Frequency-Domain Supply Current Macro-Model*

Srinivas Bodapati
ECE Dept. and Coordinated Science Lab.
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801, USA
bodapati@uiuc.edu

Farid N. Najm
ECE Department
University of Toronto
Toronto, Ontario, Canada M5S 3G4
f.najm@toronto.edu

ABSTRACT

In order to perform block level analysis of the on-chip power distribution network, a high-level model is required that captures the dependence of the current waveform drawn by a logic block, per cycle, on its input vector pair. We present a frequency domain macro-modeling technique for capturing this dependence. The macro-model is based on estimating the *Discrete Cosine Transform* (DCT) of the current waveform and then taking the inverse transform to estimate the time domain current waveform.

1. INTRODUCTION

The International Technology Roadmap for Semiconductors (ITRS-99) indicates that the total power supply current delivered to a high performance integrated circuit will grow to about 150 Amperes in 2006, and 300 Amperes in 2012. These large current values will seriously complicate the design of the on-chip power/ground distribution network, which we refer to simply as the *power grid*. Large current density in the grid would reduce its physical reliability, and can lead to significant voltage drop causing the circuit to slow down and not meet the performance spec. Therefore, it is important to do early design planning of the power grid, and to reduce the chances of having to redesign large parts of the grid [1]. In an environment where design blocks are being reused (hard IP blocks), it becomes highly advantageous to have block-level models that can give the current waveform drawn by a logic block in response to a given input vector stream. With these current models, one can perform early and fast block-level analysis of the currents and voltages in the power grid.

We propose a bottom-up current waveform macro-model for logic blocks. Existing bottom up macro-modeling techniques have targeted either the average power [11, 9, 3, 6, 2] or energy-per-cycle [4, 7]. Current waveform macro-modeling is difficult because of large variations in current waveform shapes in time domain. To overcome this problem, we have developed an approach for current waveform modeling that is based on a transformation to the frequency domain. We create a model for predicting the frequency-domain transform, which we then inverse-transform to obtain the time-domain current waveform.

Specifically, large variations in waveform shapes in the time domain translate to variations only in the parameters of the frequency-domain transforms, but not in their overall shape. Given a certain transform, we propose to construct a model that captures the dependence of its parameters on the input vector pairs. We have found that one can use low-order polynomial models to capture this dependence, and we use regression to generate these polynomials, based on a number of randomly generated vector pairs for which the circuit is simulated in HSPICE, in a process that is similar to cell library characterization.

*This research was supported by the Semiconductor Research Corporation, under SRC 97-DJ-484, with funds from Texas Instruments Inc. and SRC 99-TJ-682, with funds from IBM Corp.

ISLPED'01, August 6-7, 2001, Huntington Beach, California, USA.

2. DISCRETE COSINE TRANSFORM

The supply current waveform obtained from SPICE is a discrete time signal. The sampling time-period [10] of the signal is the time step specified in the transient simulation. The length of this discrete signal, denoted by N is the total number of samples obtained over the entire transient simulation period, for a given input vector pair. The 1-dimensional Discrete Cosine Transform [8] (DCT) of a sequence $\{x[n], 0 \leq n \leq N-1\}$ is defined as:

$$X[k] = \alpha(k) \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi(2n+1)k}{2N}\right], 0 \leq k \leq N-1$$

where:

$$\alpha(0) = \sqrt{\frac{1}{N}},\ \alpha(k) = \sqrt{\frac{2}{N}} \text{ for } 1 \leq k \leq N-1$$

The inverse transformation is given by:

$$x[n] = \sum_{k=0}^{N-1} \alpha(k) X[k] \cos\left[\frac{\pi(2n+1)k}{2N}\right], 0 \leq n \leq N-1$$

In our case, $x[n]$ is the sequence of current samples, and $X[k]$ is the sequence of DCT values. The DCT is related to the Discrete Fourier Transform (DFT), as follows. The DCT can be computed by extending the input sequence of N samples to a $2N$-sample sequence with even symmetry, taking a $2N$-point DFT, and saving only N terms of it.

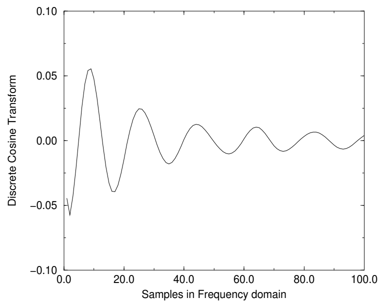


Figure 1: DCT plot of a typical current waveform.

For example, the DCT of a simple triangular current waveform $i(t) = t - a$, for $0 \leq t \leq a$ (and zero otherwise), can be obtained by uniformly sampling the following continuous transform (here, x is frequency):

$$I(x) = \frac{2ax\cos(ax) - 2\sin(ax)}{x^2}$$

This simple example is instructive because it shows an essential feature of a sinusoidal function that decays in amplitude according to $1/x^2$. This feature turns out to be common with the DCT of typical current waveforms, as shown in Fig. 1. We picked the DCT (as opposed to FFT or other transforms) because it has the most regular shape, offering an excellent candidate for model construction. Fig. 1 shows only the first 100 points of the 10,000 point DCT, for clarity, and is typical of most current waveform shapes that we have seen. Instead of constructing a model for each point on the DCT, we estimate the parameters of the decaying sinusoid as a function of input variables. We use a generalized function of a decaying sinusoid as: $I(k) = D(k)A\cos(\omega(k-1)+\pi), k = 1, 2, \ldots$ In order to completely describe a function of this form, we need four parameters:

Decay Factor $D(k)$: Motivated by the transform of the simple triangular waveform, seen above, we include a $1/k^2$ dependence in this function, among other things, as explained in section 3.1, below.

Amplitude A: We take the first largest sample point as the amplitude, and then all the sample points are measured with respect to this reference. This is further explained in section 3.2, below.

Frequency ω: This is the frequency of the sinusoid, which turns out to be not quite a constant. Our approach involves estimating the first few periods of the sinusoid as separate parameters, as explained in section 3.2, below.

DC Value $I(0)$: More on this below in section 3.2.

3. MACRO-MODEL CONSTRUCTION

In order to use the input vector pair as an input variable to the model, we need a function which maps the vector pair to a set of real numbers that can themselves be the input variables to the model. For a single input node, we want to perform a mapping $f(\cdot)$ to the real line as follows:

$$f(y) \to \Re, \text{ where } y \in \{0 \to 1, 1 \to 0, 1 \to 1, 0 \to 0\}$$

Table 1: Function for large Hamming distances.

y	$f_1(y)$
$0 \to 1$	1
$1 \to 0$	2
$1 \to 1$	3
$0 \to 0$	4

If one proposes to build a regression-based model on y, then it must be viewed, strictly speaking, as a categorical variable [5] and 3 variables would be needed to capture the 4 possible values of y ($f(\cdot)$ would have to be a vector-valued function with 3 elements). This is undesirable because it increases the number of variables. We have found that, depending on the Hamming distance of the vector pair, we can introduce a less expensive solution as follows. For large Hamming distances (more the 70% of the inputs are switching) we use $f_1(y)$, as shown in Table 1, where each input variable y, is mapped to a unique integer. In case of low Hamming distance, we use three variables as statistically required (and as shown in [4]), i.e each y is mapped to a three tuple of 0's and 1's such that the inner product between the tuples is zero.

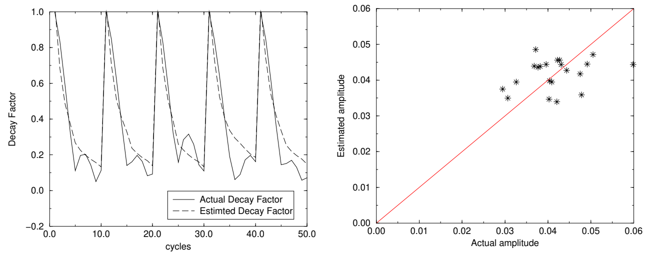
This case-analysis based on Hamming distance was found to be useful for another reason, as follows. It was found that if we build a single model for all input vector pairs (in which the Hamming distance is implicit, not an explicit variable), we would need at least a second order polynomial for most of the parameters. However, if we partition the model based on the Hamming distance, we can use just a linear model for the amplitude (A), and we do not need to to estimate the time period at all, because the time period does not vary significantly for the same Hamming distance. Therefore we partitioned our model based on the Hamming distance of the input vector pairs. It remains to identify the functions which can be used to estimate the parameters of the function $I(k)$, which can capture the internal behavior of a given circuit as a function of the input variables, thereby providing a current waveform estimate.

3.1 Decay factor

In order to represent the delay factor $D(k)$, we use a function of the following form (which is motivated by the form of $I(k)$ for the triangular waveform that we saw earlier): $D(k) = \frac{ak}{bk^2+c}$ where k is the sample point, a, b and c are linear functions of the input variables $y_i, i = 1, \ldots, p$. Thus a, b and c are given by:

$$a = \sum_{i=1}^{p} \alpha_i f_j(y_i),\ b = \sum_{i=1}^{p} \beta_i f_j(y_i),\ c = \sum_{i=1}^{p} \gamma_i f_j(y_i)$$

The values of $\alpha_i, \beta_i, \gamma_i$ are obtained by substituting the above expressions of a, b, c in the function $D(k)$ and using regression, based on the results of HSPICE simulations using randomly generated vector pairs, in a process of characterization. As part of the regression, we consider the points on the DCT corresponding to the maxima and minima of the sinusoid, and we normalize them by the amplitude A (so that $D(1) = 1$; the normalization is required because the model for $I(k)$ given above includes an explicit term for the amplitude). Furthermore, in order to reduce the computational cost of building the model, we only consider the maxima and minima in the first five cycles during regression. Some typical decay curves and their corresponding estimates obtained using our model are shown in Fig. 2(a). These were obtained from the c432 ISCAS-85 benchmark circuit.



(a) **(b)**

Figure 2. For c432 (a) decay factor and its estimate (b) typical amplitude and its estimate

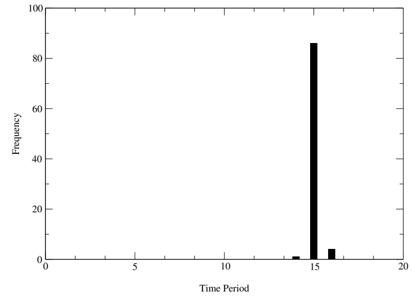


Figure 3. Distribution of the first time period, for c432

3.2 Amplitude, dc value and time Period

After partitioning the model based on Hamming distance, it was found that the amplitude A can be estimated with simply a linear model. It was also observed that for higher Hamming distances the amplitude estimation was more accurate than for low Hamming distances, but the magnitude of the current for low Hamming distance is also typically much smaller. Thus the linear model for amplitude is given by, $A = \sum_{i=1}^{p} a_i f_j(y_i)$, where the coefficients a_i are obtained using regression on the same set of waveforms used for $D(k)$. The average error in case of amplitude estimation with a linear model is around 20%. A typical plot of actual amplitude vs estimated amplitude is shown in Fig 2(b).

As far as the dc value $I(0)$ is concerned, for the same reasons as mentioned for amplitude, we use a linear model to estimate $I(0)$. Thus, $I(0)$ is also given by a function similar to A, with a different set of coefficients, obtained using linear regression.

If one attempts to build a single comprehensive model in which the Hamming distance is implicit, the time period (variation in the frequency of the sinusoidal DCT) turns out to be the most difficult parameter to model. To make things worse, this parameter has the largest impact on the current model, because any error in time period causes significant change in the frequency spectrum. However, if we partition the model based on Hamming distances, as proposed above, we find that within a Hamming distance, there is not a significant variation in the time period. For every Hamming distance, we compute during characterization a *nominal* value of the first five time periods. This is selected as the time period value that corresponds to the peak of the distribution of observed time periods for that Hamming distance. The choice of just five time periods has given sufficient accuracy, but is not a limitation of the model and can be increased if need be. A typical histogram of the first time period is shown in Fig 3.

3.3 Multiple peaks

Another issue regarding current estimation is when we have multiple peaks in time domain. In case of multiple peaks, the shape of the DCT can change from the straightforward decaying sinusoid, and can develop some high frequency changes. In such cases, our model can still offer good agreement, as seen in Figs. 5, 7, 9 & 10. We have developed an alternate flow whereby one can partition the time domain waveform into sub-intervals so that each sub-interval contains a single peak, and then we build a model for that sub-interval. This can lead to results such as in Fig. 5 where the agreement for multiple peaks is improved.

4. EXPERIMENTAL RESULTS

The current macro-model was obtained for various benchmark circuits using the methodology discussed in this paper. Basically a set of randomly generated test vectors were used to estimate the coefficients of the model, and then another set of randomly generated vectors were used to test the accuracy of the model. The resulting waveforms for some of those vector pairs are shown in Fig 4–11. The benchmark circuits used to test our model are given in Table 2. Since we are ignoring most of the high frequency components, the estimated waveforms are very smooth, and we don't see very sharp variations.

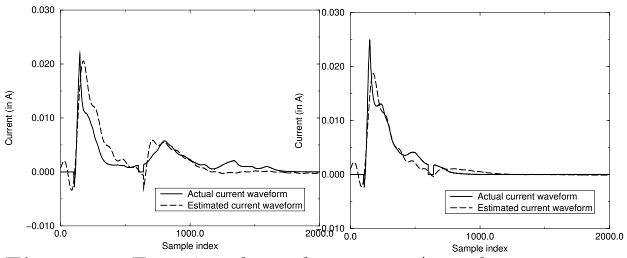


Figure 4. For c432, ham dist = 32, Actual vs. estimated current waveform

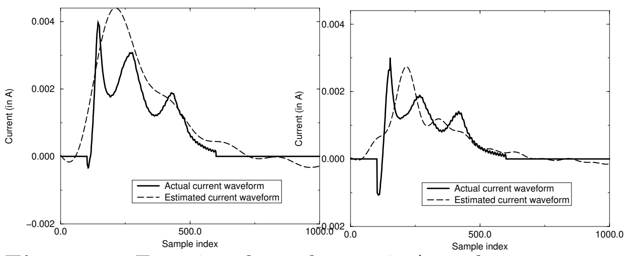


Figure 5. For c432, ham dist = 4, Actual vs. estimated current waveform

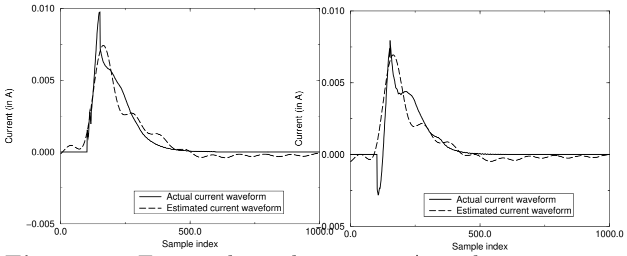


Figure 6. For cu, ham dist = 12, Actual vs. estimated current waveform

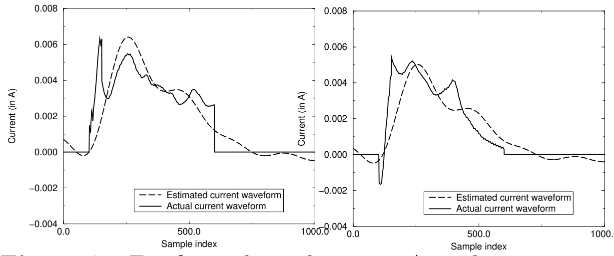


Figure 7. For f51m, ham dist = 4, Actual vs. estimated current waveform

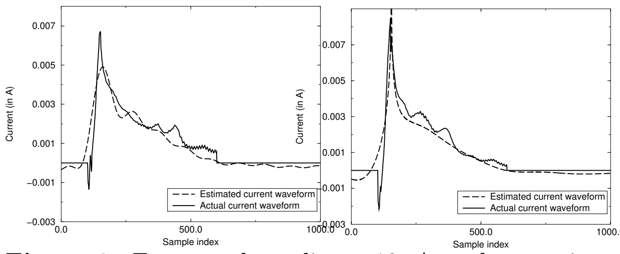


Figure 8. For mux, ham dist = 19, Actual vs. estimated current waveform

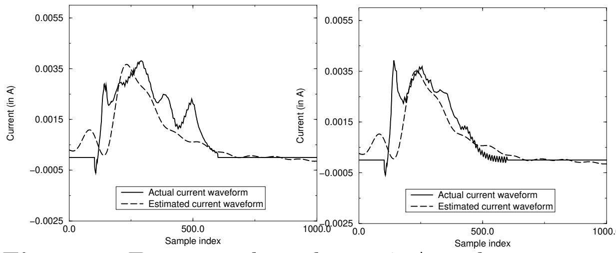


Figure 9. For parity, ham dist = 4, Actual vs. estimated current waveform

5. CONCLUSION

In order to enable early block-level analysis of the power grid, when using hard IP blocks, we have proposed a cycle-based current waveform modeling technique that involves predicting the frequency transform from the input vector pairs, and using the inverse transformation to get back the time-domain waveform. The use of frequency domain analysis is motivated by our observation that, while the time domain waveforms show large variations in shape (making them hard to model in the time domain), the frequency domain transforms (such as DCT) show much less variation and are mostly limited to variations in their parameters.

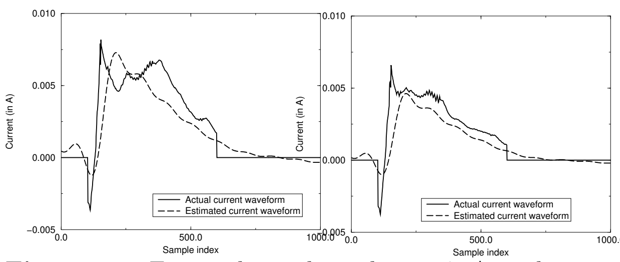


Figure 10. For random8, ham dist = 4, Actual vs. estimated current waveform

Table 2: Benchmark circuits used for characterization.

Circuit	#I	#O	#Gates
c432	36	7	217
cu	14	11	48
parity	16	1	68
mux	21	1	91
f51m1	8	3	105
random8	8	1	158
x2	10	3	50

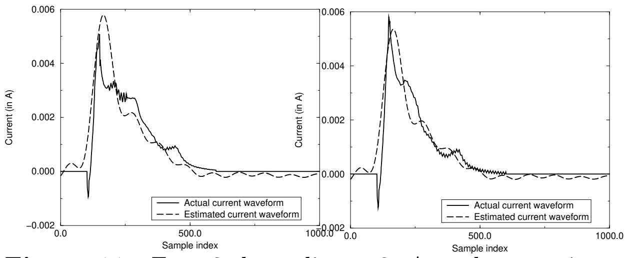


Figure 11. For x2, ham dist = 8, Actual vs. estimated current waveform

6. REFERENCES

[1] A. Dharchoudhury R. Panda D. Blaauw and D. Bearden. Design and analysis of power distribution networks in powerpc microprocessor. In *ACM/IEEE DAC*, pages 738 – 743. IEEE, 1998.

[2] A. Bogliolo and L. Benini. Node sampling: a robust rtl power modeling approach. In *IEEE ICCAD*, pages 461–467. ACM/IEEE, 1998.

[3] A. Raghunathan S. Dey and N. K. Jha. Register-transfer level estimation techniques for switching activity and power consumption. In *IEEE ICCAD*, pages 158 – 165. IEEE, 1996.

[4] Q. Qiu Q. Wu. Chih-S. Ding and M. Pedram. Cycle-accurate macro-models for RT-level power analysis. *IEEE Tans. on VLSI Systems*, 6:420–528, December 1998.

[5] R. F. Gunst and R. L. Mason. *Regression Analysis And Its Application, A Data-oriented approach.* Marcel Dekker, Inc, 1980.

[6] S. Gupta and F. N. Najm. Power macromodeling for high level power estimation. In *34th ACM/IEEE DAC*, pages 365–370. IEEE, 1997.

[7] S. Gupta and F. N. Najm. Energy-per-cycle estimation at rtl. In *ISLPED*, pages 121–126. IEEE, 1999.

[8] A. K. Jain. *The Fundamentals Of Digital Image Processing.* Prentice-Hall, Inc, 1989.

[9] P. E. Landman and J. M. Rabaey. Architectural power analysis: The dual bit type method. *IEEE Trans. on VLSI Systems*, 3:173–187, June 1995.

[10] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice-Hall, Inc, 1990.

[11] S. R. Powell and P. M. Chau. Estimating power dissipation of vlsi signal processing chips: The PFA technique. *VLSI Signal Processing IV*, pages 250–259, March 1990.