Moment-Based Power Estimation in Very Deep Submicron Technologies

Alberto Garcia-Ortiz, Lukusa Kabulepa, Tudor Murgan, Manfred Glesner
Institute of Microelectronic Systems, Darmstadt University of Technology
Karlstrasse 15, D-64283 Darmstadt (Germany)
agarcia@mes.tu-darmstadt.de

ABSTRACT

The significant power optimization possibilities in the early stages of the design flow advice the use of energy evaluation techniques at high levels of abstraction. With this aim, the present work addresses the estimation of the energy consumption in very deep submicron technologies. Using the characterization of the probability density function with a projection in an orthogonal polynomial base, and a symbolic propagation mechanism, a technique is presented to estimate the dynamic and static power consumption in digital systems. The proposed approach has been validated with circuits and excitations from realistic applications. Comparisons with reference transistor and bit level simulations are reported in order to asses the the accuracy of the technique.

1. INTRODUCTION

During the last decade, power consumption has become a major concern for the design community, due mainly to portability and thermal dissipation concerns. Moreover, with the advent of very deep submicron technologies (VDSM) some of these issues are getting even more severe, and therefore must be addressed from the very early stages of the design flow, where the optimization possibilities are larger.

At the gate level, algorithms addressing the estimation of the node switching and bit probabilities are mature enough to be integrated in commercial tools (such as Synopsys's Power Compiler^{TM} or Cadence's BuildGates^{TM}). At higher levels of abstraction however, more abstract approaches based on word-level signal parameters rather than bit-level probabilities are required. Furthermore, in modern technologies, transition activity does not suffice to provide sensible energy estimates. Submicron effects, as leakage current and coupling capacitances in long wires, affect the overall power consumption significantly, and must be therefore considered.

Different approaches [6, 1, 9, 10, 7] have been developed to estimate the power consumption in DSP architectures based on the analysis (at high levels of abstraction) of the switching activity of the signals. In these techniques, word-level statistics are propagated through the architecture, and used to perform bit level estimations of the switching activity. The main drawback of the aforementioned approaches is the underlying assumption that the signals have a distribution close to a Gaussian. This assumption reduces the usability of the estimation procedures to a relatively narrow set of practical design scenarios.

The problem of switching activity estimation in general architectures has been considered in [4], where assuming that the signals are not correlated, an accurate estimation approach is presented. Nevertheless, since this technique requires the analytical knowledge of the probability density function (i.e., a complete equation), its use is restricted to some very particular applications. An alternative methodology is described in [2]. But since it requires expensive multidimensional integration procedures, it cannot be applied for the efficient exploration of competing architectures.

Classic estimation methodologies have focused on transition activity. Although this is definitely a relevant metric for estimating dynamic power consumption in digital blocks, VDSM technologies require to take into account some other effects. Among them, coupling capacitance and leakage current are the most relevant.

Due to the increase in the aspect ratio of deep sub-micron wires [12], the coupled capacitance between adjacent bus lines tends to increase and even to dominate the total capacitance of the wire. Thus, the power consumption associated with the drivers of a bus does not depend only on the toggling of the individual lines, but also on the simultaneous transition of adjacent bus lines. Using a model similar to [8], which assumes a full-rail swing of the bit lines equals to V_{dd}, the energy dissipated in each line driver can be estimated as:

$$E_{bus}(i) = \frac{V_{dd}^2}{2}(C_t\, t_i + C_e\, te_i) \quad (1)$$

where C_t is the total ground capacitances and C_e the coupling capacitance. The factor t_i (called switching or *transition activity*) measures the probability of a toggling in the ith bus line. The equivalent factor for the coupling capacitance is the *equivalent spatial transition activity*, te_i. For accurate power estimations in interconnect structures both t_i and te_i must be estimated.

As the dimensions of the transistor get smaller, the internal electrical fields in the device tend to increase, and therefore the maximum operating voltage gets limited. In order to provide sufficient output current (and thus speed), the threshold voltage of the N- and P-MOS transistors must be also scaled down. In this scenario, leakage current increases exponentially with each technological node. The main contributor in CMOS technologies is the subthreshold current, which appears between drain and source when the transistor is OFF. A second factor (that is getting more important in the new projected thin- gate-oxide technologies) is the tunneling leakage current [5]. A key issue is that these two factors are strongly dependent on the input state of a circuit; and consequently, they can only be estimated by considering the probability density func-

ICCAD'03, November 9-12, 2003, San Jose, California, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICCAD'03, November 11-13, 2003, San Jose, California, USA.

tion (PDF) of the inputs signals. Conceptually, the leakage current in stand-by mode can be easily evaluated by a DC analysis for each possible input value, with a SPICE-like simulator. For the active-mode in combinational blocks, some preliminary results [3] suggest that a similar approach is possible, but assuming that during the signal transition periods the leakage can be neglected. Therefore, a unified framework for estimating both static (probability dominated) and dynamic (transition dominated) power contributors is highly desirable.

This work presents a stochastic data-model approach and an associated estimation algorithm suitable for energy evaluation in deep submicron technologies. It uses the statistical moments of a signal, instead of its complete PDF [4]. Since these moments can be easily propagated through the design, the approach is very efficient, and it is not restricted to particular distributions as is the case of most previous works [6, 1, 10, 7]. Furthermore, the methodology is not restricted to a particular power metric. Within a unified framework, it can be used to estimate the dynamic and static power dissipation in VDSM technologies, considering switching activity, coupling capacitances, and leakage currents.

The rest of the paper is structured as follows: Section 2 describes the proposed approach and presents the different polynomial bases used for the estimation. An algorithm for the symbolic propagation of moments is then presented in section 3. The estimation methodology provided by those two techniques is afterwards validated. Conclusions and final remarks are presented in section 5.

2. ESTIMATION APPROACH

The goal of this work is to provide efficient estimations for the bit probability (p_i), transition activity (t_i), spatial transition activity (te_i), and static power consumption. The signals are modeled by stochastic stationary random process $X[t]$, and therefore, they can be characterized by a probability density function $f(x)$. If that PDF is analytically known, p_i can be exactly evaluated [9]. Given a B-bits unsigned signal:

$$p_i = \sum_{x=0}^{2^B-1} f(x)\, Bit_i(x) \tag{2}$$

where $Bit_i(x)$ returns the value of the ith bit of x. Further on, the average static consumption can be obtained when the function $Bit_i(k)$ is substituted by another one, namely $P_{static}(x)$, which provides the static energy cost associated with each input state x.

Since the transition activity and equivalent spatial transition activity depend on the previous and actual value of the signal, they require the joint PDF of $X[n], X[n-1]$. Let $f(x,y)$ denote that PDF and let $Tran_i(x,y)$ be a function which returns a one if the ith bit of x and y differ (i.e., an XOR). Then:

$$t_i = \sum_{x=0}^{2^B-1} \sum_{y=0}^{2^B-1} f(x,y)\, Tran_i(x,y) \tag{3}$$

In the same way, te_i can be obtained by substituting $Tran_i(x,y)$ with the cost function associated with te_i. Furthermore, if the random variables $X[n]$ and $X[n-1]$ are independent, t_i and te_i can be calculated from the one-dimensional PDF $f(x)$. For example, as described in [1], $t_i = 2\,p_i\,(1-p_i)$.

From the previous analysis, it is clear that the two main obstacles for estimating p_i, t_i, te_i, and the static power consumption are the characterization of the PDF, and the efficient approximation of **Eq**. (2)–**Eq**. (3) without evaluating all the additions (the complexity is *exponential* with respect to the number of bits). For the sake of simplicity, we consider the estimation of the one-dimensional PDF $f(x)$, which allows the evaluation of p_i, the static power consumption, and both t_i and te_i when the signal is uncorrelated. However, the approach can be extended to the two-dimensional case, which provides estimations for t_i and te_i in general correlated signals.

Different approaches, both parametric and non-parametric, have been proposed to approximate an unknown density function (see [11] for a detailed exposition). In this work, we analyze the use of a linear combination of orthogonal polynomial functions as an approximation procedure. Within this framework, the PDF of the signal, $f(x)$, can be then approximated as:

$$f(x) \approx \sum_{i=0}^{N-1} c_i\, q_i(x)\phi(x) \tag{4}$$

where c_i are the fitting coefficients, $\phi(x)$ is a positive function that defines a scalar product, and $q_i(x)$ is a set of polynomials that are orthogonal for the given norm. The *approximation order N* represents the number of linear terms in the approximation, and can be used to trade accuracy with model complexity. We can calculate the discrepancy (i.e. squared average error) between the given probability function, $f(x)$, and the linear model as:

$$\varepsilon(c) = \int_{-\infty}^{+\infty} [f(x) - \phi(x) \sum_i c_i\, q_i(x)]^2\, w(x)\, dx \tag{5}$$

where $w(x)$ represents a scaling function that weights the importance of the approximation error at each particular value of x. In order to perform the estimation, we can find the set of constants c_i that minimize **Eq**. (5). Forcing the derivative to be zero, and choosing a weight function that holds $w(x)\cdot\phi(x) = 1$, we get [11]:

$$c_k = \frac{\int_{-\infty}^{+\infty} f(x)\, q_k(x)\, dx}{\int_{-\infty}^{+\infty} q_k^2(x)\, \phi(x)\, dx} \tag{6}$$

Since $q_k(x)$ is a polynomial, the numerator of **Eq**. (6) can be expressed as a linear combination of the moments of $f(x)$. Let us denote by $a_{k,j}$ the jth coefficient of the polynomial $q_k(x)$, and with γ_j the jth moment of $f(x)$, formally defined by:

$$\gamma_j = \mathbf{E}[x^j] = \int_{-\infty}^{+\infty} x^j\, f(x)\, dx \tag{7}$$

Then, c_k can be calculated as follows:

$$c_k = \frac{\sum_j a_{k,j}\, \gamma_j}{\int_{-\infty}^{+\infty} \phi(x)\, q_k^2(x)\, dx} \tag{8}$$

Once the coefficients c_k are known, the PDF $f(x)$ can be estimated using **Eq**. (4); and therefore, the different energy metrics (p_i, t_i, etc) can be obtained with **Eq**. (2)–**Eq**. (3). In order to avoid the exponential number of operations involved in those expressions, it is possible to exploit the linear form of **Eq**. (4) to provide faster estimations. The idea is to pre-compute the cost associated with each one of the terms $q_i(x)\phi(x)$. For illustration purposes, we show the procedure for p_i, but the same technique can be applied to the other metrics.

Substituting **Eq**. (4) into **Eq**. (2) we get:

$$p_i = \sum_{x=0}^{2^B-1} \left(\sum_{j=0}^{N-1} c_j\, q_j(x)\phi(x) \right) Bit_i(x) \tag{9}$$

$$= \sum_{j=0}^{N-1} BitCost_i(j)\, c_j \tag{10}$$

where $BitCost_i(j) = \sum_{x=0}^{2^B-1} q_j(x)\phi(x)\, Bit_i(x)$ is a set of N parameters that can be precalculated for each power metric. For a input

width of 10 bits, the direct use of **Eq**. (2) requires 1024 computations, whereas our technique results typically in less than 7 operations (see sec. 4 where N typically equals 7). Thus, a complexity reduction of more than two orders of magnitude is obtained. For t_i and te_i the gains are even larger.

2.1 Polynomial bases

In this work the three major families of orthogonal polynomials (i.e., Legendre, Laguerre, and Hermite) have been considered:

Legendre polynomials. This family corresponds to the norm function $\phi(x) = 1$ inside the interval $[-1,1]$. Thus, the weight function is constant, and the approximation error gets uniform distributed inside that interval. In fact, the expansion in Legendre polynomials corresponds to the minimum square error polynomial fitting of the function $f(x)$. Observe that if the density function $f(x)$ is in the range $[x_{min}, x_{max}]$, we can use the transformation $z = 2(x - x_{min})/(x_{max} - x_{min}) - 1$ to move $f(x)$ inside the working interval $[-1,1]$, where the approximation can be performed. The coefficients c_k for $k \leq 4$ are depicted in **Tab**. 1.

Laguerre polynomials. The Laguerre polynomials, $L_n(x)$, are defined in the range $[0, +\infty[$ to be orthogonal with the norm function $\phi(x) = exp(-x)$. Since this function is not constant, there is a weight factor in the definition of the error (see **Eq**. (5)). The larger values of x are more strongly pondered, and thus, the approximation tends to be more accurate in the tales of $f(x)$. The best approximation results are achieved when the function $f(x)$ is scaled in such a way that its mean equals one. As a reference, **Tab**. 1 depicts the values of c_k in this particular case.

Hermite polynomials. The Hermite polynomials are orthogonal with the norm function $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, which corresponds to the PDF of a Gaussian distribution. As in the case of the Laguerre polynomials, the weight function strengthens the accuracy for larger values of x; but in this case, the weight is even stronger. Therefore, if $f(x)$ differs significantly with respect to a Gaussian distribution, the approximation is quite poor. As we show in in the following simulations, this fact makes this base quite un-robust. The expansion of a PDF with a Gaussian kernel and Hermite polynomials has been very often used in statistics under the name of Gram-Charlier expansion. The best approximation results used to be achieved when the function $f(x)$ is linearly transformed to have zero mean, and unit variance. After this transformation $f(x)$ will resemble a normalized Gaussian distribution, and thus the second and third coefficients of the approximation are zero.

	Legendre	Laguerre
c_0	$1/2$	1
c_1	$3\gamma_1/2$	9
c_2	$3\gamma_2 - 1$	$\frac{\gamma_2 - 1}{2}$
c_3	$25\gamma_3 - 15\gamma_1$	$\frac{-\gamma_3 + 9\gamma_2 - 12}{6}$
c_4	$\frac{3(35\gamma_4 - 30\gamma_2 + 3)}{8}$	$\frac{\gamma_4 - 16\gamma_3 + 72\gamma_2 - 69}{24}$

Table 1: Coefficients as function of the signal moments for the approximation of a PDF with a Legendre and Laguerre polynomial base.

As follows, we compare the ability of the different polynomial bases to estimate $f(x)$, and thus the energy consumption. As a typical scenario, we analyze the transition activity t_i and bit probability p_i in signals with a Chi-square distribution ($\chi^2(n)$) of different degrees of freedom. Varying this parameter, a family of continuous PDFs with very different characteristics can be produced. Since the χ^2 is always positive, unsigned binary code is assumed for all the signals.

For each signal, and each polynomial family, the PDF is approximated using **Eq**. (4) and **Eq**. (8). The moments required for the estimation are calculated analytically to reduce numerical precision related errors. The so estimated parameters p_i and t_i are compared with the reference value obtained by direct numerical integration of $f(x)$. The experiment was conducted for different orders of approximation. **Fig**. 1 depicts the average error for the different distributions and its standard deviation. The results show that the Laguerre base is the one which overall performs better in terms of mean square error. Except for the simulation concerning the χ^2 with one degree of freedom, errors smaller than 0.02 can be obtained using just the first five moments of the signal. For the $\chi^2(1)$, the convergence is degraded because the PDF is not bound when $x = 0$.

For the Legendre and Laguerre bases, the quality of the approximation increases with the order of the approximation (see **Fig**. 1), which shows that they are stable. The opposite tendency is observed for the Hermite polynomials, that are very sensitive to the order of the approximation. If more than four moments are taken, and $f(x)$ differs significantly from a Gaussian distribution, the approximation gets notably deteriorated. It is worth noting that the accuracy of the Hermite base gets improved with the number of degrees of freedom. It is due to the fact that the χ^2 distributions tends to a Gaussian when the degree of freedom increases.

After this preliminary analysis, we concentrate our focus on the Laguerre and Legendre bases that seems to be the more promising ones.

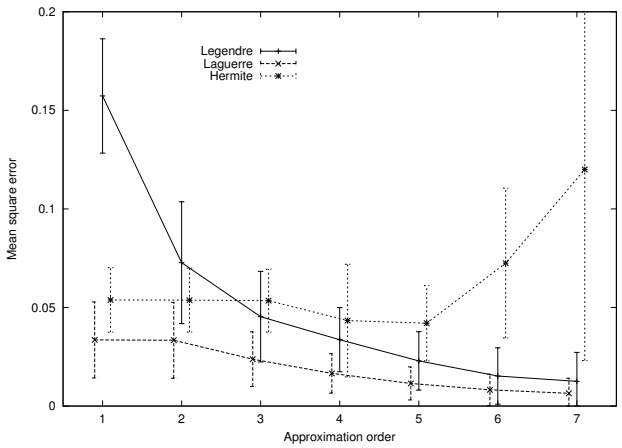

Figure 1: Mean squared error in the approximation of the transition activity. Vertical lines represent the standard deviation.

3. SYMBOLIC PROPAGATION

In order to obtain the statistical moments required by our technique, two main approaches can be employed: either by using word level simulators such as MATLAB and SystemC, or by means of

symbolic techniques. Simulation approaches are conceptually simpler and can be integrated in current industrial design flows; however they require a complete executable implementation of the system, and longer runtime executions than symbolic approaches. The main idea of the symbolic techniques is to propagate the statistical properties of the signals through the design. As follows, we describe a propagation approach for the statistical moments, focusing in different non-linear arithmetic operators (for linear operators, the propagation is simpler). Let $\gamma_{x,j}$ and $\gamma_{y,j}$ denote the jth raw moments of the inputs of a block, and $\gamma_{z,j}$ at its output, and let us denote by $\mathbf{E}[\cdot]$ the expectation or average operator. Then:
Square: Since $z = x^2$, it is straightforward that

$$\gamma_{z,j} = \mathbf{E}[x^{2j}] = \gamma_{x,(2j)}$$

Power: In general, for a positive integer power n, the output is defined as $z = x^n$. Thus, the moments at the output can be calculated as:

$$\gamma_{z,j} = \mathbf{E}[x^{nj}] = \gamma_{x,(n \cdot j)}$$

Multiplication: The raw moments at the output of the multiplier can be calculated using the input moments and the correlation between the inputs. In the particular case of spatially uncorrelated inputs, the expression gets reduced to:

$$\gamma_{z,j} = \mathbf{E}[(x \cdot y)^j] = \gamma_{x,j}\gamma_{y,j}$$

Absolute value: For a module calculating the absolute value of a signal, the moments at the output cannot be calculated using a finite number of moments at the input. The exact value can only be obtained for even orders as follows:

$$\gamma_{z,(2j)} = \mathbf{E}[|x|^{2j}] = \mathbf{E}[x^{2j}] = \gamma_{x,(2j)}$$

For smooth distributions, the odd moments can be approximated by using interpolation between two consecutive even values. It is worth noting that a direct polynomial interpolation of the raw moments cannot be used. The rationale behind is the different measuring units that each moment has. For example, $\gamma_{z,2}$ depends on the second power of z, while $\gamma_{z,4}$ does it on the fourth; and thus they cannot be combined. This problem can be overcome by normalizing the moments with the nth root function according to the moment degree. Thus,

$$\gamma_{z,(2j+1)} \approx \left(\frac{\sqrt[2j]{\gamma_{z,(2j)}} + \sqrt[2j+2]{\gamma_{z,(2j+2)}}}{2} \right)^{2j+1} \qquad (11)$$

For the raw moment of order one, the previous interpolation cannot be used. In this case, we use a linear extrapolation.

As an example of the accuracy of the technique, we compare the exact and estimated raw moments of the absolute value of a Gaussian signal. The results show an excellent approximation. The worst estimation correspond to the moment of order one where extrapolation is used. In this case, the exact value is $\sqrt{2/\pi} \approx 0.80$, whereas the approximation equals 0.84, which represents an adequate approximation.

4. EXPERIMENTAL RESULTS

In this section, the previously proposed estimation approach is experimentally assessed and compared with bit-level accurate estimations obtained by simulating long input traces (more than 2^12 samples).

First the bit probability in a set of four families of signals is analyzed. They are respectively:

G1: This family correspond to the absolute value of a set of seven Gaussian distributed signals with different standard deviations. In practice, this PDF appears very frequently when natural signals (such as music, voice, etc) are represented in sign-magnitude).

G2: The result of inverting the bits of G1.

G3: Seven Gaussian distributed signals with different standard deviations. This is the most common PDF in linear DSP systems; but it does not represent appropriately the signals in non-linear architectures.

G4: A family of χ^2 distributions with different degrees of freedom (seven values from 2 to 8). This PDF is typical for signals coming from the implementation of communication algorithms.

The sets G1, G2, and G3 concentrate the probability in the smaller, higher and middle values of the signal. Thus, they cover a complete spectrum of possibilities. The family G4 varies the shape of the PDF according to the degree of freedom. All of them are quantized with 9 bits.

The results of the experiment for estimating p_i are depicted in **Tab**. 2, where the accuracy of the proposed technique is compared with eq. 5 of [7]. In our approach, a decomposition in Legendre polynomials using seven moments is used for the estimation of the PDF (see **Eq**. (9)). The maximum and average estimation errors for each of the aforementioned family of signals are reported, together with the range of variation of the bit probability (i.e., the maximum and minimum values of p_i observed for the set of signals inside a given family). The results clearly show that the previous approach[1] [7] can only be used for signals with a PDF centered around zero (G1). If the signal deviates from this assumption, huge errors until 63% can be produced. Since the standard deviation of G1 and G3 is the same, but its p_i changes completely, similar errors are expected for any other technique which only takes the standard deviation into account. On the contrary, our technique provide very accurate results in all the scenarios with a maximum error smaller that 2%. An example of the bit accuracy of the technique is represented in **Fig**. 2. It refers to the family G4 with degree of freedom equal to 1 and 8 respectively. It highlights the excellent accuracy of the proposed moment-based technique (measurements and estimations are almost indistinguishable). Similar results are obtained for t_i and te_i. For example, **Tab**. 3 represents the maximum error induced by our approach for different approximation orders. It is observed that values of N between 5 and 7 produce adequate approximations.

	Value		% error **Eq**. (9)		% error from [7]	
Sig	Max	Min	Max	Avg	Max	Avg
G1	3.83	2.35	0.77	0.42	5.04	3.82
G2	6.66	5.16	1.27	0.28	63.00	39.00
G3	4.50	4.49	0.40	0.29	27.26	20.04
G4	4.42	3.06	0.23	0.20	12.70	7.15

Table 2: Estimation of total bit probability for different signals.

In a second experiment, the static power consumption in some circuits has been modeled and estimated. In order to prove the ability of the proposed technique to work in different scenarios, two different technology nodes, namely $0.35\mu m$ and $100nm$, have

[1]Observe, that it was developed for the magnitude of a Gaussian distributed signal; i.e., for the family G1.

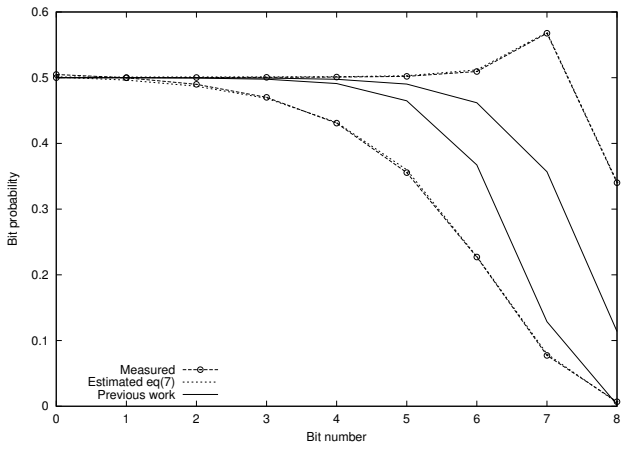

Figure 2: Measured (dots) and estimated (lines) bit probability from two chi-square distributed signals. Comparison between this work and previous approaches [8].

N	7	6	5	4	3	2
Max error (%)	1.32	1.91	2.53	5.82	6.70	32.5

Table 3: Worst case estimation error for the equivalent spatial transition activity (te_i).

been considered. (Observe that the leakage current in the $0.35\mu m$ node is negligible in comparison with the dynamic currents, and it needs only to be considered for designs with a high probability of being in stand-by mode. It is used only for the purpose of highlight the flexibility of the present estimation procedure).

The first analyzed technology corresponds to a commercial twin-well $0.35\mu m$ CMOS process, with transistors characterized for the BSIM3v3 model, level 49. The second one refers to an extrapolated $100nm$ technology with 17A gate oxide thickness [5], and characterized for the same transistor model as in the previous case . A major characteristic in this technology is that the gate-oxide tunneling leakage becomes comparable to the subthreshold leakage. The consequence is that the dependency of the power consumption with input state of the gates differs notably from previous technological nodes. In some cells (e.g., NAND gates) the effects of the input state increases, while in others the dependency almost vanishes (e.g. NOR gates) [5].

Four circuits have been analyzed. C1 and C2 correspond to a couple of comparators with the (constant) values 504 and 8 respectively. These values have been selected in such a way that the designs can be easily implemented with NAND3 and NOR3 gates. C3 is a parity checker, while C4 implements a fast incrementer. Since the last circuit is the most complex one, it has been divided in two parts (C4a and C4b) in order to present a more detailed investigation of the leakage current. Since the number of inputs on the circuits is relatively low (only 9), exhaustive simulations have been carried out to fill a table with the static power consumption associated with each input state. In order to speed up the simulations, first the single cells were simulated, and afterwards those values were used to evaluate the total static power consumption of the complete design. A simulator based on System-C was used for that purpose. For the transistor level simulations, Cadence's Spectre^{TM} was used.

Using an approach similar to **Eq**. (9), but replacing the function $Bit_i(x)$ by the aforementioned table, a template has been obtained for estimating the static power consumption with the moments of the signal. In particular, we employed a decomposition in Legendre polynomials with order 7. It is worth noting that in a real scenario a more efficient characterization process can be carried out. Instead of evaluating the leakage for all the input states, we can restrict the analysis to the N random signals $q_j(x)\phi(x)$ (with N is the order of the approximation). Moreover, that analysis can be accelerated with Monte-Carlo techniques, allowing therefore an efficient characterization of the static power consumption even for large circuits.

The results of the experiment are reported in **Tab**. 4. The entry "A" in the column labeled "Tech" states for the $100nm$ technology, while "B" refers to the $0.35\mu m$ process. Since the characterization of some cells was not available for the first case, the circuits C3, C4a, and C4b have only been estimated for one technology. In order to allow a comparison of results, the power values have been normalized to the average.

	Tech	Sig	Value		% error this work		% error previous	
			Max	Min	Max	Avg	Max	Avg
C1	A	G1	0.80	0.58	3.27	0.80	71.51	38.58
		G2	1.98	1.23	0.35	0.16	49.61	32.40
		G3	0.90	0.88	0.28	0.24	13.25	12.16
		G4	0.88	0.69	0.63	0.30	42.92	24.43
	B	G1	0.75	0.47	3.11	0.72	110.10	54.36
		G2	2.51	1.29	0.90	0.38	60.21	39.07
		G3	0.87	0.85	0.61	0.32	17.32	16.42
		G4	0.86	0.62	0.65	0.31	60.91	31.96
C2	A	G1	1.21	1.03	3.79	1.16	17.36	7.97
		G2	0.97	0.97	2.21	0.52	2.67	2.44
		G3	0.97	0.97	0.18	0.14	2.34	2.28
		G4	1.11	0.97	2.38	0.53	9.93	3.15
	B	G1	1.15	1.09	1.30	0.88	13.18	11.83
		G2	0.83	0.53	3.83	1.10	87.27	43.67
		G3	1.13	1.07	0.94	0.28	11.57	9.16
		G4	1.15	1.10	1.34	0.29	13.79	11.99
C3	B	G1	1.00	0.99	2.32	0.87	0.63	0.47
		G2	1.00	0.99	2.14	0.52	0.63	0.48
		G3	0.99	0.99	0.23	0.21	0.00	0.00
		G4	1.00	1.00	0.71	0.27	0.79	0.56
C4a	B	G1	0.88	0.72	3.49	1.45	37.18	21.96
		G2	1.69	1.13	2.93	0.66	41.04	23.64
		G3	1.00	0.96	0.67	0.51	3.62	2.24
		G4	0.96	0.80	0.45	0.38	23.54	11.99
C4b	B	G1	1.19	1.06	5.88	2.05	16.45	10.68
		G2	0.92	0.80	5.30	1.38	24.82	14.17
		G3	1.00	0.99	0.34	0.28	0.05	0.02
		G4	1.14	1.00	2.65	0.70	12.30	5.83

Table 4: Estimation of static (leakage) power for different circuit and excitations.

The accuracy of the proposed moment-based approach has been experimentally determined and compared with a naive technique that assumes a constant value for the leakage current equal to its average. The results (see **Tab**. 4) show that the accuracy achieved for our approach is in average better than 2%, with a worst case smaller than 4%. On the other hand, if a constant value of leakage is assumed, errors of more than 100% can be observed. Furthermore, it can be noticed that for some circuits (e.g., C1) there is

a notable variation of the static power with the statistical characteristics of the signal; while for others (e.g., C3), that variation is almost negligible. In addition, those effects are more significant in the 100*nm* technology, which implies that importance of an statistical characterization of the power consumption is increasing as technology improves. This observation agrees with similar conclusions in timing analysis for VDSM, where statistical methodologies are getting mandatory.

Finally, to show the importance of modeling the static power consumption, four functional-equivalent variations of C1 where created. The logical implementation of the design was adapted to provide an implementation which is optimal for a particular family of signals (i.e., G1 to G4). The modules have been then characterized, and their power consumption has been measured and estimated as in the previous experiment. The obtained results show a significant influence of the power consumption with module assignment. As a representative example, **Fig**. 3 depicts the normalized static consumption when each of this four designs is excited with a signal from the group G1. Beside the high accuracy of our estimation technique, it is observed that a sub-optimal selection of the module (for example the one optimized for G2) can increases the consumption by more than 50%, from 0.4 to 0.6. Further on, a simplistic assumption of a constant average consumption can overestimate the power by 150% (i.e., 1 instead of 0.4). In other scenarios, as for signals from G2, there is an underestimation rather than a overestimation of around 34%.

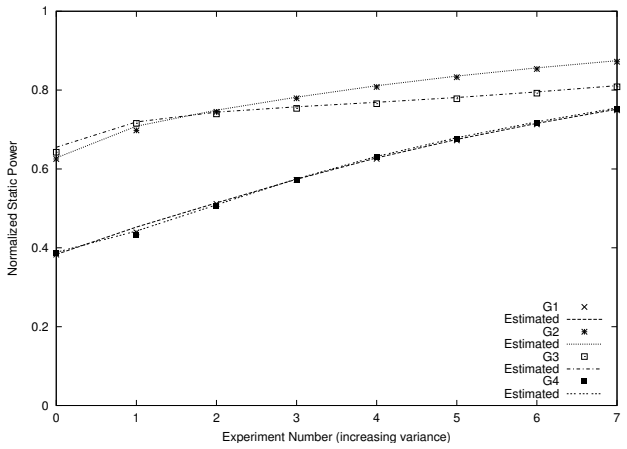

Figure 3: Measured (dots) and estimated (lines) normalized static power for different circuits excited by 8 random signals from the family G1.

5. CONCLUSIONS

In this paper, an approach has been presented to estimate the power consumption in VDSM technologies. It allows the estimation of different power metrics and VDSM effects (such as the switching activity, influence of coupling capacitances, and leakage current) within a unified framework. Implicitly, the technique has focused on data intensive applications where the internal signals can be characterized by a smooth probability density function. Since it is based on a rigorous mathematical foundation (the decomposition in an orthogonal polynomial base), it provides an excellent accuracy for a wide range of design scenarios. Moreover, the number of terms in the approximation procedure can be adjusted to provide a trade off between model complexity and accuracy. Experimental results show that appropriate values of this parameter lie between 5 and 7. Additionally, a technique has been presented to propagate the moments of the signal through the design. Thus, it avoids the need of expensive simulations (which represent a major obstacle for fast energy estimations). The combination of these two techniques represents an important step toward a mature approach for estimating the energy consumption at high levels of abstraction in data intensive applications. The current formulation restricts the estimation of t_i and te_i for the case of temporal un-correlation. Ongoing efforts to suppress that limitation are under development with very promising results.

Extensive transistor and bit-level simulations results have been reported to assess the accuracy of the proposed moment-based approach. Typical average errors of less than 5% have been measured, in contrast with previous techniques which exhibit errors of more than 100%.

6. REFERENCES

[1] S. Bobba, I. N. Hajj, and N. R. Shanbhag. Analytical expressions for average bit statistics of signal lines in DSP architectures. In *IEEE International Symposium on Circuits and Systems (ISCAS)* , pages 33 – 36, 1998.

[2] J.-M. Chang and M. Pedram. *Power optimization and synthesis at behavioral and system levels using formal methods*. Kluwer Academic Publishers, 1999.

[3] D. Eckerbert and P. Larsson-Edefors. Cycle-true leakage current modeling for CMOS gates. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 5, pages 507–510, 2001.

[4] A. García, L. D. Kabulepa, and M. Glesner. Transition activity estimation for generic data distributions. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2002.

[5] W. Kwong, D. Lee, D. Blaauw, and D. Sylvester. Analysis of simultaneous subthreshold and gate-oxide tunneling leakage current in nanometer CMOS design. In *ISQED*, Mar. 2003.

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

[7] M. Lundberg, K. Muhammad, K. Roy, and S. K. Wilson. A novel approach to high-level switching activity modeling with applications to low-power DSP system synthesis. *IEEE Trans. on Signal Processing*, 49:3157 – 3167, December 2001.

[8] C.-G. Lyuh, T. Kim, and K.-W. Kim. Coupling-aware high-level interconnect synthesis for low power. In *IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, pages 609 –613, 2002.

[9] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj. Analytical estimation of signal transition activity from word level statistics. *IEEE Trans. on CAD*, pages 718 – 733, July 1997.

[10] J. H. Satyanarayana and K. K. Parhi. Theoretical analysis of word-level switching activity in the presence of glitching and correlation. *IEEE Trans. on VLSI Systems*, pages 148 – 159, April 2000.

[11] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and applied probability. Chapman & Hall, 1986.

[12] D. Sylvester and C. Hu. Analytical modeling and characterization of deep-submicrometer interconnect. *Proc. of the IEEE*, 89(5):634–664, May 2001.