Estimation of Peak Power Dissipation in VLSI Circuits Using the Limiting Distributions of Extreme Order Statistics

(Control Number: 889, Associated Editor: Enrico Macii)

Qing Wu, Qinru Qiu, Massoud Pedram

Department of EE-Systems
Univ. of Southern California
Los Angeles, CA 90089
Email: {qwu, qinru, Pedram}@usc.edu

Abstract

In this paper, we present a statistical method for estimating the peak power dissipation in VLSI circuits. The method is based on the theory of extreme order statistics applied to the probabilistic distributions of the cycle-by-cycle power consumption, the maximum likelihood estimation, and the Monte-Carlo simulation. It enables us to predict the maximum power in the space of constrained input vector pairs as well as the complete space of all possible input vector pairs. The simulation-based nature of the proposed method allows us to avoid the limitations of a gate-level delay model and a gate-level circuit structure. Last, but not least, the proposed method produces maximum power estimates to satisfy user-specified error and confidence levels. Experimental results show that this method typically produces maximum power estimates within 5% of the actual value and with a 90% confidence level by simulating less than 2500 input vectors.

I. INTRODUCTION

Circuit reliability is an important issue in today's VLSI design. There are many sources that may cause circuit failure; one of them is large power dissipation over a short period of time. High current dissipation in a short amount of time may cause excessive heat generation resulting in permanent circuit damage or give rise to unwanted voltage change on the power supply lines resulting in temporary circuit failure. To circumvent these problems, designers have to have accurate estimates of maximum power dissipation in VLSI circuits. Estimation of maximum power in VLSI circuits is therefore essential for determining the appropriate packaging and cooling techniques or optimizing the power and ground routing networks.

In most of the previous research work, maximum power estimation refers to the problem of estimating the maximum power (or current) that the circuit may consume within any clock cycle. The problem is thus equivalent to searching for the maximum power-consuming vector pair among all possible input vector pairs. Therefore, these techniques focus on finding lower and upper bounds on the maximum power dissipation in a VLSI circuit. However, design requirements of today's complex VLSI chips make things more complicated. In general, we divide the scope of the maximum power estimation problem into two categories:

I.1. The maximum power for all possible vector pairs applied to the inputs of the circuit. We refer to this quantity as the *unconstrained maximum power*.

I.2. The maximum power for an input sequence with given average switching activity. We refer to this quantity as the *constrained maximum power*.

A number of techniques have been developed to solve the problem in Category I.1 [1]-[9]. A method using Boolean function manipulation was proposed in [1]. The problem of finding the input-pair that maximizes the weighted activity was transformed to a weighted max-satisfiability problem that can be solved exactly (albeit with exponential complexity) or approximately. Even an approximate solution of the satisfiability problem is time-consuming when the number of primary inputs is large or a complex delay model is used for the analysis.

The authors of [2] extended the method in [1] to compute the maximum power cycles of a finite state machine by modeling the problem as a calculation of maximum average length cycles of a weighted directed graph. Symbolic methods based on algebraic decision diagrams were used to compute the maximum average length cycles and the number of gate transitions in the circuit, which are needed to construct the weighted directed graph. This method is time consuming when the number of primary inputs and/or states is large.

The method proposed in [3] propagates signal uncertainty waveforms throughout the circuit to obtain a loose upper bound on the maximum power. The bound is then made tighter by computing the signal correlations in the circuit. However, the bound tightening method tends to be time consuming when the number of the primary inputs is large.

Automatic test pattern generation (ATPG) based techniques [4]-[5] try to generate an input vector pair that produces the largest switched capacitance in the circuit. The power consumption by the vector pair is then used as a lower bound on the maximum power of the circuit. ATPG-based techniques are very efficient and generate a tighter lower bound than that generated by random vector generation. However, the limitations of ATPG-based techniques are that they handle simple delay models such as the zero-delay and unit-delay models and that the analysis is done at the gate-level instead of the circuit level. Consequently, the estimation accuracy has yet to be improved.

Statistical methods have also been used for maximum power estimation. In [4] a Monte Carlo based statistical technique for maximum current estimation was briefly discussed. The method randomly generates high-activity vector pairs, and the maximum power is then estimated by the simulation. This method also suffers from low efficiency.

A continuous optimization method was proposed in [6], which treats the input vector space as a continuous real-valued vector space and then performs a gradient search to find the function maximum. The method uses the unit delay model during the function transformation. Therefore, similar to the ATPG-based techniques, the estimation accuracy is not high.

The authors of [7] proposed a technique for finding the maximum power-consuming vector using a genetic search algorithm. The authors start with a set of randomly selected vector pairs that generate large amount of power dissipation. By mixing up the highest power consuming vectors, they then generate the next population and keep on iterating until some maximum iteration count has been exceeded. The advantage of the method is that it is a simulation-based approach where the delay models do not limit the estimation accuracy. The shortcoming of this approach is that it has low efficiency, i.e., requires simulation of a large number of vectors.

The theory of order statistics has been applied in [8][9] to estimate the maximum power dissipation in a circuit by estimating a high quantile point (e.g., the 99.9% quantile point). These techniques can be very accurate because they are simulation-based techniques, but their efficiency is not much better than the random vector generation technique.

In this paper, we present a simulation-based statistical method for maximum power estimation for combinational circuits. It is a method of estimating the maximum power using the asymptotic theory of *Extreme Order Statistics*. Compared to previous work, our approach makes the following contributions:

1. Our approach is the first to provide the confidence interval for the estimated maximum power given a user-specified confidence level.

2. Our approach is the first to do maximum power estimation for any given error and confidence level.

3. Our approach can estimate the maximum power defined in both categories I.1 and I.2.

4. Because our approach is simulative, the delay model or the circuit structure does not limit its accuracy.

5. Due to statistical estimation of the extreme value distributions, the estimation efficiency of our approach is much higher than that of other existing statistical methods (including simple random sampling and quantile estimation).

On average, our proposed method can perform maximum power estimation by simulating only about 2500 vector pairs to achieve a 5% error at a confidence level of 90%. It may be of interest to the reader to note that our statistical sampling technique, which builds on the limiting distributions of extreme order statistics, can be equally useful in estimating the peak value of any other distribution. For example, one can use it to estimate the maximum delay through a circuit under process variations, temperature variations and different input populations. However, this is outside the scope of this paper. A shorter version of this paper was published in [18].

This paper is organized as follows. Section II introduces the asymptotic theory of extreme order statistics and maximum likelihood estimation. Section III describes our approaches for maximum power estimation. Section IV presents our experimental results, and Section V gives the concluding remarks.

II. BACKGROUND

In this section, we introduce some key concepts and theorems from the asymptotic theory of extreme order statistics and maximum-likelihood estimation, which will be useful for our purpose of estimating the maximum power dissipation of a VLSI circuit.

2.1 The asymptotic theory of extreme order statistics

The (cumulative) distribution function (d.f.) of a random variable (r.v.) x is defined as:

$$F(t) = P\{x \le t\} \tag{2.1}$$

The density function of x is:

$$f(x) = F'(x) \text{ if } F(x) \text{ is continuous and differentiable.}$$

The quantile function (q.f.) of a d.f. F is defined as:

$$F^{-1}(q) = \inf\{t : F(t) \ge q\}, \quad q \underset{random}{\in} [0,1] \tag{2.2}$$

where $\inf(S)$ calculates the lower bound of set S. Notice that the q.f. F^{-1} is a real-valued function, and $F^{-1}(q)$ is the smallest q quantile of F. That is, if Z is a r.v. with d.f. F, then $F^{-1}(q)$ is the smallest value t such that $P\{Z < t\} \le q \le P\{Z \le t\}$. We remark that $F(x)=\sup\{q \in [0,1]: F^{-1}(q) \le x\}$. Let $z_1, z_2, ..., z_n$ be n random units drawn from a common distribution. When they are drawn in a random manner, they are called independent identically distributed (i.i.d.) r.v.'s. If we are interested not in the order in which $z_1, z_2,...,z_n$ are drawn, but in the order of the magnitude of their values, then we have to examine the ordered sample values

$$X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$$

which are the *order statistics* of a sample of size n.

$X_{r:n}$ is called the rth order statistic, and the random vector $(X_{1:n}, X_{2:n}, ..., X_{n:n})$ is the order statistic. Note that $X_{1:n}$ is the sample minimum, and $X_{n:n}$ is the sample maximum. $X_{1:n}$ is called the *minima order statistic*,

and $X_{n:n}$ is called the *maxima order statistic*, or in general, they are called the *extreme order statistics* of a sample of size n.

Let $X_{r:n}$ be the rth order statistic of n i.i.d. random variables $z_1, z_2, ..., z_n$ with common d.f. F and density f. Then its d.f. is given by:

$$P\{X_{r:n} \leq t\} = \sum_{i=r}^{n} \binom{n}{r} F(t)^i (1-F(t))^{n-i} \tag{2.3}$$

and the density of $X_{r:n}$ is given by:

$$f_{r:n} = n! f \frac{F^{r-1}(1-F)^{n-r}}{(r-1)!(n-r)!} \tag{2.4}$$

In the special case where we are only interested in the sample maxima $X_{n:n}$, we obtain the d.f. as:

$$P\{X_{n:n} \leq t\} = F(t)^n \tag{2.5}$$

and the density of $X_{n:n}$ is:

$$f_{n:n} = nf F^{n-1} \tag{2.6}$$

Here we give three distribution types [12] that are used for studying the limiting d.f. of sample maxima (in other words, extreme value d.f.'s):

$$G_{1,\alpha}(x) = \begin{cases} 0 \\ \exp(-x^{-\alpha}) \end{cases} \quad \text{if} \quad \begin{matrix} x \leq 0 \\ x > 0 \end{matrix}, \quad \text{``Fréchet''} \tag{2.7}$$

$$G_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^{\alpha}) \\ 1 \end{cases} \quad \text{if} \quad \begin{matrix} x \leq 0 \\ x > 0 \end{matrix}, \quad \text{``Weibull''} \tag{2.8}$$

$$G_3(x) = \exp(-e^{-x}) \quad \text{for every } x, \quad \text{``Gumbel''} \tag{2.9}$$

where $\alpha > 0$ is a shape parameter. It will be convenient to write $G_{3,\alpha}$ in place of G_3 where α is understood to be always 1. Example plots of density function of these three d.f.'s are shown in Figure 1. We say that two d.f.'s G_1 and G_2 are of the same type (isomorphic under an affine transformation) if $G_1(b+ax)=G_2(x)$ for some $a>0$ and real b. The following identities show that the d.f.'s $G_{i,\alpha}$ are in fact the limiting d.f.'s of sample maxima. We have:

$$G_{1,\alpha}^n(n^{1/\alpha}x) = G_{1,\alpha}(x), \qquad G_{2,\alpha}^n(n^{-1/\alpha}x) = G_{2,\alpha}(x), \qquad G_3^n(x+\log n) = G_3(x)$$

that is, these d.f.'s are max-stable in the above sense. It is one of the notable achievements of the classical extreme value theory that one can show that the $G_{1,\alpha}, G_{2,\alpha}, G_3$ are all of the possible limiting d.f.'s of sample maxima [11].
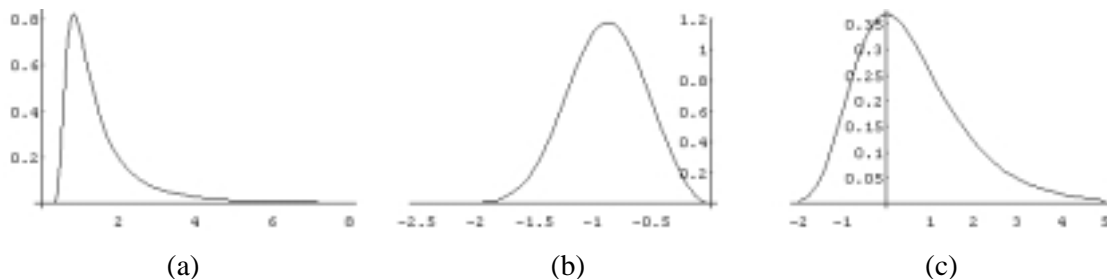

(a)   (b)   (c)

Figure 1 Plots of density functions

(a) Density function for Fréchet d.f. when $\alpha = 3$
(b) Density function for Weibull d.f. when $\alpha = 3$
(c) Density function for Gumbel d.f.

Definition 1 [11] F is said to belong to the weak domain of attraction of limiting d.f. G, if there exist constants $a_n > 0$ and reals b_n such that:

$$F^n(b_n + xa_n) \rightarrow G(x), \qquad n \rightarrow \infty \tag{2.10}$$

for every continuity point of G.

Let us define the *right endpoint* of d.f. F as:

$$\omega(F) = \sup\{x : F(x) < 1\} = F^{-1}(1) \tag{2.11}$$

Theorem 1 [11] A d.f. F belongs to the weak domain of attraction of an extreme value d.f. $G_{i,\alpha}$ if and only if one of the following conditions holds:

$$(1,\alpha): \ \omega(F) = \infty \quad \text{and} \quad \lim_{t\to\infty}\left[\frac{1-F(tx)}{1-F(t)}\right] = x^{-\alpha}, \quad x > 0 \tag{2.12}$$

$$(2,\alpha): \ \omega(F) < \infty \quad \text{and} \quad \lim_{t\downarrow 0}\left[\frac{1-F(\omega(F)+xt)}{1-F(\omega(F)-t)}\right] = (-x)^\alpha, \quad x < 0 \tag{2.13}$$

$$(3): \ \lim_{t\uparrow\omega(F)}\left[\frac{1-F(t+xg(t))}{1-F(t)}\right] = e^{-x}, \quad -\infty < x < \infty \tag{2.14}$$

where

$$g(t) = \frac{\int_t^{\omega(F)}(1-F(y))dy}{1-F(t)} \tag{2.15}$$

and $\downarrow$ represents approaching decreasingly and $\uparrow$ represents approaching increasingly.

Moreover, the constants b_n and a_n can be chosen in the following way:

$$(1,\alpha): \qquad b_n^* = 0, \qquad\qquad a_n^* = F^{-1}(1-1/n) \tag{2.16}$$

$$(2,\alpha): \qquad b_n^* = \omega(F), \qquad a_n^* = \omega(F) - F^{-1}(1-1/n) \tag{2.17}$$

$$(3): \qquad b_n^* = F^{-1}(1-1/n), \qquad a_n^* = g(b_n^*). \tag{2.18}$$

If a distribution F satisfies one of the conditions in Theorem 1, we simply call the corresponding $G_{i,\alpha}$ the asymptotic distribution of the sample maxima of distribution F. Theorem 1 gives the conditions describing to which d.f. G, the extreme distribution, will converge. In addition, it provides us with the guidelines to choose the correct asymptotic extreme distribution for a specific application.

Example [11]: We want to study the limiting distribution for the sample maxima of the common Uniform Distribution function $F(x)$ that is defined as follows:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$$

Since $\omega(F)=1$, according to Theorem 1, we exclude the Fréchet distribution.

Now let us check the condition in (2.14). Since

$$\frac{1-(t+xg(t))}{1-t} = 1 - x\frac{g(t)}{1-t}$$

cannot tend to e^{-x} as $t\to 1$ [11], (2.14) fails and the limiting distribution cannot be the Gumbel distribution.

Therefore the only possible distribution is the Weibull distribution if (2.13) holds. Let us define, for $x>0$,

$$F^*(x) = F(1 - \frac{1}{x}) = 1 - \frac{1}{x}, \qquad x > 0$$

We find that, for $x > 0$, as $t \rightarrow +\infty$,

$$\lim \frac{1 - F^*(tx)}{1 - F^*(t)} = \lim \frac{t}{tx} = x^{-1}$$

Therefore condition (2.13) holds and the limiting distribution for sample maxima of the [0,1] uniform distribution exists and follows the Weibull distribution defined in (2.8) with $\alpha = 1$.

Theorem 2 [11] The weak convergence to the limiting d.f. G holds for other choices of constants a_n and b_n if and only if

$$a_n / a_n^* \rightarrow 1 \quad \text{and} \quad (b_n - b_n^*)/a_n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty \qquad (2.19)$$

Theorem 2 gives other possibilities of choosing a_n and b_n in Theorem 1. In special cases when $F(x)$ has a finite right endpoint, by Theorem 2, the choice of b_n ($n \rightarrow \infty$) in Eqn.(2.17) is unique. This observation is very important for our discussion in Section III.

Theorem 3 [11] Let $G = G_{i,\alpha}$ for some i and $\alpha > 0$. Then the following two statements are equivalent:

(1) $F^n(b_n + xa_n) \rightarrow G(x), \qquad n \rightarrow \infty, \text{ for every } x$ $\qquad (2.20)$

(2) $(1 - F(b_n + xa_n))/(1 - G(d_n + xc_n)) \rightarrow 1, \qquad n \rightarrow \infty$ $\qquad (2.21)$

where (2.21) has to hold for every x in the support domain of G. Moreover,

$$d_n = \begin{cases} 0 \\ \log n \end{cases} \quad \text{if} \quad \begin{array}{l} i = 1,2 \\ i = 3 \end{array} \qquad \text{and} \qquad c_n = \begin{cases} n^{1/\alpha} & i = 1 \\ n^{-1/\alpha} & \text{if} \quad i = 2 \\ 1 & i = 3 \end{cases} \qquad (2.22)$$

Theorem 3 provides a more intuitive and practical way of studying the conditions in Theorem 1.

2.2 Maximum-likelihood estimation for parameters of the Weibull distribution when $\alpha > 2$

We use a maximum-likelihood estimator for parameters of a generalized Weibull distribution defined as:

$$G(x; \alpha, \beta, \mu) = \begin{cases} \exp(-\beta(\mu - x)^{-\alpha}) & \text{if} \quad \begin{array}{l} x \leq \mu \\ x > \mu \end{array} \\ 0 \end{cases} \qquad (2.23)$$

where μ is a location parameter which determines the right endpoint (i.e. maximum) of the distribution, $\beta > 0$ is a scale parameter, and α is the shape parameter. We only consider the case of $\alpha > 2$ because it holds for our application.

The maximum-likelihood estimation problem is defined as follows: Given m independent random samples $x_1, x_2, \ldots, x_m$ of $G(x; \alpha, \beta, \mu)$, find the values of α, β, μ which maximize the likelihood function [10]:

$$L_m(\alpha, \beta, \mu) = \frac{1}{m} \sum_{i=1}^{m} \log[G(x_i; \alpha, \beta, \mu)] \qquad (2.24)$$

This maximum likelihood estimator, when it exists, will be denoted by the vector $(\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m)$ and satisfies

$$\frac{\partial L_m}{\partial \alpha}(\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m) = 0, \frac{\partial L_m}{\partial \beta}(\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m) = 0, \frac{\partial L_m}{\partial \mu}(\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m) = 0 \qquad (2.25)$$

Let α_0, β_0, μ_0 denote the actual values of parameters of the distribution, G. It was proved in [10] that, when $\alpha > 2$, $m^{\frac{1}{2}}(\hat{\alpha}_m - \alpha_0, \hat{\beta}_m - \beta_0, \hat{\mu}_m - \mu_0)$ converges in distribution ($m \to \infty$) to a normal random distribution vector with mean 0 and an appropriately defined covariance matrix **VAR**. We will show how to calculate matrix **VAR** in Section 3.3.

III. THE ESTIMATION APPROACH

The problem of maximum power estimation can be stated as follows: Given a set *V* (called *population*) of input vector pairs, estimate the maximum power dissipation of the circuit under any vector pair in the population. A vector pair in *V* is called a *unit* of the population. In this paper, the population may include all possible input vector pairs applied to a circuit, or all possible vector pairs with some fixed input transition probability. Although there can only be a finite number of distinct vector pairs in the population, the size of *V*, represented by |*V*|, is assumed to be infinite since there is the possibility of repeating the vector pairs.

In this section we explain assumptions, justification, and details of our approach for maximum power estimation.

3.1 The distribution of power consumption

If we regard power consumption for a vector pair as a random variable *p*, then a distribution of *p* is formed by the power consumption values of vector pairs in set *V*. Figure 2 shows an example of power distribution function $F(p)$ and density function $f(p)$. The average power is the mean value of the distribution. The maximum power is then the right endpoint of the distribution.

Strictly speaking, the distribution function of power is different from ordinary probability distribution functions because the support domain of the power d.f. is not continuous. Practically, however, we, like most other researchers [4][8][9], make the assumption that d.f. of power in an LSI circuit is approximately continuous when the size of circuit and the size of population is large. Therefore all theories and approaches related to probabilistic distribution can be applied to the problems related to power distribution. We also assume the d.f. of power consumption in a VLSI circuit has a continuous d.f. in this paper as stated below.

Assumption 1 The d.f. of power consumption $F(p)$ is continuous and differentiable at every point *p* in its supporting domain.
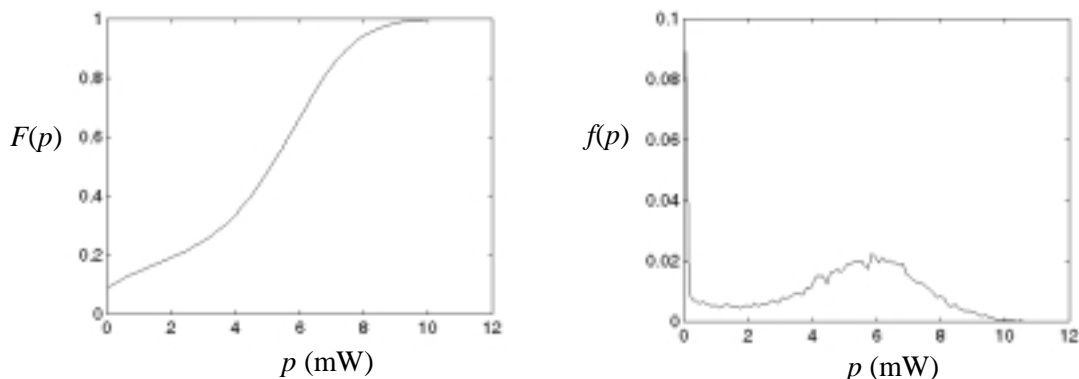

Figure 2 An example of power d.f. and density function

Consider a simple two-input NAND gate; there are at most 16 possible power consumption values. Therefore, we cannot and do not treat the power dissipation distribution of a simple NAND as a continuous distribution. However, for large circuits that consist of thousands or tens of thousands of such

simple gates, the total power distribution of the circuit can be safely modeled as a continuous function. It is, of course, possible to envision and construct circuits where the total power distribution is not a continuous function. However for most circuits, the continuity assumption holds. Note that a bimodal or multi-modal power distribution can be and often are continuous.

This assumption is indeed conservative. However, our approach is not limited only to these "nearly" continuous distributions because our interest is only in the tail of the distribution. In some situations where the power distribution is not continuous, such as a distribution with multiple abrupt modes, as long as we have samples in the mode with the highest power values, our method will still be applicable. In the extreme cases such as when there is one unit in the population that consumes much higher power than any other unit, then our method will fail. Note however that in such a case, any other method, short of an exhaustive simulation method or an intractable analytical model with a simplistic delay equation, will fail.

3.2 The asymptotic distribution of the sample maximum power

Given the population V, the ith sample for maximum power estimation is formed by the power values of n randomly selected units:

$$p_{i,1}, p_{i,2}, \cdots, p_{i,n} \qquad i = 1,2,\cdots,m$$

where n is called the sample size and m the number of samples. The maximum power in each sample is defined as:

$$p_{i,MAX} = \max\{p_{i,1}, p_{i,2}, \cdots, p_{i,n}\} \qquad i = 1,2,\cdots,m \qquad (3.1)$$

According to Eqn.(2.5), the d.f. of $p_{i,MAX}$ can be written as: $H(p_{i,MAX}) = F^n(p)$.

Now let us try to determine the asymptotic distribution of $p_{i,MAX}$ when $n \to \infty$. As mentioned in Theorem 1, $H(b_n + p_{i,MAX} \cdot a_n)$ converges to one of the three distributions defined in Eqn.'s(2.7), (2.8), and (2.9).

In the remainder of this paper, we will use $\omega(F)$ to denote the actual maximum power of the population.

Because $F(p)$ cannot be written, we cannot directly prove that the asymptotic distribution of $H(p_{i,MAX})$ exists by using the conditions in Theorem 1 or Theorem 3. What we have are the following observations which help us determine the asymptotic distribution of $H(p_{i,MAX})$.

First we know that the power consumption in a LSI circuit is always a finite value, i.e., $\omega(F) < \infty$. Therefore the condition in (2.12) is not met, and $H(b_n + p_{i,MAX} \cdot a_n)$ cannot converge to $G_{1,\alpha}$.

Second, because the upper bound of the supporting domain for G_3 is infinite while that of $G_{2,\alpha}$ is finite, the probability that the condition in Eqn.(2.21) holds for G_3 is lower than that for $G_{2,\alpha}$. Therefore $H(b_n + p_{i,MAX} \cdot a_n)$ is more likely to converge to $G_{2,\alpha}$ rather than G_3. Moreover, we can obtain the maximum power value directly from the parameter of $G_{2,\alpha}$, which simplifies the maximum power estimation problem.

Finally, it is pointed out in [11] that most frequently sed continuous distributions with a finite right endpoint $(\omega(F) < \infty)$ satisfy the condition in Eqn.(2.13). Therefore, in many engineering applications of maxima or minima estimation, it is assumed that the distribution under study belongs to the weak convergence domain of $G_{2,\alpha}$ [12]. Below we have included some examples of these works.

Indeed the Gumbel distribution could have been used as well. However, the exact choice of the limiting distribution is not critical to us because our contribution in this paper is that we design a Monte Carlo statistical simulation technique to do maximum power estimation based on our choice of the limiting distribution of the sample maxima. It will be rather straightforward to develop a similar Monte Carlo based simulation procedure starting with a Gumbel distribution instead.

Example:

1. In [13], the authors used the Weibull distribution to study the distribution of the maximum yield strength of high-tensile steel.

2. In [14] and [15], the authors used both Weibull and Gumble distributions to study the distribution of the maximum wind speed.

Therefore, the second and the last assumption of this paper is that the distribution of $p_{i,MAX}$ asymptotically follows the Weibull distribution $G_{2,\alpha}$. This means that there exist a_n and b_n such that:

$$F(b_n + a_n \cdot p_{i,MAX}) = F^n(b_n + a_n p) \to G_{2,\alpha}(p_{i,MAX}), \qquad n \to \infty, \quad i = 1,2,\cdots,m \qquad (3.2)$$

or
$$F(p_{i,MAX}) \to G_{2,\alpha}(\frac{p_{i,MAX} - b_n}{a_n}), \qquad n \to \infty, \qquad i = 1,2,\cdots,m \qquad (3.3)$$

From Eqn.'s (2.17) and (2.19), we get $b_n = \omega(F)$ where $\omega(F)$ is the maximum power consumption of the population. If we substitute the generalized Weibull distribution defined in Eqn.(2.23) into Eqn.(3.3), we get

$$F(p_{i,MAX}) \to G(p_{i,MAX}; \alpha, \beta, \mu), \qquad n \to \infty, \qquad i = 1,2,\cdots,m$$

where $\beta = (1/a_n)^{\alpha}$ and $\mu = b_n$.

Experiments have been designed to verify the asymptotic distribution of sample maximum. The distributions of sample maxima for six different sample sizes ($n = 2, 10, 20, 30, 40, 50$) were formed by 1,000 random samples ($1000 \times n$ units) from the original population. The closest Weibull distributions were obtained by using a least-square curve fitting technique. Figure 3 shows the results for circuit C3540.
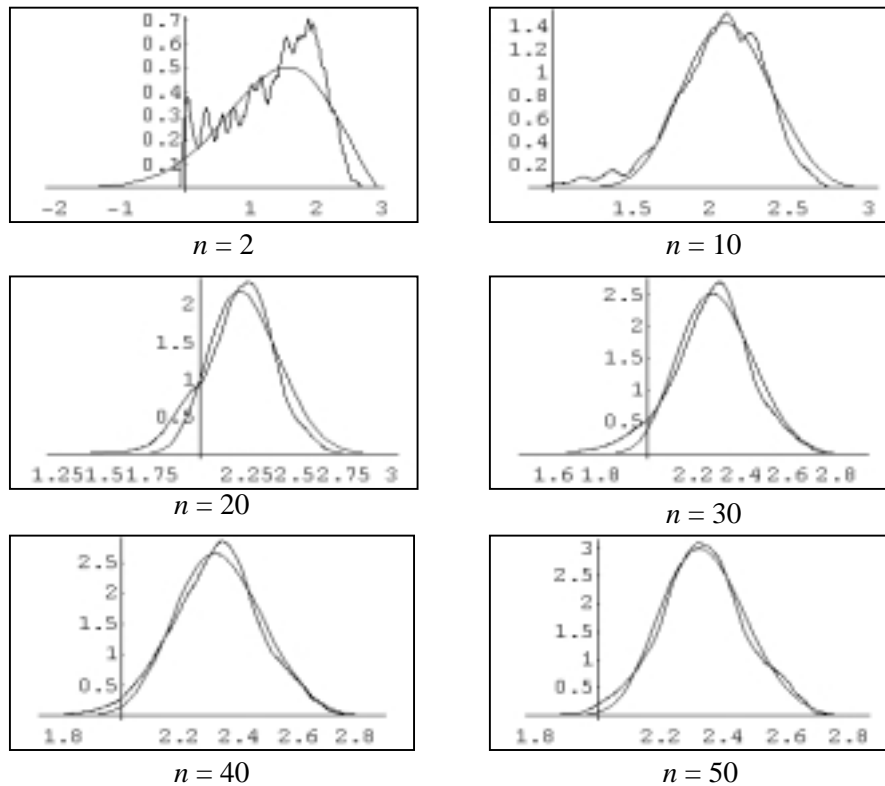

Figure 3 The comparison between distribution of sample maxima and Weibull distribution

We have repeated this experiment for other circuits and populations and obtained a similar result. From these results we conclude that the difference between distributions of $p_{i,MAX}$ and the Weibull distribution in the region near the maximum power is negligible when n is larger than or equal to some fixed value, e.g., 30. Since we are only interested in estimating the maximum power, we fix the sample size n to 30 and assume that the distribution of $p_{i,MAX}$ follows the Weibull distribution when $n \geq 30$.

Consequently, $p_{i,MAX}$'s ($i=1,2,\ldots,m$) ($n = 30$) become the samples of the generalized Weibull distribution which was defined in Eqn.(2.23). More importantly, if previous assumptions hold, we have: $\omega(F) = \mu$.

The problem of maximum power estimation is therefore equivalent to that of estimating the location parameter μ of a generalized Weibull distribution from random samples. The simplest way of doing this is to curve-fit the samples to Eqn.(2.23) to get values of α, β, and μ. Unfortunately, our study shows that the curve fitting approach is unstable since the problem becomes very difficult when we try to construct the distribution from a small number of samples. Therefore, we choose another estimation method that is more robust and has a solid theoretical support. After samples of the generalized Weibull distribution are obtained (i.e. $p_{i,MAX}$), the maximum-likelihood estimation method is used to estimate the parameters α, β, and μ.

3.3 A Maximum-likelihood estimator of maximum power dissipation

The maximum-likelihood estimator for parameters of generalized Weibull distribution for $\alpha > 2$ was described in Section II. In fact, α is always larger than 2 if the sample size n is much smaller than the population size $|V|$.

Theorem 4 [10] Let $\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m$ be the estimators that satisfy Eqn. (2.25), $\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m$ ($m \rightarrow \infty$) are unbiased estimators of α, β, μ of the Weibull distribution, which means that $\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m$ ($m \rightarrow \infty$) follow normal distributions with mean values of α_0, β_0, μ_0 and covariance matrix **VAR**. The matrix **VAR** is defined as:

$$\mathbf{VAR} = \frac{1}{m}\begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha,\beta} & \sigma_{\alpha,\mu} \\ \sigma_{\beta,\alpha} & \sigma_\beta^2 & \sigma_{\beta,\mu} \\ \sigma_{\mu,\alpha} & \sigma_{\mu,\beta} & \sigma_\mu^2 \end{bmatrix} = \frac{1}{m}\mathbf{A}^{-1} \tag{3.4}$$

where matrix **A** is symmetric and defined as:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

From Theorem 4 we know that the maximum power estimator $\hat{\mu}_m$ converges to a normal distribution with mean of μ_0 (which is the actual maximum power $\omega(F)$) and variance of σ_μ^2/m.

Theorem 5 $\hat{\mu}_m$ is an unbiased estimator for the maximum power $\omega(F)$. Given confidence level l ($l \in (0,1)$), the *confidence interval* of the estimated maximum power $\hat{\mu}_m$ ($m \rightarrow \infty$) is given by:

$$[\omega(F) - u_l \cdot \sqrt{\sigma_\mu^2/m}, \quad \omega(F) + u_l \cdot \sqrt{\sigma_\mu^2/m}] \tag{3.5}$$

where $\omega(F)$ is the actual maximum power, m is the number of samples, σ_μ^2 is defined in Eqn.(3.4), and u_l is defined as:

$$\int_{-u_l}^{u_l} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = l \qquad (3.6)$$

Proof:

Because $\omega(F)=\mu$, from Theorem 4 we know $\hat{\mu}_m$ is an unbiased estimator of μ and follows a normal distribution with mean value μ and variance σ_μ^2/m. Therefore, $\hat{\mu}_m$ is also an unbiased estimator of $\omega(F)$ and follows a normal distribution with mean value $\omega(F)$ and variance σ_μ^2/m. From [16] we know that, given confidence level l, the confidence interval of $\hat{\mu}_m$ can be calculated by Eqn.(3.5).

∎

Theorem 5 states that the probability that the estimated maximum power falls into the interval defined in Eqn.(3.5) is l. For a given l, a smaller confidence interval means higher estimation accuracy. Therefore, the relative estimation error is inversely proportional to the square root of the variance of the estimator.

The plot of σ_μ^2 (as a function of α and β) is shown in Figure 4. We can conclude from the figure that the variance (and therefore the confidence interval) of the maximum power estimator will decrease when m increases, α decreases, or β increases.
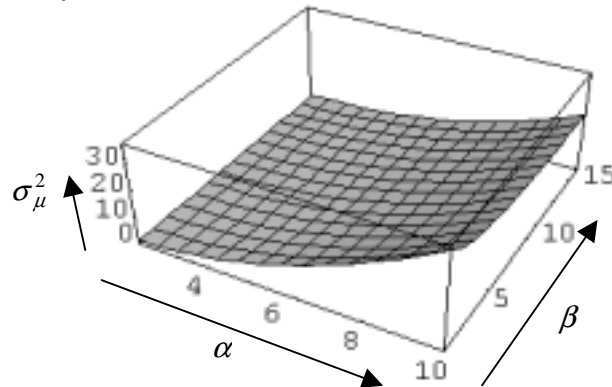
Figure 4 The plot of σ_μ^2 as a function of α and β

In practice, the theoretical confidence interval cannot be calculated directly because σ_μ^2 is unknown. Therefore, we do not know *a priori* how many samples are needed to achieve a certain confidence interval at a given confidence level. An iterative (Monte Carlo) estimation method has been designed to solve this problem.

3.4 An iterative estimation procedure

Figure 5 The distributions of estimated maximum power compared with the nearest normal distribution

11

Experiments have been designed to study the distribution of the maximum likelihood estimator for maximum power in cases when the number m of samples is finite. (We know from Theorem 4 that when $m \to \infty$, the maximum likelihood estimator for $\omega(F)$ follows a normal distribution.) The sampl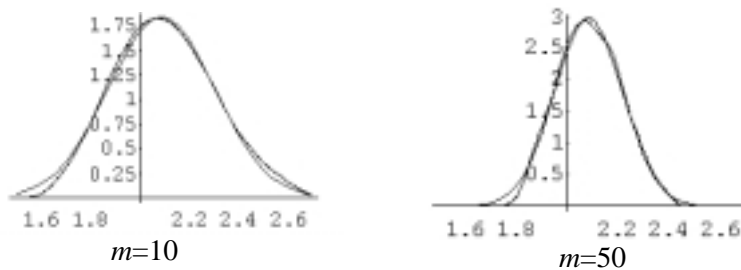e size is fixed at n=30 and different number of samples are used (m=10,50). During each single experiment, m samples with sample size n are randomly selected from the population. Maximum power is then estimated by using the maximum likelihood estimator $\hat{\mu}_m$. For each distinct m, the sampling-estimation procedure is repeated 100 times to form the distribution of estimated value. The distributions of estimated maximum power for different values of m are then formed, and their nearest normal distributions are obtained by least-square curve fitting. The resulting curves for circuit C3540 are shown in Figure 5.
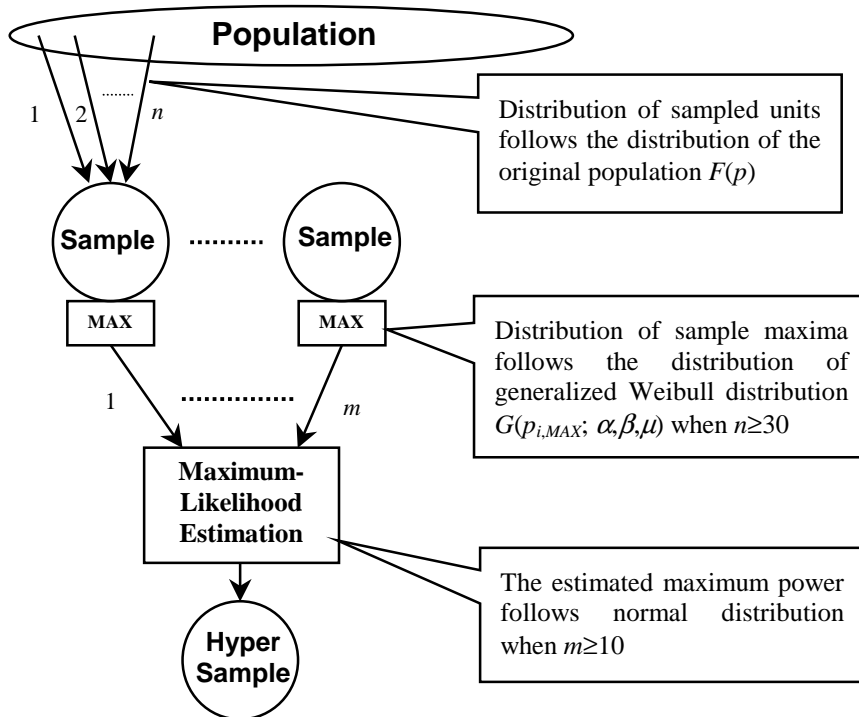
Figure 6 Synopsis of the maximum power estimation method

Similar results are obtained for other circuits. From the experimental results we conclude that the estimator for maximum power is approximately normally distributed when the number of samples is large enough (e.g., $m \geq 10$). Therefore, we assume the normal distribution of the estimator for maximum power for $m \geq 10$.

The actual distribution of the population definitely influences the efficiency and accuracy of our method. One immediate impact would be on the choice of values for n and m. They are empirically chosen based on our power measurement experiments on the testbench circuits under cases I.1 and I.2 discussed in Section I. For an arbitrary power distribution, these values will have to be re-calculated.

Before we introduce a practical procedure for maximum power estimation, we summarize our discussions in earlier part of this section as shown in Figure 6.

In Figure 6, a hyper-sample is defined as the result of one run of maximum power estimation for m samples with size n. We fix the value of n to 30 and value of m to 10, then the number of units which is needed to form a hyper-sample is 300.

Theorem 6 Let $\hat{P}_{i,MAX}$ $(i=1,2,...k)$ denote the ith hyper-sample, for $n=30$ and $m=10$, $\hat{P}_{i,MAX}$ follows the normal distribution with mean value of $\omega(F)$ and variance of $\sigma_\mu^2/10$, where σ_μ^2 is defined in Eqn.(3.4).

Proof:

Based on the two assumptions in Figure 6 and Theorem 5, $\hat{P}_{i,MAX}$ is an unbiased estimator of $\omega(F)$ and follows a normal distribution with mean value $\omega(F)$ and variance $\sigma_\mu^2/10$.

∎

Define:
$$\overline{P}_{MAX} = \frac{1}{k}\sum_{i=1}^{k}\hat{P}_{i,MAX} \qquad \text{and} \qquad s^2 = \frac{1}{k-1}\sum_{i=1}^{k}(\hat{P}_{i,MAX} - \overline{P}_{MAX})^2, \ k > 1 \qquad (3.7)$$

Theorem 7 [16] $\overline{P}_{MAX}$ and s^2 are unbiased estimators of the actual maximum power $\omega(F)$ and σ_μ^2/m, respectively. Given confidence level l, the confidence interval for the actual maximum power is given by:

$$[\overline{P}_{MAX} - \frac{t_{l,k-1}\cdot s}{\sqrt{k}}, \overline{P}_{MAX} + \frac{t_{l,k-1}\cdot s}{\sqrt{k}}] \qquad (3.8)$$

where $t_{l,k-1}$ is the $l\times100\%$ percentile point of the t distribution with degree of freedom of k-1.
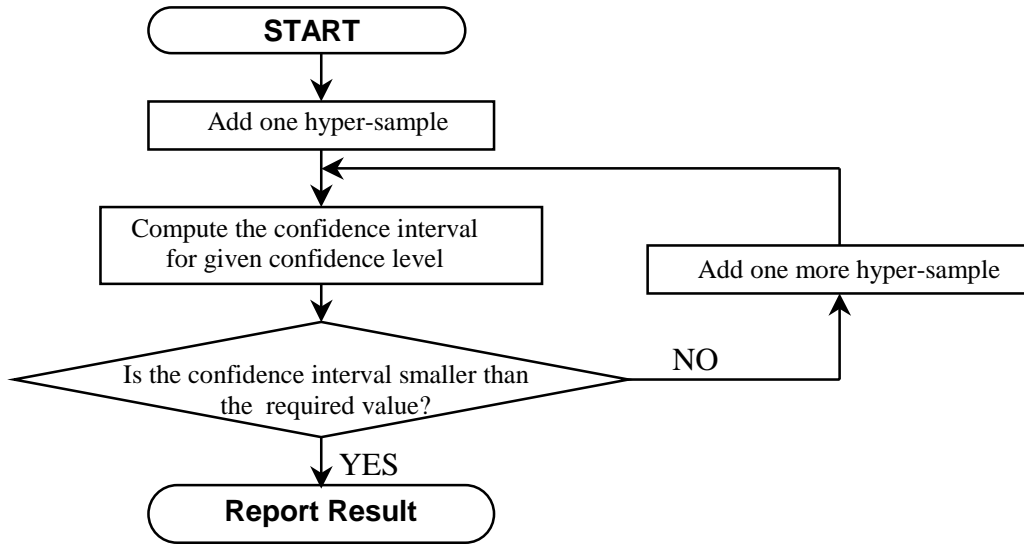
Figure 7 The basic iterative flow of maximum power estimation

Theorem 7 gives us a guideline for designing an iterative procedure for maximum power estimation subject to the required accuracy (relative error less than or equal to ε) at given confidence level l. The basic workflow is shown in Figure 7.

In Figure 7, the generation of a hyper-sample follows the procedure shown in Figure 6. The confidence interval is calculated using Eqn.(3.8). The maximum relative error is calculated using the confidence interval $\frac{t_{l,k-1}\cdot s}{\sqrt{k}}\Big/\overline{P}_{MAX}$. If this quantity is larger than the required ε, then the estimated value has not converged, and we add one more hyper-sample; otherwise, the estimation has converged, and we report the estimation result.

Theorem 8 On average, to achieve some fixed error level at a given confidence level l, the number of samples needed in theory (i.e., the number calculated based on Theorem 5) is smaller than the number needed in practice (i.e., the number calculated based on Theorem 7) by a factor of $\left(u_l/t_{l,k-1}\right)^2$, where $t_{l,k-1}$ and u_l are defined in Eqn.'s (3.5) and (3.8), respectively.

Proof:

To achieve same error level at the same confidence level l both in theory and in practice, the confidence interval defined in Eqn.(3.5) and Eqn.(3.8) must be equal. This leads to:

$$u_l \cdot \sqrt{\sigma_\mu^2/m} = t_{l,k-1} \cdot s/\sqrt{k}$$

Because s^2 is the unbiased estimator for σ_μ^2, we have $s^2 = \sigma_\mu^2$. Therefore,

$$m/k = \left(u_l/t_{l,k-1}\right)^2 .$$

Because u_l is always smaller than $t_{l,k-1}$, m is always smaller than k.

∎

Since u_l is always smaller than $t_{l,k-1}$, the theoretical number of samples (given by Theorem 5) is lower than the number obtained by the practical approach (given by Theorem 7). Moreover, other factors introduced by procedures such as the random sampling and maximum likelihood estimation, will further increase the required the number of samples in the practical approach.

3.5 Practical issue: finite population versus infinite population

The approach discussed in the earlier part of this section is designed for estimating the maximum power of an infinite population. However we must deal with a finite population in real applications. As an example, our experimental setup in the next section uses finite populations. Experimental results show that if we use the same approach for the finite population as for the infinite population, there will be some positive bias in the maximum likelihood estimation in the sense that the mean of the estimated value is always larger than the actual maximum power of the population. This happens because estimator $\hat{\mu}_m$ is estimating the maximum power of an infinite population that should, after all, have a long tail after the actual maximum power of the population. However this tail does not exist in the case of a finite population.

To solve this problem, we can regard the finite population V as a sample of size $|V|$ selected randomly from the assumed continuous distribution for the infinite population. If we assume there is only one unit in the finite population which consumes the maximum power, then the maximum power of the finite population becomes the estimated $(1-1/|V|)$ quantile point of the assumed continuous distribution. According to the tail-equivalence property between a distribution and the limiting distribution of its sample maxima [11], estimating the $(1-1/|V|)$ quantile point of the original population is equivalent to estimating the $(1-1/|V|)$ quantile point of the asymptotic Weibull distribution of the sample maxima. Therefore, when estimating the maximum power of a finite population, instead of using the theoretical $\hat{\mu}_m$ (which is the 100% quantile point of the estimated generalized Weibull distribution), we use the $(1-1/|V|)$ quantile point of the Weibull distribution (whose parameters can be estimated by using the maximum likelihood estimator) as the estimator for the maximum power. We call this the "modified estimator" for the finite population. Experimental results show that the modified estimator gives us a very good normally-distributed estimator for finite populations where the bias has been nearly eliminated.

IV. MAXIMUM POWER ESTIMATION AND EXPERIMENTAL RESULTS

Methods for estimating maximum power of different categories are slightly different from each other. Experiments have been designed to study the efficiency and accuracy of our approach on maximum power estimation. We present our experimental results and comparison with other techniques for different categories separately.

Category I.1. Estimating the *unconstrained maximum power*.

Proposed method:

In this category, the goal is to estimate the maximum power of the circuit for all possible input vector pairs. Consequently the simple random sampling procedure can be realized by randomly generating vector pairs, that is, the two methods (i.e., random vector generation and simple random sampling) are equivalent in this case. Except for the fact that the sampling technique is replaced by the random vector generation, the remaining part of our approach (cf. Figure 6 and Figure 7) remains the same.

Experimental setup A:

For circuits with a large number of inputs, we are not able to compare our experimental results with the "absolute" maximum power because that requires exhaustive simulation of an exponential number of vectors. The only thing we can do is to generate and simulate as many vector pairs in the population as we can and use the maximum of them as the basis for our comparison.

When comparing with the ATPG-based techniques, first we like to point out that the estimation speed of the ATPG-based techniques is higher than our approach; however, the accuracy of our approach will be higher because the ATPG-based techniques can only handle simple delay models at the gate-level. Second we are not able to directly compare our results with those of the ATPG-based techniques even at the gate-level because we have not been able to obtain the programs (although we have tried). As an alternative, by studying the reported results for ATPG-based techniques, we have observed that the most accurate maximum power estimation results generated by ATPG-based techniques are comparable with the results obtained by simulating (at the gate-level) 10,000 randomly generated vector pairs [4][5][6]. Therefore, we indirectly compare the estimation quality of our approach with ATPG-based techniques by comparing it with that obtained by the random vector generation approach.

We use a similar strategy to compare with other techniques in [6] and [7] for similar reasons.

Results and discussion:

When compared to the quantile estimation approach proposed in [8], we note that although the quantile estimation technique can estimate a high-quantile (e.g., 99.999%) point as the maximum power at given confidence level, it cannot give the confidence of estimating the maximum power. In other words, since the actual maximum power in the population is unknown, there is no way to tell which quantile point (99.99%, 99.999%, or 99.9999%?) to estimate such that the relative error of the estimated maximum power w.r.t. the actual maximum is less than a given value (e.g., 5%). In addition, the efficiency of using quantile estimator as the maximum power estimator is low. This is because if we use the a quantile estimator to estimate the maximum power, the procedure is equivalent to simple random sampling in the population with the maximum simulated power value as the estimation result (see below).

Now let us present a theoretical study of the efficiency of the estimation method of random vector generation, or simple random sampling. Assume we want to estimate maximum power of error less than 5% at confidence level 90% for a population. Let the size of the population be $|V|$. Define the "qualified units" as those units whose values are within 5% difference of the actual maximum. Assume the number of the "qualified units" is Z. The portion of the "qualified units" in the whole population is then $Y=Z/|V|$. If we sample x units from the population, the probability that there is at least one "qualified unit" in these x

units is given by: $P = 1-(1-Y)^x$. For P to be larger than or equal to 90%, we need, on average, $x = \log(0.1)/\log(1-Y)$ sampled units. From our experiments, we have observed that Y is very small (e.g., <0.0001). This leads to very large x (e.g., >23,000). So we conclude that the efficiency of random vector generation (and therefore quantile estimation) will be low.

The experimental setup is as follows. The population contains 160,000 randomly generated high activity (i.e., average switching activity larger than 0.3) vector pairs. Random vector generation is equivalent to generating a simple random sampling of vector pairs from the population. The whole population is simulated using Powermill [16] to get the power consumption value for each unit and, in the process, the actual maximum power. Our approach (with n=30 and m=10) and simple random sampling (SRS) have been applied to perform maximum power estimation for relative error < 5% at confidence level 90%. The experimental results are shown in Table 1 and Table 2. Our approach has been used to perform maximum power estimation 100 times for each circuit.

Table 1 shows the comparison of efficiency and accuracy of our approach versus simple random sampling. The portion of the "qualified units" in the whole population is given in the 2nd column. The maximum, minimum, and average number of units needed for our approach to converge are reported in the 3rd, 4th, and 5th columns, respectively. The 6th column gives the theoretically calculated (according to the discussion of the second paragraph from the bottom of page 14) number of units needed by simple random sampling to achieve the same error (5%) and confidence (90%) level. The 7th and 8th columns give the absolute value of the maximum and minimum estimation error of our approach. We have not given the relative error for SRS because the SRS technique is not able to predict the maximum power subject to given error and confidence levels.

Table 2 shows the estimation quality comparison. Simple random sampling techniques using 2500, 10K, and 20K units are performed 100 times, respectively. The 2nd column gives the actual maximum power of the population, which we call the "best-estimate (BE) maximum power" because it is the largest among the simulated vector pairs. Columns 3, 4, 5, and 6 give the results of the largest-error estimates for different techniques. Column 7, 8, 9, and 10 give the results of the percentage of the time when the estimated value exceeds the error level.

Table 1 Experimental results for comparing the efficiency and accuracy (I.1)

Circuit	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.0001	2700	900	1923	23024	6.0%	0.3%
C1908	0.00015	3600	1500	2412	15349	5.3%	2.4%
C2670	0.000288	1500	600	924	7993	6.2%	0.6%
C3540	0.000094	5100	600	2553	24494	5.2%	1.2%
C432	0.000038	5400	2100	3543	60593	7.7%	1.7%
C5315	0.000194	2700	600	1653	11868	5.8%	0.8%
C6288	0.000163	900	600	675	14125	6.2%	0.05%
C7552	0.00005	4500	3300	3825	46050	8.2%	0.6%
C880	0.000063	3000	2700	2859	36547	5.4%	2.9%

The experimental results show that our approach is much more efficient than the simple random sampling technique (about 12X speedup on average). More importantly, however, simple random sampling or similar techniques are not reliable because they cannot provide a confidence interval and a confidence

level for maximum power estimation. Also the estimation quality of our approach is obviously better than simple random sampling. From the results of Table 2, if we compare our approach with simple random sampling with 20K units, the average largest error is 5.3% for our approach and 10.4% for SRS. As to the average percentage of estimated value with error larger than 5%, it is 4.3% for our approach and 23% for SRS. It can be foreseen that the advantage of our approach over SRS will be more predominant for an infinite population.

Table 2 Experimental results for comparing the estimation quality (I.1)

Circuits	BE max. power (mW)	Largest estimation error				% of estimates with error > 5%			
		Our appr.	SRS			Our appr.	SRS		
			2500	10K	20K		2500	10K	20K
C1355	2.145	-6.0%	-13%	-8.5%	-6.3%	6%	80%	52%	15%
C1908	2.745	-5.3%	-14%	7.5%	-6.3%	3%	73%	28%	8%
C2670	6.529	-6.2%	-8.6%	-5.4%	-2.5%	1%	38%	2%	0%
C3540	10.732	5.2%	-14%	-10%	-8.9%	5%	80%	52%	33%
C432	1.818	-7.7%	-22%	-13%	-14%	8%	89%	73%	57%
C5315	14.372	5.8%	-9.7%	-7.7%	-6.2%	2%	73%	27%	3%
C6288	126.62	6.2%	-21%	-21%	-21%	3%	76%	26%	5%
C7552	31.237	8.2%	-14%	-10%	-7.3%	7%	92%	69%	54%
C880	4.312	5.4%	-20%	-15%	-11%	4%	88%	42%	29%

Experimental setup B: In this setup, we want to estimate the maximum power consumption of circuits based on all possible input vector pairs.

It is however impossible to do exhaustive simulation of all possible input vector pairs for the ISCAS benchmark circuits used in Setup A. This is because the number of inputs for these circuits ranges from 32 to more than 200. Therefore we arbitrarily choose 9 of the inputs to be changed exhaustively. The other inputs are fixed at some random input combination. As a result, for each circuit, we obtain a population of $2^{18}=262,144$ units. We have done experiments on four circuits: C1355, C1908, C3540 and C6288. The assignments of the fixed and changed inputs of these circuits are shown in Table 3.

Table 3 Assignment of fixed and changed inputs.

Circuits	Number of Inputs	Fixed Inputs	Changing Inputs
C1355	41	1~16, 26~41	17~25
C1908	33	1~12, 22~33	13~21
C3540	50	1~21, 31~50	22~30
C6288	32	1~12, 22~32	13~21

Table 4 Actual maximum power consumption (mW).

Circuits	Fixed Input Pattern 1	Fixed Input Pattern 2	Fixed Input Pattern 3	Fixed Input Pattern 4	Fixed Input Pattern 5
C1355	0.98	0.96	0.91	0.96	0.92
C1908	1.76	2.06	1.30	2.76	1.28
C3540	2.71	1.29	2.07	1.84	1.68
C6288	43.29	39.61	42.62	44.23	44.09

The experiments are repeated five times using five different combinations of random values for the fixed inputs of each circuit. The maximum power value (at 5Mhz clock rate) of the population for each setup is presented in Table 4.

Figure 8 shows some sample distributions for total power dissipation in different circuits.
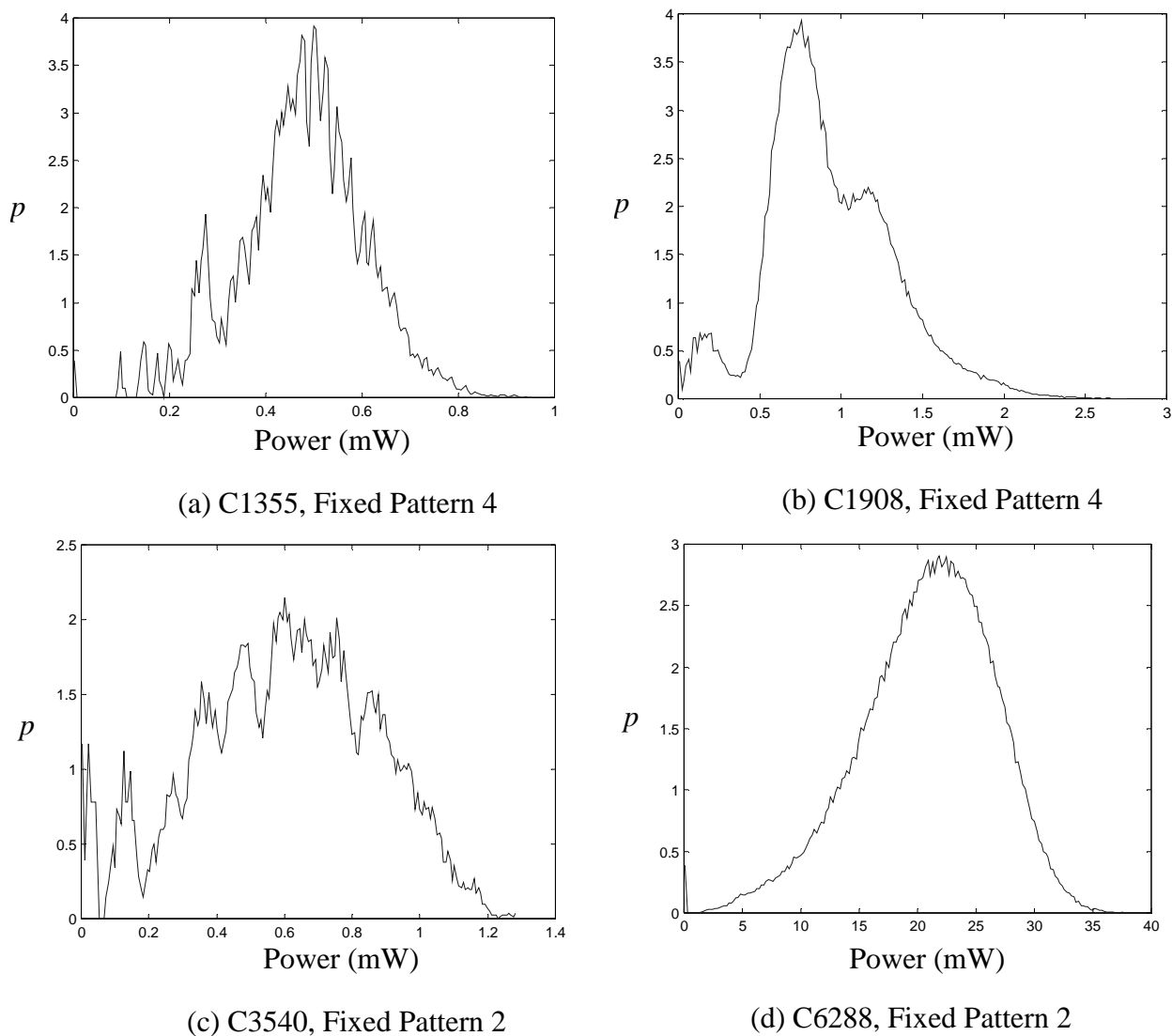

(a) C1355, Fixed Pattern 4

(b) C1908, Fixed Pattern 4

(c) C3540, Fixed Pattern 2

(d) C6288, Fixed Pattern 2

Figure 8 Example distributions of total power consumption under partially changing inputs.

We applied the same method that we used for *Setup A* to estimate the peak power dissipation in these four circuits under partially changing inputs. Column entries in these tables are defined similar to those in Table 1. The experimental results are reported in the Tables 5-9.

Table 5 Experimental results for Fixed Pattern 1.

Circuits	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.00006104	3600	600	2187	37721	10.0%	0.01%
C1908	0.00081250	1800	600	1095	2833	8.8%	2.0%
C3540	0.00125122	1800	600	1173	1839	7.0%	1.3%
C6288	0.00011444	2100	600	1335	20119	6.0%	0.1%

Table 6 Experimental results for Fixed Pattern 2.

Circuits	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.00028229	4800	600	2853	8156	10.5%	0.4%
C1908	0.00012207	3600	600	1881	18862	8.0%	0.002%
C3540	0.00109863	2100	600	1260	2095	9.9%	1.8%
C6288	0.00003052	2400	600	1092	75444	5.0%	0.05%

Table 7 Experimental results for Fixed Pattern 3.

Circuits	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.00104904	3600	600	2070	2194	10.3%	0.5%
C1908	0.00003815	3300	600	1932	60355	6.9%	0.2%
C3540	0.00048828	2400	600	1275	4715	6.2%	0.1%
C6288	0.00005722	2100	600	1188	40240	5.7%	0.3%

As we can see from data reported in these tables, compared to SRS, on average, our approach uses between 1-70 times fewer input vector pairs to produce peak power estimates within the same level of error.

Table 8 Experimental results for Fixed Pattern 4.

Circuits	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.00042725	4500	600	2583	5388	11.4%	0.9%
C1908	0.00009918	3900	600	2022	23215	8.2%	0.3%
C3540	0.00078418	2100	600	1305	2935	8.7%	2.2%
C6288	0.00008011	2100	600	1266	28742	4.5%	0.09%

Table 9 Experimental results for Fixed Pattern 5.

Circuits	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.00072098	4200	600	2055	3193	7.6%	0.2%
C1908	0.00022888	2700	600	1734	10059	9.7%	0.3%
C3540	0.00022125	1800	600	1284	10406	7.1%	2.4%
C6288	0.00015640	3000	600	1521	14721	10.0%	0.3%

Category I.2. Estimating the *constrained maximum power*.

Proposed method: Similar to Category I.1, except that vector pairs are generated under given constraints (e.g., average switching activity)

Table 10 Experimental results for comparing efficiency and accuracy (I.2)

Circuit	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.000241	3900	600	2112	9553	5.4%	1.8%
C1908	0.000378	3000	600	2403	6090	7.3%	2.0%
C2670	0.000778	900	600	675	2958	4.1%	0.5%
C3540	0.000196	1200	900	1053	11747	6.7%	4.0%
C432	0.000071	3300	1200	2259	32430	7.7%	2.2%
C5315	0.000488	1200	900	975	4717	7.1%	4.1%
C6288	0.000427	1200	600	1053	5391	4.5%	1.7%
C7552	0.000308	3900	900	2253	7475	8.0%	0.9%
C880	0.000135	2700	600	1704	17055	12%	2.1%

Experimental setup: Similar to the *Setup A* of Category I.1, this time we generate two populations (each of size 80,000) subject to the constraint that the average switching activity per input line is 0.7 and 0.3, respectively. A detailed comparison with simple random sampling has been performed as well. However,

we give only the tables for comparing efficiency and accuracy in order to save space. The experimental results for populations of average switching activity 0.7 and 0.3 are shown in Table 10 and Table 11, respectively. The meaning of the entries in different columns is the same as that in Table 1. The estimation quality comparison can be seen from the value of the portion of the "qualified units" in the 2nd columns of both tables. As expected when the number of qualified units in the population decreases, the number of units needed to estimate the maximum power dissipation in the circuit increases.

Table 11 Experimental results for comparing efficiency and accuracy (I.2)

Circuit	Portion of the "qualified units"	# of units needed				Relative error	
		Our approach			SRS	Our approach	
		MAX	MIN	AVE	AVE	MAX	MIN
C1355	0.000119	4800	1500	3348	19384	3.6%	2.2%
C1908	0.000246	2700	900	2001	9359	6.6%	3.5%
C2670	0.000313	3600	1500	2583	7355	5.3%	1.7%
C3540	0.000053	5100	600	3585	43444	7.4%	2.9%
C432	0.000179	3000	1500	2388	12862	6.8%	2.4%
C5315	0.000231	3600	1200	2622	9967	13%	3.4%
C6288	0.000079	6000	2700	5424	29145	5.1%	0.6%
C7552	0.000194	2400	1200	1977	16446	7.1%	3.3%
C880	0.000018	2700	900	1896	127920	5.0%	1.9%

V. CONCLUSIONS

A statistical approach based on the asymptotic theory of extreme order statistics is presented. This is the first maximum power estimation approach that can provide a confidence interval at a given confidence level. This is also the first approach that can perform maximum power estimation for any user-specified error and confidence level. The proposed approach can predict the maximum power in the space of constrained input vector pairs as well as the complete space of all possible input vector pairs. It is an efficient simulation-based approach with high accuracy. The generality of this approach also makes it applicable to other fields of VLSI design automation, such as, maximum delay estimation.

Revised Manuscript

REFERENCES

[1] 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*, 11(3):373-383, Mar. 1992.

[2] S.Manne, A.Pardo, R.Bahar, G.Hachtel, F.Somenzi, E.Macii and M.Poncino, "Computing the Maximum Power Cycles of a Sequential Circuit," *Proceedings of Design Automation Conference*, 1995.

[3] H. Kriplani, F. Najm, and I. Hajj, "Maximum Current Estimation in CMOS Circuits," *Proceedings of Design Automation Conference*, 1992.

[4] C.-Y. Wang and K. Roy, "Maximum Current Estimation in CMOS Circuits Using Deterministic and Statistical Techniques," *Proceedings of IEEE VLSI Design Conference*, Jan. 1996.

[5] C.-Y. Wang, T.-L. Chou, and K. Roy, "Maximum Power Estimation Under Arbitrary Delay Model," *Proceedings of IEEE International Symposium on Circuits and Systems*, 1996.

[6] C.-Y. Wang and K. Roy, "COSMOS: A Continuous Optimization Approach for Maximum Power Estimation of CMOS Circuits," *Proceedings of ACM/IEEE International Conference on Computer-Aided Design*, 1997.

[7] M.S. Hsiao, E.M. Rudnick, and J.H. Patel, "K2: An Estimator for Peak Sustainable Power of VLSI Circuits," *Proceedings of International Symposiunm on Low Power Electronics and Design*, pp.178-183, Aug. 1997.

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

[9] C.S. Ding, Q. Wu, C.T. Hsieh, and M. Pedram, "Statistical Estimation of the Cumulative Distribution Function for power dissipation in VLSI Circuits," *Proceedings of Design Automation Conference*, pp.371-376, Jun., 1997.

[10] R.L. Smith, "Maximum likelihood estimation in a class of nonregular cases," *Biometrika*, pp. 67-90, 1972.

[11] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, Robert E. Krieger Publishing Company, 1987

[12] M.R. Leadbetter, G. Lindgren, H. Rootzen, *Extremes and Releated Properties of Random Sequences and Processes*, Springer-Gerlag New York Inc., 1983.

[13] P. Öfverbeck and G. Östberg, *A Study of Outliers in Statistical Distributions of Mechanical Properties of Structural Steel*, Techn. Report, Dept. of Structural Engineering, Univ. of Lund., pp. 1-37, 1977.

[14] A. G. Davenport,*Wind Structure and Wind Climate*, Int. Res. Seminar on Safety of Structures Under Dynamic Loading, Eds. I. Holand, D. Kavlie, G. Moe, R. Sigbjörnsson, pp. 238-256, 1978.

[15] E. Simiu and J. J. Filliben, *Probability Distributions of Extreme Wind Speeds*, J. Structural Div., ASCE 102, pp.1861-1877, 1976.

[16] M. H. Hansen, W. N. Hurwitz and W. G. Madow, *Sample Survey Methods and Theory*, Wiley publications, New York, 1953.

[17] *Analog Cirsuit Engine Release Notes*, Synopsis, 1997.

[18] Q. Qiu, Q. Wu, and M. Pedram, "Maximum Power Estimation Using the Limiting Distributions of Extreme Order Statistics," *Proceedings of Design Automation Conference*, pp. 684-689, June, 1998.