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

Chih-Shun Ding
Rockwell Semiconductor Systems
Rockwell International Corporation
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 statististical 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 $\alpha \in (0, 1)$, *α-quantile point* of a cumulative distribution function is the value where the cumulative distribution function evaluates 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 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.

*This research was supported in part by DARPA under contract no. F336125-95-C1627 and NSF under contract number no. MIP-9457392.

Design Automation Conference ®

0-89791-847-9/97/0006/$3.50 DAC 97 - 06/97 Anaheim, CA, USA

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. 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 $100 \times \alpha$ percent of the population which have X values equal to or smaller than α.

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* θ 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 is also 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 θ to fall into it 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 as $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, then 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 statistic. $X_{(k)}$, $k = 1, 2, \ldots, n$, is also a random variable and its pdf $g_k(X_{(k)})$ is:

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

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 $[\beta, \gamma]^q$ 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, given a fixed n value, the optimal r value can be approximated as [10]:

$$r \approx \lceil n\alpha \rceil, \quad (2)$$

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

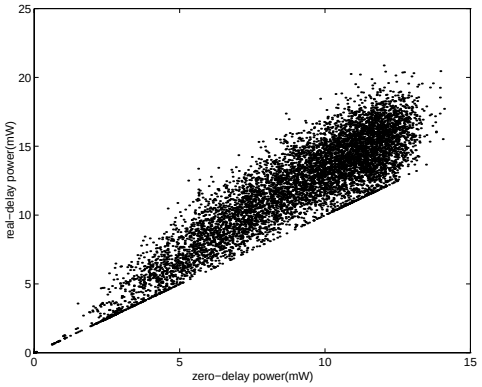


Figure 1: Scatter plot of circuit C1355.

where $\lceil\ \rceil$ denotes the ceiling function.

An empirical relation between n, ϵ and α for achieving 0.99 confidence level and $0.1 \le \alpha \le 0.9$ is derived in [11] as:

$$n \approx 6.67(\frac{1}{\epsilon})^2 \alpha(1-\alpha). \quad (3)$$

Note that given a fixed ϵ value, n is greatest when $\alpha = 0.5$ and smallest when α 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 $[\beta, \gamma]^q$, $b = F^{-1}(\beta)$, 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. $\hat{U} = U - U_M - Um$. 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 $[\beta, \gamma]^q$ 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

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

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 $\frac{|U|}{|\hat{U}|}$, 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 α and ϵ 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 $[\beta, \gamma]^q$, $i = \lfloor n\beta \rfloor$, 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 ith and jth units be d_i and

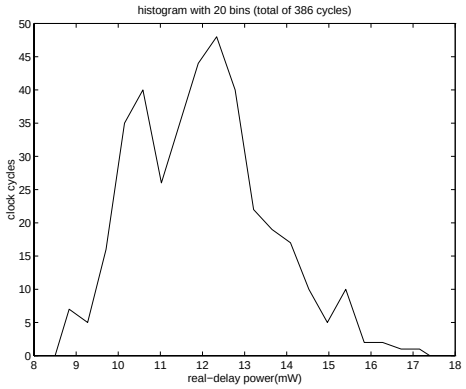


Figure 2: Histogram of units with similar predictor values.

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. However, as the estimation procedure is statistical in nature, we can never achieve 100% confidence level. While this may seem to be a drawback, it actually relaxes the requirement of of *absolutely* predictable error bounds to that of *statistically* predictable error bounds, which is a milder requirement.

We propose to use zero-delay power estimate as the predictor. In the next subsection, we will investigate the relevant issues.

3.2 Zero-delay Power as a Predictors

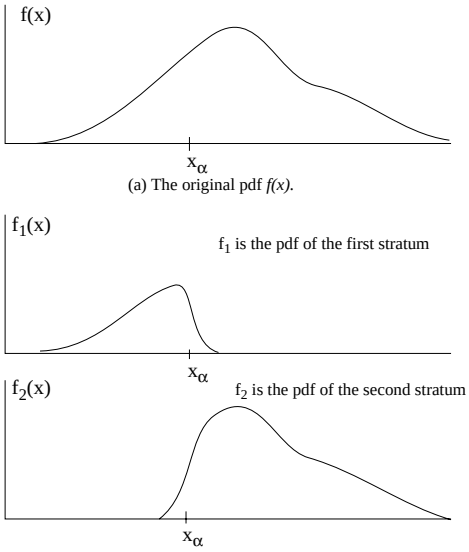
The difference between zero-delay power and real-delay power is due to glitches. In most practical circuits, except those circuits that are largely based on exclusive-or gates, the glitch power only accounts for an average of 20 to 30% of the real-delay power. It is therefore intuitive that the zero-delay 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 two power estimates. That is, from the prospective of linear regression models, the relation between real-delay power, X, and the zero-delay power P_{zd}, is

$$X = kP_{zd} + e \tag{4}$$

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 to 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 [8.6mW, 8.7mW]. The confidence level of the residue can be formally analyzed as in [12] if standard deviation σ of e is known. When σ of e is not available, our rule of thumb is to set the lower and upper pruning bounds for $\epsilon = 0.01$ to $0.6d_i$ and $1.6d_j$, 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.



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

(b) The pdfs of the strata.

Figure 3: An example of stratification.

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 $[\alpha - \epsilon, \alpha + \epsilon]^q$, 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 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, 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 (3), 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, the "importance" (or weight) of the observations drawn from each stratum should reflect this difference. Therefore, all observations from the 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 (which caused 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.

[10] suggests that if n_i is allocated such that

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

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, 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 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 becomes 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 $N = 40,000$ and $n = 4,000$, η 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 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 $F(x)$. 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.

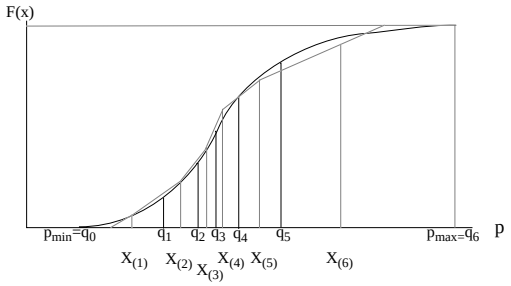


Figure 4: The approximation of cdf f(x).

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]:

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

4.1 Simple Random Sampling (SRS)

In the case of simple random sampling, we show that the maximum of (6) occurs when $\alpha_i = i/n$, 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 $n!/n^n$.*

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 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 $[(i-1)/n, i/n]^q$, 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:

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

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 (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 5.

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

[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.

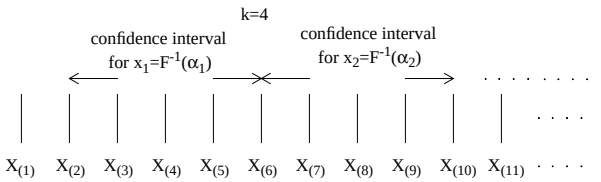


Figure 5: Selecting order statistics for quantile bound.

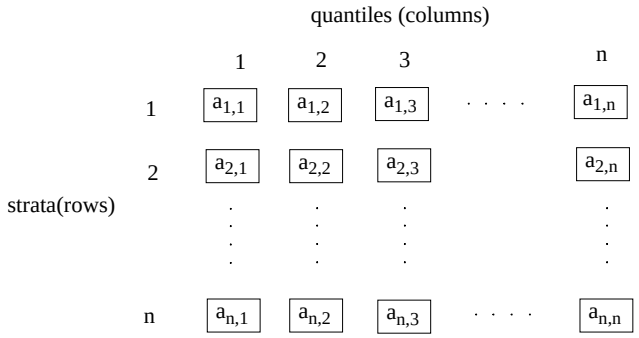


Figure 6: Stratified random sampling for distribution estimation.

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 (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 $a_{i,.}$. Similarly, the sum of all entries in a particular column is referred to as $columnsum$ and denoted by $a_{.i}$. 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 (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,*

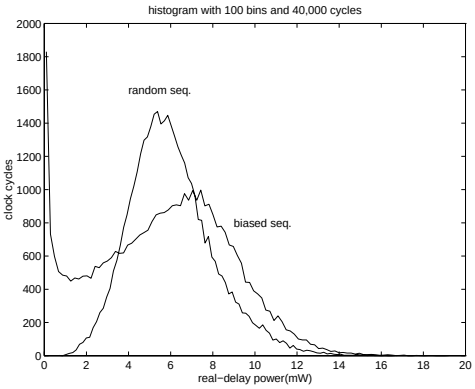


Figure 7: Typical power histogram of biased and random sequences.

then

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

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 main objective of the experiments presented in the following is to compare the relative efficiency (in terms of required sample size) of the proposed techniques vs. simple random sampling.

The circuits are mapped to a library with NAND, NOR, inverter and XOR gates. We test on two types of populations of 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 (cf. Subsection 3.2), and will be referred to as the **biased sequences**. Typical distributions of these two types of sequences are as shown in Figure 7. The distribution under the random sequence has a narrower distribution as, under the random condition, only a small range of input switching is exercised. For instance, for a 32-input circuit, 95% of the vectors has input switching in the range of [11,21]. As a result, the power dissipation is more homogeneous.

In the first set of experiments, we apply the proposed techniques to estimate α-quantile points, where $\alpha \in \{0.1, 0.5, 0.9\}$, with absolute error and confidence levels of 0.01, and 0.99, respectively. More experimental results can be found in [11]. 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 1 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 $0.6d_i$

Table 1: Results of population pruning on the biased sequences, 1000 experiments.

	α								
	0.1			0.5			0.9		
ckts	n	err	η	n	err	η	n	err	η
C432	228	0.5	26.0	10222	0.8	1.6	4023	0.5	1.5
C880	213	0.8	27.9	7784	1.7	2.1	3064	0.9	1.9
C1355	707	0.5	8.4	10462	1.6	1.6	3996	0.6	1.5
C1908	412	1.3	14.4	9884	1.0	1.6	3608	1.0	1.7
C2670	273	0.7	21.8	8963	0.7	1.8	3550	0.8	1.7
C3540	179	1.2	33.3	9180	0.9	1.8	3627	1.4	1.6
C5315	328	1.0	18.1	9453	1.4	1.7	3801	0.8	1.6
C6288	793	1.2	7.5	11858	0.9	1.4	4421	0.9	1.3
C7552	302	1.2	19.7	8867	0.9	1.8	4013	0.8	1.5
avg			19.7			1.7			1.6

Table 2: Results of combining population pruning and stratified sampling on random sequences, 1000 experiments.

	α								
	0.1			0.5			0.9		
ckts	n	err	η	n	err	η	n	err	η
C432	4700	0.7	1.3	12500	1.8	1.3	4400	2.2	1.3
C880	4100	0.3	1.5	12300	1.0	1.3	4200	2.2	1.3
C1355	4600	1.7	1.3	12800	1.2	1.3	4600	0.9	1.3
C1908	4600	1.1	1.3	12700	1.6	1.3	4600	2.5	1.3
C2670	4700	0.3	1.3	12800	0.9	1.3	4700	0.7	1.3
C3540	4600	1.1	1.3	12700	0.8	1.3	4600	1.4	1.3
C5315	4700	1.4	1.3	12800	1.2	1.3	4700	1.2	1.3
C6288	4700	0.5	1.3	12800	0.8	1.3	4700	1.0	1.3
C7552	4700	0.1	1.3	12800	0.2	1.3	4700	1.4	1.3
avg			1.3			1.3			1.3

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 'η' 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 0.99 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 η 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 0.99. 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 4. Again the improvement of STS on biased sequences is better than that on random sequences.

Table 3: Results of combining population pruning and stratified sampling on biased sequences, 1000 experiments.

	α								
	0.1			0.5			0.9		
ckts	n	err	η	n	err	η	n	err	η
C432	100	0.3	60	7500	0.9	2.2	4000	1.0	1.5
C880	100	0.1	60	4500	0.4	3.6	2800	0.6	2.1
C1355	400	0.5	15	7500	0.3	2.2	3500	0.7	1.7
C1908	250	0.9	24	7200	1.0	2.3	3600	1.2	1.7
C2670	100	0.4	60	4000	0.2	4.2	2400	1.0	2.5
C3540	100	0.6	60	5000	1.0	3.3	3300	0.8	1.8
C5315	150	1.0	40	4000	0.1	4.2	3200	0.8	1.9
C6288	450	0.8	13	6000	0.6	2.8	3600	1.2	1.7
C7552	100	0.8	60	4000	0.4	4.2	3500	1.1	1.7
avg			44			3.2			1.8

Table 4: 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

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, 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

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

[11] C.-S. Ding, Q. Wu, and M. Pedram. Statistical estimation of distribution of power dissipation in vlsi circuits. Technical Report CENG 97-09, Electrical Engineering-Systems Department, University of Southern California, March 1997.

[12] Q. Wu, C.-S. Ding, C.-T. Hsieh, and M. Pedram. Statistical design of macromodels for rt-level power evaluation. In *Proc. of ASP-DAC97*, January 1997.

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