High-level Modeling of Switching Activity With Application to Low-power DSP System Synthesis

Magnus Lundberg*, Khurram Muhammad†, Kaushik Roy† and Sarah Kate Wilson*

Email: mlg@sm.luth.se, khurram@ecn.purdue.edu, kaushik@ecn.purdue.edu, kate@sm.luth.se

**Division of Signal Processing, Luleå University of Technology, S-971 87 Luleå, SWEDEN*

†*School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA*

Abstract— **We address the issue of high-level synthesis of low-power *digital signal processing* (DSP) systems by proposing switching activity models. In particular, we present a technology independent hierarchical scheme to compare relative power performance of two competing DSP systems. The basic building blocks considered for such system are a full-adder and a one-bit delay. Estimates of switching activity at the output of these building blocks is used to model the activity in different architectural primitives used for building DSP systems. This method is very fast and simple and simulations show accuracy within 4% of extensive bit-level simulations. Therefore, it can easily be integrated into current communications/DSP CAD tools for low-power applications. The models show that the choice of multiplier/multiplicand is important when using array multipliers in a data-path. If the input signal with smaller variance is chosen as the as the multiplicand, up to 20% savings in switching activity can be achieved. This observation is verified by analog simulation.**

I. Introduction

The development of low-power mobile radio and computing systems faces tremendous challenges as increased services, faster data rates and higher processing speeds continue to dominate future trends. This motivates development of low-power synthesis tools for such applications. One widely used approach for power reduction is *supply voltage scaling* [1]. However, there are physical limitations to reduction of power using such an approach and it is useful to investigate alternative methodologies which are device independent. Several such approaches have been identified varying from architectural level investigations to device level research [1], [2]. The architectural level approaches exploit parallelism or pipelining in the algorithm and increase the throughput by employing extra hardware [1], [2]. The processing speed is then reduced in a second step by reducing the voltage supply. Other methods proposed for reducing power dissipation attempt to reduce power by substructure sharing and architectural transformations [3].

It is well known that the major contribution to the total power dissipation in present-day technology is caused by internal switching activity [1] (*dynamic power*). Large savings in switching activity is possible if this is one of the design concerns at a higher level of abstraction in DSP/Communication CAD tools. Low-power application tools require methods which can predict how transformations at architectural level affect power dissipation. Hence, high-level estimation of switching activity in signals is considered in [4], [5]. [4] focuses on estimation of switching activity in signals, whereas, [5] addresses analytical estimation of signal transition activity. We observe that if a library of analytical models for primitives (such as delay, adder and multiplier) can be constructed, switching activity estimation in a variety of DSP architectures can be addressed using a hierarchical approach. Since the total switching activity in DSP architectures is dominated by arithmetic units, low-power synthesis can be addressed in a simple manner.

In this paper, we present a novel and fast method of estimating relative power performance of data-paths in DSP applications implemented in CMOS. Our objective is to provide a simple and efficient scheme which can compute relative power dissipation in two competing architectures with good accuracy. For this purpose, we explore approaches which can be used for higher level synthesis where transistor level knowledge is not available. For reasons outlined earlier, we restrict our attention to dynamic power only. Since, larger switching activity leads to higher power dissipation, we first identify parameters which contribute directly to switching activity, and then, we outline approaches which can be used to compare two competing architectures in terms of switching activity (and hence power). In addition to being technology independent, a major advantage of higher level modeling and power estimation is the small computational requirement since transistor level simulations are eliminated.

This paper is organized in five sections. Section II introduces the basic approach used in our work and defines basic building blocks and primitives for general DSP systems. Section III presents switching activity models for these primitives based on signal statistics. Some numerical results are presented in section IV. Section V presents the effect of reordering inputs of a multiplier on the power dissipation and demonstrates the viability of our approach for high-level synthesis applications. Finally, section VI concludes this paper.

II. Basic Approach Towards Modeling

We restrict our attention to high-speed data-paths using parallel implementations. Further, for simplicity, signals at different points in the system are assumed stationary. As pointed out earlier, DSP architectures are primarily data-paths which are constructed using architectural primitives such as adders, multipliers and delays. The common building block of the first two primitives is a full-adder, whereas, a delay element is constructed using single-bit delays. Therefore, the switching activity in these architectural primitives can be estimated using simulation and modeled as a function

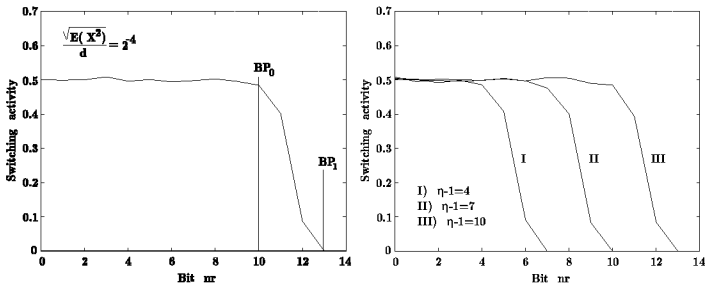

Fig. 1. Switching activity of bits of a zero-mean white Gaussian process expressed with 16 bit SM form. Sign bit is not shown.

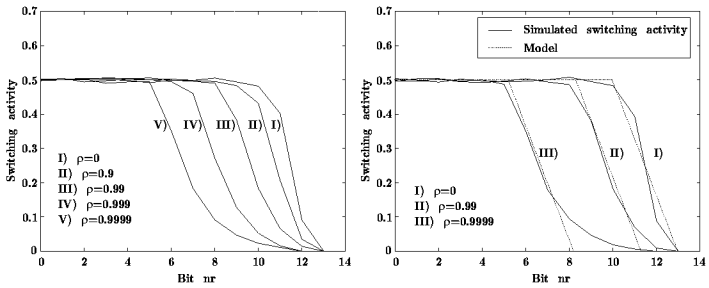

Fig. 2. Switching activity vs. different time correlations. Figure on right shows linear approximation model.

of the statistics of the signals applied at their input. Such a model can be easily constructed if the input signals are assumed stationary. Note that in fully parallel implementations, switching activity depends on present and previous states only. Hence, in this frame-work, stationarity assumption is quite valid. In contrast, a DSP processor based implementation shares computing resources between different signals, and hence, stationarity assumption is not valid.

Since our objective is to provide high-level models of power dissipation without specific knowledge of the transistor based implementation, we only consider switching at the outputs of the building blocks (full-adder and one-bit delay). As we ignore the internal switching activity, this method cannot be used for accurate power estimation. However, when comparing two systems for lower switching (and hence, lower power) the method is fast and reasonably accurate as verified in the implementation considered in section IV. Thus, it can be used to detect the effects of transformations on the power dissipation in the application level. We note that depending on implementation, an output switching of a full-adder may not consume the same power as an output switching of a one-bit delay. Consequently, we will distinguish between the two types of switching activities in our notation. The output switching of a full-adder will be denoted by a_{FA} whereas the output switching of a delay element will be represented by a_D. Appropriate weighting of the two may be done when comparing the total power dissipation.

III. Models For Architectural Primitives

In the sequel, the input of a single input primitive will be denoted by $X[n]$. Similarly, inputs of two-input primitives will be denoted by $A[n]$ and $B[n]$. For simplicity, the inputs of multiplier and adder will be assumed to be spatially uncorrelated. Finally, all signals will be assumed to be uniformly quantized in a dynamic range of $\pm d$ and represented in *sign magnitude* (SM) form using N bits.

A. Delay

A delay element consists of N one-bit delays. Since the information about signal statistics apply only to its word-level representation, whereas switching activity depends on bit-level transitions, we need a model which expresses switching of every bit given the word-level statistics. One such model is presented in [4] which models the effect of input statistics on the *most significant bits* (MSBs) and *least significant bits* (LSBs). As proposed in [4], bits in signal word are divided in to three regions; first, where switching is low (MSBs), second, where switching is high (LSBs), and, last, the region in between which is considered to be a linear transition connecting the other two. This is shown in figure 1 for SM representation. The switching of the lower ordered bits is 0.5, whereas, switching of higher order bits is 0. Two breakpoints, BP_0 and BP_1, mark the end of the first and beginning of the third region, respectively. Next, We define the number of bits needed to represent the average signal power as $\eta = \log_2\left((2^{N-1}-1)\cdot\frac{\sqrt{E(X^2[n])}}{d}+1\right)$, where N is the actual number of bits used to represent the signal value. Observe that $\sqrt{E(X^2[n])}$ equals the standard deviation of zero-mean signals. Since SM representation is used and stationarity is assumed, η may be used to find BP_0 and BP_1 from the word-level statistics.

Extensive simulation results shows that for uncorrelated signals, $BP_0 \approx \eta - 1$ and $BP_1 \approx \eta + 2$. Switching activity in each bit for uncorrelated signals is shown in figure 1 for various values of average power, $E(X^2[n])$. Next, for correlated signals, we only need to consider time correlation between successive samples, i.e $\rho_X = Cov(X[n]x[n-1])/Var(X[n])$ since switching depends only on the last signal value.

The effect of signal correlation on switching activity is shown in Figure 2. Clearly, switching activity decreases as correlation between successive samples increases. For high correlation, the transition between the LSB and MSB starts off linear close to the LSB region and becomes non-linear as it reaches the MSB region. However, in the vicinity of MSBs switching becomes very low and can be ignored for all practical purposes. Further, the slope of the linear region is constant irrespective of the value of correlation. A non-linear curve approximation gives $BP_0 = \eta - 2.1\cdot(1-\rho)^{-0.1293}+1.1$ and $BP_1 = \eta - 2.1\cdot(1-\rho)^{-0.1293}+4.1$.

Figure 2 compares these relations with results obtained using simulation. We observe that our model compares very well with simulation results. We now turn our attention to the switching activity in the sign bit. Since sign-bit switching is same for both SM and 2's complement representations, we can use the estimate in [5] to get $a_{N-1} = 2p'(1-p')(1-\rho)$ where p' is the probability that the sign-bit is equal to one (negative valued sample). The switching activity of bit i can be expressed as

$$a_D^i = \begin{cases} 0.5 & (i < BP_0) \\ 0.5 - \frac{0.5}{3}\cdot(i - BP_0) & (BP_0 \le i \le BP_1) \\ 0 & (i > BP_1) \\ 2p'(1-p')(1-\rho) & (i = N-1) \end{cases} \quad (1)$$

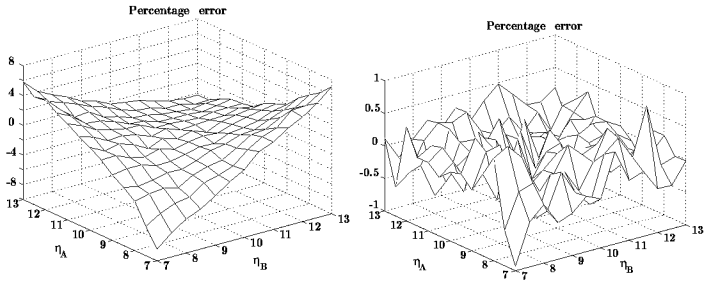

Fig. 3. Percent error in switching activity in an array multiplier. Linear model on left, quadratic model on right.

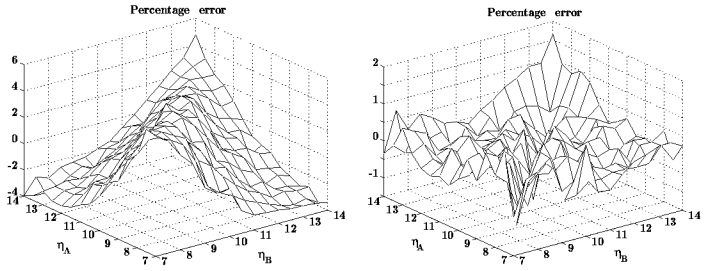

Fig. 4. Percent error in switching activity in an adder. Linear model on left, quadratic model on right.

B. 16-bit Array-Multiplier

To investigate which statistical parameters affect switching activity in an array multiplier we use the *classical factorial design* [6]. In brief, this technique identifies parameters that affect a quantity in question. One can identify main factors (sole effect of the parameters) and interaction factors (effect caused by interaction of parameters).

With the assumption that cross-correlation does not effect switching activity, the statistical parameters that may be significant are $\sqrt{E(A^2[n])}/d$, $\sqrt{E(B^2[n])}/d$, ρ_A, and ρ_B. Using the factorial design, we conclude that correlation below 0.5 does not contribute significantly to the switching activity in the array-multiplier. A full derivation and a detailed description of the technique can be found in [7]. In conclusion, correlation greater than 0.5 has negligible effect on switching activity, and hence, the model of a signed magnitude array multiplier only needs to consider the relative input powers (η_A and η_B) at the two inputs. A plot of the switching activity in a 16-bit multiplier shows that it is approximately linear function of η_A and η_B. *Least squares* (LS) linear approximation yields

$$a_{FA} = -109 + 14.8\eta_A + 9.9\eta_B \tag{2}$$

The error of the linear approximation is shown in figure 3a. It is noticed that linear approximation does not accurately model the surface, hence, considering the quadratic dependence of the error on the parameters η_A and η_B gives

$$a_{FA} = -56.15{+}7.80\eta_A{+}6.87\eta_B{-}0.035\eta_A^2{-}0.23\eta_B^2{+}0.76\eta_A\eta_B \tag{3}$$

Figure 3b shows that quadratic approximation gives a good model and yields random error within one percent of the simulations. However, linear approximation shows an interesting phenomenon; the average power of the inputs does not affect switching equally. To achieve lower switching activity, signal with higher average power should be selected as the multiplier whereas, the signal with lower average power should be selected as the multiplicand. In section V, we will show that savings of switching activity of up to 20% can be achieved by an appropriate choice of inputs in a multiplier.

C. 16-bit Adder

The adder for SM representation can be implemented either by conversion to 2's complement representation, or, by using an adder and a subtracter. For simplicity, we consider the second structure and assume that only one of the two units will be active during an operation.

Again, using the factorial design [6], we conclude that almost every effect, including interaction effects, is significant. It is, therefore, necessary to include the signal correlation of the inputs to obtain a good model for the switching activity in an adder. Recall the observation made in section III-A that switching activity due to a correlated signal resembles the switching activity of a uncorrelated signal with lower input power. Hence, the method used for modeling a multiplier can also be used for an adder. Consequently, we first model switching activity in an adder for uncorrelated signals, and then compensate for signal correlation by appropriately decreasing the input signal power.

A plot of switching activity in an adder shows that the switching activity in an adder is very nearly linearly dependent on η_A and η_B. The LS linear approximation yields $a_{FA} = 1.59 + 0.49\eta_A + 0.49\eta_B$. Figure 4a shows the error due to this approximation. As observed in the case of multiplier, there are non-linear dependencies. A close observation of figure 4a shows that the error is symmetric around the diagonal which represents equal input powers. The error seems to depend on the distance between η_A and η_B. An LS approximation with linear terms of η_A and η_B and $|\eta_A - \eta_B|$ and quadratic term of $|\eta_A - \eta_B|$ yields

$$a_{FA} = 1.14{+}0.49\eta_A{+}0.49\eta_B{+}0.25|\eta_A{-}\eta_B|{-}0.017(|\eta_A{-}\eta_B|)^2 \tag{4}$$

The error is shown in figure 4b. Though slightly more complex, this model reduces error significantly, thereby providing much better model for the switching activity.

Note that the above models are valid only for uncorrelated signals. To compensate for correlation, signal power is decreased appropriately. In section III-A, we showed that a good model for compensated input power is $\eta_X^{comp} = \eta_X - 2.1(1-\rho_X)^{-0.1293} + 2.1$. Using the above expression to compute η_A^{comp} and η_B^{comp}, we obtain the linear and quadratic models as $a_{FA}^{linear} = 1.59 + 0.49\eta_A^{comp} + 0.49\eta_B^{comp}$ and $a_{FA} = 1.14 + 0.49\eta_A^{comp} + 0.49\eta_B^{comp} + 0.25|\eta_A^{comp} - \eta_B^{comp}| - 0.017(|\eta_A^{comp} - \eta_B^{comp}|)^2$, respectively.

IV. Numerical Results

We consider an adaptive filter with filter coefficients $c_k[t]$ used for equalizing a channel with impulse response $h[k] = \sqrt{2/9} \cdot [1 + cos2\pi(k-2)/3], k = 1, 2, 3$. $h[k] = 0$ for all other values of k. We will assume that adaptation is done using the *least mean square* (LMS) algorithm. A random sequence

	7 Taps, l=3		15 Taps, l=7	
σ^2	0.01	0.1	0.01	0.1
Switching Estimates For The Delay				
Simulation	1.76×10^4	1.77×10^4	4.11×10^4	4.14×10^4
Model	1.73×10^4	1.74×10^4	4.04×10^4	4.06×10^4
Error	-1.50%	-1.81%	-1.79%	-1.84%
Switching Estimates For The Multiplier				
Simulation	8.05×10^5	8.68×10^5	1.63×10^6	1.82×10^6
Linear model	8.24×10^5	8.87×10^5	1.68×10^6	1.86×10^6
Error	2.31%	2.20%	2.85%	2.34%
Non-lin. model	7.98×10^5	8.68×10^5	1.61×10^6	1.82×10^6
Error	-0.93%	0.02%	-1.16%	0.24%
Switching Estimates For The Adder				
Simulation	4.92×10^4	5.02×10^4	9.64×10^4	1.04×10^5
Linear model	4.95×10^4	4.93×10^4	9.40×10^4	1.01×10^5
Error	0.61%	-1.71%	-2.43%	-2.89%
Non-lin. model	5.08×10^4	5.07×10^4	9.60×10^4	1.03×10^5
Error	3.28%	0.91%	-0.42%	-0.81%

TABLE I

PERFORMANCE OF PROPOSED MODELS FOR BASIC BUILDING BLOCKS.

of length 400 consisting of symbols taken from a 2-PAM constellation ($I[t] = \{\pm 1\}$) is applied to the channel input and *additive white Gaussian noise* (AWGN) with variance σ^2 is added to the channel output. The signals are represented using 16-bits SM representation and the dynamic range is set to ± 4 ($d = 4$). The output, v[t], of the $2l + 1$ tap linear equalizer is given as $v[t] = \sum_{k=-l}^{l} c_k[t] \cdot r[t-k]$. The LMS update equation for updating the filter coefficients is given as $c_k[t+1] = c_k[t] + \mu \cdot (I[t] - v[t]) \cdot r[t] = c_k[t] + \mu \cdot e[t] \cdot r[t]$ where μ is positive number chosen small enough to ensure convergence (We assume $\mu = 0.05$.). e[t] is the error after equalization. For simplicity, $I[t]$ is used during the simulations to avoid error propagation.

To estimate the switching activity, we first need to compute $E(r^2[t])$, $\rho_{r[t]}$, $E(e^2[t])$, $\rho_{e[t]}$, $E(c_k^2[t])$ and $\rho_{c_k[t]}$, for $t = 1, \cdots, 400$ and $k = -l, \cdots, l$. The first five of these can easily be obtained using any communications CAD tool which evaluates the *bit error rate* (BER) performance of such a system. The correlation between successive values of filter coefficients can be assumed to be 0.99 during the adaptation face and 0.9999 when the filter converges. We then use the models developed in section III to estimate the switching activity of the equalizer. Table I compares our models with the results obtained using extensive bit-level simulation. Clearly, higher level models yield good estimates of output switching activity. Higher level modeling yields a simple and fast approach which can be easily incorporated in current CAD tools for communication and DSP applications. This is because the statistical parameters of signals can be easily calculated at different points in a system when calculating BER performance. Switching activity estimates can then be obtained using our models. Thus, the relative power performance of competing designs can be compared when low-power is an important concern.

V. Reordering of Multiplier Inputs

Recall equation 2 which shows that the switching activity in an array-multiplier depends on our choice of multiplier and multiplicand inputs. This observation is further verified in table II which shows the relative power savings obtained for the two choices of inputs. These results were obtained by laying out a 16 bit array multiplier in a 0.6μ process operating with 3.3V supply at $25°C$. Analog simulations were used

d	$\rho_A = 0, \rho_B = 0$	$\rho_A = 0.95, \rho_B = 0.95$
3	15.76%	17.72%
4	12.08%	14.15%

TABLE II

RELATIVE POWER SAVINGS FOR THE TWO CHOICES OF INPUTS IN THE MULTIPLIER. $E(A^2[n]) = 1$ AND $E(B^2[n]) = 0.001$.

	7 Taps, l=3		15 Taps, l=7	
σ^2	0.01	0.1	0.01	0.1
Good choice	8.51×10^5	9.27×10^5	1.71×10^6	1.90×10^6
Bad choice	9.93×10^5	1.06×10^6	2.03×10^6	2.20×10^6
Gain	14.3%	12.5%	16.1%	13.3%

TABLE III

DIFFERENCE IN FULL-ADDER SWITCHING ACTIVITY IN AN ADAPTIVE LINEAR EQUALIZER FOR THE CHOICES OF INPUTS TO THE MULTIPLIER.

to compute the total power dissipation using a run of 256 input vectors. If the input signal statistics are known a priori (e.g. computed using a communications/DSP CAD tool), it is evident that proper selection of multiplier and multiplicand can have significant effect in reducing the power, especially in architectures which implement multiplication dominant algorithms. This observation is further elaborated in table III which shows the difference in full-adder switching activity for the example in section IV for the two choices of the multiplier and multiplicand input signals. Clearly, gains of up to 20% are possible just by making a good choice of inputs.

VI. Conclusion

We present a hierarchical scheme to compare relative power performance of two competing i DSP systems for high-level synthesis. Using one-bit delay and a full-adder as the basic building blocks, we construct models for primitives such as N-bit delay, adder and multiplier. This method is simple, fast and efficient and can be easily integrated in to current communication/signal processing CAD tools and specifically targets efficient higher level synthesis. We demonstrate this by showing that the switching in a SM multiplier can be reduced by 20% if input with the smaller variance is chosen as the multiplicand. This gain is obtained without any overhead if signal statistics are known a priori. Further, we verify this observation by analog simulation of a layout of $16x16$ array multiplier implemented in a 0.6μ process at $25°C$ with supply voltage of 3.3V.

References

[1] J. M. Rabaey, "Digital Integrated Circuits: A Design Perspective," Prentice Hall, New Jersey, 1996.

[2] A. P. Chandrakasan, S. Sheng, R. W. Brodersen, W. Robert, "Low-power CMOS digital design," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 4, pp. 473–484, Apr 1992.

[3] D. A. Parker and K. K. Parhi, "Low-Area/Power Parallel FIR Digital Filter Implementations," *Journal of VLSI Signal Process. Syst. Signal Image Video Technol. (Netherlands)*, Vol. 17, No. 1, pp. 75–92, Sept. 1997.

[4] Landman, L.L. and Rabaey, J.M., "Power Estimation for High Level Synthesis," *In Proc. European Conference on Design Automation with the European Event in ASIC Design (EDAC EUROASIC '93)*, pp. 361-367, Feb. 22-25, 1993.

[5] Ramprasad, S., Shanbhag, N.R. and Hajj, I.N., "Analytical Estimation Of Transition Activity From Word-Level Signal Statistics," *In Proc. 34:th Design Automation Conference (DAC '97)*, pp. 582-587, June 9-13, 1997.

[6] Box, E.P., Hunter,W.G., and Hunter, J.S. "Statistics for experimenters," Jon Wiley and Sons, Inc. 1978.

[7] Lundberg, M., "Fast Power Estimation Methods," Master Thesis, Luleå University of Technology, May 1998.