

Copyright © 1996, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**A STATISTICAL APPROACH TO TRANSISTOR
LEVEL POWER ESTIMATION**

by

Premal Buch

Memorandum No. UCB/ERL M96/7

3 March 1996

COVER PAGE

**A STATISTICAL APPROACH TO TRANSISTOR
LEVEL POWER ESTIMATION**

by

Premal Buch

Memorandum No. UCB/ERL M96/7

3 March 1996

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

We present a Monte Carlo based statistical approach to power estimation at the transistor level. While transistor level power estimation can be very accurate, for practical usage, it suffers from two disadvantages: large runtimes, and pattern dependency. We propose a two-point approach to overcome these problems: A Monte Carlo based technique is used to obtain pattern-independent, statistically meaningful estimate. A partitioning for estimation strategy is proposed to speed-up the estimation by exploiting the multi-rate power dissipation behavior of circuits and to improve modelling accuracy for circuits which are *stiff* from a power perspective. The main contributions of this work are extension of transition density for 0-1 signals to a similar notion for nodal waveforms, a divide-and-conquer approach to power estimation and a statistical modelling technique for propagating information between partitions of a circuit.

1 Introduction

Power minimization is becoming very important for a number of reasons ranging from an increasing demand for portable computing and telecommunication equipment and the problem of “hot chips” due to increasing clock frequencies and device counts of ICs. Minimizing power dissipation of chips has an impact not only on energy savings, but also helps create more reliable chips.

Power estimation is essential to systematically guiding the design process to meet its power goals. In the initial stages of the design, power estimation is required to obtain feedback about design decisions, while in the later stages power estimation can be used to help identify potential hot spots in the design before it is fabricated. The existing work on power estimation can be classified under three broad classes of techniques: empirical techniques, probabilistic techniques and simulation based methods. Each of these class of methods offers a different speed-accuracy trade-offs.

The main advantages of the empirical and probabilistic techniques is their short runtimes and input-independence. The big drawback of these methods however is their lack of accuracy. The probabilistic techniques use a stochastic model of logic signals of a circuit and propagate the probabilities of logic values through the combinational logic modules in order to compute the average switching rate of the circuit. These methods trade-off accuracy for speed by ignoring spatio-temporal correlations between internal node values. Methods which attempt to model these effects suffer from blow-ups in time and memory requirements. Empirical techniques generally use various statistical measures of the circuit to calculate the power dissipation of the circuit based on simple models. Although these methods can be very fast, since very little of the implementation details are accounted for, the errors can be unacceptably high.

Characterizing a design for power requires very accurate estimation tools which can model all physical effects in the device models (particularly so for submicron technology) and identify main sinks of power in the design. Thus, while empirical and probabilistic techniques are useful in obtaining rough estimates of power dissipation at early stages of the design, due to above disadvantages they are not suitable for very accurate estimation.

The most direct approach of obtaining the power dissipation of a circuit is to use circuit simulators such as HSPICE[6] to simulate the design [1][7]. While offering good accuracy, this approach suffers from the drawback of large runtimes. Furthermore, this approach is strongly pattern-dependent since it requires user-specified input stimuli for simulation. It is often impossible to obtain a large enough input pattern set such that the power estimate is statistically meaningful. Even in cases where such inputs are available, it is not easy to determine the size of the vector set required to obtain a meaningful estimate of the design power dissipation in a typical operating environment. A digital design with n inputs can have 2^n possible combinations at the primary inputs. For an analog design or when considering temporal correlations between input values in different cycles in the digital case, the size of the possible input vector set would increase exponentially. Clearly, it would be infeasible to perform exhaustive simulation of all possible input combination to obtain a power estimate of the design. Thus we need a method for fast estimation

with a formal technique to limit the number of required input patterns.

In this work, we propose an approach to handle these problems that allows us to take advantage of the high accuracy of circuit simulation based methods without compromising on runtimes or pattern independence. A Monte Carlo based approach is used to determine the required input vector set and a divide-and-conquer estimation approach is combined with a fast simulation technique to reduce runtimes.

We develop a stochastic model for power dissipation at the transistor level and apply this to set up a Monte Carlo based power estimation process. The Monte Carlo approach consists of applying random inputs to the system and monitoring its output. This is continued until a value of power is obtained with a desired accuracy, at a specified confidence level. Any standard circuit simulator can be used to simulate power dissipation within the Monte Carlo framework. We propose the use of the stepwise equivalent conductance [10] based approach to best speed-up the power simulation without sacrificing any accuracy or adding computational overhead/circuit modification [1].

To further speed-up the estimation process, we propose a divide-and-conquer approach to break down the problem in sizes that become practically feasible for transistor-level simulation. Each sub-problem can still be further partitioned to gain speed-ups in the circuit simulation process itself as in [11][12]. The main advantage of partitioning for estimation is that we can exploit the multi-rate behavior and stiffness of a circuit from the power perspective. Depending upon the switching frequency at internal nodes, different parts of the circuit may require different number of vectors to yield a statistically significant estimate of the power dissipation. Partitioning the circuit allows us to use different input vector sets of appropriate size for different parts of the circuit. This not only speeds-up the estimation process by reducing computation, but also improves the accuracy of the stopping criteria of the Monte Carlo estimation in case of circuits where a single power dissipation model does not fit well.

We propose a statistical model to propagate information between partitions of a circuit. This model can also be used to bias the vector generation at the primary inputs when any user provided information about external inputs is available.

This paper is organized as follows: In Section 2 we review the circuit simulation background relevant to the transistor level power estimation problem. In Section 3, we extend the concept of transition density to obtain a similar measure for switching activity in analog waveforms. Section 4 presents some experimental results and Section 5 concludes with a summary of this work.

2 Background

The process of power simulation involves solving nonlinear, time-varying system of circuit equations. In the following we briefly describe the simulation engine used in this estimation framework. This approach speeds-up simulation by using the stepwise equivalent conductance model to approximate the conductance of a nonlinear device by a constant equivalent conductance during each time-step of the transient simulation. This allows us to transform the nonlinear, time-varying system of circuit equations to a linear, time-invariant system at every time-point in transient simulation. This transformation eliminates the time consuming Newton-Raphson iterations for implicit integration. Further speed-up is obtained by using the piecewise linear approximation for voltage waveforms.

From the power simulation standpoint, the advantage of this approach is that the power can be measured directly (without introducing any additional powermeter circuitry [10]) by monitoring the conductance and the voltage waveform during each time-step. Using the constant equivalent conductance for each non-linear device and the piecewise linearity of the voltage waveforms, the power dissipation in a device in one time-step h_n , from t_n to t_{n+1} is given by

$$P_d = \frac{1}{h_n} \int_{t_n}^{t_{n+1}} V(t) \mathcal{F}(V(t)) dt = \frac{1}{h_n} \int_{t_n}^{t_{n+1}} \mathcal{G} \cdot V^2(t) dt = \frac{\mathcal{G}}{h_n} \int_0^{h_n} \left[V(t_n) + t \cdot \left. \frac{dV}{dt} \right|_{h_n} \right]^2 dt \quad (1)$$

where $V(t)$ is the voltage across a device, $\mathcal{F}(V(t))$ is the time-varying current through a nonlinear device, and \mathcal{G} is the constant stepwise equivalent conductance of the device during the time-step.

Nonlinear capacitors and inductors can be handled similarly. The power dissipation of the circuit is obtained by summing the average power dissipated for the simulation period over all nodes.

3 Proposed Approach

The Monte Carlo method is a technique based on performing sampling experiments on the model of a system. It can be used not only for solution of stochastic problems, but also for solution of deterministic problems which have the same formal expression as some stochastic process. In Section 3.1, we show how the transistor-level power estimation problem exhibits this characteristic. Section 3.2 describes the set-up of the resulting Monte Carlo problem and its stopping criteria. Section 3.3 presents a divide-and-conquer estimation strategy, and Section 3.4 proposes a statistical modeling technique for propagating information between each partitions in the divide-and-conquer approach.

3.1 A Measure of Switching Activity for Analog Waveforms

In general, the instantaneous power dissipation $p(t)$ in a MOS device can be represented as a polynomial function $f(t)$ of the analog voltage waveform at the drain-source nodes.

$$p(t) = f(V_{ds}(t), V_{gs}(t)) \quad (2)$$

We denote the total average power dissipated in the circuit during time interval $(-T/2, T/2]$ as $P_T = \frac{1}{T} \int_{-T/2}^{T/2} p(t) dt$. The average power dissipation P_{avg} is then given by

$$P_{avg}(t) = \lim_{T \rightarrow \infty} P_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} p(t) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (3)$$

In the following, we show how the power estimation problem can be reduced to a mean estimation problem. We construct a companion stochastic process to $f(t)$ and prove that this process is strict-sense stationary (SSS) and mean-ergodic. The transformation to mean estimation will follow as a result.

We will denote the probability of an event A by $Prob\{A\}$ and, if x is a random variable, we denote its mean by $E[x]$ and its distribution function by $F_x(a) \equiv \mathcal{P}\{x \leq a\}$. We assume that the values of the analog waveforms are bounded by some arbitrary constant K_v . This is not a restrictive constraint for any real circuit, and in any case, the power estimation problem is not well-defined in presence of unbounded node voltages.

A stochastic process $f(t)$ is said to be SSS if its statistical properties are invariant to a shift of the time origin [15]. This essentially means that the mean $E[f(t)]$ of such a process is a constant and independent of time (we will denote it by $E[f]$). By definition, a constant-mean stochastic process is said to be mean-ergodic [15] if its time average tends to its constant-mean as $T \rightarrow \infty$.

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt = {}_1E[f] \quad (4)$$

where $=_1$ is used denote convergence everywhere with probability 1.

Let $\tau \in (-\infty, \infty)$ be a random variable with the probability function $F_\tau(t) = 1/2$ for any finite t , $F_\tau(-\infty) = 0$, and $F_\tau(\infty) = 1$. If $F_{\tau T}(t)$ is the uniform distribution over $[-T/2, T/2]$, then $\lim_{T \rightarrow \infty} F_{\tau T} = F_\tau$. Thus, one might say that τ is uniformly distributed over the whole real line \mathcal{R} . We now use τ to build from $f(t)$, a stochastic process $f(t)$, defined as follows:

Definition: Given a polynomial function $f(t)$ of an analog signal, and a random variable τ , uniformly distributed over \mathcal{R} , define a stochastic process $f(t)$ called the companion process of $f(t)$ given by: $f(t) \equiv f(t + \tau)$.

Proposition 1: Let $f(t)$ be a polynomial function of an analog signal. If $f(t)$ is the companion process of $f(t)$, then the following “convergence everywhere” results are true:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (5)$$

Proof: To prove (4), we need to show that for any give finite $\tau_1 \in \mathcal{R}$, the difference

$$\Delta_a \equiv \frac{1}{T} \int_{-T/2}^{T/2} f(t + \tau_1) dt - \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (6)$$

tends to zero as $T \rightarrow \infty$. This can be written as

$$\Delta_a = \frac{1}{T} \int_{-T/2 + \tau_1}^{T/2 + \tau_1} f(t) dt - \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt = \frac{1}{T} \int_{T/2}^{T/2 + \tau_1} f(t) dt - \frac{1}{T} \int_{-T/2}^{-T/2 + \tau_1} f(t) dt \quad (7)$$

Since $f(t) \leq K_v$, then $|\Delta_a| \leq K_v \frac{|\tau_1|}{T}$ must go to 0 as $T \rightarrow \infty$. Since this is true for any $\tau_1 \in \mathcal{R}$, then the convergence is everywhere, in the sense that every value of τ will lead to convergence.

Proposition 2: The companion process $f(t)$ of a polynomial function $f(t)$ of an analog signal is SSS and mean-ergodic with

$$E[f] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (8)$$

Proof: At $t = 0$, we have $E[f(0)] = E[f(\tau)]$. An interesting property of τ is that if a is a constant than $a + \tau$ has the same distribution as τ . Indeed, if $F_{a+\tau}(t)$ is the distribution function of $a + \tau$, then $F_{a+\tau}(t) = \mathcal{P}\{a + \tau \leq t\} = \mathcal{P}\{\tau \leq t - a\} = 1/2 = F_\tau(t)$. Therefore, since $t + \tau$ and τ are identically distributed, we have $E[f(t + \tau)] = E[f(\tau)]$, which means that $f(t)$ is a constant-mean process with $E[f(0)] = E[f(t)] = E[f(\tau)]$ for any time t .

To prove mean-ergodicity, consider the random variable

$$\eta_T \equiv \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (9)$$

from (4), we have

$$\lim_{T \rightarrow \infty} \eta_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (10)$$

where this convergence is everywhere. Therefore,

$$\lim_{T \rightarrow \infty} E[\eta_T] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (11)$$

By linearity of the expected value operator, this can be rewritten as

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} E[f(t)] dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (12)$$

But $E[f(t)]$ is a constant. Therefore, the left-hand side of (11) is simply $E[f]$, and mean-ergodicity follows, with $E[f]$ given by (7).

Thus, using (3),

$$P_{\text{avg}} = E[f] \quad (13)$$

and the problem of estimating P_{avg} is reduced to the task of computing $E[f]$.

3.2 A Monte Carlo Approach to Transistor Level Power Estimation

The mean estimation problem corresponding to the power estimation problem can be efficiently solved by a Monte Carlo based approach involving monitoring simulation results over a length of time. In order to estimate the expected value of the mean $E[f](= P_{\text{avg}})$, we observe N samples of the power dissipation process and use their average μ_N as a point estimate of $E[f]$

$$\mu_N = \frac{1}{N} \sum_{i=1}^N f(t_i, \tau) \quad (14)$$

In order to guarantee an error bound on this estimate μ_N with a certain confidence level, we need to obtain an interval estimate of μ_N . This requires determining the distribution μ_N . In general, this is a difficult problem involving multiple convolutions. To simplify it, we shall assume μ_N is normal. This is true if f is normal.

Theorem 1 (Central Limit Theorem): Let X_1, X_2, \dots, X_n be a random sample of size n from a population whose distribution has finite mean and variance μ and σ^2 respectively, and let \bar{X} be the sample

mean. The random variable $Z = \sqrt{N}(\bar{X} - \mu)/\sigma$ has as its limiting distribution as $n \rightarrow \infty$, the standard normal distribution.

Since P_T is the sum of power dissipations at the m devices in the circuit, under the above theorem, a sufficient condition for normality of P_T is that m be large, and power dissipation in each device be independent. This is true under fairly general conditions irrespective of the individual distributions making up P_T [7]. In the context of power dissipation, this assumption is reasonable for the following reason:

In CMOS circuits, there are two components that contribute to power dissipation [3]: static dissipation (due to leakage current) and dynamic dissipation (due to switching transient current and charging and discharging of load capacitance). Ideally, CMOS circuits dissipate no static power since in the steady state there is no direct path from V_{dd} to ground. In practice, since the MOS transistor is not a perfect switch, there are always leakage currents and substrate injection currents, which give rise to a static component of CMOS power dissipation. Since the substrate current reaches its maximum for gate voltages near $0.4V_{dd}$ and since the gate voltages only reside in this range during the switching transients, the actual power contribution of the substrate injection current is a function of the switching of gate-source voltage V_{GS} of the MOS device and is quite small [6]. Another source of static power dissipation is sub-threshold currents of the transistors. Again, this contribution is dependent on V_{GS} and is very small for current technologies. Thus the static power dissipation for each MOS device in a static CMOS circuit is a strong function of the corresponding V_{GS} . The dynamic power dissipation of a circuit is proportional to the switching frequency of its nodes. The power dissipation during one transition from low-high-low depends on the V_{ds} and V_{gs} waveforms during the transition.

Thus, while power dissipation in each device in a circuit is a complex, non-linear function of many parameters, it is primarily controlled by its terminal node voltage waveforms. Although node voltages may be locally correlated, in a large circuits under random inputs, there is very little correlation among arbitrary node voltage pairs. Thus the Central Limit theorem can be applied in the power estimation context. [2] discusses a similar proposition for logic level power estimation. Experimental justification for this assumption and the validity of the stopping criterion when this assumption is violated are presented in Section 4.

Accordingly, suppose P_T is normally distributed with a mean μ and variance σ^2 . Let μ_N be the observed mean and s_N be the observed standard deviation from N simulations of the circuit, each of length T . The parameters μ and σ of the distribution of P_T are unknown, and are estimated by μ_N and s_N . Since μ_N is the mean of N stochastically independent observations from a normally distributed population with parameters (μ, σ^2) , $\sqrt{N}(\mu_N - \mu)/\sigma$ is normally distributed with parameters $(0, 1)$. Since σ is unknown, if we use the estimate s_N instead of σ , then the variable $\sqrt{N}(\mu_N - \mu)/s_N$ has a t -distribution with $N-1$ degrees of freedom.

Thus, the hypothesis $\mu = \mu_N$ can be tested by the t -test of significance. The critical region at the α level of significance is now given by

$$\frac{|\mu - \mu_N| \sqrt{N}}{s_N} < t_{\alpha/2} \quad \Rightarrow \quad \frac{|\mu - \mu_N|}{\mu_N} < \frac{t_{\alpha/2} s_N}{\mu_N \sqrt{N}} \quad (15)$$

Thus for a specified percentage error ϵ in power estimate, and a given confidence level $(1-\alpha)$, we must simulate the circuit until (13) is satisfied.

$$\frac{t_{\alpha/2} s_N}{\mu_N \sqrt{N}} < \epsilon \quad (16)$$

The samples in a Monte Carlo method, as a rule, are independent. We set-up the power estimation problem as a Monte Carlo estimation problem by sampling the power dissipation of a circuit using a circuit simulator. Random numbers are used at the primary inputs. These can be biased to reflect any additional information provided by the user. Independent samples are taken by allowing a settling time between two samples as in [2].

3.3 Partitioning for Estimation

In the context of power estimation via simulation, it is important to note that while we want the accuracy that can only be provided by circuit simulation, the final aim is to obtain a power estimate, a scalar quantity, out of the given system. Thus, we do not need to preserve the waveform vectors at each internal node of the circuit as long as we can extract statistically significant information about the switching behavior. Motivated by this observation, we propose a partitioning for estimation approach to speed-up the estimation process. This divide-and-conquer strategy is used to partition the estimation problem itself and is independent of any circuit partitioning used to speed-up the simulation.

A big advantage of our divide-and-conquer approach is that we can model power dissipation in each subcircuit separately. This allows us to exploit the multi-rate behavior and stiffness of a circuit from the power perspective. Depending upon the node switching activity, different parts of the circuit may need different number of node vectors to yield a statistically significant power estimate. In general, a very active node provides more information in the sample of a given length than does a quiet node. If we sampled the entire circuit together, we would be constrained to use the maximum length input vector set required among all node. By partitioning the design, we can use vector sets of appropriate length for each partition, thus gaining an over-all speed-up.

Another benefit of partitioning is the improved accuracy in estimating power dissipation in circuits where the normal distribution assumption of the previous section is not a good fit. In [2], it was observed that many circuits have a *double normal* distribution (a special case of *bimodal* distribution where each of the two distribution is normal). This can be caused by circuits which have different parts with widely different power dissipation behavior or different functional modes with different power consumption. Our approach can easily handle cases where different parts of the circuit have different behavior, since power dissipation in a normal distribution can be fitted to each part separately.

We perform static circuit partitioning based on channel connected components to obtain a fine grain partitioning of MOS circuits. (In the case of static CMOS circuits, this would correspond to partitioning the circuit in simple logic gates). Since the granularity of this partitioning is very fine, we cluster together many such components to obtain a few partitions covering the entire circuit. A Monte Carlo based approach is used as before to perform statistical estimation of each cluster.

In theory, the number of simulations required to guarantee a confidence level is only weakly dependent on the circuit size [2]. However, in practice, the simulation time will strongly depend on the granularity and quality of partitioning. Since the statistical estimate for each subcircuit is obtained by multiple runs of simulation, signal correlations and all physical affects inside the subcircuit are accounted for in the estimate. However, signal correlations between partitions are only weakly accounted for (via input biasing).

3.4 Information Propagation Between Partitions

The fanouts of a subcircuit can be fanins to other subcircuits in the design. Thus, the inputs to each subcircuit are not independent random variables. This poses two problems: scheduling of partitions for Monte Carlo estimation, and information propagation among partitions.

A cycle-free schedule is generated using a signal flow graph for the circuit with each partition represented by a node in the graph. Selective trace algorithm [11] is used to schedule the partitions for Monte Carlo estimation. Note that local feedback is not a problem as all tightly coupled nodes are clustered in the same component. Global feedback can cause the signal flow graph to be cyclic. We solve this problem by clustering the components forming a global cycle into one big partition. This is not a major constraint since the number of partitions does not have to be very large.

We also need to account for the fact that inputs to a partition may be fanouts of other partitions in the design. One solution would be to simply propagate the output waveforms of one partition to all its fanout partitions. However, this would constraint that all partitions use the same number of vectors, thus prohibiting us from exploiting the multi-rate behavior and stiffness of the circuit. To solve this problem we construct a statistical model to each output and use this model to generate samples of this signals as required by the partitions it is fanning out to. We fit a normal distribution model to the analog signal waveforms at the outputs. The mean is obtained by computing the area under the waveform and the variance is obtained using this mean and the average waveform value during each clock cycle of the simulation. This model is consistent with our initial conjecture that device (and consequently nodal) waveforms are distributed normally. These parameters of normal distribution are then used to bias the input vector generation of the fanouts of the subcircuit.

Note that this approach does not account for spatiotemporal correlations between signals crossing subcircuit boundaries. This creates a trade-off between speed and accuracy. A very fine granularity partitioning speeds-up the simulation process but can be inaccurate since the error cause by neglecting the spatiotemporal correlations can dominate over the accuracy advantage gained by partitioning above; while

in absence of partitioning, the runtimes can be very high. Spatiotemporal correlations can be approximately accounted by using pairwise correlations between signals to bias the input patterns generations for such signals.

4 Experimental Results

The algorithms outlined in this paper were implemented in a power estimation program called *agni*. There is no constraint on the choice of the internal circuit simulation engine for the Monte Carlo samplings. In our implementation the SWEC circuit simulator was used since it provides good speed-ups at demonstrated high accuracy [1][12]. For the purpose of this benchmarking, we used a suite of industrial circuits described in Table 1. The netlists were extracted from the layout of real designs using industry ASIC cell libraries. We compare the power estimation results from *agni* against two metrics: direct circuit simulation results using very long random input vector sets and vector sets with maximum switching at the primary inputs. The first comparison is intended to demonstrate the accuracy of the proposed Monte Carlo approach while the second comparison demonstrates the inadequacy of direct simulation in the absence of

Circuit Name	# MOS	Description
count32	174	32-bit counter
mux2b16	214	1 bit 2-to-1 mux
ripple4	442	16 bit comparator
cla16	1200	16 bit carry look ahead adder
mult8	2691	8 bit wallace tree multiplier
mult16	9778	16 bit wallace tree multiplier
multp16	11314	16 bit pipelined multiplier
m16	6323	processor block

Table 1: Benchmark circuit descriptions

Circuit Name	Steady State Power Dissipation			Monte Carlo Estimation			% error
	# vectors	runtime	power (mW)	# vectors	runtime (sec)	power (mW)	
count32	10000	19.20 hr	4.25	1857	12461.15	4.14	2.59
mux2b16	10000	5.86 hr	190.11	103	261.22	189.77	0.18
ripple4	10000	33.17 hr	1587.28	468	2783.13	1665.31	4.91
cla16	10000	36.49 hr	795.46	80	1196.10	821.82	3.31
mult8	10000	26.06 hr	3951.45	48	1763.77	4118.40	4.22
mult16	2500*	46.46 hr	17888.04	38	6168.53	18404.70	2.81
multp16	5000*	77.44 hr	13173.72	43	8028.01	13805.30	4.79
m16	-*	-	-	297	44.84 hr	751.04	-

Table 2: Benchmark results for industry circuits (*: less than 10000 vectors due to cpu time constraints)

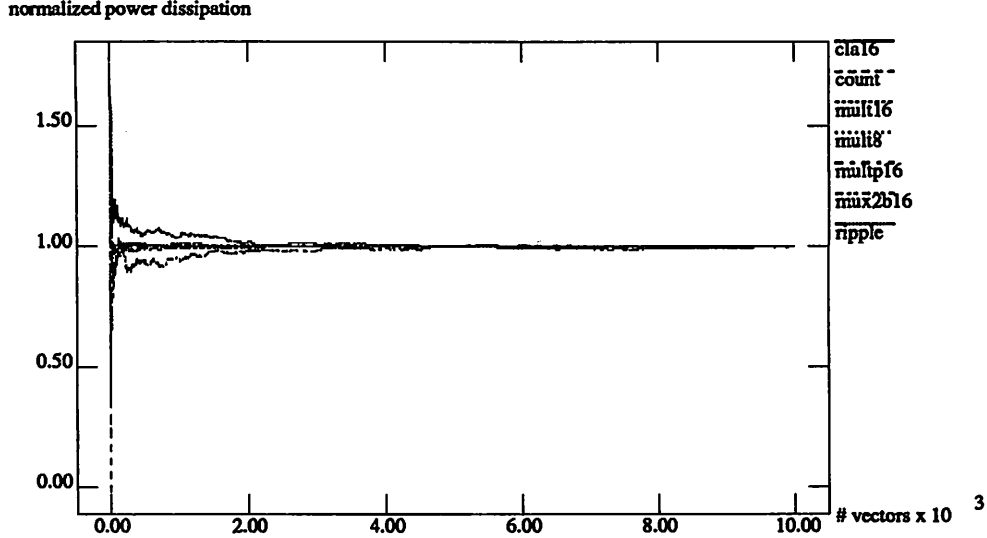


Figure 1: Normalized power dissipation of benchmark circuits

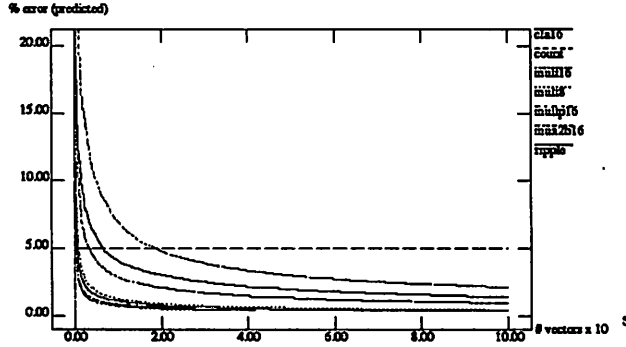


Figure 2: Predicted error for benchmark circuits

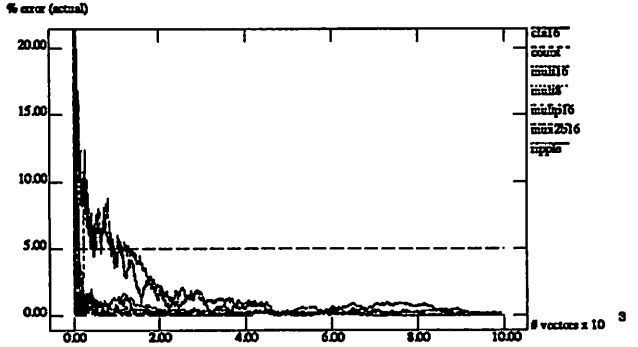


Figure 3: Actual error for benchmark circuits

a formal vector set selection process. The results were obtained on a DEC 5000 platform with a 24 Mbyte main memory.

Table 2 contains the results of the statistical power estimator on the benchmark circuits. In the first three columns, we present the runtime and power dissipation data for direct simulation of benchmark circuits with 10000 vectors. This was used to represent the real power dissipation of the circuit since it is an average of power dissipation over an extremely long simulation sample. The next three columns contain the runtime and power dissipation data from the Monte Carlo estimator. For these results, a 5% error tolerance with a 90% confidence interval was used to determine the stopping criteria. The last column contains actual error in the Monte Carlo approach when compared with the 10000 vector simulation with the above stopping criteria. All errors were found to be within the required bounds. The results clearly indicate that the stopping criteria predicts the error well, and indeed only a small set of vectors is required to obtain power estimates with good accuracy. This can result in large runtime savings while meeting the error specifications of the user, and without compromising the confidence in the estimate.

Figures 1, 2, and 3 further illustrate the effectiveness of the Monte Carlo approach. Figure 1 plots the

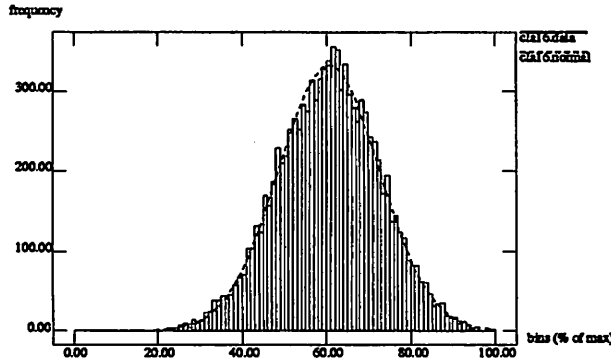


Figure 4: Power dissipation distribution of the 16 bit carry look ahead adder

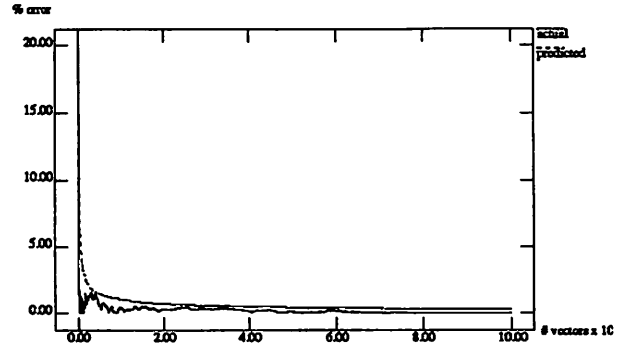


Figure 5: Predicted and actual error for the 16-bit carry look ahead adder

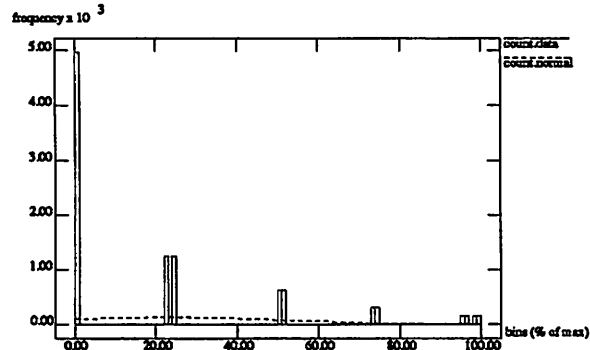


Figure 6: Power dissipation distribution of the 32 bit counter

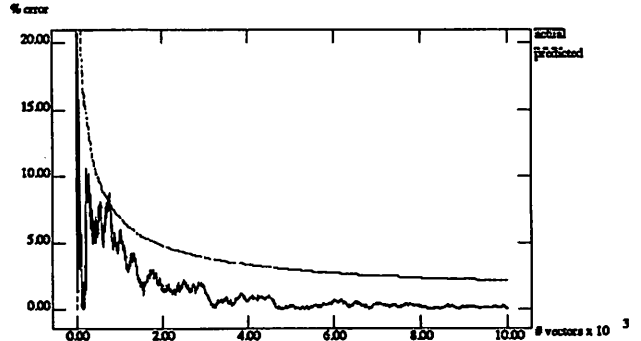


Figure 7: Predicted and actual error for the 32 bit counter

power dissipation waveforms for each circuit under test. All waveforms have been normalized with respect to their final, steady state value at the end of 10000 cycles. These waveforms demonstrate that the power dissipation asymptotically approaches its steady state value and number of vectors required for a desired accuracy can be predicted if a reasonable bound were available. Figures 2 and 3 represent the predicted error bound and the actual error relative to the steady state value. Clearly, the bound and stopping criteria proposed in this work are conservative and valid.

One of the key assumptions in the derivation of the stopping criteria of the Monte Carlo estimator was that P_T , the power dissipation in a clock cycle, is normally distributed. In Section 3 we indicated that while this cannot be proved theoretically, under most general conditions this is a good approximation. Figures 4,5,6 and 7 examine this assertion in more detail on two cases among our test circuits which are the best and the poorest fit for our normal distribution assumption used to derive the stopping criteria.

Figure 4 compares the distribution of power dissipation data from a 10000 vector random simulation of a 16-bit carry look ahead adder circuit with a normal distribution (with the mean and standard deviation derived to fit the data). It can be seen that the normal distribution is indeed a good approximation to the sampled data. To further test our proposition on this example case, we applied the χ^2 Goodness-of-Fit test [7][16] to the simulation data. It was found that the null hypothesis that P_T is normally distributed satisfied the χ^2 test with a type I error α of 0.05. For this circuit the predicted error bound and the actual error are

Circuit Name	power (μ W)		
	Simulation for maximum switching at PI	Steady State Dissipation	Monte Carlo Estimation
count32	9.89	4.25	4.14
mux2b16	687.25	190.11	189.77
ripple4	2702.15	1587.28	1665.31
cla16	1619.02	795.46	821.82
mult8	12154.30	3951.45	4118.40
mult16	54067.00	17888.04	18404.70
multp16	42325.30	13173.72	13805.30
m16	1871.33	-	751.04

Table 3: Comparison between direct simulation with worst case primary input switching, and the statistical approach

plotted in Figure 6. Clearly, the error bound is a conservative, tight bound on the actual error.

In Figure 6, we compare the distribution of power dissipation data from a 10000 vector random simulation of a 32-bit counter with a normal distribution (with the mean and standard distribution derived to fit the data). This circuit has five distinct modes depending upon which bit of the counter is high and clearly, the normal distribution is a very poor approximation to the actual power dissipation. Even in this case, we found that the predicted error bound was a conservative bound on the actual error (although not as tight as in the best fit case).

In general, it is possible that P_T is not normally distributed. [2] describes some such examples encountered in logic level power estimation, and their effect on overall accuracy. We expect that at the transistor level this effect would be even less pronounced due to the number of variables (nodal power dissipation) being higher than at the logic level, thus causing the total power dissipation to be *better behaved* from the perspective of applying the Central Limit theorem as described in Section 3.

Table 3 compares the power estimates from the Monte Carlo approach and the 10000 vector direct simulation to a direct simulation using a small vector set with the worst case switching at the primary inputs. The results demonstrate that if the inputs are not biased properly and/or the input vector set is not statistically large enough, a simulation based result can be off the actual power dissipation by as much as 300%.

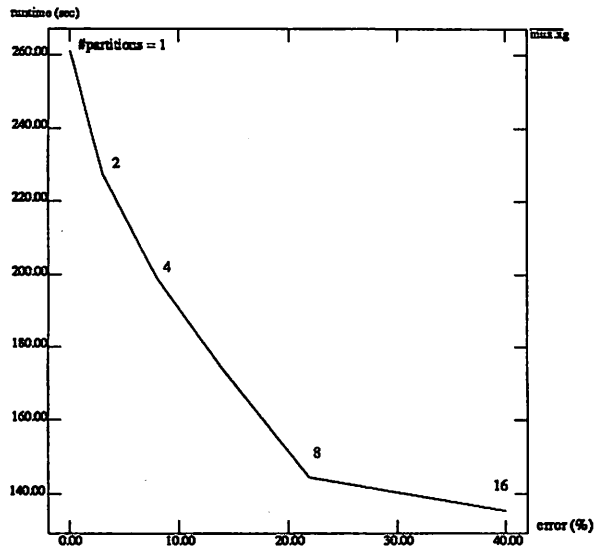


Figure 8: Runtime vs. err or trade-off as a function of number of partitions: 16-bit 2-to-1 mux

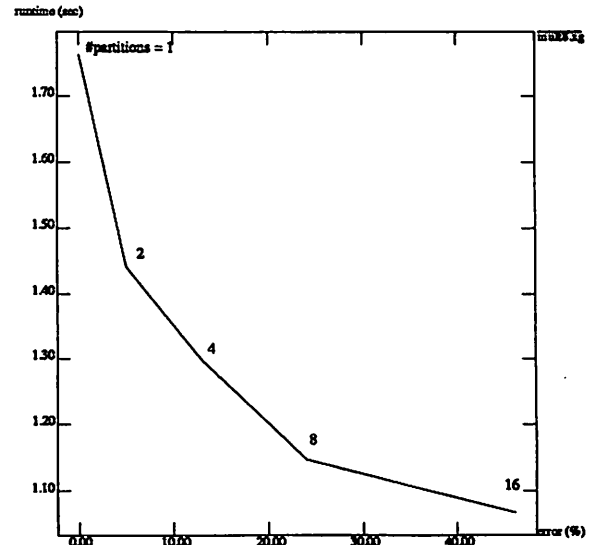


Figure 9: Runtime vs. err or trade-off as a function of number of partitions: 8 bit wallace multiplier

Figures 8 and 9 show the effect of varying the granularity of the partitioning for estimation on the accuracy and runtime of the program. The results demonstrate that partitioning does speed-up the estimation process. As mentioned in Section 3.3, since the statistical model used for information propagation between partitions does not account for the spatiotemporal correlations, accuracy considerations may constrain the number of partitions employed in this approach.

5 Conclusions

We have applied a statistical approach to power estimation at the transistor level. While transistor level power estimation can be very accurate, for practical usage, it suffers from two disadvantages: large runtimes, and pattern dependency. We proposed a two-point approach to overcome these problems: we use a Monte Carlo based method for estimation without requiring external input stimuli and a partitioning for estimation approach to speed-up the estimation process.

In this work we have,

- extended the concept of transitional density [14] (used at the gate level power estimation [2]) to analog waveforms. We have shown that a similar approach can be adopted for power estimation at the transistor level. Specifically, we have proved that for each device voltage variable, a companion stochastic process can be constructed which converges to the device power dissipation everywhere and is strict-sense stationary and mean-ergodic.
- Developed a formal stopping criterion to guarantee a desired error bound with a specified confidence level, under the normal strobilation assumption. We demonstrated the validity of this approach with experimental results and analyzed the implications of its violation.
- proposed the use of partitioning for estimation to speed-up the power estimation problem. We showed that this approach can be used to exploit the multi-rate behavior and stiffness of a circuit from a power perspective to reduce the number of vectors required for the estimation. We showed that this can also improve the accuracy of the power estimate in cases of *stiff* circuits (where a single normal distribution may not be an appropriate model for the power dissipation of the entire circuit).
- presented a statistical approach for propagating signals and associated switching probabilities between partitions. We fit a statistical model to the output waveforms of a subcircuit and use this model to generate input patterns for all its fanout subcircuits.

Experimental results were presented on a set of CMOS industry circuits.

6 Acknowledgments

I would like to thank Prof. Ernest S. Kuh and Dr. Vijay Nagasamy for many useful discussions and guidance throughout this work.

References

- [1] P. Buch, S. Lin, V. Nagasamy, and E. S. Kuh, "Techniques for fast circuit simulation applied to power estimation of CMOS circuits," *Proc. of the 1995 International Symposium on Low Power Design*, pp. 135-138, April 1995.
- [2] R. Burch, F. Najm, P. Yang, and T. Trick, "A Monte Carlo approach for power estimation," *IEEE Transactions on VLSI Systems*, Vol. 1, No. 1, pp. 63-71, March 1993.
- [3] A. P. Chandrakasan, S. Sheng, and R. Brodersen, "Low power CMOS digital design," *IEEE Transactions on Solid-State Circuits*, Vol. 27, No. 4, pp. 473-483, April 1992.
- [4] M. A. Cirit, "Estimating dynamic power consumption of CMOS circuits," *IEEE International Conference on Computer-Aided Design*, pp. 534-537, Nov. 1987.
- [5] A.-C. Deng, "Power analysis for CMOS/BiCMOS circuits," *Proc. of the International Workshop on Low Power Design*, pp. 1-6, 1994.
- [6] P. Gray, H.-S. Lee, J. Rabaey, C. Sodini, B. Wooley, "Challenges and opportunities in low power integrated circuit design," *SRC Publication Id S94019*, November 1994.
- [7] A. Hald, *Statistical Theory with Engineering Applications*, New York, John Wiley & Sons, Inc., 1952.
- [8] J. M. Hammersley, and D. C. Handscomb, *Monte Carlo Methods*, New York, John Wiley & Sons, Inc., 1964.
- [9] *HSPICE Version H92 User's Manual*, Meta-Software Inc., Campbell, CA.
- [10] S. M. Kang, "Accurate estimation of power dissipation in VLSI circuits," *IEEE Journal of Solid-State Circuits*, Vol. SC-21, pp. 889-891, Oct. 1986.
- [11] J. Kleckner, R. Saleh, and A. R. Newton, "Electric consistency in schematic simulation," *Proc. of the IEEE International Conference on Circuits and Computers*, pp. 30-34, October 1982.
- [12] S. Lin, E. S. Kuh, and M. Marek-Sadowska, "Stepwise equivalent conductance circuit simulation technique," *IEEE Transactions on CAD*, Vol. 12, No. 5, pp. 672-683, May 1993.
- [13] F. Najm, "Estimating power dissipation in VLSI circuits", Technical Report, University of Illinois at Urbana-Champaign, May 1994.
- [14] F. Najm, "Transition density, a new measure of activity in digital circuits", *IEEE Transactions on CAD*, Vol. 12, No. 2, pp. 310-323, Feb. 1993.
- [15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, New York, McGraw-Hill Inc., 1991.
- [16] K. Pearson, "On the theory of contingency and its relation to association and normal correlation," *Philos. Mag. Series*, Vol. 50, pp. 157-175, 1900.
- [17] R. Rubinstein, *Monte Carlo Optimization, Simulation and Sensitivity of Queuing Networks*, John Wiley & Sons, Inc., New York, 1986.