A CDMA and PAM Signaling Interconnect Architecture

Goran S. Jovanović[1], Tatjana R. Nikolić[2] and Mile K. Stojčev[3]

***Abstract* – The system performance of complex CMOS VLSI ICs are limited nowadays by its interconnect bandwidth in both on-chip and off-chip communications. In order to improve the overall system performance without drastically increasing communication resources and complexity, communication protocol, and signaling technology that exploits parallelism at on- and off-chip system level, resulting in higher concurrency on a shared bus or direct link, a merged CDMA and PAM interconnect is proposed in this paper. The CDMA logic uses different PN codes to separate the information carried on different bus lines over shared bus, while PAM approach is based on multilevel signaling, so their combination offers an interesting potential solution. The transceiver circuit has been developed using Xilinx FPGA technology for realization of the CDMA coder/decoder, and 0.35 μm CMOS four layers process technology for implementation of the 4_PAM coder/decoder block. Simulation results show that the effective communication bandwidth is 1.1 Gbps, with an average power consumption of 25 mW per single channel at 200 MHz operating frequency and power supply of 3.3 V.**

***Keywords* – Signaling Interconnect, CDMA&PAM shared bus, SoC**

I. INTRODUCTION

The digital system performance consists of two parts: computation performance and communication performance. With rapid development in CMOS technology scaling, the computation performance of a chip has been increased drastically. As computation performance goes up, the required communication throughput needs also increase with the same rate. However, until now, the communication bandwidth has been scaling much more slowly. Therefore, the communication between multiple on- or off-chip semiconductor intellectual property, IP, blocks and electronic modules is becoming a dominant cost, performance, and power factor in contemporary electronic systems [1-2]. To achieve a high data transfer rate, the bus bandwidth must be increased. Increasing the bus bandwidth, however, increases the pin count and enlarges the chip area. It also leads to complicated routing between different IP blocks within the VLSI IC and among modules on the same printed circuit board.

The concept of transferring multiple bits over each symbol through modulation techniques has been proposed to efficiently solve the aforementioned problems [2-3]. Code Division Multiple Access, CDMA, and Pulse-Amplitude Modulation, PAM, technique that incorporate multilevel amplitudes rather than binary signals, with order to increase data rate, is proposed in this paper. The basic idea is as follows: If two CDMA coded consecutive bits in the sequence are grouped and converted to one of four levels, 4-PAM signaling, then each level is twice as long as a bit period, demanding only half the bandwidth required for transmission of the binary system.

II. TAXONOMY OF ON-CHIP COMMUNICATION

On-chip communication architectures, see Fig. 1, can be divided into the following three main classes:

a) Point-to-point (P2P) interconnects - pairs of processing units communicate directly over dedicated physically wired connections. Because of its simplicity, this architecture has been widely adopted in many applications. P2P communication architecture can provide the utmost in communication performance at the expense of dedicated channels among all the communicating IP pairs. However, these architectures suffer from lack of scalability in terms of high complexity, cost, and design effort. Custom interconnect (see Fig. 1), sometimes referred as ad-hoc interconnect, is simply connecting processing elements by wires when there is a necessity. On the other hand, uniform interconnect often has well defined interconnect topology, which can be precisely specified by equations or graphs [4], [5].

b) Bus architectures - long wires are grouped together to form a single physical communication channel, which is shared among different logical channels. The bus consists of a set of shared parallel wires to which various components are connected. An arbitration mechanism is used to control sharing of the bus. Only one component on the bus can have control over the shared wires at any given time in order to perform data transfer. A hierarchical shared bus (see Fig. 1) defines segmented bus architecture. Bus segments are connected via a bridge, which may buffer data. An alternative approach is to use a split-bus. This refers to a set of custom-design segmented buses. Bus segments can be interconnected in an ad-hoc manner, or based on a systematic approach [4], [5], [6].

[1]Goran S. Jovanović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedova 14, 18000 Niš, Serbia, Email: goran.jovanovic@elfak.ni.ac.rs

[2]Tatjana R. Nikolić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedova 14, 18000 Niš, Serbia, Email: tatjana.nikolic@elfak.ni.ac.rs

[3]Mile K. Stojčev is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedova 14, 18000 Niš, Serbia, Email: mile.stojcev@elfak.ni.ac.rs

c) Network-on-Chip (NoC) - is an architecture inspired by data communication networks, such as LANs and WANs. Inter-processor communication is achieved by sending message packets between IP blocks using an on-chip packet-switched network [3], [7], [9].

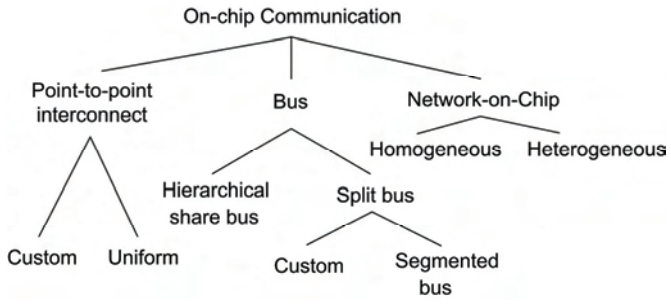

Fig. 1. Taxonomy of on-chip communication architectures

III. WIRES AND INTERCONNECTS

As interconnects have become a key limiting factor in VLSI design in deep submicron regime, chip design must first consider the constraints of interconnects in each design stage. A direct way to increase the information carrying capacity of an interconnect is to simply make the interconnect wider or faster. However, wider buses require more interconnect resources and packaging pins, and the extra drivers use more silicon real estate, dissipate more power and create more noise; faster speed leads to higher power dissipation, noise and radiation can be worse. All these issues must be considered in early decisions of bus width and speed [1], [2].

Differential signaling offers the highest signal integrity in which the common-mode noise introduced by return paths is rejected by the differential receiver. It forms the foundation of many high-performance signaling standards, including LVDS and ECL. The price paid for this is more complex drivers and receivers plus twice the amount of wires. Besides, mixed standards are used. For example, in many systems the clock is transmitted fully differential to minimize clock skew, but the signals are not differential, such signals are termed single-ended [1], [9].

Although there are different high-level abstract interconnect schemes, such as point-to-point interconnect, bus architecture, switched network, etc, the underlying physical global wire structure will be the same, namely the parallel wire structure. For point-to-point connections, each link is realized by a certain number of parallel wires; for bus architecture, the bus itself is a number of wires in parallel and the same applies to a link connecting two switches in a network-on-chip. Therefore it is of great importance to study the interconnect in general, the parallel wire structure in particular, and model them accurately. The following discussion serves this purpose.

IV. SIGNALING OVER PARALLEL WIRES

The most common method of reducing the delay over long interconnects is to insert repeaters (inverters) at appropriate points. However, delay is not the only concern associated with the interconnect. Another major issue is the bandwidth supported by the interconnect under certain constraints, such as limited area, limited power consumption, and limited freedom in choosing repeater insertion strategy. In the sequel we study how delay and bandwidth are related and derive a suitable bandwidth under different constraints.

Ideally future interconnect systems must encompass the following important features:

- ultra high data rates, usually > 100 Gbps,
- concurrent multi input-output service for simultaneous and bidirectional communications on a shared transmission medium,
- real-time re-configurability in connectivity and bandwidth for optimized channel efficiency and fault-tolerance, and
- the fabrication of interconnect systems must be compatible with the current SoC and SiP (System in Package) technologies for low-cost system production [1], [9].

The concept of transferring multiple bits over each wire has been proposed to solve the problem of high data transfer rate. This concept can be classified into two types. One is so-called bidirectional I/O interface [10], and the other is accomplished through modulation techniques [1-2]. In some applications, data and clock channels are combined as single channel to reduce the wire count. It necessitates the clock/data recovery circuits to extract the clock from the incoming data stream [2]. On the other hand, by reconsidering the representations of binary data, several kinds of modulation techniques were addressed, including the pulse-width modulation, PWM [11], a combinational of PWM and PAM signaling technology [5], code division multiple access, CDMA [12], frequency division multiple access, FDMA [9], and CDMA/TDMA (Time Division Multiple Access) [13].

In this paper a 1.1 Gbps interface circuit intended for data transfer over single interconnect channel using a combination of CDMA and PAM scheme is presented. In order to reduce the number of wires over the bus and increase the bus bandwidth the binary data are first CDMA coded and after that 4-PAM which incorporates multilevel amplitudes rather than binary signals is implemented.

A. Signaling issues and comparison

To achieve higher computational performance at the system level, instead of just increasing processor clock speed, parallelism in computation is widely used for chip multiprocessing, CMP, and symmetric multiprocessor, SMP, system design. These systems achieve computational concurrencies by using scheduling techniques such as multitasking and multithreading. However, the on- and off-chip interconnect communication architectures are still mostly based on conventional time-division multiplexing (TDM)-based communication protocols (typically known as time-division multiple access (TDMA) or time interleaving) that do not allow real communication parallelism. So the key motivation of this paper is to explore how to exploit parallelism and concurrency in communication without increasing resources complexity. This paper suggests how the

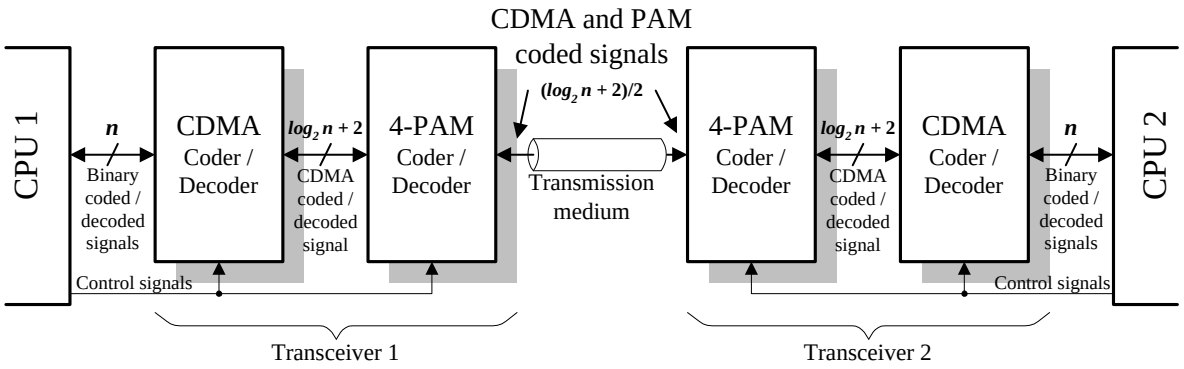

Fig. 2. Block diagram of the CDMA and PAM signaling interconnect structure

standard I/O interfaces should be modified for higher system performance with less cost. Using our results, designers of systems with multiple heterogeneous semiconductor-based processing devices will be able to take advantage of parallelism at both the computation and communication levels.

Current- or voltage-mode 2-PAM binary signaling has been used most widely in applications for high-speed on- and off-chip communications due to its simplicity, high noise immunity, and reliability. 4-PAM multilevel signaling has also been introduced, since its lower baud rate can double the eye width and also reduce frequency-dependent channel losses without increasing bus speed. However, for a given power, 4-PAM is more susceptible to channel interference noise such as crosstalk, since the multiple signal levels are more closely spaced than 2-PAM, resulting in higher BEER. Also the effective eye width is further reduced due to transitions between signal levels that are not adjacent.

V. MERGED CDMA AND PAM INTERCONNECT ARCHITECTURE

A block diagram of the merged CDMA and PAM signaling interconnect architecture, called CPIA, intended for data transfer over transmission medium, such as for example SoC and multiprocessor shared bus, is pictured in Fig. 2. Both building blocks, the CDMA– and 4_PAM–coder/decoder are constituents of a single transceiver. During data transfer from CPU_1 to CPU_2 the CDMA-coder/decoder of a Transceiver_1 is driven with n bus binary coded signals. At its output it generates $\log_2 n + 2$ binary coded signals. The 4_PAM coder converts two-level input signals to four-level output signals and send them through a transmission medium over $(\log_2 n + 2)/2$ lines. At the other end of a transmission line the decoder logic of a Transceiver_2 performs an opposite action.

In the sequel we will briefly describe the hardware structure of the CDMA-coder/decoder and 4_PAM-coder/decoder blocks.

B. CDMA-coder/decoder

The encoding scheme is illustrated in Fig. 3(b). Data from different bus lines drive the encoder. Each data bit is spread into S bit encoded data generated by XOR operations with a unique S-bit spreading code. Each bit of the S-bit encoded data generated by XOR operations is called a data chip. Then, the data chips which come from different data bus lines are added together arithmetically according to their bit positions in the S-bit sequences. After the add operations, we will get S sum values of S bit encoded data. Binary equivalents of the sum values are transferred over $\log_2 n + 2$ - bus lines to the 4_PAM coder/decoder block.

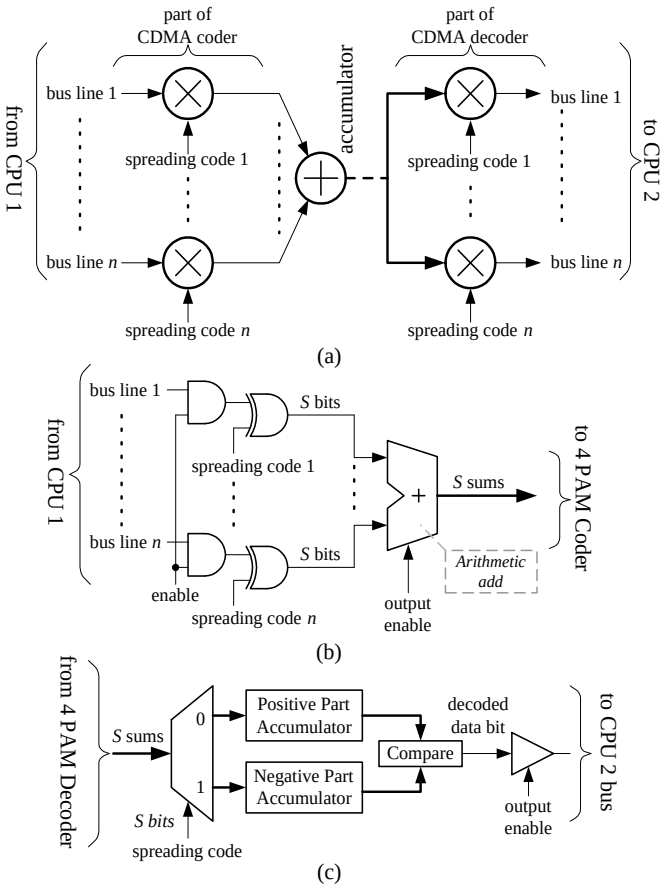

Fig. 3. CDMA coder/decoder: (a) CDMA technique principle; (b) Digital CDMA encoding scheme; (c) Digital CDMA decoding scheme

The decoding scheme is sketched in Fig. 3(c). The decoder accumulates the received sum values into two separate parts, a positive part and a negative part, according to the bit value of the spreading code used for decoding. After accumulating the sum values, the original data bit can be decoded by comparing the values between the two accumulators. For more details concerning the hardware structure of the CDMA coder/decoder see Reference [14].

C. 4 PAM coder/decoder

The encoding scheme of the 4_PAM coder is pictured in Fig. 4(b). It consists of three building blocks, 2-to-4 Decoder, Buffer-Stage, and Output-Stage. The 2-to-4 Decoder converts 2-bit CDMA coded signals, D_i and D_{i+1}, into four binary signals P_0, P_1, P_2 and P_3. At a given instant only one of the signals from P_0 up to P_3 is active. The Buffer-Stage is used for driving the Output_Stage. A control signal EN enables (disables) the Output_Stage, i.e. when EN=0 at a pin Out we have high-impedance. The Output-Stage is composed of four switches, referred as S_0, S_1, S_2 and S_3. M=8 points to the fact that a corresponding switch is realized with 8 transistors that operate in parallel. In order to generate a multi-level signal at the pin Out we have that $V_{D3} > V_{D2} > V_{D1} > 0$ V. Switches S_3 and S_0 are realized as a single level–, while S_2 and S_1 as double level–switches. The voltage level at the pin Out is defined according to the following Table I.

TABLE I
VOLTAGE LEVEL AT PIN OUT

switch				Out
S_3	S_2	S_1	S_0	
off	off	off	off	High-impedance
off	off	off	on	0 V
off	off	on	off	1.1 V
off	on	off	off	2.2 V
on	off	off	off	3.3 V

The decoding scheme of the 4_PAM decoder is given in Fig. 4 (c). The comparator block and thermometer coder logic are its basic building blocks. Principle of the thermometer coder logic is described according to the truth Table II.

TABLE II
THERMOMETER TRUTH TABLE

inputs			outputs	
o3	o2	o1	out2	ouit1
0	0	0	0	0
0	0	1	0	1
0	1	1	1	0
1	1	1	1	1

The Comparator block is composed of three identical comparator units C_1, C_2 and C_3.They compare the input signal, which feed their non-inverting inputs, with referent voltages derived from the resistor network. In our case we have V_1=0.5 V, V_2= 1.65 V, and V_3= 2.75 V. All comparator units are active during the time interval when the clock pulse CLK is positive.

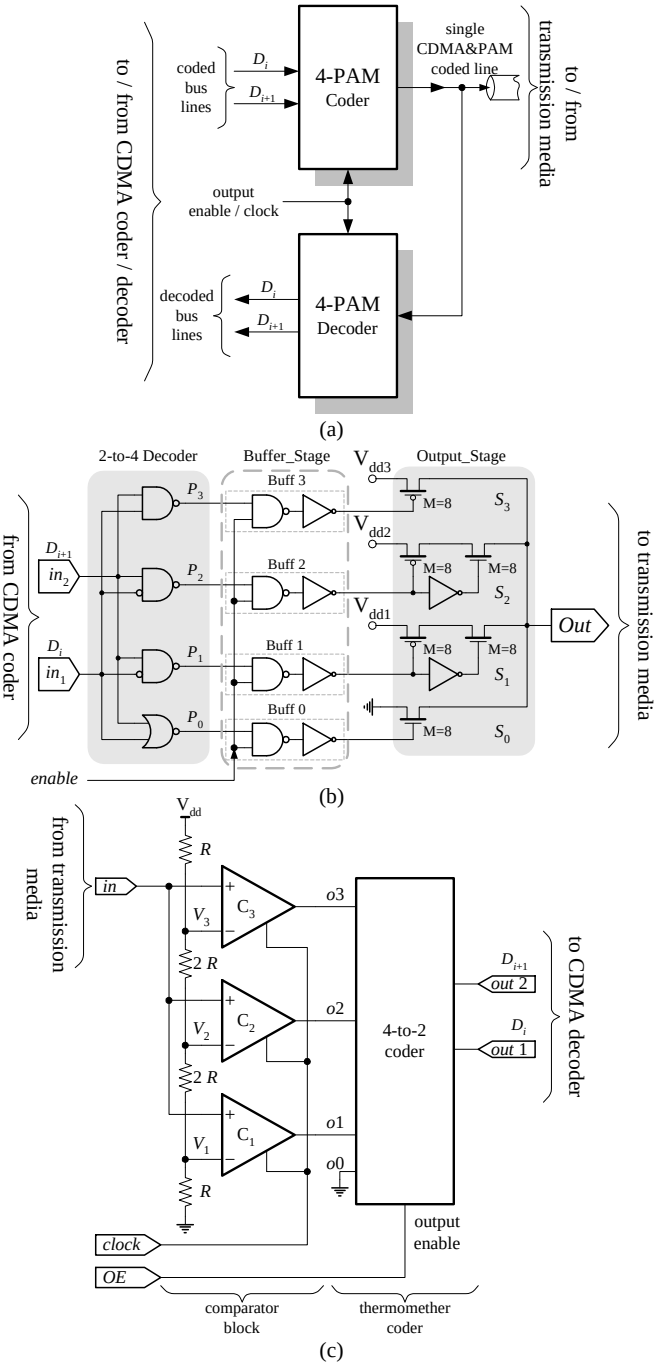

Fig. 4. PAM coder/decoder; (a) PAM coder/decoder structure; (b) 4 PAM coder; (c) 4 PAM decoder

D. Comparator block architecture

A detailed hardware structure of the Comparator block is given in Fig. 5. It consists of two constituents, Linear-Block, LB, and Digital-Block, DB.

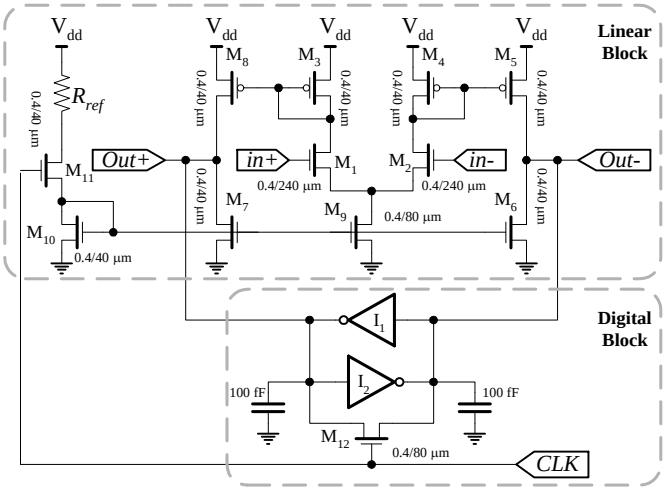

Fig. 5. Hardware structure of the Comparator block

The LB is composed of a differential stage M_1 and M_2, a current sink M_9, two pairs of current mirrors M_3–M_8 and M_4–M_5, two pairs of output stages M_5–M_6 and M_7–M_8, and a bias circuit M_{10}–M_{11}. The transistor M_{11} acts as a switch. M_{11} is in on-state when *CLK*=1, during which time period it provides condition for correct operation of the LB.

The DB logic consists of two weak-inverters I_1 and I_2. When *CLK*=1, M_{12} is on, the inverters I_1 and I_2 enter in linear mode of operation. During this period capacitors C_1 and C_2 are charged to values present at the LBs' outputs Out+ and Out-, respectively. When *CLK*=0, M_{12} is off, the LB switches to inactive mode of operation, I_1 and I_2 remain in active state and maintain the charges of C_1 and C_2 unchangeable.

VI. Experimental results

In order to estimate the performance attainable with the multilevel signaling method the design solution of a merged CDMA and PAM interconnect architecture has been implemented using: a) Xilinx FPGA technology for realization of the CDMA coder/decoder block; and b) conventional 0.35 µm length 2 polysilicon 4 metal layers with maximum voltage range of 3.5 V CMOS technology [15] for realization of the 4_PAM coder/decoder block. Starting from the fact that the proposed architecture consists of two different building blocks connected in cascade, a completely digital CDMA coder/decoder and a mixed-signal 4_PAM coder/decoder, various optimization techniques were used for their performance evaluation.

As a first, the CDMA coder/decoder constituent was realized as wrapper interface logic located between the CPU and 4_PAM coder/decoder block. It was described at RTL level using VHDL. For synthesis, routing and mapping a Xilinx development CAD tool WebPack 9.2i was used [16]. Design verification was performed using testbenches intended for parallel excitation of all data lines. The wrapper was implemented on FPGA devices from Spartan 2, Spartan 3, Virtex 4, Virtex 5 and Virtex E series. A pair of CDMA coder/decoder block seems to occupy appropriate space, in average 2000 equivalent gates, considering CPU cost of about 30 000 gates, what is less than 8% of hardware overhead per CPU. Let note that the communication bandwidth of the CDMA coder/decoder is much higher with respect to that of the 4_PAM coder/decoder circuit. Bearing this in mind these performance metrics will be omitted here. More details concerning performance of the CDMA coder/decoder logic are given in Reference [14].

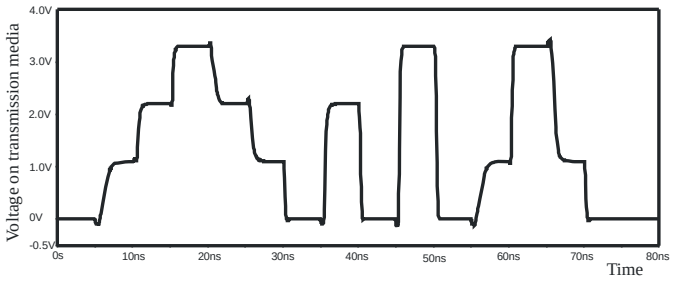

Fig. 6. Voltage on transmission media

The design of a 4_PAM coder/decoder circuit for mixed-signals was verified using HSpice simulation. Model level 47 that correspond to 0.35 µm CMOS technology was used. During simulation several different supply V_{DD} voltages (1.1 V, 2.2 V and 3.3 V) needed for correct operation of the Output-Stage were used. The supply voltage for all other logic circuits was fixed to 3.3 V. Fig. 6 deals with HSpice simulation which relates to dynamics of multi-level signals at the output of the Output-Stage. All types of signal transitions are presented.

TABLE III

SIGNAL TRANSITIONS AT THE NEAR END OF A TRANSMISSION LINE

Transition	Capacitive load *C* = 1 pF			
	raising edge [ps]	raising edge delay [ps]	falling edge [ps]	falling edge delay [ps]
0 ⇔ 1	619	960	189	183
0 ⇔ 2	190	552	234	275
0 ⇔ 3	170	454	212	408
1 ⇔ 2	210	704	347	655
1 ⇔ 3	223	409	344	829
2 ⇔ 3	196	361	549	900
	Capacitive load *C* = 3 pF			
0 ⇔ 1	939	1175	219	216
0 ⇔ 2	308	615	252	317
0 ⇔ 3	223	504	257	446
1 ⇔ 2	337	751	533	796
1 ⇔ 3	251	445	539	967
2 ⇔ 3	244	394	718	956
	Capacitive load *C* = 5 pF			
0 ⇔ 1	1280	1300	255	240
0 ⇔ 2	430	670	290	350
0 ⇔ 3	270	540	305	480
1 ⇔ 2	460	800	710	880
1 ⇔ 3	280	470	740	1090
2 ⇔ 3	290	420	890	1000

TABLE IV
DELAYS INVOLVED BY THE RECEIVER

Transition	raising edge [ps]	falling edge [ps]
0 ⇔ 1	446	0
0 ⇔ 2	0	445
0 ⇔ 3	437	192
1 ⇔ 2	0	193
1 ⇔ 3	444	203
2 ⇔ 3	439	448

The obtained results, presented in Table III for three different capacitive load of a transmission line (C=1 pF, 3pF, and 5 pF), relate to: i) pulse rising (falling) time (from 10% up to 90% pulse amplitude); and ii) pulse delay of a rising (falling) edge (50% to 50% pulse amplitude). By analyzing the results given in Table III we can conclude that maximum delays for rising edges are obtained for 0↔1 signal transitions, while for falling edges for transitions 1↔3 and 2↔3. All these case relate to *C*=5 pF capacitive load of transmission line. Maximum delay for falling edge is obtained for transitions 1↔3 and 2↔3. According to Table III we conclude that largest delays are obtained for 0↔1, 1↔3 and 2↔3 signal level transitions.

Delays involved by the receiver for all types of signal transitions are given in Table IV. In Table V the power consumption of a transmitter-receiver pair (single channel) for 200 MHz operating frequency and V_{DD}= 3.3 V are presented.

TABLE V
POWER CONSUMPTION OF A TRANSMITTER-RECEIVER

Capacitive load C_L [pF]	Power consumption [mW]
1	25.51
3	25.92
5	26.37

According to the results given in Table III and Table IV we can conclude that the maximal operating frequency of a transceiver is 571 MHz, what corresponds to 1.75 ns signal delay (including delay of a transmitter, receiver and transmission line loaded with C=5 pF). This implies that a communication bandwidth of 1.1 Gbps per single channel can be achieved.

VII. CONCLUSION

The goal of the on- and off-chip interconnect designs is to obtain the optimum combination of bandwidth, latency, number of wires and cost. Traditional inter-chip and intra-chip communications are based solely on TDMA approach. To overcome the limitations of the TDMA techniques a number of new interconnect schemes, mainly based on CDMA, PAM, PWM and FDMA have been investigated recently in order to increase the data rate, concurrency, and reduce latency and power consumption. In this paper a combined CDMA and PAM signaling interconnect architecture is proposed. Use of this method allows the required number of bus lines, n, for a given bandwidth to be reduced to $(\log_2 n + 2)$. The CDMA logic has been implemented in Xilinx FPGA family while the PAM interface in 0.35 μm CMOS technology. Simulation results show that the proposal can operate at 1 Gbps per single channel (wire) over transmission line loaded with C = 5pF. The scheme is generally applicable to any IP block – to – IP block link within the shared SoC bus or chip-to-chip link that can be implemented in a point – to – point fashion.

REFERENCES

[1] J. Kim, I. Verbauwhede, M-C Frank Chang, "Design of an Interconnect Architecture and Signaling Technology for Parallelism in Communication", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 15, No. 8, pp.881-893, August 2007.

[2] S. Pasrisha, N. Dutt, *On-Chip Communication Architectures: System-on-Chip Interconnect*, Morgan Kaufmann, Boston, 2008.

[3] D. Foley, M. Flynn, "A Low-Power 8-PAM Serial Transceiver in 0.5 μm Digital CMOS", IEEE Journal of Solid-State Circuits, Vol. 37, No. 3, pp.310-316, March 2002.

[4] TST. Mak, P. Sedcole, P. Y. K. Cheung, W. Luk, "On-FPGA communication architectures and design factors", In: Proc. International Conference on Field Programmable Logic and Applications, Vol 1, pp. 1-8, Madrid, 2006.

[5] M. Mitic, M. Stojcev, "An overview of on-chip buses", Facta Universitatis, Series: Electronics and Energetics, Vol.19, No.3, pp.405-428, November 2006.

[6] S. Dey, K. Lahiri, A. Raghunathan, "Design of communication architectures for high-performance and energy-efficient systems-on-chips", In: Jerraya AA, Wolf W, editors. Multiprocessor Systems-on-Chips, Morgan Kaufman, pp. 187-222, San Francisco, 2004.

[7] L. Benini, G. De Micheli, *Networks on Chips*, Elsevier, Boston, 2006.

[8] A. Jantsch, H. Tenhunen, Introduction. In: Jantsch A, Tenhunen H, editors. *Networks on Chip*, Kluwer Academic Publishers, pp. 1-24, Boston, 2003.

[9] M-c Frank Chang, I. Verbauwhede, C. Chien, Z. Xu, J. Kim, J. Ko, Q. Gu, B-C Lai, "Advanced RF/Based Interconnect Schemes for Inter- and Intra-ULSI Communications", IEEE Transactions on Electron Devices, Vol. 52, No. 7, pp.1271-1285, July 2005.

[10] J.Y. Sim, Y.S. Sohn, S.C. Heo, H.J. Park, S.I. Cho, "A 1-Gb/s Bidirectional I/O Buffer Using the Current Mode Schemes", IEEE Journal of Solid-State Circuits, Vol. 34, No. 4, pp.1271-1285, April, 1999.

[11] W-H. Chen, G-K. Debug, J-W. Chen, S-I. Liu, "A CMOS 400/Mb/s Serial Link for AS-Memory Systems Using a PWM Scheme", IEEE Journal of Solid-State Circuits, Vol. 36, No. 10, pp. 1498-1505, October 2001.

[12] X. Wang, T. Ahonen, J. Nurmi, "Applying CDMA Technique to Network-on-Chip", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 15, No. 10, pp.1091-1100, October 2007.

[13] B. C. Lai, P. Schaumont, I. Verbauwhede, CT-bus, "A Heterogeneous CDMA/TDMA Bus for Future SoC", In: Proc. 38th Annu. Asilomar Conference on Signals, Systems, and Computers, Vol. 2, pp. 1868-72, 2004.

[14] T. Nikolić, M. Stojčev, G. Đorđević, "CDMA Bus-Based On-Chip Interconnect Infrastructure", Microelectronics Reliability, Vol. 49, No.4, pp.448-459, April 2009.

[15] 0.35 μm CMOS SPICE Parameters, available at https://www.eecs.berkeley.edu/~boser/courses/software/spice/CMOS35/index.html, February 2010.

[16] ISE 9.2i Quick Start Tutorial Xilinx ISE 9.2i, Software Manuals Helps, http://www.xilinx.com/support/sw_manuals/xilinx92/download