An AER handshake-less modular infrastructure PCB with x8 2.5Gbps LVDS serial links

T. Iakymchuk[1], A. Rosado[1], T. Serrano-Gotarredona[2], B. Linares-Barranco[2]

[1]ETSE. GPDS. Dpto. Ing. Electrónica. University of Valencia.
[2]Instituto de Microelectrónica de Sevilla, IMSE-CNM (CSIC and Univ. Sevilla), SPAIN. bernabe@imse-cnm.csic.es

A. Jiménez-Fernández, A. Linares-Barranco, G. Jiménez-Moreno

Robotic and Technology of Computers Lab., Univ. Sevilla, SPAIN.

***Abstract*— Nowadays spike-based brain processing emulation is taking off. Several EU and others worldwide projects are demonstrating this, like SpiNNaker, BrainScaleS, FACETS, or NeuroGrid. The larger the brain process emulation on silicon is, the higher the communication performance of the hosting platforms has to be. Many times the bottleneck of these system implementations is not on the performance inside a chip or a board, but in the communication between boards. This paper describes a novel modular Address-Event-Representation (AER) FPGA-based (Spartan6) infrastructure PCB (the AER-Node board) with 2.5Gbps LVDS high speed serial links over SATA cables that offers a peak performance of 32-bit 62.5Meps (Mega events per second) on board-to-board communications. The board allows back compatibility with parallel AER devices supporting up to x2 28-bit parallel data with asynchronous handshake. These boards also allow modular expansion functionality through several daughter boards. The paper is focused on describing in detail the LVDS serial interface and presenting its performance.**

I. INTRODUCTION

Address Event Representation (AER) is a communication protocol born at the end of the eighties [1] when neuro-inspired researchers faced the problem of communicating off-chip thousands of on-chip VLSI neurons with only a few pins. Originally, AER multiplexed in time the neurons spike activity into a high-speed asynchronous handshaked digital bus. Since then, AER based systems have grown in complexity and resources, requiring higher and higher communication bandwidths and logic resources. This situation has enabled the construction of ambitious infrastructures that could emulate processes in a similar way as a brain does. From sensors [2]-[4], through filters and processors [5],[8]-[14], to actuators [15]-[16], the AER protocol is present in this neuromorphic world.

With the AER evolution, a set of interfaces, bridges, communication infrastructures and protocol adaptations have arisen in the literature. From the first PCI-AER interface able to capture, monitor and map AER activity at 1Meps [17] in 2001, to really fast and powerful infrastructures capable of communicating at 3Geps [18] in 2010, many solutions can be found like a versatile and stand-alone USB-AER interface for debugging AER systems (generating, sequencing, mapping, monitoring, datalogging and processing at up to 16-bit 10Meps) [5], or serial oriented communication and processing platforms [19].

With the growth of neuromorphic systems, processing and communication infrastructures performance (event rates, data transmission requirements between boards and computers) has also increased. In AER-based systems the connectivity is the main performance bottleneck because it is the core of the AER and spike-base processing philosophy: to have small processing cells working in parallel that massively communicate with each other. Event-based processing speed is often limited by the throughput of the hardware, which is even worse when transmission is established between different boards.

Interfaces with parallel digital buses have limited bandwidth due to bus frequency limitations, inter-bit jitter and skew, and bus length design restrictions. Serial interfaces advantages are: (1) they operate at much faster clock speeds that directly imply much higher bandwidths; (2) they can transmit variable word lengths without complex hardware changes; (3) handshake protocol per event is substituted by a flow control mechanism between event streams, implying important speed ups. Since clock signals inside a chip or an FPGA can be multiplied without risk by x5 factor, it is feasible and desirable to mix these two concepts for next-generation of AER processing: hundreds of MHz clocks for parallel on-chip processing, and units of GHz for LVDS serial communications between boards. Nevertheless, GHz serial communications give rise to new problems, like unaligned words, bit errors, clock phase shifting, … that must be solved.

This paper presents an FPGA based solution, which we call "AER-Node" PCB, for high speed serial communications between boards that does not limit the performance of the event-based processing system implemented on many-boards systems. The work has been designed, synthesized and tested on a set of AER-Node PCBs, whose heart is a powerful but relatively low-cost Spartan6-150t FPGA with GTP serial transceivers. The board was designed as a universal spike processing platform, which is scalable by connecting to a set of daughter boards (see Fig.1).

Next Section presents a brief description of the AER-Node board and their daughter boards. Section III, is focused on the LVDS board-to-board communication interface, replacing classical handshake by flow control (as proposed elsewhere [19]). Section IV gives implementation details, and Section V the conclusions.

II. AER-NODE PCB

The AER-Node board was designed in the context of a Spanish Government funded project, where the aim was to demonstrate that spike-processing under AER is feasible and convenient for high speed frame-free vision, filtering, processing and actuation, using spikes from vision sensors to DC motors. Under this project we also developed an improved dynamic vision sensor [3], an AER convolution chip [20], a spike based filter for object detection and tracking [21], a scalable mesh network multi-PCB technique of spike convolutions for cortex operation emulation [9], and spike based motor controllers (proportional-integral-derivative [15] and neuro-inspired open-loop VITE [16]). The AER-Node board was designed to allow multi-PCB communication with conventional parallel-handshaked-AER chips (retinas and convolutions) or robots with the adequate motor interfaces. To achieve these requirements a Spartan6 XC6S1500FXT FPGA was selected, and a set of daughter boards were designed to increase the functionality (interface to convolution chips, to multiple DVS retinas, USB-computer, Controller-Area-Network, embedded computer and monitoring interfaces). Scalability is provided by four SATA connectors for bidirectional LVDS high-speed communications to enable a mesh of AER-Node boards [9]. Through two parallel 28-bit connectors and two 8-bit data connectors, functionality can be increased with proper daughter boards.

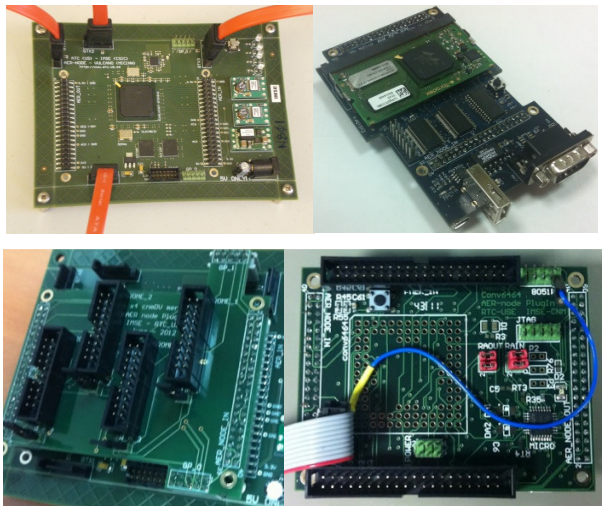

Fig. 1. AER-Node board with 4 SATA and 2 parallel AER connectors (left-top) and daughter boards: Toradex embedded processor with CAN interface (right-top), multi-retina receiver (left-bottom) and convolution-chip + USB computer interface (right-bottom).

Two clocks of 50MHz (single wired) and 100MHz (LVDS) allow the board to implement high frequency designs, using the FPGA internal Digital Clock Multipliers (DCM), and high-speed LVDS[1] serial transmission through embedded GTP transceivers and IP-core supported by Xilinx for managing communication issues like clock recovery/ correction, data loss detection and event alignment.

[1] With a 100MHz LVDS reference clock, this Spartan6 allows for a line frequency of up to 2.5Gbps. For maximum 3.2Gbps line frequency, a reference clock of 160MHz is required.

III. CONVENTIONAL BIT-SERIAL TRANSMISSION IN FPGAS

Modern Xilinx FPGAs have several high-speed differential physical interfaces [6]. These interfaces provide fast data rate (up to 3.125 Gbps per line) and are compatible with popular SATA or PCI Express standards. Differential line transmission hardware has embedded buffers, encoders, decoders and all necessary circuitry for reliable serial data transmission with wide range of physical signal parameters and timings to match various data transmission protocols and standards. To allow for embedded clock transmission with improved line electrical characteristics and bit error rate (BER), an 8b/10b encoding is normally used. This encoding decreases effective data rate by 20%, but allows much higher speeds and longer transmission lines, improving robustness.

Fig. 2 shows a simplified block diagram of one of these physical interfaces, called "GTP Transceiver wrapper", provided by Xilinx's Core Generator utility. It includes a transmitter and a receiver.

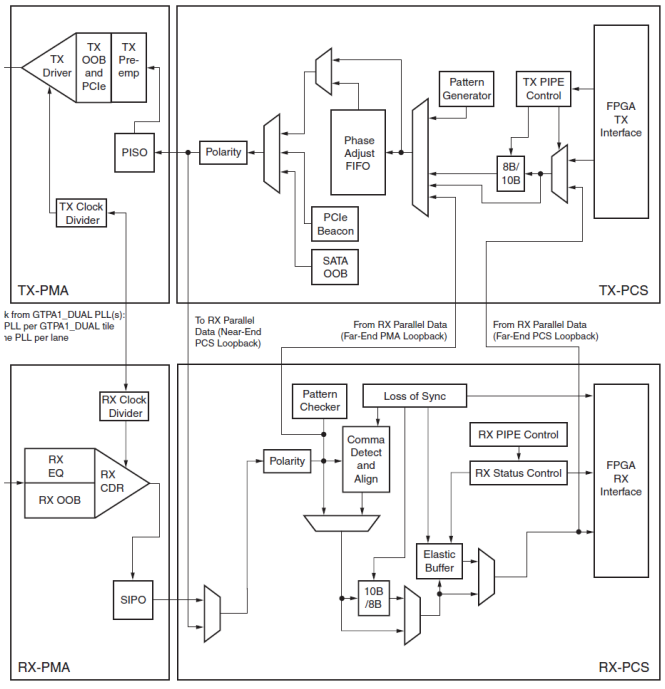

Fig. 2: Simplified Xilinx LVDS GTP Transceiver wrapper diagram [7]

A. *The GTP wrapper Transmitter.*

The transmitter (Fig 2. top) captures parallel data in the FPGA TX interface (with configurable depth) on rising edges of the TXUSRCLK2 clock signal. A second clock is used internally by the transceiver (TXUSRCLK = n·TXUSRCLK2 with n=1,2,4 the number of data bytes). Depending on the selected data rate and usage or not of 8b/10b encoding, data are serialized and sent synchronously with the selected clock rate. Input data sampling clock can be calculated as:

$$TXUSRCLK\ rate = \frac{line\ rate}{data\ width}$$

8b/10b encoding maps any 8-bit word into a 10-bit one trying to balance the number of '1s' and '0s' [22]. A few spare 10-bit balanced combinations are available that are not mapped to any 8-bit word. These are called K-chars. We use some of them for keeping the channel synchronized while it is idle or to perform flow control [9]. A *Phase-adjust-FIFO*

block is in charge of synchronizing phases of two different clock governed circuits. A *Pattern generator* can be used to generate pseudo-random sequences useful for standard compliance and communication testing purposes. *Polarity* of the LVDS signal can be configured. A high speed clock for the serial transmission can be configured by a clock divider, and finally voltage levels of the LVDS signal can be adapted to a particular supported standard, like PCIe, SATA, etc.

B. *The GTP wrapper Receiver*

The receiver has to detect the correct phase of the signal, decode 8b/10b encoding, and recognize/extract K-chars. Due to mismatch in the TX and RX clock frequencies, the phase alignment of the data tends to drift in time, requiring an internal elastic buffer for clock correction. Each 32-bit event is a four-byte word.

Crystal oscillators of different boards tend to change their frequency with temperature or voltage supply drift. At 2.5 Gbps data rate these changes imply data loss and channel misalignment every several thousands of transmitted words. The *RX elastic buffer* resolves these differences. There are two possible situations: (1) If the clock (XCLK) of the sending Transceiver is slower than the one in the receiving Transceiver (RXUSRCLK), the receiver clock correction module eventually inserts a spare K-char while activating clock correction signal 'ClockCorr'; (2) if XCLK is faster than RXUSRCLK a spare K-char needs to be removed periodically, slightly decreasing effective data rate, but making the link robust. Inserting/removing these K-chars requires including additional control at the receiver user circuitry to realign multi-byte words.

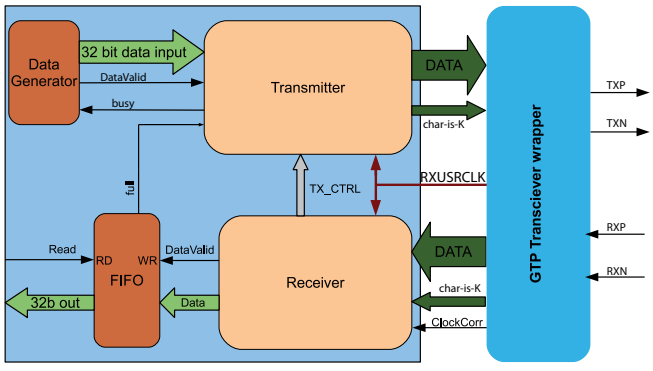

Fig. 3: Global transceiver schematic.

IV. IMPLEMENTATION

The Transceiver was synthesized with a 32-bit parallel synchronous data interface and a number of additional control/state signals. This handshake-less interface is synchronized with the recovered clock. From the user point of view, after successful channel setup, the GTP transceiver is a "pipe" and all data set on the input synchronously with the clock will be serialized, encoded and sent to the other side through SATA cables, where they will be deserialized, decoded and put on a 32-bit parallel output port by another Transceiver.

After the physical connection, the channel needs alignment. The alignment is made by repeatedly sending a special four K-char sequence. The receiver is monitoring the incoming stream and after the detection of this sequence, it aligns the stream to the closest 4-byte word boundary signalizing readiness for data reception.

For maximum throughput testing purposes an internal VHDL 32-bit data generator was used. The global concept of the transceiver design can be seen in the Fig. 3. A full bidirectional link is made of two such Transceivers where lines {TXP/N} and {RXP/N} are cross connected.

After startup the transmitter sends repeatedly the 4 K-char alignment word 0x1CBCBCBC. The '1C' header allows determining the beginning of the word (an example of received word is BC1CBCBC). Once words are kept aligned, data communication can start. Then the receiver sends 32-bit words (events) to the FIFO (which are synchronously read from outside) when signal "Read" is active. If the outside receiver is slower than the incoming data speed, the "full" signal is activated and the Transmitter sends another special K-char to the sending Transceiver to stop data transmission. In this case, the receiver (at the other Transceiver) uses TX_CTRL to tell the Transmitter to temporarily freeze acknowledge signal TX_ACK. When the FIFO has space again, the "full" signal is deactivated and a resume K-char is sent to the other Transceiver to resume acknowledge signaling.

The FIFO is a cyclic buffer and can handle read and write operation in one clock cycle. The length of the FIFO must be greater than the deserialization delay (typically 20 words).

V. EXPERIMENTAL RESULTS

In order to test 32-bit data transmission on full speed, an internal data generator was added to the design, as shown in Fig. 3.

The test setup consisted of two AER-Node boards (A and B), connected by SATA cable. On each board the design had two transmitter-receiver pairs. Clock frequency values were measured by an oscilloscope: A (Min, Max, delta) clock frequency was (99.92, 100.05, 0.13), and B was (99.97, 100.3, 0.06). As can be seen, precise quartz resonators have noticeable fluctuations, which are multiplied by the PLLs, causing channel desynchronization. Clock jitter of the LVDS channel of AER-Node board A can be seen in Fig. 4 (average eye width is 283.5ps out of a 400ps period).

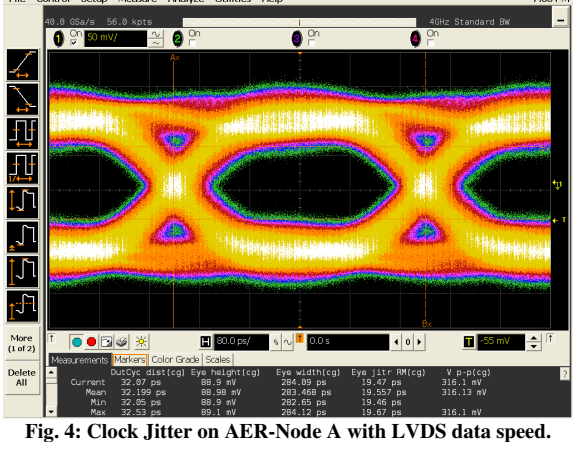

Fig. 4: Clock Jitter on AER-Node A with LVDS data speed.

During the experiment, clock differences were always the same, i.e. one board constantly reported underflow, while the other overflow. In a first test, the goal was to measure K-chars rate versus clock correction rate. The results for

communication between A and B boards are presented in Table 1. Valid Word rate expresses the number of valid data words transmitted before a clock correction is signaled by the GTP wrapper. It shows max, min and standard deviation values.

The larger the Valid Word rate and K-char distance, the better the communication performance. We observed that for stable clock correction, comma word separations should be below half of clock correction separation. Therefore a K-char every 40K data words is a top limit.

Comma each X words, board A to B	Valid Word rate (Max)	Valid Word rate (Min)	Valid Word rate (Std)
5	71991.2793	70596.2902	650.8272
1024	84279.1190	83836.4359	190.3470
32767	83919.4519	83361.1175	394.8020
40000	85953.3874	85652.3278	212.8813

Table 1. Clock correction rate in communication channels.

A second robustness test for determining error rate was accomplished. Two 10 hour runs were conducted. Errors in data or invalid K-char were counted as a single transmission fault. There were no transmission faults detected during the experiment (10 hours of 2.5Gbps transmission, 1comma word per 1024 words), thus zero channel misalignments were detected, meaning that the channel was stable all the time.

VI. CONCLUSIONS

A high-speed serial interface for 32-bit handshake-less serial AER event communication with 2.5Gbps bandwidth and a peak rate of 62.5Meps is presented. It has been implemented for a Spartan6 FPGA on the AER-Node board infrastructure PCBs, which allows implementing complex spike-based processing systems by interconnecting as many AER-Node boards as necessary through these serial links with SATA cables. The infrastructure functionality can be enhanced by using proper daughter boards. On the target device, the AER-Node board, the design works in a stable manner at 2.5Gbps, with a safe comma rate of one each 1024 words. This results in an effective speed of 62,439,000 words per sec, which is almost equal to the maximum theoretical limit of 62.5Meps. Using lower comma rates such as 1:40000 would approximate to the maximum throughput speed.

ACKNOWLEDGMENTS

This work has been supported by Spanish grants (with support from the European Regional Development Fund) VULCANO (TEC2009-10639-C04-02/01) and BIOSENSE (TEC2012-37868-C04-02/01), Andalusian grant NANO-NEURO (TIC-6091), and EU CHIST-ERA grant PNEUMA (PRI-PIMCHI-2011-0768).

REFERENCES

[1] M. Sivilotti, “Wiring considerations in analog VLSI systems with application to field-programmable networks,” Ph.D. dissertation, Computation and Neural Systems, California Inst. Technol., Pasadena, CA, 1991.

[2] P. Lichtsteiner, C. Posch, and T. Delbruck, A 128x128 120dB 15us latency asynchronous temporal contrast vision sensor, IEEE J. Solid State Circuits, 43(2) 566-576, 2007.

[3] Serrano-Gotarredona, T. ; Linares-Barranco, B. “A 128x128 1.5% Contrast Sensitivity 0.9% FPN 3µs Latency 4mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers”. Solid-State Circuits, IEEE Journal of. Vol. 48, No. 3. Pp: 827 - 838. 2013.

[4] Chan, V.; Shih-Chii Liu; van Schaik, A., "AER EAR: A Matched Silicon Cochlea Pair With Address Event Representation Interface," *Circuits and Systems I: Regular Papers, IEEE Transactions on* , vol.54, no.1, pp.48,59, Jan. 2007

[5] R. Serrano-Gotarredona, et al. “CAVIAR: A 45k-Neuron, 5M-Synapse, 12G-connects/sec AER Hardware Sensory-Processing-Learning-Actuating System for High Speed Visual Object Recognition and Tracking”, IEEE Trans. on Neural Networks, vol. 20, No. 9, pp. 1417-1438, September 2009.

[6] DS160. Spartan-6 Family Overview, v.2.0. Xilinx, October 2011.

[7] UG386 Spartan-6 FPGA GTP Transcievers. Advance Product Specification, v.2.2. Xilinx, April 2010.

[8] L. Camuñas-Mesa, C. Zamarreño-Ramos, A. Linares-Barranco, A. Acosta-Jiménez, T. Serrano-Gotarredona, and B. Linares-Barranco, “An event-driven multi-kernel convolution processor module for event-driven vision sensors,” *IEEE J. Solid-State Circuits*, vol. 47, no. 2, pp. 504–517, Feb. 2012.

[9] C. Zamarreño-Ramos, A. Linares-Barranco, T. Serrano-Gotarredona, B. Linares-Barranco, “Multicasting Mesh AER: A Scalable Assembly Approach for Reconfigurable Neuromorphic Structured AER Systems. Application to ConvNets”. IEEE TRANS. ON BIOMEDICAL CAS, VOL. 7, NO. 1, pp 82-102. FEBRUARY 2013.

[10] Merolla, P.A. ; Arthur, J.V. ; Shi, B.E. ; Boahen, K.A. “Expandable Networks for Neuromorphic Chips”. Circuits and Systems I: Regular Papers, IEEE Transactions on. Vol. 54, No. 2. Pp 301 – 311. 2007.

[11] S.B. Furber et al. “Overview of the SpiNNaker System Architecture,” IEEE Trans. Computers, doi 10.1109/TC.2012.142, 2012.

[12] http://brainscales.kip.uni-heidelberg.de/

[13] http://www.stanford.edu/group/brainsinsilicon/neurogrid.html

[14] S. Mitra, S. Fusi, and G. Indiveri, “Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI,” IEEE Trans. Biomed. Circuits Syst.,vol.3,no.1,pp.32–42,Feb.2009.

[15] Jimenez-Fernandez A, Jimenez-Moreno G, Linares-Barranco A, Dominguez-Morales MJ, Paz-Vicente R, Civit-Balcells A. A Neuro-Inspired Spike-Based PID Motor Controller for Multi-Motor Robots with Low Cost FPGAs. Sensors. 2012; 12(4):3831-3856.

[16] Perez-Peña, Fernando; Morgado-Estevez, Arturo; Linares-Barranco, Alejandro; Jimenez-Fernandez, Angel; Gomez-Rodriguez, Francisco; Jimenez-Moreno, Gabriel; Lopez-Coronado, Juan. 2013. "Neuro-Inspired Spike-Based Motion: From Dynamic Vision Sensor to Robot Motor Open-Loop Control through Spike-VITE." Sensors 13, no. 11: 15805-15832.

[17] Vittorio Dante, Paolo Del Giudice, and Adrian M. Whatley. “Interfacing with address events”. The Neuromorphic Engineer. DOI: 10.2417/1200503.0021. 2005. http://www.ine-web.org

[18] Hartmann, S. ; Schiefer, S. ; Scholze, S. ; Partzsch, J. ; Mayr, C. ; Henker, S. ; Schiiffny, R. “Highly integrated packet-based AER communication infrastructure with 3Gevent/S throughput”. Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on. Page(s): 950 - 953.

[19] Daniel B. Fasnacht, Adrian M. Whatley, Giacomo Indiveri. “A Serial Communication Infrastructure for Multi-Chip Address Event Systems”. *Circuits and Systems (ISCAS), 2008 IEEE International Symposium on.*

[20] L. Camuñas-Mesa, C. Zamarreño-Ramos, A. Linares-Barranco, A. Acosta-Jiménez, T. Serrano-Gotarredona, and B. Linares-Barranco, “An Event-Driven Multi-Kernel Convolution Processor Module for Event-Driven Vision Sensors,” *IEEE J. of Solid-State Circuits*, vol. 47, No. 2, pp. 504-517, Feb. 2012.

[21] Gómez-Rodríguez, F. ; Miró-Amarante, L. ; Diaz-del-Rio, F. ; Linares-Barranco, A. ; Jimenez, G. “Real time multiple objects tracking based on a bio-inspired processing cascade architecture”. ISCAS-2010. pp 1399-1402.

[22] P. A. Franaszek, et al., “Byte oriented DC balanced (0,4) 8b/10b partinioned block transmission code,” US Patent 4,486,739, Dec. 4, 1984.