Design Consideration of 6.25 Gbps Signaling for High-Performance Server

Jian Hong Jiang, Weixin Gai, Akira Hattori, Yasuo Hidaka,
Takeshi Horie, Yoichi Koyanagi, Hideki Osone

Platform Innovation Group
Fujitsu Laboratories of America, Inc.
1240 East Arques Avenue
Sunnyvale, CA 94085
Tel: 408-530-4600
Fax: 408-530-4501
e-mail:jianhong.jiang@us.fujitsu.com

Abstract - As network data rate increases rapidly, high-speed signaling circuits for server communication pose many design challenges due to various system requirements using different interconnect mediums. This paper discusses main problems and solutions of high-speed circuits for server interconnect. Then, it presents a high-speed circuit implementation for such interconnect using 90nm CMOS technology that achieved data rate at 6.25 Gbps in a backplane environment.

1. INTRODUCTION

Modern servers must meet diverse set of I/O network requirements. Server networks such as CAN (Cluster Area Network), SAN (Storage Area Network), and LAN (Local Area Network) differ in applications, but all strive for same desirable performance features such as low latency, high throughput, high density and low power [1]. Within the server network, components are connected through several types of interconnects. These interconnects must meet requirements for scalability, distance, and low bit error rate (BER) specified by each standard. Currently, there are three main interconnect mediums that link server networks. First is PCB trace, which tends to be for short distance up to 50cm. Second type is backplanes, which consist of minimum two connectors and three segments of PCB traces [2], with length up to 1m. Third medium is high quality cable such as InfiniBand cable, which can be up to 15m between servers.

Many issues arise in designing high-speed signaling circuits for server links due to the server's requirements and interconnect environments. First is power consumption in each I/O link because server network throughput is directly related to number of I/O ports. Another is the noise both on chip and in the communication channels including reflection from impedance mismatches, and cross talk between the transmitting signals. These noise components reduce the effective transmitting voltage and timing margins, which cause high BER, and system failure. Last, all copper server interconnect components suffer from frequency related signal loss due to skin effect and dielectric loss. As transmitting data speed and distance between servers increase, power, noise, and transmission channel frequency-dependant losses become dominant issues of high-speed signaling circuits.

2. HIGH-SPEED SIGNALING DESIGN APPROACH FOR SERVERS

To communicate between different server network components over various channels, signaling must be carefully defined and implemented. A signaling is the methods of transmit digital information from one location to another location correctly and effectively [3]. There are several signaling architectures. One way to transmit signal is to use a parallel interface where all outputs are directly transmitted to receiving devices, there are as many pins as number of transmitting signals. This approach, while simple in implementation, has several drawbacks. Server system usually has limited interconnect channels, which do not scale well with increasing signals. For example, PCB traces, cable wires and their associated connectors numbers are not easily scaled up due to both space and costs. Another issue is the large signal and clock skews due to variations in interconnect paths. The high pin count inside chip also means additional circuits and power. Another choice to transmit signals is to use a serialized interface, which puts many outputs in a serial data string by carefully clocking data into a multiplexed path. It can transmit very high data rate through the communication channel while requires just few pins. To achieve high bandwidth, data rate is higher in serialized transmission per channel comparing to parallel one, and it is the only choice when system requirements are considered. There are several common signaling methods of transmitting and receiving the high-speed data in CMOS technology. A preferred one is a differential, small swing (transmitter voltage amplitude often in the range of 500mv-1600mv peak to peak differential), current-mode CMOS signaling, which consumes less power, has better noise rejections, and outputs better quality signals.

A design example of high-speed signaling I/O macro fabricated in 90nm CMOS process is shown below to illustrate various signaling circuit components and techniques that achieved good performance.

3. AN EXAMPLE OF IMPLEMENTATION FOR 6.25 GBPS DATA RATE SIGNALING CIRCUITS FOR SERVER

We have designed a multi-gigabit/s CMOS serial transceiver operating at 6.25 Gbps that provides high bandwidth data communication. This macro can also support

1-4244-0630-7/07/$20.00 ©2007 IEEE.

various high-speed I/O transmission standards such as gigabit Ethernet, or 3.125 Gbps XAUI/CX4 over different transmission mediums such as cable, or backplane.

The high-speed transceiver macro consists of 4 transmitters, 4 receivers and a clock unit.

3.1. TRANSMITTER CIRCUITS

The transmitter block (Fig. 1) receives 32 bits data and a clock from the core interface, and then synchronizes them with local transmitter clock. The first stage high-speed multiplexer and synchronizer unit performs 32:8 multiplexing and local clock domain data synchronization. After that, either the data generated by PRBS pattern generator or the multiplexed data is processed through delay stage controlled by logic. Delay stage outputs input data that has same or opposite polarity with different bit delay to form pre-emphasis with later stages. The 8:1 multiplexer finally converts data rate up to 6.25 Gbps and provides inputs to the transmitter output stage. The output driver converts the full swing high-speed digital inputs into smaller analog amplitude waveforms to drive the communication channel and receiver.

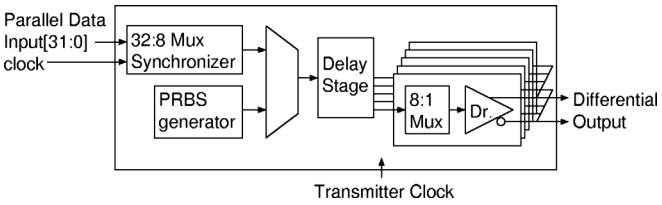

Fig. 1. One channel transmitter block diagram.

To minimize power consumption, we designed the entire mixed-signal transmitter using a single 1.2V power supply. To reduce power even further, the transmitter output driver stage uses an LVDS type circuit instead of a conventional CML type of driver. Fig. 2 shows the LVDS output driver circuits with interconnect, and a terminated receiver resistive load with its common mode control node VDR. For any data rate, as soon as the switching IN and IP are completed, only one current path flows through a PMOS, termination resistors and a NMOS and the power is I_0^2*Rt. For the same voltage swing, the LVDS type of driver consumes half of the current of a conventional CML type driver, which has two conducting current paths with power consumption of $2*I_0^2$*Rt. However, there is a big challenge for LVDS driver design when supply voltage is as low as 1.2V, because various I/O standards such as XAUI/CX4, 6.25G OIF require single end voltage swing at package node up to 0.6V. This requirement means that the driver must provide as much as 0.7V single end voltage at chip output for data transmission. In our design, the unusable overhead voltage across both current sources was minimized, and control circuits were applied to keep output centered at half supply voltage.

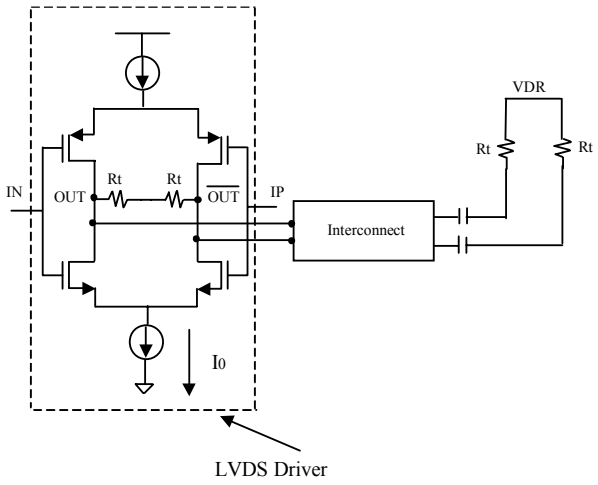

Fig. 2. LVDS output driver.

The transmitter output stage has five identical output fingers, each consisting of a pre-driver and an associate output driver (Fig. 3). The output fingers can be turned on or off depending on pre-driver logic values. Table 1 shows the logic function of the associated pre-driver in Fig.3 of an output finger. To address frequency-dependant losses in the transmitter driver itself and the server interconnect, the multiplexer, delay stage, and the driver output stage form pre-emphasis function. The driver output fingers, which are controlled by pre-drivers, can be configured as a multi-tap discrete FIR filter when selected individually or collectively. The resulting filter has the following property in time domain:

$$Y[n]=C0*X[n+DL0] + C1*X[n+DL1] + C2*X[n+DL2] + C3*X[n+DL3] + C4*X[n+DL4] \quad (1)$$

Where, C0~C4 = -63 to 63, DL0~DL4 = 0 to 7, n = integer index of UI, X[n] = FIR filter input data sequence, Y[n] = FIR filter output at time n. The coefficients C0~C4 are controlled by their associated current sources. Each current source is a digital to analog converter (DAC) with six-bits resolution.

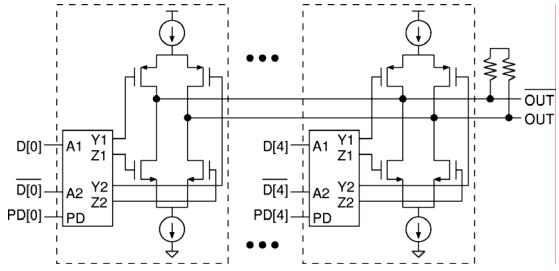

Fig. 3. Driver output stage with pre-driver control.

Table I
PRE-DRIVER LOGIC CONTROL FUNCTIONS

PD	A1	A2	Y1	Y2	Z1	Z2
0	0	1	0	1	0	1
0	1	0	1	0	1	0
1	X	X	1	1	0	0

Besides the flexibility in forming FIR filter, the driver output stage can control the output swing and loadings as needed by different requirements. When an output driver finger is turned off, there is total isolation between output nodes and its current source nodes, which have large capacitance. When the transmitter has to output large amplitude but at slower data rate, as many as 5 fingers can be turned on, all providing driving currents. If smaller amplitude and fast data rate is desired, some of the output fingers can be turned off. This feature effectively provides an additional slew rate control capability.

3.2. RECEIVER CIRCUITS

For receiver design, a block diagram is shown in Fig. 4. The outputs of the linear equalizer are sampled by 2 sets of 2-ways interleaved decision latches to obtain both data and boundary information. Data are sampled at center of eye diagram and boundaries are sampled at transitions. After sampling, data and boundary values are de-multiplexed down to 32+32 way. Clock recovery using digital filter and phase interpolator recovers clock from incoming data so that the sampling clock is aligned correctly. To check if the receiver achieves required BER, a PRBS checker in receiver channel is implemented.

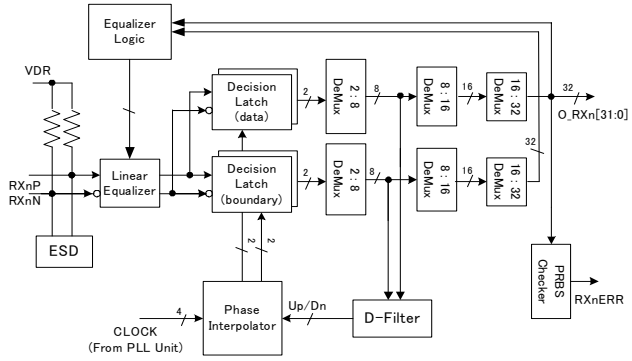

Fig. 4. One channel receiver block diagram

The linear equalizer is a first order system. This unit compensates both channel and receiver frequency-dependant losses and amplifies the receiver input signal amplitude, so that they meet decision latch sensitivity requirement. The frequency transfer function up to 3.125 GHz (6.25 Gbps) can be described as:

$$H(s)=a*s+b \quad (2)$$

The equalizer consists of a DC and a first order derivative path that realizes the above function. A source-coupled pair with capacitive degeneration provides the high pass filter function for the derivative path, while a simple source coupled pair provides DC value with similar path delay [4]. Gains for each path are adjusted in blocks a and b so the two paths can be mixed later. The block diagram of the equalizer and the frequency responses of two paths before mixer are shown in Fig. 5. There is also an equalizer bypass option. When the frequency losses in the interconnect are small and receiver decision latches can sample the incoming data correctly, the equalizer DC and derivative path can be shut off to save power.

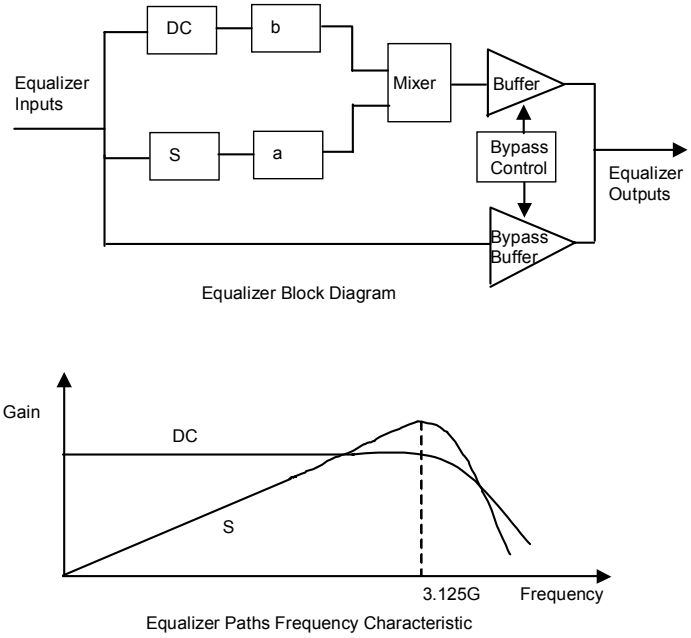

Fig. 5. Linear equalizer architecture and frequency response.

An equalizer logic unit using sampled boundary and data information forms a negative feedback loop with linear equalizer. This control unit will adjust various linear equalizer parameters so that correct data are provided before decision latches.

3.3. CLOCK UNIT

Clock unit provides multiple frequency and phase clocks for both the transmitter and receiver channels. The main functional component in the clock unit is PLL clock generator, frequency dividers and clock buffers. Since the PLL generates high frequency clocks for both transmitters and receivers, any noises generated in it are propagated to circuits where a clock signal is required. To minimize PLL clock noise, we carefully designed key PLL component, a voltage controlled oscillator (VCO), which is a main source of high frequency noise. The VCO of the PLL uses an on chip spiral inductor and varactor to form an LC tank oscillator. A precision bandgap reference generated bias current sources are used for fine control of VCO. This VCO

has small phase noise. For clock delivery, all buffers use fully differential circuits to reduce power and coupling noise from other signals in the delivery paths.

3.4. MEASUREMENTS

The measurements were performed by sending PRBS23 data pattern from transmitter, through backplane test kit to the receiver. In one test case, the measurement setup included approximate 40" of backplane PCB trace and 2 backplane connectors for the complete test path. The transmitter chose a 3 taps pre-emphasis for filter function in this experiment. The resulting transmitter near end (Fig. 6) and far end (Fig. 7) eye diagrams are shown below. The far end eye diagram in Fig 7 shows clear eye opening. The transceiver in this case has achieved BER < 10^{-15}.

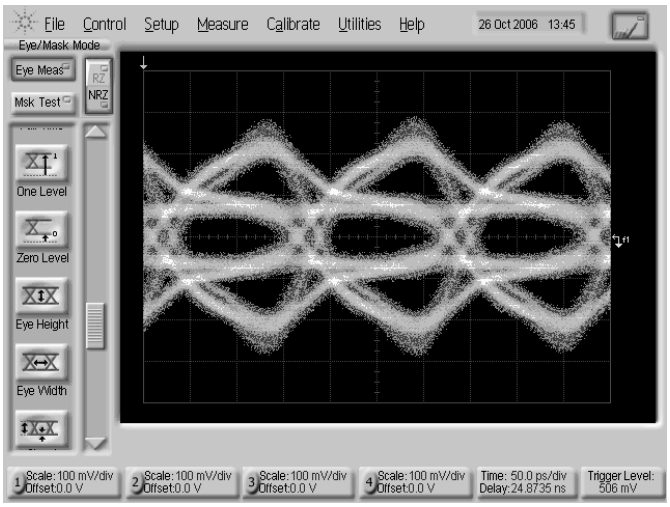

Fig. 6. Transmitter near end eye diagram.

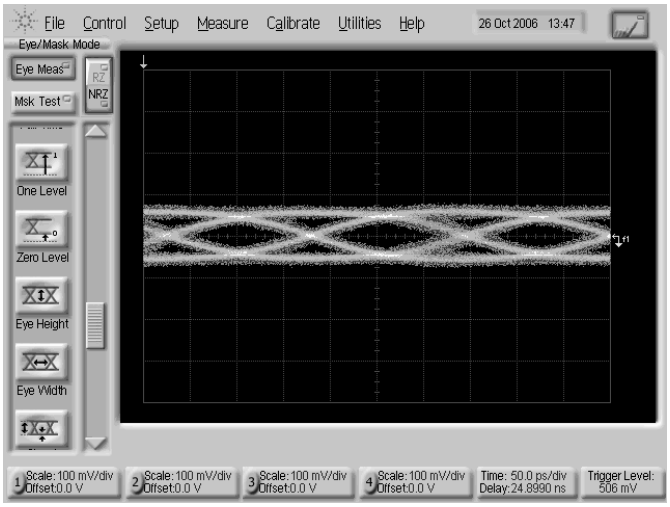

Fig. 7. Transmitter far end eye diagram.

4. SIMULATION CHALLENGE FOR HIGH-SPEED SIGNALING CIRCUITS

The simulation for this transceiver was quite challenging. Many signaling circuits containing feedback loops require long simulation time because they operate in frequency that are several magnitude lower than other part of high-speed circuits. Since there were no quick ways to force those low frequency loops settle at their final values, SPICE simulation consumes long time, in some cases becomes impossible when the loop time constants are in milliseconds. Other major problems we had to deal with are the high frequency transmission line model inaccuracies, which often differ from real measurements. With transmission line elements included in circuit, the mix-mode circuit simulator did not work; so simulation could only be done in SPICE, which often took days to complete. The entire layout of high-speed analog circuit were manually drawn from scratch because lack of ways to automatically generate them. In order to ensure design reliability, a lot of efforts were put in meeting electrical migration rules, because of lack of accurate and user-friendly commercial tools. As high-speed interconnect get further complicated, more efficient tools and better design methodologies addressing these problems would greatly improve design qualities and increase productivities.

5. CONCLUSION

As projected data rate increases further, it is even more important for high-speed signaling technology to focus on improving in the areas of transmission bandwidth, frequency-dependant loss, overall power consumption and signal noise. Many aspects of high-speed signaling circuit techniques for servers are discussed. A serial link that achieves low power by using a single 1.2V supply and compensates for various frequency-dependant losses has been shown to work at 6.25 Gbps in a backplane environment.

ACKNOWLEDGEMENTS

The development was partially funded by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] R. J. Recio, "Server I/O Networks Past, Present, and Future," *Proceedings of the ACM SIGCOMM 2003 Workshops*. Pp. 163 - 178.

[2] D. N. de Araujo, E. Matoglu, M. Cases, N. Pham, "Complex High Speed Links in Server Environments," *Proceedings of 2005 Electronic Components and Technology Conference*, Vol, 2, pp. 1756 - 1761.

[3] W. J. Dally, J. W. Poulton, *Digital Systems Engineering*, Cambridge University Press.

[4] W. Gai, Y. Hidaka, Y. Koyanagi, J. H. Jiang, H. Osone, and T. Horie, "A 4-channel 3.125Gb/s/ch CMOS transceiver with 30dB equalization," *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 138 - 141, June 2004.