Statistical Estimation of Distribution of Power Dissipation in VLSI Circuits

Chih-Shun Ding   Cheng-Ta Hsieh   Massoud Pedram

Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089

Abstract

In the past, average and peak power dissipations have been the primary focus of power estimation techniques and tools. It has however become important to estimate the power distribution of the circuit over a large number of clock cycles. This information is especially useful for determining the circuit reliability, performing dc/ac noise analysis, and choosing appropriate packaging and cooling techniques for IC's. In this paper, we propose to use quantile points of the cumulative distribution function for power consumption to provide information about the power distribution. We will show that the quantile points can be estimated using order statistics and simple random sampling. Next, we describe two techniques based on population pruning and stratification to improve the efficiency of quantile estimation. Both population pruning and stratification are based on a low cost predictor, such as zero-delay power estimate. We also present an efficient technique to directly estimate the cumulative distribution function for power dissipation. Experimental results show the effectiveness of the proposed techniques in providing detailed power distribution information.

1 Introduction

The power consumption per clock cycle of a circuit is regarded as a random variable. Associated with this random variable is a cumulative distribution function. Thus, the tasks of estimating the average and maximum power dissipation reduce to that of estimating the mean and upper bound of the random variable.

A number of techniques have been proposed to estimate the average and maximum power consumption. For estimating the average power consumption, two classes of techniques have been proposed: *statistical* techniques and *probabilistic* techniques. Statistical techniques [1, 2, 3] use statistical sampling methods while probabilistic techniques [4, 5, 6, 7] use probabilistic models to propagate the average value of the switching activity from circuit inputs throughout the circuit. The average power consumption is then calculated from the switching activity of each circuit node.

For estimating the maximum power consumption, existing techniques can also be classified into two classes: *statistical* techniques and *deterministic* techniques. In statististical techniques [8], the maximum power consumption is estimated using the order statistics from a simple random sample. In deterministic techniques, the maximum power is estimated either by solving the max-satisfiability problem [9] or by using approximation techniques [10, 9, 11] to obtain the upper bound on the maximum power consumption. The disadvantage of the statistical technique is that the size of the sample can be high and therefore not as efficient as the deterministic techniques. The disadvantage of the deterministic techniques of [10, 11] is that the upper bound can become very loose when the circuit level count is high or when the circuit is comprised of several circuit blocks and the correlations between the inputs to different circuit blocks are too complicated to model.

The probability distribution function of power consumption in a circuit is difficult to predict. This is because probability distribution functions of power dissipation of two different circuits can be very different in shape. For instance, researchers have found that the distribution function can be uni-modal or multi-modal and the tails of the distribution can be either short or long [2]. Unfortunately, these characteristics cannot be encapsulated by merely the average and the maximum power dissipation values. Some questions that today's chip designers are interested in answering, can be stated as follows:

Q1 Find the minimum power value x such that the circuit power dissipation is smaller than x in percentage of the time y (as measured by, say, number of clock cycles in a synchronous circuit).

Q2 Find the peak power dissipation in the circuit over all time (or clock cycle).

Q3 Find the percentage of time where power dissipation in the circuit is between two known values x_1 and x_2, inclusively.

In this paper, we address the problem of estimating the quantile point in the cumulative probability function (to answer Q1 and Q2 above) and deriving the cumulative probability function itself (to answer Q3). For $\alpha \in (0, 1)$, We assume that an input sequence is given. The given sequence is first broken into consecutive vector pairs and these consecutive vector pairs constitute the population for the estimation. The power consumption of each vector pair is regarded as a random variable. *α-quantile point* of a cumulative distribution function is the value where the cumulative distribution function evaluate to α. For estimating a single quantile point in the cumulative probability function, we can use order statistics of a simple random sample. However, the efficiency of this approach is rather low. We propose two techniques: population pruning and stratified random sampling to improve the efficiency. The objective of population pruning is to remove those units from the population that are not in the quantile interval of interest. Stratified random sampling partitions the population into two strata: one for those units which are likely to reside above the quantile point of interest, one for those which are likely to reside below the quantile point of interest. The objective of stratification is to give a better cross-sectional view of the population to gain higher efficiency. Both of these techniques use the zero delay power estimate as a predictor. The issue of estimating the cumulative probability function is also addressed by simultaneously estimating a set of quantile points in the cumulative probability function followed by a curve-fitting. The proposed technique uses strata of equal weight and equal size sample allocation. We show that the accuracy of this technique is no worse than that of simple random sampling. Experimental results demonstrate that the proposed technique provide detailed power distribution information efficiently.

The organization of this paper is as follows. Section 2 reviews the basic concepts and background material. The problems of estimating a single quantile point and a set of quantile points in the cumulative probability function are addressed in Sections 3 and 4, respectively. Experimental results are presented in Section 5, followed by the conclusion in Section 6.

2 Background

We are given a collection (called *population*), $U = \{u_1, u_2, \ldots, u_N\}$ of objects (called *units*) on which some property (called *characteristic*) y_i is defined for each u_i. For power evaluation purpose, the unit u_i is a vector pair and the characteristic y_i is the power consumption of a combinational circuit C under u_i. In practice, if the vector pairs is specified by a finite vector sequence of length n, we can break the sequence into $n-1$ consecutive vector pairs and the collection of these $n-1$ consecutive vector pairs becomes the population.

The characteristic (or power dissipation) associated with each unit can be regarded as a *random variable*, denoted by X. On this random variable, a discrete *probability density function* (*pdf*) $f(x)$ and a discrete *cumulative distribution function* (*cdf*) $F(x)$[1] can be defined. To simplify the presentation, we assume that these functions can be approximated by continuous functions. Discrete pdfs can be easily handled as well. Another issue is whether or not $F(x)$ is strictly increasing. When $F(x)$ is strictly increasing, the inverse function $F^{-1}(x)$ is well defined. For $\alpha \in (0,1)$, *α-quantile point* of a cdf $F(x)$ is defined as the value $x(\alpha)$ such that $F(x(\alpha)) = \alpha$, or $x(\alpha) = F^{-1}(\alpha)$. In another words, there are exact 100α percent of the population have X values equal to or smaller than α. If $F^{-1}(x)$ does not exist, $x(\alpha)$ can be any value in some interval. In the discussion of this paper, we assume that $F^{-1}(x)$ exists.

Let $X_1, X_2, \ldots, X_n$ be a sample of n observations sampled from the population. n is referred to as the *sample size*. An *estimator* θ is a function of the random variable values on these n selected units that is used to estimate the parameters (such as the mean value and the quantile point, etc.) of the population. An estimator is also a random variable and may take different values from sample to sample. A *confidence interval* is an interval $[b, c]$ which the probability for the estimator θ value to fall into is δ, that is, $Prob(b \le \theta \le c) = \delta$, where δ is referred to as the *confidence level.* Note that the larger the confidence interval, the higher the confidence level.

If the X_i's are sorted and ordered from the smallest to the largest values to be $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, they are defined as the *order statistics* of the sample. The *ith* element in this sorted list is referred to as the *ith* order statistic of the sample, and i is the rank of this order statitic. $X_{(k)}$, $k = 1, 2, \ldots, n$, is also a random variable and its *pdf* $g_k(X_{(k)})$ is:

$$g_k(X_{(k)} = y_k) = \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1 - F(y_k)]^{n-k} f(y_k) \qquad (1)$$

The intuitive interpretation of (0.1) is as follows. For the *kth* order statistic to be y_k, we must have exactly one observation be y_k (the probability for this condition is $f(y_k)$), and $(k-1)$ observations smaller than y_k and $(n-k)$ observations larger than y_k (the probability is $\frac{(n-1)!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-$

[1]In this paper, we use lower and upper case functions to represent the $pdfs$ and $cdfs$, respectively.

$F(y_k)]^{n-k}$) and there are n possible elements to have value y_k. Therefore the pdf is as shown in (0.1).

If we define a new function $Z = F(x)$ and $z_k = F(y_k)$, after using variable substitution, the above function can be rewritten as a beta function $(k, n-k-1)$, and denoted by $h_k(Z)$:

$$h_k(Z = z_k) = \frac{n!}{(k-1)!(n-k)!} z_k^{k-1}(1-z_k)^{n-k} \tag{2}$$

The domain of Z is $[0,1]$. As a result, there are two ways to define the confidence levels of an estimator that is based on order statistics. One is on X domain, and the other is on quantile domain (Z). Throughout this paper we will use $[\beta, \gamma]^q$ to denote an interval in the quantile domain.

3 Quantile Estimation Techniques

In this section, we address the problem of estimating an α-quantile point of cdf $F(x)$. A straightforward approach is to use the rth order statistic from a simple random sample of size n as the estimator θ for the α-quantile point. It has been shown in [12] that when $n \to \infty$, $r/n \to \alpha$ with $0 < \alpha < 1$, and $F(x)$ is differentiable with density function $f(x_\alpha)$ at x_α where $x_\alpha = F^{-1}(\alpha)$, then $X_{(r)}$ is asymptotically normal with mean x_α and variance $\alpha(1-\alpha)/nf^2(x_\alpha)$. Intuitively, this implies that given a fixed n value, the optimal r value can be approximated as [12]:

$$r \approx \lceil n\alpha \rceil, \tag{3}$$

where $\lceil\rceil$ is ceiling function.

Table 0.1 shows the values of the smallest n and its corresponding optimal r that achieve 0.99 confidence level on some commonly used confidence intervals $[\alpha - \epsilon, \alpha + \epsilon]^q$. $\Delta^{\lceil n\alpha \rceil}$ is the actual confidence level using $\lceil n\alpha \rceil$ approximation. While $\lceil n\alpha \rceil$ is sometimes different from the optimal r value, the error is often very small.

The empirical relation between n, ϵ and α for achieving 0.99 confidence level from Table 0.1 is:

$$n \approx 6.67(\frac{1}{\epsilon})^2 \alpha(1-\alpha). \tag{4}$$

Note that given a fixed ϵ value, n is greatest when $\alpha = 0.5$ and smallest when α is close to 0 or 1.

A different way to define the confidence interval is on the X domain by using two specific order statistics, $X_{(i)}$ and $X_{(j)}$ [13]. The analysis of this type of confidence interval selection is very similar

Table 1: The n, r values to achieve 99% confidence level on some commonly used quantile and confidence interval values

α	0.1			0.25			0.5			0.75			0.9		
ϵ	0.05	0.02	0.01	0.05	0.02	0.01	0.05	0.02	0.01	0.05	0.02	0.01	0.05	0.02	0.01
n	221	1478	5955	492	3105	12435	661	4143	16585	492	3105	12435	221	1478	5955
r	21	124	594	122	776	3108	331	2072	8293	371	2330	9328	201	1332	5362
$\lceil n\alpha \rceil$	23	148	596	124	777	3109	331	2072	8293	370	2329	9327	199	1331	5260
$\Delta^{\lceil n\alpha \rceil}$	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99

and both of them are population independent. However, it returns an interval for $u_\alpha = F^{-1}(\alpha)$, instead of a single value.

Lastly, we would like to give an intuitive explanation why Monte Calro simulation is not suitable for quantile estimation. Order statistics are normal distributions only asymptotically and even then under very stringent conditions (cf. beginning of this section). The required n value is very population dependent ($\gg 2000$, compared to 30 on average power estimation). Since we need at least two samples to converge and we cannot determine the appropriate n value, it makes the approach very inefficient.

In the remainder of this section, we present two techniques to improve the efficiency of quantile point estimation.

3.1 Population Pruning

Given a confidence interval $[\beta, \gamma]^q$, and let $b = F^{-1}(\beta)$, $c = F^{-1}(\gamma)$. If we remove as many units with X value greater than c or smaller than b as possible, we can improve the efficiency as explained next. Let U_M be a subset of U such that all the units in U_M have X values greater than c, and U_m be a subset of U such that all the units in U_m have X values smaller than b. $\hat{U} = U - U_M - Um$. We can derive a more efficient estimator using the order statistics of a sample that is drawn from $\hat{U}$ as implied in the following Lemma.

Lemma 3.1 *Let U be the original population and $|U|$ be the number of units in the population. Let X be a random variable defined on U. Let $[\beta, \gamma]^q$ be the confidence interval and b, c, U_m, U_M, and*

$\hat{U}$ be defined as above. Let the new quantile of b and c on $\hat{U}$ be β' and γ', respectively, then

$$\beta' = \frac{\beta|U| - |U_m|}{|\hat{U}|},$$

$$\gamma' = \frac{\gamma|U| - |U_m|}{|\hat{U}|},$$

and

$$\gamma' - \beta' = \frac{|U|}{|\hat{U}|}(\gamma - \beta)$$

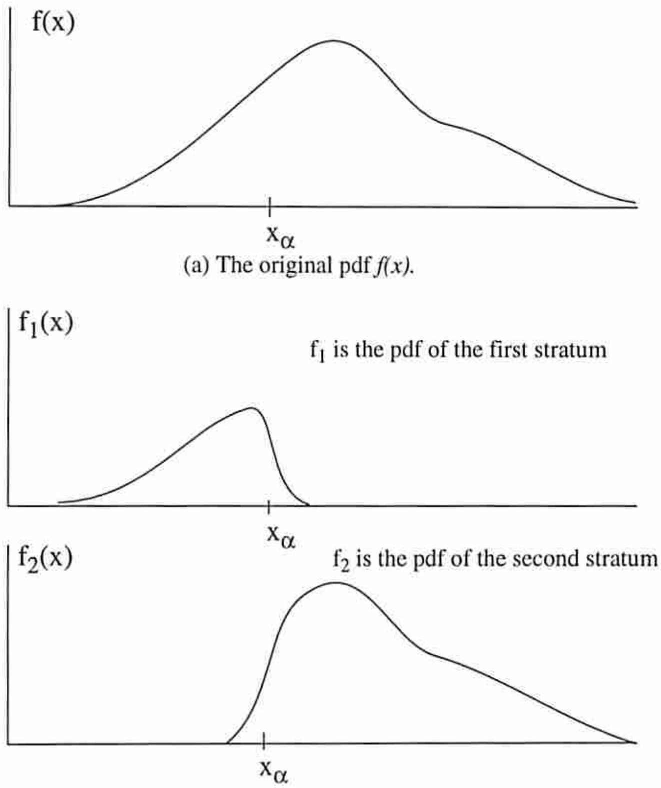
Proof Straightforward. ∎

Since the new quantile interval has been increased by $\frac{|U|}{|\hat{U}|}$, a sample with fewer observations is now needed to achieve a given confidence level(cf. (0.4)). The required sample size to achieve 99% confidence level can be calculated by first computing the new α and ϵ values on $\hat{U}$ and then plugging them into (0.4).

In practice, this "population pruning" procedure can be accomplished using a predictor with predictable error bounds. Let the confidence interval be $[\beta, \gamma]^q$, $i = \lfloor n\beta \rfloor$, and $j = \lceil n\gamma \rceil$, where $\lfloor\rfloor$ is the flooring function. We can sort the population based on predictor values. Let the predictor values on the ith and jth units be d_i and d_j, respectively. In another word, they are the β-quantile and γ-quantile points on the predictor domain, respectively. If the error bounds of the predictor is $[-20\%, 100\%]$, all units with predictor values smaller than $0.8d_i/2$ or greater than $2d_j/0.8$ can be removed from the population. The effectiveness of population pruning depends on how tight the error bounds are and the range of X values. In general, it may be difficult to derive tight error bounds. However, when carefully used, one can derive the error bounds using statististical inference at the cost of slight reduction in the confidence level due to the fact that some units in the quantile interval of estimation interest being removed incorrectly.

3.2 Stratified Sampling (STS)

Another technique to improve the efficiency is based on stratified sampling [14]. In stratified sampling, the population is first partitioned into a number of disjoint subpopulations, called *strata*, of known weights (representing the percentage of units in the strata). Then a predetermined number of units are drawn from each stratum. These units collectively constitute a sample. In [12], an estimator for quantile points based on stratified sampling of two strata is investigated, however,



(a) The original pdf *f(x)*.

(b) The pdfs of the strata.

Figure 1: An example of stratification.

the author makes no comments about how the strata should be designed. This is however the key problem that must be addressed. In the following, we present a method of stratifying the population into two strata to obtain the optimal estimator for a given quantile point using zero delay power estimate as the predictor.

Given the confidence interval $[\alpha-\epsilon, \alpha+\epsilon]^q$, the population is stratified into two strata with weights w_1 and w_2, and $w_1 + w_2 = 1$. The way we construct these two strata is as follows. We first sort the population according to the predictor values. All the units on the left-hand side of the rth unit, where $r = \lceil n\alpha \rceil$, are put in one stratum, and remaining units in the other stratum. The reason of doing so is that units on the left-hand side of the rth unit are likely to have X values smaller than $x_\alpha = F(\alpha)$. Similarly, the units on the right-hand side are likely to have X values greater than x_α. Therefore, the units that reside in $[\alpha-\epsilon, \alpha+\epsilon]^q$ will be moved to either the upper quantiles of the first stratum or the lower quantiles of the second stratum. From (0.4), this minimizes the required sample size.

Let the number of observations in a sample drawn from each stratum be n_i, and $n_1 + n_2 = n$. Since n_i's and w_i's could be different for both strata, the "importance" (or weight) of the observations drawn from each stratum should reflect this difference. Therefore, all observations from ith stratum are assigned a weight of w_i/n_i. After the sample is sorted to form the order statistics, we sum from the left (smallest) to right (largest) the weights of the order statistics. As soon as this sum becomes greater than α, we stop and return the corresponding order statistic (that cause this) as the estimator. We now give an example to show how the estimator is selected when stratified sampling is applied.

Example:

Assume $w_1 = 0.2$, $w_2 = 0.8$, $n_1 = 4$, $n_2 = 8$, and $\alpha = 0.35$. Let the observations drawn from the each stratum be (1.2, 3.4, 2.7, 0.5), (0.7, 2.3, 1.4, 0.9, 1.6, 2.4, 1.5, 2.9). The order statistics from this sample and their associated weights, when represented as a tuple, are: (0.5, 0.05),(0.7, 0.1), (0.9, 0.1), (1.2, 0.05), (1.4, 0.1), (1.5, 0.1), (1.6, 0.1), (2.3, 0.1), (2.4, 0.1), (2.7, 0.05), (2.9, 0.1), (3.4, 0.05). The first order statistics with accumulated weight exceeding α is 1.4, therefore $\theta = 1.4$.

The main difference between this technique and population pruning is: 1) no units are removed from the population; they are just moved to different strata, and 2) the rank of the order statistic that is used as the estimator cannot be determined in advance, that is, the rank changes from one sample to next.

[12] suggests that if n_i is allocated such that

$$n_i \propto w_i \sqrt{F_i(x_\alpha)(1 - F_i(x_\alpha))}, \tag{5}$$

one could obtain the optimal estimator, where $F_i()$ is the cdf of the ith stratum, and $x_\alpha = F^{-1}(\alpha)$. In practice, the difficulty of applying this criterion is that $F_i(x_\alpha)$ is not known in advance. However, as the population is partitioned at the α-quantile point on the predictor domain, we expect that $\sqrt{F_i(x_\alpha)(1 - F_i(x_\alpha))}$ is approximately the same for both strata. This reduces the above criterion to proportional allocation, which allocates $n_1 = n\,w_1$ and $n_2 = n\,w_2$.

The merit of population pruning over stratification is that the confidence level can be calculated before sampling. If the confidence level needs to be accurately calculated, stratification may not be a good choice. On the other hand, stratification does not require the predictor to have predictable error bounds.

Population pruning and stratified sampling can be combined to further improve the accuracy of the estimator and the combined technique can be regarded as a special case of stratified sampling with 4 strata in which two of them have $\sqrt{F_i(x_\alpha)(1 - F_i(x_\alpha))} = 0$ and we allocate no observation to these two strata.

3.3 Efficiency Analysis

While population pruning and stratified sampling techniques can reduce the sample size when compared with simple random sampling and thus reduce the run time of power simulation (using PowerMill for instance), there is an overhead for these two techniques to calculate the predictor. In the following, we derive the condition where population pruning and stratified sampling techniques are more efficient. The *relative efficiency* of two sampling techniques, denoted by η, is defined as the inverse ratio of the required sample sizes in these two techniques when achieving the same confidence level. Let the population size and the sample size required by simple random sampling be N and n, respectively, and η be the relative efficiency of population pruning (or stratified sampling) over simple random sampling. Therefore, the required sample size in population pruning (or stratified sampling) is n/η. In addition, let the cost of zero delay simulation and power simulation for one clock cycle be C_{zd}, C_{ps}, respectively. The population pruning (or stratified sampling) technique is more efficient than simple random sampling when:

$$\begin{aligned} nC_{ps} &> NC_{zd} + nC_{ps}/\eta, \\ \eta &> \frac{1}{1 - \frac{NC_{zd}}{nC_{ps}}}. \end{aligned}$$

When using PowerMill to perform power simulation, $\frac{C_{zd}}{C_{ps}} \cong \frac{1}{4000}$. If $N = 4000$ and $n = 400$, η needs to be greater than 1.003. When N is large, the required value of η could be large. One can adapt the concept of multi-stage sampling [14] to first obtain a subpopulation using simple random sampling and then draw the sample from this subpopulation. The size of this subpopulation becomes N in the above equations and thus reduces the required η value. The empirical rule is to set the subpopulation size to be 5 to 10 times of n. Fortunately, most of the estimator based on order statistics are population independent, thus n can be determined beforehand and so is the size of the subpopulation.

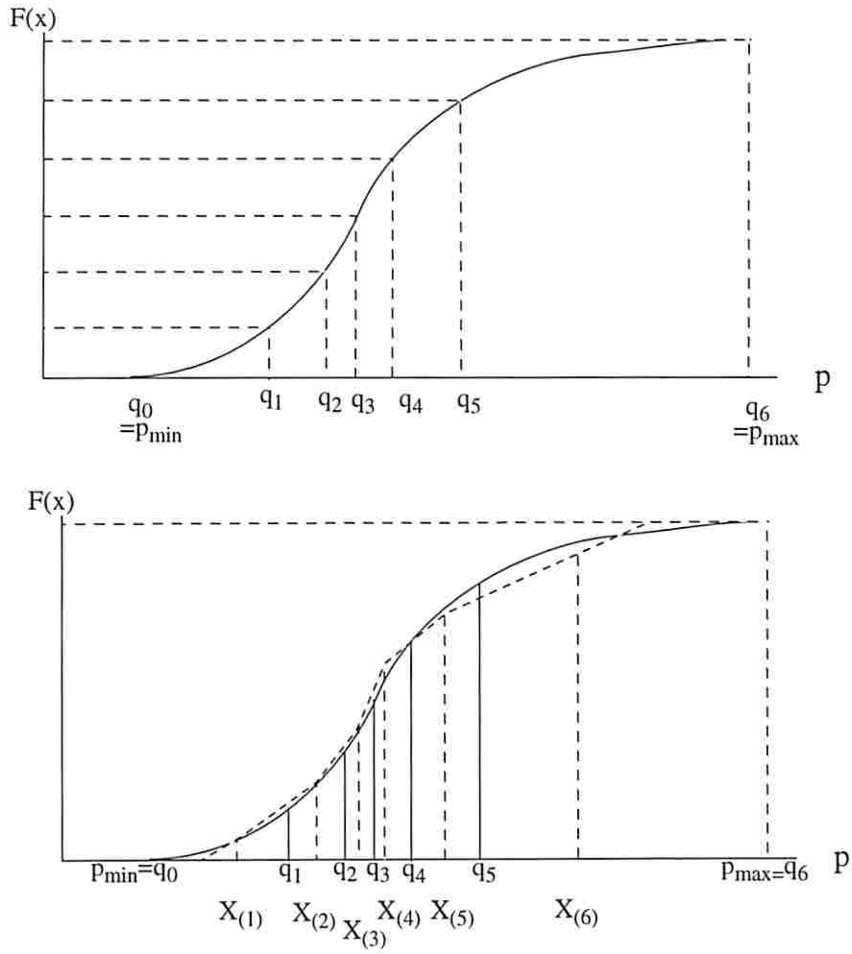


Figure 2: The approximation of cdf f(x).

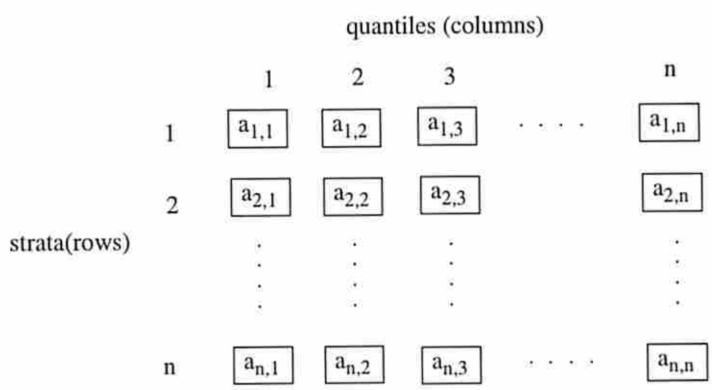


Figure 3: Stratified random sampling for distribution estimation.

4 Distribution Estimation Techniques

In this section, we address the problem of estimating the cdf of power consumption. Our approach is to construct an empirical *cdf* $\hat{F}(x)$ by simultaneously estimating a set of α_i-quantile points that cover the domain $[0,1]^q$, e.g. $\alpha_i = i/n, i = 1, .., n-1$. If the sample size n is adequately large, the empirical $\hat{F}(x)$ will approach the true $F(x)$. Although one may also construct $\hat{F}(x)$ by estimating each quantile point separately, it is very inefficient when the number of quantile points is large. This is mainly because the order statistics of a sample that is used to estimate a specific quantile point can be also used to estimate other quantile points. This type of information is lost when the quantile points are estimated separately.

Given a cdf $F(x)$, sample size n, and a set of monotonically increasing α_i values, $i = 1, \ldots, n-1$. For every α_i-quantile points, we want to use the ith and $(i+1)$st order statistics as the confidence interval for $x_i = F^{-1}(\alpha_i)$. Therefore, we need to maximize the following probability[2]:

$$Prob(X_{(1)} < x_1 < X_{(2)} < x_2 \ldots < x_{n-1} < X_{(n)}). \tag{6}$$

4.1 Simple Random Sampling (SRS)

In the case of simple random sampling, we can show that the maximum of (0.6) occurs when $\alpha_i = i/n$, as stated in the following theorem.

Theorem 4.1 *Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics of a simple random sample of size n from a population U on which a random variable X is defined. Let $\alpha_k, k = 1, 2, \ldots, n-1$ be a sequence of monotonically increasing real number between 0 and 1. Assume that $F^{-1}(x)$ exists and $x_i = F^{-1}(\alpha_i)$. $Prob(X_{(1)} < x_1 < X_{(2)} < x_2 \ldots < w_{n-1} < X_{(n)})$ is maximized when $\alpha_i = i/n$. The maximum value of this probability is $n!/n^n$.*

Proof Let the quantiles corresponding to $X_{(k)}$ be q'_k, that is, $q'_k = F(X_{(k)})$.

$$Prob(X_{(1)} < x_1 < X_{(2)} < x_2 \ldots < x_{n-1} < X_{(n)}) = Prob(q'_1 < \alpha_1 < q'_2 < \alpha_2 < \ldots < \alpha_{n-1} < q'_n)$$

[2] Alternatively, one can also maximize the average of the confidence levels of all quantile points. However, the analysis is more complicated. Empirically, we found these two confidence levels are related.

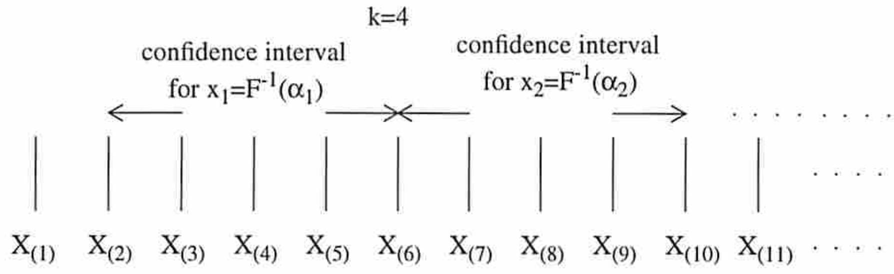


Figure 4: Selecting order statistics for quantile bound.

$$= n! \prod_{i=1}^{n} (\alpha_i - \alpha_{i-1})$$

where $\alpha_0 = 0$, $\alpha_n = 1$. If we define a set of new variables w_i, $i = 1, 2, \ldots, n$, where $w_i = \alpha_i - \alpha_{i-1}$, we have:

$$\text{Maximize} \quad n! \prod_{i=1}^{n} w_i \quad \text{subject to} \quad \sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0$$

The solution is apparent. ∎

The above theorem implies that the order statistics can be used to estimate a set of α_i-quantile points simultaneously and it is most efficient when $\alpha_i = i/n$. In Figure 0.2, we show the $\hat{F}(x)$ constructed by a piece-wise linear function of the order statistics. Since the actual quantile of $X_{(i)}$ is between $[(i-1)/n, i/n]^q$, we assume that it is at the midpoint of $[(i-1)/n, i/n]^q$, that is, $F(X_{(i)}) = i/n - 1/2n$.

The efficiency of simple random sampling is not very high. More specifically, using Sterling's formula for $n!$, one can find that the maximum value stated in the above theorem is:

$$\frac{n!}{n^n} \approx \sqrt{2\pi n}(\frac{n}{e})^n \cdot \frac{1}{n^n} = \frac{\sqrt{2\pi n}}{e^n} \tag{7}$$

where e is the base of natural logarithm.

One can increase sample size to improve the confidence level. However, we have more order statistics than the quantile points. We need to select a subset of those order statistics to bound the quantile points in (0.6). Let the number of quantile points to be estimated be n and the size of the sample be $k(n+1)$, we can pick $X(r)$ where $r = \lceil k/2 \rceil + kr, k = 1, \ldots, n$, as shown in Figure 0.4.

In the following we present a technique based on stratified sampling to improve the estimation efficiency.

4.2 Stratified Sampling (STS)

To give an intuitive motivation for using stratified random sampling for estimating a set of quantile points, let us assume that we want to take a sample of size 10 to estimate the α_i-quantile points in increments of 1/10. In addition, let us assume that the population can be perfectly partitioned into ten strata (using ideal predictor function) and each stratum contains all units in the population between two consecutive quantile points (this case corresponds to the case of *perfect stratification*). If we use equal allocation and take one observation from each stratum, the probability stated in (0.6) is 1. It is true that this is the ideal case, and in practice, the maximum probability of 1 is never achieved. However, we will show that the probability obtained using STS is never less than what can be achieved using SRS. First we give some definition on matrices.

Let A be an $n \times n$ matrix and $a_{i,j}$ be the ith row and jth column entry in A. The sum of all entries in a particular row i is referred to as *rowsum* and denoted by $a_{i\cdot}$. Similarly, the sum of all entries in a particular column is referred to as *columnsum* and denoted by $a_{\cdot i}$. The permanent of A, denoted as $per(A)$, is "the determinant without the sign", calculated as:

$$per(A) = \sum_{p \in permu} \prod_{i=0}^{n} a_{i,p(i)} \tag{8}$$

Let the sample size and number of strata be n. A is a $n \times n$ matrix. Each entry $a_{i,j}$ in A represents the portion of units in the ith stratum that are located in jth quantile interval $[(j-1)/n, j/n]^q$ in the original population. When drawing an observation from the ith stratum, the probability that this observation is from quantile interval $[(j-1)/n, j/n]^q$ is $a_{i,j}$. Therefore, matrix A has the following two properties: 1) all entries $a_{i,j}$ are non-negative, and 2) all columnsums and rowsums are 1. The estimation is correct only when no two observations are from the same quantile interval. Therefore (0.6) calculates $per(A)$.

Theorem 4.2 *Let A be an $n \times n$ matrix with the following properties:*

1. $0 \leq a_{i,j} \leq 1$,
2. *each columnsum and rowsum is 1,*

then

$$\frac{n!}{n^n} \leq per(A) \leq 1 \tag{9}$$

Proof For upper bound,

$$\begin{aligned} per(A) &= \sum_{i=0}^{n} a_{i\cdot} - \sum_{p \notin permu} \prod_{i=0}^{n} a_{i,p(i)} \\ &\le \sum_{i=0}^{n} a_{i\cdot} = 1. \end{aligned}$$

The lower bound is also known as the *Van der Waerden Conjecture* and is proved in 1981. For the proof, please refer to [15]. ∎

While the upper bound is 1, the actual confidence level depends on the correlation between the predictor and power value. If needed, one can also increase the sample size to further improve the confidence level as in SRS. In practice, we stratify the population into k strata with equal weights. The population is first sorted based on predictor values. The first $1/k$ of the units are put in the first stratum, and the second $1/k$ of the units are put in the second stratum, etc. We then draw equal number of observations from each stratum. Lastly, estimation of power distribution can be combined with average power estimation, that is, the same sample can be also used to estimate the average power.

5 Experimental Results

The proposed techniques have been implemented in C and tested on ISCAS85 benchmarks. We use a gate level simulator to calculate the actual power and a bit-parallel zero delay power simulator as the predictor. The reason that we use a gate level simulator, instead of a circuit simulator such as PowerMill, is that it will take too long to finish the experiments. The main objective of the experiments presented in the following is to compare the relative efficiency of different techniques.

The circuits are mapped to a library with NAND, NOR, inverter and XOR gates. We test on two types of populations with size $40,000$. One is comprised of random vectors with 0.5 signal and transition probabilities, and will be referred to as the **random sequences**. Another one is a mixture of random vectors with 0.5 signal probability but variable transition probabilities, and will be referred to as the **biased sequences**. Typical distributions of these two types of sequences are as shown in Figure 0.5. The minimum, maximum and the quantile points for $\{0.1, 0.25, 0.5, 0.75, 0.9\}$ are listed in Table 0.2.

In the first set of experiments, we apply the proposed techniques to estimate α-quantile points in the population, where $\alpha \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$, with absolute error and confidence levels of

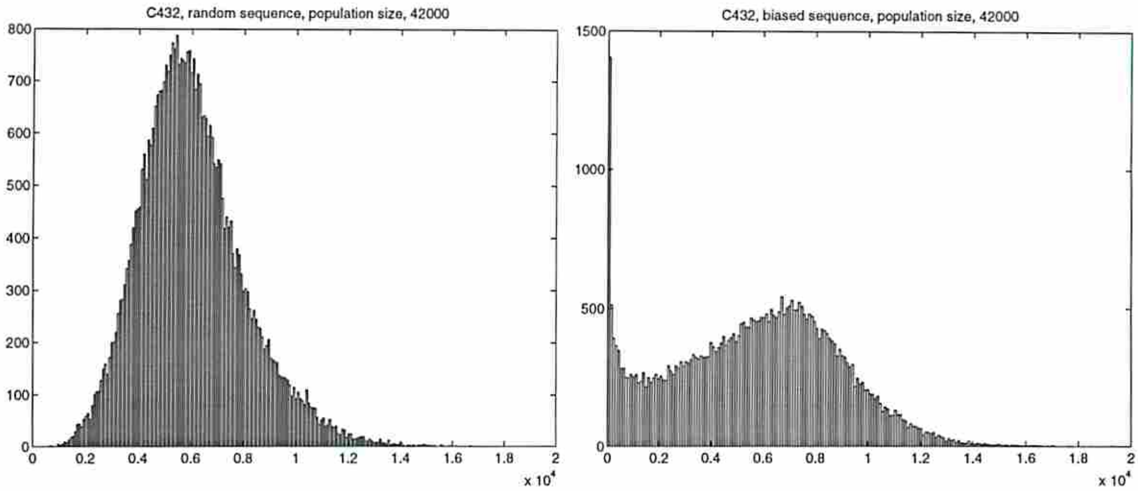


Figure 5: Typical power distribution of biased and random sequences.

Table 2: The quantile points of the distributions on random and biased sequences.

	random sequences							biased sequences						
ckts	min	0.1	0.25	0.5	0.75	0.9	max	min	0.1	0.25	0.5	0.75	0.9	max
C432	0.6	3.8	4.7	5.8	7.2	8.7	16.7	0.0	0.9	3.2	5.8	7.9	9.6	19.3
C880	2.8	7.6	9.0	10.8	12.9	15.1	35.6	0.0	2.2	5.8	10.3	14.9	17.9	35.2
C1355	5.8	10.6	11.8	13.2	14.5	15.6	19.9	0.0	5.7	9.5	12.6	14.8	16.3	21.8
C1908	2.4	10.7	11.7	13.9	16.8	19.4	32.8	0.0	3.9	8.8	13.6	17.0	20.1	32.7
C2670	9.9	17.2	18.8	20.5	22.3	23.8	32.4	0.0	5.4	12.3	20.5	26.2	29.0	38.5
C3540	5.3	26.8	33.0	39.9	46.6	52.4	79.0	0.0	7.6	22.2	39.5	52.3	60.5	84.4
C5315	27.3	46.8	51.0	55.8	60.8	65.6	87.3	0.0	15.1	33.9	55.6	70.3	78.0	101.6
C6288	74.4	293.9	336.7	379.0	416.2	447.3	574.8	0.0	113.7	250.5	371.6	439.0	478.4	609.1
C7552	38.4	64.9	72.9	95.6	121.7	130.1	160.8	0.0	20.0	49.4	94.5	137.9	152.5	196.2

Table 3: Results of population pruning on the biased sequences, 1000 experiments.

	α														
	0.1			0.25			0.5			0.75			0.9		
ckts	n	err	η	n	err	η	n	err	η	n	err	η	n	err	η
C432	228	0.5	26	4164	0.8	2.9	10222	0.8	1.6	8281	0.8	1.5	4023	0.5	1.5
C880	213	0.8	27.9	1759	0.8	7.0	7784	1.7	2.1	6293	0.9	2.6	3064	0.9	1.9
C1355	707	0.5	8.4	5102	1.0	2.4	10462	1.6	1.6	8157	1.0	2.0	3996	0.6	1.5
C1908	412	1.3	14.4	3616	1.0	3.4	9884	1.0	1.6	7491	1.2	2.2	3608	1.0	1.7
C2670	273	0.7	21.8	2532	0.7	4.9	8963	0.7	1.8	7196	0.7	2.3	3550	0.8	1.7
C3540	179	1.2	33.3	2644	1.0	4.7	9180	0.9	1.8	7424	1.1	2.2	3627	1.4	1.6
C5315	328	1.0	18.1	3218	1.1	3.9	9453	1.4	1.7	7700	1.2	2.2	3801	0.8	1.6
C6288	793	1.2	7.5	7636	1.8	1.6	11858	0.9	1.4	9160	0.9	1.8	4421	0.9	1.3
C7552	302	1.2	19.7	2875	1.2	4.3	8867	0.9	1.8	7865	1.0	2.1	4013	0.8	1.5
avg			19.7			3.9			1.7			2.1			1.6

Table 4: Results of combining population pruning and stratified sampling on random sequences, 1000 experiments.

	α														
	0.1			0.25			0.5			0.75			0.9		
ckts	n	err	η	n	err	η	n	err	η	n	err	η	n	err	η
C432	4700	0.7	1.3	9400	1.0	1.3	12500	1.8	1.3	9400	0.9	1.3	4400	2.2	1.3
C880	4100	0.3	1.5	9300	0.1	1.3	12300	1.0	1.3	9600	2.2	1.4	4200	2.2	1.3
C1355	4600	1.7	1.3	9600	1.4	1.3	12800	1.2	1.3	9600	1.1	1.3	4600	0.9	1.3
C1908	4600	1.1	1.3	9600	1.5	1.3	12700	1.6	1.3	9600	2.1	1.3	4600	2.5	1.3
C2670	4700	0.3	1.3	9600	0.2	1.3	12800	0.9	1.3	9700	0.8	1.3	4700	0.7	1.3
C3540	4600	1.1	1.3	9600	0.6	1.3	12700	0.8	1.3	5300	1.1	2.3	3300	1.4	1.3
C5315	4700	1.4	1.3	9700	1.0	1.3	12800	1.2	1.3	9700	1.2	1.3	4700	1.2	1.3
C6288	4700	0.5	1.3	9600	1.1	1.3	12800	0.8	1.3	9700	0.9	1.3	4700	1.0	1.3
C7552	4700	0.1	1.3	9600	0.0	1.3	12800	0.2	1.3	9700	1.0	1.3	4700	1.4	1.3
avg			1.3			1.3			1.3			1.3			1.3

Table 5: Results of combining population pruning and stratified sampling on biased sequences, 1000 experiments.

	α														
	0.1			0.25			0.5			0.75			0.9		
ckts	n	err	η	n	err	η	n	err	η	n	err	η	n	err	η
C432	100	0.3	60	1700	0.6	7.3	7500	0.9	2.2	7500	1.0	1.6	4000	1.0	1.5
C880	100	0.1	60	600	1.2	20.7	4500	0.4	3.6	4200	1.1	3.0	2800	0.6	2.1
C1355	400	0.5	15	2600	0.7	4.8	7500	0.3	2.2	6200	1.1	2.0	3500	0.7	1.7
C1908	250	0.9	24	2200	1.0	5.7	7200	1.0	2.3	6500	1.2	1.9	3600	1.2	1.7
C2670	100	0.4	60	600	1.1	20.7	4000	0.2	4.2	2600	1.1	4.8	2400	1.0	2.5
C3540	100	0.6	60	1200	1.0	10.3	5000	1.0	3.3	5300	0.8	2.3	3300	0.8	1.8
C5315	150	1.0	40	1000	0.9	12.4	4000	0.1	4.2	4000	1.1	3.1	3200	0.8	1.9
C6288	450	0.8	13	2900	0.9	4.3	6000	0.6	2.8	6700	0.4	1.9	3600	1.2	1.7
C7552	100	0.8	60	700	0.4	17.7	4000	0.4	4.2	5000	0.7	2.5	3500	1.1	1.7
avg			44			12			3.2			2.6			1.8

0.01, and 0.99, respectively. In all of our experiments, the error levels (or confidence intervals) are defined on the quantile domain. We perform 1000 experiments for each circuit and quantile point combination. Table 0.3 lists the results using only population pruning on the biased sequences. The random sequences are not listed here, because only less than 2% of the population is pruned. The pruning criterion is such that all units with zero delay power estimates smaller than $0.6d_i$, or greater than $1.6d_j$, are removed, except for C432 in which we set the values to be $0.6d_i$ and $2d_j$, respectively (cf. Subsection 0.3.1). The 'err' columns list the percentage of the experiments that violate the error level. Some of those errors are slightly greater than 1%, mainly because the error bounds on the predictor are not tight. The required sample size for simple random sampling can be found from Table 0.1. The 'η' columns list the relative efficiency of population pruning over simple random sampling as defined in Subsection 0.3.3.

Next, we combine population pruning with stratified sampling. Unlike population pruning technique, stratified sampling cannot predict the required sample size in advance. The way that we conduct this set of experiments is to try different n values until it achieves approximately 0.99 confidence level. Then we compare the n values to get the relative efficiency. The results are summarized in Table 0.4 and Table 0.5. It shows that stratified sampling can further reduce the required sample size. The reason for η to be smaller on random sequences is that the variances of the power dissipations on this

Table 6: Results of 1,000 experiments on distribution estimation.

	random sequence			biased sequence		
	n			n		
circuit	SRS	STS	η	SRS	STS	η
C432	13000	9500	1.37	13000	6000	2.16
C880	12500	7500	1.66	13000	4000	3.25
C1355	13000	10000	1.30	13000	6000	2.16
C1908	13000	10500	1.24	12500	6500	1.92
C2670	12500	7000	1.78	12500	2500	5.00
C3540	12500	8500	1.47	13000	4500	2.88
C5315	12500	9500	1.32	13000	3200	4.06
C6288	12500	9000	1.39	13000	4500	2.89
C7552	12500	6000	2.08	13000	3500	3.71
avg			1.66			3.11

type of sequences are not high. Therefore the correlations between actual power and the predictor is lower than those on the biased sequences.

For estimating the cdfs, we set the quantile increments to 1/50, which correspond to an error level of 0.01. The confidence level is 0.99. This is the average confidence levels of all 50 quantile points. The number of experiments and strata are set to 1000 and 500, respectively. Since we cannot predict the required sample size in stratified sampling in advance, we try different n values until it achieves the confidence level, i.e. the average error violation rate is less than 0.01. The results are listed in Table 0.6. Again the improvement of STS on biased sequences is better than that on random sequences.

6 Conclusion

In this paper, we have proposed to use quantile points of the cumulative distribution function for power consumption to provide information about the power distribution. We proposed two technques: population pruning and stratified sampling, both of which are based on a low cost predictor. The experimental results showed that the proposed techniques provide detailed power distribution information efficiently.

References

[1] 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.

[2] 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.

[3] F. N. Najm, S. Goel, and I. N. Hajj. Power estimation in sequential circuits. In *Proceedings of the 32nd Design Automation Conference*, pages 635–640, 1995.

[4] A. A. Ghosh, S. Devadas, K. Keutzer, and J. White. Estimation of average switching activity in combinational and sequential circuits. In *Proceedings of the 29th Design Automation Conference*, pages 253–259, June 1992.

[5] F. N. Najm, R. Burch, P. Yang, and I. N. Hajj. Probabilistic simulation for reliability analysis of cmos circuits. IEEE *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 9(4):439–450, April 1990.

[6] F. N. Najm. Transition density: A new measure of activity in digital circuits. IEEE *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 12(2):310–323, February 1993.

[7] J. Monteiro, S. Devadas, B. Lin, C-Y. Tsui, M. Pedram, and A. M. Despain. Exact and approximate methods of switching activity estimation in sequential logic circuits. In *Proceedings of 1994 International Workshop on Low Power Design*, pages 117–122, April 1994.

[8] A. Hill, C.-C. Teng, and S. M. Kang. Simulation-based maximum power estimation. In *Proceeding of ISCAS-96*, 1996.

[9] S. Devadas, K. Keutzer, and J. White. Estimation of power dissipation in CMOS combinational circuits using Boolean function manipulation. IEEE *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11(3):373–383, March 1992.

[10] S. Chowdury and J. Barkatullah. Estimation of maximum currents in MOS IC logic circuits. IEEE *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 9(6):642–654, April 1990.

[11] H. Kriplani, F. N. Najm, and I. N. Hajj. Pattern independent maximum current estimation in power and ground buses of CMOS VLSI circuits: Algorithm, signal correlations, and their resolution. *Bull. Inter. Statist. Inst.*, 22:1–26, 1926.

[12] R. M. Loynes. Some aspects of the estimation of quantiles. *J. R. Statistic. Soc.*, B28:497–512, 1966.

[13] R. V. Hogg and A. T. Allen. *Introduction to Mathematical Statistics.* Prentice-Hall, 1995.

[14] R. L. Scheaffer, W. Mendenhall, and L .Ott. *Elementrary survey sampling.* PWS-KENT Publishing Company, 1995.

[15] J. H. van Lint and R. M. Wilson. *A course in Combinatorics.* Cambridge University Press, 1992.