REDUCING NOISE IN HIGH-PERFORMANCE ON-CHIP BUS SEGMENTS

P. Liljeberg, J. Tuominen, S. Tuuna, J. Plosila, and J. Isoaho

Electronics and Communication Systems
Dept. of Information Technology
University of Turku, Finland
{pakrli | joeltu | satatu | juplos | jisoaho}@utu.fi

ABSTRACT

In this paper we present techniques and analysis on reducing crosstalk and switching noise in segments of a pipelined on-chip bus. Each segment in such a bus architecture consists of two high-performance links transfering data simultaneously in the opposite directions. The segments are isolated by self-timed transfer stages which pipeline data flow. Time-interleaving, dual-rail and 1-of-4 data encoding techniques are used for reducing problems caused by interconnect signal coupling, current peaks, and the probability for erroneous states.

1. INTRODUCTION

Power supply noise or unwanted fluctuation of the supply voltage within a digital ULSI chip mainly originates from simultaneous clock-induced switching of CMOS circuits which causes high peak current draws from the power source. The total power supply noise is the sum of two major components: the resistive voltage drop IR and the inductive switching noise $L\Delta I/\Delta t$ [2]. Here R and L are the effective supply wire resistance and inductance, respectively, and ΔI is the total current change during the rise or fall time Δt of the concurrently transitioning signals. Hence, dealing with high current spikes, rather than average current, is a key issue to minimize voltage fluctuations in the power supply network. Typically, large on-chip decoupling capacitors are needed to provide a stable power supply for every system module on a chip [3]. The area required by these capacitors naturally increases with the size and complexity of the system. In order to decrease the need for decoupling capacitance, current peaks must be lowered. This can be accomplished by decreasing the number of simultaneous switching events in the system. Such an approach, based on *de-synchronization* of inter-module communication, is presented in this paper.

Noise caused by capacitive and inductive crosstalk between on-chip signal wires is another fundamental problem in modern high-speed system-on-chip design. Capacitive coupling is currently dominant, but inductive coupling becomes more and more significant as signal frequencies and on-chip wire lengths increase. Crosstalk has two major detrimental effects. Firstly, if the magnitude and duration of the coupled noise is sufficient, a signal may temporarily assume an erroneous logic value which in turn may lead to a logical failure. Secondly, crosstalk also affects timing. The delay of a wire not only depends on the properties of the wire itself but also on how the wires that are capacitively or inductively coupled to it are switching. If a wire and another wire coupled to it switch simultaneously in opposite directions, crosstalk increases the delay of the wires because twice as much charge must be transferred across the coupling capacitance. On the other hand, if the coupled wires switch in the same direction, the delay is reduced.

Crosstalk can be kept within appropriate limits by sufficient wire spacing. The spacing requirements can be significantly eased, resulting in a more compact bus implementation, by controlling transmission of signals in such a way that crosstalk is minimized. The noise reduction methodology described in this paper combines de-synchronization techniques with appropriate asynchronous data encoding schemes to achieve this goal in on-chip bus design.

2. METHODOLOGY FOR NOISE REDUCTION

The purpose of the methodology is to prevent spurious transitions in the on-chip wires and thereby the gates that are connected to those wires. If this is not handled properly, it could potentially cause static erroneous states to the circuits and lead to a malfunction of the entire system. Even though the noise might not always cause false switching, it could still degrade the performance by slowing down the signals. In addition to above, the method concentrates on minimizing current peaks by employing asynchronous techniques and encoding schemes. In the first phase of the method the system is partitioned into appropriate regions, illustrated in Figure 1. A noise budget is created to define noise margin for each of these partitions. Doing so the effort can be focused on the most critical regions separately. In the second phase the number of simultaneous switching events

is decreased in each region so that the current peaks will be lower, consequently decreasing power supply noise and electromagnetic interference (EMI) [4]. This could also be referred to as de-synchronization, because the timing of the system is re-tuned so that simultaneous clock related events are distributed into a larger time frame. The third phase deals with minimizing crosstalk. Different asynchronous encoding methods and de-synchronization can be used to reduce crosstalk on a bus as will be explained in Section 4.

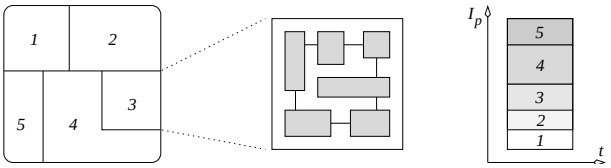

Fig. 1. System partitioning for decreasing concurrent events

3. PIPELINED BUS

In this paper both crosstalk and switching noise characteristics are analyzed on an advanced pipelined bus structure [5] targeted for GALS based system-on-chip design. The bus architecture is a distributed organization based on self-timed communication. Unlike in the case of a conventional bus, the pipelined bus can be simultaneously accessed by all the attached processing elements. The physical wires that implement the bus are divided into N-1 segments, where N is the number of modules connected to the bus. The segments are isolated from each other by N *transfer stages*, one attached to each module. The arbitration and control are evenly distributed among the transfer stages which contain internal FIFO queues for pipelining the data flow. A bus segment between adjacent stages consists of two separate unidirectional point-to-point interconnects which transfer data asynchronously between the stages in opposite directions. These two links of a segment can operate in parallel, and due to pipelining, all segments of the system bus can transfer data concurrently. The pipelined bus architecture is illustrated in Figure 2.

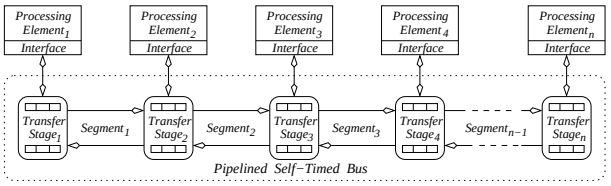

Fig. 2. Pipelined bus architecture

In the following paragraphs crosstalk and switching noise in a segment of the pipelined bus is investigated. Different signaling methods are compared against each others in order to minimize noise, and effect to the performance is analyzed. The pipelined bus was implemented by using a 0.13 μm technology with 1.2 V supply voltage. The wires in the bus segment in question are placed 0.6 μm apart from each other and they have length of 2 mm, width of 0.6 μm and thickness of 0.32 μm. The size of transmitted messages are 32-bit in all the cases with 150 ps rise and fall times.

4. NOISE REDUCTION IN BUS

The clock dictated operation forces a great deal of gates and flip-flops in the chip to change their states nearly at the same moment. As a consequence, the current profile of the circuit is dominated by the clock induced high peaks, which are the main source for power supply noise. Utilizing asynchronous techniques those peaks can be folded to a longer period of time by adjusting the timing of the circuit. Furthermore, in a synchronous design a large number of capacitively and inductive coupled interconnects along the chip also switch simultaneously. Crosstalk between those interconnect wires can be reduced by employing self-timed protocols and encoding methods. The noise analysis was performed for both crosstalk and power supply noise. The worst-case switching activity was expected.

4.1. Noise characteristics of the bus segment

In the first phase noise characteristics on a bus segment are analyzed when employing the bundled data convention with the two-phase signaling protocol [7]. This protocol was chosen in order to minimize the signaling events, and thereby the delay, in the rather long interconnect with considerable parasitic properties. However, in this case the signal protocol could be as well synchronous from the noise analysis point of view, because all wires may switch at the same time and no encoding techniques are used. This signaling method serves as a reference point of the noise comparison against the other methods.

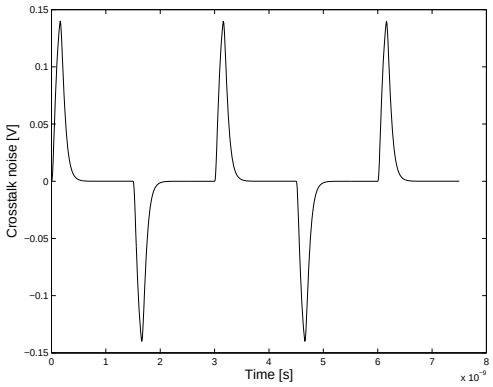

Fig. 3. Crosstalk voltage on a bus segment.

The worst-case voltage coupling between adjacent wires, crosstalk, in such a bus is shown in Figure 3, where three transfer cycles are considered. It occurs when all the wires in the bus except one in the middle switches simultaneously from zero to one while the intended value for the wire in the middle is zero. The maximum crosstalk voltage that is coupled to that wire is 144 mV or 12 % of the supply voltage. One might think 12 % is not a significant amount of noise, but it could easily be crucial when added to the other noise sources such as switching noise, electromagnetic interference, receiver and transmitter offset and so on. The current draw of the bus drivers is shown in Figure 4 by the curve a, the upper one. As can be seen, the current profile is clearly dominated by the current peaks induced by simultaneous signal switching.

4.2. Time-Interleaving

In this approach data bits are send in a time-interleaved fashion. A message is divided into bit groups which are transmitted at slightly different times with respect to each other. Hence, the power hungry bus drivers will not switch exactly at the same moment of time, instead only a set of drivers will switch simultaneously. This considerably reduces the peak current draw and therefore the switching noise. Each of these data groups are formed so that a group does not contain adjacent bits. For example if a 32-bit message is divided into four groups, the first group contains bits 0, 4, 8, 12, 16, 20, 24, 28. This reduces crosstalk since neighboring bus wires do not switch exactly at the same moment. In addition to attenuated crosstalk characteristics, such a bit division decreases the switching noise even further. This is because the bus drivers involved in transmitting a certain group of bits are not located next to each other. Therefore the concurrent current draw of the drivers is spread into a larger area, reducing the local peak current. The number of wires in the bus are kept same as before, n-bit messages requires n wires for the data and 2 wires for the handshake signals. Obviously, this method sacrifices the performance to the increased noise margin. However, by keeping the delays between bit groups relatively small, considerable reduction of noise can be achieved with a minor performance loss.

The current profile of the bus segment with time-interleaving is shown in Figure 4 by the curve b. Each message was divided into four 8-bit groups which were transmitted in an interleaved fashion. The time interval between groups was 80 ps, obtained by using an average size driver. As can be seen, the peak values are considerably lower when compared to the synchronous curve. Furthermore, the profile is significantly smoother and the peaks are reduced by 43%.

In addition to the above data partitioning, the effect on the noise when the data transmission is partitioned into two and three groups were studied. The crosstalk and peak current values with different message partitioning are shown in Figure 5, where crosstalk values are illustrated as solid lines and peak current values are presented as dashed lines. The analysis was performed with different interleaving times up to 200 ps, after which the reduction in both values can be thought saturated. As can be expected, the amount of noise decreases when the time between bit groups increases. The noise characteristic with different interleaving times when the data is partitioned into four 8-bit groups is presented as the curves denoted by c. This method provides the lowest noise coefficients of the three different partitioning methods that were studied. However, this is the slowest one since it contains most groups, and therefore the time consumed by interleaving is largest. For example, by using 100 ps time intervals the transfer time is increased by 300 ps but at the same time the crosstalk is reduced by 50 %, from 12 % to 6 %, and the peak current has decreased by 49 %. Similar behavior but with a smaller decrease in noise characteristics can be seen when the data is divided into three or two groups, the curves labeled by b and a, respectively. However the decrease in crosstalk is rather small when the data transfer is divided into two groups.

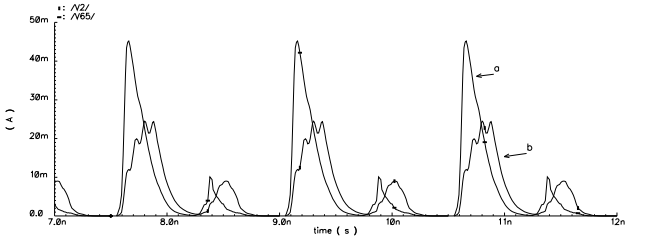

Fig. 4. Current profiles of a bus segment a) synchronous, b) time-interleaved.

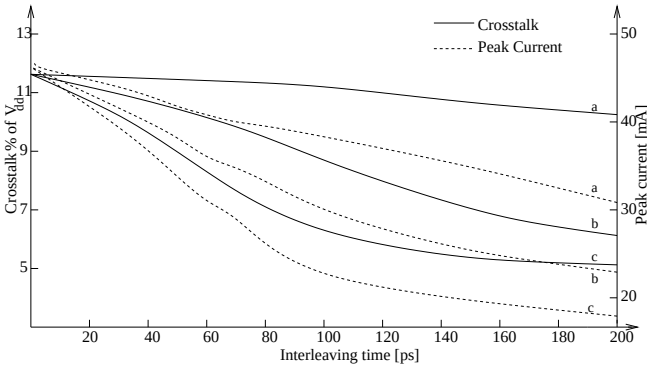

Fig. 5. Peak current in a function of time-interleaving

4.3. Dual-rail and 1-of-4 data encoding

An alternative to the bundled data method is *dual-rail* encoding in which each data bit act as a separate request signal [6]. A data bit is encoded onto two wires; transmitting

n-bit data then requires $2n + 1$ wires, $2n$ for data and one for acknowledge. An often used 4-phase dual-rail encoding has three legal states: '00' for idle, '10' for valid zero, and '01' for valid one. The combination '11' is the illegal state. Transmission of a bit requires transition from the idle state to either the valid 0 or valid 1 state. After the sender has received the acknowledge, it must made transition back to the idle state. In the 2-phase dual-rail encoding there are no idle and illegal states. The transmitted value is encoded into events so that only of the two wires is allowed to make a transition during a cycle. A transition on one wire indicates the sending of '0' while a transition on the other indicates sending of '1'. After the receiver has acknowledged the data, a new transfer cycle may start immediately by a transition in either of the two wires. Notice that in this case the value of the code word is not important, only the mutually exclusive events matter. The dual-rail encoding is insensitive to the wire delays and hence there is no need for any timing assumptions like in the bundled data signaling. This is advantageous when one is using automated routing particularly for long on-chip interconnects. Compared to synchronous bus the reduction in power supply noise and crosstalk was rather small.

In addition to above commonly used encoding methods there are plenty of other techniques [8]. A particularly interesting one is a 1-of-4 delay insensitive data encoding scheme [1]. It bears a resemblance to the dual-rail encoding and can use both transition and level signaling protocols, even though the decoding of the transition signaling is quite a lot of more complex. In addition to that, the number of required wires is the same. In the 1-of-4 data encoding a two-bit symbol is transmitted by using four wires. A two-bit code, '00', '01', '10', or '11', is transmitted by changing the signal level on just one of the four wires. 1-of-4 encoding as well as all the other 1-of-N encoding methods are delay insensitive [8].

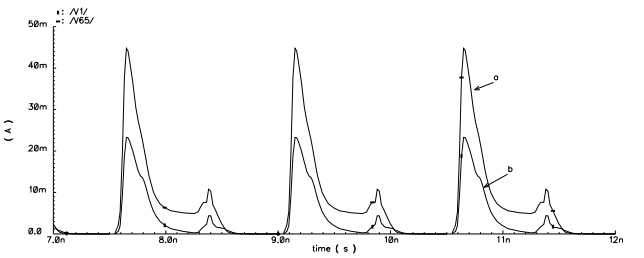

Fig. 6. Current profiles of a bus segment a) synchronous, b) 1-of-4 encoding.

The current profiles of the synchronous and 1-to-4 encoded bus are shown in Figure 6. The current peaks are significantly lower compared to the synchronous bus, while the effect on crosstalk is rather small. However, the crosstalk should not be as detrimental to performance for a 1-to-4 encoded interconnect as it is for single-rail implementation [1], since the adjacent wires cannot switch in opposite directions. The 1-of-4 encoding is attractive in the low-power perspective because it transmits two bits of information using two transitions as opposed to the dual-rail encoding that requires four transitions. The average current consumption of the 32-bit bus is 7.5 mA for the synchronous one and 3.5 mA for the 1-of-4 encoding interconnect, the reduction being 53 %. This demonstrates the low-power characteristics of the 1-of-4 encoding.

5. CONCLUSION

A methodology to minimize crosstalk and switching noise characteristics in a high performance bus was presented in this paper. The methodology is based on de-synchronization of the system and asynchronous encoding schemes. With de-synchronization, time-interleaved bit groups guarantee that all drivers do not switch simultaneously. This reduces both crosstalk and switching noise considerably. The peak current is decreased 43 % and the crosstalk is reduced 37 % when the bus transactions are divided into four signal groups with 70 ps time-interleaving. In addition to above the 1-of-4 delay insensitive data encoding scheme was applied. It is attractive from the low-power design point of view since the average current is reduced by 53 % compared to the synchronous bus. The study considered in this paper revealed the possibility to decrease the crosstalk and power supply noise by utilizing asynchronous techniques.

6. REFERENCES

[1] W. J. Bainbridge and S. B. Furber " Delay Insensitive System-on-Chip Interconnect Using 1-of-4 Data Encoding", in *Proc. of International Symposium on Asynchronous Circuits and Systems (ASYNC)*, March 2001.

[2] H. H. Chen and D. D. Ling. "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design", in *Proc. of Design Automation Conference,* Anaheim, California, 1997.

[3] W.J. Dally and J.W. Poulton. *"Digital System Engineering"*, Cambridge University Press, 1998.

[4] S. Hayashi and M.Yamada. "EMI-Noise Analysis Under ASIC Design Environment", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol.19, No.11, November, 2000.

[5] J. Plosila, P. Liljeberg, and J. Isoaho. "Pipelined On-Chip Bus Architecture with Distributed Self-timed Control", in *Proc. of IEEE International Symposium on Signals, Circuits and System*, July 2003.

[6] J. Sparso and S. Furber. *Principles of Asynchronous Circuit Design - A System Perspective.* Kluwer Academic Publishers, 2001.

[7] I. E. Sutherland. "Micropipelines. The 1988 Turing Award Lecture", *Communications of the ACM*, 32(6): pp.720-738, June 1989.

[8] T. Verhoeff. "Delay-insensitive codes - an overview", *Distributed computing*, 3(1):1-8, 1988.