Cycle-Accurate Macro-Modeling: Algorithm and Implementation

Qing Wu, Qinru Qiu, Chih-Shun Ding
and Massoud Pedram

CENG 98-11

Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, California 90089-2562
(213 740-4458)
May 1998

Cycle-Accurate Macro-Modeling: Algorithm and Implementation

Qing Wu, Qinru Qiu, Massoud Pedram
Department of Electrical Engineering-Systems
University of Southern California

Chih-Shun Ding
Rockwell Semiconductor Systems
Newport Beach Division

Abstract

In this paper we present a methodology and techniques for generating cycle-accurate macro-models for RT-level power analysis. The proposed macro-model predicts not only the cycle-by-cycle power consumption of a module, but also the moving average of power consumption and the power profile of the module over time. We propose an exact power function and approximation steps to generate our power macro-model. First order temporal correlations and spatial correlations of up to order 3 are considered in order to improve the estimation accuracy. A variable reduction algorithm is designed to eliminate the "insignificant" variables using a statistical sensitivity test. Population stratification is employed to increase the model fidelity. Experimental results show our macro-models with 15 or fewer variables, exhibit <5% error for average power and <20% errors for cycle-by-cycle power estimation compared to circuit simulation results using Powermill.

I. INTRODUCTION

Due to rapid progress in the semiconductor manufacturing, the device density and operating frequency have greatly increased, making power consumption a major design concern. High power consumption exacerbates the reliability problem by raising the die temperature and by increasing current density on the supply rails. It also reduces the battery life which is a key concern in portable devices. Therefore, low power design requirements are driving a new breed of computer aided design methodologies and tools which in turn rely on accurate and efficient estimation tools at various design abstraction levels.

Power estimation at RT level is crucial in achieving a short design cycle. The standard hierarchical simulation approach to RT-level power estimation consists of three steps: 1) functionally simulate the RT-level description and collect the input sequences for each circuit block. 2) simulate each block at gate or circuit-level using the collected input sequences. 3) add the power consumption for all blocks to produce the power consumption of the whole circuit. The disadvantage of this approach is that it requires the interaction between RT-level simulators and low-level simulators and that power evaluation is actually done at gate-level or circuit-level where the simulation speed is low.

Alternatively, one could use the macro-modeling technique for power estimation at RT-level. In this technique, low-level simulations of modules under their respective input sequence is replaced by power macro-model equation evaluation (which can be performed very fast).

Macro-modeling techniques use capacitance models for circuit modules and activity profiles for data or control signals [1-5]. The simplest form of the macro-model equation is given by:

$$Power = \frac{1}{2} V^2 \cdot f \cdot C_{eff} \cdot SW \tag{1.1}$$

where C_{eff} is the effective capacitance, SW is the module activation ratio, and f is the clock frequency. The Power Factor Approximation (PFA) technique [1] uses an experimentally determined weighting factor, called the power factor, to model the average power consumed by a given module over a range of designs.

To improve the accuracy, more sophisticated macro-model equations have been proposed. Dual Bit Type model, proposed in [2], exploits the fact that, in the data path or memory modules, switching activities of high order bits depend on the temporal correlation of data while lower order bits behave similarly to white noise data. Thus a module is completely characterized by its capacitance models in the MSB and LSB regions. The break-point between the two regions is determined based on the signal statistics collected from simulation runs. The Activity-Based Control (ABC) model [4] is proposed to estimate the power consumption of random-logic controllers. An Input-Output model has been proposed in [5] to capture the relation between power and input signal probability, input transition density, and output transition density. By introducing variables related to output activity, the Input-Output model improves the estimation accuracy compared to the models which do not make use of the output information. One common feature of the above macro-model techniques is that, they only provide information about average power consumption over a relatively large number of clock cycles.

The above techniques, which are suitable for estimating the average-power dissipation, are referred to as *cumulative power macro-models*. In some applications, however, estimation of average power only is not sufficient. Other important tasks include the estimation of the k-cycle moving average of the power, power profiling on a cycle-by-cycle basis, and estimation of the rate of current change from one cycle to next. This information is crucial for circuit reliability (maximum current limits, heat dissipation and temperature gradient calculation, latch-up conditions) analysis, DC/AC noise analysis (DC drop and inductive bounce on power and ground lines), and design optimization (power/ground net topology, construction and sizing, number and placement of decoupling capacitors, buffer insertion, etc.). For example, the k-cycle average power can provide power consumption information for any given window of time. To perform these tasks requires knowing the power consumption value for every clock cycle. If the macro-modeling technique does not provide such information, the circuit designers will have to resort to gate-level or circuit-level simulation again. Consequently, cumulative macro-models are considered to have limited use.

The notion of *cycle-accurate macro-models* which was proposed in [6] is described next. Let P_k denote the power consumption of some module in clock cycle k, then we can write:

$$P_k = F(V_{k-1}, V_k) \tag{1.2}$$

where V_k and $V_{k\text{-}1}$ denote the input vectors applying to the module at cycles k and k-1, and F is some function of the input vector pairs. The goal of power macro-modeling is to find function F, given an input vector sequence V (the so called *training set*) for the module and given the corresponding power consumption values.

In this paper, we propose the methodology of building cycle-accurate macro-models for circuit modules. Compared to previous work, our approach makes the following tangible contributions:

1. The macro-model generated by our approach can predict the cycle-based power consumption, as well as the average power. Our cycle-accurate macro-model equation can be easily transformed into a cumulative macro-model equation.

2. We present an exact power consumption function which captures the relation between the circuit power consumption and the spatial-temporal correlations of primary inputs. This is used as the starting point for our macro-model generation. It can be a useful guide for generating macro-models for other purpose.
3. We introduce piecewise linear power macro-model equations to increase the fidelity of the macro-model.
4. A statistical significance test is proposed to eliminate "insignificant variables" and therefore simplify the macro-model equation.
5. Because of the statistical nature of our macro-model, it can be validated and improved by statistical methods. The estimation error can be predicted for given confidence level.

Experimental results show that, our cycle-accurate macro-models having 15 or fewer variables, exhibit <5% (3.2% maximum and 1.1% minimum) error in average power, and <20% (19.3% maximum and 6.2% minimum) error in cycle power compared to circuit simulation results using Powermill [7].

This paper is organized as follows. Section II gives the theoretical background for regression analysis. Section III discusses a procedure of building the macro-model whereas Section IV presents the experimental results. Section V will discuss some applications of cycle-accurate macro-models. Section VI is the conclusion of our work.

II. BACKGROUND

2.1 Introduction to linear regression analysis

We define a cycle-accurate power macro-model as a linear function between estimated power dissipation of a vector pair and the characteristic values of the vector pair, that is, we write:

$$\hat{P} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{2.1}$$

where $\hat{P}$ is the predicted power dissipation, $\beta_0, \beta_1, \cdots, \beta_k$ are constants called the regression coefficients or parameters of the macro-model, and $X_1, X_2, \cdots, X_k$ are characteristic variables extracted from the input vector pair. The methods for extracting values of $X_1, X_2, \cdots, X_k$ will be discussed in Section 3.1. The regression parameters are calculated by doing least-squares-fit during the linear regression analysis.

Based on the theory of linear regression analysis [8], we can define the relation between the actual power P (e.g. the power value simulated by Powermill) and the estimated power as:

$$P = \hat{P} + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \tag{2.2}$$

where ε is called the residual term of the linear regression model and follows a normal distribution with mean value 0 and variance σ^2. Equation (2.2) means that P is a random variable which follows a normal distribution with mean value $\hat{P}$ and variance σ^2.

Assume that we have been given the equation form of the macro-model as in Eqn.(2.1) and have performed Powermill simulations (observations) on m randomly sampled vector pairs in the population (this set of m vector pairs is referred to as the *training set*) so that we have obtained m simulation results (observation values) of power consumption. The linear regression model for vector pairs from the training set can be written as:

$$P_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k} + \varepsilon_i, \quad i = 1,2,\cdots,m \tag{2.3}$$

or in matrix form as:

$$\mathbf{P} = \mathbf{X}\beta + \varepsilon \tag{2.4}$$

where P_i's are random variables corresponding to observations: $(x_{i,1}, x_{i,2}, \cdots, x_{i,k})$ $i = 1,2,..,m$; $\beta_0, \beta_1, \cdots, \beta_k$ are the regression coefficients; $x_{i,1}, x_{i,2}, \cdots, x_{i,k}$ are known values derived from the input vector pair $(V_{i,1}, V_{i,2})$; and ε_i's are independent random variates representing deviation from the mean value of power with variance $\mathbf{VAR}[\varepsilon_i] = \sigma^2$, and $\mathbf{Cov}[\varepsilon_i, \varepsilon_j] = 0$, for $i \neq j$. Consequently, the random vector $\mathbf{P}$ has an expected value of $\mathbf{E}[\mathbf{P}]=\mathbf{X}\beta$ and the variance-covariance matrix of $\mathbf{P}$ is $\mathbf{Cov}[\mathbf{P}] = \sigma^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

The β coefficients are estimated using the least squares estimator by substituting the actual power values for $\mathbf{P}$:

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{P} \tag{2.5}$$

where

$$\underset{(k+1)\times 1}{\mathbf{b}} = [b_0, b_1, \cdots, b_k]^{\mathrm{T}} \tag{2.6}$$

It has been proven in [8] that the least squares estimator is an unbiased estimator for β, i.e., $\mathrm{E}[\mathbf{b}]=\beta$.The estimated (fitted) power from macro-model is given by the multiplication of input variables and the estimated coefficients:

$$\hat{\mathbf{P}} = \left[\hat{P}_1, \hat{P}_2, \cdots, \hat{P}_m\right] = \mathbf{Xb} \tag{2.7}$$

and the residual terms (error) are defined as the difference between the fitted power and observed (actual) power:

$$\mathbf{e} = [e_1, e_2, \cdots e_m] = \mathbf{P} - \hat{\mathbf{P}} = \mathbf{P} - \mathbf{Xb} \tag{2.8}$$

In the following, we define some relevant terms for regression analysis:

sum of squares error: $SSE = \sum_{i=1}^{m} e_i^{\,2}$

mean squares error: $MSE = SSE/(m-k-1)$

regression sum of squares: $SSR = \sum_{i=1}^{m} (\hat{P}_i - \overline{P})^2$

regression mean squares: $MSR = SSR/k$

coefficient of multiple correlation: $R = \sqrt{SSR/(SSR+SSE)}$

The above terms will be used in many discussions in the rest part of this paper. The statistical nature of the macro-model enables us to predict the accuracy level of fitted power value as follows. Given any input vector pair, the values of its characteristic variables (x_1, x_2, …, x_k) are first computed. The fitted (predicted) power is given by $\hat{P} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$. Given a confidence level 1-α, the confidence interval of the actual power P is defined as an interval [P_1, P_2] such that the probability that the actual power value lies inside this interval is 1-α. From this definition, we can compute the confidence interval for actual power P at any confidence level 1-α as:

$$[\hat{P} - t(1-\alpha/2; m-k-1)\cdot s[P], \hat{P} + t(1-\alpha/2; m-k-1)\cdot s[P]] \tag{2.9}$$

where $t(1\text{-}\alpha/2;m\text{-}k\text{-}1)$ is the $(1\text{-}\alpha/2)\times 100$ percentile point of the t distribution with degree of freedom of $(m\text{-}k\text{-}1)$ and $s[P]$ is the standard deviation of the new observation which is given by:

$$s[P] = \sqrt{MSE \cdot (1 + X^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} X)} \tag{2.10}$$

where **X** and *MSE* are the variable matrix and mean squares error of the training set, respectively.

2.2 Evaluating the quality of a macro-model

The quality of the macro-models can be evaluated in terms of the following criteria:

1. Correlation factor: From the coefficient of multiple correlation R, we derive a similar quantity r as:

$$r = 1/(1-R^2) = 1 + SSR/SSE \tag{2.11}$$

We call r the correlation factor of the macro-model. r is a monotonic-increasing function of R. In many applications of linear regression, r is a general measure of the quality of a regression model since it represents linearity of the model and the magnitude of the error. It also reflects the stability or fidelity of a macro-model. The higher the r value, the better the quality of the regression model. The r value may differ from one population to next for the same macro-model. Therefore, the r values of different macro-models should be compared only when they are subjected to the same input population.

2. Errors: Error in cycle power (ECP) gives the average error when estimating power on cycle by cycle basis while error in average power (EAP) gives the average error when estimating the average power. More precisely, we can write:

$$ECP = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{P}_i - P_i}{P_i}\right|, \qquad EAP = \frac{\sum_{i=1}^{n}\hat{P}_i - \sum_{i=1}^{n}P_i}{\sum_{i=1}^{n}P_i} \tag{2.12}$$

In most cases the training set only represents a small portion of the target population, and in some extreme cases, the training set will be totally different from the population. Designed based on such training sets, the macro-models can exhibit good quality on the training set and on populations which have similar characteristics as the training set, but not on other populations. To assess the quality of macro-model, accuracy comparison of macro-models should be carried out on populations (the set of input vector pairs and corresponding Powermill power values) whose behavior is different from that of the training set. On the other hand, the design of training set is very important. The more closely the training set represents the target population, the more accurate the resulting macro-model will be. In addition, it is very difficult to predict in advance what type of population that macro-model will be subjected to. Therefore, careful design of a good macro-model equation form is the key to reducing the estimation error. Last, but not least, extracting the $X_1, X_2, \cdots, X_k$ variables in Eqn.(2.1) from the input vector pair is important in designing a good macro-model.

III. MACRO-MODEL CONSTRUCTION

The overall macro-model generation procedure is described in Figure 1. The macro-model generation procedure consists of four major steps: variable selection, training set design, variable reduction, and least squares fit. Notice that the macro-model equation design refers to not only the form of the equation (linear function, values of coefficients), but also the procedure for extracting variable values from the input vector pair.

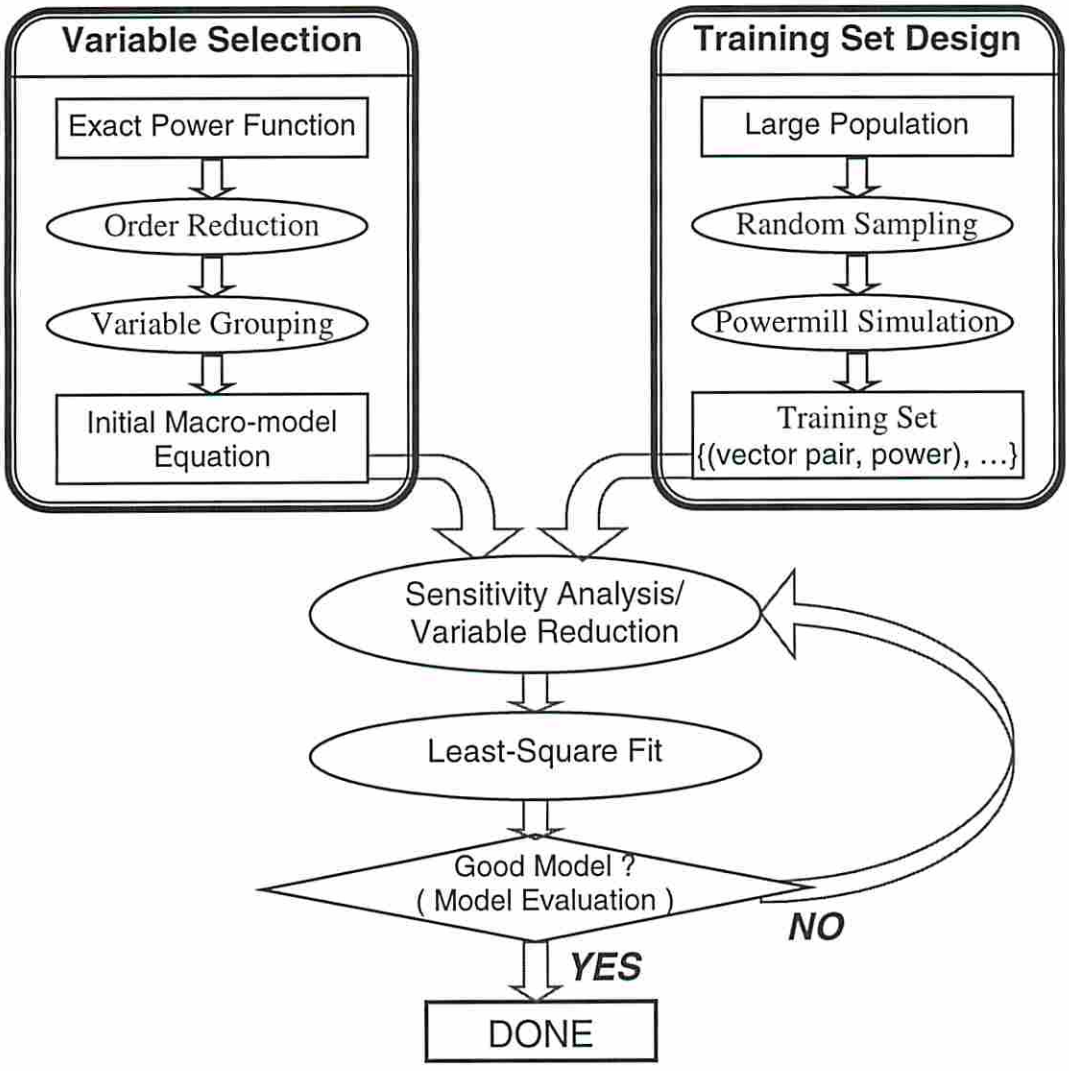


Figure 1 The workflow of generating a cycle-accurate macro-model

In the variable selection step, we start with an exact function that relates the cycle power and the input vector pair. The function terms are organized according to the order of spatial correlations between bits of the input vector pair. During order reduction, high order terms are dropped based on a performance-cost trade-off consideration. Variable grouping is performed to collapse the variables which have similar influence on power into the same groups. At the end of this step, we obtain an initial macro-model equation which is a reduced form of the original exact function and define the procedure for extracting variable values from the input vector pair.

The training set design step is simple and straight forward compared to other parts. It starts from a very large set of vector pairs, which is in turn obtained either from real application or is synthetically generated. Random sampling is performed to obtain a much smaller sub-set of vector pairs which is representative of the original set. Finally, the module under analysis is simulated using Powermill and by applying the vector pairs in the sub-set. The vector pairs in the sub-set and their power values form the so called training set.

In the third step of the flow, a statistical variable reduction algorithm is applied on the initial macro-model equation using the training set. The goal of this algorithm is to eliminate the variables which have the least impact on the circuit power dissipation, and therefore, limit the number of variables in the final macro-model equation. Subsequently, we obtain a final macro-model equation consisting of the most power-significant variables, that is, we obtain the final macro-model equation in the form of Eqn.(2.1) with a relatively small number of variables ($k \leq 15$).

In the fourth step of the flow, the training set is used once again to form the linear regression model in Eqn.(2.2). By using Eqn.(2.5), least-squares fit is performed to calculate the regression parameters of the macro-model. Power estimation and error calculation for the training set is done using Eqn.(2.7) and Eqn.(2.8), respectively. When estimating power consumption for any arbitrary vector pair, the estimation error is predicted using Eqn.(2.9) and Eqn.(2.10) for any given confidence level.

Model evaluation should be carried out before the macro-model is used in real applications. The standards for evaluating the quality of a macro-model were discussed in the previous section.

In our approach, to improve the fidelity of macro-model, we build different macro-models for different ranges of variable values. We call this procedure the population stratification approach (cf. Section 3.2) and the resulting macro-model the piecewise linear macro-model.

In the remainder of this section, we will focus on discussing variable selection and variable reduction steps which are the key procedures for building a good macro-model. Other steps either will be discussed briefly, or have been addressed in previous paragraphs.

3.1 Variable selection

3.1.1 The exact functional relation between the cycle power and the input vector pair

If we ignore the power consumption of floating nodes within gates (it is less than 5% in practice), the power consumption of a combinational module is only a function of transitions at the primary inputs. We can thus write:

$$P = f(\bar{t}_1, \bar{t}_2, \cdots, \bar{t}_k) \tag{3.1}$$

where k is the number of primary inputs (notice that this "k" is different from the "k" that was used in Section II) and $\bar{t}_1, \bar{t}_2, \cdots, \bar{t}_k$ are the so called transition variables which are encoded by a bit vector as follows:

$$\begin{aligned}
&\bar{t}_i = [a \quad b \quad c], \qquad i = 1,2,\cdots,k \\
&a=0, b=0, c=0 \qquad \text{if input } i: 0 \rightarrow 0 \\
&a=1, b=0, c=0 \qquad \text{if input } i: 0 \rightarrow 1 \\
&a=0, b=1, c=0 \qquad \text{if input } i: 1 \rightarrow 0 \\
&a=0, b=0, c=1 \qquad \text{if input } i: 1 \rightarrow 1
\end{aligned} \tag{3.2}$$

We use 3-bit encoding scheme instead of 2-bit encoding (which is the minimum length encoding) because 3-bit encoding is more effective in expressing the exact power function.

Define the $\otimes$ operation and + operation (normal addition) between two vectors as follows:

$$\begin{aligned}
&[u_1, u_2, \cdots, u_m] \otimes [v_1, v_2, \cdots, v_l] = [u_1 v_1, u_1 v_2, \cdots, u_1 v_l, u_2 v_1, u_2 v_2, \cdots, u_2 v_l, \cdots, u_m v_1, u_m v_2, \cdots, u_m v_l] \\
&[u_1, u_2, \cdots, u_m] + [v_1, v_2, \cdots, v_m] = [u_1 + v_1, u_2 + v_2, \cdots, u_m + v_m]
\end{aligned} \tag{3.3}$$

We give the exact functional relation between the cycle power and the input vector pair (the transition variables) as:

$$P = a_0 + \sum_{i=1}^{k} \bar{t}_i \cdot \begin{bmatrix} a_i^{0\to1} \\ a_i^{1\to0} \\ a_i^{1\to1} \end{bmatrix} + \sum_{i=1}^{k} \sum_{j=i+1}^{k} \bar{t}_i \otimes \bar{t}_j \cdot \begin{bmatrix} a_{i,j}^{0\to1,0\to1} \\ a_{i,j}^{0\to1,1\to0} \\ \vdots \\ a_{i,j}^{1\to1,1\to1} \end{bmatrix}$$

$$+ \cdots + \bar{t}_1 \otimes \bar{t}_2 \otimes \cdots \otimes \bar{t}_k \cdot \begin{bmatrix} a_{1,2,\ldots,k}^{0\to1,0\to1,\cdots,0\to1,0\to1} \\ a_{1,2,\ldots,k}^{0\to1,0\to1,\cdots,0\to1,1\to0} \\ \vdots \\ a_{1,2,\ldots k}^{1\to1,1\to1,\cdots,1\to1,1\to1} \end{bmatrix} \tag{3.4}$$

$$= a_0 + \sum_{i=1}^{k} \bar{t}_i \cdot \bar{a}_i^{\mathrm{T}} + \sum_{i=1}^{k} \sum_{j=i+1}^{k} \bar{t}_i \otimes \bar{t}_j \cdot \bar{a}_{i,j}^{\mathrm{T}} + \cdots + \bar{t}_1 \otimes \bar{t}_2 \otimes \cdots \otimes \bar{t}_n \cdot \bar{a}_{1,2,\ldots,k}^{\mathrm{T}}$$

where $\bar{t}_i$ is called *order* 1 *transition variable* of input *i*, $\bar{t}_i \otimes \bar{t}_j$ is called *order* 2 *joint transition variable* of inputs *i* and *j*, etc. $a_0, \bar{a}_i^{\mathrm{T}}, \cdots, \bar{a}_{1,2,\cdots,k}^{\mathrm{T}}$ are constant real numbers. Entries of these vector variables are either 0 or 1 and the sum of entries in each vector add up to 1. Notice that each vector variable includes multiple scalar variables, when we are talking about the "number of variables" in the function, we refer to the number of scalar variables.

Example:

Given an input vector pair (101→011) with bit 1 to bit 3 listed from left to right, $\bar{t}_1 = \{010\}$, $\bar{t}_2 = \{100\}$, $\bar{t}_3 = \{001\}$, $\bar{t}_1 \otimes \bar{t}_2 = \{000100000\}$, $\bar{t}_1 \otimes \bar{t}_3 = \{000001000\}$, $\bar{t}_2 \otimes \bar{t}_3 = \{001000000\}$, $\bar{t}_1 \otimes \bar{t}_2 \otimes \bar{t}_3 = \{000000000001000000000000000\}$.

The power consumption calculated by Eqn.(3.4) is:

$$P = a_0 + a_1^{1\to0} + a_2^{0\to1} + a_3^{1\to1} + a_{1,2}^{1\to0,0\to1} + a_{1,3}^{1\to0,1\to1} + a_{2,3}^{0\to1,1\to1} + a_{1,2,3}^{1\to0,0\to1,1\to1}.$$

Theorem 1 *Equation (3.4) gives the exact power consumption for any vector pair applied to the inputs of any combinational module with k inputs. Furthermore, coefficients in the equation are unique for a given module.*

Proof: We first do some analysis on (3.4). The number of coefficients in Eqn. (3.4) is:

$$1 + 3 \times C_k^1 + 3^2 C_k^2 + \ldots + 3^k C_k^k = (1+3)^k = 4^k$$

A *k*-input circuit has 4^k different input transition combinations and corresponding power consumption values. Inserting these values to (3.4) we obtain a linear system of equations (A.1).

$$\mathbf{X}_{4^k \times 4^k} \cdot \mathbf{A}_{4^k \times 1} = \mathbf{P}_{4^k \times 1} \tag{A.1}$$

where $\mathbf{X}_{4^k \times 4^k}$ is the input switching matrix. Each row represents a different input vector pair which is denoted by the set of variables $\bar{t}$ in Eqn. (3.4), i.e., $\mathbf{X}_{1\times4^k} = [\bar{t}_1, \bar{t}_2, \cdots, \bar{t}_n, \bar{t}_{1,2}, \bar{t}_{1,3}, \cdots, \bar{t}_{n-1,n}, \cdots, \bar{t}_{1,2,\cdots,n}]$. $\mathbf{A}_{4^k\times1}$ is the coefficient matrix which consists of the coefficients $\bar{a}$ in Eqn. (3.4), i.e., $\mathbf{A}_{4^k\times1} = [\bar{a}_1{}^T, \bar{a}_2{}^T, \cdots, \bar{a}_n{}^T, \bar{a}_{1,2}{}^T, \bar{a}_{1,3}{}^T, \cdots, \bar{a}_{n-1,n}{}^T, \cdots, \bar{a}_{1,2,\cdots,n}{}^T]^T$. To prove the theorem we only need to prove that the solution of Eqn. (A.1) exists and is unique.

Lemma 1 For any sub-matrix **X'** of **X** which consists of a subset of row vectors of **X**, there is at least one column, which has only one nonzero entry.

Proof: Each column of matrix **X** can be denoted by $\mathbf{X}_{\mathbf{I}}^{x}, \mathbf{I} \in$ the power set of inputs, *x* is the transition type of **I**. Each row of matrix **X** represents a different input vector pair. The entry of

column $\mathbf{X}_{\mathbf{I}}^{x}$ of row j is nonzero if and only if the input set **I** of the vector pair j has transition x. Because each row represents a different input transition, for any set of input vector pairs (that is, for any x), there must be a set of switching input **I** and its transition type x, which is different from others. The column corresponding to this combination has a single nonzero entry.

We mark the row vectors of **X** to be $r_1, r_2, \cdots, r_{4^k}$. To prove that they are linearly independent, let us give a contradict assumption: they are linear dependent. We can write:

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + \cdots + c_{4^k}\mathbf{r}_{4^k} = \mathbf{0}_{4^k\times 1} \tag{A.2}$$

Assume that $\{c_{n_1}, c_{n_2}, \cdots, c_{n_l}\}$ is the subset of $\{c_i\}$, which are not 0, Eqn.(A.2) can be rewritten as:

$$c_{n_1}\mathbf{r}_{n_1} + c_{n_2}\mathbf{r}_{n_2} + \cdots + c_{n_l}\mathbf{r}_{n_l} = \mathbf{0}_{4^k\times 1} \tag{A.3}$$

According to Lemma1, for any sub-matrix, which consist of row vectors of **X**, there is at least one column which has only one nonzero entry, so we can not find a set of $c_i \neq 0$ which causes Eqn.(A.3) to hold. This leads to a contradiction to our assumption. Therefore, we conclude that all the row vectors are linearly independent and the rank of matrix **X** is 4^k.

Let $\mathbf{X}^{+}_{(4^k)\times(4^k+1)} = \left[\mathbf{X}_{4^k\times 4^k}\mathbf{P}_{4^k\times 1}\right]$. Because of its dimension, we know that

$$\text{Rank of } \mathbf{X}^{+} \le 4^k \tag{A.4}$$

Because **X** is a sub-matrix of $\mathbf{X}^{+}$,

$$\text{Rank of } \mathbf{X}^{+} \ge \text{Rank of } \mathbf{X}. \tag{A.5}$$

From (A.4), (A.5) we get: Rank of $\mathbf{X}^{+}$ = Rank of **X** = The number of the variables.

Therefore the equation system in Eqn. (A.1) is consistent and has a unique solution. The theorem is proved.

■

3.1.2 Transitive fanout correlation between primary inputs

It is obvious that $a_0 = 0$ since power consumption for vector pair (00…0)→(00…0) must be zero. All other coefficients in Eqn.(3.4) can be uniquely determined from circuit-level simulation on some specific vector pairs.

For example, to compute coefficient $a_1^{0\to 1}$, we simulate the module using the vector pair {(0,0…0), (1,0…0)} and obtain the power consumption value of $P_1^{0\to 1}$. From equation (3.4) we know:

$$P_1^{0\to 1} = f(t_1^{0\to 1}, t_2^{0\to 0}, \cdots, t_k^{0\to 0}) = a_1^{0\to 1}$$

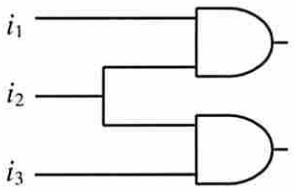
To compute $a_{1,2}^{0\to 1,0\to 1}$ we simulate the module using the vector pair {(0,0…0), (1,1,0…0)} and obtain the power consumption value of $P_{1,2}^{0\to 1,0\to 1}$. Again from Eqn.(3.4) we know:

$$P_{1,2}^{0\to 1,0\to 1} = f(t_1^{0\to 1}, t_2^{0\to 1}, t_3^{0\to 0}, \cdots, t_k^{0\to 0}) = a_1^{0\to 1} + a_2^{0\to 1} + a_{1,2}^{0\to 1,0\to 1}$$
$$\Rightarrow a_{1,2}^{0\to 1,0\to 1} = P_{1,2}^{0\to 1,0\to 1} - P_1^{0\to 1} - P_2^{0\to 1}$$

and so on.

Definition Inputs $i_1, i_2, \ldots, i_j$ are *transitive fanout correlated* if and only if their transitive fanout cones in the circuit have at least one common node, that is, there exists at least one node (internal node or output) of the module whose logic function includes all inputs $i_1, i_2, \ldots, i_j$. j is called the *order* of the correlation.

For the sake of simplicity, we use "correlation" to stand for "transitive fanout correlation" in the remainder of this paper.



(a)

order 1 corr.	order 2 corr.	order 3 corr.
i_1; i_2; i_3	(i_1, i_2); (i_2, i_3)	None

(b)

Figure 2 Example of Transitive Fanout Correlation

Figure 2(a) shows a simple 3-input 2-gate circuit. Since it has only 3 inputs, the highest possible order of correlation between inputs is 3. The table in (b) shows the correlated inputs for different orders. Notice that the input pair (i_1, i_3) is not on the list of order 2 correlations and triplet (i_1,i_2,i_3) is not in the list of order 3 correlations because the corresponding inputs have no common nodes among their transitive fanout cones.

The coefficients in Eqn.(3.4) essentially reflect the correlation between the corresponding (joint) transition probabilities and the power consumption in a circuit.

Proposition 1 *If $i_1,i_2,\ldots,i_j$ are not correlated, all entries of $\bar{a}_{i_1,i_2,\cdots,i_j}$ are zero.*

Proof : Consider the power-transition function for each node in the circuit. These power functions are written in the form of Eqn.(3.4). The variables are the appropriate circuit input transitions for the nodes. Since inputs i_1, $i_2 \ldots i_j$ are not (transitive fanout) correlated, there is no node whose logic transition includes all of these inputs. So the entries of $\bar{a}_{i_1,i_2,\cdots,i_j}$ are zero in the power functions of these nodes. The circuit power function is the sum of the power functions for each node in the circuit. Therefore, the entries of $\bar{a}_{i_1,i_2,\cdots,i_j}$ are also zero in the circuit power function.

∎

Corollary *If J is the highest order of correlation among inputs of a module, the first J+1 terms of Eqn.(3.4) are sufficient to model the exact power for any input vector pair applied to the module.*

Proof : All coefficients are zero for other terms.

∎

3.1.3 Function order reduction

Eqn.(3.4) gives an exact representation of the relation between power and input transition. However, this form is too complicated for practical use. In this section, we will discuss the first step to simplify the macro-model function, which is order reduction.

From Eqn.(3.4), we know that the complexity of the macro-model increases exponentially with the order of the input correlation we which consider. Evidently, ignoring the high order term leads to some estimation error. The first question, is how much error will be introduced if we drop certain high order terms. The second question is what the cost will be if we keep more terms in the original function. Table 1 shows some examples of the percentage error caused by ignoring the high order input correlations. Column 1 gives the circuit name. Circuit A is a 4-bit multiplier, B is a 4-bit ripple carry adder without carry in, and C is an 8-input random logic circuit. In the experiment, all the coefficients of the exact power function in Eqn.(3.4) are calculated in the way discussed in section 3.1.2. Then we do power calculation on the population of all possible (4^k, k is

the number of inputs) input vector pairs to the module using the reduced-order functions of Eqn(3.4), i.e., ignoring the high order terms (by assuming that the coefficients of high order transition variables to be all zero). The power values calculated by the reduced-order functions are compared with the actual power values. The average relative errors are then reported in Table 1 from the 1st data column to the 8th data column. The integer number i on top of each column indicates the maximum order to which the function terms are kept. For example, in the 4th data column, number "4" means that, the reduced-order function is the same as Eqn.(3.4) except that transition variables (and their coefficients) with order higher that 4 are ignored. The last row of the table shows the total number of variables in the reduced-order functions for each of the circuits.

Table 1 Average percentage error on power calculation by reduced-order functions

Circuit	1	2	3	4	5	6	7	8
A	99.8%	42.6%	31.7%	23.9%	13.7%	4.1%	0.8%	0.0%
B	40.3%	9.2%	11.9%	9.6%	8.4%	4.3%	1.3%	0.0%
C	37.6%	13.1%	8.9%	7.7%	5.1%	4.3%	1.0%	0.0%
	24	276	1788	7458	21066	41478	58974	65535

Table 1 shows that keeping higher order terms tends to, although not monotonically, to provide more accurate power estimation results. One significant improvement shown in the table is from the 1st data column (keeping only order 1 terms) to the 3rd data column (keeping order 1, 2 and 3 terms). From this point on, the complexity of the reduced-order function increases much faster than the percentage error decreases.

We therefore approximate Eqn.(3.4) by ignoring terms with order higher than 3. The reduced-order function is written as:

$$P = a_0 + \sum_{i=1}^{k} \bar{t}_i \cdot \begin{bmatrix} a_i^{0\to1} \\ a_i^{1\to0} \\ a_i^{1\to1} \end{bmatrix} + \sum_{i=1}^{k}\sum_{j=i+1}^{k} \bar{t}_i \otimes \bar{t}_j \cdot \begin{bmatrix} a_{i,j}^{0\to1,0\to1} \\ a_{i,j}^{0\to1,1\to0} \\ \vdots \\ a_{i,j}^{1\to1,1\to1} \end{bmatrix} + \sum_{i=1}^{k}\sum_{j=i+1}^{k}\sum_{l=j+1}^{k} \bar{t}_i \otimes \bar{t}_j \otimes \bar{t}_l \cdot \begin{bmatrix} a_{i,j,l}^{0\to1,0\to1,0\to1} \\ a_{i,j,l}^{0\to1,0\to1,1\to0} \\ \vdots \\ a_{i,j,l}^{1\to1,1\to1,1\to1} \end{bmatrix} + \varepsilon \tag{3.5}$$

$$= a_0 + \sum_{i=1}^{k} \bar{t}_i \cdot \bar{a}_i^{\mathrm{T}} + \sum_{i=1}^{k}\sum_{j=i+1}^{k} \bar{t}_i \otimes \bar{t}_j \cdot \bar{a}_{i,j}^{\mathrm{T}} + \sum_{i=1}^{k}\sum_{j=i+1}^{k}\sum_{l=j+1}^{k} \bar{t}_i \otimes \bar{t}_j \otimes \bar{t}_l \cdot \bar{a}_{i,j,l}^{\mathrm{T}} + \varepsilon$$

where ε is the error caused by approximation.

We can minimize error ε by re-computing the coefficient values by doing least-squares fit for Eqn.(3.5). However Eqn.(3.5) is too complicated to be our macro-model equation since the number of variables in it is $3 \cdot k + 9 \cdot C_k^2 + 27 \cdot C_k^3$, which is too high! Furthermore, the use of 0-1 variables in (3.5) makes it very difficult to significantly reduce the number of variables using a regression approach.

3.1.4 Variable grouping

To further reduce the function complexity in Eqn.(3.5), we use a variable grouping approach as will be described next. This approach offers two advantages: 1) uses integer variables which are

easier to work with and offer more flexibility compared to 0-1 variables, 2) has a constant number of variables which is independent of number of primary inputs, k.

We define G_1 as the set of all inputs, G_2 as the set of all possible combinations of two inputs, G_3 as the set of all possible combinations of three inputs:

$$G_1 = \{1,2,\cdots,k\},$$
$$G_2 = \{(1,2),(1,3),\cdots,(1,k),(2,3),\cdots,(k-1,k)\},$$
$$G_3 = \{(1,2,3),(1,2,4),\cdots,(1,2,k),(1,3,4),\cdots,(k-2,k-1,k)\}$$

Note that G_i consists of indices for order i transition variables. The variable grouping technique forms N_1 subsets of G_1, N_2 subsets of G_2, and N_3 subsets of G_3 such that:

$$\bigcap_{g=1}^{N_1} G_{1,g} = \phi,\ \bigcup_{g=1}^{N_1} G_{1,g} \subseteq G_1,\ |G_{1,g}| \le K_1 \qquad \bigcap_{g=1}^{N_2} G_{2,g} = \phi,\ \bigcup_{g=1}^{N_2} G_{2,g} \subseteq G_2,\ |G_{2,g}| \le K_2 \qquad \bigcap_{g=1}^{N_3} G_{3,g} = \phi,\ \bigcup_{g=1}^{N_3} G_{3,g} \subseteq G_3,\ |G_{3,g}| \le K_3 \tag{3.6}$$

where K_1, K_2, K_3 bounds are given. The size constraints are specified to manage the complexity of macro-model equation characterization and evaluation.

We approximate equation (3.5) by assuming that:

$$\begin{bmatrix} a_i^{0\to1} \\ a_i^{1\to0} \\ a_i^{1\to1} \end{bmatrix}_{3\times1} \equiv \begin{bmatrix} b_{1,g}^{0\to1} \\ b_{1,g}^{1\to0} \\ b_{1,g}^{1\to1} \end{bmatrix}_{3\times1} \qquad \forall i \in G_{1,g}, \qquad g = 1,2,\cdots,N_1$$

$$\begin{bmatrix} a_{i,j}^{0\to1,0\to1} \\ a_{i,j}^{0\to1,1\to0} \\ \vdots \\ a_{i,j}^{1\to1,1\to1} \end{bmatrix}_{9\times1} \equiv \begin{bmatrix} b_{2,g}^{0\to1,0\to1} \\ b_{2,g}^{0\to1,1\to0} \\ \vdots \\ b_{2,g}^{0\to1,1\to1} \end{bmatrix}_{9\times1} \qquad \forall (i,j) \in G_{2,g}, \qquad g = 1,2,\cdots,N_2$$

$$\begin{bmatrix} a_{i,j,l}^{0\to1,0\to1,0\to1} \\ a_{i,j,l}^{0\to1,0\to1,1\to0} \\ \vdots \\ a_{i,j,l}^{1\to1,1\to1,1\to1} \end{bmatrix}_{27\times1} \equiv \begin{bmatrix} b_{3,g}^{0\to1,0\to1,0\to1} \\ b_{3,g}^{0\to1,0\to1,1\to0} \\ \vdots \\ b_{3,g}^{0\to1,1\to1,1\to1} \end{bmatrix}_{27\times1} \qquad \forall (i,j,l) \in G_{3,g}, \qquad g = 1,2,\cdots,N_3$$

where $b_{\{1,2,3\},g}^{\cdots}$ are constant real numbers. To minimize the error introduced by the above approximation, we should do a careful variable grouping. We firstly calculate the coefficients for terms of orders 1,2 and 3 in Eqn.(3.4) by using the method discussed in Section 3.1.2. In our approach, the grouping criteria are based on the following quantities:

$c_i = \frac{1}{3}(a_i^{0\to1} + a_i^{1\to0} + a_i^{1\to1}),\quad i = 1,2,\cdots,k$ for grouping single inputs,

$c_{i,j} = \frac{1}{9}(a_{i,j}^{0\to1,0\to1} + a_{i,j}^{0\to1,1\to0} + \cdots + a_{i,j}^{1\to1,1\to1}),\quad i,j = 1,2,\cdots,k,\ i < j$ for grouping pairs of inputs,

$c_{i,j,l} = \frac{1}{27}(a_{i,j,l}^{0\to1,0\to1,0\to1} + a_{i,j,l}^{0\to1,0\to1,1\to0} + \cdots + a_{i,j,l}^{1\to1,1\to1,1\to1}),\ i,j,l = 1,2,\cdots,k,\ i < j < l$ for grouping triplets of inputs.

Transition variables of order i are sorted in increasing order of the corresponding c-values for order i. The domain of the c-values is divided into several sub-domains such that the number of transition variables with c-values in different sub-domains is approximately equal, but is less than the corresponding K_i values. The indices of transition variables with c-values in different sub-

domains define the groups. The first N_1 groups with largest absolute c-values are adopted as $G_{1,g}$ (g=1,2,…,N_1) as defined in Eqn.(3.6). Other groups with smaller c-values are abandoned. Similarly, the first N_2 or N_3 groups with largest absolute c-values are adopted as $G_{2,g}$ or $G_{3,g}$ (g=1,2,…,N_2 or N_3).

Example Let $N_1 = 3$, $N_2 = 4$, $N_3 = 2$. Assume we want to do variable grouping for a macro-model equation corresponding to a 6-input circuit. The group size constraints are set as: $K_1 = K_2 = K_3 = 3$. Firstly we calculate the value of c_i (i = 1,2,…,6). Assume, the c-values are given as:

$c_1 = 0,\ c_2 = 0.5,\ c_3 = 1.9,\ c_4 = 2.0,\ c_5 = 0,\ c_6 = 0$

We divide the domain of c-values, [0, 2.0], into 3 sub-domains: [0, 0.5), [0.5, 1.5), and [1.5, 2.0]. The grouping for single inputs is: $G_{1,1} = \{1,5,6\}, G_{1,2} = \{2\}, G_{1,3} = \{3,4\}$. We keep $G_{1,1}$ through $G_{1,3}$.

Then we compute the value of $c_{i,j}$ (i,j = 1,2,…,k, $i < j$). Again, suppose:

$c_{1,2} = 1.0,\ c_{1,3} = 0.3,\ c_{1,4} = -0.2,\ c_{1,5} = -0.4,\ c_{1,6} = 1.6,\ c_{2,3} = 1.6,\ c_{2,4} = 0.8,\ c_{2,5} = 1.4,$

$c_{2,6} = 0.7,\ c_{3,4} = -1.1,\ c_{3,5} = 0.2,\ c_{3,6} = 0.2,\ c_{4,5} = 1.6,\ c_{4,6} = -0.7,\ c_{5,6} = 0.9$

The division of c-values is: [-1.1, -0.5), [-0.5, 0), [0, 0.5), [0.5, 1.0), [1.0, 1.5), and [1.5, 1.6]
The grouping for input pairs is: $G_{2,1} = \{(1,6),(2,3),(4,5)\}$, $G_{2,2} = \{(2,5),(1,2)\}$

$G_{2,3} = \{(2,4),(2,6),(5,6)\}$, $G_{2,4} = \{(1,3),(3,5),(3,6)\}$, $G_{2,5} = \{(1,4),(1,5)\}$, $G_{2,6} = \{(3,4),(4,6)\}$

We only keep $G_{2,1}$, $G_{2,2}$, $G_{2,3}$, and $G_{2,6}$,

The case for grouping variables of order 3 is solved similarly.

■

Let's introduce some notation:

$T_{1,g}^{i\to j}$: the total number of transitions of type $i\to j$ in group $G_{1,g}$

$T_{2,g}^{i\to j,k\to l}$: the total number of pair-wise joint transitions of type ($i\to j$ $k\to l$) in group $G_{2,g}$

$T_{3,g}^{i\to j,k\to l,m\to n}$: the total number of joint transitions of type ($i\to j$, $k\to l$, $m\to n$) in group $G_{3,g}$

$$\bar{T}_{1,g} = \sum_{i\in G_{1,g}} \bar{t}_i = \left[\ T_{1,g}^{0\to1} \quad T_{1,g}^{1\to0} \quad T_{1,g}^{1\to1}\ \right]_{1\times3}$$

$$\bar{T}_{2,g} = \sum_{(i,j)\in G_{2,g}} \bar{t}_i \otimes \bar{t}_j = \left[\ T_{2,g}^{0\to1,0\to1} \quad T_{2,g}^{0\to1,1\to0} \quad \cdots \quad T_{2,g}^{1\to1,1\to1}\ \right]_{1\times9}$$

$$\bar{T}_{3,g} = \sum_{(i,j,l)\in G_{3,g}} \bar{t}_i \otimes \bar{t}_j \otimes \bar{t}_l = \left[\ T_{3,g}^{0\to1,0\to1,0\to1} \quad T_{3,g}^{0\to1,0\to1,1\to0} \quad \cdots \quad T_{3,g}^{1\to1,1\to1,1\to1}\ \right]_{1\times27}$$

We can thus write our initial cycle-accurate macro-model as follows:

$$P = b_0 + \sum_{g=1}^{N_1} \begin{bmatrix} T_{1,g}^{0\to1} & T_{1,g}^{1\to0} & T_{1,g}^{1\to1} \end{bmatrix} \cdot \begin{bmatrix} b_{1,g}^{0\to1} \\ b_{1,g}^{1\to0} \\ b_{1,g}^{1\to1} \end{bmatrix} + \sum_{g=1}^{N_2} \begin{bmatrix} T_{2,g}^{0\to1,0\to1} & T_{2,g}^{0\to1,1\to0} & \cdots & T_{2,g}^{1\to1,1\to1} \end{bmatrix} \cdot \begin{bmatrix} b_{2,g}^{0\to1,0\to1} \\ b_{2,g}^{0\to0,1\to0} \\ \vdots \\ b_{2,g}^{1\to1,1\to1} \end{bmatrix}$$
$$+ \sum_{g=1}^{N_3} \begin{bmatrix} T_{3,g}^{0\to1,0\to1,0\to1} & T_{3,g}^{0\to1,0\to1,1\to0} & \cdots & T_{3,g}^{1\to1,1\to1,1\to1} \end{bmatrix} \cdot \begin{bmatrix} b_{3,g}^{0\to1,0\to1,0\to1} \\ b_{3,g}^{0\to0,0\to1,1\to0} \\ \vdots \\ b_{3,g}^{1\to1,1\to1,1\to1} \end{bmatrix} \quad (3.7)$$

In terms of N_1, N_2, N_3, values, the number of variables in the macro-model is $3N_1+9N_2+27N_3$, which is independent of the number of circuit inputs k.

Table 2 shows the experimental results for three macro-models using different number of groups and using different grouping strategies. For Macro-model 1, $N_1 = 1$, $N_2 = 1$, $N_3 = 1$; For Macro-model 2, $N_1 = 8$, $N_2 = 8$, $N_3 = 2$, and the single inputs, input pairs, and input triplets are grouped randomly; For Macro-model 3, $N_1 = 8$, $N_2 = 8$, $N_3 = 2$, and our variable grouping heuristic is used. The input sequence is randomly generated.

Table 2 Experimental results of variable grouping

	Macro-model 1		Macro-model 2		Macro-model 3	
Module	r	ECP (%)	r	ECP (%)	r	ECP (%)
C1355	1.98	8.07	1.57	9.19	2.54	7.76
C1908	1.41	15.36	1.42	15.04	2.76	11.16
C2670	1.18	11.66	1.19	11.64	1.63	10.38
C3540	1.42	17.47	1.76	15.48	2.37	12.19
C432	1.11	29.07	1.12	29.00	2.46	20.15
C5315	1.21	9.87	1.30	9.4	2.79	8.1
C6288	2.15	8.10	2.47	7.6	2.79	6.82
C7552	1.04	33.00	1.11	30.94	6.39	9.24
C880	1.42	19.82	1.31	20.78	1.95	16.34
Mul16	2.34	8.90	2.57	8.32	2.96	7.04
Adder16	2.05	8.63	2.12	8.44	4.08	6.15

Results show that macro-models 1 and 2 have similar correlation factors and ECP errors, while the quality of macro-model 3 is clearly better than the other two.

From Table 2, we can draw the following conclusions:

- Using more groups in variable grouping improves the quality of macro-models.
- A good variable grouping technique is very important to obtain a high quality macro-model.

3.3 Population stratification

From our experiments we found that the regression factor r between the estimated power and the actual power is different for different ranges of power dissipation. This means that the regression model is not strictly linear over the range of all possible power values. This phenomenon occurs in many practical situations. One reason for the lack of linearity is that the macro-model equation is only an approximation to the power-transition function. During the variable selection, we discard the high order terms in the power-transition function and group subsets of variables of given order together. The approximation introduces some non-linearity into the macro-model equation. This effect is more pronounced when the number of variables is small.

To improve the quality of our macro-model, we refine the macro-model to a *piece-wise linear regression model*. At the first step, we stratify the training set into several disjoint subsets (strata) based on the switching activity of the vector pairs in the training set. A vector pair will fall into exact one of these strata. Then the macro-model is trained separately for each subset of the training set. When we apply this piece-wise linear macro-model to estimate the power for a given vector pair, we first examine the switching activity range of the vector pair, and then invoke the macro-model equation which was trained using vector pairs with a similar switching activity.

Theorem 2 The correlation measure r_{str} of the macro-model obtained by the population stratification is no worse than that without population stratification r_{nostr}, i.e.,

$$r_{str} \geq r_{nostr}$$

Proof : Since r is a monotonic-increasing function of R, we can prove the theorem by proving:

$$R_{str} \geq R_{nostr}$$

We first introduce some properties of least squares [8]:

1. The sum of the observed values Y equals the sum of the fitted values $\hat{\mathbf{Y}}$:

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i \tag{B.1}$$

2. The sum of the weighted residues is zero if the residue in the ith trial is weighted by the fitted value of the corresponding variable in the ith trial:

$$\hat{\mathbf{Y}}^{\mathrm{T}} \cdot \mathbf{e} = 0, \text{ where } \hat{\mathbf{Y}} = \left[\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_n\right]^{\mathrm{T}}, \quad \mathbf{e} = [e_1, e_2, \cdots, e_n]^{\mathrm{T}} \tag{B.2}$$

Since $\mathbf{e} = \hat{\mathbf{Y}} - \mathbf{Y}$, Eqn.(B.2) can be written as $\hat{\mathbf{Y}}^{\mathrm{T}} \cdot \hat{\mathbf{Y}} = \hat{\mathbf{Y}}^{\mathrm{T}} \cdot \mathbf{Y}$ (B.3)

As for the proof, assume that we have a macro-model with n variables. From [8] we know that Without population stratification, the SSE and SSR of the macro-model are given by [8]:

$$SSE_{nostr} = \mathbf{e}^{\mathrm{T}} \cdot \mathbf{e} = \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{Y} - \hat{\mathbf{Y}}^{\mathrm{T}} \cdot \mathbf{Y}$$

$$SSR_{nostr} = \hat{\mathbf{Y}}^{\mathrm{T}} \cdot \mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}^{\mathrm{T}}\mathbf{J}\mathbf{Y}$$, where J is a $n \times n$ matrix of 1s.

$$SSE_{nostr} + SSR_{nostr} = \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}^{\mathrm{T}}\mathbf{J}\mathbf{Y} \tag{B.4}$$

After stratification, we have k group of observed values Y_1, Y_2,..Y_k and k group of fitted values $\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, .., \hat{\mathbf{Y}}_k$. Each group has n_i trials such that $\sum_{i=1}^{k} n_i = n$.

$$SSE_{str} = \sum_{i=1}^{k} SSE_i = \sum_{i=1}^{k} (\mathbf{Y}_i^{\mathrm{T}} \cdot \mathbf{Y}_i - \hat{\mathbf{Y}}_i^{\mathrm{T}} \cdot \mathbf{Y}_i)$$

Notice that $\sum_{i=1}^{k} \mathbf{Y}_i^{\mathrm{T}} \cdot \mathbf{Y}_i = \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{Y}$

$$SSE_{str} = \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{Y} - \sum_{i=1}^{k} \hat{\mathbf{Y}}_i^{\mathrm{T}} \cdot \mathbf{Y}_i \tag{B.5}$$

$$SSR_{str} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\hat{Y}_i - \overline{Y})^2 \tag{B.6}$$

Let $\mathbf{L} = [1,1,\cdots,1]_{1\times n}^{\mathrm{T}}, \mathbf{L}_i = [1,1,\cdots,1]_{1\times n_i}^{\mathrm{T}}$, (B.5) can be written as:

$$SSR_{str} = \sum_{i=1}^{k} [\hat{\mathbf{Y}}_i - \frac{1}{n}\mathbf{L}_i \cdot \mathbf{L}^{\mathrm{T}}\mathbf{Y}]^{\mathrm{T}} [\hat{\mathbf{Y}}_i - \frac{1}{n}\mathbf{L}_i \cdot \mathbf{L}^{\mathrm{T}}\mathbf{Y}]$$

$$= \sum_{i=1}^{k} (\hat{\mathbf{Y}}_i^{\mathrm{T}}\hat{\mathbf{Y}}_i - \frac{1}{n}\hat{\mathbf{Y}}_i^{\mathrm{T}}\mathbf{L}_i\mathbf{L}^{\mathrm{T}}\mathbf{Y} - \frac{1}{n}\mathbf{Y}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}_i\hat{\mathbf{Y}}_i + \frac{1}{n^2}\mathbf{Y}^{\mathrm{T}}\mathbf{L}\mathbf{L}_i^{\mathrm{T}}\mathbf{L}_i\mathbf{L}^{\mathrm{T}}\mathbf{Y})$$

Because of (B.1), SSR_{str} can be simplified as:

$$SSR_{str} = \sum_{i=1}^{k} \hat{\mathbf{Y}}_i^{\mathrm{T}}\hat{\mathbf{Y}}_i - \frac{\sum_{i=1}^{k} n_i}{n^2}\mathbf{Y}^{\mathrm{T}}\mathbf{J}\mathbf{Y} \tag{B.7}$$

Using (B.3), we get:

$$SSR_{str} = \sum_{i=1}^{k} \hat{\mathbf{Y}}_i^{\mathrm{T}}\mathbf{Y}_i - \frac{1}{n}\mathbf{Y}^{\mathrm{T}}\mathbf{J}\mathbf{Y} \tag{B.8}$$

So $$SSE_{str} + SSR_{str} = \mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \frac{1}{n}\mathbf{Y}^{\mathrm{T}}\mathbf{J}\mathbf{Y} = SSE_{nostr} + SSR_{nostr} \tag{B.9}$$

Next we will prove that $SSE_{str} \leq SSE_{nostr}$:

Assume that the population is stratified into k disjoint strata. The form of the macro-model and the variable selection is the same in each stratum. Let $\mathbf{O}_{\mathrm{nostr}}$ denote the un-stratified population which includes both the vector pairs and the corresponding power consumption, also let $\mathbf{O}_1$, …, $\mathbf{O}_k$ denote the sub populations in the k strata.

$$\mathbf{O}_{nostr} = \mathbf{O}_1 \cup \mathbf{O}_2 \cup \ldots \cup \mathbf{O}_k$$

$$\mathbf{O}_i \cup \mathbf{O}_j = \Phi,\ i \neq j$$

Let the variable matrix $\mathbf{X}_{nostr}$ and $\mathbf{P}_{nostr}$ denote the variables drawn from $\mathbf{O}_{nostr}$ and $\mathbf{X}_1$, …, $\mathbf{X}_k$, $\mathbf{P}_1$, …, $\mathbf{P}_k$ denote the variable matrixes and power values drawn from $\mathbf{O}_1$, …, $\mathbf{O}_k$. Because each stratum is disjoint:

$$\mathbf{X}_{nostr} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_k; \tag{B.10}$$

$$\mathbf{P}_{nostr} = \mathbf{P}_1 + \mathbf{P}_2 + \ldots + \mathbf{P}_k. \tag{B.11}$$

The macro-model without population stratification can be written as:

$$f_{nostr} = \mathbf{B}_{nostr} \cdot \mathbf{X}_{nostr}$$

where $\mathbf{B}_{nostr}$ is obtained by least squares fit using a set of training values $\mathbf{P}_{nostr}$ and $\mathbf{X}_{nostr}$. The function can also be written as:

$$f_{nostr} = \mathbf{B}_{nostr} \cdot \mathbf{X}_1 + \mathbf{B}_{nostr} \cdot \mathbf{X}_2 + \cdots + \mathbf{B}_{nostr} \cdot \mathbf{X}_k \tag{B.12}$$

The macro-model function for each stratum can be written as:

$$f_1 = \mathbf{B}_1 \cdot \mathbf{X}_1,\ f_2 = \mathbf{B}_2 \cdot \mathbf{X}_2,\ \ldots,\ f_k = \mathbf{B}_k \cdot \mathbf{X}_k$$

Because of (B.11),the sum of the power values for each stratum is the power of the whole population, we can thus obtain the macro-model function for the whole population as follows:

$$f_{str} = f_1 + f_2 + \cdots + f_k = \mathbf{B}_1 \cdot \mathbf{X}_1 + \mathbf{B}_2 \cdot \mathbf{X}_2 + \cdots + \mathbf{B}_k \cdot \mathbf{X}_k \qquad \text{(B.13)}$$

Because of the macro-model coefficients generation procedure, we know that $\mathbf{B_1}$, $\mathbf{B_2}$, …, $\mathbf{B_k}$ are least squares fit coefficients for $f_1, f_2, \ldots, f_k$. Since, however, the coefficients $\mathbf{B}_{nostr}$ are not the least squares fit coefficients with the respect to variables $\mathbf{X_1}$, $\mathbf{X_2}$, …, $\mathbf{X_k}$, $SSE_{str} \leq SSE_{nostr}$.

Because of (B.9), we know that $SSR_{str} \geq SSR_{nostr}$. Hence $R_{str} \geq R_{nostr}$

■

Experimental results in Table 3 shows the improvement on the regression factor r of the macro-model with the population stratification approach (Macro-model 1) and without it (Macro-model 2). We use Eqn.(3.7) as the macro-model equation, and the input sequence contains both biased (non-random, wider range of switching activity and power) and random vectors.

Table 3 Experimental results of population stratification approach

Module	Macro-model 1 (w/ str)		Macro-model 2 (w/o str)	
	r	ECP (%)	r	ECP (%)
C1355	26.0	7.86	16.4	8.76
C1908	12.8	9.34	10.4	11.19
C2670	23.6	8.77	20.0	10.22
C3540	19.7	11.45	11.8	12.88
C432	6.5	19.07	5.3	22.96
C5315	27.9	7.64	26.8	8.72
C6288	46.5	6.03	37.0	7.16
C7552	43.7	6.58	39.0	7.36
C880	10.4	14.19	10.4	15.32
Mul16	30.0	6.32	27.9	6.90
ADDER16	38.1	5.64	18.4	6.73

Notice that population stratification can be done not only according to the switching activity of the input vector pair, but also according to the transition behavior of some special inputs such as clock, mode control, etc. However, this is not the subject of this paper.

3.2 Variable reduction

In the initial macro-model equation (3.7), the number of variables is about 150. Although the large number of variables improves the quality of the macro-model, we would like to avoid

evaluating a large macro-model equation for every clock cycle. Therefore, we must reduce the number of variables in the equation without incurring a large error.

In our approach, the modified forward stepwise regression procedure [8] is used to reduce the number of variables. The search method develops a sequence of regression models. At each step, one X variable is added to or deleted from the final macro-model equation. The criterion used for adding or deleting variables is the F^* *statistics* of the regression theory. The algorithm is described next:

Input of the algorithm: Given are a set of candidate variables { X_1, X_2, …, X_n } which is in the initial macro-model, a training set (values of variables for input vector pair and corresponding Powermill power value), a low threshold t_0 for deleting a variable, a high threshold t_1 for adding a variable, an upper bound of number of variables $\mathrm{MAX}_{\mathrm{var}}$, S is the set of selected variables.

Step 0 (Initialization) : Set $S = \Phi$ and $C = \{ X_1, X_2, \ldots, X_n \}$

Step 1 (Find the first variable) : Fit a one-variable linear regression model for each variable X_i in C. The F^* test for each model is given by:

$$F_i^* = \frac{MSR(X_i)}{MSE(X_i)}, \quad i = 1,2,\cdots,N$$

where MSR and MSE were defined in Section II. Assume that X_j is the variable with the maximum F^* value. If $F_j^* \geq t_1$ then move X_j from C to S and denote it as X_1^*. Otherwise, no macro-model can be found for the given t_1 value (t_1 must be reduced). The algorithm terminates.

This step finds the first variable for the final macro-model. The F^* test is used to find the "most significant" variable (as far as power dissipation in the module is concerned).

Step 2 (Add a variable) : Assume $S = \{\}$, for each X_i remaining in C, fit the regression model with a+1 variables $X_1^*, X_2^*, \cdots, X_a^*$ and X_i. For each of them, the partial F test statistics is:

$$F_i^* = \frac{MSR(X_i \mid X_1^*, X_2^*, \cdots, X_a^*)}{MSE(X_i, X_1^*, X_2^*, \cdots, X_a^*)} = \left(\frac{b_i}{s\{b_i\}}\right)^2$$

where b_i is the estimated value of β_i coefficient and $s\{b_i\}$ is the standard deviation of b_i. Let X_j be the variable with the maximum F_i^* value. If $F_j^* \geq t_1$ then move X_j form C to S and denote it as X_{a+1}^*, increase a by 1, and go to Step 3; Otherwise the algorithm terminates.

This step adds one more variable into the final macro-model. The F^* test is used to find the "most significant" variable to add to the set of existing (already selected) variables.

Step 3 (delete a variable) : Assume S={ $X_1^*, X_2^*, \cdots, X_a^*$ }, and X_a^* is the latest variable added in Step 2. Compute the partial F test statistics for all other variables in S:

$$F_i^* = \frac{MSR(X_i^* \mid X_1^*, X_2^*, \cdots, X_{i-1}^*, X_{i+1}^*, \cdots, X_a^*)}{MSE(X_i^*, X_1^*, X_2^*, \cdots, X_a^*)} = \left(\frac{b_i}{s\{b_i\}}\right)^2$$

Let X_j^* be the variable with minimum F^* value. If $F_j^* < t_0$ then remove X_j^* form S.

After adding a new variable into the macro-model, the "significance" of some old variable may be reduced due to the joint effect of the newly added variable and other old variables. In such a case, we have to remove the old variable from the macro-model. The F^* test is used to find the "most insignificant" variable to delete from the set of existing (already selected) variables.

Step 4 : Repeat Steps 2 and 3 until one of following conditions is true:

1. Algorithm terminates in Step 2.
2. $C=\Phi$.
3. The number of variables in S equals to MAX_{var}.

In our approach, the number of variables in the candidate set is 162 at the beginning (since we set N_1=8, N_2=8, and N_3=2). We choose $t_0 = t_1 = 10.0$, $MAX_{var} = 15$. For most macro-models, the algorithm terminated at the 3rd condition at step 4 when the number of variables equals to MAX_{var}. Only for one of the macro-models the algorithm terminated at step 2 when $F_j^* < t_1$.

3.4 Evaluation speed of the macro-model

The time complexity of macro-model evaluation in real applications is an important measure of macro-model performance. In applications where the macro-model has to be evaluated every clock cycle, the speed of the evaluation is, of course, a big concern. We divide the macro-model evaluation into 2 stages: *variable extraction* and *equation evaluation*. The variable extraction stage computes the variable values from an input vector pair. The equation evaluation stage multiplies the variable values with their coefficients and sums up all the terms. Our macro-model generation method guarantees that the computational time in each stage is constant, independent of the input number and the module size. To study the "absolute" speed of evaluation, we wrote a fully optimized macro-model evaluation program which takes an input vector sequence and produces power estimates for the given module at each clock cycle. The speed of execution is about 5000 vectors per second* which is acceptable (compare this with execution speed of Powermill simulation which is on average 150 gate-vector/second* which means that for a module with 1500 gates, Powermill can simulate only 0.1 vector per second). The macro-model used by the program is generated by our macro-model generation procedure and stored in a library. Notice that the absolute evaluation speed can be further improved by setting smaller number to the constants in model generation procedure, such as N_1, N_2, N_3, and K_1, K_2, K_3 in variable grouping and MAX_{var} in variable reduction. The loss in accuracy depends on the characteristics of the module itself (number of primary inputs, highest order of correlated inputs, etc).

IV. EXPERIMENTAL RESULTS

We have built our cycle-accurate macro-models for several modules, including the ISCAS-89 benchmarks. In our macro-models, we also included variables representing transitions on circuit outputs, but only for two of the circuits (C432 and C880) variables related to outputs survived the variable reduction phase.

The experimental setup is as follows. For each circuit, the population size is set to 80,000 vector pairs (including both random and non-random sub-sequences). We first simulate each circuit for the entire sequence using Powermill and record the cycle-by-cycle power. Size of the training set is set to 3,000 . The macro-model is then trained using the training set. After the macro-model is built, we apply it to different subsets of the population. These subsets are selected such that their power behaviors are different from that of the training set. Average ECP and EAP are computed by averaging the ECP's and EAP's of all sub-sets. The correlation factor r is computed based on

---

* On an UltraSparc II workstation with 256MB of memory

the fitted results on the entire population. Experimental results for our cycle-accurate macro-models is summarized in Table 4.

Experimental results shows that our macro-model technique are very accurate when estimating power consumption at RT-level. The average ECP and EAP are 10.2% and 2.0%, respectively. Meanwhile, if we compare the results with those of macro–model 1 in Table 3, which is the full-length macro-model before variable reduction, we see that our variable reduction algorithm is significantly reducing the number of variables without incurring large error.

Table 4 Experimental results of cycle-accurate macro-models

Circuit	No. of Variables	r	ECP (%)	EAP (%)
C1355	15	13.2	9.3	2.7
C1908	15	7.9	11.6	2.0
C2670	15	19.8	9.6	2.0
C3540	15	9.7	12.5	2.0
C432	14	5.1	19.3	3.1
C5315	15	27.4	7.8	1.6
C6288	15	45.9	6.2	1.9
C7552	15	6.58	6.9	1.1
C880	15	8.7	14.3	3.2
Mul16	15	34.3	6.5	1.6
ADDER16	15	28.8	6.4	1.1
AVE.				

V. OTHER APPLICATIONS OF CYCLE-ACCURATE MACRO-MODELS

Application 1 : Estimation of moving average power

Cycle-accurate macro-models can be used to provide the average power consumption of a set of input vector pairs $\{(V_{11}, V_{12}), (V_{21}, V_{22}), \ldots, (V_{n1}, V_{n2})\}$. As a special case, we want to estimate the n+1-clock-cycle moving average of power consumption, which is $\{V_1V_2\ldots V_{n+1}\}$, the set becomes $\{(V_1, V_2), (V_2, V_3), \ldots, (V_n, V_{n+1})\}$.

The estimated average error is computed by:

$$\hat{\overline{P}} = \frac{1}{n}\sum_{i=1}^{n} \hat{P}_i \tag{5.1}$$

where $\hat{P}_i$ is the estimated power for vector pair (V_{i1}, V_{i2}). The estimation error is given by:

$$e = \hat{\overline{P}} - \overline{P} = \frac{1}{n}\sum_{i=1}^{n} \hat{P}_i - \frac{1}{n}\sum_{i=1}^{n} P_i = \frac{1}{n}\sum_{i=1}^{n} (\hat{P}_i - P_i) = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{5.2}$$

where P_i is the actual power value for vector pair (V_{i1}, V_{i2}).

As mentioned in Section II, the error terms e_i are random variables which follow the same normal distribution with mean of 0 and variance of $\mathbf{VAR}(e_i) = \sigma^2$. As a result, the error for average power e also follows the normal distribution with mean value of 0 and variance of $\mathbf{VAR}(e) = \frac{1}{n}\sigma^2$.

Proposition [8] *For a given confidence level, the confidence interval for e is smaller than the confidence interval of e_i by a factor of $\sqrt{1/n}$.*

Intuitively, the error for estimating average power will become smaller as n increases.

To demonstrate the impact of n on estimation error, Figure 3 shows the scatter plots of the estimated power versus actual power for the circuit C1908. Figure 3(a) shows the plot for power of each clock cycle (n = 1), Figure 3(b) shows the plot for average power over each 5 consecutive clock cycles (n = 5), Figure 3(c) shows the plot for average power over each 10 consecutive clock cycles (n = 10).

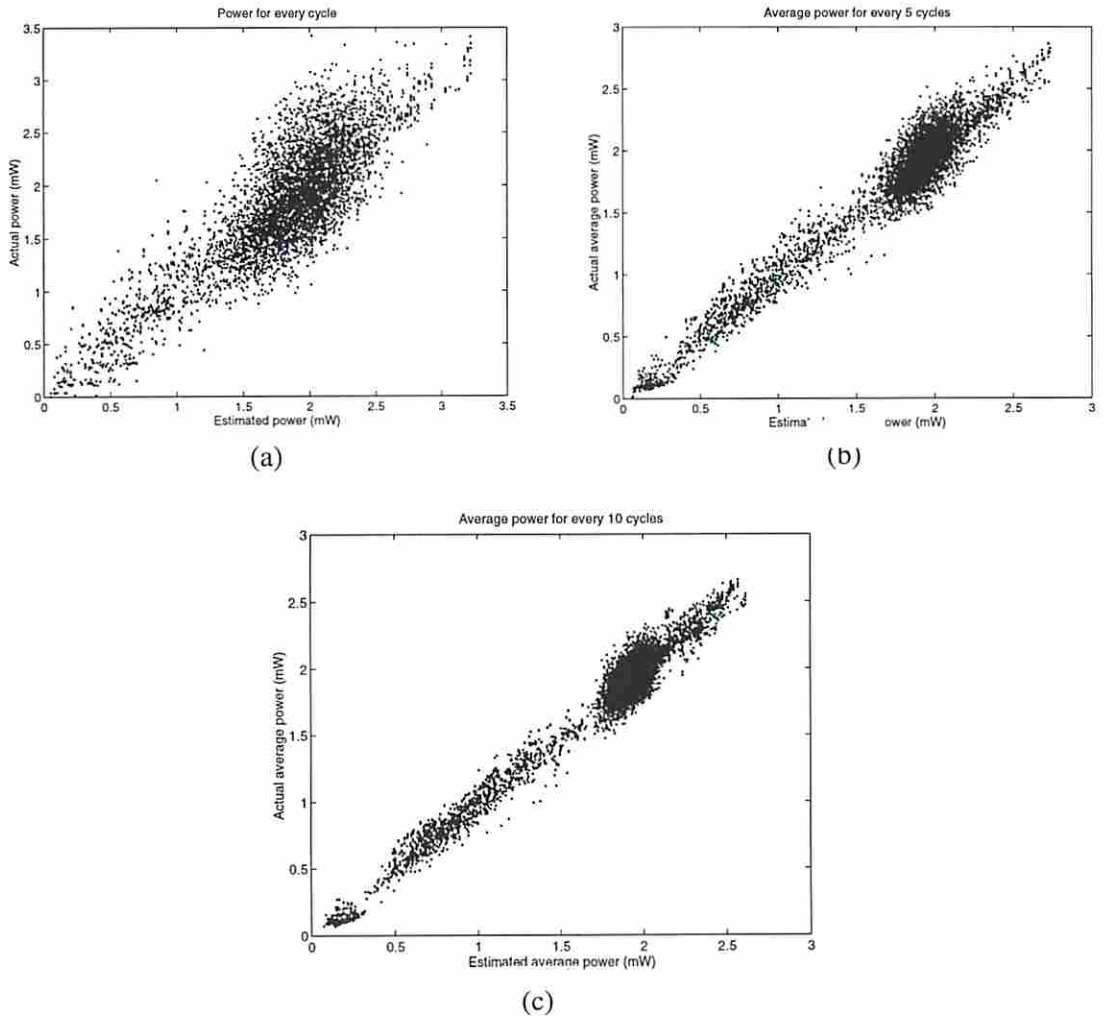


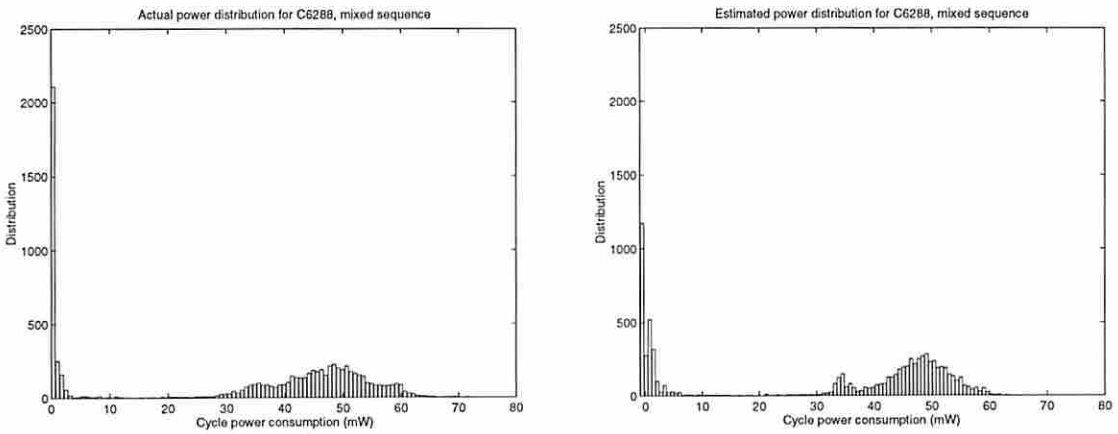
Figure 3 Scatter plots of estimating average power

It can be seen obviously from Figure 3 that, estimation error decreases when n increases from 1 to 10. Therefore we can predict (and we have already seen) that the error in estimating the average circuit power (i.e., the moving average power for large n) is very small.

Cycle-accurate macro-models can be transformed directly to cumulative macro-models which can estimate the average power consumption, given average transition probabilities and joint transition probabilities of the input sequence [6].

Application 2 : Estimation of power distribution

A good cycle-accurate macro-model is the basis for doing power analysis at RT-level. Figure 4 shows the experimental results when applying cycle-accurate macro-model to estimate the power distribution of module C6288 under a 8000-vector sequence. Figure 4(a) shows the actual power distribution, and Figure 4(b) shows the estimated power distribution. It can be seen that they are quite close.

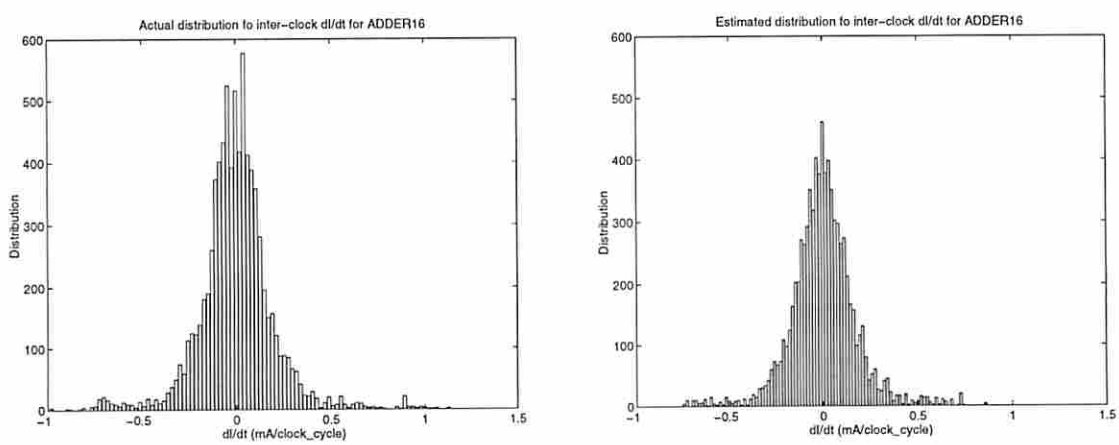


(a) (b)

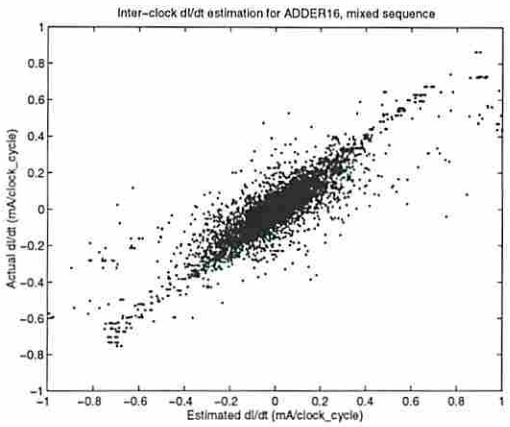
Figure 4. Cycle-accurate macro-model for estimating power distribution

Application 3 : Estimation of rate of current change (d*i*/d*t*)

Noise analysis requires information about the change in current flow between two clock cycles. Figure 5 shows the experimental results of estimating the rate of current change and its distribution using cycle-accurate macro-model for ADDER16.



(a) (b)



(c)

Figure 5. Cycle-accurate macro-model for estimating current

VI. CONCLUSION

In conclusion of our work, we present a method for generating cycle-accurate macro-models for RT-level power analysis. The proposed macro-model predicts not only the cycle-by-cycle power consumption of a module, but also the moving average of power consumption and the power profile of the module over time. We present an exact power consumption function to derive our final macro-model equation. A variable reduction algorithm has been proposed to eliminate the "insignificant" variables based on statistical sensitivity test. First order temporal correlations and spatial correlations of up to order 3 are considered in order to improve the estimation accuracy. Population stratification has been used to increase the fidelity of the macro-model. Experimental results show that, the macro-models have 15 or fewer variables and exhibit <5% error in average power, and <15% errors in cycle-by-cycle power compared to circuit simulation results using Powermill.

REFERENCES

[1] S. Powell and P. Chau, "Esitmating power dissipation of VLSI signal processing chips: The PFA techniques", *Proceeedings of IEEE Workshop on VLSI Signal Processing IV*, vol. IV, pp.250-259, 1990.

[2] P.Landman and J. Rabaey, "Architectural Power Analysis: The Dual Bit Type Method", *IEEE trans. VLSI*, Vol. 3, pp. 173-187, Jun. 1995.

[3] D. Liu and C. Svensson, "Power consumption estimation in CMOS VLSI chips", *IEEE Journal of Solid State Circuits*, vol. 29, pp.663-670, Jun. 1994.

[4] P. Landman and J. Rabaey, "Activity-Sensitive Architectural Power Analysis", *IEEE trans. CAD*, Vol. 15, pp. 571-587, Jun. 1996.

[5] S. Gupta and F.F. Najm, "Power Macromodeling for High Level Power Estimation", *Proceedings of Design Automation Conference*, pp.365-370, Jun. 1997

[6] Q. Wu, C. Ding, C. Hsieh, and M. Pedram, "Statistical Design of Macro-models For RT-Level Power Evaluation", *Proceedings of the Asia and South Pacific Design Automation Conference*, pp.523-528, Jan. 1997.

[7] PowerMill User manual, Release 3.1, EPIC Design Technology, 1994

[8] J. Neter, W. Wasseman, and M. H. Kutner, *Applied Linear Regression Models*, Second Edition, Richard D. Irwin, Inc, 1989.