On Mixture Density and Maximum Likelihood Power Estimation via Expectation-Maximization

R. Chandramouli

Department of ECpE
Iowa State University
Ames, IA 50011
e-mail: mouli@eng.iastate.edu

Vamsi K. Srikantam

Hewlett Packard Labs
Palo Alto
CA 94304-1317
email: vamsi@hpl.hp.com

Abstract— A maximum-likelihood estimation procedure for computing the average power consumption of VLSI circuits is proposed. The method can handle data that has a mixture-density with multiple components unlike most of the previous approaches. An iterative computational procedure based on the expectation-maximization principle is also discussed. This can be used to estimate the parameters of an arbitrary (but finite) number of components of the probability distribution of the simulated power data. Experimental results for ISCAS '85 benchmark circuits and a large industrial circuit are given in order to validate the efficiency and practicality of the algorithm. Comparisons show that the proposed method estimates the multiple components (even those with a low probability of occurrence) while the Monte Carlo estimate captures *only* the most probable component.

I. Introduction

The increasing demand for personal computing devices and wireless communication equipment with real time applications has resulted in the need for designing circuits with low power consumption. Until recently, designers were mainly concerned with area and throughput as the important design parameters. The addition of power as a third parameter to the design search space has consequently lead to the exploration of the tradeoff between area, delay and power. Low power designs can be realized at various levels of the design cycle. Therefore, accurate estimation of power consumption at each level is necessary for the synthesis of these designs. However, power dissipation is an input dependent phenomenon and would require exhaustive simulation to get an accurate estimate. This becomes intractable for circuits with large number of inputs. Apart from dependence on the input pattern applied, accurate estimation of power dissipation is difficult due to its dependence on the the delay model and the circuit structure. Hence, two approaches for power estimation have been investigated in the literature: non-simulative (probabilistic) and simulative.

The non-simulative techniques use probabilities to describe the set of logic signals. These logic signals are modeled as stochastic processes having signal and transition probabilities associated with them. The signal probability is defined as the probability of having a logic one and the transition probability represents the proportion of transitions on that signal. But this approach has the drawback that it is not very accurate since the correlations (temporal, spatial and spatio-temporal) at the inputs and the internal nodes have to be accurately modeled. This is further complicated by the fact that real delay models should be used to get an accurate estimate. This is very expensive and hence these approaches trade off accuracy for speed.

The simulative or statistical methods take care of the input dependence by proper choice of the input vectors. Most of the previous work in the literature has focussed on developing a power estimation methodology which consists of the following three parts: input vector generation, a power estimator and stopping criterion. The input pattern generator captures the spatial and temporal correlations of the environment that the circuit is placed in, and generates the required inputs. The power estimation or simulation engine computes the power of the circuit under consideration. The stopping criterion determines when to stop the simulation when a desired accuracy is obtained with a specific confidence level. One of the earliest works that proposes such a power estimation engine is described in [1] which is based on *Monte Carlo Simulation*. The idea is to use simulation and compute the average power consumption of the circuit repeatedly. Then an adaptive update rule is used to estimate a final average power value. A stopping criterion determines when to stop the iterative estimation. In [1], the power estimation problem is reduced to that of mean estimation. In other words, the power is estimated as the mean of several P_T (where P_T is the average power observed for a time interval of length T) values. The stopping criterion for the interval estimator is derived assuming that P_T is normally distributed. This gives the total average power but not the individual gate power. The work in [2] provides the node densities along with total power. The stopping criterion is based on the absolute error bound on the low activity nodes in

the circuit. These nodes may take a long time to converge but their contribution to overall power is very less since they are low activity nodes. The main drawback of this approach is that it may take a lot of time for large circuits.

Two techniques recently investigated are population pruning and stratified random sampling [3, 4]. Population pruning removes the area of that interval which is not in the quantile interval of interest. In population pruning, given a confidence interval, samples whose probability of belonging to that family of observations are small are removed. Stratified sampling [3] uses ideas from *survey sampling*. The simulated samples are partitioned such that each partition is more homogeneous. This way, the sample distributions have a good chance of being normal distributed. Once normality is achieved the estimation problem becomes easier. A non-parametric sampling method has been proposed in [5] to achieve a trade off between accuracy and computational efficiency. Order statistics are used to handle a circuit with any power distribution. But this technique may lead to *over-sampling*. Power estimation for sequential circuits is complicated due to the presence of feedbacks. Some of the works that discuss this problem can be found in [6]-[9]. These show that the proper choice of initial states and the length of the *warm-up* period of Monte Carlo method give accurate estimates. An overview of the various *state-of-the-art* power estimation techniques can be found in [9].

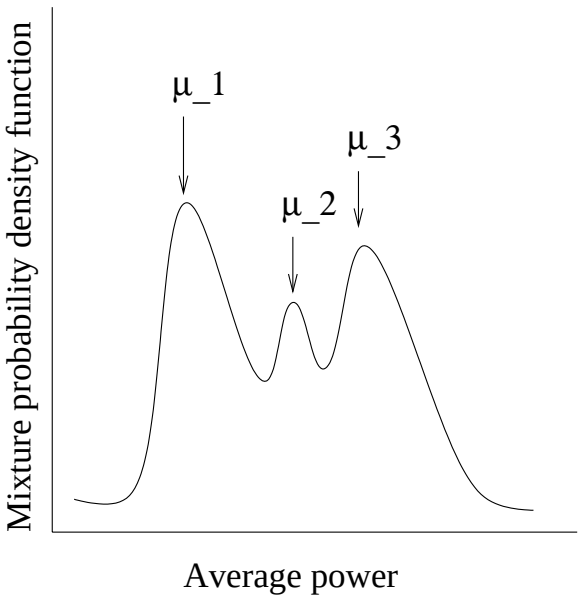

Fig. 1. A normal mixture density with three components

Most of the methods discussed so far assume normality for the probability distribution of the power consumption of the circuits. It is justified to a large extent by the central limit theorem and a excessive simulation of various kinds of circuits. Unfortunately, this may not be the case in many practical scenarios. The probability distribution may be bi-modal or even multi-modal, called, *mixture density models* as shown in Fig. 1. The reasons for this behavior could be

- Structure of the circuit and/or inputs
- Limited sampling due to cost
- The conditions for central limit theorem to hold may not be valid for the given data

For example, a circuit may be enabled or disabled by means of a signal and based on the activity on that signal, the circuit can show very different power consumptions. Hence, such power distributions need to be captured for accurate power estimation and optimization. In general, the power estimation techniques based on simulation and Monte Carlo methods produce estimates that correspond to most probable component of the mixture density. This is because, the stopping criterion does not reflect the fact that there could be more than component in the data distribution. Hence, when the estimator observes sufficient number of samples from the most probable component the stopping criterion is reached and the estimation procedure stops.

The aim of this paper is to develop and discuss an algorithm that accounts for, and estimates the parameters of the multiple components (if any) in the distribution of the power samples. The estimate is *maximum likelihood*. This means, the estimate that the algorithm produces is the most probable given the observed samples. In Section II we discuss the idea behind mixture density followed by maximum likelihood estimation. A general procedure for parameter estimation called *expectation-maximization* is discussed for an arbitrary mixture density in Section III. This is followed by the derivation of iterative parameter estimation algorithm for normal mixture densities based on expectation-maximization. Some properties of this algorithm are also discussed. This is followed by validating the algorithm's predicted performance by experiments in Section IV. Two kinds of experiments are performed. First, the performance for ISCAS '85 benchmark circuits is discussed. This is done so that comparison of other algorithms with the proposed one is justified. We compare and discuss our results with the Monte Carlo based approach. This is followed by experiments with a large industrial circuit to demonstrate the practicality of the algorithm. The paper is summarized along with brief conclusions in Section V.

II. ML Estimation for Mixtures

In this section we briefly describe the mathematical ideas and concepts behind mixture distributions of a ran-

dom variable followed by the maximum-likelihood estimation procedure. Suppose a random variable, P, takes values in a sample space, $\mathcal{P}$, and admits a probability density function (pdf). Let the pdf be of the form

$$f(p) = \pi_1 f_1(p) + \pi_2 f_2(p) + \ldots + p_k f_k(p) \quad (p \in \mathcal{P}), \quad (1)$$

where $\pi_j > 0$, $j = 1, 2, \ldots, k$; $\sum_j \pi_j = 1$ and $f_j(.) \geq 0$, $\int_{\mathcal{P}} f_j(p)dp = 1$. We then say that P has a *finite mixture distribution* and that $f(p)$ in Eq. (1) is a *finite mixture density function*. The parameters $\pi_1, \pi_2, \ldots, \pi_k$ are called the *mixing weights* and $f_1(.), f_2(.), \ldots, f_k(.)$ the *component densities* of the mixture. Frequently, $f_1(.), f_2(.), \ldots, f_k(.)$ will have parametric forms, parameterized respectively by the elements of the set $\theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$. Then from Eq. (1) we have

$$f(p) = \pi_1 f_1(p|\theta_1) + \pi_2 f_2(p|\theta_2) + \ldots + \pi_k f_k(p|\theta_k) \quad (2)$$

Let the complete collection of all the parameters in the mixture model be $\Psi = \{\pi_1, \pi_2, \ldots, \pi_k, \theta_1, \theta_2, \ldots, \theta_k\}$. The finite mixture density then takes the form

$$f(p|\Psi) = \sum_{j=1}^{k} \pi_j \tilde{f}(p|\theta_j) \quad (3)$$

where, each of $\theta_1, \theta_2, \ldots, \theta_k$ belongs to the same parameter space, Θ (say) and $\tilde{f}(.|\theta)$ denotes a generic density function. Then, $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ may be thought of as defining a probability distribution over Θ, where $\pi_j = Pr(\theta = \theta_j)$, $j = 1, 2, \ldots, k$. If we denote the probability measure defined by π over Θ by $G_\pi(.)$ then Eq. (3) may be written in general as

$$f(p|\Psi) = \int_{\Theta} \tilde{f}(p|\theta) dG_\pi(\theta) \quad (4)$$

Let the observed values of the power consumption for various inputs be $\{p_1, p_2, \ldots, p_n\}$. Each of these is described by a parametric pdf of the form given in Eq. (3). The first issue is to decide the optimum number of components, k, in the mixture. We do not address this problem explicitly. However, we note that a good estimate of k can be obtained by methods such as *histogram segmentation*, *hypothesis testing* [11], or stratified sampling [3]. We treat k as a parameter for the problem under study. Given n statistically independent observations from the mixture their joint probability density function is the product of the individual densities. Therefore the likelihood function is given by

$$L(\Psi) = \prod_{i=1}^{n} \left[\sum_{j=1}^{k} \pi_j \tilde{f}(p_i|\theta_j) \right] \quad (5)$$

The maximum likelihood estimate (MLE) of Ψ, say, Ψ^* is then defined as

$$\Psi^* = \max_{\Psi} L(\psi) \quad (6)$$

Usually, for computational simplicity the log-likelihood, $\mathcal{L}(\Psi) = log_e L(\Psi)$ is maximized to obtain the MLE. MLE is a popular estimation technique due to the following reasons : (a) well studied asymptotic theory, (b) estimates can be computed easily, (c) can be combined with the likelihood-based statistical inference methods.

III. Expectation-Maximization Algorithm

The EM algorithm for MLE was first proposed in [12]. It produces the maximum likelihood estimate of the unknown parameters iteratively. Two steps – expectation and maximization (EM) are iterated until some convergence condition is met. It does not depend on gradient computations like the stochastic approximation methods. Note that the maximization of $\mathcal{L}$ numerically is difficult because it contains terms involving the log of a sum. If we know the component of the mixture from which an observed data point is generated then the problem would be simpler. Since this information is not known when the data is observed the observations are termed as *incomplete*. The idea is to maximize $\mathcal{L}$ using the incomplete data. Define the *complete* data to be the fully categorized data, *i.e.*, $\{y_i = 1, 2, \ldots, n\} = \{(p_i, \mathbf{z}_i);\ i = 1, 2, \ldots, n\}$ where each $\mathbf{z}_i = (z_{ij}, j = 1, 2, \ldots, k)$ is an indicator vector of length k with 1 in the position corresponding to the mixture component to which p_i belongs. Now, the likelihood for the complete data can be written as

$$g(y_1, y_2, \ldots, y_n|\Psi) = \prod_{i=1}^{n} \prod_{j=1}^{k} \pi_j^{z_{ij}} [f_j(p_i|\theta_j)]^{z_{ij}} \quad (7)$$

which, with the logarithm becomes

$$\begin{aligned} \mathcal{L}_0 &= \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} log_e \left(\pi_j f_j(p_i|\theta_j)\right) \\ &= \sum_{i=1}^{n} \mathbf{z}_i^T V(\pi) + \sum_{i=1}^{n} \mathbf{z}_i^T U_i(\theta) \end{aligned} \quad (8)$$

where $V(\pi)$ has $log_e \pi_j$ as its jth component and $U_i(\theta)$ has the jth component $log f_j(p_i|\theta_j)$[1]. We now note that each sample is associated with its component density. The maximization of this new likelihood can be decoupled into a set of simpler maximizations. Maximization for each of the densities in the mixture model can be performed separately. The data observations from each of the densities can be used to estimate its parameters. However, we do not know the value of z_{ij} *a priori* making the data incomplete. To overcome this difficulty the expected value of the log-likelihood, $\mathcal{L}_0$, can be maximized instead. In [12] it is shown that if a certain value of the parameter θ increases the expected value of the log-likelihood (Eq. (8)) then the log-likelihood also increases. Our intention is to

[1] T denotes the transpose of a vector

compute $\Psi^* = \max_\Psi L(\Psi)$ for the incomplete data. Let y denote a complete version of p and $\mathcal{Y}(p)$ the set of all possible such y. Clearly, the cardinality of $\mathcal{Y}(x)$ is equal to k^n. Then, starting from some initial value, $\Psi(0)$, a sequence of estimates, $\{\Psi^{(m)}\}$ is obtained. Each iteration of the EM algorithm consists of the following two steps : Iteration count, $m = m + 1$

E Step:
Evaluate

$$\begin{aligned} Q(\Psi, \Psi^{(m)}) &= E[log\ g(y|\Psi)|p, \Psi^{(m)}] \\ &= \sum_{i=1}^{n} w_i(\Psi^{(m)})^T V(\pi) + \sum_{i=1}^{n} w_i(\Psi^{(m)}) U_i(\theta) \end{aligned}$$

where $w_i(\Psi^{(m)}) = E[z_i|p_i, \Psi^{(m)}]$ and $w_{ij} = [w_i(\Psi^{(m)})]_j = \pi_j^{(m)} f_j(p_i|\theta_j^{(m)})/f(p_i|\Psi^{(m)})$ for each i, j. These are the weights correspond to the probabilities of the ith observation generated from j component conditional on p_i and $\Psi^{(m)}$.

M Step:
Compute $\Psi^{(m+1)} = \arg\max_\Psi Q(\Psi, \Psi^{(m)})$. When π and θ are distinct it can be shown that the M step for π is

$$\pi_j^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} w_{ij}(\Psi^{(m)}), \quad j = 1, 2, \ldots, k \tag{10}$$

The iterations are stopped when $|\pi_j^{(m+1)} - \pi_j^{(m)}| < \epsilon$, $j = 1, 2, \ldots, k$ for an arbitrary $\epsilon > 0$. Note that the M step is problem dependent. Using Jensen's inequality we see that $L(\Psi^{(m+1)}) \geq L(\Psi^{(m)})$, $m = 0, 1, 2, \ldots$ This shows that the likelihoods of the incomplete data are monotonic. Further theoretical properties of the EM procedure for mixture density estimation are not discussed here due to space limitations.

A. EM for Gaussian Mixture Density

In this section we explicitly derive the EM steps for the d-dimensional Gaussian mixture density. The d-dimensional Gaussian mixture density is given by

$$f(\mathbf{p}) = \sum_{j=1}^{k} \frac{\pi_j}{((2\pi)^d |\Sigma_j|)^{1/2}} exp\left\{ -\frac{1}{2}(\mathbf{p} - \bar{\mu}_j)\Sigma_j^{-1}(\mathbf{p} - \bar{\mu}_j)^T \right\}$$

where $\bar{\mu}_j$, Σ_j, $j = 1, 2, \ldots, k$ contain parameters of the component densities and π_j, $j = 1, 2, \ldots, k$ are the unknown probability of occurrence of each component in the mixture. Σ_j is a positive definite symmetric matrix. Let $\{\mathbf{p}_i\}_{i=1}^n$ denote n, d-dimensional Gaussian random vectors. Assume that each Gaussian component has a covariance matrix $\Sigma_j = \sigma_j^2 \mathbf{I}$. Then the parameters to be estimated are $\theta_j = (\bar{\mu}_j, \sigma_j)$, and π_j, $j = 1, 2, \ldots, k$. In our notation, $\Psi = (\pi_j, \bar{\mu}_j, \sigma_j;\ j = 1, 2, \ldots, k)$. We can then show that

E Step:

$$\begin{aligned} w_{ij} &= E[z_{ij}|\mathbf{p}_i, \Psi(m)] \\ &= \frac{\sigma_j^{-d} exp\{-||\mathbf{p}_i - \bar{\mu}_j(m)||^2/2\sigma_j^2(m)\}}{\sum_{l=1}^{k} \sigma_j^{-d}(m) exp\{-||\mathbf{p}_i - \bar{\mu}_l(m)||^2/2\sigma_l^2(m)\}} \end{aligned} \tag{11}$$

M Step:

$$\pi_j(m+1) = \frac{1}{n} \sum_{i=1}^{n} w_{ij} \tag{12}$$

$$\bar{\mu}_j(m+1) = \frac{\sum_{i=1}^{n} w_{ij} \mathbf{p}_i}{\sum_{i=1}^{n} w_{ij}} \tag{13}$$

$$\sigma_j(m+1) = \frac{\sum_{i=1}^{n} w_{ij} ||\mathbf{p}_i - \bar{\mu}_j(m+1)||^2}{\sum_{i=1}^{n} w_{ij}} \tag{14}$$

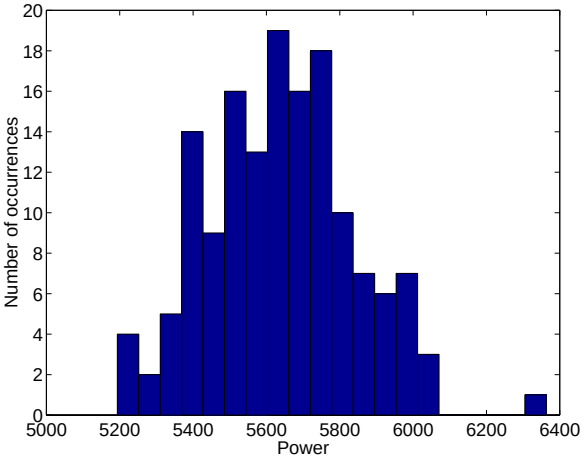

Fig. 2. Histogram of average power consumption for C2670

The algorithm is summarized in the pseudo-code below.

{

E-Step:

1. Compute the number of components k by histogram (of the simulated average power values) segmentation.

2. Compute initial estimates for π_j, μ_j, and σ_j, $j = 1, 2, \ldots, k$ by the k-means algorithm and start the E-Step iteration.

3. Given the parameter estimates at iteration m and the simulated data, compute the new weights w_{ij} using Eq. (11).

M-Step:

TABLE I
COMPARISON OF EM POWER ESTIMATES (IN μW) FOR A TWO COMPONENT MIXTURE MODEL AND MONTE CARLO (MC) APPROACH

Circuit	(π_1, π_2)	(μ_1, μ_2)	(σ_1, σ_2)	MC estimate
C880	(0.896,0.104)	(1353.3,1453.4)	(59.3,14.12)	1367.08
C2670	(0.263,0.73)	(5497.9,5690.0)	(144,195.6)	5621.37
C1355	(0.028,0.972)	(3245.5,3007.8)	(16.87,88.32)	3017.57
C1908	(0.013,0.987)	(6025.2,5192.4)	(4.35,284.2)	5194.05
C499	(0.083,0.917)	(1219.0,1216.4)	(4.9,33.47)	1212.50
C432	(0.018,0.982)	(1069.4,888.35)	(5.03,60.79)	883.47

TABLE II
EM POWER ESTIMATE FOR AN INDUSTRIAL CIRCUIT

Component No.	Prob.	Mean
1	0.128774	102.000000
2	0.061386	109.003657
3	0.071737	111.999419
4	0.269919	115.000597
5	0.400315	117.999408
6	0.067869	128.000000

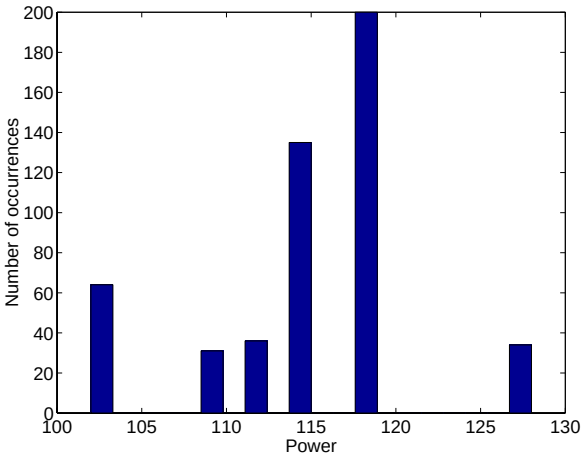

Fig. 3. Histogram of power consumption for an industrial circuit

1. Use the weight estimate from E-Step to compute the component probabilities at $m+1$ iteration, $\pi_j(m+1)$, $j = 1, 2, \ldots, k$ using Eq. (12).
2. Compute the component means at iteration $m+1$, $\mu_j(m+1)$, $j = 1, 2, \ldots, k$ using Eq. (13) and the current weights.
3. Compute the component variances at iteration $m+1$, $\sigma_j(m+1)$, $j = 1, 2, \ldots, k$ using Eq. (14)
4. If absolute value of the (current estimate of π_j - previous estimate of π_j <)ϵ, $j = 1, 2, \ldots, k$ for a given $\epsilon > 0$ then stop.
If not go to E Step and iterate again.
}

IV. EXPERIMENTAL RESULTS

To test the proposed algorithm, we performed two experiments. In the first experiment, ISCAS '85 benchmarks were used. The circuits were mapped and simulated in SIS [10]. The circuits were mapped to the *lib2.genlib* library and power was estimated using simulation. We used a small sample size of 150. There are two reasons for this. First, it is known that for small sample sizes the central limit theorem may not hold. This is because the rate of convergence of the distribution of the samples to the Gaussian distribution is affected by the sample size and other factors. Also, in may circumstances obtaining a large number of samples could be costly. So, our aim is to study the performance of our estimator algorithm for small number of observations and compare it with a popular existing estimator. The histogram of the power consumption of each circuit is segmented to get an estimate for the number of components, k, in the mixture density. Fig. 2 is the histogram for the power consumption of the C2670 benchmark circuit. We can see that a two component mixture model would be a good starting point. After the initial estimate for the number of components, we then used the k-means clustering algorithm [13] to get an initial estimate for the prior probabilities of each mixture component, means and variances of the mixture density. This estimate is then used by the EM algorithm as initial values along with the training set of power observations to produce the final parameter estimates. Table I shows the comparison of the estimated power for various benchmark circuits using the EM algorithm and the Monte Carlo estimator. Clearly, the proposed estimate is more general in nature. Even the small probability components are captured in the approach. However, the Monte Carlo method produces an estimate of only the most probable component–a well-known disadvantage of this approach.

In the second experiment, an industrial circuit was used. The circuit is very large and was synthesized and mapped using commercial tools. The circuit was described in Verilog and synthesized using Design Compiler from Synopsys. Technology mapping was performed using the HP 0.25 micron CMOS libraries. The gate level netlist with all delays annotated was then simulated using Verilog and the power estimates were obtained using Sente's WattWatcher tool. The circuit was simulated extensively to obtain 500 samples of the consumed average power.

The histogram of the simulated power is shown in Fig 3. As can be seen from the figure, the power distribution is a mixture of six components. The multi-modal power distribution is due to the fact that the circuit is large and consists of many feedback (or sequential) elements along with a RAM. Hence, many parts of the circuit have different power dissipation under different input conditions. For each simulation run enough vectors were given for the circuit to be in a steady state before applying the vectors for which the power was measured. Some clock gating is also used in the circuit; this further causes the power dissipation fluctuation based on the activity of the control signal. The estimates of all six components as seen in the histogram is summarized in Table II. The table shows the probability of occurrence of the various components along with the value of their means. We also note that k is a free parameter in the algorithm. The user can choose the value of k. If a component has a very small probability of occurrence, it can be merged with its neighboring component. This analysis with a real large design using the state of the art CAD tools validates the practicality of the proposed algorithm. In deep sub-micron regimes, power consumption can be affected by many secondary factors which contribute to bi-modal or multi-modal power distribution. The proposed algorithm estimates these multiple components (or modes) thereby giving a better power distribution estimate which will be very useful in power optimization procedures.

V. Summary and Conclusions

A maximum-likelihood power estimation procedure is proposed and discussed for simulated power data with a normal mixture density. Unlike most of the previous approaches, this method can be used to estimate the average power consumption with an arbitrary (but finite) number of components in a mixture density. An efficient iterative computational procedure based on the EM algorithm is given. Estimates based on simulations for the ISCAS '85 benchmark circuit and a large industrial circuit are given. From these experiments we conclude that the proposed algorithm is efficient and, can handle the randomness in simulated power values in both small and large circuits. Comparison with Monte Carlo estimation shows that the proposed method is preferable due to its generalization capacity.

Acknowledgments

The authors thank Srinath Chavali at ViewLogic for providing the data and Monte Carlo estimates for the benchmark circuits.

References

[1] R. Burch, F. Najm, P. Yang, and T. Trick, "A Monte Carlo Approach for Power Estimation", *IEEE Transactions on VLSI systems*, Volume 1, pp.63-71, March 1993.

[2] M. Xakellis and F. Najm, "Statistical estimation of switching activity in digital circuits," *Proc. of IEEE/ACM Design Automation Conf.*, pp 728-733, 1994.

[3] C. S. Ding, C. T Hsieh, Q. Wu, and M. Pedram, "Stratified random sampling for power estimation," *Proc. of IEEE/ACM ICCAD*, San Jose, CA, pp. 577-582, 1996.

[4] C. S. Ding, Q. Wu, C. T Hsieh, and M. Pedram, "Statistical estimation of the cumulative distribution function for power dissipation in VLSI circuits," *Proc. of Design Automation Conference*, pp. 371-376, March 1997.

[5] L. P. Yuan, C. C. Teng, and S. M. Kang, "Statistical estimation of average power dissipation using non-parametric techniques," *IEEE transactions on VLSI*, vol 6, number 1, pp 65-73, March 1998.

[6] F. Najm, S. Goel, and I. Hajj, " Power estimation in sequential circuits,"*Proc. Design Automation Conference*, pp 635-640, 1995.

[7] T. L. Chou, and K. Roy, "Accurate power estimation of CMOS sequential circuits," *IEEE Transactions VLSI systems*, vol 4, no 3, pp 369-380, Sept 1996.

[8] L. P. Yuan, C. C. Teng, and S. M. Kang, "Statistical Estimation of Average Power dissipation in Sequential circuits", *Proc. Design Automation Conference*, pp.377-382, 1997.

[9] M. Pedram, "Power simulation and estimation in VLSI circuits," *The VLSI Handbook,* Edited by W-K. Chen. The CRC Press and the IEEE Press, 1999. (Available at *http://atrak.usc.edu/ massoud*)

[10] E. M. Sentovich et al., "Sequential circuit design using synthesis and optimization," *Proc. of IEEE ICCD*, pp. 328-333, 1992.

[11] E.L. Lehmann, *Testing statistical hypotheses*, Wiley, New York, 1986.

[12] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)," *J.R. Statist. Soc. B*, Vol. 39, pp. 1-38, 1977.

[13] A. Gersho and R.M. Gray, *Vector quantization and signal compression*, Kluwer, Boston, 1992.