A 3-Gb/s/ch Transceiver for 10-mm Uninterrupted *RC*-Limited Global On-Chip Interconnects

Daniël Schinkel, *Student Member, IEEE*, Eisse Mensink, *Student Member, IEEE*, Eric A. M. Klumperink, *Member, IEEE*, Ed (A. J. M.) van Tuijl, and Bram Nauta, *Senior Member, IEEE*

***Abstract*—Global on-chip data communication is becoming a concern as the gap between transistor speed and interconnect bandwidth increases with CMOS process scaling. Repeaters can partly bridge this gap, but the classical repeater insertion approach requires a large number of repeaters while the intrinsic data capacity of each interconnect-segment is only partially used. In this paper we analyze interconnects and show how a combination of layout, termination and equalization techniques can significantly increase the data rate for a given length of uninterrupted interconnect. To validate these techniques, a bus-transceiver test chip in a 0.13-μm, 1.2-V, 6-M copper CMOS process has been designed. The chip uses 10-mm-long differential interconnects with wire widths and spacing of only 0.4 μm. Differential interconnects are insensitive to common-mode disturbances (e.g., non-neighbor crosstalk) and enable the use of twists to mitigate neighbor-to-neighbor crosstalk. With transceivers operating in conventional mode, the chip achieves only 0.55 Gb/s/ch. The achievable data rate increases to 3 Gb/s/ch (consuming 2 pJ/bit) with a pulse-width pre-emphasis technique, used in combination with resistive termination.**

***Index Terms*—Crosstalk, data bus, duty cycle, interconnect, intersymbol interference (ISI), on-chip communication, pre-emphasis, pulse-width, repeater insertion, transceivers.**

I. Introduction

ON-CHIP communication is getting more attention, as global interconnects are rapidly becoming a speed, power and reliability bottleneck for digital CMOS systems [1]. While gate speed increases under scaling, smaller cross-sectional wire dimensions will decrease the interconnect bandwidth for a given length. As pointed out in [1], a clear distinction should be made between local and global interconnects. The local wires connect gates inside a functional block and the length of these wires scales down together with the gates. Global wires do not scale down in length as the perimeter of large-scale digital ICs has remained roughly constant over different technologies. Technological advances such as copper interconnects and low-k dielectrics are by themselves not sufficient to keep the global interconnect bandwidth in pace with the advances in transistor speeds.

From a circuit design perspective, a general solution to the limited interconnect bandwidth is the use of repeaters, which make the repeated wire delay linear with length instead of the quadratic dependency of an unrepeated wire [2]. However, the number of repeaters should be kept to a minimum as they cost area and power and make floorplanning more difficult as portions of active area all over the chip have to be reserved for large repeater circuits. Furthermore, the classical approach to repeater insertion [2] using plain buffers/inverters as repeaters has serious limitations for global communication. With plain non-clocked buffers as repeaters, delay optimization requires closely spaced repeaters and delay variations due to crosstalk and due to process variations will accumulate and limit the achievable data rate. With such a classical repeater scheme, only a small portion of the intrinsic data capacity of each line segment is actually used.

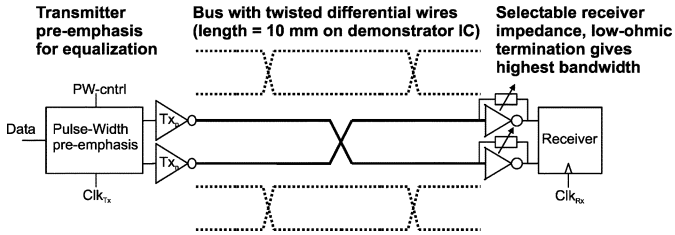

Fig. 1. Transceiver system overview.

These arguments motivate the search for more advanced solutions that can increase the data rate for a given length or can increase the unrepeated wire length for a given data rate, preferably in combination with a decrease in crosstalk sensitivity and power consumption.

Examples of recent advances in on-chip communication include [3] and [4]. Low-swing overdrive signaling over differential 10-mm aluminum interconnects is described in [3], but with the requirement of a dedicated supply and with clocked switches along the wire (increasing the already troublesome clock load). In [4], it has been proposed to use 16-μm-wide differential wires (20 mm long) and exploit the *LC* regime (transmission-line behavior) of these wires, but at the expense of a significant increase in power consumption and interconnect area. Both papers achieve 1 Gb/s/ch in a 0.18-μm CMOS technology.

It is shown in [5] that a new form of pre-emphasis—pulse-width pre-emphasis—enables high data rates without the need for a dedicated supply. In combination with low-ohmic termination and twisted differential interconnects, 3 Gb/s/ch is achieved over 10 mm of uninterrupted wire of only twice the minimum pitch. Fig. 1 shows an overview of the transceiver system. This paper discusses the motivations behind the various design choices of the presented transceiver. A detailed analysis of the interconnects (optimal dimensions, termination, twisting)

Manuscript received May 13, 2005; revised September 2, 2005. This work was supported by the Technology Foundation STW, applied science Division of NWO, and the technology programme of the Ministry of Economic Affairs, under project TCS.5791.

The authors are with the IC-Design Group, University of Twente, Enschede 7500 AE, The Netherlands (e-mail: d.schinkel@utwente.nl; e.mensink@utwente.nl).

Digital Object Identifier 10.1109/JSSC.2005.859880

is presented in Section II. Section III discusses pulse-width pre-emphasis and gives an analysis of the ideal pulse-width settings and robustness toward parameter variations. Section IV describes the implementation of the transceiver circuits. Section V compares the transceiver to classical repeater insertion. Section VI shows the results and compares them to predictions. Section VII gives the conclusions.

II. Interconnect Analysis and Dimensioning

In this paper, the communication structure is assumed to consist of point-to-point buses with all signals traveling in the same direction. For the demonstrator IC, the length of the bus is chosen to be 10 mm, to represent a typical global interconnect and allow for easy comparison with prior work. This section analyzes the interconnect and describes how the bandwidth of interconnects can be optimized and how crosstalk can be minimized.

A. Interconnect Model

Fig. 2 shows the model of the interconnects. The bus is placed in metal 5 as it is assumed that the thick top-metal (metal 6) is reserved for clock and power routing. The bus model is simulated with a 3-D EM-field solver to analyze the behavior of the interconnects and extract distributed *RLC* parameters. For 10-mm-long, 0.4-μm–wide wires these parameters are $R' = 0.15\ \mathrm{k}\Omega/\mathrm{mm}$, $L' = 0.25$ nH/mm and $C' = 0.23$ pF/mm ($C' = 0.27$ pF/mm for differential wires due to Miller-multiplication of the side-plate capacitance).

In the EM-field solver model, metal 4 and metal 6 plates approximate the capacitance of other perpendicular interconnects (assuming a Manhattan routing style), as a large-scale IC usually has a high wire density in all layers. In the actual demonstrator IC, ground- or V_{dd}-connected metal stripes are used in metal 4 and 6 to model the capacitance of these other interconnects.

Fig. 3 shows simulated transfer functions for single-ended interconnects with both low-ohmic and high-ohmic receiver termination (50 Ω transmitter impedance). Also included in the figure is the crosstalk transfer function from one wire to a direct neighbor. The transfer functions show three regions, where region I and II are caused by the *RC* behavior and region III by the *LC* behavior. An interesting aspect of *RC*-limited interconnects is that they have a single dominant pole, giving a high resemblance to a first-order roll-off in region one (this aspect will be exploited for the pre-emphasis). The higher-order part of the distributed *RC*-line transfer starts to dominate in region two. Only in the third region, for frequencies where $\omega L \geq R$ does the inductance begin to play a roll, creating the typical transmission line behavior. For these thin and long wires this frequency region is useless for data transmission as the attenuation is more than 150 dB.

The interconnect is *RC*-limited and the inductance can be neglected as long as the *RC* time constant $(\mathrm{length}^2 R'C')$ is much larger than the L/R time constant (L'/R'), which is true in our case for lengths larger than 0.7 mm. Lumped-element *RC* line models (100 lumps) are hence used for practical transceiver simulations; results from these lumped models are nearly indistinguishable from EM-field solver results.

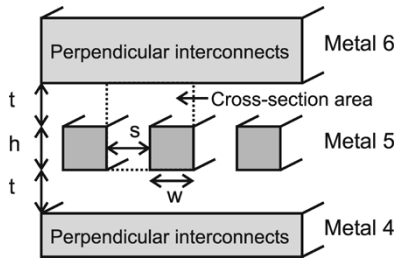

Fig. 2. Interconnect model.

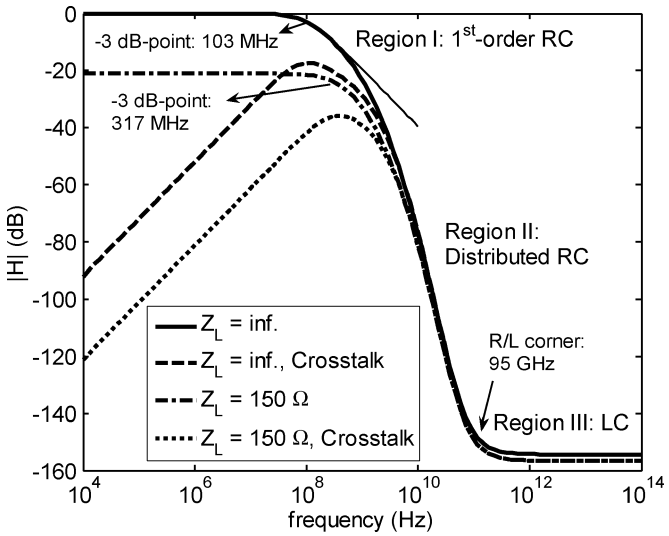

Fig. 3. Interconnect transfer functions and crosstalk transfer function for a 10-mm-long single-ended wire, terminated with either infinite resistance or with 150 Ω.

B. Bandwidth Versus Termination

The corner frequency of the dominant pole (and hence the -3-dB bandwidth) of an *RC*-limited interconnect depends on the impedance of the transmitter (R_S) and receiver (R_L) and can be approximated by [6]

$$\mathrm{BW}_{-3\ \mathrm{dB}} = \frac{1}{\frac{1}{2}R_{\mathrm{wire}}C_{\mathrm{wire}}\cdot 2\pi} \cdot \frac{R_S + R_{\mathrm{wire}} + R_L}{R_S + \frac{R_{\mathrm{wire}}}{3} + R_L + \frac{2R_S R_L}{R_{\mathrm{wire}}}} \quad (1)$$

The transmitter impedance R_S can be neglected, provided that a sufficiently large driver is used. The conventional form of receiver termination is a small capacitive load from a gate (modeled as an infinite impedance). With this form of termination, the bandwidth corresponds to the well known approximation $\omega_{-3\ \mathrm{dB}} = 1/(\frac{1}{2}R_{\mathrm{wire}}C_{\mathrm{wire}})$ [2] and is only 100 MHz for the example in Fig. 3 (80 MHz for differential wires). If, instead, a resistor is used as receiver termination with a value sufficiently lower than the wire resistance then the bandwidth can improve; up to a factor of three according to (1) in the limit of zero Ohm termination $(R_{\mathrm{S}} = R_L = 0)$. In practice, current-sensing amplifiers with low input impedance are used to create the resistive input impedance [7], [8].

C. Cross-Sectional Dimensions

The bandwidth of the interconnect depends of course also on its cross-sectional dimensions. The chosen interconnect width of 0.4 μm and spacing of 0.4 μm are optimized to give the

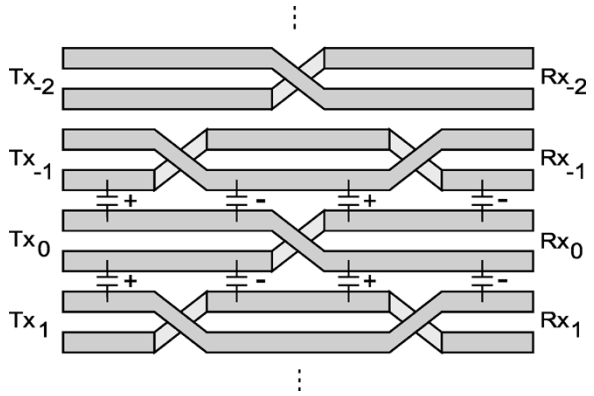

Fig. 4. Twisted differential bus.

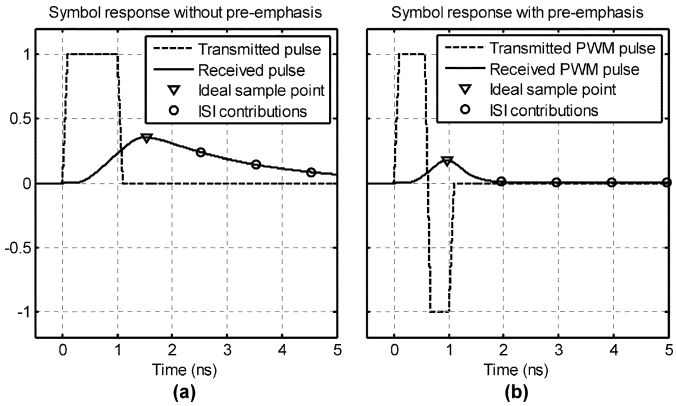

Fig. 5. Symbol responses of a (capacitively terminated) 10-mm interconnect with 1-ns symbol period.

highest bandwidth per cross-sectional area (BW/Area). A bus with these optimized interconnects will have the highest aggregate data rate for a certain bus area.

Analysis and EM-field simulations show that the BW/Area peaks when all the wire and spacing dimensions (w, h, s, and t in Fig. 2) are about equal. This result is illustrated with the simplified equations below (neglecting fringe-capacitance):

$$C' = 2C'_{\text{side}} + 2C'_{\text{topbottom}} \rightarrow C' \propto \left(\frac{h}{s} + \frac{w}{t}\right) R' \propto \frac{1}{wh} \quad (2)$$

$$BW \propto \frac{1}{R'C'} \qquad Area = (w+s)(h+t)$$
$$\frac{BW}{Area} \propto \frac{1}{R'C'} \frac{1}{(w+s)(h+t)} \quad (3)$$

$$\frac{BW}{Area} \propto \frac{wh}{\left(\frac{h}{s} + \frac{w}{t}\right)(w+s)(h+t)} = \frac{1}{\left(\frac{h+t}{s} + \frac{h+t}{w} + \frac{w+s}{h} + \frac{w+s}{t}\right)}. \quad (4)$$

The partial derivatives of (4) are all zero if $w = h = s = t$. Second-order effects such as fringe capacitance and differential signaling (which increases the C_{side}) give a minor alteration of the optimum. Simulations were used to fine-tune w and s (h and t are fixed by the process).

Most new technologies use a hierarchical wiring system with increasing wire thickness for higher metal layers. This is beneficial as the use of a thicker metal layer with larger interlayer dielectrics will give a higher bandwidth for a single interconnect (2), (3). With optimal dimensions ($d_{\text{opt}} = w = h = s = t$), the BW/Area is independent on d_{opt} (4). So the required data rate per single interconnect can determine the choice of metal layer, with little impact on aggregate data rate per cross-area.

D. Differential Signaling and Crosstalk Minimization

Crosstalk between different interconnects is a serious problem that decreases the integrity of the data at the receiving end. Crosstalk also limits the data rate as the ratio between crosstalk-interference to received signal power increases with frequency (see crosstalk transfer function in Fig. 3). Of the many types of crosstalk, crosstalk between neighboring wires in a bus has the most severe impact on reception. Fortunately, this form of crosstalk is also the most predictable and can be mitigated by proper use of twisted differential interconnects.

Other reasons to use differential signaling include the low offset of a differential sense-amplifier and the robustness toward common-mode disturbances (e.g., V_{dd} disturbances or crosstalk from perpendicular wires).

In [3], twisted interconnects are used for communication over global interconnects but the additional via resistance due to the eight twists was overlooked. In our demonstrator IC, we use only a single twist in the even channels and two twists in the odd channels, as shown in Fig. 4. It can be shown [9] that the optimal positions of these twists depend on the type of termination. With resistive termination, the suppression of crosstalk is most efficient and in this case, the optimal position of the twist in the even channels is at 50% of the length. This single twist mitigates the crosstalk from a differential aggressor to the differential-mode signal of the neighboring victims (with complete cancellation if equal transmitter and receiver resistance are used). The two twists in the odd channels are used to also mitigate crosstalk from a differential aggressor to the common-mode signal of the neighboring victims. This common-mode crosstalk is minimized if the twists are placed at about 30% and 70% of the length [9].

III. Pulse-Width Pre-Emphasis

Intersymbol interference (ISI), caused by the finite bandwidth of the interconnect is the primary limitation to the achievable data rate. A way to analyze ISI is to look at the response of a channel to a single isolated symbol [10]. Examples of such symbol responses are shown in Fig. 5, assuming bipolar signaling with normalized amplitude. Note that the zero level only occurs in the symbol response; a *sequence* of bipolar symbols would switch between one and minus one. Fig. 5(a) shows for example that plain binary signaling can not be used for reliable transmission at 1 GHz as the remaining energy of each symbol will give too much interference for the reliable detection of successive symbols.

Pre-emphasis (or de-emphasis) is a well-known equalization technique to reduce the amount of ISI and increase the data rate for a given channel bandwidth. The dominance of the first-

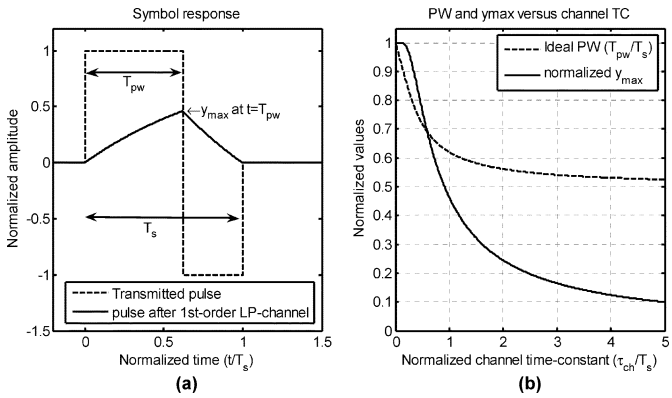

Fig. 6. Pulse-width pre-emphasis for a first-order channel.

order roll-off, as discussed in Section II-A, makes an on-chip interconnect very suitable for simple pre-emphasis transmission schemes (e.g., two-taps FIR). However, conventional pre-emphasis schemes (overdrive signaling), as used in interchip communication [10] are less suitable for on-chip implementation. As discussed in Section II-B, a low-ohmic driver impedance is desired to obtain a high interconnect bandwidth, while simple current-summing pre-emphasis transmitters have a high driver impedance. Voltage-mode overdrive transmitters on the other hand require the availability of additional supply voltages or require a large static current to create low output impedance, while slew-rate limits their equalizing performance at high speeds.

As a robust alternative, the use of pulse-width (PW) pre-emphasis is proposed [5]. As shown in Fig. 5(b), PW pre-emphasis can greatly reduce the amount of ISI by using the second part of the symbol-time to compensate for the remaining line charge. An advantage of a PW pre-emphasis circuit is that it only needs to switch between two voltage levels, which allows the use of simple transmitters (e.g., inverters) and reduces the influence of finite slew rates. The emphasis on timing accuracy instead of amplitude accuracy (conventional pre-emphasis) also facilitates the scaling to future deep-submicron, high-speed, low-voltage CMOS technologies. A drawback of PW pre-emphasis is the fact that the power consumption does not scale with data activity, as the transition inside the symbol always consumes power.

For first-order low-pass channels, PW pre-emphasis can completely cancel the ISI as shown in Fig. 6(a). The required pulse-width for zero ISI is a function of the symbol-time $(\mathrm{T_s})$ and of the time constant of the channel (τ_{ch}) and can be found by writing the symbol response as a summation of three step responses:

$$x(t) = \mathrm{step}(t) - 2\mathrm{step}(t - T_{\mathrm{pw}}) + \mathrm{step}(t - T_s). \quad (5)$$

The step response terms of a first-order channel are simple exponential functions (valid from the start of the step):

$$y(t) = \left(1 - e^{-\frac{t}{\tau_{\mathrm{ch}}}}\right)\Big|_{t\geq 0} - 2\left(1 - e^{\frac{-t+T_{\mathrm{pw}}}{\tau_{\mathrm{ch}}}}\right)\Big|_{t\geq T_{\mathrm{pw}}} + \left(1 - e^{\frac{-t+T_s}{\tau_{\mathrm{ch}}}}\right)\Big|_{t\geq T_s}. \quad (6)$$

The response will be zero for $t \geq T_s$, meaning no ISI, on the following condition:

$$e;^{-\frac{t}{\tau_{\mathrm{ch}}}}\left(-1 + 2e^{\frac{T_{\mathrm{pw}}}{\tau_{\mathrm{ch}}}} - e^{\frac{T_s}{\tau_{\mathrm{ch}}}}\right) = 0$$
$$\rightarrow \frac{T_{\mathrm{pw}}}{\tau_{\mathrm{ch}}} = \ln\left(\frac{1}{2} + \frac{1}{2}e^{\frac{T_s}{\tau_{\mathrm{ch}}}}\right) \rightarrow$$
$$PW;= \frac{T_{\mathrm{pw}}}{T_s} = \ln\left(\frac{1}{2} + \frac{1}{2}e^{\frac{T_s}{\tau_{\mathrm{ch}}}}\right)\frac{\tau_{\mathrm{ch}}}{T_s}. \quad (7)$$

So the ideal pulse-width is a function of the ratio between T_s and τ_{ch}. The maximum value of the response is found at $t = T_{\mathrm{pw}}$ (which is hence, the ideal detection instant) and can be found by substituting (7) into the first term of (6)

$$y_{\max} = y(T_{\mathrm{pw}}) = 1 - \left(\frac{1}{2} + \frac{1}{2}e^{\frac{T_s}{\tau_{\mathrm{ch}}}}\right)^{-1}. \quad (8)$$

For symbol times much smaller than the channel time constant, a high amount of de-emphasis is needed and the ideal pulse-width approaches 50% while $y_{\max}$ approaches zero as shown in Fig. 6(b). As an example, the ideal pulse-width for 2 Gb/s data rate, with a channel-corner frequency of 80 MHz ($\tau_{\mathrm{ch}}/T_{\mathrm{s}} = 2$ ns/0.5 ns $= 4$) is 53% and the receiver swing is only 12% of the transmitter swing.

To analyze the achievable data rate with higher-order channel transfer functions, a high-level numerical eye-diagram analysis has been carried out. A lumped *RC*-model (100 lumps) is used to simulate the symbol response, assuming a perfect transmitter with infinite slew rate. The receiver samples the received symbol a certain time after the start of the transmission, which defines the latency of the system. By summing the absolute values of the symbol response at multiples of the symbol interval after this sample moment, the worst-case amount of ISI is determined (see Fig. 5). Subtraction of the worst-case ISI from the sampled value gives the vertical eye-opening (eye-height) at the sample moment. By evaluating these results at different sample moments, both the ideal sample moment (latency) and the horizontal eye-opening (eye-width) can be determined. The analysis is repeated for different data rates and different pulse-widths, giving information such as eye-opening versus data rate.

Results of the analysis are shown in Fig. 7 for the case of a 10-mm differential interconnect with capacitive termination, with wire parameters as measured on the prototype ($R' = 0.19\ \mathrm{k\Omega/mm}$, $C' = 0.25$ pF/mm). Fig. 7(a) shows that without pre-emphasis, the eye at the receiver side will be completely closed at rates exceeding 600 Mb/s. With the use of PW pre-emphasis, the theoretically achievable data rate increases to about 4.2 Gb/s as shown in Fig. 7(b). However, at a rate of 2 Gb/s, the optimal pulse-width is only 53%. At this rate, the higher-order part of the channel transfer decreases the signal swing at the receiver (V_{pp}) to only 6%, instead of the 12% predicted by (7) and (8). Both the swing and the eye-opening relative to the swing rapidly decrease further for rates higher than about 2 Gb/s and effects such as receiver offset will start to degrade detection, making it nearly impossible to reach the theoretical limit of about 4 Gb/s.

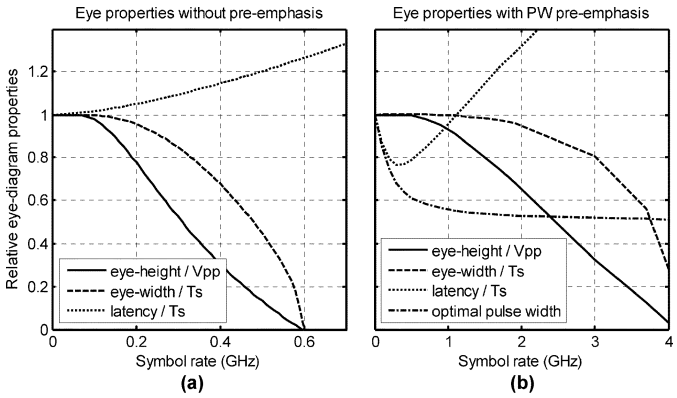

Fig. 7. Eye-diagram properties relative to receiver swing ($V_{\rm PP}$) or symbol time (T_s) for a capacitively terminated 10-mm interconnect.

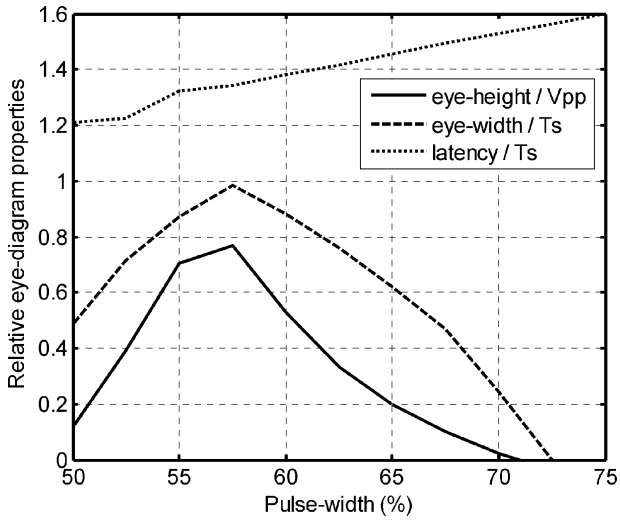

Fig. 8. Eye-diagram properties relative to receiver swing ($V_{\rm PP}$) or symbol time (T_s) for a resistively terminated 10 mm interconnect with a data rate of 2.5 Gb/s.

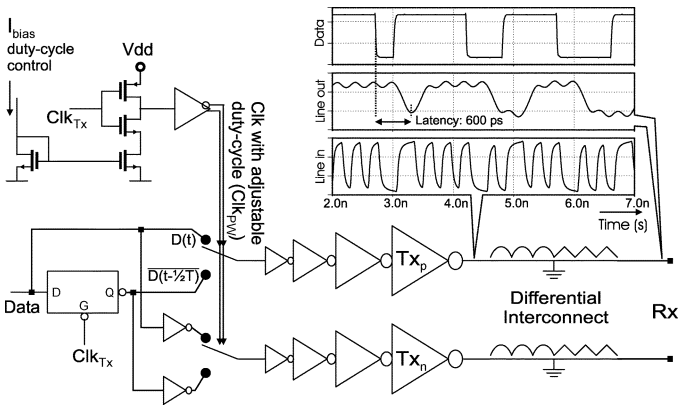

Fig. 9. Transmitter schematic and signal waveforms.

If a low-impedance current-sensing receiver is combined with PW pre-emphasis, then the theoretical achievable data rate increases to about 7 Gb/s. This increase is less than the factor of three predicted by (1), again due to the higher-order components in the wire transfer function. At a given data rate, the higher bandwidth of the resistively terminated interconnects leaves more room for mismatch between the time constant of the symbol-shape and the time constant of the wire. In applications, room for mismatch is necessary to be robust for variations of the line-length, spread in wire parameters and spread in actual pulse-width. The numerical analysis results in Fig. 8 illustrate how the pulse-width affects the eye-opening at a fixed data rate. For easy comparison with measurements, a data rate of 2.5 Gb/s with 150-Ω receiver termination is used and it is visible that the eye remains open for large variations of the pulse-width. At the optimal pulse-width of about 58%, the vertical eye-opening is about 75% of the swing. The actual voltage swing at the detector will be determined by the chosen gain of the current-sensing amplifier.

The eye-diagram analysis can be extended to include interference from crosstalk, showing that crosstalk would reduce the achievable data rate by about a factor of 1.5, in case differential wires without twists would be used.

IV. Implementation

An externally configurable demonstrator IC has been designed to validate the various techniques as described in the previous sections and to compare results with analysis. The schematic of the PW pre-emphasis transmitter is shown in Fig. 9, together with some signal waveforms. Conceptually, PW pre-emphasis involves the creation of a clock $(\mathrm{Clk_{PW}})$ with the correct duty cycle and XOR this clock with the incoming data. In the prototype IC, the duty cycle of $\mathrm{Clk_{PW}}$ is controlled by an external current source $(I_{\rm bias})$. The $I_{\rm bias}$ controls the slew rate of the falling edge of the output of an inverter, driven by a normal 50% duty cycle clock $(\mathrm{Clk_{Tx}})$. The controllable slew rate is converted to a controllable falling-edge delay by the second buffer.

The resulting clock with adjustable duty cycle selects either Data or not(Data), thereby implementing the XOR operation. A latch delays the not(Data) by half a clock-cycle to increase the timing margin. With this setup, the signals at the input of the switches (transmission gates) are stable during a transition of the $\mathrm{Clk_{PW}}$ and the Tx_p and Tx_n path have matched delay. To drive the wire, a cascade of four scaled inverters is used and the last inverter has an effective output resistance $R_{\rm out}$ of about 60 Ω. The size of the differential transmitter is about 300 μm^2. Dynamic latches, low-V_T transistors and small fan-outs ($\leq$3) are used to meet the target data rate of 3 Gb/s even at high temperature (100 °C) and at the slow process corner. Monte Carlo simulations show that transistor spread mainly causes common-mode offset with little change in latency and eye-opening. With a transmitter delay of 170 ps, the latency of the transmitter and channel amounts to 600 ps, as shown in Fig. 9.

The external current provides programmability. With $I_{\rm bias} = 0$, data is transmitted conventionally without PW pre-emphasis. At a 3-GHz clock an $I_{\rm bias}$ of 80, 200, or 400 μA results in transmitted symbols with pulse-widths of 75%, 58%, or 52%, respectively. The external control over the pulse-width allows for a comparison of the analysis from the previous sections with the results. In an actual application, the $I_{\rm bias}$ can be fixed at design time as the analysis indicates that the transmission scheme is robust toward (circuit and wire) parameter deviations if the data rate is chosen sufficiently below the theoretical limit. An automatic calibration or adaptation algorithm could also be used to

set the $I_{\rm bias}$ (for a bus) at run-time, but the benefit of a higher data rate would probably not outweigh the associated costs.

The schematic of the receivers is shown in Fig. 10. The input inverters use transmission gates as selectable feedback resistors, similar to [8]. In this way, either conventional capacitive termination or (active) resistive termination ($R_{\rm in} \approx 150\ \Omega$) can be selected. A regenerative sense amplifier (clocked comparator) followed by a dynamic latch samples the received data and restores it to full-swing.

With the transmission gates turned on, the input inverters behave as transimpedance amplifiers with an input impedance roughly equal to the $1/G_m$. The ratio between the feedback resistance $R_{\rm fb}$ of the transmission gates and the wire resistance $R_{\rm wire}$ controls the voltage gain from the transmitter to the input of the sense amplifier. A similar ratio $(R_{\rm wire} + R_{\rm fb})/R_{\rm wire}$ determines the output equivalent value of the offset voltage. The output-equivalent differential offset voltage of the transimpedance amplifiers is about 7.5 mV (one-sigma) and is comparable to the offset of the subsequent sense amplifier, giving a total one-sigma offset of about 10 mV. The design of the receiver has been optimized for a balance between minimal offset and maximal speed. The sense amplifier, as shown in Fig. 10, consists of a differential input pair and a cross-coupled pair with an NMOS reset transistor. The bias voltages V_{b1} and $\rm V_{b2}$ are generated locally with current mirrors from a single resistor-current as their exact value is not critical. The latch after the sense amplifier converts the regenerated data to full-swing data that is stable for a full clock period. The receiver adds only 50-ps delay, making the latency of the total transceiver equal to 650 ps at 3 Gb/s.

As with the transmitter, the receiver also uses low-V_T transistors to ensure correct operation at 3 Gb/s over different process corners. Only the input inverters are normal-V_T, as a lower overdrive voltage improves the G_m versus current ratio. The total size of the prototype receiver is about 1000 μm^2, not optimized and including the reference circuits.

V. Comparison With Repeaters

A classical repeated single-ended interconnect system has been simulated in the same technology and Table I shows a comparison with the presented transceiver. The length of the interconnect segments and the size of the drivers in the repeated system are optimized for minimal delay [2]. This optimization requires as much as ten repeaters, each with a driver size (20 μm NMOST width) that is larger than the single driver used in this work (7 μm NMOST width). The wire dimensions are in both cases equal to the optimized values (0.4 μm width and 0.4 μm spacing). This would amount to roughly the same wiring resources per channel for both systems, as the repeater system needs shields between the signal lines to avoid a severe eye degradation due to neighbor to neighbor crosstalk (see Table I).

Although the power consumption of the repeated system is modest (3.1 pJ/transition, giving 1.6 pJ/bit with 50% data activity), the many repeaters need much layout resources and the long chain of inverters creates a high static variation in delay (430 ps) for different process, voltage, and temperature (PVT)

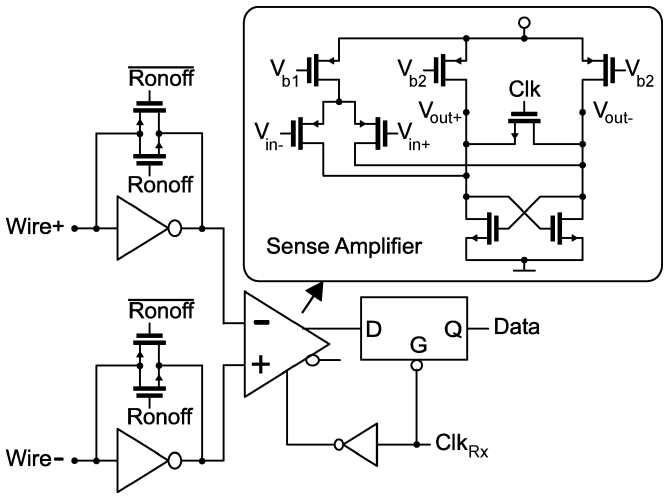

Fig. 10. Receiver schematic.

TABLE I
Comparison of Presented System Versus Repeater System (Simulation Results)

	Repeater system	This work
Wire segment length	1mm (near optimal)	10mm
Driver / Repeater NMOS width	20µm (optimal)	7µm
Energy consumption	3.1pJ/transition	2pJ/bit
Nominal delay (50% crossing)		
PW pre-emp circuit	n.a.	80ps
Transmitter driver cascade	120ps	90ps
Interconnect	680ps (10 segments)	270ps
Total	800ps	440ps
Delay variation due to neighbor-to-neighbor crosstalk	-120ps to +140ps (without shielding)	none
PVT delay spread		
Random mismatch (one-sigma)	20ps	4ps
Slow process, Temp = 90°C, Vdd = 1.1V	+280ps	+100ps
Fast process, Temp = 0°C, Vdd = 1.3V	-150ps	-50ps

corners. Without additional measures, the data rate should be lower than 1/430 ps $=$ 2.3 Gb/s to keep the delay variation within one clock-cycle over all corners.

Larger transistors in combination with fewer repeaters could reduce the effect of PVT variations on delay, but at the expense of an increase in power or an increase in nominal delay. With a receiver that samples data at the centre of the eye, half a symbol-time after the 50% crossing, the latency is already 800 ps + 165 ps $=$ 965 ps, while the latency of the presented transceiver is only 650 ps (including the receiver). The latency of the presented transceiver is also much less affected by PVT variations.

In their current form, both systems need special receiver clocking strategies as the latency is higher than one clock-cycle. On the demonstrator IC, the receiver clock is supplied externally to be able to change its phase relative to the transmitter clock and measure the eye-width. In an application, one could transmit clock information alongside the data bus (source-synchronous) or one could choose to use shorter wire segments and pipeline the communication. In that case, the presented system would require far fewer pipeline stages than the conventional (clocked) repeater system as the presented techniques increase the achievable data rate and lower the latency for a given length of interconnect.

VI. Experimental Results

A. Measurement Setup

The micrograph of the demonstrator IC is shown in Fig. 11. The chip has been fabricated in a standard 1.2-V, 6-M, 0.13-μm CMOS process with copper interconnects. A seven-channel differential bus with twisted wires (width and spacing of 0.4 μm each, optimized as explained in Section II) is placed in metal 5 and is completely surrounded by GND/V_{dd}-connected metal stripes. An additional seven-channel single-ended bus with perpendicular orientation is placed below the differential bus for additional characterization purposes (a variety of wire pitches is used in this bus) and to provide an indication of interlayer crosstalk. An external single-channel 3.2 Gb/s pattern generator/analyzer is used for the data generation and BER measurement. Large on-chip delay lines (chains of flip-flops, ten per channel) provide all bus channels with pseudo-independent data. This setup allows for random-data BER testing in a realistic crosstalk environment while deterministic data patterns (for e.g. step response measurement) can also be applied.

Different twisting patterns and receiver configurations are used for the different channels of the differential bus, as shown in Fig. 12. Channels 1, 4, and 6 are equipped with 50 Ω output buffers and pads for measurements. The output buffers can accommodate a full-swing input range, with some large-signal compression. They attenuate the signal by about 6 dB (small-signal) to 9 dB (large-signal). Channel 4 is used for the BER measurements while the other two channels are used for e.g. crosstalk and eye-diagram measurements. The receiving ends of the single-ended bus interconnects are directly connected to pads to enable measurements directly on the interconnect. The chip has been measured in a probe station using 50 Ω GSSG probes for the high-speed signals. At the receiver side, dedicated GSSG pads are available for the various channels to enable wide-band measurements directly on the specific channel.

B. Performance

The measured interconnect parameters are $R' = 0.19\ \mathrm{k\Omega/mm}$ and $C' = 0.25\ \mathrm{pF/mm}$ (for a differential interconnect). These values agree with the EM-field simulations, given the tolerance bounds of the process. Indirect measurement of the capacitance suggests that it is composed of roughly 0.05 pF/mm to each of the four sides; the part of the capacitance between the differential wires is doubled due to Miller multiplication.

The configurability of the transmitter and receiver enables measurements with or without PW pre-emphasis and with capacitive or resistive termination. Eye-diagrams for each of the four settings are shown in Fig. 13, as measured at the output of channel 6. BER measurements (with PRBS data patterns) for the four settings were carried out at channel 4 and Table II shows the highest data rate at which bit-errors are not yet measurable $(\mathrm{BER} < 1\mathrm{e}^{-12})$. At the boundary of error-free operation the BER drops sharply, as the primary bit-error sources are deterministic (ISI) or static (offset) and a BER much lower than $1\mathrm{e}^{-12}$ is expected at the shown data rates.

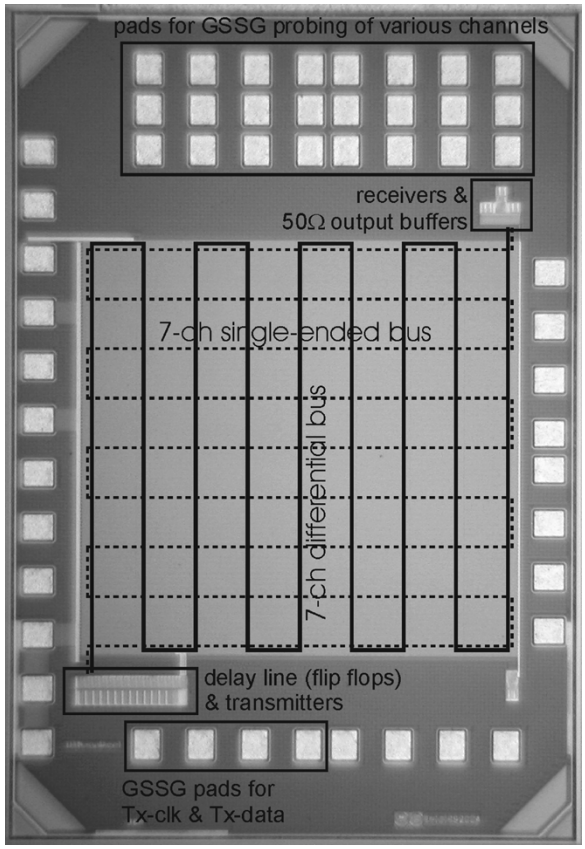

Fig. 11. Chip micrograph.

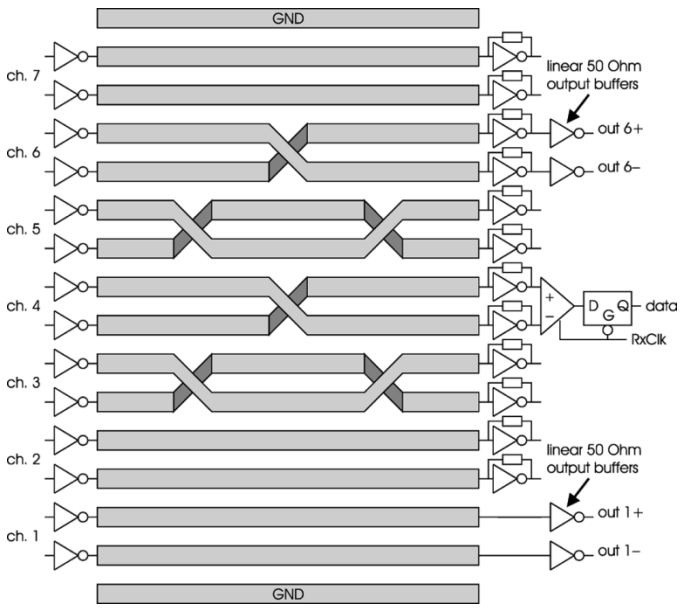

Fig. 12. Differential bus configuration as implemented on demonstrator IC.

The transceiver circuits of channel 4 have a dedicated supply, and the energy consumption of the channel is also shown in Table II (measured with PRBS data patterns, giving 50% data activity). Simulated values for the energy consumption of the various parts of the transceiver are also shown. The Tx and Rx circuits consume more power than necessary for a given mode

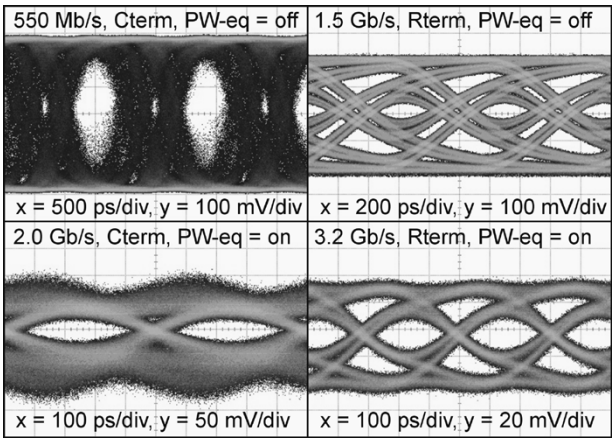

Fig. 13. Eye-diagrams for various transceiver settings. The output buffers compress the vertical scale by 6 to 9 dB.

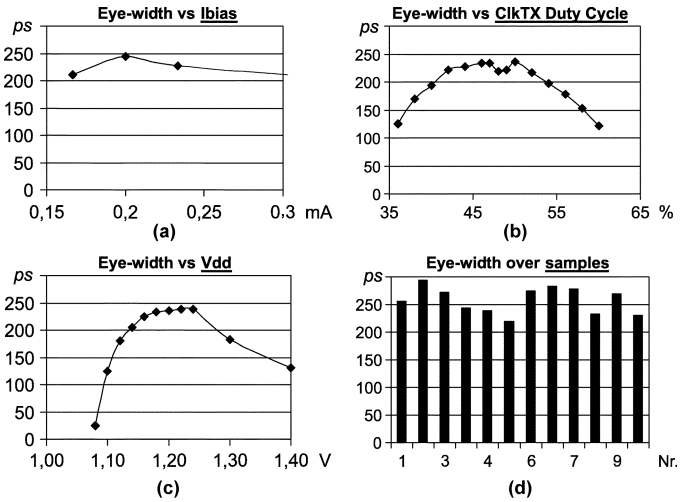

Fig. 14. Measured eye-width versus parameters and over different samples at 2.5 Gb/s.

TABLE II
ACHIEVABLE DATA RATE (BER $< 1e^{-12}$) AND ENERGY CONSUMPTION; BETWEEN BRACKETS, SIMULATED ENERGY CONSUMPTION VALUES ARE SHOWN FOR THE TRANSMITTER (Tx), WIRE, TRANSIMPEDANCE AMPLIFIER (TIA) AND SENSE AMPLIFIER WITH LATCH (SA)

	Capacitive termination	Resistive termination
Without PW pre-emphasis	550 Mb/s 3.4 pJ/bit (Tx: 0.2pJ/b, Wire: 1.2pJ/b, TIA: 0.7pJ/b, SA: 1.4pJ/b)	1.5 Gb/s 2.5 pJ/bit (Tx: 0.2pJ/b, Wire: 0.8pJ/b, TIA: 0.9pJ/b, SA: 0.5pJ/b)
With PW pre-emphasis	2 Gb/s 2.5 pJ/bit (Tx: 0.5pJ/b, Wire: 0.8pJ/b, TIA: 0.7pJ/b, SA: 0.5pJ/b)	3 Gb/s 2 pJ/bit (Tx: 0.5pJ/b, Wire: 0.6pJ/b, TIA: 0.5pJ/b, SA: 0.3pJ/b)

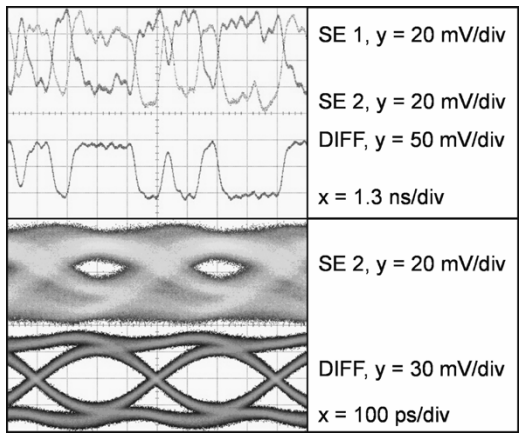

Fig. 15. Effect of crosstalk on the single-ended and twisted differential output of channel 6 at 2.5 Gb/s. On-chip signals are about 6 dB larger.

of operation, as they are designed to function in all modes and are optimized for speed.

The results show good agreement with the analysis. The 550 Mb/s achieved in the conventional case is only slightly lower than the theoretical limit of 600 Mb/s. Resistive termination improves the achievable data rate by nearly a factor of three. The improvement of PW pre-emphasis together with conventional termination is a factor of four and is a factor of two if used in combination with resistive termination. The eye at the receiver is still open at 3.2 Gb/s as visible in the bottom right of Fig. 13, but the opening is so small (40 m$V_{\rm pp}$) that effects such as hysteresis and offset in the clocked receiver prevent reliable detection at this data rate $(\text{BER} = 10^{-8})$. At 3 Gb/s, error-free operation is possible for all ten measured samples ($I_{\rm bias} = 400\ \mu$A) but without much $V_{\rm dd}$ or $I_{\rm bias}$ tolerance. At 2.5 Gb/s, the design is robust and the BER remains immeasurable with large external parameter deviations. Fig. 14 illustrates this robustness by plotting the measured eye-width as a function of an external parameter while keeping the other parameters at their nominal value ($V_{\rm dd} = 1.2$, $\text{Clk}_{\rm TX}$ duty cycle $= 50\%$ and $I_{\rm bias} = 200\ \mu$A). To measure the eye-width, a phase-shifter was used to vary the skew of the receiver clock and find the phase-shifts where the BER just becomes measurable. The optimal bias current of 200 μA (giving a PW pre-emphasis duty cycle of about 58%) agrees with predictions from Fig. 8. The measured relationship between the external $\text{Clk}_{\rm TX}$ duty cycle and the eye-width also behaves as expected, except for a small drop in eye-width around 50% $\text{Clk}_{\rm TX}$ duty cycle which can probably be attributed to measurement tolerances. The highest measured eye-width of 250 ps is lower than the theoretical value of almost 400 ps due to the required setup and hold times of the sense amplifier.

The influence of crosstalk on the eye-diagram is shown in Fig. 15. This figure shows both the output of the single-ended (SE) halves and the differential output of channel 6 at a rate of 2.5 Gb/s. Each SE half of channel 6 receives crosstalk mainly from the wire-piece that runs alongside channel 7 (but the other channels in the bus and the perpendicular bus also generate some common-mode crosstalk). The eye-closure due to the crosstalk in the single-ended output is clearly visible in the figure, while the crosstalk is mitigated in the differential output. If the twist in channel 6 would not be present, then the crosstalk on both SE halves would be even higher and it would not be canceled in the differential output.

VII. Conclusion

Techniques to improve global on-chip data communication for given lengths of uninterrupted interconnect are presented and analyzed in this paper. A transceiver system has been implemented on a demonstrator IC with 10 mm long, 0.4 μm wide, twisted differential interconnects as robust channels that are insensitive to crosstalk. Measurements show that the data rate of 0.55 Gb/s/ch, obtained with the transceiver operating in conventional mode, can be increased to 3 Gb/s/ch (2 pJ/bit) with a combination of pulse-width pre-emphasis and resistive termination. At 2.5 Gb/s/ch, the system is tolerant to parameter deviations. Analysis, such as the predicted tolerance in pulse-width, agrees well with measurements.

The presented pulse-width pre-emphasis technique is a simple and robust equalizing approach, suitable for advanced deep-submicron processes. The technique can also be applied to interchip or wire-line [12] communication.

The power consumption (6 mW at 3 Gb/s) of the transceiver in its current form has little dependence on data activity. However, the transmitter and receiver circuits are well suited for power management as the speed-enhancing, but power consuming techniques can be easily turned on and off dynamically [8].

When compared to classical repeater techniques targeted at comparable data rates, the presented transceiver can bridge an uninterrupted wire length that is a factor of ten higher and has a lower and more predictable latency (650 ps versus 965 ps). Pulse-width pre-emphasis and resistive termination should enable much higher data rates for shorter lengths of uninterrupted interconnect (or for interconnects placed in higher/larger metal layers), as long as the wire bandwidth is the bottleneck. The presented techniques can thus be used for repeaterless global communication or can be used to improve the trade-off between data rate and repeater spacing.

Acknowledgment

The authors would like to thank Philips Research for chip fabrication.

References

[1] R. Ho, K. W. Mai, and M. A. Horowitz, "The future of wires," *Proc. IEEE*, pp. 490–504, April 2001.

[2] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.

[3] R. Ho, K. W. Mai, and M. A. Horowitz, "Efficient on-chip global interconnects," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2003, pp. 271–274.

[4] R. Chang, N. Talwalkar, C. Yue, and S. Wong, "Near speed-of-light signaling over on-chip electrical interconnects," *IEEE J. Solid-State Circuits*, vol. 38, no. 5, pp. 834–838, May 2003.

[5] D. Schinkel, E. Mensink, E. A. M. Klumperink, A. J. M. van Tuijl, and B. Nauta, "A 3 Gb/s/ch transceiver for RC-limited on-chip interconnects," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2005, pp. 386–387.

[6] E. Seevinck, P. van Beers, and H. Ontrop, "Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAMs," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 525–536, Apr. 1991.

[7] A. Katoch, E. Seevinck, and H. Veendrick, "Fast signal propagation for point to point on-chip long interconnects using current sensing," in *Proc. ESSCIRC*, Sep. 2002, pp. 195–198.

[8] R. Bashirullah, W. Liu, R. Cavin, and D. Edwards, "A 16 Gb/s adaptive bandwidth on-chip bus based on hybrid current/voltage mode signaling," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2004, pp. 392–393.

[9] E. Mensink, D. Schinkel, E. A. M. Klumperink, A. J. M. van Tuijl, and B. Nauta, "Optimally-placed twists in global on-chip differential interconnects," in *Proc. ESSCIRC*, Sep. 2005, pp. 475–478.

[10] R. Farjad-Rad *et al.*, "A 0.4-μm CMOS 10-Gb/s 4-PAM pre-emphasis serial link transmitter," *IEEE J. Solid-State Circuits*, vol. 34, no. 5, pp. 580–585, May 1999.

[11] H. Tenhunen and D. Pamunuwa, "On dynamic delay and repeater insertion," in *Proc. IEEE ISCAS*, May 2002, pp. I-97–I-100.

[12] J. H. R. Schrader, E. A. M. Klumperink, J. L. Visschers, and B. Nauta, "CMOS transmitter using pulse-width modulation pre-emphasis achieving 33 db loss compensation at 5-Gb/s," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2005, pp. 388–391.

Daniël Schinkel (S'03) was born in Finsterwolde, The Netherlands, in 1978. He received the M.Sc. degree in electrical engineering (with honors) from the University of Twente, The Netherlands, in 2003. He is currently pursuing the Ph.D. degree on the subject of high-speed on-chip communication at the same university.

During his studies, he worked on various occasions as a trainee at the Mixed-Signal Circuits and Systems Department of Philips Research, Eindhoven, The Netherlands. This work resulted in a number of publications and two patent filings. His research interests include analog and mixed-signal circuit design, sigma-delta data converters, class-D power amplifiers, and high-speed communication circuits.

Eisse Mensink (S'03) was born on January 10, 1979, in Almelo, The Netherlands. He received the M.Sc. degree in electrical engineering (*cum laude*) from the University of Twente, The Netherlands, in 2003. He is currently working toward the Ph.D. degree at the same university on the subject of high-speed on-chip communication.

Eric A. M. Klumperink (M '98) was born on April 4, 1960, in Lichtenvoorde, The Netherlands. He received the B.Sc. degree from HTS, Enschede, The Netherlands, in 1982.

After a short period in industry, he joined the Faculty of Electrical Engineering of the University of Twente, Twente, The Netherlands, in 1984, where he was mainly engaged in analog CMOS circuit design and research. This resulted in several publications and a Ph.D. thesis, in 1997, on the subject of "Transconductance Based CMOS Circuits." He is currently an Assistant Professor at the IC-Design Laboratory and also involved in the MESA+ Research Institute. He holds four patents and authored and coauthored more than 50 journal and conference papers. His research interest is in design issues of HF CMOS circuits, especially for the front-ends of integrated CMOS transceivers.

Dr. Klumperink was a co-recipient of the ISSCC 2002 Van Vessem Outstanding Paper Award.

Ed (A. J. M.) van Tuijl (M'97) was born in Rotterdam, The Netherlands, on June 20, 1952.

He joined Philips Semiconductors, Eindhoven, The Netherlands, in 1980. As a Designer, he worked on many kinds of small-signal and power audio applications, including A/D and D/A converters. In 1991, he became Design Manager of the audio power and power-conversion product line. After many years at Philips Semiconductors he joined Philips Research Eindhoven, The Netherlands in 1998 as a Principal Research Scientist. In 1992, he joined the University of Twente, Enschede, The Netherlands, as a part-time Professor. His current research includes data conversion, high-speed communication and low-noise oscillators. He is an author or co-author of many papers and holds many patents in the field of analog electronics and data conversion.

Bram Nauta (M'91–SM'03) received the M.Sc. degree (*cum laude*) in electrical engineering and the Ph.D. degree from the University of Twente, Enschede, The Netherlands in 1987 and 1991, respectively. His dissertation focused on analog CMOS filters for very high frequencies.

In 1991, he joined the Mixed-Signal Circuits and Systems Department, Philips Research, Eindhoven, The Netherlands, where he worked on high-speed A/D converters. Starting in 1994, he led a research group in the same department, working on analog key modules. In 1998, he returned to the University of Twente as a full Professor heading the IC Design Group in the MESA+ Research Institute and Department of Electrical Engineering. His current research interest is analog CMOS circuits for transceivers. In addition, he is also a part-time consultant in industry and, in 2001, he co-founded ChipDesignWorks. His Ph.D. dissertation was published as the book *Analog CMOS Filters for Very High Frequencies* (Kluwer, 1993). He holds 11 patents in circuit design.

Prof. Nauta served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II, ANALOG AND DIGITAL SIGNAL PROCESSING from 1997 to 1999, and in 1998 he served as Guest Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS. In 2001, he became an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS, and he is a member of the technical program committee of ESSCIRC and ISSCC. He was the co-recipient of the ISSCC 2002 Van Vessem Outstanding Paper Award, and he was the recipient of the Shell Study Tour Award for his Ph.D. work.