ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Nuclear Instruments and Methods in Physics Research A 570 (2007) 518–524

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH
Section A

www.elsevier.com/locate/nima

# Remoting field bus control by means of a PCI Express-based optical serial link

M. Bellato*, R. Isocrate, G. Meng, M. Passaseo, G. Rampazzo, A. Triossi, S. Ventura

*INFN Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy*

Received 22 September 2006; accepted 24 October 2006
Available online 16 November 2006

## Abstract

A major evolution moved into the I/O architecture of modern computers, where the multi-drop buses have been replaced by a network of point-to-point links. Besides the increased throughput and the inherent parallelization of the data flows, the serial nature of those links and the packet-based protocols allow an easy geographical decoupling of a peripheral device. In the context of the LINCO project, we investigated the possibility of using an optical physical layer for the PCI Express, and we built a bus adapter which can bridge, through such a link, remote buses (>100 m) to a single-host computer without even the need of a specialized driver, given the legacy PCI compatibility of the PCI Express hardware. By the choice of suitable components and dedicated control logic, the adapter has been made tolerant to harsh environmental conditions, like strong magnetic fields or radiation fluxes, that the data acquisition needs of high-energy physics experiments often require.

*PACS:* 07.50.Ek; 07.05.Wr; 07.05.Hd; 85.60.−q

*Keywords:* PCI Express; I/O Virtualization; VME; Optical link

## 1. Introduction

PCI Express is a new I/O technology for desktop, mobile, server and communications platforms designed to increase the levels of computer system performance. It represents a radical move from traditional I/O architectures and it replaces parallel multi-drop buses with serial switched point-to-point links. The promise of bandwidth scaling is committed to the choice of the number of lanes (individual serial pairs) that compose the link. Being each lane bi-directionally driven at 2.5 Gbit/s, the capacity of the channel after 8b/10b encoding is fixed at 250 Mbyte/s times the number of lanes [1]. While this bandwidth outperforms the capacity of former standards, the points worth considering for our goal are not only concerned with speed. The new serial technology adopts a "communication centric" approach: the load-store operations between two nodes are performed exchanging framed packets in accordance to a suite of stacked protocol layers taking care of the physical, link and transaction issues of the channel. In case of PCI Express, all these activities are carried out at the hardware level, with no software intervention. Clearly, this load-store model logically matches the model of field bus control in which a host and a networked peer node exchange software arranged packets to access memory and registers of the field bus for I/O operation. To investigate and further extend this parallelism to its practical consequences, we addressed a number of activities in the context of the INFN Gr. V-funded project LINCO. The goals pursued were:

- investigate a non-standardized physical medium for the PCI Express protocol, namely running on a optical fiber link,
- manufacture an adapter to translate PCI Express signals to/from the optical physical layer and which, using commercial bridges, could be fitted into legacy bus standards (PCI, CompactPci, VME),

---

*Corresponding author. Tel.: +39 049 8277252; fax: +39 049 8277087.
*E-mail address:* marco.bellato@pd.infn.it (M. Bellato).

0168-9002/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2006.10.190

- assert its effectiveness in a real environment by deploying it as a basic building block of the local data acquisition and control system of a high-energy physics detector (Fig. 1).

## 2. The new physical medium

As per the PCI Express specifications version 1.0a: "*the interconnect comprises everything between the pins at of a Transmitter package and the pins of a Receiver package*". No assumptions are made on the nature of the medium although it is clearly stated that… "*often, this will consist of traces on a printed circuit board or other suitable medium, AC coupling capacitors and perhaps connectors*" [2]. The three key parameters to be considered in the choice of an alternative interconnection are basically:

1. loss (i.e. the attenuation of the differential voltage swing that can be tolerated between a transmitter and a receiver),
2. jitter, being specified both at the transmitter and the receiver implicitly constrains the in-between medium, and
3. reference clock distribution.

Loss versus frequency, in the case of optical fibers, of course is not a concern because commercial optical transceivers normally used for telecom applications are tuned for operation in the 1310 or 1550 nm bands where, notably, insertion loss has local minima and, even in the short-reach grade, covering distances of 2–3 km is straightforward.

Instead, the figure of total jitter reserved for interconnection by the specifications deserves careful attention. These state that a maximum of 0.3 Unit Intervals (UI) is allowed, e.g. 120 ps of total (random plus deterministic) jitter at a Bit Error Ratio (BER) level of $10^{-12}$ [2] introduced by the chain of transmission side connector, electrical-to-optical conversion, fiber, optical-to-electrical conversion and receiver side connector. To this regard, important design parameters become line impedance matching, power and signal integrity along the chain and a low-phase noise of the reference clock. Optical transceivers fall into the same considerations as their contribution to the overall jitter is in the critical path; for this reason, we evaluated two types of transceivers: a JDS Uniphase multirate (up to 2.7 Gbit/s) transceiver qualified for OC-48 operation on a 1310 nm single mode fiber and an Intel transceiver qualified for Infiniband operation at 2.5 Gbit/s in a multimode fiber. Fig. 2 shows the contribution of the optical interface by a direct measure of the voltage and timing margins at the entrance and exit of the electrical–optical–electrical chain taken with an Agilent 54855A 6 GHz real-time scope, showing that the contribution of the medium to total jitter is $4.88 \pm 0.21$ ps. In order to measure the eye opening at a BER level of $10^{-12}$ ,we used the tail-fit method. We obtained the jitter's Probability Density Function (PDF) by plotting a histogram of the data Time Interval Error (TIE), i.e. the time difference between the recovered clock (from the data stream) and the data signal.

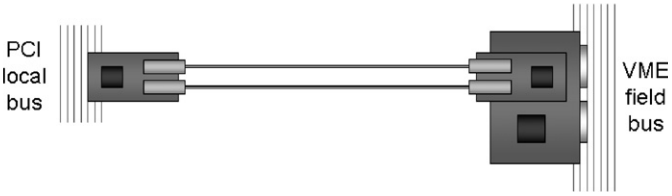

Fig. 1. Block diagram of the hardware setup.

The timing total jitter $\mathrm{PDF_T}$ can be further deconvolved into a deterministic component $\mathrm{PDF_D}$ and a random one ($\mathrm{PDF_R}$), defined as Gaussian,

$$\mathrm{PDF_T}(t) = \int_{-\infty}^{+\infty} \mathrm{PDF_D}(\tau)\mathrm{PDF_R}(t-\tau)\,\mathrm{d}\tau. \tag{1}$$

Since $\mathrm{PDF_D}$ is bounded below a certain jitter range, $\mathrm{PDF_T}$ comprises just random jitter processes. Through a Gaussian fit on these tail regions, we extrapolated the PDF analytic function concerning rare events [3]. In Fig. 3, we can see the value of the mean ($\mu$) and $\sigma$ of the Gaussian distributions related to the left and right tails. Fig. 3 gives an illustration on the relationship between jitter PDF and BER. The two vertical dashed lines denote the zero crossing time locations for the bit cell. As we can see from the jitter PDF, if the sampling point falls away from the edge transition a bit error occurs with a lower probability. The BER function is the sum of the two jitter PDF areas and it can be considered the cumulative distribution function, essentially the integral of the

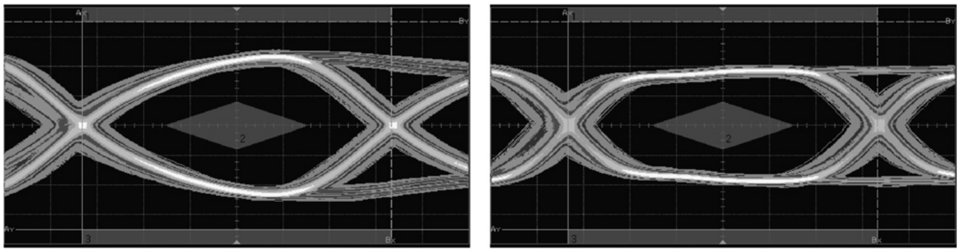

Fig. 2. Contribution of the optical transport medium to signal jitter.