

Crosstalk-Aware PWM-Based On-Chip Global Signaling in 65nm CMOS

Jae-sun Seo, Dennis Sylvester, and David Blaauw

University of Michigan, Ann Arbor, MI

Abstract

This paper proposes two crosstalk-aware signaling techniques based on pulse width modulation (PWM) for energy-efficient on-chip global busses. Two bits of information are encoded into transition type and pulse width for transmission over one wire. Measurements from 5mm on-chip links in 65nm CMOS show that the proposed schemes simultaneously achieve 15% performance improvement, 46% peak energy reduction, up to 25% average energy reduction, and >2X leakage reduction compared to conventional repeaters.

Introduction

Continuous technology scaling results in tighter wiring pitch with higher coupling capacitance and crosstalk noise, which directly impacts maximum clock frequency and chip power consumption. Previous work on pulsed signaling [1-4] showed improvements in energy and/or performance over conventionally repeated interconnect, but incurred peak energy consumption penalties [1], wide transmission lines [2], design complexity of dual supplies [3], or additional delay elements and signals [4]. We propose two crosstalk-aware encoding techniques to improve both performance and energy consumption in on-chip global interconnect. Exploiting the controllability of pulse widths in pulsed signaling, two bits of information are sent on one wire. By halving the number of wires in multi-bit busses, the wire width and spacing can be doubled for the same overall footprint, leading to a decrease in wire resistance and coupling capacitance, and thus improved delay and energy consumption.

Monotonic PWM and Hybrid PWM Scheme

Figure 1(a) shows the configuration of the conventional bus, where long interconnect between flip-flops is optimally repeated. The bus system for both proposed schemes is shown in Figure 1(b), where one wire replaces two wires of the conventional bus through an encoder and decoder. The concept of the two PWM-based signaling approaches is illustrated in Figure 2. Both proposed schemes use transition-based encoding to suppress power consumption with non-switching inputs. When the data is idle, the encoders of both mono-PWM and hybrid-PWM will stay idle as well, resulting in no transition.

The monotonic PWM scheme realizes pulsed signaling based on pulse width modulation. Figure 3 shows the encoder and decoder circuits of the mono-PWM scheme. When at least one bit switches, the encoder generates a pulse with three different pulse widths depending on the bit switching pattern, and this pulse propagates through the repeated interconnect. The decoder evaluates the pulse width and converts this information back to the original 2-bit data.

The hybrid PWM scheme combines single transition signaling with PWM-based signaling to further reduce energy consumption. In one of three switching cases, the hybrid-PWM encoder generates a single transition instead of a pulse, while the two remaining cases lead to two different pulse widths. Similar to the PWM-mono system, the decoder interprets the pulse width information and generates the original 2-bit data. The hybrid-PWM encoder and decoder circuits (Figure 4) are slightly more complex than the mono-PWM scheme since the quiescent state can be either high or low. However, the average energy consumption for random data is reduced due to the use of single transitions.

To minimize variability between the encoder and decoder in both schemes, identical variable delay chains are used and sized up to minimize RDF. Hence, the encoder pulse width and decoder delay track global PVT variation, but for calibration against local mismatch they are separately controllable by scan chain. Note that there is no clock overhead in the encoder and decoder in either scheme. Repeaters are optimally inserted in both schemes, but they cannot be skewed as aggressively as in [1,3] since the repeated wire cannot alter the pulse width appreciably. Therefore, all repeaters use a beta ratio of 2. Since the wires in the proposed schemes have much less coupling capacitance and resistance, the size and number of repeaters are significantly less than in the conventional bus.

Crosstalk-Aware Signaling

One of the main challenges in the proposed PWM-based signaling schemes is to mitigate the effect of crosstalk from adjacent wires on the transmitted pulse width. This challenge arises because the first edge

(always rising in mono-PWM) of the pulses are aligned between adjacent signals, but the second edge (always falling in mono-PWM) of the pulses can be separated depending on the switching pattern, thereby modulating the pulse width. Ignoring this effect would result in excessive guardbanding of pulse width margins, negating the speed and energy improvements. Crosstalk effects could be avoided by shielding each wire, at the cost of degraded delay and energy consumption compared to double pitch wires. To address this challenge without shielding, encoding circuits in Figures 3 and 4 are designed to enable crosstalk-aware signaling. Since pulse width dependency on data switching is deterministic, shorter pulses are generated through variable delay chains in the encoders if the pulse will be lengthened by crosstalk, and vice versa. The pulse shortening (or lengthening) for each data pattern is analyzed through extracted SPICE simulations and can be controlled digitally in our design for testability purposes. Also, the pulse width for each input case was chosen such that the shortened (or lengthened) pulse due to crosstalk will not overlap with the next pulse width, resulting in a pulse step of 80-100ps.

Measurement Results

8-b 5mm links using the conventional, mono-PWM, and hybrid-PWM schemes were fabricated in 65nm CMOS (Figure 9). The conventional scheme used M5 minimum pitch wires, and the proposed schemes used M5 2X minimum pitch wires for identical routing area. The peak energy versus delay characteristics are shown in Figure 5. Despite the encoder/decoder overhead, the considerable reduction in wire parasitics results in both a 15% delay improvement and 46% peak energy reduction in the mono-PWM scheme. The improvements are smaller in the hybrid-PWM scheme due to the additional logic to enable single transition and a peak MCF of 2. The bit error rate is less than 10^{-13} .

To measure average energy, traces from a microprocessor cache and memory bus were tested continuously through the in/out data registers. The data patterns were generated using the M5 full system simulator [5] for a GZIP workload from the Spec2000 benchmark suite. The results are shown in Figure 6, where up to 21% and 25% average energy reduction is achieved for the mono-PWM and hybrid-PWM scheme, respectively, compared to the conventional bus. The proposed schemes are observed to be more energy efficient in address traces since adjacent address bits tend to switch simultaneously. Overall, the hybrid-PWM scheme exhibits better average energy savings than mono-PWM scheme due to single transition signaling in 33% of the switching cases, which is most evident in the DATA1 trace of Figure 6.

An on-chip oscilloscope similar to the one proposed in [6] is implemented to capture waveforms at internal nodes. Figure 7 shows measured timing waveforms of the encoder output and decoder input in mono-PWM for different incoming data patterns. It can be seen that the pulse widths are well preserved across the 5mm link. The effect of crosstalk-aware signaling is shown in Figure 8. The y-axis depicts the pulse width change at the decoder input between the case when adjacent wires are idle and the case when adjacent wires are switching with different data. Crosstalk-aware signaling suppresses the effect of crosstalk on pulse width by 71%, enhancing the performance improvement of the proposed techniques. Leakage power is compared in Table 1, where the proposed schemes achieve >2X reduction over conventional repeaters primarily because the number of optimal repeaters is significantly reduced due to both fewer wires and lower resistance and capacitance of the double pitch wires.

The low sensitivity of the proposed system to global supply and temperature variation is shown in Figure 10(a). Functionality down to 700mV also demonstrates good robustness to process-induced mismatch, which is emphasized at low V_{dd}. To consider local supply voltage variation, different voltages are supplied to the encoder and decoder, and the contour plot of measured functionality at 40°C is shown in Figure 10(b). Calibrating the timing margin by altering the pulse widths trades off performance gains for robustness under supply variation.

Acknowledgement

The authors acknowledge fabrication support by ST Microelectronics.

References

[1] M. Khellah, et al., Symposium on VLSI Circuits, 2002.
 [2] P. Wang, et al., IEEE Transactions on VLSI, 2004.
 [3] H. Deogun, et al., ISLPED, 2006.

[4] D. Bojort, et al., Linköping University, 2005.
 [5] N. Binkert, et al., IEEE Micro, 2006.
 [6] S. Pant, et al., ESSCIRC, 2008.

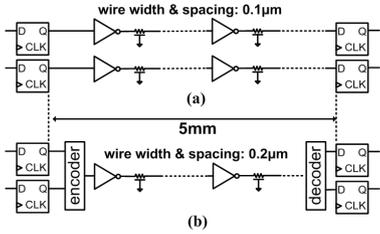


Figure 1. (a) Conventional bus with repeaters
 (b) Proposed PWM-based bus with repeaters
 (both schemes have same footprint).

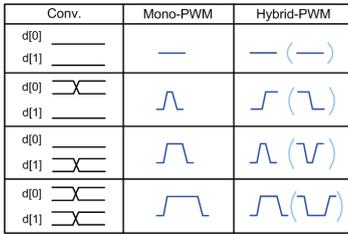


Figure 2. Proposed signaling schemes.

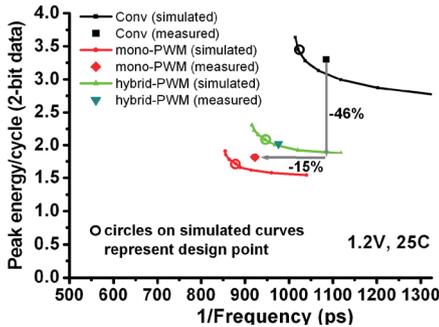


Figure 5. Peak energy and performance comparison.

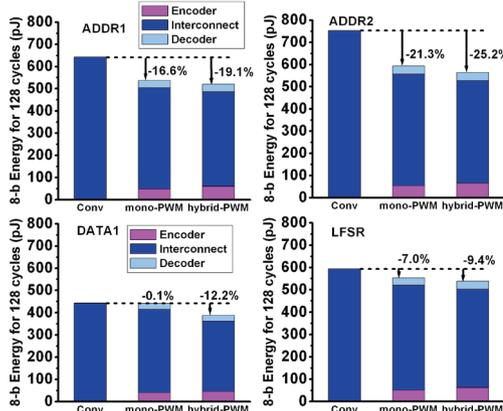


Figure 6. Average energy comparison for memory bus traces and LFSR.

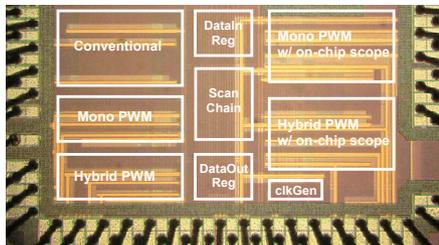


Figure 9. Die photo (1.5mmX0.7mm).

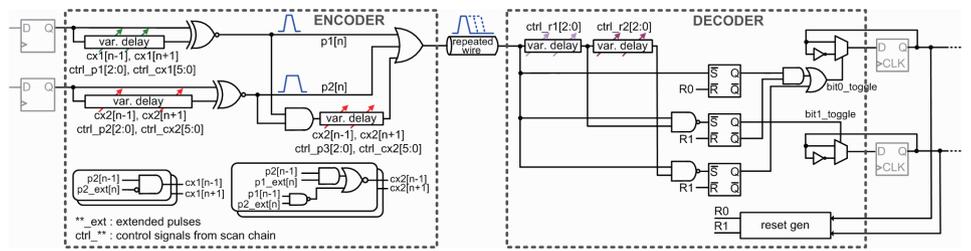


Figure 3. Monotonic PWM encoder and decoder circuit diagram.

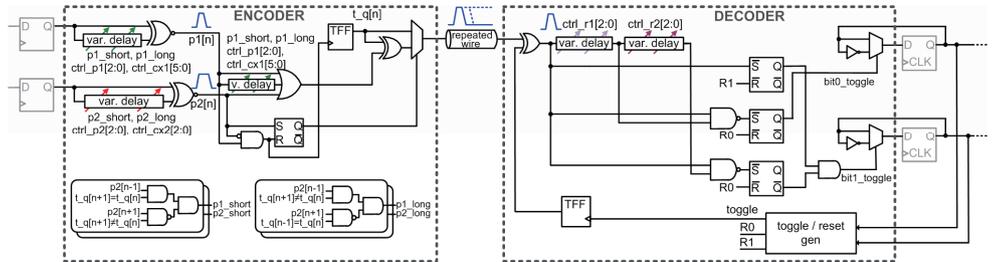


Figure 4. Hybrid PWM encoder and decoder circuit diagram.

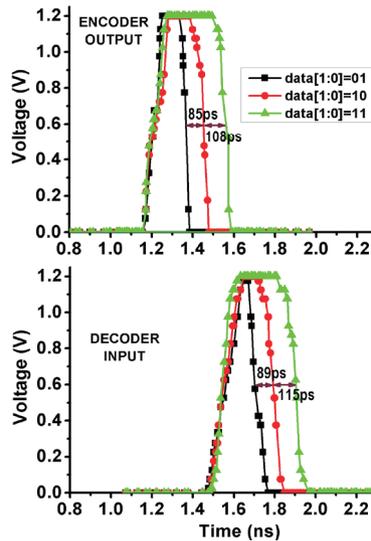


Figure 7. Measured waveforms of mono-PWM using on-chip oscilloscope.

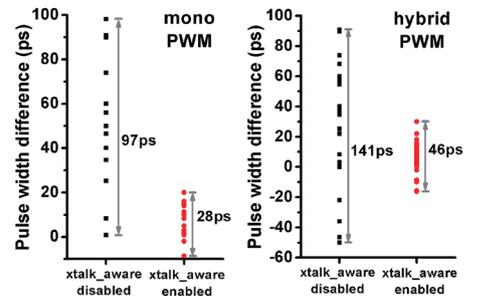


Figure 8. Comparison of crosstalk-aware signaling on decoder input pulse width spread for all possible data patterns.

Table 1. Leakage power measurement (units: µW).

	Enc.	Wire	Dec.	Total
Conv.	-	35.2	-	35.2
Mono-PWM	3.4	8.5	2.3	14.2 (-60%)
Hybrid-PWM	4.6	8.3	2.7	15.6 (-56%)

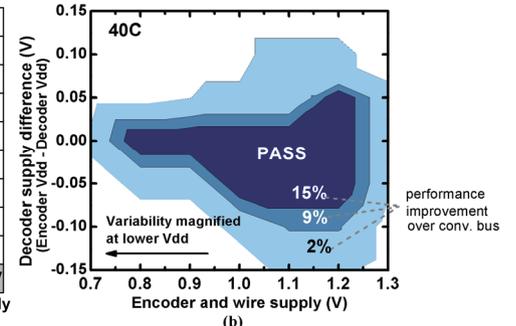
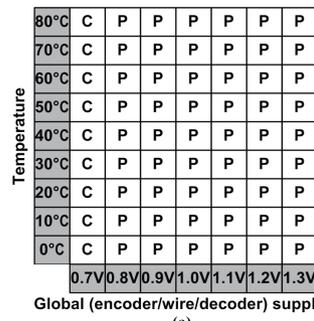


Figure 10. For variation experiments of mono-PWM, functionality is measured over 10^{11} bits. (a) Sensitivity of mono-PWM system to global voltage and temperature variation. 'P' represents pass with default control signals and 'C' represents pass with control signal adjustment. (b) Contour plot showing functionality and performance with local supply variation at 40°C. Additional guardbanding improves robustness at the expense of performance gains.