

RECENT ADVANCES IN HIGH-SPEED SERIAL I/O TRENDS, STANDARDS AND TECHNIQUES

Peter Noel, Farhad Zarkeshvari and Tad Kwasniewski
Department of Electronics, Carleton University
<pnoel, fzarkes, tak> @doe.carleton.ca

Abstract

The goal of this paper is to provide the reader with an overview of the recent advances (in the past year) made in the industry with respect to high-speed data transport. The latest developments in the industry's main high-speed I/O protocols and the attempted standardization of the high-speed physical interface will be presented. In reviewing these advanced techniques, examples will use the basic circuit-level building blocks within a high-speed serial transceiver and will give the basics behind the design techniques required to successfully design and implement a typical multi-Gigahertz serial I/O device.

Keywords: High-Speed Data, Serial I/O, Rapid I/O, Advanced Switching, PCI Express.

1. INTRODUCTION

High-speed data transport and device integration are two main requirements of network development and installation. As applications for high-speed data communications evolves, so must the protocol, standards and techniques. The past year has seen such advancement with further developments in the protocol implementation of such standards as Rapid I/O and PCI Express/Advanced Switching. As protocols mature, the interfaces to the physical layer on to which the data from such advanced implementations are launched and from which the signal is retrieved must be designed to be more robust. Such is the basis of the recently minted Unified 10 Gbps Physical-layer Initiative (UxPI).

The increasing demand for more bandwidth to support inter-device communication or, from a more general viewpoint, to provide advanced voice, data and video applications via media interconnect is continuing to drive the development of high-speed serial transceivers. A typical networking installation switching multiple 10 Gigabit Ethernet sources and destinations, for example, would most likely implement the Rapid I/O or Advanced Switching protocol and would certainly benefit from a common SERDES standard such as the

proposed UxPI. Similarly, the interconnection of a cluster of high-performance microprocessors might very well incorporate Hypertransport, Rapid I/O or PCI Express as the inter-processor interconnect protocol, thereby allowing the full horse-power of such a super-computer to be utilized.

2. Recent High-Speed Interconnect Trends

The physical interconnection of ASICs, SOCs, microprocessors, DSPs and the boards on which these devices are placed are continuing to involve increased use of high-speed serial I/O. With an emphasis on efficiency, robustness and economy, the multi-lane high-speed serial interface is becoming common place in system design. Traditional interconnect designs required numerous (64 or 128) data pins, address pins, control signals and clocking signals. The overall pin counts for each individual interconnection often exceed 200 pins. These interconnect or bus based designs could provide point-to-multipoint interconnections but as the number of devices connected to a bus increased, so did the associated capacitance or loading. This results in a reduced data rate even if the communication was only destined for one device.

The design of a system requires the interconnection of devices on a single printed circuit board and the physical linking of these circuit boards to each other through either mezzanine or backplane connections. To manufacture the system, the printed circuit card increases in complexity and cost if more circuit traces are required as does the connectors associated with the mezzanine or backplane. It is preferred to reduce the number of circuit traces on the cards and in the connectors.

Recent trends in high-speed system interconnect concentrates on reducing the pin count, increasing the overall throughput and decreasing complexity and cost. By moving to a lower pin count, higher frequency interconnect, the above requirements can be achieved but at the loss of the multi-point interconnect ability. To obtain the high data throughput while reducing the pin

count or data bus width, the clock frequency must be increased. Increasing the clock frequency makes the channel more sensitive to the capacitive loading effects, including reflections, as would be experienced by adding multiple drop nodes on the interconnect. Typical high-speed low-pin count interconnect designs tend to incorporate point-to-point arrangements in response to the adverse loading effects at the higher frequencies.

To address the reduction in the allowable number of devices that can “hang” off of a high speed interconnect, systems with multiple devices that must communicate, tend to use an integrated switching module. In such an arrangement, each device is connected to a switching fabric with a high speed interconnect. Typically a star configuration is implemented. This restricts the full-duplex ability of single device to simultaneously communicate with multiple devices. A single device, may however broadcast to more than one device, through the switching fabric as long as all intended destinations are capable of receiving.

3. Evolving Standards

For the implementation of high-speed data transport, the role of the RapidIO and PCI-Express continue expand and appear to lead in the development of support through protocol and circuit availability.

PCI-Express continues to be popular with proponents of the legacy PCI bus and with Intel orientated designers. While the PCI-Express is not fully backward compatible with the early PCI variants or with PCI-X, the protocol similarities are often sufficient to peak the interest of developers. The Intel suite of advanced processors utilizes the PCI-Express standard for processor interconnection and for interfacing to support circuitry end peripherals such as memory.

The Advanced Switching brand of Vitesse is based on PCI-Express. Vitesse provides a family of switch fabrics and support circuitry for use in designing PCI-Express systems.

The last year has seen significant proliferation of the RapidIO standard and associated interface circuitry. RapidIO is intended to provide a standard for inter-processor communication as well as board-to-board communications. Two variants are available: Serial RapidIO and Parallel RapidIO. The designer is free to choose either implementation but the choice is often made based on the distance between communicating devices. The Serial RapidIO implementation uses an 802.3 XAUI like transceiver in either a short run or a long run configuration. The long run configuration is essential for board-to-board communication. The Parallel

RapidIO used LVDS transceivers, is configurable as an 8-bit wide or 16-bit wide bus and scales in frequency from 250 MHz to 1 GHz providing a maximum throughput of 8 Gb/s. The Parallel RapidIO is typically only used for short device-to-device interconnect [1]. Tundra Semiconductor has developed a series of RapidIO switching fabrics as well as several bridge devices.

In addition to the development of standards for the high-speed interconnect protocol, several semiconductor and telecommunications companies have formed an alliance to provide a common physical implementation standard for the high-speed interconnect at 10 Gb/s. The alliance has formed the Unified 10 Gb/s Physical-layer Initiative (UxPI) [2]. Simply stated, the goal of the UxPI is to provide a common physical layer across all high-speed IO standards, organizations and markets.

4. High-Speed Circuit Techniques

The demand for speed and performance in broadband systems continues to increase. The increasing speed in high-performance ICs and the strong tendency in the market to use the existing infrastructures (e.g., FR4-dielectric boards, multi-mode fiber, legacy connectors, etc.) motivates circuit designers to overcome the non-idealities of the transmission channels and to push electrical interconnect speeds higher. Effects such as bandwidth loss, reflections and crosstalk can distort the signal to such an extent that robust data recovery requires equalizer-based backplane transceiver designs. The popular backplane transceiver designs in the 1 to 3Gb/s range use the power-efficient non-return to zero (NRZ) signaling schemes and equalization at the transmitter and/or receiver side. New signaling schemes, with better spectral efficiency, such as PAM4 and duobinary signaling, are of more interest as industry-standard data rates have passed 3Gb/s and approach 5 to 12Gb/s. Duobinary [3] is a type of partial response signaling that can be helpful in reducing the required bandwidth, as it allows for a controlled amount of ISI to be removed afterward. The duobinary signal bandwidth is 2/3 that of NRZ signaling (also known as PAM-2) and has only one cross-point between the symbols. This makes the clock recovery easier than for the PAM-4 scheme. PAM-4 [4,5] has 1/2 of the bandwidth of NRZ but suffers from interoperability issues and reduced voltage margins that exacerbate crosstalk concerns. This is due to the fact that a PAM-4 signal includes the maximum transition between the lowest and the highest levels, however, the duobinary signal only includes the transitions between adjacent levels. Therefore duobinary signaling has better immunity to crosstalk and reflection than PAM-4, which is proportional to the maximum transition. In [6] duo-

binary signaling is employed with a 10-tap x2-oversampled equalizer implemented in 90nm CMOS technology to achieve 12Gb/s over a 75cm low-k PCB trace. The measured eye height is 3dB larger than for NRZ signaling.

4.1 Equalization

In the last few decades, several equalization techniques have been proposed to compensate for the low-pass nature of the transmission channels. The transmission channels introduce inter-symbol interference (ISI) for high bit rates. Equalization can be performed in either the digital or analog domains, at the transmitter or at the receiver, with feed-forward or feedback topologies, and in a linear or non-linear manner. For very high data rates a combination of techniques are used to get the best possible results. Transmit equalization (often called pre-emphasis or de-emphasis) is a simple and often effective way of coping with dispersion-induced ISI. In transmit equalization low frequencies are attenuated relative to the Nyquist signaling frequency, thus flattening the overall system response and removing ISI.

The optimum receiver, in terms of symbol error rate (SER) is the maximum likelihood sequence estimation detector [7], but its high complexity makes it impractical for use in many applications. Linear equalizers (LEs) followed by a symbol-by-symbol detector are attractive in terms of reduced complexity, although these might excessively enhance the noise if the channel frequency response presents deep nulls [8]. In general linear equalizers address ISI but not crosstalk, as such equalizers amplify high-frequency noise as well as the signal. Non-linear equalizers, however, are capable of boosting the high-frequency signal energy while rejecting noise. The feed-forward equalizer (FFE) cancels the precursor ISI while a decision feedback equalizer (DFE) [9] uses a linear combination of past decisions to cancel the post-cursor ISI. In the analog domain, a FFE is realized by summing the outputs of a fixed-gain DC path and a variable-gain AC path, resulting in a continuous-time finite-impulse-response (FIR) filter [10] or an infinite-impulse-response (IIR) filter [11].

Decision feedback based receiver equalization (DFE) can be effective in dealing with configuration dependent reflections as well as ISI induced loss and dispersion. A linear DFE uses a FIR filter to cancel any ISI that is a linear function of past decisions. Recent results have shown that a linear DFE can perform as well or better than a partial response maximum likelihood (PRML) detector under certain conditions [12,13].

Nonlinear DFEs provide post-cursor ISI cancellation with reduced noise enhancement and are widely

recognized to offer better steady-state performance than linear equalizers [14]. In addition, the nonlinear DFE has an advantage of not reducing the transmit power at lower frequencies in order to achieve an equalization response. However, due to the presence of a nonlinear decision device inside the DFE feedback loop, erroneous decisions can result in error bursts that degrade SER performance. Also, a DFE cannot cancel precursor ISI and post-cursor cancellation is limited by filter length. The DFE can be implemented using either FIR or IIR filters. A continuous-time IIR structure has less area and consumes less power with respect to the FIR structure. At high data rates, the latency of the feedback loop in the standard DFE can present a serious bottleneck. Moreover, while feasible at around 6 Gb/s, the difficulty of feeding back the first tap fast enough currently precludes cancellation of the first post-cursor at higher bit rates (i.e. around 12Gb/s). Using an FFE and DFE together addresses these concerns [10,11,15-17]. In [10], a backplane link transceiver architecture, implemented in 0.13 μ m CMOS technology, incorporates a 4-tap FFE (FIR) that in conjunction with a DFE enables 6.25/12.5Gb/s data transmission. In [11] a 2-tap pre-emphasis network in the transmitter, along with a 1-tap FFE (IIR) and 3-tap DFE structure in the receiver is used in a 5Gb/s NRZ transceiver to achieve a BER of less than 10^{-15} in the presence of crosstalk.

The adaptation of transmit and receive equalization can be added to the link with minimal overhead. To enable system independent calibration of the transmit pre-emphasis, a common-mode back-channel is included in the link to enable communication of the updates in the reverse direction of the high speed data flow. A popular adaptive algorithm is sign-sign LMS (a derivation of the least-mean algorithm) [18]. As an algorithm that leads to a very simple implementation, it creates updates for the tap coefficients based only on the sign of the data and the measured error [19]. A transceiver core operating from 0.6 to 9.6 Gb/s using adaptive receive equalization with a 1-tap DFE followed by a linear equalizer, is described in [16]. The core dissipates only 150mW at 6.25 Gb/s. An analog adaptive equalizer implemented in [20] can compensate for loss in up to 30 inches of FR4 transmission line dissipates only 25mW.

Manufacturing variations, environmental conditions and voltage variations have significant impact on the performance of high-speed backplane systems. The links will have various trace lengths and via stub-lengths on the line, switch and backplane PCB modules and chip packages. Typically, the SerDes circuits on the IC are designed to minimize the impact of process, voltage and temperature variations on the performance of transmitters and receivers [21].

4.2 Clock and Data Recovery

The clock data recovery (CDR) circuit plays a major role in the serial link receiver by extracting the clock and regenerating the data from the input stream. Phase-tracking CDRs have been used for several Gb/s rates [22,23] because they do not suffer from phase quantization errors. Compared to other kinds of phase detection methods, the binary CDR is more suitable for high-speed operation than the linear CDR as it does not suffer from the timing offset caused by setup/hold-timing uncertainty of the sampler [24].

The jitter of a binary CDR circuit is set by the minimum resolution of the phase interpolator because of its bang-bang operation [25]. The high gain of the bang-bang phase detector at zero phase error suppresses the static phase error due to charge-pump offset and enables superior phase alignment for re-timing the data symbols [26]. The bang-bang CDR measures the phase error using a single slicer that has a zero threshold. In the locked condition, the falling edge of the VCO is aligned to the zero crossings of the data, and the rising edge of the clock retimes the data. In the case of an ideal CDR circuit with no delay, which immediately updates the timing, the recovered clock jitter is limited by the minimum resolution of the phase interpolator. If there are delays in the recovery loop, the jitter will be greater than the minimum resolution as these delays prevent immediate timing updates. This refers to the jitter tolerance of the CDR at high frequencies, which is only 0.15 UI.

Conclusions

This paper attempts to provide an overview of recent advances in the field of high-speed data interconnect. The trends point to a reduction of the interconnect bus width and an ever increasing clock frequency. The clock frequency must increase by a ratio greater than the bus width reduction ratio to allow for an improvement in average data through-put. It would appear that the PCI-Express and RapidIO standards are the more popular implementations from the protocol perspective. The past year has seen an increase in interest and membership of the UxPI, originally launched in 1999, in an attempt to drive the standardization of the physical layer for 10 Gb/s interconnect. The final section of the paper provides an overview of the more recent circuit techniques utilized to implement the physical layer of high-speed interconnection. Undoubtedly, equalization and clock/data recovery will continue to play significant roles in the evolution of multi-Gigahertz serial data communication.

References

- [1] Sam Fuller, "RapidIO, The Embedded System Interconnect", John Wiley and Sons, New York, 2005.
- [2] "UxPI, Unified 10 Gbps Physical-layer Initiative, Executive Overview", www.uxpi.org
- [3] Sonntag et al., "An Adaptive PAM-4 5Gb/s Backplane Transceiver in 0.25um CMOS," *IEEE CICC*, pp. 363-366, May, 2002.
- [4] S. Wu et al., "Design of a 6.25Gb/s Backplane SerDes with TOP-down Design Methodology," *DesignCon2004*, Feb., 2004.
- [5] J. L. Zerbe et al., "Equalization and Clock Recovery for a 2.5-10-Gb/s 2-PAM/4-PAM Backplane Transceiver Cell," *IEEE J. Solid-State Circuits*, vol. 38, pp. 2121-2130, Dec., 2003.
- [6] K. Yamaguchi, "12Gb/s Duobinary Signaling with x2 Oversampled Edge Equalization," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 70-71.
- [7] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
- [8] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349-1387, Sept. 1985.
- [9] M.E. Austin, "Decision-feedback equalization for digital communication over dispersive channels," M.I.T. LL, Aug. 1967.
- [10] P. Landman, "A Transmit Architecture with 4-Tap Feedforward Equalization for 6.25/12.5Gb/s Serial Backplane Communications," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 66-67.
- [11] N. Krishnapura, "A 5Gb/s NRZ Transceiver with Adaptive Equalization for Backplane Transmission," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 60-61.
- [12] K. Han and R. Spencer, "Comparison of different detection techniques for digital magnetic recording channels," *IEEE Trans. Magn.*, pp. 1128-1133, Mar. 1995.
- [13] P. S. Bednarz, N. P. Sands, C. S. Modlin, S. C. Lin, I. Lee, and J. M. Cioffi, "Performance evaluation of an adaptive RAM-DFE read channel," *IEEE Trans. Magn.*, pp. 1121-1127, Mar. 1995.
- [14] C. Belfiore and J. Park, Jr., "Decision feedback equalization," *Proc. IEEE*, vol. 67, pp. 1143-1156, Aug. 1979.
- [15] M. Sorna, "A 6.4Gb/s CMOS SerDes Core with Feedforward and Decision-Feedback Equalization," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 62-63.
- [16] D. Yokoyama-Martin, "A 0.6 to 9.6Gb/s Binary Backplane Transceiver Core in 0.13um CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 64-65.
- [17] R. Payne, "A 6.25Gb/s Binary Adaptive DFE with First Post-Cursor Tap Cancellation for Serial Backplane Communications," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 68-69.
- [18] Lei Lin, Peter Noel and Tad Kwasniewski, "Implementing a Digitally Synthesized Adaptive Pre-emphasis Algorithm for use in a High-Speed Backplane Interconnection, CCECE 2004, Niagara Falls
- [19] J. Zerbe et al., "Comparison of adaptive and non-adaptive equalization methods in high-performance backplanes," *DesignCon 2004*, Feb. 2004.
- [20] S. Gondi, "A 10Gb/s CMOS Adaptive Equalizer for Backplane Applications," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 328-329.
- [21] R. Kollipara et al., "Comparison of adaptive and non-adaptive equalization methods in high-performance backplanes," *DesignCon 2004*, Feb. 2004.
- [22] P. Larsson, "An offset-cancelled CMOS clock-recovery/demux with a half-rate linear phase detector for 2.5 Gb/s optical communication," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2001, pp. 74-75.
- [23] S. B. Anand and B. Razavi, "A CMOS clock recovery circuit for 2.5 Gb/s NRZ data," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 432-439, Mar. 2001.
- [24] S.-H. Lee et al., "A 5-Gb/s 0.25-um CMOS jitter-tolerant variable-interval oversampling clock/data recovery circuit," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1822-1830, Dec. 2002.
- [25] M. Fukaiishi et al., "A 20-Gb/s CMOS multichannel transmitter and receiver chip set for ultra-high-resolution digital displays," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1611-1618, Nov. 2000.
- [26] R. Walker, "Designing Bang-bang PLLs for Clock and Data Recovery in Serial Data Transmission Systems," *Phase-Locking in High-Performance Systems*, IEEE Press, pp. 34-45, 2003.