Maximum Power Estimation Using the Limiting Distributions of Extreme Order Statistics

Qinru Qiu, Qing Wu and Massoud Pedram

Department of Electrical Engineering – Systems
University of Southern California
Los Angeles, CA 90089
EMAIL: {qinru, qingwu, massoud}@zugros.usc.edu

Abstract

In this paper we present a statistical method for estimating the maximum power consumption 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 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. The simulation-based nature of the proposed method allows it to avoid the limitations of a gate-level delay model and a gate-level circuit structure. Last, but not least, the proposed method can produce maximum power estimates to satisfy user-specified error and confidence levels. Experimental results show that this method, on average, produces maximum power estimates within 5% of the actual value and with a 90% confidence level, by simulating only about 2500 vector pairs.

I. Introduction

Circuit reliability is an important issue in today's VLSI manufacturing. There are many sources that may cause circuit failure, one of them is excessive power dissipation over a short period of time. Unexpected high current in short time may lead to permanent circuit damage, sudden voltage change on the supply source and ground nets, or temporary circuit failure. To design a circuit with high reliability, designers have to rely on the efficient and accurate estimation of maximum cycle-by-cycle power (or for short maximum power). Maximum power estimation in VLSI circuits is also essential to determine the IR drop and bounce noise on supply lines and to optimize 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 looking for the maximum-power-consuming vector pair among all possible input vector pairs. Therefore, these techniques focus on finding the lower bound and upper bound of the maximum power. However, different design requirements of today's VLSI chips make things a little bit more complicated. In general, we divide the scope of 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 as the *unconstrained maximum power*.

I.2 The maximum power for given transition/joint-transition probability specification for the circuit inputs. We refer to this as the *constrained maximum power*.

A number of techniques have been developed to solve the problems in Category I.1 [1]-[8] and Category I.2. The method proposed in [1] propagates the signal uncertainty through the circuits to obtain a loose upper bound on the maximum power. The bound is then made tighter by doing a detailed search on part of the primary inputs. The bound tightening method tends to be time consuming when the number of the primary inputs is large.

The Automatic Test Pattern Generation (ATPG) based techniques [3]-[4] 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. The ATPG based techniques are very efficient and generate a tighter lower bound than that generated by random vector generation. The limitations are however that the ATPG based techniques can only handle simple delay models such as the zero-delay and unit-delay models and that the analysis is done at the gate-level. Consequently, the estimation accuracy is not high. A continuous optimization method was proposed in [5], which treats the input vector space as a continuous real-valued vector space and then performs a gradient search to find the function maximum. Similar to the ATPG based techniques, the estimation accuracy is not high.

The authors of [6] proposed a technique for finding the maximum power-consuming vector using a genetic search algorithm. The limitation of this approach is that it requires simulation of a lot of vectors, i.e., its efficiency is not high. Statistical methods have been studied for maximum power estimation. In [3] 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 simulation. This method also suffers from low efficiency.

The theory of order statistics has been applied in [7][8] to estimate maximum power by estimating the high quantile point. Their efficiency is as low as 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 theory of *Asymptotic Extreme Order Statistics*. Compared to previous work, our approach makes the following tangible contributions:

1. Our approach is the first approach that provides the confidence interval for the estimated maximum power for the user-specified confidence level.
2. Our approach is the first approach which can do maximum power estimation for any given error and confidence levels.
3. Our approach can estimate the maximum power defined in both categories I.1 and I.2.
4. Because it is a simulation-based technique, the delay model or the circuit structure do not limit its accuracy.
5. By efficient statistical estimation of the extreme distributions, the estimation efficiency is largely improved compared to existing statistical methods (including simple random sampling or quantile estimation).

On average, the method can do maximum power estimation by simulating only about 2500 vector pairs to achieve a 5% error at a confidence level of 90%.

This paper is organized as follows, Section II introduces the theory of asymptotic 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

2.1 The asymptotic theory of extreme order statistics

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

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

The quantile function (in short 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. If they are drawn in a random manner, they are called independent identically distributed (in short i.i.d.) r.v.'s. If one is not interested in the order in which $z_1,z_2,...,z_n$ are drawn, but interested in the order of the magnitude of their values, one has 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.

The distribution function of the sample maxima $X_{n:n}$, is given by:

$$P\{X_{n:n} \le t) = F(t)^n \tag{2.3}$$

Three distribution functions are given for studying the limiting d.f. of sample maxima (in other words: extreme value d.f.'s):

$$G_{1,\alpha}(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \exp(-x^{-\alpha}) & \text{if } x > 0 \end{cases}, \text{ “Fréchet”} \tag{2.4}$$

$$G_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^{-\alpha}) & \text{if } x \le 0 \\ 0 & \text{if } x > 0 \end{cases}, \text{ “Weibull”} \tag{2.5}$$

$$G_3(x) = \exp(-e^{-x}) \quad \text{for every } x, \quad \text{“Gumbel”} \tag{2.6}$$

Definition 1 [10] F is said to belong to the weak domain of attraction of limiting d.f. G, if there exist serious of constants $a_n > 0$ and reals b_n such that:

$$F^n(b_n + xa_n) \to G(x), \qquad n \to \infty \tag{2.7}$$

for every continuity point of G.

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

$$\omega(F) = \sup\{x : F(x) < 1\} = F^{-1}(1) \tag{2.8}$$

Theorem 1 [10] A d.f. F belongs to the weak domain of attraction of an extreme value d.f. $G_{i,\alpha}$ iff., one of the following conditions holds:

$$(1,\alpha):\ \omega(F) = \infty \text{ and } \lim_{t\to\infty}\left[\frac{1-F(tx)}{1-F(t)}\right] = x^{-\alpha}, x > 0 \tag{2.9}$$

$$(2,\alpha): \omega(F) < \infty \text{ and } \lim_{t\downarrow 0}\left[\frac{1-F(\omega(F)+xt)}{1-F(\omega(F)-t)}\right] = (-x)^{\alpha}, x < 0 \tag{2.10}$$

$$(3):\ \lim_{t\uparrow\omega(F)}\left[\frac{1-F(t+xg(t))}{1-F(t)}\right] = e^{-x}, \quad -\infty < x < \infty \tag{2.11}$$

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

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

$$(1,\alpha):\quad b_n^* = 0, \qquad a_n^* = F^{-1}(1 - 1/n) \tag{2.12}$$

$$(2,\alpha):\quad b_n^* = \omega(F), \qquad a_n^* = \omega(F) - F^{-1}(1 - 1/n) \tag{2.13}$$

$$(3):\quad b_n^* = F^{-1}(1 - 1/n), \qquad a_n^* = g(b_n^*) \tag{2.14}$$

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 not only gives the conditions of to which d.f. G the extreme distribution will converge, but also the guidelines for us to choose a correct asymptotic extreme distribution for a specific application.

Theorem 2 [10] The weak convergence to the limiting d.f. G holds for other choices of constants a_n and b_n if, and only if,

$$a_n / a_n^* \to 1 \quad \text{and} \quad (b_n - b_n^*)/a_n \to 0 \ \text{ as } n \to \infty \tag{2.15}$$

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 ($n\to\infty$) in Eqn.(2.13) is unique.

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

For reasons that will be made clear in a later section, we are interested in developing a maximum-likelihood estimator for parameters of a generalized Weibull distribution defined as:

$$G(x;\alpha,\beta,\mu) = \begin{cases} \exp(-\beta(\mu - x)^{-\alpha}) & \text{if } x \le \mu \\ 0 & \text{if } x > \mu \end{cases} \tag{2.16}$$

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.

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 [9]:

$$L_m(\alpha,\beta,\mu) = \frac{1}{m}\sum_{i=1}^{m}\log[G(x_i;\alpha,\beta,\mu)] \tag{2.17}$$

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

$$\frac{\partial L_m}{\partial\{\alpha,\beta,\mu\}}(\hat{\alpha}_m,\hat{\beta}_m,\hat{\mu}_m)=0 \quad (2.18)$$

Let α_0,β_0,μ_0 denote the actual values of parameters of the distribution, G. It was proved in [9] that, when α > 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.

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 that the circuit may exhibit for 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 either all possible input vector pairs applied to a circuit, or all possible vector pairs under some input transition probability constraints. Although there could only be a finite number of distinct vector pairs in the population, but the size of V, represented by $|V|$, is assumed to be infinite since there is the possibility of repeating the vector pairs.

3.2 The asymptotic distribution of the sample maximum power

If we regard power consumption for a vector pair as a random variable p, a distribution of p is then formed by the power consumption values of vector pairs in set V. The average power is the mean value of the distribution. The maximum power is then the right endpoint of the distribution. Like other papers on statistical power estimation, we assume the d.f. of power consumption in a large LSI circuit as a continuous distribution.

Given population V, the ith sample for max 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:

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

According to Eqn.(2.5), the d.f. of $p_{i,MAX}$ can be written as: $H(p_{1,MAX})=F^n(p)$. As mentioned in Theorem 1, $H(b_n+p_{i,MAX}\cdot a_n)$ asymptotically 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 $\omega(F)$ denoting the actual maximum power of the population.

We know that power consumption in a LSI circuit is always a finite value, i.e., $\omega(F)<\infty$. Therefore the condition in (2.9) is not met and $H(b_n+p_{i,MAX}\cdot a_n)$ can not converge to $G_{1,\alpha}$. Also, because the upper bound of supporting domain for G_3 is infinite while that of $G_{2,\alpha}$ is finite, The condition in (2.10) is more likely to hold than that in (2.11). Therefore $H(b_n+p_{i,MAX}\cdot a_n)$ is more likely to converge to $G_{2,\alpha}$ rather than G_3.

It is pointed out in [10] that, most frequently used continuous distributions with finite right endpoint ($\omega(F)<\infty$) satisfy the condition in Eqn.(2.10). Therefore, in many engineering applications of maxima estimation, it is assumed that the distribution under study belongs to the weak convergence domain of $G_{2,\alpha}$. This statement has also been empirically proved to be true by our experiments (cf. later this section).

Therefore, we state 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:

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

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

From Eqn.'s (2.13) and (2.15), 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.(2.16) into Eqn.(3.3), we get

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

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

Experiments have been designed to verify the asymptotic distribution of sample maxima. The distributions of sample maxima for different sample size (n = 2, 20, 30, 50) was formed by 1,000 random samples from the population. Their closest Weibull distributions are obtained by using least-square fitting techniques. Figure 1 shows the results for circuit C3540.

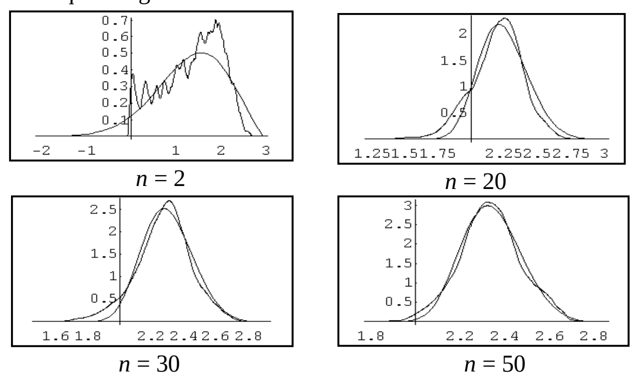

Figure 1 The comparison between distribution of sample maxima and Weibull distribution

Experiments are done for other circuits and populations and the similar results are obtained. From these results we concluded that the difference between distributions of $p_{i,MAX}$ and Weibull distribution in the region near the maximum power is negligible when $n\geq 30$. Since we are only interested in estimating the maximum power, we fix the sample size n to 30 and assum that the distribution of $p_{i,MAX}$ follows Weibull distribution when $n\geq 30$. Consequently, $p_{i,MAX}$ (i=1,2,…,m) (n = 30) become the samples of the generalized Weibull distribution in (2.16). Most importantly, if previous assumptions hold, we have: $\omega(F)=\mu$.

Therefore, the problem of maximum power estimation is equivalent to the problem 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.(2.16) to get values of α, β, and μ. However, our study shows that the curve fitting approach is unstable since the problem becomes difficult when we trying to construct the distribution from small number of samples. We choose another estimation method that is more robust and has a solid theoretical support.

3.3 A Maximum-likelihood estimator of maximum power dissipation

The maximum-likelihood estimators for parameters of generalized Weibull distribution for α > 2 have been introduced in Section II. In fact, α is always large than 2 if the sample size n is much smaller than the population size $|V|$. Consequently, let $\hat{\alpha}_m,\hat{\beta}_m,\hat{\mu}_m$ be the estimators that satisfy Eqn. (2.18), we can prove the following result.

Theorem 3 $\hat{\alpha}_m,\hat{\beta}_m,\hat{\mu}_m$ ($m\to\infty$) are the unbiased estimators of α,β,μ of the Weibull distribution, which means that

$\hat{\alpha}_m, \hat{\beta}_m, \hat{\mu}_m$ ($m \to \infty$) follow normal distributions with mean values of α_0, β_0, μ_0 and covariance matrix **VAR**. The matrix **VAR** is defined as:

$$\mathbf{VAR} = \frac{1}{m}\begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha,\beta} & \sigma_{\alpha,\mu} \\ \sigma_{\beta,\alpha} & \sigma_\beta^2 & \sigma_{\beta,\mu} \\ \sigma_{\mu,\alpha} & \sigma_{\mu,\beta} & \sigma_\mu^2 \end{bmatrix} \quad (3.4)$$

From Theorem 3 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 4 $\hat{\mu}_m$ is an unbiased estimator for maximum power $\omega(F)$. Given confidence level l ($l \in (0,1)$), the *confidence interval* of the estimated maximum power $\hat{\mu}_m$ ($m \to \infty$) is given by:

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

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:

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

Theorem 4 states that the probability that the estimated maximum power falls into the interval defined in Eqn.(3.5) is l. For a given l, 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.

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 certain confidence interval at given confidence level. An iterative (Monte-Carlo) method has been designed to solve this problem.

3.4 The iterative estimation procedure

Experiments have been designed to study the distribution of the maximum likelihood estimator for maximum power in cases when the number of samples m is finite (we know from Theorem 3 that when $m \to \infty$ this maximum likelihood estimator for $\omega(F)$ follows a normal distribution). The sample 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 μ_m. The program for maximum likelihood estimation can be found in many places. 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 results for circuit C3540 are shown in Figure 2.

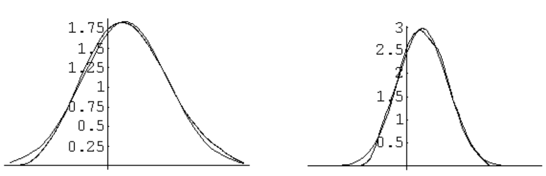

Figure 2 The distributions of estimated maximum power compared with the nearest normal distribution

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

Before we introduce our practical maximum power estimation procedure, we summarize our discussions in earlier part of this section as shown in Figure 3.

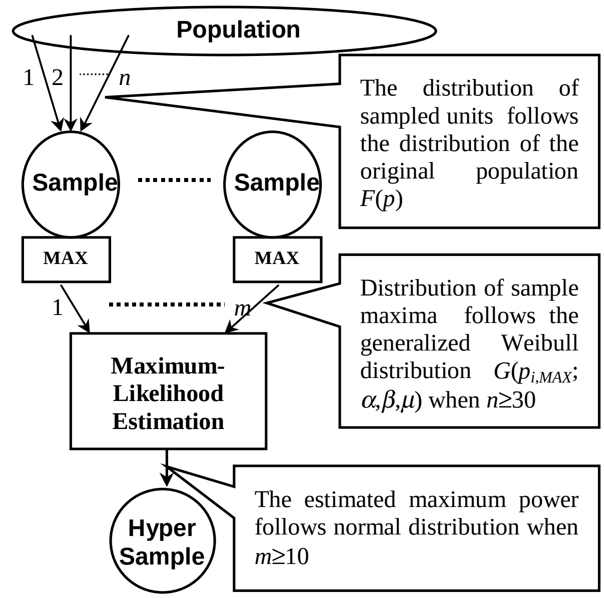

Figure 3 Synopsis of maximum power estimation method

In Figure 3, 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 5 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). By define:

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

Theorem 6 $\overline{P}_{MAX}$ and s^2 are unbiased estimators of the actual maximum power $\omega(F)$ and $\sigma_\mu^2 / 10$, respectively. Given confidence level l, the confidence interval for the actual maximum power is given by:

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

where $t_{l,k-1}$ is the $l \times 100\%$ percentile point of the t distribution with degree of freedom of k-1.

Theorem 6 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 4.

In Figure 4, the generation of a hyper-sample follows the procedure shown in Figure 3. Confidence interval is calculated using Eqn.(3.8). The maximum relative error is calculated using the confidence interval as $\frac{t_{l,k-1} \cdot s}{\sqrt{k}} \Big/ \overline{P}_{MAX}$. 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.

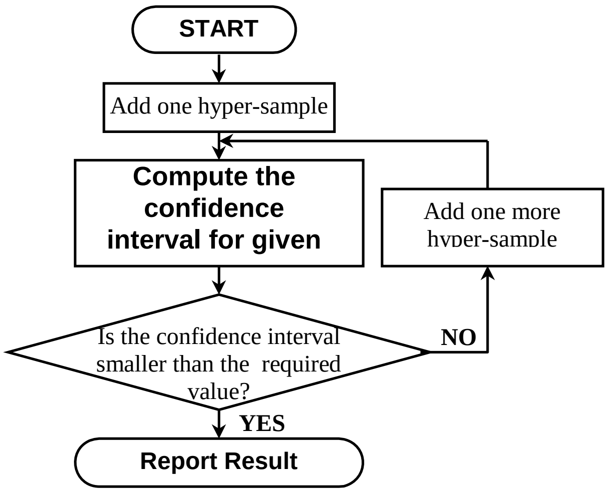

Figure 4 Iterative flow of maximum power estimation

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 shows that, if we use the same approach for finite population as for the infinite population, there will be bias in 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 μ_m is estimating the maximum power of an infinite population that should have (with some probability) an even tail after the actual maximum power of the population. While 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. 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 [10], 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 μ_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 an unbiased estimator for finite populations.

IV. Experimental results

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 method (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 3 and Figure 4) remains the same.

Experimental results and discussion:

Let us give a theoretical study on 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 "qualifies 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).

The experimental setup is as follows. The population contains 160,000 randomly generated high activity (average switching activity larger that 0.3) vector pairs. Random vector generation is equivalent to simple random sampling vector pairs from the population. The whole population is simulated using Powermill [11] to get the power consumption value for each unit and in the process the actual maximum power. Our approach (*n*=30, *m*=10) and simple random sampling (SRS) have been applied to do 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 applied to do 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 comparison for the estimation quality. 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. Columns 3, 4, 5 and 6 give the results of 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.

The experimental results shows 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 confidence interval and 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 infinite population.

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	1924	23024	6.0%	0.3%
C1908	0.00015	3600	1500	2410	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	3544	60593	7.7%	1.7%
C5315	0.000194	2700	600	1653	11868	5.8%	0.8%
C6288	0.000163	900	600	676	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%

Table 1 Results for comparing the efficiency and accuracy

Circuits	Actual 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%

Table 2 Results for comparing the estimation quality

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	1054	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	1052	5391	4.5%	1.7%
C7552	0.000308	3900	900	2252	7475	8.0%	0.9%
C880	0.000135	2700	600	1703	17055	12%	2.1%

Table 3 Results for comparing efficiency and accuracy

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

Proposed method: Similar to Category I.1, except that vector pairs are generated under given constraints.

Experimental setup: Similar to 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. 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 3 and Table 4, respectively. The meaning of 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 he number of qualified units in the population decreases, the number of units needed to estimated the maximum power dissipation in the circuit increases.

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	2584	7355	5.3%	1.7%
C3540	0.000053	5100	600	3587	43444	7.4%	2.9%
C432	0.000179	3000	1500	2389	12862	6.8%	2.4%
C5315	0.000231	3600	1200	2623	9967	13%	3.4%
C6288	0.000079	6000	2700	5424	29145	5.1%	0.6%
C7552	0.000194	2400	1200	1976	16446	7.1%	3.3%
C880	0.000018	2700	900	1897	127920	5.0%	1.9%

Table 4 Results for comparing efficiency and accuracy

V. Conclusion

In conclusion, a statistical approach has been proposed based on the asymptotic theory of extreme order statistics. This is the first maximum power estimation approach which can provide confidence interval at given confidence level. This is also the first approach which can do maximum power estimation for any user-specified error and confidence levels. 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 makes it also applicable to other fields of VLSI design automation, for example, maximum delay estimation.

REFERENCES

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

[4] 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.

[5] 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, to appear.

[6] 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.

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

[8] 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.

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

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

[11] Analog Cirsuit Engine Release Notes, Synopsis, 1997.