A Monte Carlo Approach for Maximum Power Estimation Based on Extreme Value Theory

Nestoras E. Evmorfopoulos, *Student Member, IEEE*, Georgios I. Stamoulis, *Member, IEEE*, and John N. Avaritsiotis, *Senior Member, IEEE*

***Abstract*—A Monte Carlo approach for maximum power estimation in CMOS very large scale integration (VLSI) circuits is proposed in this paper. The approach is based on the largely unexploited area of statistics known as extreme value theory. Within this framework, it attempts to appropriately model the extreme behavior of the probability distribution of the peak instantaneous power drawn from the power supply bus, in order to yield a close estimate of its maximum possible value. The approach features a relatively small number of necessary input patterns that does not depend on the circuit size, user-specified accuracy, and confidence levels for the final estimate, simplicity in the algorithmic implementation, noniterative single-loop execution, highly accurate simulation-based operation, and easy integration within the design flow of CMOS VLSI circuits. Experimental results establish the above claims and demonstrate the overall efficiency of the proposed approach to address the problem of maximum power estimation.**

***Index Terms*—Extreme value theory, maximum power estimation, statistical simulation.**

I. Introduction

DURING the past few years, reliability analysis of CMOS digital circuits has drawn considerable attention and is steadily becoming a major part of the design process. The issues regarding reliability are even more important as we move into deep submicrometer where the safeguarding measures that have been used up to now (plain fabrication of reliable transistors and wires) are no longer acceptable. In order to address these issues efficiently, there is an increasing need for tools and methods that perform fast and accurate estimation of the circuit's maximum power requirements during the design phase. This has given the spark for substantial research activity in the field and has led to the development of various relevant techniques, which typically fall into the broad categories of simulative and nonsimulative approaches [1].

Nonsimulative approaches [2]–[7] attempt to create and exploit deterministic or stochastic information (such as switching activity) about the circuit's inputs, in order to perform power estimation and evaluation without explicitly simulating the circuit. While they have proved extremely fast in their operation, they make some simplifying assumptions that often result in loss of accuracy. In addition, their speed and algorithmic efficiency generally degrade with large or complex circuit structures. These particular shortcomings may prove crucial for their successful integration within the design process of the next-generation VLSI systems, as the ongoing decrease of feature size and dimensions coupled with the increase of circuit complexity and packing density render the reliability problems even more pronounced and therefore demand the maximum achievable accuracy in power analysis. To the best of our knowledge, no such method of nonsimulative nature is capable of providing the accuracy that is actually needed for maximum power estimation in deep submicrometer ICs.

Simulative approaches [8]–[10], on the other hand, are based on simulation of the circuit under typical input vectors. They have an edge over their nonsimulative counterparts in terms of precision because highly accurate transistor or gate-level simulators can be used in order to handle general gate delay models and arbitrary circuit structures. In addition, they can capture second-order effects such as finite slew rate, spurious intermediate transitions (glitches or hazards), and reconvergent fanout nodes, which are typically neglected by nonsimulative approaches. Finally, they take into account signal correlations, the modeling of which can be very expensive or oversimplified in nonsimulative approaches; in fact, an entirely satisfactory solution for a wide range of circuit types has yet to appear. However, simulative approaches do possess a couple of drawbacks. The most important is that output results are highly pattern dependent since they are a function of the input vectors. As a consequence, two simulation sessions with even the slightest difference in the input vectors may exhibit large deviations in their outcomes. Moreover, in contrast to nonsimulative approaches which implicitly cover numerous patterns at once, the selected vectors must be applied sequentially and this imposes a significant timing overhead.

Since it is prohibitively expensive to examine the entire set of all possible input vectors, we have to resort to statistical techniques, which transfer the burden from a large population of units to a much smaller sample, in order to overcome the aforementioned problems. The objective here is to obtain a reasonable estimate of the circuit's maximum power through efficient statistical analysis of a sample of randomly drawn (or computer-generated) input vectors from the entire population. A key point in this analysis is the use of *asymptotic* properties and distributions (as established through appropriate limit theorems) that can be safely assumed to hold for samples of moderate size, in order to extract results that are independent of the underlying

Manuscript received March 7, 2001; revised September 6, 2001 and December 15, 2001. This work was supported in part by Intel Corporation under an equipment grant. This paper was recommended by Associate Editor E. Macii.

N. E. Evmorfopoulos and J. N. Avaritsiotis are with the Department of Electrical and Computer Engineering, National Technical University of Athens, Athens 15773, Greece (e-mail: abari@cs.ntua.gr).

G. I. Stamoulis was with Intel Corporation, Santa Clara, CA 95052 USA. He is now with the Department of Electrical and Computer Engineering, Technical University of Crete, Chania 73100, Greece.

Publisher Item Identifier S 0278-0070(02)02474-0.

statistical model which is, most of the time, unknown. In general, methods which are based on a computer's ability to generate *random* patterns and subsequently use them to extract information about a much bigger (practically infinite) population, are known as *Monte Carlo* methods [11].

Although statistical techniques seem to offer the means to address the particular estimation problem we are dealing with, they have not seen a widespread application as yet, mainly due to the inadequacy of common statistical distributions to model extreme situations such as in maximum power analysis. Extreme value distributions and the corresponding theory, which have only lately started to appear in engineering applications, suggest a viable alternative in this respect. Within this framework, the approach adopted in the current paper attempts to appropriately model the extreme behavior of the probability distribution of the peak instantaneous power drawn from the power supply bus, in order to yield a close estimate of its maximum possible value.

An approach based on the same theoretical framework was previously proposed in [12]. Although the idea behind it was novel, the implementation was not optimal regarding a couple of points, and thereby one of the objectives of the current approach is to improve upon these points. In particular, we have the following.

- Only one of the two major disciplines of extreme value theory (maxima) was considered for the general case of estimation in [12]. However, as we will see later, the other discipline (exceedances) that is followed here exhibits superior characteristics for the specific problem of estimation of a maximum, in terms of the problem order and the quality of the selected subsample.
- The approach in [12] employed a time-consuming iterative procedure of successive approximation in order to perform maximum power estimation subject to a certain level of accuracy (which is taken equal to the relative estimation error). On the other hand, the current approach achieves the same goal using a number of units that is determined *up-front*, on the basis of a *theoretically* established confidence interval for the estimated value. Moreover, the total number of units needed by [12] to carry out the whole iterative procedure is considerably larger than the equivalent number of units given by the theoretical confidence interval for the same level of accuracy because the practical distribution which is used to construct the confidence interval in [12] is characterized by one additional sampling session and a larger variance compared to its theoretical counterpart (a fact that was also acknowledged in [12]).

Apart from the above, the main drawback of [12] is that it overlooks an important special case of extreme value theory which cannot be addressed efficiently by the conventional treatment of the general case, as this would result in either an arbitrarily large estimation error or an arbitrarily large number of units needed to bring this error to acceptable levels. It must be pointed out that this special case is *not* merely of academic interest (as one might think) but, quite the opposite, it is generally the rule and not the exception in the vast majority of practical situations. In particular, for the current problem which is of interest, our experimental results show that this case actually characterizes the 90% (nine out of ten) of the circuits that were examined. Therefore, the second (but essentially primary) objective of the current paper is to deal with this special case of extreme value theory in an efficient and theoretically sound way. A step toward a thorough approach to the maximum power estimation problem was made in [13], but the treatment there was on the heuristic side as it was based on information acquired from statistical plots. The approach proposed here is not only rigorous and based on a solid theoretical foundation, but it also introduces some completely new and original concepts to the underlying theory.

The distinctive features of the proposed approach are the following.

- The approach is simulation based, and consequently it overcomes the limitations imposed by simplified gate delay models, second-order effects, and signal correlations.
- The approach is independent of the particular simulation program employed and therefore compatible to any design flow. This is further enhanced due to the overall simplicity of the algorithm.
- Under a given confidence level, the approach is capable of providing the error in the final estimate (in the form of a confidence interval) for a fixed number of units or, what is more useful and frequently wanted, the number of units required to obtain an estimate for which the error approximates a fixed predefined level. In the latter case, as already mentioned, the number of units is determined up-front and does not need time-consuming iterative procedures.
- The number of units that need to be drawn from the population to attain a desired level of accuracy is independent of the circuit size. This property, known as *dimension independence*, is a most attractive additional feature that is generally pursued in Monte Carlo implementations. An important consequence of this is that the approach is of *polynomial* time, meaning that its execution time is roughly in the same order for all circuits, either small or large (actually, the execution time may increase slightly for some very large circuits due to the computational needs of the simulation program to handle them, but this has nothing to do with the approach itself). At the very opposite, most competing approaches impose a computational load of input units that grows exponentially with circuit size and may prove prohibitively expensive to apply to large circuits.

The only major limitation of the proposed approach comes as a result of its statistical nature since, however smart and efficient use of the theory it makes, its performance always depends on the quality of the acquired sample and to what extent it can be representative of the population.

The rest of the paper is organized as follows. Section II presents an overview of maximum power and its impact on reliability of CMOS very large scale integrated (VLSI) circuits. Section III establishes the mathematical framework of extreme value theory and provides the fundamental concepts that will form the basis of the subsequent steps. Section IV introduces

some additional and unpublished concepts to extreme value theory, which are essential to construct the overall estimation approach. Their presentation was chosen to be brief so as not to impair readability and the full details are given in Appendices A, B, and C, for the interested reader. Section V describes the proposed maximum power estimation procedure and justifies any decisions made. Section VI presents our experimental results together with an empirical comparison to previous work and discusses some important facts that were derived. Finally, Section VII summarizes the significant points quoted throughout the paper and makes suggestions on potential future work and applications.

II. Maximum Power in CMOS VLSI Circuits

Excessive instantaneous power dissipation is a potential source of many reliability and performance problems. Certain issues of major importance today are physical failures due to overheating or electromigration and various functional failures, along with reduced noise margins and performance degradation, mainly caused by voltage drop (or IR-drop). In particular, voltage drop on power rails lowers the circuit's effective supply voltage and imposes an adverse effect on noise immunity and switching speed, or even worse, generates faulty logic signals that may cause circuit malfunction. These problems become even more pronounced in the deep-submicrometer domain where the feature size and power supply levels are considerably smaller. The maximum value of voltage drop depends, in general, on the peak of the current waveform and the effective resistance of the power grid. Thus, the need for early estimation of maximum power is crucial, not only to prevent reliability problems, but also to perform efficient design and optimization of the power distribution network.

The term *maximum power* refers to the worst case instantaneous power (or equivalently current) that is drawn from the supply bus. In order to render the estimation problem precise, we have to consider the sources of power dissipation in digital CMOS circuits. These are $P_{\rm sw}$, $P_{\rm sc}$, and $P_{\rm lk}$, which correspond to switching, short-circuit, and leakage power, respectively [14]. The first two quantities are the dynamic components of total power since they only occur at transitions between logic states, whereas the third one is the static component as it is a constant source of dissipation. The simulation-based nature of our approach allows all three components to be taken into consideration. In contrast, only $P_{\rm sw}$ is typically considered by nonsimulative approaches, even though $P_{\rm sc}$ also constitutes a significant fraction of total power, especially in deep submicrometer circuits. The $P_{\rm lk}$ component, because of its static nature, may be regarded as an offset to the total instantaneous power, while at the same time it is two to three orders of magnitude smaller than the other two components. It is thus clear that the worst case power conditions arise strictly on transition intervals and that any corresponding analysis should be restricted inside a clock cycle which marks the transition between two consecutive input vectors. Because of this, maximum power estimation is also referred to as *cycle-accurate* power estimation.

Within this context, estimation of maximum power is a fundamentally different problem than the conventional one of average power estimation. The former is related to the maximum value, among all possible vector pairs, of the *peak* instantaneous power consumed within a clock cycle, whereas the latter involves the mean power of a sequence of input patterns applied over a large (theoretically infinite) number of clock cycles that cover an extended period of time. Also, from a statistical point of view, estimation of average power concentrates on the *center* of an unknown probability distribution and therefore involves *averaging* operations that lead to samples which are modeled by normal (Gaussian) distributions, as a direct consequence of the well-known *central* limit theorem. On the other hand, however, the interest during maximum power estimation is focused on the *tail* of the unknown distribution where other types of distributions, established through the less famous *extreme value* limit theorems, are applicable. Consequently, existing Monte Carlo statistical techniques [15], though proven successful for average power estimation, are not adequate to handle maximum power and thus a radically different approach, exploiting the aforementioned extreme value distributions, had to be adopted in this paper to deal with the problem.

III. Mathematical Framework

A. Preliminaries and Notation

Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be a sample of size n from a given *infinite* population with cumulative distribution function (cdf) $F(x)$, where $F(x)$ is assumed to be *continuous and differentiable*. The units X_i, $(i = 1, 2, \ldots, n)$ are drawn in random so that they constitute *independent and identically distributed* (*iid*) random variables with common cdf equal to $F(x)$. A new sequence of the ordered units $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ is formed by subsequently rearranging them in an increasing order of magnitude. The pth term $X_{p:n}$ of this new sequence is often called the pth *order statistic* of the sample.

Extreme value theory is particularly interested in the upper (and in some cases lower) order statistics, also known as extreme order statistics or simply extremes, as they convey important information about the behavior of $F(x)$ in the neighborhood of its (right) tail. There are two possible methods of modeling extreme behavior on the basis of upper order statistics. The first one is to acquire k samples of size m and create another sample $\underline{X}_{mx}$ which is composed of the *maxima* units (or the mth order statistics) $X_{m:m} = \max(X_1, X_2, \ldots, X_m)$ from each of the initial samples. The other method is to pick the r largest units (or the r upper order statistics) $X_{n-r+1:n} \leq X_{n-r+2:n} \leq \cdots \leq X_{n:n}$ from a sample of size n and consider them as a separate sample $\underline{X}_{ex}$ of the *exceedances* over a predetermined high threshold u which is taken equal to the order statistic $X_{n-r:n}$.

Theorem 1 [16]: The maxima units from multiple samples of size m with parent cdf $F(x)$ follow a distribution with cdf $F_{X_{m:m}}(x)$ that is given by

$$\begin{aligned} F_{X_{m:m}}(x) &= P\left[\max_{1\leq i\leq m} X_i \leq x\right] \\ &= P\left[X_1 \leq x, \ldots, X_m \leq x\right] = F^m(x). \end{aligned} \quad (1)$$

The exceedances over threshold u from a sample with parent cdf $F(x)$ follow a distribution with (conditional) cdf $F^{[u]}(x)$ that is given by

$$F^{[u]}(x) = P[X \le x | X > u] = \frac{P[X \le x, X > u]}{P[X > u]} = \frac{F(x) - F(u)}{1 - F(u)}, \quad x \ge u. \tag{2}$$

In practice, although some samples (one or more) are readily available most of the time, the analytic form of their parent cdf $F(x)$ is usually unknown. In that case, the above results are not useful and we have to turn to certain asymptotic (or limit) distributions of extremes that emerge for large sample sizes m and high thresholds u.

B. Extreme Value Distributions and Limit Theorems

We need first to define the concept of the *upper end point*, which plays a central role in extreme value theory and particularly in this paper.

Definition 1: The upper (or right) end point $\omega(F)$ of cdf $F(x)$ is defined as the upper bound of the support of $F(x)$

$$\omega(F) = \sup\{x : F(x) < 1\}. \tag{3}$$

The upper end point represents the maximum value that the associated random variable can acquire and becomes $\omega(F) = F^{-1}(1)$ if the random variable is bounded or $\omega(F) = +\infty$ in the opposite case.

Definition 2: Functions $G(x)$ and $W(x)$ are defined as the limit distributions (or the domains of attraction) for maxima and exceedances of cdf $F(x)$, respectively, when they satisfy

$$\lim_{m \to +\infty} \left[F^m(x) - G\left(\frac{x - a_m}{b_m} \right) \right] = 0 \tag{4}$$

$$\lim_{u \to \omega(F)} \left[\frac{F(x) - F(u)}{1 - F(u)} - W\left(\frac{x - a_u}{b_u} \right) \right] = 0 \tag{5}$$

for given *location* and *scale* parameters a_m, a_u and $b_m, b_u > 0$, which depend on $F(x)$ and on m or u. The above are typical cases of *weak* convergence, also known as convergence in law or convergence in distribution.

Theorem 2 (Limit Theorem) [17], [18]: If $F(x)$ is continuous and differentiable and the limits in (4) and (5) exist and are nondegenerate, then $G(x)$ and $W(x)$ always have the following forms (irrespective of the type of $F(x)$):

$$G(x) = \exp\left(-(1 - cx)^{1/c} \right), \quad 1 - cx \ge 0 \tag{6}$$

$$W(x) = 1 - (1 - cx)^{1/c}, \quad x \ge 0 \text{ and } 1 - cx \ge 0,\ c \ne 0 \tag{7}$$

for a specific (common) value of *shape* parameter c, which depends on $F(x)$. Furthermore, the parameters a_m, a_u and b_m, b_u are given by

$$a_m = F^{-1}\left(1 - \frac{1}{m} \right)$$

and

$$b_m = \begin{cases} -cF^{-1}\left(1 - \frac{1}{m}\right), & \text{if } c < 0 \\ c\left(\omega(F) - F^{-1}\left(1 - \frac{1}{m}\right)\right), & \text{if } c > 0 \end{cases} \tag{8}$$

$$a_u = u \text{ and } b_u = \begin{cases} -cu, & \text{if } c < 0 \\ c(\omega(F) - u), & \text{if } c > 0 \end{cases}. \tag{9}$$

The above limit theorem is the heart of extreme value theory. In effect, it demonstrates that for a large m and a high u, the samples composed of the maxima and exceedances from other samples with *any* continuous and differentiable cdf $F(x)$ approximately follow the *extreme value distributions* (6) and (7) with specific values in place of location, scale, and shape parameters.

Note that the location parameter a_u of $W(x)$ is always equal to the preselected (and thus known beforehand) threshold u. Based on this, a_u may be substituted by the fixed value of u and not be treated as a parameter in the subsequent calculations.

It is also interesting to observe that $G(x)$ and $W(x)$ are connected through the simple analytical relationship $W(x) = 1 + \log(G(x))$, which highlights the inherent interdependence between the two different methods (maxima and exceedances) of extreme value theory.

Functions $G(x)$ and $W(x)$ are not written directly in probability distribution forms in (6) and (7) and as such, the latter expressions are not very useful in practice. In order to formulate more appropriate expressions for $G(x)$ and $W(x)$, we split each of these functions into two parametric families of distributions according to the sign of parameter c as shown in (10) and (11) at the bottom of the page, or, by substituting $\mu_m = a_m + b_m/c$

$$\sigma_{m,u} = \begin{cases} -b_{m,u}/c, & \text{if } c < 0 \\ b_{m,u}/c, & \text{if } c > 0 \end{cases}$$

and

$$\beta = \begin{cases} -1/c, & \text{if } c < 0 \\ 1/c, & \text{if } c > 0. \end{cases}$$

See (12)–(15) at the bottom of the next page.

$$G\left(\frac{x - a_m}{b_m}\right) = \exp\left(-\left(1 - c\frac{x - a_m}{b_m} \right)^{1/c} \right) = \begin{cases} \exp\left(-\left(\frac{x - \left(a_m + \frac{b_m}{c}\right)}{-\left(\frac{b_m}{c}\right)} \right)^{-(-1/c)} \right), x \ge a_m + \frac{b_m}{c}, & \text{if } c < 0 \\ \exp\left(-\left(-\frac{x - \left(a_m + \frac{b_m}{c}\right)}{\left(\frac{b_m}{c}\right)} \right)^{1/c} \right), x \le a_m + \frac{b_m}{c}, & \text{if } c > 0 \end{cases} \tag{10}$$

$$W\left(\frac{x - u}{b_u}\right) = 1 - \left(1 - c\frac{x - u}{b_u} \right)^{1/c} = \begin{cases} 1 - \left(\frac{x - \left(u + \frac{b_u}{c}\right)}{-\left(\frac{b_u}{c}\right)} \right)^{-(-1/c)}, x \ge u, & \text{if } c < 0 \\ 1 - \left(-\frac{x - \left(u + \frac{b_u}{c}\right)}{\left(\frac{b_u}{c}\right)} \right)^{1/c}, u \le x \le u + \frac{b_u}{c}, & \text{if } c > 0 \end{cases} \tag{11}$$

Distributions (12)—(15) are known as the *Frechet*, *Weibull*, *Pareto*, and *Beta* families, respectively. Note that the Beta family in (15) is a special case of the common Beta cdf defined by the incomplete Beta integrals $F(x;a,b) = (1/B(a,b)) \int_0^x y^{a-1}(1-y)^{b-1}dy$, $0 < x < 1$, where we have set $b = 1$ and added location and scale parameters.

Expressions for the location and scale parameters μ_m and σ_m, σ_u of the partition of $G(x)$ and $W(x)$ into (12)–(15) also follow by applying the former substitutions to (8) and (9)

$$\mu_m = 0 \text{ and } \sigma_m = F^{-1}\left(1 - \frac{1}{m}\right), \quad c < 0 \tag{16}$$

$$\mu_m = \omega(F)$$

$$\sigma_m = \omega(F) - F^{-1}\left(1 - \frac{1}{m}\right), \quad c > 0 \tag{17}$$

$$\sigma_u = u, \quad c < 0 \tag{18}$$

$$\sigma_u = \omega(F) - u, \quad c > 0. \tag{19}$$

Theorem 3 [17]: The sign of parameter c, which characterizes the type of the extreme value distribution of a cdf $F(x)$, is determined as follows.

i) $c < 0$ *if and only if* $\omega(F) = +\infty$.
The value of c may then be inferred by computing the following equivalent limit to both (4) and (5), where the $c < 0$ part of (6)–(8) and (7)–(9) has been substituted, respectively

$$\lim_{t \to \omega(F)} \frac{1 - F(t - xt)}{1 - F(t)} = (1-x)^{1/c}, \; x > 0. \tag{20}$$

ii) $c > 0$ *if and only if* $\omega(F) < +\infty$.
The value of c may then be inferred by computing the following equivalent limit to both (4) and (5), where the $c > 0$ part of (6)–(8) and (7)–(9) has been substituted, respectively

$$\lim_{t \to \omega(F)} \frac{1 - F\left(t + x\left(\omega(F) - t\right)\right)}{1 - F\left(t\right)} = (1-x)^{1/c}, \; x < 0. \tag{21}$$

An important fact derived from this theorem is that a parent cdf with an infinite upper end point ($\omega(F) = +\infty$) can only have an extreme value distribution with $c < 0$, whereas a cdf with a finite upper end point ($\omega(F) < +\infty$) suggests an extreme value distribution with $c > 0$.

Since parameter c cannot be strictly equal to zero, the potential case $c = 0$ has not been considered in the first place. However, it is perfectly possible for c to be *approximately* equal to zero, and thereby two important extreme value distributions emanate from (10) and (11) by assuming that $c \to 0$ (where the limit) may be taken either increasingly (i.e., $c \uparrow 0$) or decreasingly (i.e., $c \downarrow 0$), as both cases $c < 0$ and $c > 0$ of (10) or (11) lead to

$$G_0(x;\mu_m,\sigma_m) = \exp\left(-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right), \quad -\infty < x < +\infty \tag{22}$$

$$W_0(x;\sigma_u) = \begin{cases} 0 & \text{if } x < u \\ 1 - \exp\left(-\frac{x-u}{\sigma_u}\right) & \text{if } x \geq u \end{cases} \tag{23}$$

where $\mu_m = a_m$ and $\sigma_{m,u} = b_{m,u}$. Distributions (22) and (23) are the *Gumbel* and *Exponential* families, respectively. These distributions effectively constitute an asymptotic result within another (more general) asymptotic result, in the sense that after functions $G(x)$ and $W(x)$ of (6) and (7) are generally established as the adequate limit distributions satisfying (4) and (5), respectively, they are found to exhibit a certain asymptotic behavior for $c \to 0$ due to their specific functional form.

Because no parameter substitutions were made during the extraction of (22) and (23), the corresponding location and scale parameters μ_m and σ_m, σ_u remain as originally quoted in (8) and (9). Since, however, parameters $\sigma_m = b_m$ and $\sigma_u = b_u$ are *scale* parameters, they cannot be used in any practical application in the forms (8) and (9) where a zero-approaching parameter c appears as a multiplicative factor. Fortunately, this inconvenience may be alleviated by introducing appropriate alternatives σ_m^* and σ_u^* which, as pointed out in [17], can safely substitute σ_m and σ_u as long as they satisfy $\lim_{m \to +\infty} \sigma_m^*/\sigma_m = 1$ and $\lim_{u \to \omega(F)} \sigma_u^*/\sigma_u = 1$, respectively. Specifically for the maxima method, a suitable parameter σ_m^* to be used in place of $\sigma_m = b_m$ of (8) for $c \to 0$ is the following [17]:

$$\sigma_m^* = m \int_{F^{-1}(1-1/m)}^{\omega(F)} (1 - F(y))\, dy \tag{24}$$

which can be proved to satisfy

$$\lim_{m \to +\infty} \frac{m \int_{F^{-1}(1-1/m)}^{\omega(F)} (1 - F(y))\, dy}{\lim_{c \uparrow 0} \left(-cF^{-1}\left(1 - \frac{1}{m}\right)\right)} = 1, \text{ if } c < 0$$

$$G_-(x;\mu_m,\sigma_m,\beta) = \begin{cases} 0, & \text{if } x < \mu_m \\ \exp\left(-\left(\frac{x-\mu_m}{\sigma_m}\right)^{-\beta}\right), & \text{if } x \geq \mu_m \end{cases} \tag{12}$$

$$G_+(x;\mu_m,\sigma_m,\beta) = \begin{cases} \exp\left(-\left(-\frac{x-\mu_m}{\sigma_m}\right)^{\beta}\right), & \text{if } x \leq \mu_m \\ 1, & \text{if } x > \mu_m \end{cases} \tag{13}$$

$$W_-(x;\sigma_u,\beta) = \begin{cases} 0, & \text{if } x < u \\ 1 - \left(\frac{(x-u)+\sigma_u}{\sigma_u}\right)^{-\beta}, & \text{if } x \geq u \end{cases} \tag{14}$$

$$W_+(x;\sigma_u,\beta) = \begin{cases} 0, & \text{if } x < u \\ 1 - \left(-\frac{(x-u)-\sigma_u}{\sigma_u}\right)^{\beta}, & \text{if } u \leq x \leq u + \sigma_u \\ 1, & \text{if } x > u + \sigma_u \end{cases} \tag{15}$$

and

$$\lim_{m\to+\infty}\frac{m\int_{F^{-1}(1-1/m)}^{\omega(F)}(1-F(y))\,dy}{\lim_{c\downarrow 0}c\left(\omega(F)-F^{-1}\left(1-\frac{1}{m}\right)\right)}=1,\ \text{if } c>0.$$

The purpose of creating a separate pair of extreme value distributions for $c\to 0$ is that if $F(x)$ is such that parameter c is close to zero, it might sometimes be convenient (or even mandatory) to assume that the samples of maxima and exceedances approximately follow (22) and (23) instead of (12), (13) and (14), (15), respectively. Events and conditions under which this assumption is both advantageous and valid will be seen in practice later. For the sake of completeness now, a theoretical condition to decide whether $c\to 0$ for a given parent cdf $F(x)$ may be derived from either (20) or (21), both of which lead to the following expression for $c\to 0$ [17]:

$$\lim_{t\to\omega(F)}\frac{1-F\left(t+x\frac{\int_t^{\omega(F)}(1-F(y))dy}{1-F(t)}\right)}{1-F(t)}=\exp(-x). \quad (25)$$

Definition 3: Two distribution functions $F(x)$ and $H(x)$ are referred to as *right tail equivalent* if $\omega(F)=\omega(H)$ and $\lim_{x\to\omega(F)}(1-F(x))/(1-H(x))=1$.

Theorem 4 [19]: Distribution functions $F(x)$ and $H(x)$ are right tail equivalent if and only if their extreme value limit distributions (for either maxima or exceedances) exist and coincide, i.e., they have the same parameters

$$\lim_{m\to+\infty}\left[F^m(x)-\exp\left(-\left(1-c\frac{x-a_m}{b_m}\right)^{1/c}\right)\right]$$
$$=\lim_{m\to+\infty}\left[H^m(x)-\exp\left(-\left(1-c\frac{x-a_m}{b_m}\right)^{1/c}\right)\right]=0$$
$$\lim_{u\to\omega(F)}\left[\frac{F(x)-F(u)}{1-F(u)}-1+\left(1-c\frac{x-u}{b_u}\right)^{1/c}\right]$$
$$=\lim_{u\to\omega(F)}\left[\frac{H(x)-H(u)}{1-H(u)}-1+\left(1-c\frac{x-u}{b_u}\right)^{1/c}\right]=0.$$

The practical implication of this theorem is that *any* common cdf $H(x)$ with the same extreme value distribution parameters as those observed for an unknown cdf $F(x)$ may replace it in extreme value problems without affecting the associated properties and particularly the upper end point. Note that the above is also valid in the occasion where $c\to 0$ (an exclusive reference for this case was made in [19]), as long as $F(x)$ and $H(x)$ lie on the same side with respect to the sign of c, i.e., their extreme value distributions have both $c\uparrow 0$ or both $c\downarrow 0$.

C. Sample-Based Parameter Estimation

Definition 4: Let $\underline{X}=(X_1,X_2,\ldots,X_n)$ be a sample of size n from a population with parametric cdf $F(x;\underline{\theta})$ and probability density function (pdf) $f(x;\underline{\theta})=\partial F(x;\underline{\theta})/\partial x$, where parameter $\underline{\theta}=(\theta_1,\theta_2,\ldots,\theta_p)$ is generally a vector. The joint density of the n units of the sample $L_f(\underline{\theta})=\prod_{i=1}^n f(X_i;\underline{\theta})$ and its logarithm $\log L_f(\underline{\theta})=\sum_{i=1}^n\log f(X_i;\underline{\theta})$, considered as a function of $\underline{\theta}$, are called the likelihood and log-likelihood functions of the sample, respectively. The value $\hat{\underline{\theta}}$ of $\underline{\theta}$ that maximizes $\log L_f(\underline{\theta})$ is the *maximum likelihood* (ML) estimate of $\underline{\theta}$ and is a random variable itself.

Theorem 5 [20]: Under certain general conditions known as regularity conditions, the ML estimate $\hat{\underline{\theta}}$ of parameter $\underline{\theta}$ is unique, uniformly unbiased with minimum variance, and asymptotically (i.e., for $n\to+\infty$) follows a multivariate normal distribution with mean $\underline{\theta}$ and covariance matrix $\Sigma=1/nI^{-1}$, where I is the (*Fisher*) *information matrix* with the following entries evaluated at $\hat{\underline{\theta}}$:

$$e_{ij}=E\left[\frac{\partial\log f(X;\underline{\theta})}{\partial\theta_i}\frac{\partial\log f(X;\underline{\theta})}{\partial\theta_j}\right]$$
$$=-E\left[\frac{\partial^2\log f(X;\underline{\theta})}{\partial\theta_i\partial\theta_j}\right]$$
$$=-\int_{-\infty}^{+\infty}\frac{\partial^2\log f(x;\underline{\theta})}{\partial\theta_i\partial\theta_j}f(x;\underline{\theta})dx$$
$$(i,j=1,2,\ldots,p). \quad (26)$$

IV. Contributions to Extreme Value Theory

A. Introductory Discussion

It is evident from Theorem 2 that the set of possible (asymptotic) extreme value distributions, with parameter c as the independent variable, spans the union $(-\infty,0)\cup(0,+\infty)$ of two one-dimensional sections of the real line. The location (i.e., the value of c) of the extreme value distribution of a particular parent cdf $F(x)$ within this set is related to the rate with which the right tail of $F(x)$ approaches 1 as $x\to\omega(F)$ (rate of ascent). This rate starts slow and smooth for $c\to-\infty$, with the right tail of $F(x)$ going to infinity ($\omega(F)=+\infty$), and then gradually rises with increasing values of c to become ultimately very steep for $c\to+\infty$, while $F(x)$ now ends at a finite point $\omega(F)<+\infty$ (Fig. 1). Halfway at the center, where the discontinuity point $c=0$ lies, the transition from $\omega(F)=+\infty$ to $\omega(F)<+\infty$ takes place. This central position is occupied by the Gumbel/Exponential family being the asymptotic result for $c\to 0$ and, as already demonstrated in Section III-B, constitutes a potential domain of attraction for parent distributions from both sides i.e., with either infinite (when $c\uparrow 0$) or finite (when $c\downarrow 0$) upper end points. We have some strong reasons to believe that the latter is actually the case for the *majority* of empirical distributions encountered in engineering applications. First of all, theory itself proves that the two undoubtedly most important and common probability distributions, namely the *normal* and *gamma* families, have extreme value distributions with $c\to 0$[17], [21]. On top of this, since the value of parameter c of an extreme value distribution is determined exclusively by the *overall* shape of the right tail of its parent distribution, it follows that all other distributions with similar right tails to the previous ones will also exhibit $c\to 0$. Common practical experience confirms that indeed most empirical distributions approximate the tails of certain normal or gamma distributions (even if they do not come from one of these families), as they are neither long tailed [Fig. 1(a)] nor short tailed [Fig. 1(c)–1(d)] but rather lie somewhere between [Fig. 1(b)]. To further support this, experiments performed on various sets

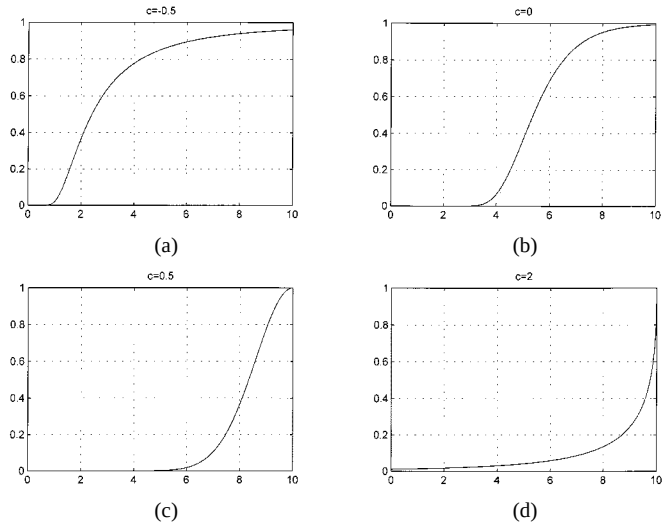


Fig. 1. Graphs of the extreme value distribution function $G(x)$ for certain values of parameter c, showing the rising rate of ascent at the right tail with increasing c. For most empirical distributions, the rate of ascent and the shape of the right tail are similar to (b).

of engineering data in [21] (which deals exclusively with engineering applications of extreme value theory), resulted in seven out of eight situations yielding extreme value distributions with $c \to 0$. This prevalence of the family of extreme value distributions with $c \to 0$ was confirmed for the particular problem at hand, as shown by our experimental results in Section VI.

As will become clear later, the problem that is being investigated in the current paper may be reduced to the equivalent problem of estimation of a *finite* upper end point $\omega(F)$. Theorem 3 then implies that the parent cdf $F(x)$ falls into the extreme value case $c > 0$, and thus one of distributions (13) or (15), in conjunction with parameters (17) or (19), should be adopted for the estimation of $\omega(F)$. However, we will argue in the next section that if c approaches zero (decreasingly) for this particular cdf $F(x)$ (which, as previously stated, happens in most practical situations), then $\omega(F)$ cannot be efficiently or accurately estimated through the previous set of distributions and parameters constructed for the general case $c > 0$ and we need to exploit one of distributions (22) or (23) with corresponding parameters, which are related to the special case $c \downarrow 0$. Consequently, a thorough procedure of estimation of $\omega(F)$ when it is finite must separately take into account the cases $c > 0$ and $c \downarrow 0$. This problem is not adequately addressed in the literature and will be the subject of the rest of Section IV.

B. Estimation of the Upper End Point in the General Case $c > 0$

This case is straightforward as, depending on the extreme value method that we choose to follow (maxima or exceedances), we may obtain an estimate $\hat{\omega}(F)$ of the upper end point from either (17) or (19)

$$\hat{\omega}(F) = \hat{\mu}_m \tag{27}$$

$$\hat{\omega}(F) = \hat{\sigma}_u + u \tag{28}$$

where $\hat{\mu}_m$ and $\hat{\sigma}_u$ are the ML estimates of parameters μ_m and σ_u as extracted from the samples of maxima and exceedances, respectively. For reasons that will be made clear in Section V, we will proceed with the exceedances method and estimate $\omega(F)$ through (28). The pdf of the Beta family (15), which asymptotically models the sample of exceedances when $c > 0$, is given by

$$w_+(x;\sigma_u,\beta) = \frac{\beta}{\sigma_u}\left(-\frac{(x-u)-\sigma_u}{\sigma_u}\right)^{\beta-1} \quad u \le x \le u+\sigma_u. \tag{29}$$

The corresponding log-likelihood function of a Beta-distributed sample of size r is

$$\log L_{w_+}(\sigma_u,\beta) = \sum_{i=1}^{r}\left(\log\frac{\beta}{\sigma_u} + (\beta-1)\log\left(-\frac{(X_i-u)-\sigma_u}{\sigma_u}\right)\right). \tag{30}$$

As established in Section III-C, maximization of (30) with respect to σ_u and β yields estimates $\hat{\sigma}_u$ and $\hat{\beta}$ which generally follow normal distributions with known parameters given by Theorem 5. However, due to some regularity problems appearing when the actual value of β is less than or equal to two (details are cited in Appendix A), $\hat{\sigma}_u$ is asymptotically normal only if $\beta > 2$. Assuming that this holds true (see Appendix A for the contrary) and since u is constant with $\mathrm{var}(u) = 0$, it follows from (28) that $\hat{\omega}(F)$ is also normally distributed with mean $\omega(F)$ and variance equal to the variance of $\hat{\sigma}_u$. The latter is computed to be (see Appendix A)

$$\mathrm{var}\left(\hat{\omega}(F)\right) = \mathrm{var}(\hat{\sigma}_u) = \frac{1}{r}\hat{\sigma}_u^2\frac{(\hat{\beta}-2)(\hat{\beta}-1)^2}{\hat{\beta}}. \tag{31}$$

Since $\hat{\omega}(F)$ is normally distributed with known mean and variance, the confidence interval that corresponds to a confidence level of $(1-\delta)\times 100\%$ is [20]

$$|\hat{\omega}(F) - \omega(F)| \le \frac{z_{\delta/2}}{\sqrt{r}}\hat{\sigma}_u(\hat{\beta}-1)\sqrt{\frac{\hat{\beta}-2}{\hat{\beta}}} \tag{32}$$

where $z_{\delta/2}$ is the $\delta/2$ quantile point of the standard normal distribution $N(0,1)$.

In the occasion where $c \downarrow 0$, we have $\beta = 1/c \to +\infty$ and the variance of the estimate in (31) tends to infinity (the same also applies if the maxima instead of the exceedances method were followed). Obviously, this is an undesirable effect since, as dictated by statistical theory, small variance is one of the basic features that an estimate must carry (recall from Theorem 5 that the previous ML estimate exhibits the minimum variance among all possible estimates) [20]. From another point of view, which exploits the notion of the confidence interval, the absolute estimation error $|\hat{\omega}(F) - \omega(F)|$ lies (with probability $1-\delta$) inside an interval of nearly infinite size, and therefore the deviation of the estimated $\hat{\omega}(F)$ from the actual $\omega(F)$ may turn out to be extremely large. Of course one can always counterweigh this effect by increasing the size r of the sample of exceedances, but this would require a huge number of units in order to attain an acceptable estimation error, which is hardly of any practical value. The moral is that the above procedure proves to be either inefficient or inaccurate for the estimation of $\omega(F)$ when $c \downarrow 0$ and an alternative one, which is specifically tailored to this latter case, must be formulated instead.

C. Estimation of the Upper End Point in the Special Case $c \downarrow 0$

In a similar way as in the preceding section, it is possible to express the upper end point as a function of the pertinent extreme value distribution parameters and then obtain an estimate $\hat{\omega}(F)$ by substituting the corresponding ML parameter estimates. However, the current case is considerably more complicated than the previous one and a thorough approach has yet to appear in the literature. We notice at first that the lack of a parametric expression of $\omega(F)$ in the exceedances approach renders the maxima method, for which $\omega(F)$ appears as a parameter in (24), our only option when $c \downarrow 0$. Since, however, we do not possess analytic knowledge of the parent cdf $F(x)$ which is required in order to exploit (24), we have to turn to the tail equivalence principle (Theorem 4) and use it to replace $F(x)$ with an appropriate cdf $H(x)$ in the subsequent calculations (bear in mind that this substitution is confined to the *tail* of $F(x)$ and it is by no means assumed that the two distributions coincide). In particular, $H(x)$ can be any location–scale parametric cdf that features a finite upper end point and an extreme value distribution with $c \downarrow 0$. Certain candidate functions with the aforementioned properties have been examined and the most convenient out of them proved to be the following [17], [21]:

$$H(x) = 1 - \exp\left(-\frac{\sigma_H^2}{(x-\mu_H)^2}\right) \tag{33}$$

where μ_H and $\sigma_H > 0$ are, as usual, the location and scale parameters, while the upper end point $\omega(H)$ is equal to μ_H. Theorem 4 guarantees that $H(x)$ and $F(x)$ will have *identical* upper end points, i.e., $\omega(F) = \omega(H) = \mu_H$, if and only if their extreme value distribution parameters $\mu_m = a_m$ and σ_m in (8) and (24) are equal. Therefore, our next step is to express μ_H and σ_H with respect to these parameters so that, for given values of μ_m and σ_m (either known or estimated from the sample of maxima for $F(x)$), we may compute the values of μ_H and σ_H which lead to an extreme value distribution for $H(x)$ with the same given parameters. The estimate $\hat{\omega}(F)$ for $c \downarrow 0$ then follows directly from the derived expression of μ_H. After the appropriate calculations (details are cited in Appendix B), we end up with the following:

$$\hat{\omega}(F) = \hat{\mu}_H = \hat{\mu}_m + \frac{\hat{\sigma}_m}{1 + m\sqrt{\pi \log m}\left(\operatorname{erf}(\sqrt{\log m}) - 1\right)} \tag{34}$$

where $\hat{\mu}_m$ and $\hat{\sigma}_m$ are the ML estimates of μ_m and σ_m, respectively, and $\operatorname{erf}(x)$ is the *error function* (see Appendix B). The pdf of the Gumbel family (22), which asymptotically models the sample of maxima when $c \downarrow 0$, is

$$g_0(x; \mu_m, \sigma_m) = \frac{1}{\sigma_m} \exp\left(-\frac{x-\mu_m}{\sigma_m}\right) \cdot \exp\left(-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right), \quad -\infty < x < +\infty \tag{35}$$

and the corresponding log-likelihood function of a Gumbel-distributed sample of size k is

$$\log L_{g_0}(\mu_m, \sigma_m) = -\sum_{i=1}^{k}\left(\frac{X_i - \mu_m}{\sigma_m} + \exp\left(-\frac{X_i - \mu_m}{\sigma_m}\right) + \log \sigma_m\right). \tag{36}$$

In this case, ML estimation poses no regularity problems and therefore, both $\hat{\mu}_m$ and $\hat{\sigma}_m$ obtained by maximization of (36) are asymptotically normal. The variance of the estimate is then given by (see Appendix B) (37), as shown at the bottom of the page, where γ is the *Euler gamma* constant (see also Appendix B). The confidence interval corresponding to a confidence level of $(1-\delta) \times 100\%$ finally becomes [20]; see (38), also shown at the bottom of the page.

D. Decision on the Behavior of Parameter c From a Sample

Since the parent cdf $F(x)$ is, most likely, unknown and condition (25) is not possible to check, we need to have the means to determine whether $c > 0$ or $c \downarrow 0$ on the basis of a sample $\underline{X}$ in order to distinguish between the two cases addressed in the previous sections. This problem may be formulated in the context of *hypothesis testing*[20], which is commonplace in statistics. In particular, we wish to test the null hypothesis $H_0 : c \downarrow 0$ against the alternative hypothesis $H_1 : c > 0$. Generally, a decision between H_0 and H_1 is based on the construction of a suitable test statistic $T(\underline{X})$ of $\underline{X}$ and the formation of a critical region $C(a)$ which corresponds to a given significance level a. If $T(\underline{X})$ falls into $C(a)$, then H_0 is rejected at this particular significance level. Given the cdf $F_T(x)$ of $T(\underline{X})$ under the null hypothesis, the critical region takes the form $C(a) = \left\{\underline{X} : T(\underline{X}) \geq F_T^{-1}(1-a)\right\}$ so that the probability of $T(\underline{X})$ falling into $C(a)$ is equal to a. In Appendix C, we investigate in detail the application of these concepts onto the specific problem we are facing. The result is summarized in the subsequent test statistic which, when calculated on the r upper order statistics (exceedances) part $\underline{X}_{ex}$ of the sample $\underline{X}$, asymptotically (i.e.,

$$\operatorname{var}\left(\hat{\omega}(F)\right) = \frac{1}{k}\frac{6\hat{\sigma}_m^2}{\pi^2}\left((\gamma-1)^2 + \frac{\pi^2}{6} + \frac{2(1-\gamma)}{1 + m\sqrt{\pi \log m}\left(\operatorname{erf}(\sqrt{\log m}) - 1\right)} + \frac{1}{\left(1 + m\sqrt{\pi \log m}\left(\operatorname{erf}(\sqrt{\log m}) - 1\right)\right)^2}\right) \tag{37}$$

$$|\hat{\omega}(F) - \omega(F)| \leq \frac{z_{\delta/2}}{\sqrt{k}}\hat{\sigma}_m\frac{\sqrt{6}}{\pi}\sqrt{(\gamma-1)^2 + \frac{\pi^2}{6} + \frac{2(1-\gamma)}{1 + m\sqrt{\pi \log m}\left(\operatorname{erf}(\sqrt{\log m}) - 1\right)} + \frac{1}{\left(1 + m\sqrt{\pi \log m}\left(\operatorname{erf}(\sqrt{\log m}) - 1\right)\right)^2}} \tag{38}$$

for $r \to +\infty$) follows a standard normal distribution under the null hypothesis H_0

$$T(\underline{X}_{ex}) = \frac{\frac{(m_{\underline{X}_{ex}} - u)^2}{s^2_{\underline{X}_{ex}}} - 1}{\frac{2}{\sqrt{r}}} \sim N(0,1) \tag{39}$$

where $m_{\underline{X}_{ex}}$ and $s_{\underline{X}_{ex}}$ are the mean and standard deviation of $\underline{X}_{ex}$. Accordingly, the critical region for a one-tailed test [20] is

$$C(a) = \{\underline{X}_{ex} : T(\underline{X}_{ex}) \geq z_a\} \tag{40}$$

where z_a is, as usual, the a quantile point of the standard normal distribution $N(0,1)$.

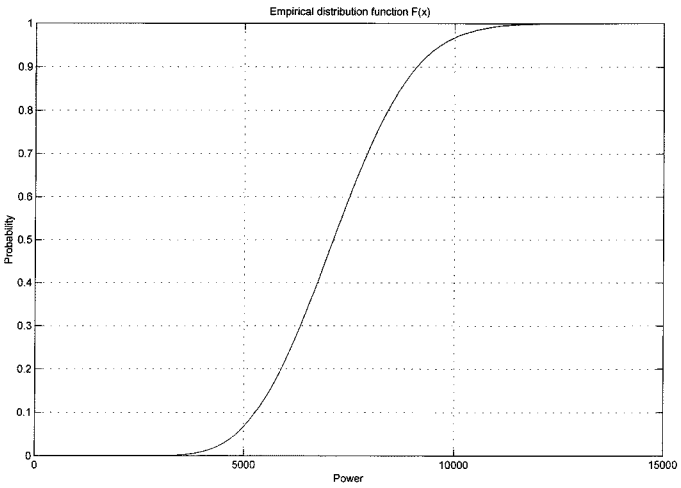
V. Estimation Procedure

A. Problem Formulation

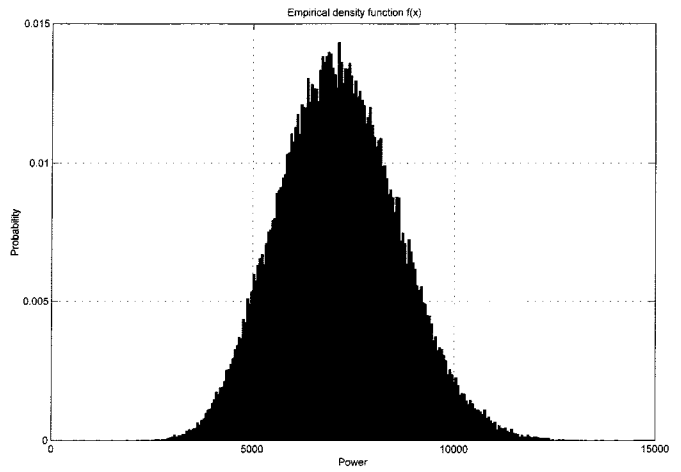
The cycle-accurate peak instantaneous power dissipation in a CMOS circuit depends upon the specific pair of the input binary vectors (v_1, v_2) that generate a transition on a particular clock edge. Maximum power is then defined as the maximum value of the peak instantaneous power among all possible vector pairs, as already cited in Section II. If the number of primary inputs is n, then an input vector consists of n bits and obviously the population of input vector pairs includes $2^n \cdot 2^n = 4^n$ units. We will restrict our analysis to combinational circuits, and furthermore we will assume that all primary inputs execute their transitions simultaneously on the same clock edge. However, the generalization to a sequential circuit is straightforward, if we expand the *previous* input vector v_1 to encompass the state bits (or outputs) of all sequential elements (flip-flops) of the circuit, following [22]. In that case, if the number of state bits is m, then the input population will contain $2^{n+m} \cdot 2^n = 4^n \cdot 2^m$ units. We should also remark that the input population may not always include all possible vector pairs but only selected ones that correspond to typical circuit operation or input switching activity requirements (constrained problem).

Having defined an input population of vector pairs, we may regard the cycle-accurate peak instantaneous power consumption of a *particular* circuit as a random variable with a certain probability distribution $F(x)$ (see Fig. 2 for an example of such a distribution). Its maximum value will then be represented by the upper end point $\omega(F)$ of $F(x)$ which, obviously, is always finite due to physical limitations. This (probability-based) aspect of the problem allows us to avoid an exhaustive search within the input population in order to locate the maximum power-consuming vector pair and, instead, formulate an estimation procedure for $\omega(F)$ by applying the pertinent maximizing operators on a *random sample* rather than the entire population. As is thereupon dictated by extreme value theory (Theorem 2), the samples created by applying maximizing operators on other samples are directly related to certain extreme value distributions of known form.

The proposed approach closely follows the concepts presented in Sections III and IV. However, these only apply under the assumptions of a continuous and differentiable $F(x)$ and an input population of infinite size, as well as the condition of



(a)



(b)

Fig. 2. Empirical cdf $F(x)$ and pdf $f(x)$ of the cycle-accurate peak instantaneous power for benchmark circuit c1355, based on a sample of 100 000 units.

a sample of power values that is comprised of iid units [each with cdf $F(x)$]. Both assumptions are most frequently justified in practice for circuits of reasonable size and with a moderate number (eight or more) of primary inputs. The additional condition is guaranteed if the sampling from the population of vector pairs is completely random so that all pairs have equal probabilities of being selected. In statistical terms, the sampling process must yield a *uniform* distribution of input vector pairs. Random sampling (or computer-based random pattern generation) is the essence of Monte Carlo methods, which generally rely on the fact that if random variables $U_1, \ldots, U_n$ are uniform, then the transformed random variables $X_1 = F^{-1}(U_1), \ldots, X_n = F^{-1}(U_n)$ follow a distribution with cdf $F(x)$ (Monte Carlo law) [11].

Before closing this section, we have to point out that when dealing with sequential circuits, a phenomena of clustering of power values may occur, leading to a probability distribution with more than one mode (multimodal). Although this is a major concern for average power estimation, as heavy concentration of units inside one mode will certainly affect the estimation results, it has no consequence in maximum power estimation. That is

because, in the latter, we are only interested in the right tail of the distribution, which always exhibits a unique upper end point that may be estimated from a sample, provided that enough units fall into the rightmost mode.

B. Procedure Description

A flow chart of the estimation procedure is depicted in Fig. 3. The procedure begins with the generation of n vector pairs in a random way or under certain constraints. This is equivalent to random sampling out of the unconstrained or constrained population of input vector pairs. The circuit due for maximum power analysis is then entered in a transistor-level or gate-level simulator such as SPICE or PowerMill. Each generated vector pair is fed as input for transient analysis to the simulator and the peak power during transition time is recorded. Special care has to be taken for the clock period to be longer than the maximum delay along any path from the inputs to the outputs of the circuit, so that there is enough time for the transition effects to spread to all internal nodes. The set of power values obtained for all generated vector pairs forms the main sample to be used hereafter. Extreme value analysis then begins with the decision on the behavior of parameter c, which is made in accordance with Section IV-D. In short, we pick the r upper order statistics from the main sample and create a sample of exceedances on which we calculate the test statistic T from (39) and compare it with the critical value z_a (for significance level a) in order to decide for or against the hypothesis that $c \downarrow 0$. Experiments performed on various random samples drawn from an Exponential distribution (see Appendix C) have resulted to a critical value of $z_a \approx 7$ which corresponds to a significance level of $a = 10^{-12}$. We subsequently follow one of two possible paths on the basis of the outcome of the previous step. The sequence of actions for each path was studied in detail in Sections IV-B and IV-C and we briefly rehearse the basic points here. If $c > 0$, we restore the previously formed sample of exceedances on which, assuming a Beta asymptotic model, we perform ML estimation of σ_u and β by maximizing (30) and then compute the maximum power estimate and its confidence interval (for a given confidence level $1-\delta$) from (28) and (32), respectively. If, on the other hand, $c \downarrow 0$, we create a sample of maxima by dividing the main sample into k subsamples of size m and by subsequently taking the maxima units out of each subsample. The remaining steps are similar to the other path as, assuming a Gumbel limit distribution on the sample of maxima, we obtain the ML estimates of μ_m and σ_m by maximization of (36) and then determine the maximum power estimate and its confidence interval (for confidence level $1-\delta$) from (34) and (38), respectively. In either case, the procedure ends with a report of the estimation results. Before closing this section, we should remark that the choices concerning the sizes of all samples (main, exceedances, maxima) that take part in the estimation procedure of Fig. 3 are discussed in Section V-D.

C. Implementation Issues

This section takes a brief look at the implementation and the computational aspects of the estimation procedure and its constituent operations. If we exclude the simulation part, the most

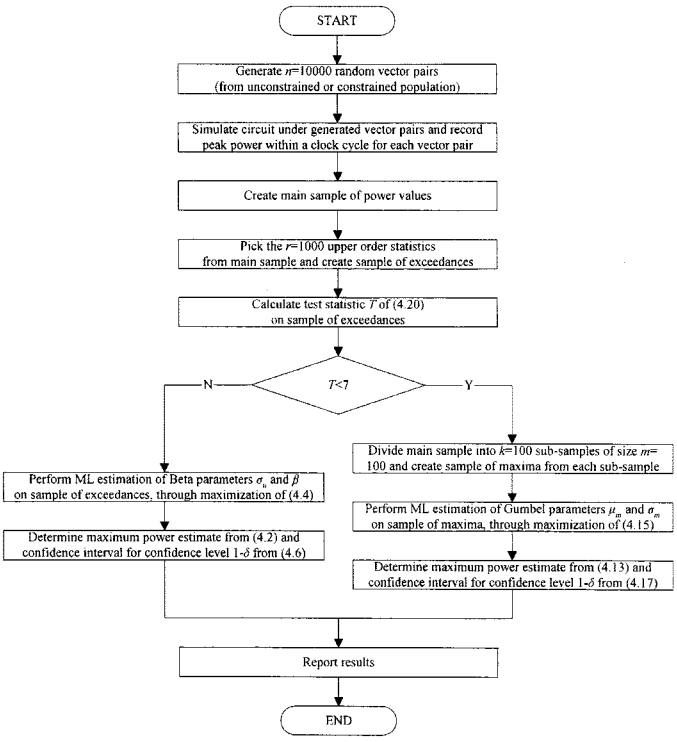


Fig. 3. Flow chart of the proposed approach for maximum power estimation.

time-consuming operation (due to its iterative nature) is the ML estimation process, while the rest are simple numerical calculations. It is, therefore, of utmost importance that ML estimation is performed only once during the whole procedure. Fortunately, the next section demonstrates that it is possible to obtain a final estimate with the desired characteristics via a single-loop execution of the procedure. This fact obviates the need to iteratively refine the confidence interval so as to meet the accuracy specifications, which would require repeated iterations of the ML estimation process.

ML estimation involves maximization (or optimization) of a scalar function of several continuous variables. In general, if the objective function to be optimized is nonlinear, then exact expressions for the values of variables that optimize it are hard or impossible to compute and *iterative* numerical methods have to be employed. All of these methods start with certain initial estimates of the variables and iteratively improve upon them until a stopping criterion is satisfied. Two broad classes of algorithms and methods for numerical optimization are available in the literature. The first class concentrates on the numerical solution of the system of equations obtained by setting the partial derivatives with respect to the parameters equal to zero. The second class attempts a direct maximization of the objective function, either by simplex search methods (such as the Nelder–Mead algorithm and its variants) or by methods that exploit gradient information (such as the steepest ascent, the secant, or the various Newton and quasi-Newton methods).

Generally, constraints have to be imposed during ML estimation, which ensure that the log-likelihood function to be optimized is real-valued. Therefore, the above methods must be adapted to the so-called *constrained* optimization problem. The assessment and implementation details of the algorithms, as

well as their adaptation to constrained optimization, are beyond the scope of this paper and the interested reader is referred to [23]–[25]. The necessary constraints for the Beta and Gumbel log-likelihood functions in (30) and (36) are

$$\sigma_u > \max_i X_i - u \text{ and } \beta > 0 \tag{41}$$

$$\sigma_m > 0. \tag{42}$$

Initial estimates for the independent variables can be obtained by equating *sample* moments $m_{\underline{X}}$ and $s_{\underline{X}}$ with *population* moments m_f and s_f. The mean m_{g_0} and the standard deviation s_{g_0} of the Gumbel distribution are given by [21]

$$m_{g_0} = \mu_m + \gamma\sigma_m \text{ and } s_{g_0} = \sigma_m \frac{\pi}{\sqrt{6}} \tag{43}$$

from which we have for the sample of maxima $\underline{X}_{mx}$

$$\sigma_{\text{m,init}} = \frac{s_{\underline{X}_{mx}}\sqrt{6}}{\pi} \text{ and } \mu_{\text{m,init}} = m_{\underline{X}_{mx}} - \gamma\sigma_{\text{m,init}}. \tag{44}$$

The corresponding quantities m_{w_+} and s_{w_+} for the Beta distribution are [16]

$$m_{w_+} = u + \frac{\sigma_u}{\beta+1} \text{ and } s_{w_+} = \frac{\sigma_u}{\beta+1}\sqrt{\frac{\beta}{\beta+2}}. \tag{45}$$

Subsequent application of the same procedure on the sample of exceedances $\underline{X}_{ex}$ yields

$$\beta_{\text{init}} = \frac{2}{\left(\frac{m_{\underline{X}_{ex}}-u}{s_{\underline{X}_{ex}}}\right)^2 - 1} \text{ and } \sigma_{\text{u,init}} = (m_{\underline{X}_{ex}} - u)(\beta_{\text{init}} + 1). \tag{46}$$

We now return to Section IV-B and provide the first of two reasons for following the exceedances method instead of the maxima method when $c > 0$ (the other and more important reason is related to the sample size and is examined in the next section). This one stems from the fact that the Beta distribution for exceedances in (15) is a function of two parameters, whereas the Weibull distribution for maxima in (13) is a three-parameter function. Thus, the overall problem order is lower in the former than in the latter case, and consequently the Beta log-likelihood function is more convenient to optimize. Generally, a two-parameter function is expected to exhibit a better and more efficient attitude to optimization algorithms in terms of speed and general convergence properties, as well as be less sensitive to numerical problems such as truncation or round-off errors. In addition, a global (instead of a local) maximum is more easily located and initial estimates have less influence in the final outcome.

D. Sample Sizes and Dimension Independence

This section focuses on the efficient choice of sizes n, r, and k of main, exceedances, and maxima samples, respectively. On this basis, it also demonstrates the dimension independence property of the proposed approach. The efficient choice of sample sizes attempts to determine the minimum number of units required to attain a desired level of accuracy in the estimated value, while at the same time ensure the validity of the asymptotic models established through (4) and (5).

We start with case $c \downarrow 0$ and (4) which, in order to hold in general, requires the size m of subsamples where maxima are taken to be large enough. The worst case is encountered for the normal distribution, which needs subsamples of about 100 units to approach its Gumbel-type limit distribution for maxima [26]. Therefore, we set $m = 100$. This permits calculation of sizes k and $n = m \cdot k$ for a desired level of accuracy reflected in the relative estimation error $\varepsilon = |\hat{\omega}(F) - \omega(F)|/\hat{\omega}(F)$. By substituting expressions (34) and (38) (where we keep the equality part of (38) that gives the maximum possible absolute error) and by subsequently solving with respect to k, we have (47), shown at the bottom of the page. It is obvious from this expression that, in order to determine the value and the general behavior of k (and n), we have to investigate the ratio $\hat{\mu}_m/\hat{\sigma}_m$ of the ML estimates or, since $\hat{\mu}_m$ and $\hat{\sigma}_m$ are *unbiased* estimates of μ_m and σ_m, the equivalent ratio μ_m/σ_m of the actual parameters of the limit distribution for maxima. The latter is more suitably written with respect to the ratio m_{g_0}/s_{g_0} of the population moments (for maxima) by using (43) as follows:

$$\frac{\mu_m}{\sigma_m} = \frac{\pi}{\sqrt{6}}\frac{m_{g_0}}{s_{g_0}} - \gamma. \tag{48}$$

Before going into detail, let us illustrate the dimension independence property of our approach. Indeed, looking at the previous set of expressions, the total number $n = m \cdot k$ of units that need to be drawn is a function of ε and δ which are fixed and the ratio m_{g_0}/s_{g_0} which typically does *not* depend on the circuit size but rather is an intrinsic characteristic of the specific population at hand (i.e., the set of vector pairs for the specific circuit under test) and its probability distribution. We have to remark here that, although the expression for n was extracted based on the assumption that $c \downarrow 0$, the *same* n that is derived from (47) and (48) by substituting the appropriate ratio of the population moments for maxima will suffice to ensure equal or greater accuracy under the given confidence level when it is actually $c > 0$. In that case, parameter $\beta = 1/c$ is small (less than four most of the time) and this implies that the variance of the estimate in (31) is much smaller than the variance in (37) which holds when $c \downarrow 0$.

$$k = \left(\frac{z_{\delta/2}}{\varepsilon}\right)^2 \frac{\frac{6}{\pi^2}\left((\gamma-1)^2 + \frac{\pi^2}{6} + \frac{2(1-\gamma)}{1+m\sqrt{\pi\log m}\left(\text{erf}(\sqrt{\log m})-1\right)} + \frac{1}{\left(1+m\sqrt{\pi\log m}\left(\text{erf}(\sqrt{\log m})-1\right)\right)^2}\right)}{\left(\frac{\hat{\mu}_m}{\hat{\sigma}_m} + \frac{1}{1+m\sqrt{\pi\log m}\left(\text{erf}(\sqrt{\log m})-1\right)}\right)^2} \tag{47}$$

Since the distribution of the population is unknown, it is not possible to calculate the precise number of required units for given accuracy and confidence levels ε and $1-\delta$ from (47) and (48). Therefore, our aim is to establish in advance an *approximate* number of units leading to a final estimate for which the *exact* confidence interval that is subsequently calculated through (32) or (38) corresponds to an accuracy figure close to the desired one. In order to achieve this, we have to rely on *empirical* evidence and approximate the ratio m_{g_0}/s_{g_0} through measurements of the sample moments $m_{\underline{X}_{mx}}$ and $s_{\underline{X}_{mx}}$ conducted on one or more test samples of maxima from the circuit under consideration. Surprisingly, a single number of units that is reasonable for a wide range of circuits may be given, as the samples of maxima in all the examined circuits have been found to exhibit the same relative magnitude of $m_{\underline{X}_{mx}}$ to $s_{\underline{X}_{mx}}$ of approximately

$$\frac{m_{\underline{X}_{mx}} - s_{\underline{X}_{mx}}}{m_{\underline{X}_{mx}}} \approx 0.95. \tag{49}$$

The above result also makes clear that m_{g_0}/s_{g_0} (and consequently n) is independent of the circuit size, as both small and large circuits were included among the ones that have been examined. Generalizing (49) on the population we have $m_{g_0}/s_{g_0} \approx 20$ and further from (48) $\mu_m/\sigma_m \approx 25.074$, which when substituted in (47) (in place of $\hat{\mu}_m/\sigma_m$) and accompanied by the appropriate calculations gives

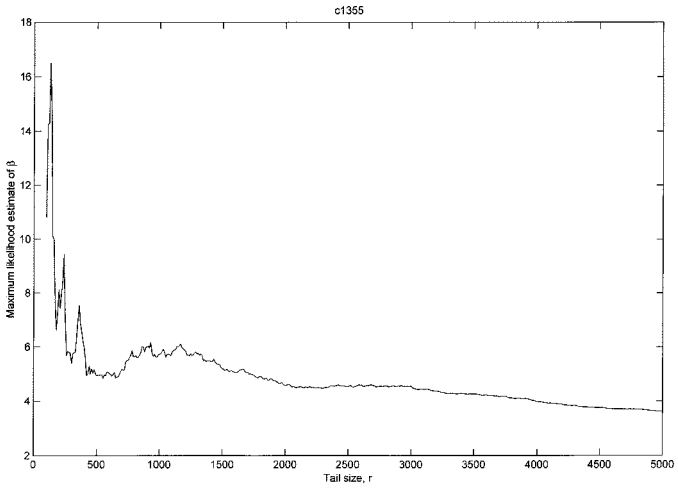
$$k \approx 0.0676 \cdot \left(\frac{z_{\delta/2}}{\varepsilon}\right)^2 \tag{50}$$

and ultimately

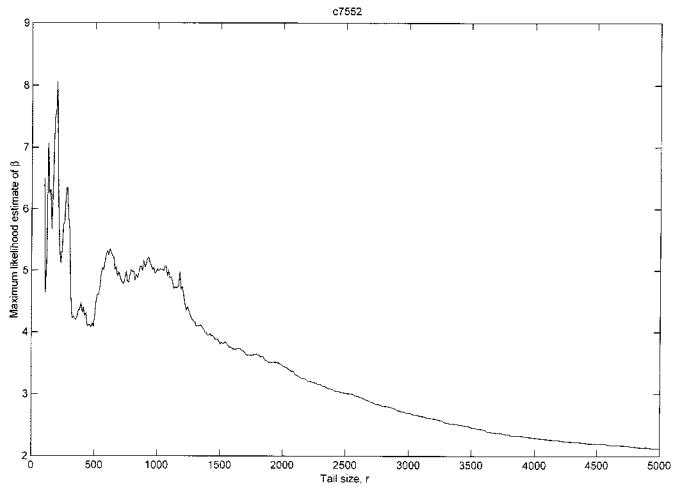
$$n = m \cdot k \approx 6.76 \cdot \left(\frac{z_{\delta/2}}{\varepsilon}\right)^2. \tag{51}$$

The latter expressions yield the approximate values $k = 100$ and $n = 10\,000$ for choices of $\varepsilon = 0.05$ and $1-\delta = 0.95$, respectively. If greater confidence and/or accuracy levels are required, then the sample sizes will be increased accordingly. We stress that (50) and (51) only hold for the range of circuits satisfying (49) and that, even though no counterexamples have been found among the available ones, other circuits (populations) may exist where this approximation is not plausible. In such cases, the accuracy specifications may not be met for the above k and n and their values have to be recalculated (the same also applies to size r discussed below). Yet one should emphasize that no matter what sample sizes are finally chosen and what approximations are made in the process, the confidence interval obtained from (32) or (38) is still the *actual* one and always gives the correct accuracy figure.

When $c > 0$, the size r is determined through a tradeoff which dictates that the upper order statistics (exceedances) must constitute a portion of the sample that is both large enough so that the ML estimates $\hat{\sigma}_u$, $\hat{\beta}$ and the test statistic (39) are approximately normal (Sections III-C and IV-D) and small enough in order for (5) to hold (recall that r is the number of units greater than the threshold $u = X_{n-r:n}$, which implies that $r \to 0$ when $u \to \omega(F)$). In strict mathematical terms, the size $r(n)$, when viewed as a function of the size n of the whole sample, must satisfy $r(n) \xrightarrow{n\to+\infty} +\infty$ and $r(n)/n \xrightarrow{n\to+\infty} 0$. Particularly for the Beta distribution in (15), a theorem from [27] sets $r(n)$ in the order of $r(n) = o(n^{2/3})$. Subsequently, we have chosen $r = 1000 (\approx 2n^{2/3})$. Our choice is visually supported by the so-called *diagrams of estimates* [16], where the ML estimates of the distribution parameters are plotted against size r. The optimal values of r lie within a "plateau" of the diagram that is surrounded by strong fluctuations on the left and a general decline on the right. Examples of these diagrams for two samples of power values from different benchmark circuits are displayed in Fig. 4, where such a plateau becomes visible around $r = 1000$.



(a)



(b)

Fig. 4. Diagrams of the estimates of parameter β for two samples of 10 000 units from benchmark circuits c1355 and c7552. Observe the plateau formation around $r = 1000$ in both diagrams.

The fact that a fair portion (10%) of the sample is considered in the tail constitutes the second reason for rendering the exceedances method superior to the maxima one when $c > 0$. Indeed, if the last method were followed, the sample of maxima would comprise no more than $k = 100$ units, which would make the final estimate sensitive to differences between samples as well as different arrangements of a particular sample into subsamples for maxima. In contrast, a sample of $r = 1000$ units in the exceedances method is certain to give a better quality in the final estimate.

TABLE I
RESULTS FOR THE AVERAGE ESTIMATED VALUES OF MAXIMUM POWER AMONG 10 RUNS

Circuit	Value of test statistic	Type of extreme model	Sample maximum (mA)	Observed maximum (mA)	Estimated maximum (mA)	Confidence interval (mA)				95% relative error
						90%	95%	99%	99.99%	
c432	4.046	I	7.081	9.053	9.407	±0.519	±0.618	±0.812	±1.227	6.6%
c499	4.430	I	16.075	17.778	21.042	±1.064	±1.267	±1.666	±2.516	6.0%
c880	4.529	I	10.492	12.254	14.592	±0.846	±1.008	±1.325	±2.001	6.9%
c1355	5.679	I	12.866	15.175	17.406	±0.890	±1.061	±1.394	±2.106	6.1%
c1908	6.974	I	12.491	14.760	17.028	±0.810	±0.965	±1.268	±1.916	5.7%
c2670	4.870	I	16.757	20.034	21.193	±0.931	±1.109	±1.457	±2.201	5.2%
c3540	4.263	I	22.208	25.774	30.928	±1.713	±2.041	±2.682	±4.051	6.6%
c5315	4.591	I	36.991	41.511	47.358	±2.103	±2.506	±3.293	±4.974	5.3%
c6288	3.475	I	42.843	46.570	54.103	±2.407	±2.869	±3.770	±5.694	5.3%
c7552	9.194	II	40.656	41.323	41.866	±1.848	±2.202	±2.894	±4.371	5.2%

TABLE II
RESULTS FOR THE MAXIMUM ESTIMATED VALUES OF MAXIMUM POWER AMONG 10 RUNS

Circuit	Value of test statistic	Type of extreme model	Sample maximum (mA)	Observed maximum (mA)	Estimated maximum (mA)	Confidence interval (mA)				95% relative error
						90%	95%	99%	99.99%	
c432	4.443	I	6.996	9.053	9.746	±0.569	±0.678	±0.891	±1.346	6.9%
c499	2.415	I	16.480	17.778	21.582	±1.131	±1.347	±1.770	±2.674	6.2%
c880	2.832	I	10.609	12.254	15.276	±0.939	±1.119	±1.471	±2.222	7.3%
c1355	3.131	I	14.072	15.175	18.546	±1.047	±1.248	±1.640	±2.477	6.7%
c1908	6.868	I	12.839	14.760	18.555	±0.992	±1.182	±1.554	±2.347	6.4%
c2670	2.258	I	17.276	20.034	22.672	±1.107	±1.319	±1.733	±2.618	5.8%
c3540	4.007	I	22.751	25.774	31.743	±1.827	±2.177	±2.861	±4.322	6.8%
c5315	3.947	I	37.655	41.511	48.608	±2.263	±2.697	±3.544	±5.353	5.5%
c6288	3.680	I	44.550	46.570	55.810	±2.647	±3.154	±4.146	±6.262	5.6%
c7552	8.277	II	41.122	41.323	42.519	±2.249	±2.679	±3.521	±5.319	6.3%

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed approach was applied on the ISCAS85 benchmark circuits, and the results for the average, maximum, and minimum estimated values, among 10 runs of the procedure of Fig. 3 for each circuit, are shown in Tables I–III, respectively. Note that the *current* instead of the power figures were recorded, assuming that the circuit is supplied by a fixed voltage level. The tables have common layout, with the main information reported including the value of the test statistic in column 2, the type of the resulting extreme model in column 3, the maximum current estimate in column 6, and the confidence intervals for confidence levels 90%, 95%, 99%, and 99.99% in columns 7–10. Note that the type-I and type-II extreme models in column 3 stand for cases $c \downarrow 0$ and $c > 0$, respectively. The two intermediate columns 4 and 5 contain additional information to help evaluate the performance of the approach. In particular, column 4 reports the maximum current in the sample used for estimation and column 5 ("observed maximum") gives our best estimate for the maximum current of each circuit after a very long simulation comprising of 1 000 000 units. Finally column 11, which is actually column 8 (95% confidence interval) divided by column 6 (estimated maximum), displays the relative estimation error for a 95% degree of confidence.

Some important observations can be made by careful examination of these tables. First of all, it is evident by looking at columns 2 and 3 that only one of the circuits examined (c7552) belongs to the type-II extreme model, while the rest (nine out of ten) are of type-I, a result that was largely expected considering the discussion held in Section IV-A. A visual explanation of this result is provided by the *empirical* cdf (or sample cdf) [20] of the peak instantaneous power (Fig. 5), the right tail of which is, more or less, similar to the tail of a normal (or gamma) distribution [Fig. 6(a)] for all circuits except c7552. For the latter circuit, the corresponding right tail in the diagram of Fig. 5 is considerably shorter and appears to come closer to the tail of a uniform distribution [Fig. 6(b)], which is the most representative distribution of case $c > 0$[17], [21]. Another thing to observe in the tables and particularly in columns 4–6 containing the sample maximum, observed maximum, and estimated maximum, respectively, is that, while the capability of the approach to look beyond the maximum of the sample used for estimation is clear and unquestionable, the results of estimation in type-I circuits may appear, at first glance, somewhat pessimistic when compared to the observed maximum. This should be of no concern, though, since the entire population of vector pairs is extremely large (many orders of magnitude larger than the 10^6 units used to extract the observed maximum), and thus, considering the (quite long) tails of type-I distributions, it is more likely that the actual maximum value is ultimately increased rather than remain at the observed level. Our next observation comes from column 11, which confirms that a fixed total number of 10 000 units (as calculated in Section V-D) can provide a relative estimation error of approximately 5% under a given confidence level of 95% for all type-I circuits, either small ones (like the c432) or

TABLE III
RESULTS FOR THE MINIMUM ESTIMATED VALUES OF MAXIMUM POWER AMONG 10 RUNS

Circuit	Value of test statistic	Type of extreme model	Sample maximum (mA)	Observed maximum (mA)	Estimated maximum (mA)	Confidence interval (mA)				95% relative error
						90%	95%	99%	99.99%	
c432	5.840	I	6.960	9.053	8.955	±0.469	±0.559	±0.734	±1.109	6.2%
c499	5.898	I	15.119	17.778	20.033	±0.937	±1.116	±1.467	±2.216	5.6%
c880	5.838	I	9.873	12.254	13.695	±0.729	±0.868	±1.141	±1.723	6.3%
c1355	3.793	I	13.154	15.175	16.372	±0.748	±0.891	±1.171	±1.769	5.4%
c1908	6.751	I	13.605	14.760	16.527	±0.740	±0.881	±1.158	±1.750	5.3%
c2670	5.400	I	16.218	20.034	20.664	±0.854	±1.017	±1.337	±2.019	4.9%
c3540	4.321	I	22.336	25.774	29.973	±1.580	±1.883	±2.475	±3.738	6.3%
c5315	5.693	I	35.421	41.511	45.939	±1.937	±2.308	±3.033	±4.582	5.0%
c6288	4.108	I	42.549	46.570	52.516	±2.191	±2.610	±3.430	±5.181	5.0%
c7552	8.417	II	38.589	41.323	39.553	±0.754	±0.899	±1.181	±1.784	2.3%

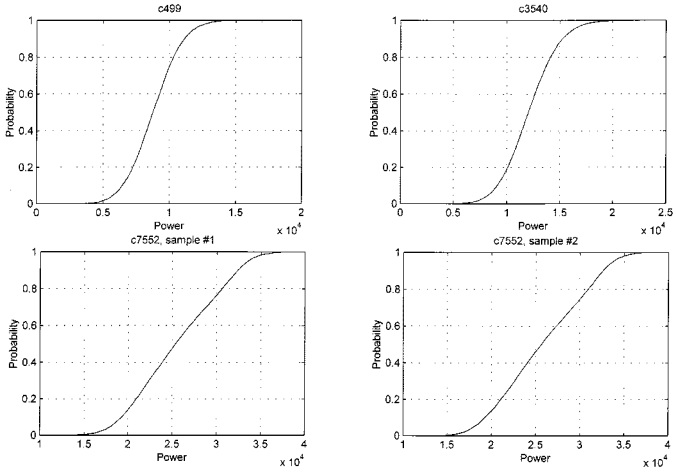


Fig. 5. Empirical cdf (based on samples of 10 000 units) of the peak instantaneous power for circuits c499, c3540, and c7552 (two samples). Observe the faster rate of ascent at the right tail in the case of c7552.

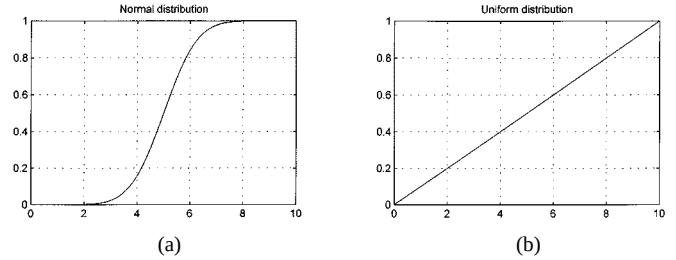


Fig. 6. Cumulative distribution functions of the normal and uniform probability distributions.

large ones (such as the c6288). As also expected, the error for the same number of units is typically even lower in type-II circuits, reaching as low as 2.3% for the only circuit (c7552) that fell into this category. As a final observation, we may notice that the results for c7552 (the only type-II circuit) are too tight, with nearly half of the runs yielding estimated values below the observed maximum. With that in mind, along with the fact that type-II circuits are not so frequent, we may state that it is generally safer to follow the treatment of type-I circuits (Section IV-C) in all circumstances and still obtain satisfactory (albeit on the pessimistic side) results in type-II cases. From another point of view, the estimate obtained by assuming $c \downarrow 0$ may be regarded as a (sample-dependent) *statistical* upper bound of maximum power, in the sense that it is the worst possible (i.e., most pessimistic) estimate of the upper end point when the latter is finite. For the sake of comparison, the average estimated value among the same ten runs for c7552 treated as a type-I circuit was 47.096 mA, with the 90%, 95%, 99%, and 99.99% confidence intervals being ± 1.570 mA, ± 1.870 mA, ± 2.458 mA, and ± 3.713 mA, respectively.

In addition to the experimental validation of the current approach itself, an empirical comparison with the previous approach proposed in [12] has been carried out and the results are presented next. Due to the different implementation of the benchmark circuits for the simulation process, a direct comparison (in terms of absolute numbers) of our experimental results to the corresponding ones in [12] is not possible. Therefore, our only option was to apply the estimation procedure given in [12] on the experimental data that were at our disposal. As pointed out in the introduction of the paper, the approach in [12] differs from the current approach in that it did not separately consider the case $c \downarrow 0$, but adopted in all situations the same general procedure developed for $c > 0$. Furthermore, compared to the corresponding procedure described in Section IV-B, the latter followed the Weibull-maxima model instead of the Beta-exceedances model and estimated maximum power through (27) instead of (28). In order to obtain proper results for the comparison, we employed the same total number of units 10 000 and, additionally, calculated the theoretical confidence interval of the estimate for the Weibull-distributed sample of maxima in a similar way as the corresponding interval (32) for the Beta-distributed sample of exceedances was calculated in Section IV-B and Appendix A (note that the two intervals exhibit the same asymptotic behavior for $c \downarrow 0$). The comparison between the two approaches was performed on the same samples that were used to construct Tables I–III and the results are summarized in Tables IV–VI, respectively. In the latter tables, the leftmost columns 2–4, convey the results for the observed and estimated maximum and the estimation error from the former tables (columns 5, 6, and 11) in order to facilitate the comparison, while the actual results of applying the previous approach are reported in columns 5–10 and include the estimated maximum, the 90%, 95%, 99%, and 99.99% confidence intervals, and the 95% relative estimation error, respectively.

Since we effectively treated the nine type-I circuits in a type-II way during the application of [12], the infinity-approaching (in view of Section IV-B) variance of the estimate is expected to lead to large estimation errors and/or large deviations of the estimated value between different samples for these circuits. Indeed, that kind of behavior was observed in practice, with the experimental results showing error figures that reached as high as 203.5% for c432 and estimates that ranged from 47.954 up to 68.904 in the case of c6288. Some good results did turn up occasionally, but they were scarce and needed samples of excellent quality. On the other hand, the results for c7552 (which was originally of type-II) were, in all cases, reasonable as expected, even though they may appear slightly worse than previously due to the smaller sample size employed in the maxima method than in the exceedances method when $c \downarrow 0$ (Section V-D). Another problem which might be overlooked at first sight is that the large variance of the estimate (or, equivalently, the large estimation interval) in type-I circuits frequently results in an estimated maximum that is substantially lower than the observed maximum. This is a typical case of underestimating maximum power, which is clearly unacceptable for the reliability design stage that naturally follows. Thus, even if we disregard the large estimation error characterizing all type-I circuits, the underestimation effect is yet another reason for rendering the previous approach inappropriate to be used as part of the design process for circuits of this type. In summary, the above comparison and its results clearly demonstrate that the case $c \downarrow 0$ cannot be left out of the estimation process but should be taken into

TABLE IV
COMPARATIVE RESULTS OF APPLYING THE PREVIOUS APPROACH OF [12] ON THE SAMPLES USED IN TABLE I

Circuit	Observed maximum (mA)	Estimated max. (mA) (our appr.)	95% relative error (our appr.)	Estimated max. (mA) (prev. appr.)	Confidence interval (mA) (previous approach) 90%	95%	99%	99.99%	95% relative error (prev. appr.)
c432	9.053	9.407	6.6%	14.180	±24.220	±28.860	±37.928	±57.288	203.5%
c499	17.778	21.042	6.0%	17.582	±2.461	±2.932	±3.854	±5.821	16.7%
c880	12.254	14.592	6.9%	13.906	±5.701	±6.793	±8.928	±13.485	48.8%
c1355	15.175	17.406	6.1%	15.103	±2.888	±3.441	±4.523	±6.831	22.8%
c1908	14.760	17.028	5.7%	13.301	±0.753	±0.897	±1.179	±1.781	6.7%
c2670	20.034	21.193	5.2%	18.169	±2.205	±2.628	±3.454	±5.217	14.5%
c3540	25.774	30.928	6.6%	43.170	±61.370	±73.130	±96.110	±145.170	169.4%
c5315	41.511	47.358	5.3%	40.932	±5.655	±6.738	±8.856	±13.376	16.5%
c6288	46.570	54.103	5.3%	58.105	±32.434	±38.648	±50.792	±76.717	66.5%
c7552	41.323	41.866	5.2%	43.418	±6.895	±8.216	±10.797	±16.309	18.9%

TABLE V
COMPARATIVE RESULTS OF APPLYING THE PREVIOUS APPROACH OF [12] ON THE SAMPLES USED IN TABLE II

Circuit	Observed maximum (mA)	Estimated max. (mA) (our appr.)	95% relative error (our appr.)	Estimated max. (mA) (prev. appr.)	Confidence interval (mA) (previous approach) 90%	95%	99%	99.99%	95% relative error (prev. appr.)
c432	9.053	9.746	6.9%	8.055	±1.586	±1.889	±2.483	±3.750	23.4%
c499	17.778	21.582	6.2%	27.781	±30.880	±36.795	±48.357	±73.040	132.4%
c880	12.254	15.276	7.3%	14.348	±6.070	±7.232	±9.505	±14.357	50.4%
c1355	15.175	18.546	6.7%	17.760	±7.321	±8.724	±11.465	±17.317	49.1%
c1908	14.760	18.555	6.4%	14.816	±1.887	±2.249	±2.955	±4.464	15.2%
c2670	20.034	22.672	5.8%	21.285	±6.715	±8.002	±10.516	±15.884	37.6%
c3540	25.774	31.743	6.8%	30.289	±12.665	±15.091	±19.833	±29.957	49.8%
c5315	41.511	48.608	5.5%	45.028	±11.763	±14.017	±18.421	±27.824	31.2%
c6288	46.570	55.810	5.6%	68.904	±67.670	±80.630	±105.970	±160.060	117.0%
c7552	41.323	42.519	6.3%	43.800	±6.906	±8.229	±10.815	±16.335	18.8%

TABLE VI
COMPARATIVE RESULTS OF APPLYING THE PREVIOUS APPROACH OF [12] ON THE SAMPLES USED IN TABLE III

Circuit	Observed maximum (mA)	Estimated max. (mA) (our appr.)	95% relative error (our appr.)	Estimated max. (mA) (prev. appr.)	Confidence interval (mA) (previous approach) 90%	95%	99%	99.99%	95% relative error (prev. appr.)
c432	9.053	8.955	6.2%	8.290	±2.490	±2.967	±3.899	±5.890	35.8%
c499	17.778	20.033	5.6%	15.806	±0.880	±1.049	±1.378	±2.082	6.6%
c880	12.254	13.695	6.3%	11.343	±1.727	±2.057	±2.704	±4.084	18.1%
c1355	15.175	16.372	5.4%	14.921	±3.039	±3.621	±4.759	±7.189	24.3%
c1908	14.760	16.527	5.3%	15.089	±3.151	±3.754	±4.934	±7.453	24.9%
c2670	20.034	20.664	4.9%	18.338	±2.680	±3.193	±4.196	±6.338	17.4%
c3540	25.774	29.973	6.3%	33.810	±25.136	±29.951	±39.363	±59.454	88.6%
c5315	41.511	45.939	5.0%	38.689	±3.586	±4.272	±5.615	±8.481	11.0%
c6288	46.570	52.516	5.0%	47.954	±8.885	±10.587	±13.914	±21.016	22.1%
c7552	41.323	39.553	2.3%	39.537	±1.177	±1.402	±1.843	±2.783	3.5%

consideration along with the general case, $c > 0$, a fact that provides enough evidence about the actual contribution of the proposed approach.

VII. Conclusion

The problem of maximum power estimation in CMOS VLSI circuits has been investigated in this paper. During the course of the paper, we have progressed through a sequence of logic steps and actions which are summarized here. We initially established the critical need for fast and accurate estimation of maximum power requirements in order to locate reliability problems early in the design phase and rejected the prospect of employing existing average power estimation techniques for this purpose. We stated that a sound approach has to combine simulation, which guarantees the accuracy needed for deep-submicrometer IC design and statistics so as to overcome the pattern dependence obstacle. We identified extreme value theory as the pertinent field of statistics for the problem at hand and presented the theoretical background behind it. We subsequently formulated the problem of maximum power estimation and provided the means to address it within the framework of extreme value theory. In particular, we have exhaustively examined the two cases of extreme value theory that may arise in practice, constructed an appropriate discrimination criterion, and integrated them into a procedure based on Monte Carlo random sampling. We went on by demonstrating the dimension independence property of the approach, along with its capability to predict a value for maximum power that approximately meets a specified level of accuracy at a certain degree of confidence. Finally, we have derived some very attractive features of the approach such as straightforward integration within the design flow of CMOS VLSI circuits, single-loop execution which obviates the need for iterative improvement and avoids potential convergence problems and overall algorithmic simplicity as plain numerical calculations are mostly involved. All the above claims have been supported by experimental results from benchmark circuits.

The proposed approach was applied to the problem of global maximum power estimation in the power supply bus. However, it is also applicable to local maximum power estimation in every node of the distribution network, where the objective is to uncover weak spots that may become the source of potential problems. The latter is a considerably more demanding problem than the former one, since every node must be treated individually and the proposed approach exhibits properties that render it capable of handling the computational load in reasonable time. Apart from this, the approach may be combined with an algorithm that performs optimization of the power grid, in order to locate and modify regions of overdesign where valuable silicon area is being wasted. Finally, the core of the algorithm can be easily adapted, with only minor modifications, to the general problem of estimating the upper end point of a bounded random physical process.

Appendix A
Details on the Estimate and Confidence Interval for $c > 0$

The statistical behavior of the ML estimates $\hat{\sigma}_u$ and $\hat{\beta}$ of σ_u and β is determined by the following theorem.

Theorem 6 [28]

The ML estimate $\hat{\beta}$ of β satisfies the regularity conditions and therefore complies with the conclusions of Theorem 5. On the other hand, the behavior of the ML estimate $\hat{\sigma}_u$ of σ_u varies according to the actual value of β and is distinguished in the following cases.

i) If $\beta > 2$, then the regularity conditions hold and $\hat{\sigma}_u$ also complies with the conclusions of Theorem 5.
ii) If $1 \leq \beta \leq 2$, then the regularity conditions are not satisfied and $\hat{\sigma}_u$ converges to a distribution with no simple analytical form, whereas the corresponding information term is infinite.

Note that the case $\beta < 1$ cannot exist in practice because the Beta pdf (29) tends to infinity as $x \to \omega(F) = u + \sigma_u$, and consequently almost every unit X_i inside a sample from this hypothetical distribution would be equal to $\omega(F)$.

In the regular case $\beta > 2$, the variance of the normally distributed $\hat{\sigma}_u$ involves computation of the information matrix with elements which are determined from (26); see the equation at the bottom of the page.

The covariance matrix is then obtained by inverting the information matrix and evaluating at $(\hat{\sigma}_u, \hat{\beta})$

$$\begin{aligned}\Sigma &= \begin{pmatrix} \operatorname{var}(\hat{\sigma}_u) & \operatorname{cov}(\hat{\sigma}_u, \hat{\beta}) \\ \operatorname{cov}(\hat{\sigma}_u, \hat{\beta}) & \operatorname{var}(\hat{\beta}) \end{pmatrix} \\ &= \frac{1}{r} \begin{pmatrix} \frac{\beta}{(\beta-2)\sigma_u^2} & -\frac{1}{\sigma_u(\beta-1)} \\ -\frac{1}{\sigma_u(\beta-1)} & \frac{1}{\beta^2} \end{pmatrix}^{-1} \Bigg|_{\substack{\sigma_u = \hat{\sigma}_u \\ \beta = \hat{\beta}}} \\ &= \frac{1}{r} \begin{pmatrix} \hat{\sigma}_u^2 \frac{(\hat{\beta}-2)(\hat{\beta}-1)^2}{\hat{\beta}} & \hat{\sigma}_u \hat{\beta}(\hat{\beta}-2)(\hat{\beta}-1) \\ \hat{\sigma}_u \hat{\beta}(\hat{\beta}-2)(\hat{\beta}-1) & \hat{\beta}^2(\hat{\beta}-1)^2 \end{pmatrix}\end{aligned}$$

from which we finally obtain (31).

$$-E\left[\frac{\partial^2 \log w_+(X;\sigma_u,\beta)}{\partial \sigma_u^2}\right] = \int_u^{u+\sigma_u} \left(\frac{\beta-1}{(x-u-\sigma_u)^2} - \frac{\beta}{\sigma_u^2}\right) \frac{\beta}{\sigma_u} \left(-\frac{(x-u)-\sigma_u}{\sigma_u}\right)^{\beta-1} dx = \frac{\beta}{(\beta-2)\sigma_u^2}$$

$$-E\left[\frac{\partial^2 \log w_+(X;\sigma_u,\beta)}{\partial \beta^2}\right] = \int_u^{u+\sigma_u} \frac{1}{\beta^2} \frac{\beta}{\sigma_u} \left(-\frac{(x-u)-\sigma_u}{\sigma_u}\right)^{\beta-1} dx = \frac{1}{\beta^2}$$

$$-E\left[\frac{\partial^2 \log w_+(X;\sigma_u,\beta)}{\partial \sigma_u \partial \beta}\right] = \int_u^{u+\sigma_u} \left(\frac{1}{x-u-\sigma_u} + \frac{1}{\sigma_u}\right) \frac{\beta}{\sigma_u} \left(-\frac{(x-u)-\sigma_u}{\sigma_u}\right)^{\beta-1} dx = -\frac{1}{(\beta-1)\sigma_u}$$

$$-E\left[\frac{\partial^2 \log g_0(X;\mu_m,\sigma_m)}{\partial \mu_m^2}\right] = \int_{-\infty}^{+\infty} \frac{1}{\sigma_m^2}\exp\left(-\frac{x-\mu_m}{\sigma_m}\right) \cdot \frac{1}{\sigma_m}\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\exp\left(-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right)dx\frac{1}{\sigma_m^2}$$

$$-E\left[\frac{\partial^2 \log g_0(X;\mu_m,\sigma_m)}{\partial \sigma_m^2}\right] = \int_{-\infty}^{+\infty} \left(2\frac{x-\mu_m}{\sigma_m}\left(1-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right)+\frac{(x-\mu_m)^2}{\sigma_m^4}\exp\left(\frac{x-\mu_m}{\sigma_m}\right)-\frac{1}{\sigma_m^2}\right) \cdot \frac{1}{\sigma_m}\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\exp\left(-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right)dx = \frac{1}{\sigma_m^2}\left((\gamma-1)^2+\frac{\pi^2}{6}\right)$$

$$-E\left[\frac{\partial^2 \log g_0(X;\mu_m,\sigma_m)}{\partial \mu_m \sigma_m}\right] = \int_{-\infty}^{+\infty} \left(\frac{1}{\sigma_m^2}\left(1-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right)+\frac{x-\mu_m}{\sigma_m^3}\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right) \cdot \frac{1}{\sigma_m}\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\exp\left(-\exp\left(-\frac{x-\mu_m}{\sigma_m}\right)\right)dx = \frac{\gamma-1}{\sigma_m^2}$$

When $1 \leq \beta \leq 2$ the problem of ML estimation is nonregular. Consequently, the above results do not hold anymore and the construction of a confidence interval for $\hat{\omega}(F)$ is not possible. We must point out, however, that this circumstance is very rare; in fact, it has not been encountered in practice among the situations examined in the current paper. Nevertheless, it is still a possibility that should not be neglected and we have thus formulated a suitable approach to follow whenever such a situation occurs. Specifically, we proceed as in the regular case and estimate $\omega(F)$ through (28), where $\hat{\sigma}_u$ is still obtained by maximization of (30). Since, however, the corresponding information term is infinite, the variance of $\hat{\sigma}_u$ will be zero [this can be verified by setting $\beta = 2$ in (31)], which means that $\hat{\sigma}_u$ exhibits negligible deviation with respect to its actual value σ_u and thus we do not need to construct a confidence interval.

Appendix B
Details on the Estimate and Confidence Interval for $c \downarrow 0$

Evaluation of the parametric expressions in (8) and (24), with $H(x)$ of (33) in place of the unknown $F(x)$, yields

$$\mu_m = \mu_H - \frac{\sigma_H}{\sqrt{\log m}} \tag{52}$$

(that is the solution of μ_m which complies with $\mu_m < \mu_H$ and $\sigma_H > 0$)

$$\begin{aligned}\sigma_m =& m\int_{\mu_H-\sigma_H/\sqrt{\log m}}^{\mu_H}\exp\left(-\frac{\sigma_H^2}{(y-\mu_H)^2}\right)dy\\ =& m\left((y-\mu_H)\exp\left(-\frac{\sigma_H^2}{(y-\mu_H)^2}\right)\right.\\ &\left.+\ \sigma_H\sqrt{\pi}\,\mathrm{erf}\left(\frac{\sigma_H}{y-\mu_H}\right)\right)\Bigg|_{y=\mu_H-\sigma_H/\sqrt{\log m}}^{y=\mu_H}\\ =& \sigma_H\frac{1+m\sqrt{\pi\log m}\left(\mathrm{erf}(\sqrt{\log m})-1\right)}{\sqrt{\log m}}\end{aligned} \tag{53}$$

where $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x \exp\left(-y^2\right)dy$ is the well-known *error function* that bears a close relationship with the cdf of the normal distribution and which is found in tabular form in many mathematical textbooks. Note that (53) makes use of the properties $\lim_{x\uparrow\mu}\mathrm{erf}(\sigma/(x-\mu)) = -1$ for $\sigma > 0$ and $\mathrm{erf}(-x) = -\mathrm{erf}(x)$.

Equations (52) and (53) may be solved with respect μ_H to σ_H and as follows:

$$\sigma_H = \frac{\sigma_m\sqrt{\log m}}{1+m\sqrt{\pi\log m}\left(\mathrm{erf}(\sqrt{\log m})-1\right)} \tag{54}$$

$$\mu_H = \mu_m + \frac{\sigma_m}{1+m\sqrt{\pi\log m}\left(\mathrm{erf}(\sqrt{\log m})-1\right)} \tag{55}$$

from which we finally obtain the estimate (34).

In order to derive the variance of the estimate, the elements of the information matrix are computed from (26); see the equation at the top of the page, where $\gamma \approx 0.5772\ldots$ is the *Euler gamma* constant which is defined as the limit $\gamma = \lim_{n\to+\infty}(\sum_{i=1}^{n} 1/n - \log n)$. The covariance matrix is then given by

$$\begin{aligned}\Sigma =& \begin{pmatrix}\mathrm{var}(\hat{\mu}_m) & \mathrm{cov}(\hat{\mu}_m,\hat{\sigma}_m)\\ \mathrm{cov}(\hat{\mu}_m,\hat{\sigma}_m) & \mathrm{var}(\hat{\sigma}_m)\end{pmatrix}\\ =& \frac{1}{k}\begin{pmatrix}\frac{1}{\sigma_m^2} & \frac{\gamma-1}{\sigma_m^2}\\ \frac{\gamma-1}{\sigma_m^2} & \frac{1}{\sigma_m^2}\left((\gamma-1)^2+\frac{\pi^2}{6}\right)\end{pmatrix}^{-1}\Bigg|_{\sigma_m=\hat{\sigma}_m}\\ =& \frac{1}{k}\frac{6\hat{\sigma}_m^2}{\pi^2}\begin{pmatrix}(\gamma-1)^2+\frac{\pi^2}{6} & 1-\gamma\\ 1-\gamma & 1\end{pmatrix}.\end{aligned}$$

The variance of $\hat{\omega}(F)$ can be computed from the above result by using the following theorem.

Theorem 7

[20] If the random variables X, Y are normal with means $m_X = E[X]$, $m_Y = E[Y]$ and covariance matrix $\begin{pmatrix}\mathrm{var}(X) & \mathrm{cov}(X,Y)\\ \mathrm{cov}(X,Y) & \mathrm{var}(Y)\end{pmatrix}$, then the random variable $Z = aX + bY$ (a, b: constants) is also normal with mean $m_Z = am_X + bm_Y$ and variance $\mathrm{var}(Z) = a^2\mathrm{var}(X) + 2ab\mathrm{cov}(X,Y) + b^2\mathrm{var}(Y)$.

By setting $X = \hat{\mu}_m$, $Y = \hat{\sigma}_m$, and $Z = \hat{\omega}(F)$ extracting a, b from (34), we finally obtain (37).

APPENDIX C
DETAILS ON THE DECISION BETWEEN $c > 0$ AND $c \downarrow 0$ FROM A SAMPLE

From the analysis presented in Section III-B it follows that the problem of the decision between $H_0 : c \downarrow 0$ and $H_1 : c > 0$ is equivalent to identifying the type of distribution (Exponential or Beta) that characterizes the r upper order statistics (exceedances) of the sample. Thus, any test statistic that will be eventually employed has to be calculated on the exceedances part $\underline{X}_{ex}$ and not on the main sample as a whole. Moreover, since both hypotheses H_0 and H_1 correspond to location-scale parametric families of distributions, we need a test statistic that is location and scale invariant and only depends on shape parameter $c = 1/\beta$. We consider the *coefficient of variation* $v_f = s_f/m_f$ which, for a particular distribution with pdf $f(x)$, is defined as the ratio of the standard deviation s_f to the mean m_f. Specifically for the Exponential and Beta families in (23) and (15), the latter are given by [16]

$$m_{w_0} = u + \sigma_u \text{ and } s_{w_0} = \sigma_u \tag{56}$$

$$m_{w_+} = u + \frac{\sigma_u}{\beta+1} \text{ and } s_{w_+} = \frac{\sigma_u}{\beta+1}\sqrt{\frac{\beta}{\beta+2}}. \tag{57}$$

If, in both (56) and (57), we divide the standard deviations by the means decreased by u, we obtain the location and scale invariant quantities $v_{w_0} = 1$ in the Exponential case ($c \downarrow 0$) and $v_{w_+} = \sqrt{\beta/(\beta+2)}$ in the Beta case ($c > 0$). We also observe that $v_{w_+} \xrightarrow{\beta \to +\infty} v_{w_0}$, which was expected since (23) is the asymptotic result of (15) for $c = (1/\beta) \downarrow 0$. From the above we conclude that a suitable test statistic is the *sample* coefficient of variation $v_{\underline{X}_{ex}} = s_{\underline{X}_{ex}}/(m_{\underline{X}_{ex}} - u)$, where $m_{\underline{X}_{ex}}$ and $s_{\underline{X}_{ex}}$ are the mean and standard deviation of the sample $\underline{X}_{ex}$, respectively. Indeed, $v_{\underline{X}_{ex}}$ is expected to take values near unity under H_0 (i.e., when $\underline{X}_{ex}$ follows an Exponential distribution) and deviate accordingly when H_1 holds true. In the following, we will use a variant of this test statistic with a well-defined distribution under the null hypothesis, so that the formation of a critical region is feasible. Specifically, it was proved in [29] (which in turn was based on an older result) that, under H_0, the quantity $(1/r)(m_{\underline{X}_{ex}} - u)^2/s^2_{\underline{X}_{ex}} = (1/r)1/v^2_{\underline{X}_{ex}}$ asymptotically (i.e., for $r \to +\infty$) follows a normal distribution with mean $1/r$ and variance $4/r^3$. Based on this result and after some calculations, we finally arrive at the test statistic of (38).

REFERENCES

[1] M. Pedram, "Power minimization in IC design: Principles and applications," *ACM Trans. Design Automation Electron. Syst.*, vol. 1, no. 1, pp. 3–56, 1996.

[2] H. Kriplani, F. Najm, and I. Hajj, "Pattern independent maximum current estimation in power and ground buses of CMOS VLSI circuits: Algorithms, signal correlations and their resolution," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 998–1012, Aug. 1995.

[3] C.-Y. Wang and K. Roy, "Maximum power estimation for CMOS circuits using deterministic and statistical approaches," *IEEE Trans. VLSI Syst.*, vol. 6, pp. 134–140, 1998.

[4] C.-Y. Wang, T.-L. Chou, and K. Roy, "Maximum power estimation for CMOS circuits under arbitrary delay model," in *IEEE Int. Symp. Circuits Systems*, 1996.

[5] C.-Y. Wang and K. Roy, "COSMOS: A continuous optimization approach for maximum power estimation of CMOS circuits," in *IEEE/ACM Int. Conf. Computer-Aided Design*, 1997.

[6] A. Krstic and K.-T. Cheng, "Vector generation for maximum instantaneous current through supply lines for CMOS circuits," in *ACM/IEEE Design Automation Conf.*, 1997.

[7] Y.-M. Jiang, K.-T. Cheng, and A. Krstic, "Estimation of maximum power and instantaneous current using a genetic algorithm," in *IEEE Custom Integrated Circuits Conf.*, 1997.

[8] Y.-M. Jiang and K.-T. Cheng, "Exact and approximate estimation for maximum instantaneous current of CMOS circuits," in *Design, Automation Test in Europe*, 1998.

[9] A. Hill, C.-C. Teng, and S. Kang, "Simulation-based maximum power estimation," in *IEEE Int. Symp. Circuits Systems*, 1996.

[10] C. Ding, Q. Wu, C. Hsieh, and M. Pedram, "Statistical estimation of the cumulative distribution function for power estimation in VLSI circuits," in *ACM/IEEE Design Automation Conf.*, 1997.

[11] G. Fishman, *Monte Carlo: Concepts, Algorithms and Applications*. New York: Springer, 1996.

[12] Q. Qiu, Q. Wu, and M. Pedram, "Maximum power estimation using the limiting distributions of extreme order statistics," in *ACM/IEEE Design Automation Conf.*, 1998.

[13] N. Evmorfopoulos and J. Avaritsiotis, "A new statistical method for maximum power estimation in CMOS VLSI circuits," *Active and Passive Electronic Components*, vol. 22, no. 3, pp. 215–233, 2000.

[14] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design: A Systems Perspective*, 2nd ed. Reading, MA: Addison-Wesley, 1993.

[15] R. Burch, F. Najm, P. Yang, and T. Trick, "A Monte Carlo approach for power estimation," *IEEE Trans. VLSI Syst.*, vol. 1, pp. 63–71, Jan. 1993.

[16] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values*. Boston, MA: Birkhauser, 1997.

[17] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed: Krieger, 1987.

[18] J. Pickands, "Statistical inference using extreme value order statistics," *Ann. Stat.*, vol. 3, pp. 119–131, 1975.

[19] S. Resnick, "Tail equivalence and its applications," *J. Appl. Probability*, vol. 8, pp. 136–156, 1971.

[20] G. Roussas, *A Course in Mathematical Statistics*, 2nd ed. New York: Academic, 1997.

[21] E. Castillo, *Extreme Value Theory in Engineering*. New York: Academic, 1988.

[22] G. Stamoulis, "A Monte Carlo approach for the accurate and efficient estimation of average transition probabilities in sequential logic circuits," in *IEEE Custom Integrated Circuits Conf.*, 1996.

[23] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.

[24] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. New York: Academic, 1981.

[25] Y. Evtushenko, *Numerical Optimization Techniques*. New York: Springer, 1985.

[26] G. Hahn and S. Shapiro, *Statistical Models in Engineering*. New York: Wiley, 1994.

[27] P. Hall, "On estimating the endpoint of a distribution," *Ann. Stat.*, vol. 10, no. 2, pp. 556–568, 1982.

[28] R. Smith, "Maximum likelihood estimation in a class of nonregular cases," *Biometrika*, vol. 72, no. 1, pp. 67–90, 1985.

[29] A. Hasofer and Z. Wang, "A test for extreme value domain of attraction," *J. Amer. Statistical Assn. (JASA)*, vol. 87, no. 417, pp. 171–177, 1992.

Nestoras E. Evmorfopoulos (S'99) was born in Athens, Greece, in 1971. He received the B.E. and M.E. degrees in electrical and computer engineering and the M.Sc. degree in engineering economics from the National Technical University of Athens (NTUA), where he is currently pursuing the Ph.D. degree.

His research interests include the mathematical aspects of CAD tools for VLSI design. He has published four papers in international journals and conferences.

Mr. Evmorfopoulos is a member of the Technical Chamber of Greece.

Georgios I. Stamoulis (S'94–M'95) received the Ph.D. degree from the University of Illinois, Urbana-Champaign, in 1994.

He then joined the Electrical and Computer Engineering Department of the University of Iowa as a Visiting Assistant Professor and in 1995 he joined Intel Corp. Since 2001, he has been an Assistant Professor at the Technical University of Crete, Greece. His research interests include the area of design and analysis for low power and reliability of ICs.

John N. Avaritsiotis (SM'92) was born in Greece in 1948. He received the B.Sc. (Hon.) in physics from the Department of Physics, University of Athens, Greece, in 1972, and the M.Sc. and Ph.D. degrees from Loughborough University of Technology, in 1974 and 1976, respectively.

From 1976 to 1979, he was employed as a Research Fellow in the Thin Film Group of Loughborough University of Technology. From 1980 to 1986, he was appointed Lecturer at the Electronics Laboratory of Athens University and in 1986 he was elected Associate Professor in the Department of Electrical Engineering and Computer Engineering of the National Technical University of Athens. In 1990, he was elected Professor of Microelectronics in the same department. He has worked as a Technical Consultant to various British and Greek industrial firms for the incorporation of new thin-film deposition techniques in their production processes. He has published over 80 technical articles in various scientific journals and has presented more than 30 papers at international conferences. His research activities include the field of microelectronics materials and thin film deposition techniques for use in the fabrication of semiconductor devices. He is the Director of two R&D Laboratories: the Microelectronics Lab and the Electronic sensors Lab of NTUA.

Dr. Avaritsiotis is the Associate Editor of *Active and Passive Electronic Components* and Guest Editor of IEEE TRANSACTIONS ON COMPONENTS, PACKAGING, AND MANUFACTURING TECHNOLOGY. He is a member of IOP and ISHM.