Modeling and Minimization of Interconnect Energy Dissipation in Nanometer Technologies *

Clark N. Taylor, Sujit Dey, and Yi Zhao
Department of Electrical and Computer Engineering
University of California, San Diego
{cntaylor, dey, yizhao}@ece.ucsd.edu

ABSTRACT

As the technology sizes of semiconductor devices continue to decrease, the effect of nanometer technologies on interconnects, such as crosstalk glitches and timing variations, become more significant. In this paper, we study the effect of nanometer technologies on energy dissipation in interconnects. We propose a new power estimation technique which considers DSM effects, resulting in significantly more accurate energy dissipation estimates than transition-count based methods for on-chip interconnects. We also introduce an energy minimization technique which attempts to minimize large voltage swings across the cross-coupling capacitances between interconnects. Even though the number of transitions may increase, our method yields a decrease in power consumption of up to 50%.

1. INTRODUCTION

Currently, the semiconductor industry is continuing to develop and implement smaller technology sizes, enabling a host of new and more powerful applications. However, as technology sizes continue to decrease, many new effects are being observed due to the use of nanometer technologies. Some significant deep sub-micron (DSM) effects are caused by increasing cross-coupling capacitance and inductance, the effects being most significant in global interconnects (such as buses connecting the components in a system-on-chip) [1].

Recent studies on the effects caused by nanometer technologies focus primarily on timing [2] and noise/signal integrity [3]. In this paper, we study energy dissipation in nanometer technology interconnects. Our results show that not only is there a significant increase in the energy consumed in DSM technology interconnects, but also the energy consumed is very dependent on the nature of the signal transitions.

Current techniques for estimating the energy consumption of interconnects typically rely on the number of transitions to compute average power[4]. However, because energy dissipation due to cross-coupling capacitances can vary depending on the type of transitions that are transmitted across the interconnects, traditional power estimation techniques do not accurately predict the energy dissipation of nanometer technology interconnects. To more accurately estimate the power consumption in long interconnects, we propose a method which, in addition to the number of transitions, takes into account the *type* of transitions which occur.

Moreover, recent methods that have been proposed to reduce energy dissipation in on-chip interconnects focus on minimizing the number of transitions on the interconnects [5, 6]. However, for interconnects in nanometer technologies, techniques need to be developed which take into account DSM effects, not just the number of transitions. In this paper, we propose an interconnect energy optimization technique which distinguishes between the types of transitions, minimizing the ones which lead to a larger voltage swing across the cross-coupling capacitances. We have tested the proposed technique with several different types of data transmissions across buses, and have found that the method can significantly reduce energy dissipation in the interconnects.

In [7], a power estimation and minimization approach which considers DSM effects has been proposed. However, the power estimation technique in [7] assumes lumped capacitances and instantaneous driver switching, while our method considers the distributed nature of cross-coupling capacitance, intrinsic capacitance, and resistance in the interconnects, along with the rise and fall times of the interconnect drivers. In addition, our power minimization technique differs from the technique in [7] because it has simpler encoding hardware, and can help reduce timing problems due to DSM effects in long interconnects.

The paper is organized as follows: section 2 presents results demonstrating the importance of considering DSM effects in estimating energy dissipation. Section 3 presents a new power estimation model which takes into account DSM effects. Section 4 presents our methodology for minimizing the energy consumption of interconnects in nanometer technologies. Section 5 concludes the paper.

2. EFFECT OF NANOMETER TECHNOLOGIES ON ENERGY DISSIPATION

To ascertain the effect of nanometer technologies on the energy consumption of interconnects, we conducted experiments using a 100 nm, 1.2V, 5-interconnect bus system running at 1 GHz. The technology parameters used are from the International Technology Roadmap for Semiconductors[8]. Under the assumption that all input transitions are launched simultaneously and have the same slope (100 ps), we simulated a transmission-line model of the bus interconnects using HSPICE [9]. Figure 1 plots the power consumed by the interconnects, varying in length. The plot labeled "without crosstalk" does not take into consideration the coupling capacitances and inductances due to the use of nanometer technology, while the plot labeled "with crosstalk" does. The divergence between the two plots, and hence the power consumed due to crosstalk effects, increases significantly as the length of the interconnect increases. For example, at 1mm, the power consumption due to crosstalk effects is .19mW, while 1.23 mW is consumed due to crosstalk at 5mm. These results clearly show the need to take into account DSM crosstalk effects for accurate power estimation, especially in long interconnects such as busses and inter-core communication structures in a system-on-chip.

Because DSM crosstalk effects occur between the interconnects, as opposed to between an interconnect and ground or V_{dd}, the energy consumed in nanometer technology interconnects depends on more than just the number of transitions. Figure 2 shows how different types of transitions can have a profound effect on the amount of energy dissipated. Judging only by the number of transitions, a RFRFR bus event should consume 1.2 times the power of a RR0RR bus event, where R is a rising transition, F is a falling transition, and 0 indicates no transition. However, as shown in Figure 2, an RFRFR transition consumes 6.7 times the energy of an RR0RR transition for a set of 5mm interconnects. The difference in energy

* This work is supported by SRC under contract #98-TJ-648, and NSF under grant #CCR-9912414

DAC 2001, June 18-22, 2001, Las Vegas, Nevada, USA.

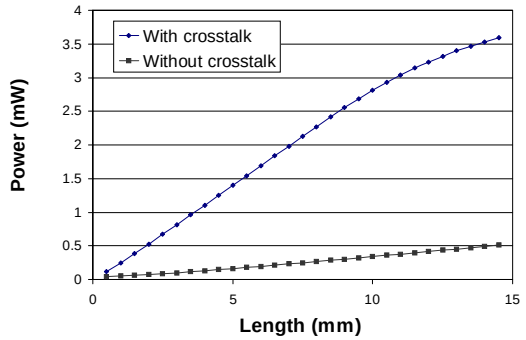

Figure 1: Result of DSM crosstalk effects on energy dissipation in interconnects

consumption between different types of transitions increases as the interconnect length increases.

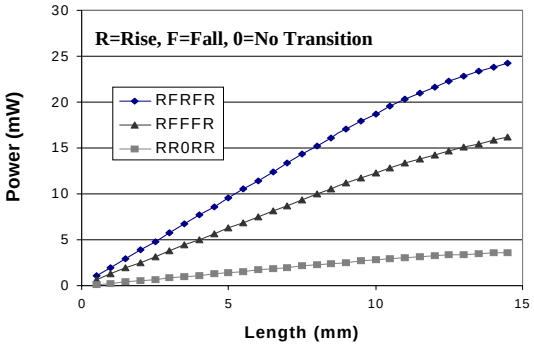

Figure 2: The effect of different types of transitions on energy consumption in interconnects

Our experiments demonstrate that the energy dissipation of interconnects in nanometer technologies depends on not only the number of transitions, but also the types of transitions. However, current power estimation techniques typically analyze only the number of transitions. While the effect of crosstalk on energy dissipation is not significant (and can possibly be ignored) for short interconnects, the DSM effects are significant and have to be considered for long interconnects, such as global interconnects in a system-on-chip. In the next section, we propose a method which analyzes both the number and type of transitions to obtain a more accurate estimate of energy dissipation.

3. A NEW DSM-AWARE ENERGY DISSIPATION MODEL

Since DSM effects can significantly impact the energy consumed in long interconnects, it is necessary to model the effect of nanometer technologies on interconnect energy dissipation at a high level so that designers can be aware of the energy consumption of their designs. In this section, we propose a DSM-aware power estimation model for interconnects in nanometer technologies. We then present results which demonstrate the improved estimation accuracy of our method over a traditional power estimator.

3.1 Interconnect power estimation model

To implement our proposed power estimation model on a group of interconnects (such as a bus), we first estimate how much energy is dissipated on each individual interconnect. To compute the energy dissipation of a single interconnect, all interconnects with a significant effect on the interconnect being evaluated must be considered. Because the cross-coupling capacitance between interconnects declines sharply the further apart the interconnects are, only an interconnect and its immediate neighbors need be considered when estimating energy dissipation.

Using detailed SPICE simulations of all possible types of transitions on a group of three interconnects, a *three-wire lookup table* is created. By looking up the type of transition which occurs on three interconnects in the lookup table, an accurate estimate of the center interconnect's energy consumption is returned. For example, Figure 3 contains the energy dissipation of the center interconnect for any type of transition on three, 1mm, minimum-width interconnects in .18μm Cu technology [10].

Trans.	Energy (pJ)	Trans.	Energy (pJ)
s=no transition, t=transition, t'=opposing transition			
s s s	0.0	t t t	0.002
s t s	0.156	s t t	0.045
s s t	0.033	t t t'	0.153
t s t	0.131	s t t'	0.327
t s t'	0.0	t' t t'	0.560

Figure 3: A three-wire lookup table for 1mm, .18μm Cu interconnects

To determine the energy dissipation of a group of interconnects, the sum of all individual interconnect's energy dissipation is computed. Because three interconnects are considered at a time, the lookup table required is small, allowing for fast power estimation, even at a high level. In addition, because the lookup table is generated from detailed SPICE simulations, the transmission line nature, along with the rise/fall times of the interconnect drivers, is accounted for in this power estimation method. However, the accuracy of this model depends on the accuracy of the three-wire lookup table.

Two main possibilities exist for generating an accurate three-wire lookup table for any set of interconnects. First, for a given set of interconnects, a complete SPICE simulation of three wires can be run. This is the best method for interconnects with unique inter-wire spacing, intrinsic capacitance, or other parameters making the set of interconnects unique.

However, during design of a large chip, some basic assumptions about the parameters of the interconnects (such as inter-wire spacing) can be made, leaving length as the primary variable. In this case, SPICE simulations can be run a-priori for several lengths of the interconnects, storing a lookup table for each length. To compute a three-wire lookup table for a wire length which is not already stored from previous SPICE simulations, the two three-wire lookup tables representing the two closest lengths are used for linear interpolation.

3.2 Results of proposed power estimation model

To determine the accuracy of our proposed power estimation model, we compared the energy dissipation estimates of our model, and a transition-count based model, against a full SPICE simulation of a 32-bit bus. Three-wire lookup tables from SPICE simulations were stored for several interconnect lengths, in .1mm increments. We derived the interconnect technology parameters from the UMC Group's specifications of their .18μm Cu technology [10], assuming minimum spacing between wires. For the transition-count based model, the energy dissipated by a transition on the middle interconnect, with both neighboring interconnects static, is multiplied by the total number of transitions.

In Figure 4, we show the percentage error of our proposed method and a traditional transition-based method compared to full SPICE simulations for random bus transitions. We tested our energy estimation technique on 10 different sets of 100 random vectors, testing each set of vectors at three random bus lengths. From the simulations, we observe that the transition-count based method has an average error of -32%, while our model exhibits an average error of -6.7%.

The inaccuracy of transition-count based methods for estimating energy dissipation is even more apparent with systematic (non-

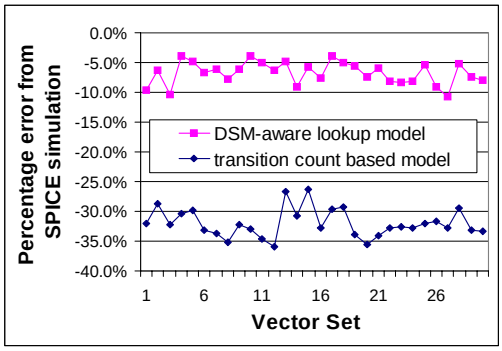

Figure 4: Percentage error of transition-count and DSM-aware power estimation methods for random bus transactions

random) data. In Figure 4, the transition-count based method is inaccurate, but still has a small standard deviation (2.4%) in percentage error. However, for data patterns with different types of transitions, the accuracy of the transition-count based method varies widely. Figure 5 shows the percentage error, compared to full SPICE simulations, of our proposed method and the traditional transition count-based method for 10 different non-random vector sets, tested at 3 random lengths each. The vector sets include 2 vector sets of address traces from software benchmarks, and 8 vector sets generated by a system-on-chip during three different stages of executing the JPEG image compression algorithm[12]. The average error observed when using a transition-count based power estimation method is 178%, with a standard deviation of 284%, while our method has an average error of -11% with a standard deviation of 6%. These results demonstrate the need for a power estimation model, like the one proposed in this paper, which considers the effect of nanometer technologies on energy dissipation. It also empirically validates the accuracy of our proposed power estimation model.

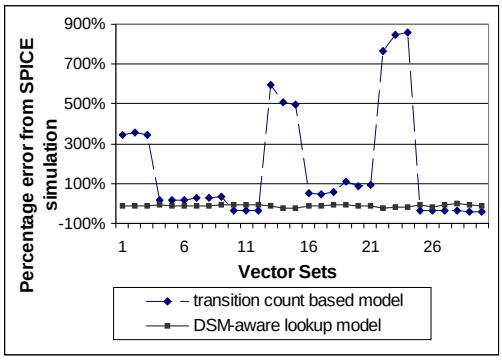

Figure 5: Percentage error of transition-count and DSM effect based power estimation methods for non-random bus transactions

With the use of our new power estimation model, it is possible to design and evaluate new methods for minimizing the energy dissipation of interconnects. In the next section, we propose a method which significantly reduces the energy consumption in interconnects, even though the number of transitions increase.

4. DSM-AWARE ENERGY MINIMIZATION

Most power minimization techniques available today do not consider the effects of nanometer technologies on power consumption. In this section, we propose a methodology for minimizing energy consumption which considers DSM effects. We also present results which demonstrate the effectiveness of our technique at minimizing energy consumption.

To minimize energy consumption while considering DSM effects, it is useful to note the role that the voltage swing across cross-coupling capacitances plays. *Opposing transitions* on neighboring interconnects, defined as one interconnect transitioning from 0 to 1, with its neighbor transitioning from 1 to 0 (or vice versa), may cause up to a $2V_{dd}$ voltage swing on the cross-coupling capacitances, leading to high energy consumption. Meanwhile, a *same direction transition* on neighboring interconnects causes no voltage swing on the cross-coupling capacitances, requiring much less energy consumption.

To eliminate opposing transitions and increase the number of same direction transitions to save energy, a *non-data line* interconnect can be inserted between neighboring interconnects. The non-data line is driven by an and gate of the data interconnects on either side of the non-data line. By driving the non-data line with an and gate, opposing transitions between neighboring interconnects are eliminated, while preserving same direction transitions. In addition to saving energy, the insertion of non-data lines also helps reduce the timing and noise problems due to DSM effects, much like the method of inserting shielding interconnects tied to V_{dd} or ground as presented in [13].

However, the insertion of non-data lines also causes a significant increase in the area consumed by the interconnects. Therefore, when using non-data lines to minimize energy dissipation, it is necessary to insure that the power savings achieved are worth the increased area overhead. In the subsection that follows, we describe our DSM-aware interconnect energy minimization methodology that judiciously inserts non-data lines to minimize interconnect energy, with and without area overhead constraints.

4.1 DSM-aware methodology to minimize interconnect energy dissipation

The goal of our methodology is to minimize the energy consumption of the DSM interconnects, while satisfying area overhead constraints. If no area constraint is given, then the methodology attempts to find the best placement of non-data lines to minimize power dissipation.

The inputs to our methodology, as shown in Figure 6, include a vector set used for energy estimation of the current non-data line placement configuration, and an area constraint specifying the maximum number of non-data lines to add to the bus. The methodology first evaluates the original case where no non-data lines are added. It then adds a non-data line between two neighboring interconnects with the largest number of opposing transitions. The energy consumption of the entire bus with the new non-data line is evaluated. If the energy consumption has not increased (gain > 0), and the maximum number of non-data lines has not been inserted, then the methodology repeats. Otherwise, the methodology returns the bus configuration last evaluated before the insertion of the last non-data line. If no area constraint is specified, then the methodology returns when the power starts to increase with the insertion of non-data lines (Gain $<= 0$) or when every data interconnect is separated from its neighbors by a non-data line.

4.2 Result of energy minimization methodology

To evaluate the effectiveness of our energy minimization methodology, we tested the methodology with several different types of vector sets on a 32 bit bus. Our vector sets included a random set, a set of address vectors coming from software benchmarks, and three image data traces obtained from an image compression system-on-chip that we are designing [11]. The system-on-chip executes the JPEG image compression algorithm[12], and the traces come from three different stages of the JPEG image compression algorithm (before any compression, "*raw*"; after the DCT transform step, "*DCT*"; and after the quantization step, "*quant*").

In Figure 7, we present the results of our methodology when run without area constraints. Each row represents a different test vector set at a random interconnect length. Each vector set was tested at 2 random lengths. The column titled **# non-data lines**

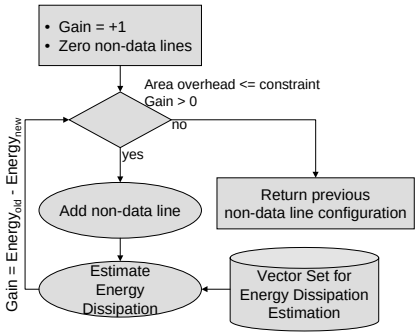

Figure 6: Energy minimization methodology using non-data lines

represents the number of non-data lines that the methodology uses to minimize energy consumption. In the columns representing the number of transitions (**#trans**), the worst case timing delay for the signals to travel across the bus (**T**), the energy consumed (**E**), and the area, timing delay, and energy product (**ATE**), the numbers are presented as percentages of the original case without any non-data lines inserted.

Vector Set	Line Length	# non-data lines	% of original			
			#trans	T	E	ATE
random	2.084	31	172%	79%	51%	80%
	2.866	31	172%	78%	52%	80%
address	2.408	31	189%	79%	63%	98%
	0.618	11	126%	100%	63%	87%
image-raw	1.225	31	167%	90%	50%	86%
	3.073	31	167%	78%	51%	78%
image-DCT	3.359	15	136%	100%	57%	84%
	0.975	15	136%	100%	57%	84%
image-quant	3.168	1	101%	100%	95%	97%
	3.525	1	101%	100%	94%	97%

Figure 7: Effects of minimization methodology without any area constraints

We make the following observations from Figure 7. First, the energy consumption of the bus configuration output by the unconstrained minimization methodology can be up to 50% lower than without any non-data lines inserted. Second, for the cases when a non-data line is inserted between every interconnect (indicated by a 31 in the # non-data lines column), the timing delay of the interconnects is significantly reduced leading to an average reduction of 19% in worst case timing delay. Third, the ATE product averages 87% of the original case. The fact that the ATE product after running the methodology is lower is especially important because the ATE product directly weighs the area overhead of the methodology against the savings in energy consumption and timing delay. Fourth, to minimize energy consumption for some types of data, a non-data line is not needed between every interconnect. For example, both the image-DCT and image-quant traces achieved minimum energy consumption with the insertion of fewer than 31 (15 and 1, respectively) non-data lines. Lastly, the number of transitions may increase significantly even though the energy dissipation is lower, demonstrating the importance of accounting for DSM effects when attempting to lower energy consumption.

Figure 8 presents the results of using our methodology with area constraints. The energy savings plotted are averaged over two random lengths for each set of data. The Figure shows that as the constraint on the number of non-data lines allowed increases (x axes), the energy savings also increase (y axes). It is interesting to note that for the address and image-DCT data traces, the energy savings quickly increase with the first few non-data lines inserted, but more slowly increases with more non-data lines. This demonstrates that for certain types of data, significant power savings can be achieved through the insertion of relatively few non-data lines.

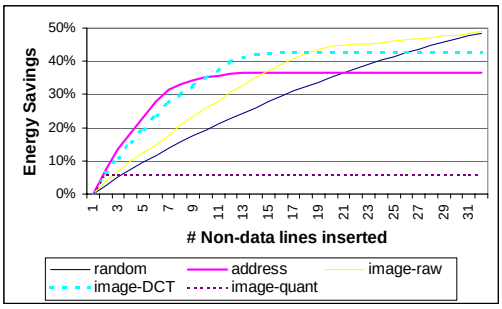

Figure 8: Energy savings available with varying area constraints

The results presented from using our non-data line insertion methodology for minimizing interconnect energy show that significant energy savings are available through the insertion of non-data lines. In addition, our methodology is effective in finding minimum energy configurations with or without area overhead constraints. It is important to note that to effectively minimize energy consumption, the method for estimating energy consumption must consider the effect of nanometer technologies on power, as the number of transitions actually increase while the power usage decreases.

5. CONCLUSION

We have demonstrated the importance of considering DSM effects when estimating the energy consumption of long interconnects in DSM system-on-chips. DSM effects must be considered during both estimation and minimization of energy. We have proposed a novel lookup table method for estimating energy dissipation in interconnects, which is significantly more accurate than traditional, transition-count based methods. We have also introduced a methodology for minimizing energy dissipation with and without area constraints. While the estimation and minimization methods proposed here have been targeted towards capacitive coupling effects, we will investigate in the future other DSM effects, like IR drop, and their effect on energy consumption.

6. REFERENCES

[1] M. Cuviello, S. Dey, X. Bai, and Y. Zhao, "Fault Modeling and Simulation for Crosstalk in System-on-Chip Interconnects", in *Digest of Technical Papers, 1999 ICCAD*, p. 297, 1999.

[2] X. Zhang, "Coupling Effects on Wire Delay. Challenges in Deep Submicron VLSI Design", *IEEE Circuits and Devices Magazine*, vol. 12, November 1996.

[3] P. Nordholz, et. al, "Signal Integrity Problems in Deep Submicron arising from Interconnects between Cores", in *Proc. of IEEE VTS*, 1998.

[4] D. Liu and C. Svensson, "Power Consumption Estimation in CMOS VLSI Chips", *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 663–70, June 1994.

[5] M. R. Stan and W. P. Burleson, "Low-Power Encodings for Global Communication in CMOS VLSI", *IEEE Trans. on VLSI Systems*, vol. 5, pp. 444–55, December 1997.

[6] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj, "A Coding Framework For Low-Power Address and Data Busses", *IEEE Trans. on VLSI Systems*, 1998.

[7] P. P. Sotiriadis and A. Chandrakasan, "Low Power Bus Coding Techniques Considering Inter-wire Capacitances", in *CICC*, May 2000.

[8] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors", 1999.

[9] Avant! Corporation, HSPICE, v.98.2.

[10] UMC Group, *0.18um 1P6M Logic Process Interconnect Capacitance Model (Rev. 1.2)*, July 1999.

[11] S. Dey, et. al, "Using a Soft Core in a SOC Design: Experiences with picoJava", *IEEE Design and Test*, July/September 2000.

[12] G. K. Wallace, "The JPEG still picture compression standard", in *IEEE Trans. on Consumer Electronics*, vol. 38, February 1992.

[13] A. P. Chandrakasan, F. Fox, and W. Bowhill, eds., *Design of high-performance microprocessor circuits*, 2001.