A Deep Submicron Power Estimation Methodology Adaptable to Variations Between Power Characterization and Estimation

Daniel Eckerbert and Per Larsson-Edefors
daniel.eckerbert@ieee.org, perla@ce.chalmers.se
Chalmers University of Technology, Göteborg, Sweden

Abstract— Traditionally, RTL power estimation techniques are characterizing a component for a fixed environment (most importantly load capacitance, activity, and operating frequency). This article presents a solution to problems originating from the ineluctably changing operating conditions such as differing load capacitance due to different applications; different activity and operating frequency as power reduction techniques are more frequently employed.

I. Introduction

Power macro models being developed so far are first characterized using power simulation data from a gate-level simulator. The macro model is then used during actual estimation and power values form the output of these macro models. The only data that is fed to these macro models are the input vectors which in effect assume that the component environment is the same during estimation as it was during characterization.

Since gate-level simulation is used as a basis for building the power macro models, no deep submicron (DSM) phenomena are visible to the power macromodels. For DSM processes (approximately spanning from the 0.25 μm technology node) subthreshold leakage has become important. Going beyond 0.10 μm, gate leakage will start to become a problem. All of the leakage power consumption mechanisms, which have appeared as we have moved into the DSM regime, force us to change the RTL power estimation paradigm; this article will propose such a paradigm shift and analyze the effects which have forced this paradigm shift into existence.

II. Traditional Power-Estimation Methodologies

An obvious way to express the power consumption would be to use an equation such as Eq. 1

$$\mathcal{P}_{tot} = \sum_{n=0}^{N} \beta_n \gamma_n = \beta_0\gamma_0 + \beta_1\gamma_1 + \ldots + \beta_N\gamma_N \qquad (1)$$

γ_n denotes an event, e.g. a transition, and β_n denotes the corresponding cost for that particular event.

In [1], Gupta and Najm presented a efficient power macro model based on a 3D table. Variables that could be a part of the equation based model are here the different dimensions of the table.

Three independent variables were shown to be the most prominent in predicting the power consumption of a component: average input signal probability (P_{in}), average input transition density (D_{in}), and average output signal density (D_{out}). The table then provides the power estimation mapping $\mathcal{P}_{tot} = f(P_{in}, D_{in}, D_{out})$.

Three years after the publication of [1], another article, [2], was published by Gupta and Najm extending the 3D model to incorporate yet another independent variable constituting a fourth dimension. That variable is spatial correlation, SC_{in} which is the average of all SC_{ij} which is defined $SC_{ij} = P(in_i \wedge in_j)$.

III. Impact of Environment Variations from Characterization to Estimation

A common trait for all of the power estimation methodologies described in Sec. II is the assumption that the same conditions are valid from characterization to actual estimation. This article focuses on two conditions where this assumption is likely to not hold true.

A. Load Capacitance

There are two scenarios, both of them very likely to occur during typical digital system design, in which the load capacitance of a component is different during estimation as compared to characterization. 1) The component is driving a non-negligible interconnect. 2) The component is driving a number of other types of components. The impact of change in dynamic power consumption is easy to compensate for. There is however a second effect to changing the load capacitance between characterization and estimation, the short-circuit power consumption will change. This was investigated in [3].

Nose and Sakurai did publish predictions on the scaling of short-circuit power consumption in [4]. The conclusion was that since short-circuit power scales as $P_{sc} \propto V_{cc}/V_t$, the short-circuit power consumption will decrease as the threshold voltage does not scale as aggressively as does the supply voltage.

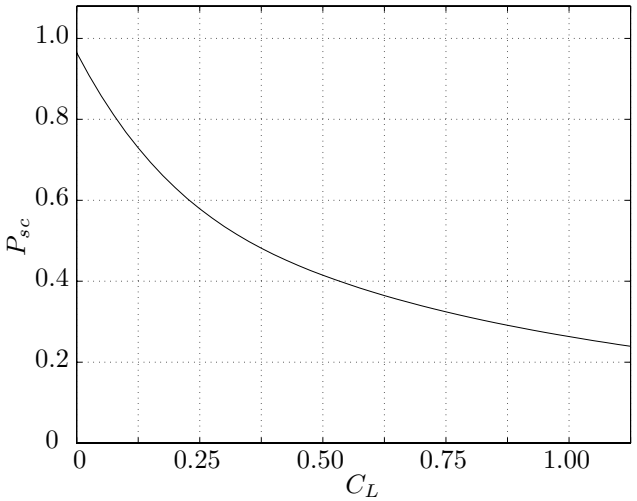

Fig. 1. The short-circuit power consumption plotted as a function of the load capacitance.

Although the short-circuit power consumption is expected to decrease, there are however components in which it is prominent. A buffer array is one such example which is found in great abundance throughout most designs; driving feedback buses in ALUs, long interconnects, etc. Fig. 1 illustrates the influence the load capacitance has on the short-circuit power consumption.

B. Activity and Switching Frequency

Changes in switching frequency (f_{sw}) can be the result of a change in operating frequency (f) *or* activity (α). Typically, the power consumption of a component is characterized close to the maximum frequency. Most components will however run at considerably lower frequency to minimize power consumption. This has become even more true with modern power reduction schemes such as gating and dynamic operating frequency control.

Looking at a data sheet for a standard cell library the power consumption is typically given as $\mu W/MHz$ which is logical considering that $\mathcal{P}_{sw} = fCV_{cc}^2$. This assumes that $\mathcal{P}_{tot} \approx \mathcal{P}_{dyn} \propto f_{sw}$, where $\mathcal{P}_{dyn} = \mathcal{P}_{sw} + \mathcal{P}_{sc}$. This assumption held true quite well until recently; but, with the emergence of deep submicron processes, care has to be taken so that the leakage power is included, $\mathcal{P}_{tot} = \mathcal{P}_{dyn} + \mathcal{P}_{leak}$ where $P_{dyn} \propto f$ and P_{leak} can be considered to be independent of f. The $\mu W/MHz$ approach would give zero power consumption for a component in sleep mode which is not the case in reality. The methodology proposed in Sec. IV utilizes the information available from $\mathcal{P}_{tot} = \mathcal{P}_{dyn} + \mathcal{P}_{leak}$, making sure that the leakage power consumption is taken into account.

Subthreshold leakage is growing exponentially with reduced threshold voltage and is hence a big problem both in terms of power consumption and circuit robustness. With the emerging sub-100 nm technology nodes, gate leakage has grown to become an important leakage power consumption mechanism. The severity of the gate leakage stems from the fact that it scales exponentially with decreasing gate oxide thickness. That subthreshold leakage has become and will continue to be important has been known for some time; in [5] the gate leakage is shown to be comparable to the subthreshold leakage. These leakages will add up to severe problems[6] and for the sub-100 nm technologies, leakage power consumption is expected to overtake the switched power consumption as the major power consumption component.

IV. Proposed Power-Estimation Methodology

As shown in the analysis in Sec. III, there is a need to move from a power-estimation methodology where the power is modeled as $\mathcal{P}_{tot}$ to one where power is modeled as $\mathcal{P}_{sw}+\mathcal{P}_{sc}+\mathcal{P}_{leak}$. All of these three power components show a dependence on the input vector set. The reason for considering them separately is that they are all influenced by changing component environment in different ways such that $\mathcal{P}_{tot}(f, C_L) = \mathcal{P}_{sw}(f)+\mathcal{P}_{sc}(f, C_L)+\mathcal{P}_{leak}$.

By separating the different components of power consumption according to their underlying mechanisms, we have achieved a model which is valid although the component environment might differ from that used during characterization. The first steps towards this separation of power sources were presented in [3] for short-circuit power and in [7] for the leakage power macro model.

Fig. 2 shows the power estimates using traditional and proposed methodologies where the proposed methodology will correctly estimate the power consumption over the entire C_L–f-surface; the error of the traditional approach is also plotted. If the load capacitance C_L is fixed, a linear scaling of $\mathcal{P}_{tot}$ can be observed which is anticipated as $\mathcal{P}_{dyn} = fCV_{cc}^2$. If the frequency f is fixed, the curved surface gives a $\mathcal{P}_{tot}$ which is identically shaped as the curve in Fig. 1.

A. Characterization and Estimation Flows

Going to a macro model consisting of several separated power macro models ($\mathcal{P}_{sw}$, $\mathcal{P}_{sc}$, etc) causes the complexity of the total power estimation methodology to increase. This is however inevitable in order to achieve a method which can take into account changes between characterization and estimation.

The first step of characterization is choosing a simulator which actually simulates the different mechanisms discussed in the analysis section: short-circuit, subthreshold leakage, and gate leakage. A contributing factor as to why the short-circuit and leakage problems have not been addressed earlier is that gate simulators have been used for power characterization. These gate simulators do not include the DSM phenomena which are the reason for the discrepancies which this article remedies.

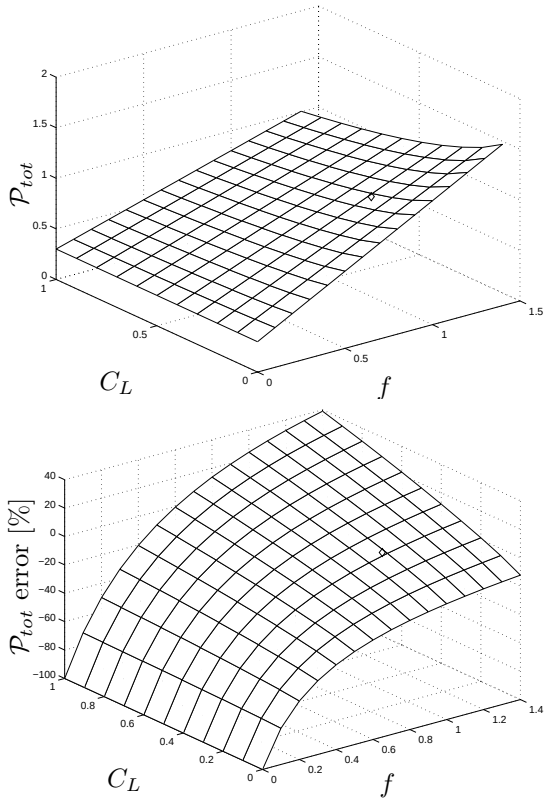

Fig. 2. Power consumption plotted at the top as a function of node switching frequency, f_{sw}, and load capacitance, C_L. The lower plots the error of the traditional approach. The traditional characterization is done where the plot is marked '$\diamond$'.

How the actual characterization is to be performed is then depending on which power macro models are chosen for the different power components. Our proposed methodology does however require extra characterization steps as a sensitivity analysis is needed for the variables C_L, t_e and f. For the load capacitance C_L, only three different load capacitances have to be simulated in order to fit the model parameters. For the edge rate t_e only two edge rates need be simulated. For the leakage, a straightforward traditional characterization can be performed with the relaxation that it is input vector dependent $\mathcal{O}(2^n)$, not *transition* dependent $\mathcal{O}(2^{2n})$. Frequency dependence is in accordance with linear scaling for dynamic power consumption and independent for the static power consumption. Power estimation is then performed by applying the function $\mathcal{P}_{tot}(in, f, t_e, C_L) = \mathcal{P}_{sw}(in, f) + \mathcal{P}_{sc}(in, f, t_e, C_L) + \mathcal{P}_{leaksub}(in) + \mathcal{P}_{gate}(in)$.

B. Switched Power Macro Model, $\mathcal{P}_{sw}(f)$

The switched power consumption is a transient event that occurs every time the input vector changes and is hence depending on the activity and operating frequency. Since it is depending on *which* input vectors are applied before and after the transition, the complexity for the model is 2^{2n} where n is the number of inputs.

Fortunately, the traditional power macro models were developed under the assumption that $\mathcal{P}_{tot} \approx \mathcal{P}_{sw}$. These power macro models are reported to have good accuracy wherefore e.g. the Gupta-Najm 4D model can be used for switched power modeling with good results.

C. Short-Circuit Power Macro Model, $\mathcal{P}_{sc}(f, t_e, C_L)$

Like the switched power, the short-circuit power is a transient event and is hence depending on both currently and previously applied input vector causing a model complexity of $\mathcal{O}(2^{2n})$.

Compared to switched power, short-circuit power has two additional dependent variables: t_e and C_L. t_e is the edge rate of the input signals which causes short-circuit power consumption in gates at the input of the component. C_L is the load capacitance at the outputs of the gate which modulates the short-circuit power consumption in the boundary gates at the outputs of the component.

$$\mathcal{P}_{sc,out} = P_{sc_0} \frac{\kappa}{\kappa + C_L^{\gamma}} \tag{2}$$

where all parameters can be characterized by simulating three different load capacitances.

$$\mathcal{P}_{sc,in} = P_{sc_0} \left(1 + \frac{t_e - t_{e_0}}{\beta\, t_{e_0}}\right) \tag{3}$$

where t_{e_0} is the edge rate used during characterization and β is a fitted parameter which roughly corresponds to the ratio of total power to short-circuit power consumption in the input boundary gates.

D. Leakage Power Macro Model, $\mathcal{P}_{leak}$

Contrary to both switched and short-circuit power, subthreshold leakage power is not dependent upon the frequency at which the input vectors are changing; it is only dependent upon the currently applied input vector. This fact greatly reduces the complexity of the model to 2^n.

Since the leakage power consumption model simply needs to map an input vector on a power consumption value, any of the traditional power models for switched power consumption can be applied with good results.

V. Results

To evaluate the proposed methodology, a number of components were characterized using the proposed methodology. The components were simulated over a number of input vectors in the same manner as for the traditional methodology. From this data, leakage figures

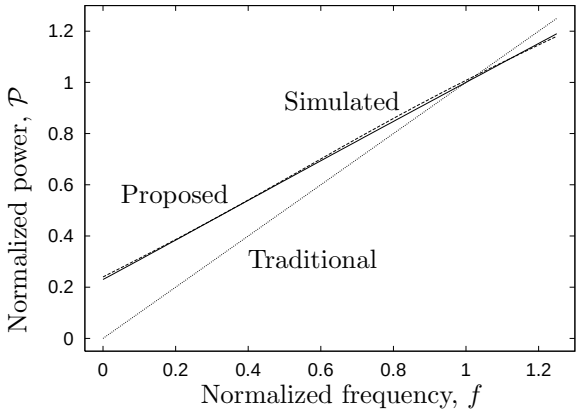

Fig. 3. The figure shows power estimation when the node switching frequency changes from characterization to estimation. Traditional methodology (dotted), proposed methodology (dashed), and simulated (solid) results are plotted.

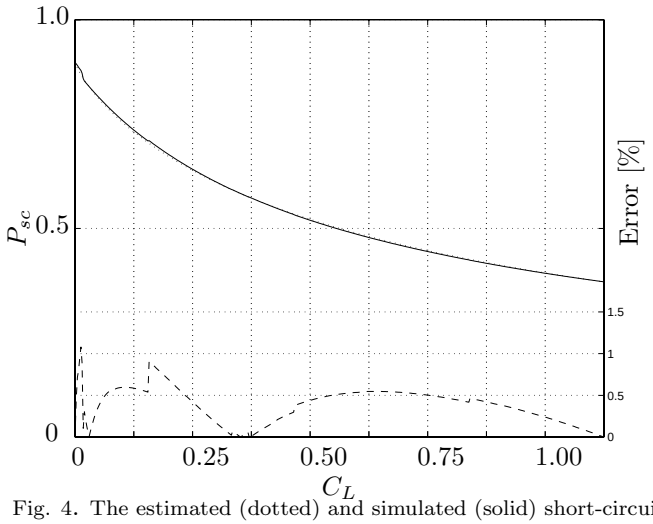

Fig. 4. The estimated (dotted) and simulated (solid) short-circuit power. The error (dashed) is plotted against the right axis.

were extracted which allowed characterization of the leakage power model. In addition, a smaller number of input vectors were simulated over a number of load capacitances. The results of the proposed model were then compared to those achieved from a traditional methodology. No gate leakage has been modeled since available process parameters are only BSIM3v3. An exhaustive table was used to evaluate the proposed methodology so as not to bias the result with the errors of any given method.

Fig. 3 shows the error due to changing switching frequency between characterization and evaluation/estimation. Fig. 4 shows the result of the short-circuit model as it depends on the load capacitance. The estimated short-circuit is within 1% of the simulated one.

As shown in Table I, for some components, such as the buffer array, changing the load capacitance from characterization to evaluation or estimation will significantly affect the accuracy. For other components, such as arithmetic components, which have a large logic depth, a changing load capacitance will not significantly influence the overall accuracy while a changing operating frequency or activity will seriously degrade the accuracy.

TABLE I
Accuracy for a buffer array and an arithmetic component.

Component	Short-Circuit	Leakage
Buffer array	17%	34%
Adder	4%	29%

There is a cost associated with the increased accuracy; that cost is increased complexity for both characterization and estimation. The actual increase in complexity is depending on the power macro models chosen.

VI. Summary and Conclusions

In this article, an analysis was done regarding the implications of changing component environment (load capacitance and node switching frequency). Very large errors are introduced if load capacitance and node switching frequency are not taken into account during power estimation. If a power reduction scheme reduces the system frequency to a very low frequency, the error will be the entire leakage power consumption; in sub-100 nm technology nodes and beyond, the active leakage is even greater than the active switched power consumption. A methodology was also presented for a power characterization and estimation flow which takes these changes into account.

References

[1] S. Gupta and F.N. Najm, "Power macromodeling for high level power estimation," in *Proc of DAC* , pp. 365–370, 1997.

[2] S. Gupta and F.N. Najm, "Power modeling for high-level power estimation," *IEEE Trans. on VLSI Systems*, pp. 18-29, 2001.

[3] D. Eckerbert and P. Larsson-Edefors, "Interconnect-driven short-circuit power modeling," *Proc. of the 2001 Euromicro Digital System Design*, pp. 414–421, 2001.

[4] K. Nose and T. Sakurai, "Analysis and future trend of short-circuit power," *IEEE Trans. on CAD of Integrated Circuits and Systems*, Vol. 19, pp. 1023–1030, 2000.

[5] Y.-S. Lin and C.-C. Wu and C.-S. Chang and R.-P. Yang and W.-M. Chen and J.-J. Liaw and C.H. Diaz, "Leakage scaling in deep submicron CMOS for SoC," *IEEE Trans. on Electron Devices*, Vol. 49, pp. 1034–1041, 2002.

[6] The Semiconductor Industry Association, "The international technology roadmap for semiconductors, 2001 edition," , 2001.

[7] D. Eckerbert and P. Larsson-Edefors, "Cycle-true leakage current modeling for CMOS gates," *Proc. of the International Symposium on Circuits and Systems*, pp. V 507–510, 2001.