Using an energy-time index for reducing power consumption in buses according to their application

Valentin Gies
ENSTA
32, Bd Victor
75015 Paris, FRANCE
Email: contact@vgies.com

Thierry M. Bernard
ENSTA
32, Bd Victor
75015 Paris, FRANCE

***Abstract*— Reducing power dissipation is a key issue in designing embedded systems. Nowadays, most of the energy is dissipated in communication buses. In this paper, we propose a new method for optimizing the number of wires in a bus of fixed width, together with its corresponding communication scheme. It allows an application dependent trade-off between data throughput and power consumption. Then, an energy-time index is introduced to measure the relative capabilities of a bus in terms of speed and power-savings. We propose a method to design a bus with a desired index value depending on the application. Furthermore, this index is meaningful, therefore valuable, for characterizing any existing bus.**

I. Introduction

Low power and miniaturization are key ingredients for developing embedded systems such as mobile phones, personal digital assistants (PDAs) and embedded computers for mobiles robots.

Let's recall the causes of power dissipation in electronic devices. A chip is composed of many blocks dedicated to signal processing and computations such as arithmetic and logic units (ALUs), or information storage such as memories. These blocks are interconnected using groups of wires called *buses*.

There are two main sources of power consumption in CMOS circuits :

- The first one is the energy spent during the commutation of transistors inside computational blocks. Between commutations, CMOS circuits exhibit a practically negligible static current. The power dissipated is proportional to the clock frequency. It is also proportional to a polynomial function of the supply voltage.
- The second one is the energy spent for data transmission on buses. Each wire of a bus is a capacitor charged or discharged every time wire potential is modified. The power dissipated is consequently directly proportional to the bus capacitance and to the clock frequency. It also behaves quadratically with the supply voltage.

Both sources of power dissipation are proportional to the operating frequency. Consequently, higher frequencies linearly impact power dissipation. Fortunately, along the evolution of CMOS technology, this has been compensated for by the power supply voltage decrease.

However, shrinking has reduced the space between adjacent wires, and it has therefore increased the inter-wire capacitances. As a consequence, power consumption for data transmission in buses, proportional to bus capacitance has increased because of the scale reduction of devices. Until the 90's, power consumption was mainly caused by transistor commutations in computing or data storage blocks. For several years, in the deep sub-micron era, power consumption for transmitting communicating between blocks has become comparable or even larger than computing power consumption.

This paper focus on power consumption in buses. In section II, we give a summary of related works. In section III, we define the bus model used for the rest of the paper. In section IV, we estimate power consumption and time needed for transmitting one bit of information. In section V, we introduce an energy-time index characterizing whether a bus is designed to maximize speed or to reduce power consumption, an application using this index is proposed.

II. Related work

In order to lower data transmission energy consumption, many works has been proposed in different directions :

- Circuit parameter optimization : bus capacitances reduction is possible optimizing physical parameters such as dielectric permittivity. Such an optimization is technology specific and cannot be adapted easily when upgrading to a newer technology.
- Bus Encoding scheme : reducing energy loss can be performed by using buses in energetically cheap configurations [1][2][3][4]. These methods allow to decrease energy consumption, but can also reduce the transmission rate[5].
- Energy recycling : energy spent for loading capacitances can be reduced by using *adiabatic* techniques [6][7]. One of them amounts to split a voltage transition in a bus wire into N successive smaller transitions. Such a technique allow to reduce energy loss by factor of N at the expense of slowing down by a factor of N.

The rest of the paper focuses on the trade-off between speed and energy. It is shown that, unlike the above adiabatic technique, data throughput is not bound to be reduced by the same factor as power consumption.

III. Bus model

Let's define the bus model used throughout this paper. In the past, the pin number and the communication protocol of IC components was fixed. Nowadays, to increase efficiency, designers interconnect pre-designed cores often having a flexible interface. That means the number of wires composing interconnection buses can be adapted as needed. However, cores are placed in order to minimize space, and buses are placed later. Consequently, buses width is often highly constrained and the only parameter that can be set freely is the number of wires composing the bus.

In deep sub-micron technology, there are 2 different types of capacitances responsible for power loss in buses, *wire-metal planes capacitances* and *inter-wire capacitances*. The first one is fixed for a given technology, whereas the second one depends on the distance between adjacent wires. Using more wires in fixed-width bus allows to communicate faster, but also reduce the distance between wires, increasing the inter-wire capacitance, and therefore increasing power consumption. This trade-off is studied in section V. The bus model used is a simplified version of [8] and is depicted in Fig. 1. The potential of a wire is 0 or V depending on its binary value: 0 or 1

Considering a n-wire bus of fixed width L, distance between

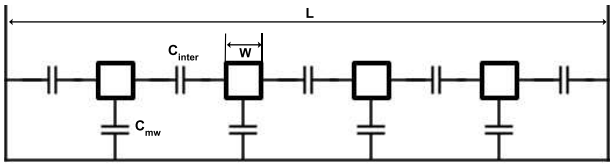


Fig. 1. Bus model

2 wires is :

$$D = \frac{L - n\ W}{n+1}$$

Consequently, the inter-wire capacitance is :

$$C_{inter} = \frac{\alpha}{\frac{L}{nW} - 1}$$

with α the inter-wire capacitance when distance between wires is equal to W. This constant depends on the technology used. Metal-wire capacitance C_{mw} also depends on technology but is supposed not to depend on n. We have :

$$C_{mw} = \beta$$

IV. Calculating bus power and time needed for transmitting data

As shown in previous sections, energy dissipated in buses is related to wire capacitances and to the number of wires used for communicating. To reduce the power consumption, a first solution is to reduce the number of wires composing the fixed-width bus, thus reducing the inter-wire capacitance. A second solution is to reduce the number of wires for which the potential changes at each clock cycle, thus reducing the number of capacitances to load. Obviously, these reductions also affect the amount of data sent at each clock cycle.

Actually, these reductions allow to trade-off speed for power consumption, as will be shown in section V. For the moment, we evaluate the power needed to transfer data through a n-wire bus with at most k wires changing at each clock cycle and the time needed for this transmission. Values are calculated per bit transmitted.

First, average power consumption is evaluated in the case of changing exactly k out of n wire potentials, and then in the case of changing at most k out of n wire potentials.

A. Power for changing exactly k out of n wire potentials

Energy is spent in buses when the capacitance loads of wires changes. Power for changing the potential of a wire is computed using the following formula :

$$P\ =\ C\ V^2$$

where C is the capacitances of wires having there loads modified, V is the supply voltage. C includes metal-wire capacitance C_{mw} and inter-wires capacitances C_{inter}. When a wire potential is changed, C_{mw} load always changes simultaneously. Concerning C_{inter}, its load changes only if adjacent wires are not simultaneously modified in the same way. Depending on the situation, power used for modifying a wire potential is finally :

$$P\ =\ (C_{mw} + x\ C_{inter})\ V^2, \quad x \in [0;4]$$

x depends on the transitions of the wire neighbors (for example, $x = 4$ for a wire going from 0 to V between two wires going from V to 0). Using low-value x has been investigated in bus encoding methods (cf. section II).

Our work does not deal with this kind of power reduction (however they can be used complementarily), therefore every bus transition is equiprobable [9]. Thus, in the next sections,it will be correct to base derivations on the average power consumption. Now, the average energy spent for a wire transition is :

$$P\ =\ (C_{mw} + 2\ C_{inter})\ V^2$$

Changing k wire potentials will statistically cost k times as much as changing 1 wire potential. Power consumption when using k wires is then :

$$P_k(n) = k\beta + k\frac{\alpha}{\frac{L}{nW} - 1}$$

B. Power for changing at most k out of n wire potentials

The above formula shows that the consumption increases linearly with k. Using several low k values for communicating among a bus seems to be a interesting solution for optimizing power consumption. Let's study a communication scheme allowing to change at most k out of n wire potentials

The average power needed for changing the state of at most k wires out of n is an arithmetical mean of $P_m(n)$, $m \in [0;k]$. Each term is weighted by its corresponding number of transition configurations. This number is $\binom{k}{n}$ for

a transition involving exactly k out of n wires. Finally, the power consumption corresponding to a transmission of at most k out of n wires is :

$$P(n,k) = \frac{\sum_{m=0}^{k} \binom{m}{n} P_m(n)}{\sum_{m=0}^{k} \binom{m}{n}}$$

C. Power and time needed per bit transmitted

The power consumption evaluated in IV-B allows to change at most k out of n wires at each clock cycle. This bus transition encodes $\sum_{m=0}^{k} \binom{m}{n}$ combinations, corresponding to the following number of bits :

$$Nb_{bits}(n,k) = log_2 \left(\sum_{m=0}^{k} \binom{m}{n} \right)$$

Consequently, the average power needed for transmitting each data bit is :

$$P_{1bit}(n,k) = \frac{\sum_{m=0}^{k} \binom{m}{n} P_m(n)}{\left(\sum_{m=0}^{k} \binom{m}{n} \right) log_2 \left(\sum_{m=0}^{k} \binom{m}{n} \right)}$$

The average time needed for transmitting each data bit is :

$$T_{1bit}(n,k) = \frac{1}{f \ log_2 \left(\sum_{m=0}^{k} \binom{m}{n} \right)}$$

where f is the frequency of the clock cycles.

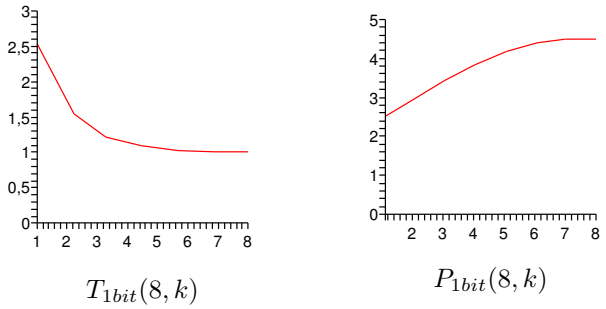
D. Application

An application example is presented in Fig.2 and Fig.3 considering a bus having a fixed width $L = 9\lambda$, with wires having a width $W = 1\lambda$, where λ is a length unit depending on the technology chosen. Under these conditions, the bus can be composed of a maximum of 8 wires.

Fig.2 shows for different values of k, the power and the time needed to send each data bit in a 8-wire bus using at most k wires. As explained before, increasing k increases the power consumption (there is more activity on the bus) and reduces the transmission time. Fig.3 shows for different values of n, the power and the time needed to send each data bit in a n wire bus using at most 1 wire. Increasing n increases the power consumption (inter-wire capacitances are bigger), but reduces transmission time (there are more possible combinations on the bus). Both examples are an illustration of the trade-off between speed and power consumption.

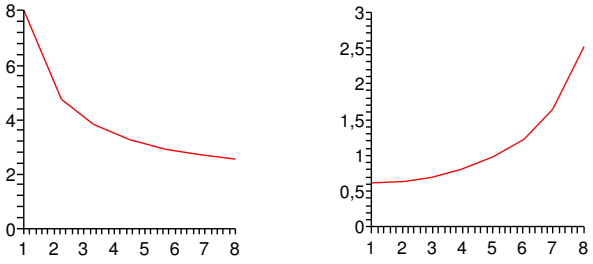
Let's determine the range of variation for power consumption and transmission speed corresponding to the situation above($L = 9\lambda$ and $W = \lambda$). The minimum power consumption (corresponding to $n = 1$ and $k = 1$) is 16 times smaller than the biggest one (corresponding to $n = 8$ and $k = 8$). However, between this 2 extreme cases, speed as been divided by 8.

As shown in this basic example, the range of power consumption and communication speed is very wide even in a small size bus and our method allows to adapt the bus set-up to a wide range of communication cases.



$T_{1bit}(8,k)$ $P_{1bit}(8,k)$

Fig. 2. Power and time consumption for transmitting each data bit in a 8 wire bus using at most k wires at each clock cycle



$T_{1bit}(n,1)$ $P_{1bit}(n,1)$

Fig. 3. Power and time consumption for transmitting each data bit in a n wire bus using at most 1 wire at each clock cycle

To determine whether a bus is a power-saving one or a fast-operating one, we propose in next section to introduce an energy-time index characterizing its behavior.

V. Energy-time index for bus optimization and characterization

Depending on the application, the trade-off between energy consumption and information throughput can be adjusted for reducing power or increasing throughput. This choice has an important impact on the design of the bus. Obviously, high-speed buses will have as much wires as possible and will exploit all transition configurations including the most power consuming ones (where there is a transition on each wire). In contrast, power-saving buses will use a smaller number of wires with transitions on only a small proportion of them. To characterize the behavior of a bus design, we introduce an energy-time index indicating whether a bus design allows fast communications or is power-saving.

A. Energy-time index and bus optimization

Let's define $\gamma \in [0;1]$ as an energy time index for characterizing the relative weight of speed and power-saving in a bus design. $\gamma = 1$ will lead to optimize the bus for communication speed only, whereas $\gamma = 0$ will lead to minimize power consumption only. The bus set-up will be determined by maximizing an efficiency function given a constraint represented by γ on the bus behavior. To do so, we propose to use the following efficiency function :

$$Eff(n,k) = \frac{(Nb_{bits}(n,k))^{\gamma}}{(P_{1bit}(n,k))^{1-\gamma}}$$

We recall that Nb_{bits} is the number of data bits transmitted at each clock cycle (i.e. that is the information throughput) and P_{1bit} the power consumption for transmitting 1 data bit (cf. section IV).
For a given γ, the efficiency function is computed for every (n,k) possible couple. The couple corresponding to the maximal efficiency is the best bus configuration corresponding to the chosen energy-time index γ. When $\gamma = 1$, we can verify that the transmission speed is being optimized (in this case $E_{ff}(n,k) = Nb_{bits}(n,k)$), and when $\gamma = 0$, the power consumption will be optimized ($E_{ff}(n,k) = 1/(P_{1bit}(n,k))$).

Let's consider an example. Considered bus width is fixed to $L = 65\lambda$ and $W = \lambda$, $\alpha = \beta = 1$. Such a configuration is commonly used for 32-bit buses where wires are evenly spaced at a distance $W = \lambda$ from each other. Table I provides n and k corresponding to different values of γ. Highest γ

TABLE I

OPTIMIZED N AND K FOR DIFFERENT γ VALUES

γ	0	0.2	0.3	0.4	0.5	0.55	0.6	0.7	0.8	0.9	1
n	9	12	17	21	28	32	35	43	51	58	64
k	1	1	3	5	9	11	13	19	26	35	64

values correspond to highest n, because speed is predominant. Lowest γ values correspond to power saving configurations using low n values (for reducing inter-wire capacitances) and only $k = 1$ wire. Setting $\gamma = 0.5$ leads to a bus composed of $n = 28$ wires and to using only $k = 9$ wires at each clock cycle. Compared with a 32-bit bus, using every wire, power consumption is reduced by 40% whereas transmission speed is reduced by 26%. This confirm that data throughput does not have to suffer from the same factor as the power consumption decrease, as mentioned before.
The introduced energy-time index allows to efficiently design a bus depending on the importance of speed or power saving in an application.

B. Characterization of a bus configuration

Using an efficiency function optimization method is not limited to optimize a bus for a speed or low power dedicated application depending on the energy-time index γ. Another interesting application is to characterize existing bus designs by calculating their corresponding γ. It will allow to check whether a bus is suitable for a power-saving design or not.
The algorithm used is the following one. First, L, W, α and β are determined considering the physical design of the considered bus. Then a table containing the optimized values of n and k for different values of γ is generated (cf. Table I). The γ corresponding to the n equal the real number of wire of the bus is taken as the energy-time index of the bus.
For example, using a 32-wire bus with $L = 64\lambda$, $W = \lambda$, $\alpha = \beta$ lead to $\gamma = 0.55$ (cf.I). Using a 16-wire bus instead of a 32-wire bus without changing the bus parameters would lead to $\gamma = 0.3$. The 16-bit bus energy-time index is lower than the 32-bit one, that means the 16-bit bus is more power-saving than the other one. This result can be confirmed calculating the cost for each bit transmitted ; power reduction is about 40% using the 16-wire bus.
The proposed algorithm extracts an energy-time index for bus characterization. Such an index can be used to classify different bus designs, or to determine if a bus is well designed for a given application. For example in a power-saving application, a bus energy-time index higher than 0.5 indicates that the bus is not suitable for the application and must be modified.

VI. CONCLUSION

In this paper we have proposed a new method for optimizing a bus structure depending on its application. An energy-time index has been introduced to quantify the relative weights of speed and power-saving in a bus design. We have proposed two applications for this index. The first one is to design an optimized bus structure and to define its corresponding communication scheme. The second one is to check if an existing bus structure is suitable for its corresponding high speed or power saving application. Future works will focus on integrating additional bus encoding techniques to improve power reduction.

REFERENCES

[1] M. Stan and W. Burleson, "Bus-invert coding for low-power i/o," 1995. [Online]. Available: citeseer.ist.psu.edu/stan95businvert.html
[2] L. Benini, A. Macii, E. Macii, M. Poncino, and R. Scarsi, "Architectures and synthesis algorithms for power efficient bus interfaces," 2000.
[3] P. P. Sotiriadis and A. Chandrakasan, "Low power bus coding techniques considering inter-wire capacitances."
[4] F. Najm, "Transition density: A new measure of activity in digital circuits," 1992. [Online]. Available: citeseer.ist.psu.edu/article/najm92transition.html
[5] P. P. Sotiriadis, A. Wang, and A. Chandrakasan, "Transition pattern coding: An approach to reduce energy in interconnect." [Online]. Available: citeseer.ist.psu.edu/510479.html
[6] L. S. Karafyllidis, "A new adiabatic technique for designing low power array architectures." [Online]. Available: citeseer.ist.psu.edu/439669.html
[7] W. Athas, J. Koller, and L. Svensson, "An energy-efficient cmos line driver using adiabatic switching," 1994. [Online]. Available: citeseer.ist.psu.edu/athas93energyefficient.html
[8] S.-C. Wong, G.-Y. Lee, and D.-J. Ma, "Modeling of interconnect capacitance, delay, and crosstalk in vlsi," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, no. 1, 2000.
[9] A. Shen, A. Ghosh, S. Devadas, and K. Keutzer, "On average power dissipation and random pattern testability."