	0.08	0.30	0	2.05	0.01
pair	1.45	0.07	0.01	0	0.13	0.01
t481	1.68	0.04	0.03	0	1.94	0.01

TABLE IV

Results of Monte Carlo simulation vs. two-stage stratified sampling.

ctks	simulated vec. pairs: Markov based max	Markov based avg	two-stage STS max	two-stage STS avg	error violation percentage: Markov based 5%	two-stage STS 5%	improve-ment
C432	1890	1083	1020	467	2.3	0.9	2.3
C880	1230	726	540	193	1.8	0.3	3.8
C1355	600	246	330	171	0.3	0.3	1.4
C1908	1510	729	660	307	2.6	0.7	2.4
C2670	510	254	210	126	0.7	0.0	2.0
C3540	1230	694	480	257	2.1	0.4	2.7
C5315	540	256	270	152	1.1	0.0	1.7
C6288	1230	442	420	195	2.0	0.4	2.3
C7552	1320	558	270	144	2.3	0.0	3.9
apex6	1950	995	240	138	2.2	0.1	7.2
dalu	2100	1106	450	192	1.9	0.1	5.7
des	540	261	180	111	0.6	0.1	2.3
i10	1110	546	540	273	2.1	0.8	2.0
i8	2730	1824	510	204	1.8	0.1	8.9
pair	690	305	240	132	1.0	0.1	2.3
t481	2670	1569	660	248	1.8	0.9	6.3
avg							3.6

infinite-size population (i.e., only the signal and transition probabilities at circuit inputs are given). The results are presented as follows.

A. Experiment I

We performed this experiment on two type of sequences, each of length 4000. The first sequence (random sequence) is generated randomly by assuming 0.5 signal and transition probabilities for every circuit input. The second sequence (biased sequence) is a non-random sequence obtained from industry. The circuits were mapped to a library with NAND, NOR, inverter and XOR gates. Since performing simulation on PowerMill is very time-consuming, we simulate whole sequence once to extract the power consumption for every pair of consecutive vectors. The zero delay power estimates are calculated using a bit-parallel algorithm. The average run time of this algorithm on a Sun SS-20 is 5M gate-vector/sec. Figure 2 shows typical power histograms for both sequences.

The stratification scheme is as described in Section IV-B. The sample size allocation is equal-size allocation. n is set to 30 for both sampling methods. The results are summarized in Tables I, II and III. We first evaluate the 'theoretical' (expected) efficiency improvement based on Equations (1) and (3) when $K = 10$, as shown in the "exp. rel. eff." columns. After performing more experiments we found that the improvement on relative efficiency is not very significant when K is increased beyond 10. Therefore, we used $K = 10$ in the simulation runs. We performed 100,000 simulation runs with a confidence level of 0.99 and an error level of 5% for each circuit. The maximum and average numbers of required vector pairs that satisfy Equation (2) are reported.

We also show the projected run time improvement over PowerMill based on the ratio of the average number of simulated vectors and it is listed the last column of Table I and II.

We observe two different error violation characteristics. One for violating the specified error level. The other one is intentionally set to a higher level to detect the long tails in the sample distribution. We set this level to be 20% and the error violation of this level is referred to as *high-error violation*. The percentages of error violation are summarized in Table III. It shows that the high-error violation does not exist for this set of circuits when using stratified random sampling (although, for biased sequence, simple random sampling results show high-error violation). This demonstrates that stratified sampling can handle difficult population distributions. As expected, there is no high-error violation observed in the random sequence.

The efficiency improvement of stratified sampling for the random sequence is much smaller than that for biased sequence. This is because the transition probability of each circuit input was assumed to be 0.5; From the law of large numbers, the distribution of the number of input bit changes has a high peak around a value equal to half of the circuit input count. Since the number of input bit changes has a direct impact on circuit power consumption, the power distribution is already very homogeneous in this case. Stratification is less effective in homogeneous populations.

B. Experiment II

In this experiment, we compare the efficiency of the proposed two-stage stratified sampling technique to the technique proposed in [5], which is based on a Markov process model. Since this experiment cannot be performed in a reasonable time with PowerMill, we use a real-delay gate-level power estimator instead. The signal and transition probabilities at the circuit inputs was assumed to be 0.5 and 0.25, respectively. The subpopulation size and the sample size for the two-stage sampling approach was set to 4,000 and 30, respectively. In the first stage of the two-stage sampling approach, the random number generator is run twice for each circuit input to generate the vector sequence for each unit in the selected subpopulation. The first run determines the initial input value while the second run determines if the input changes. In the Markov-based approach, only the initial value of the first unit in a sample is randomly generated, the initial values of the remaining units are assumed to follow the previous unit, as described by the Markov process.

We performed 1,000 simulation runs with 0.99 confidence and 5% error levels. Results are summarized in Table IV. There is no high-error violation observed in either technique. The proposed two-stage stratified sampling technique is more than three times as efficient as the Markov-based technique.

VII. Conclusion

We have presented new statistical sampling techniques for circuit-level power estimation. Compared with existing statistical power estimation techniques, not only the efficiency of sampling is improved, but also difficult population distributions can be handled more effectively. Although we have used zero delay power estimates for stratification, other predictors can be used, depending on the trade-off between computation overhead and efficiency improvement. The proposed two-stage stratified sampling technique can be easily extended to multi-stage sampling. However, the predictors used at different stages should be as uncorrelated as possible.

References

[1] S. M. Kang. Accurate simulation of power dissipation in VLSI circuits. IEEE *Journal of Solid State Circuits*, 21(5):889–891, Oct. 1986.

[2] G. Y. Yacoub and W. H. Ku. An accurate simulation technique for short-circuit power dissipation. In *Proceedings of the International Symposium on Circuits and Systems*, pages 1157–1161, 1989.

[3] C. M. Huizer. Power dissipation analysis of CMOS VLSI circuits by means of switch-level simulation. In *IEEE European Solid State Circuits Conf.*, pages 61–64, 1990.

[4] C. Deng. Power analysis for CMOS/BiCMOS circuits. In *Proceedings of 1994 International Workshop on Low Power Design*, pages 3–8, April 1994.

[5] R. Burch, F. N. Najm, P. Yang, and T. Trick. A Monte Carlo approach for power estimation. IEEE *Transactions on VLSI Systems*, 1(1):63–71, March 1993.

[6] J. Neyman. On the two different aspects of the representative method. *J. R. Statist. Soc.*, 97:558–606, 1934.

[7] V. K. Sethi. A note on optimum stratification for estimating the population means. *Aust. J. Statist.*, 5:20–33, 1963.

[8] T. Dalenius and M. Gurney. The problem of optimum stratification ii. *Skand. Akt.*, 34:133–148, 1951.

[9] B. W. Lindgren. *Statistical Theory.* Chapman & Hall, 1993.