Statistical Estimation of the Cumulative Distribution Function for Power Dissipation in VLSI Circuits

Chih-Shun Ding
Rockwell Semiconductor Systems
Newport Beach, CA 92660

Qing Wu  Cheng-Ta Hsieh  Massoud Pedram
Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089

ABSTRACT

This paper proposes to use quantile points of the cumulative distribution function for power consumption to provide detailed information about the power distribution in a circuit. The paper also presents two techniques based on population pruning and stratification to improve the efficiency of estimation. Both population pruning and stratification are based on a low cost predictor, such as zero-delay power estimate. Experimental results show the effectiveness of the proposed techniques in providing detailed power distribution information.

1 Introduction

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.

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 [1, 2, 3, 4, 5]. For estimating the maximum power consumption, existing techniques can be classified into two classes: statistical techniques and deterministic techniques. In statistical techniques [6], the maximum power consumption is estimated using order statistics derived from a simple random sample. In deterministic techniques, the maximum power is estimated either by solving the max-satisfiability problem [7] or by using approximation techniques [8, 7, 9] to obtain an 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 it may require large simulation times. The disadvantage of the deterministic techniques of [8, 9] 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 complex to model.

The cumulative distribution function of power consumption in a circuit is difficult to predict. This is because probability density functions of power dissipation of two different circuits can be very different in shape. For instance, researchers have found that the probability density 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. The following questions are frequently asked by today's chip designers:

Q1 What is the minimum power value \( x \) such that in \( y \) percentage of the time the circuit power dissipation is smaller than \( x \)?

Q2 What is the peak power dissipation in the circuit?

Q3 What percentage of time power dissipation in the circuit is between two known values \( x_1 \) and \( x_2 \)?

In this paper, we address the problem of estimating the quantile point in the cumulative distribution function for power dissipation (to answer Q1 and Q2 above) and deriving the cumulative distribution function itself (to answer Q3). 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.

For \( \alpha \in (0, 1) \), \( \alpha \)-quantile point of a cumulative distribution function is the value where the cumulative distribution function evaluates to \( \alpha \). 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 thus 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. 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. 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) \( F(x) \) 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. When \( F(x) \) is strictly increasing, the inverse function \( F^{-1}(x) \) is well defined. For \( \alpha \in (0, 1) \), the quantile point of \( F(x) \) 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 100 \( \alpha \)% percent of the population which have \( X \) values equal to or smaller than \( \alpha \).

Let \( X_1, X_2, \ldots, X_n \) be a sample of \( n \) observations taken from the population. \( n \) is referred to as the sample size. An estimator \( \hat{x} \) is a function of the random variable values on these \( n \) selected units which is used to estimate the parameters (such as the mean value and the quantile point, etc.) of the population. An estimator also is a random variable and may take different values from sample to sample. A confidence interval is an interval \([b, c]\) where the probability for the estimator value \( x \) to fall into it is \( \delta \), that is, \( \Pr(b \leq x \leq c) = \delta \), where \( \delta \) 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 as \( X_{(1)} < X_{(2)} < \cdots < X_{(n)} \), then they are defined as the order statistics of the sample. The \( i \)th element in this sorted list is referred to as the \( i \)th order statistic of the sample and is a random variable and its pdf \( g_0(X_{(i)}) \) is:
\[
g_0(X_{(i)}) = \frac{n!}{(k - 1)!(n - k)!}[F(y_k)]^{i-1}[1 - F(y_k)]^{n-i}f(y_k)
\]

In addition, if we define a new function \( Z = F(x) \), \( Z \) gives the quantile value associated with power value \( x \). The domain of \( Z \) is \([0, 1]\). Throughout this paper we will use \([\beta, \gamma]\) to denote an interval in the quantile domain.

3 Quantile Estimation Techniques

In this section, we address the problem of estimating an \( \alpha \)-quantile point \( F(x) \). A straightforward approach is to use the \( r \)th order statistic from a simple random sample of size \( n \) as the estimator \( \hat{x} \) for the \( \alpha \)-quantile point. It has been shown, given a fixed \( n \) value, the optimal \( r \) value can be approximated as [10]:
\[
r \approx \lceil n\alpha \rceil,
\]
(2)

Figure 1: Scatter plot of circuit C1355.

where \( \lceil \rceil \) denotes the ceiling function.

An empirical relation between \( n, \alpha \) and \( \epsilon \) for achieving 0.99 confidence level and 0.1 \( \leq \alpha \leq 0.9 \) is derived in [11] as:
\[
n \approx 6.671 \frac{1}{\epsilon} \alpha (1 - \alpha). \tag{3}
\]

Note that given a fixed \( \epsilon \) value, \( n \) is greatest when \( \alpha = 0.5 \) and smallest when \( \alpha \) is close to 0 or 1.

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 \([\beta, \gamma]\), \( b = F^{-1}(\beta) \), and \( c = F^{-1}(\gamma) \). 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 \( b \). \( \hat{U} = U - U_M - U_m \). We can derive a more efficient estimator using the order statistics of a sample that is drawn from \( \hat{U} \) as stated by 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 \([\beta, \gamma]\) be the confidence interval and \( b, c, U_m, U_M \), and \( U \) be defined as above. Let the new quantile of \( b \) and \( c \) on \( U \) be \( \beta' \) and \( \gamma' \), respectively, then
\[
\beta' = \frac{\beta |U| - |U_m|}{|\hat{U}|},
\gamma' = \frac{\gamma |U| - |U_m|}{|\hat{U}|},
\]
and
\[\gamma' - \beta' = \frac{|U|}{|\hat{U}|}(\gamma - \beta)\]

For the sake of space, all proofs in this paper are omitted (see [11] for proofs).

Since the new quantile interval has been increased by \([\beta, \gamma]\), a sample with fewer observations is now needed to achieve a given confidence level(cf. (3)). The required sample size to achieve 99% confidence level can be calculated by first computing the new \( \alpha \) and \( \epsilon \) values on \( \hat{U} \) and then plugging them into (3).

In practice, this “population pruning” procedure can be accomplished using a predictor with predictable error bounds. Let the confidence interval be \([\beta, \gamma]\), \( i = \lceil n \beta \rceil \), and \( j = \lceil n \gamma \rceil \), where \( \lfloor \rfloor \) is the flooring function. We sort the population based on predictor values. Let the predictor values on the \( i \)th and \( j \)th units be \( d_i \) and
In the next subsection, we will investigate the relevant issues.

residue values smaller than of of this may seem to be a drawback, it actually relaxes the requirement that of real-delay power, we need to only consider the impact of the estimation with short tails. Indeed this is the case when using zero-delay power. In a good regression model, the error bound. When the required accuracy level is higher, these two are incrementally changed from 1/101 to 100/101 with 1/101 increment. The vector sequence consists of 100 subclock cycles in which the predictor values are around $d_j$, or zero-delay power, falls in $[8.6 \text{mW}, 8.7 \text{mW}]$. The confidence level of the residue can be formally analyzed as in [12] if standard deviation of $e$ is known. When $\sigma$ is not available, our rule of thumb is to set the lower and upper pruning bounds for $e$ as $0.01$ to $0.6 d_j$ and $1.6 d_j$, respectively. This is equivalent to a [44\%, -15\%] error bound. When the required accuracy level is higher, these two numbers are set to be wider. While this may seem to be a very loose bound for Figure 2, we should note that units whose predictor values are around $d_i$, or $d_j$, have the largest effect on the estimation accuracy. The farther the predictor values are from from $d_i$, or $d_j$, the less likely those units will affect the estimation accuracy. We also found Figure 2 applies to multipliers and adders as well except that the curve is shifted to right due to higher glitch activity in these circuits.

3.2 Zero-delay Power as a Predictors

The difference between zero-delay power and real-delay power is due to glitches. In most practical circuits, except those circuits that are largely based on exclusive-or gates, the glitch power only accounts for an average of 20 to 30% of the real-delay power. It is therefore intuitive that the zero-delay power estimate is a good predictor.

To demonstrate this result more clearly, we show the scatter plot of real-delay power vs zero-delay power of a vector sequence of 40,000 cycles in Figure 1. The vector sequence consists of 100 subsequences in which the transition probabilities of each circuit input are incrementally changed from 1/101 to 100/101 with 1/101 increment. This will ensure a better range coverage of power dissipation as will be explained in the section of experimental results. The scatter plot shows an approximately linear correlation between the power estimates. That is, from the perspective of linear regression models, the relation between real-delay power, $X$, and the zerodecode power $P_{zd}$, is

$$X = kP_{zd} + e$$

where $k$ is the fitting constant and $e$ is the fitting error. $e$ is also referred to as the residue. Since $k$ does not change the order relation in the power dissipation, we need only consider the impact of $e$.

In a good regression model, $e$ will resemble a bell-shaped function with short tails. Indeed this is the case when using zero-delay power estimate as a predictor. In Figure 2, we plot the histogram of real-delay power of all clock cycles in which the predictor values, or zero-delay power, falls in $[8.6 \text{mW}, 8.7 \text{mW}]$. The confidence level of the residue can be formally analyzed as in [12] if standard deviation of $e$ is known. When $\sigma$ of $e$ is not available, our rule of thumb is to set the lower and upper pruning bounds for $e$ as $0.01$ to $0.6 d_j$ and $1.6 d_j$, respectively. This is equivalent to a [44\%, -15\%] error bound. When the required accuracy level is higher, these two numbers are set to be wider. While this may seem to be a very loose bound for Figure 2, we should note that units whose predictor values are around $d_i$, or $d_j$, have the largest effect on the estimation accuracy. The farther the predictor values are from from $d_i$, or $d_j$, the less likely those units will affect the estimation accuracy. We also found Figure 2 applies to multipliers and adders as well except that the curve is shifted to right due to higher glitch activity in these circuits.

Figure 2: Histogram of units with similar predictor values.

(d) The pdfs of the strata.

3.3 Stratified Sampling (STS)

Another technique to improve the efficiency is based on stratified sampling [13]. 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 [10], an estimator for quantile points based on stratified sampling of two strata is investigated, however, 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 for 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 $[x - \alpha, x + \alpha + \varepsilon]^n$, 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 units on the left-hand side of the $r$th 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 $r$th unit are likely to have $X$ values smaller than $x_\alpha = F(\alpha)$. Similarly, units on the right-hand side are likely to have $X$ values greater than $x_\alpha$. Therefore, the units that reside in $[x - \alpha, x + \alpha + \varepsilon]^n$ will be moved to either the upper quantiles of the first stratum or the lower quantiles of the second stratum. From (3), this minimizes the required sample size.

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, the “importance” (or weight) of the observations drawn from each stratum should reflect this difference. Therefore, all observations from the $i$th 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 $\alpha$, we stop and return the corresponding order statistic (which caused this) as the estimator.
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 \( (1.2, 3.4, 2.7), (0.7, 2.3, 1.4, 0.9, 1.6, 2.4, 1.5, 2.9) \). 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 \( \alpha \) 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.

[10] suggests that if \( n_i \) is allocated such that

\[
n_i \approx w_i \sqrt{F_i(x_i)(1 - F_i(x_i))},
\]

one could obtain the optimal estimator, where \( F_i(x) \) is the cdf of the \( i \)th stratum, and \( x_i = F_i^{-1}(\alpha_i) \). In practice, the difficulty of applying this criterion is that \( F_i(x_i) \) is not known in advance. However, as the population is partitioned at the \( \alpha_i \)-quantile point on the predictor domain, we expect that \( \sqrt{F_i(x_i)(1 - F_i(x_i))} \) is approximately the same for both strata. This reduces the above criterion to proportional allocation, i.e., \( n_1 = n w_1 \) 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.

### 3.4 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 improve the estimation efficiency. The relative efficiency of two sampling techniques, denoted by \( \eta \), 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 \( \eta \) 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 \( n \eta \). In addition, let the cost of zero delay simulation and power simulation for one clock cycle be \( C_{ps}, C_{pr} \), respectively. The population pruning (or stratified sampling) technique becomes more efficient than simple random sampling when:

\[
n C_{pr} > N C_{ps} + n C_{pr} / \eta,
\]

\[
\eta > \frac{1}{1 - \frac{N C_{ps}}{n C_{pr}}}
\]

When using PowerMill to perform power simulation, \( C_{ps} \approx 1/4000 \). If \( N = 40,000 \) and \( n = 4,000 \), \( \eta \) needs to be greater than 1.003.

### 4 Estimation of CDF \( F(x) \)

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 \( \alpha_i \)-quantile points that cover the domain \([0, 1]^2\), e.g. \( \alpha_i = i/n, i = 1, \ldots, n - 1 \). If the sample size \( n \) is adequately large, the empirical \( \hat{F}(x) \) will approach the true \( F(x) \). One may also construct \( \hat{F}(x) \) by estimating each quantile point separately. This is however very inefficient when the number of quantile points is large. The reason is that 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 \( \alpha_i \) values, \( i = 1, \ldots, n - 1 \). For every \( \alpha_i \)-quantile points, we want to use the \( i \)th and \( (i + 1) \)st order statistics as the confidence interval for \( x_i \) according to \( F^{-1}(\alpha_i) \). Therefore, we need to maximize the following probability:

\[
\text{Prob}(X_{(1)} < x_1 < X_{(2)} < x_2 \ldots < x_{n-1} < X_{(n)}) = \frac{n!}{(n-1)!}\frac{1}{n^n}.
\]

#### 4.1 Simple Random Sampling (SRS)

In the case of simple random sampling, we show that the maximum of (6) occurs when \( \alpha_i = i/n, i = 1, \ldots, n - 1 \), 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 \( \eta = F^{-1}((\alpha_i)) \). \( \text{Prob}(X_{(1)} < x_1 < X_{(2)} < x_2 \ldots < x_{n-1} < X_{(n)}) \) is maximized when \( \alpha_i = i/n \). The maximum value of this probability is \( n!/n^n \).

The above theorem implies that the order statistics can be used to estimate a set of \( \alpha_i \)-quantile points simultaneously and it is most efficient when \( \alpha_i = i/n \). In Figure 4, 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 \( [(i - 1)/n, i/n] \), we assume that it is at the midpoint of \( [(i - 1)/n, i/n] \), that is, \( \hat{F}(X_{(i)}) = i/n - 1/2n \).

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:

\[
\frac{n!}{n^n} \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \cdot \frac{1}{n^n} = \frac{\sqrt{2\pi n}}{e^n}.
\]

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 (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 \( r = \lfloor k/2 \rfloor + kr, k = 1, \ldots, n \), as shown in Figure 5.

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

\[\text{Figure 4: The approximation of cdf } f(x).\]

\[\text{Figure 5: The approximation of cdf } f(x).\]
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 column sum and row sum is 1.

Let the sample size and number of strata be \( n \). \( A \) is an \( n \times n \) matrix. Each entry \( a_{i,j} \) in \( A \) represents the portion of units in the \( i\)th stratum that are located in \( j\)th quantile interval. When drawing an observation from the \( i\)th stratum, the probability that this observation is from quantile interval \([j-1]/n, j/n]\) is \(a_{i,j}\). Therefore, matrix \( A \) has the following two properties: 1) all entries \( a_{i,j} \) are non-negative, and 2) all column sums and row sums are 1. The estimation is correct only when no two observations are from the same quantile interval. Therefore (6) calculates \( \per(A) \).

\[
\per(A) = \sum_{p \in \text{rows}, i=0}^{n} a_{i,p(i)} \tag{8}
\]

Let the sample size and number of strata be \( n \). \( A \) is an \( n \times n \) matrix. Each entry \( a_{i,j} \) in \( A \) represents the portion of units in the \( i\)th stratum that are located in \( j\)th quantile interval \([j-1]/n, j/n]\) in the original population. When drawing an observation from the \( i\)th stratum, the probability that this observation is from quantile interval \([j-1]/n, j/n]\) is \(a_{i,j}\). Therefore, matrix \( A \) has the following two properties: 1) all entries \( a_{i,j} \) are non-negative, and 2) all column sums and row sums are 1. The estimation is correct only when no two observations are from the same quantile interval. Therefore (6) calculates \( \per(A) \).

\[
\per(A) = \sum_{p \in \text{rows}, i=0}^{n} a_{i,p(i)} \tag{8}
\]

Let the sample size and number of strata be \( n \). \( A \) is an \( n \times n \) matrix. Each entry \( a_{i,j} \) in \( A \) represents the portion of units in the \( i\)th stratum that are located in \( j\)th quantile interval \([j-1]/n, j/n]\) in the original population. When drawing an observation from the \( i\)th stratum, the probability that this observation is from quantile interval \([j-1]/n, j/n]\) is \(a_{i,j}\). Therefore, matrix \( A \) has the following two properties: 1) all entries \( a_{i,j} \) are non-negative, and 2) all column sums and row sums are 1. The estimation is correct only when no two observations are from the same quantile interval. Therefore (6) calculates \( \per(A) \).

\[
\per(A) = \sum_{p \in \text{rows}, i=0}^{n} a_{i,p(i)} \tag{8}
\]
and $2d_j$, respectively (cf. Subsection 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 (3). The ‘$\eta$’ columns list the relative efficiency of population pruning over simple random sampling as defined in Subsection 3.4.

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 0.90 confidence level. Then we compare the $n$ values to get the relative efficiency. The results are summarized in Table 2 and Table 3. It shows that stratified sampling can further reduce the required sample size. The reason for $\eta$ to be smaller on random sequences is that the variances of the power dissipations on this 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 0.90. 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 4. 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 techniques: 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


