A 2-byte Parallel 1.25 Gb/s Interconnect I/O Interface with Self-configurable Link and Plesiochronous Clocking

● Kohtaroh Gotoh   ● Hideki Takauchi   ● Hirotaka Tamura
(Manuscript received January 13, 2000)

An I/O transceiver for scalable multiprocessor systems[1] has been developed with a high parallel bandwidth (1.25 Gb/s × 2-byte) and low latency (7.4 ns). The transceiver performs plesiochronous clocking, and compensates for skin-effect cable loss and inter-wiring skew across cable connections of 20 m in length. We used a phase-interpolator-based clocking scheme that ensures a high skew-adjustment resolution (25 ps ± 5 ps adjustment step) and plesiochronous clocking and can tolerate slight differences in frequency between the incoming and internal reference clocks. A Differential Partial Response Detection (DPRD) receiver has also been developed to ensure a low latency equalization for a skin-effect cable loss of up to 10 dB. The receivers are equipped with deskew circuitry to tolerate an inter-wiring skew of up to 6.4 ns for 20 data bits. The data rate, driver output level, and receiver clock phase are adjusted automatically by a logic sequencer called the "Basic control." The sequencer maximizes the data rate and the minimizes power consumption without external manual adjustments, and can adapt to a wiring environment ranging from on-board PCB traces to 20 m twisted-pair cables. We designed a test chip for parallel-link interconnection using a 0.25 μm CMOS process and confirmed that it was capable of 1.25 Gb/s 2-byte parallel signal transmission over a 20 m AWG 28 twisted-pair cable.

1.   Introduction

The interconnection issue is increasingly dominating modern high-performance digital systems[1] that link commodity microprocessors, memories, and I/O components (**Figure 1**). High-performance multiprocessing servers, for example, cashe-coherent symmetric multi-processors (SMPs),
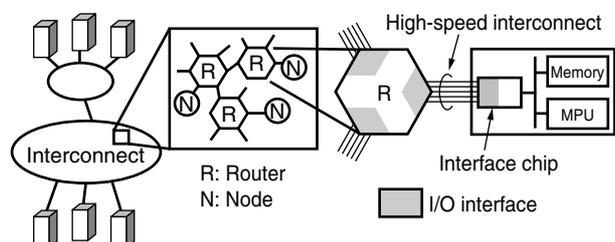
Figure 1
Interconnect of multi-processing servers.

require a high bandwidth as well as a low-latency I/O design.[2,3] The cabinet-to-cabinet interconnection for servers requires an equalization capability to compensate for the skin-effect cable loss to permit long twisted-pair cable connections of up to 20 m. The multiple reference-clock domains resulting from the interconnection of two cabinets which have their own crystal oscillators require plesiochronous clocking in which the clock frequencies on each side are slightly different.

In this paper, we propose a 1.25 Gb/s 2-byte parallel-interconnect I/O interface that meets this requirement. A Differential Partial Response Detection (DPRD) receiver enables a low-latency equalization to compensate for a skin-effect cable loss of up to 10 dB to permit a 20 m twisted-pair cable connection. A phase interpolator and phase interpolator-based clock recovery loop provide a
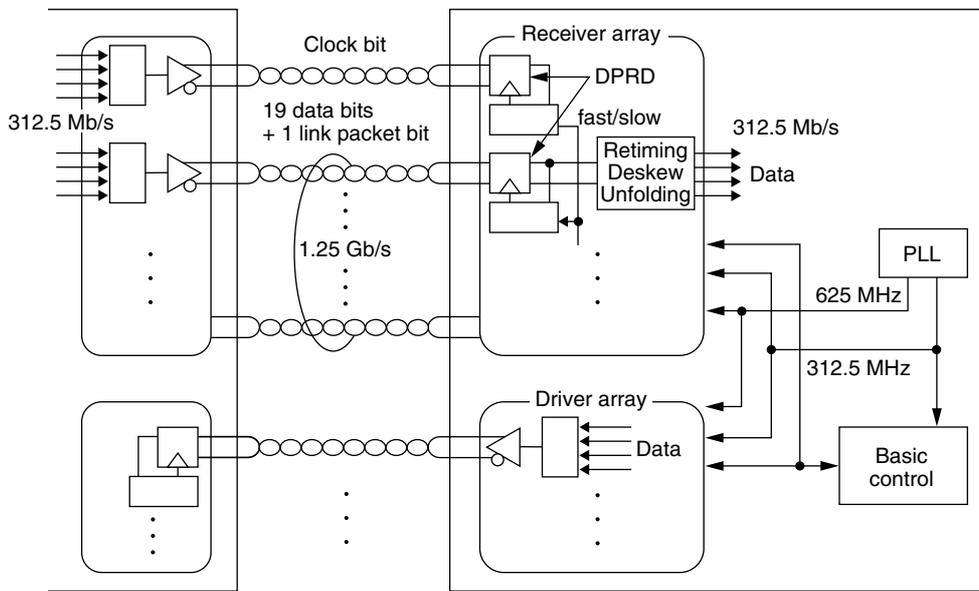
Figure 2
Parallel I/O link.

high skew-adjustment resolution and plesiochronous clocking, respectively.

2. I/O link design

The interconnect design we propose consists of 21-bit driver and receiver arrays, a logic sequencer we call the "Basic control," and a PLL (**Figure 2**). The driver/receiver arrays have a dedicated clock line, a 1-link packet bit, and 19 data signals that include 2 ECC bits and 1 Tag bit. The packet bit is used for handshaking in the I/O link tuning sequence. All of the bits are transferred using a differential mode. A single core PLL provides two different core clocks, 625 MHz and 312.5 MHz, to the I/O interfaces. The logic circuits, including the Basic control, operate at the 312.5 MHz core clock. The data from the core logic, which is synchronized with the 312.5 MHz core clock, is applied to the driver unit, which performs 4-to-1 multiplexing to output a 1.25 Gb/s data stream. The incoming 1.25 Gb/s data is subjected to 1-to-4 demultiplexing and alignment to a single incoming clock through the DPRD receiver's retiming and deskewing circuits and is then sent to the core logic. A phase interpolator in each receiver unit compensates for data-to-clock skew and provides an incoming-data sampling clock to the respective DPRD receiver. The clock recovery loop in the clock bit receiver tracks the incoming clock signal and outputs a phase code for the entire data receiver. The phase code enables the data receiver to lock onto the incoming clock. The Basic control controls the logic portion of the I/O interface and performs I/O link tuning.

3. Circuit design
3.1 Driver unit

The driver unit consists of a 4-way interleaving pre-driver and main output stages that perform the 4-to-1 folding operation to output a 1.25 Gbps differential data stream (**Figure 3 (a)**). The pre-driver stage contains four data registers which receive 4-bit data from the core logic synchronized with the 312.5 MHz core clock. It also contains a dynamic-type pre-driver operated by a 4-phase 312.5 MHz clock with a phase difference of $\pi/2$.

The main output stage employs a high-output-impedance push-pull output to reduce current consumption to a level lower than that of

(a) 4-way interleaving output stage that performs
4-to-1 folding operation

(b) Pre-driver and CMOS push-pull output stages
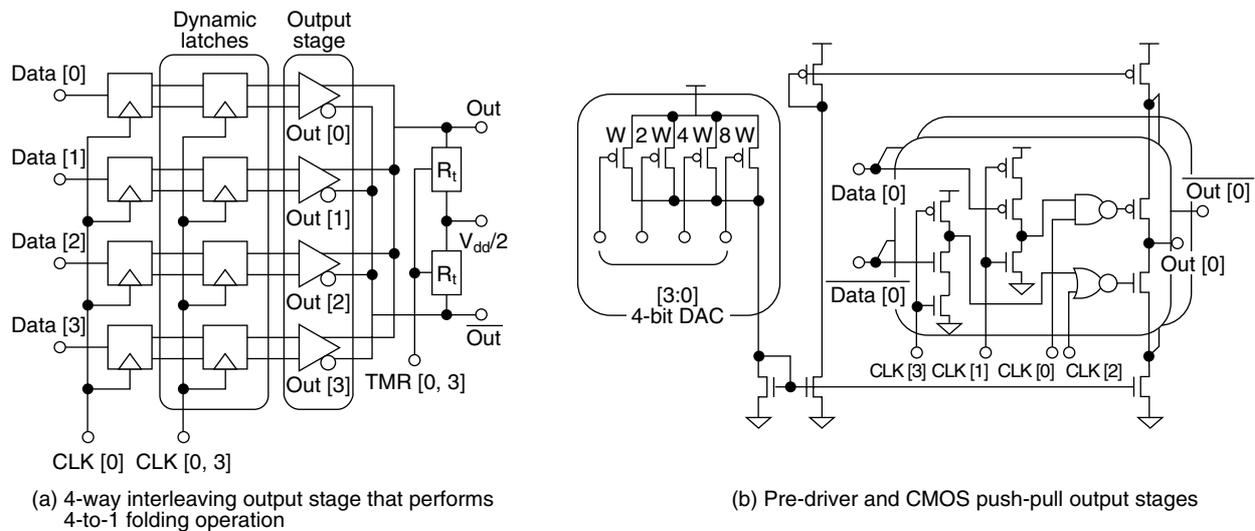
Figure 3
Driver circuit.
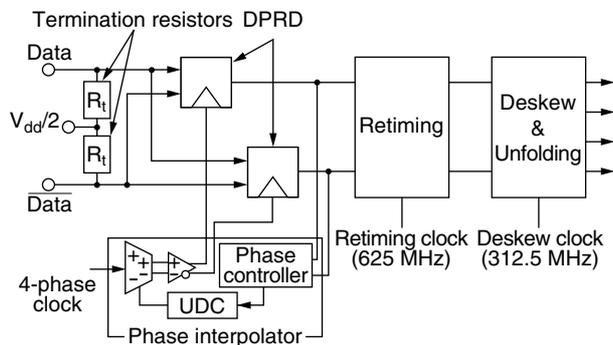
Figure 4
Receiver unit block diagram.

conventional resistor loads and NMOS current steering type transmitters (**Figure 3 (b)**). By using a dynamic-type data latch operation in synchronization with the 4-phase clock of the pre-driver stage, the output stages ensure 4-way-interleaving, a high output impedance, and a high driving current. To match the output impedance with a 50-ohm cable impedance, the output stage is parallel-terminated by on-chip CMOS transfer gate terminators. The termination resistances are controlled by a 4-bit binary code, TMR [0, 3], and are adjusted to the value of the external 50-ohm reference resistor by feedback control. The adjusted resolution is within ±5% of the value of the

reference resistor. The output current is digitally controlled using a PMOS current-source DA converter which is adjusted over the range from 0 to 21 mA using a 4-bit binary value. The adjustment is done by applying a differential DC offset current to the receiver input and detecting the input current level to ensure a signal voltage of 250 mV. This adjustment is completed during a self-configured power-on initialization and compensates for the cable loss from a PCB board trace to a 20 m twisted-pair cable while maintaining minimum current consumption. The clock skew fluctuation resulting from a supply voltage variation of 2.25 to 2.75 V was estimated by SPICE simulations to be 160 ps in the 1.25 Gb/s data stream output.

3.2 Receiver unit

The receiver unit consists of on-chip termination resistors and the DPRD receiver, phase interpolator, and retiming and deskew circuits (**Figure 4**).

The differential input is terminated by the termination resistors, which are identical to those in the driver, and applied to 2-way interleaving receivers. The phase interpolator provides a 2-phase data sampling clock with $\pi/2$ phase separation to each DPRD receiver to ensure continuous bit stream detection. Through the retiming and

deskew circuitry, the 2-way interleaved data is aligned to a single clock to compensate for bit-to-bit cable skew. The retiming circuit compensates for the cable skew within a 1-bit time period, while the deskew circuit compensates for the skew beyond a 1-bit time period.

3.3 DPRD receiver

Skin-effect resistance results in an attenuation of 6.8 dB over a 20 m AWG 28 twisted pair cable at a frequency of 625 MHz, which in turn reduces the signal bandwidth and increases the associated inter-symbol interference (ISI). Some equalization schemes have been proposed to eliminate ISI, for example, transmitter pre-emphasis.[4),5)] We implemented an equalization capability in the receiver using an 1-xD operation to compensate for the high-frequency loss (**Figure 5 (a)**). One advantage of the receiver equalization is that it enables a reduction of the power consumed in the driver for pre-emphasizing the high-frequency component of the driver output signal. Another advantage is that the capacitive coupling node in the receiver input eliminates low-frequency common-mode noise. A 1-xD operation is performed on the receiver input signal, where x is a positive number less than unity and D is a 1-bit time-delay operator. This eliminates the ISI and, as a result,

compensates for the high-frequency cable loss (**Figure 5 (b)**).

It has been reported that a PRD receiver can support a low-latency equalization scheme.[6)] In this study, we developed a differential-type PRD receiver for ISI elimination which consists of coupling capacitors and a differential latch-type sense amplifier (**Figure 6**). The differential input terminal is capacitor-coupled to the latch input nodes. The 1-xD operation is performed using the coupling capacitors and CMOS transfer-gates operated with a 2-phase interleaved clock (ϕ_1 and ϕ_2). At the previous bit time, ϕ_1, the input node voltages of the latch amplifiers are reset to a precharge level, V_{tt}. Coupling capacitors C_1 and C_2 are charged to V_{tt} and the differential signal line voltages, respectively (Figure 6 (a)). In the decision period, ϕ_2, the capacitors are connected in parallel and a weighted summing of the previous

(a) 1-xD operation   (b) ISI elimination in receiver input
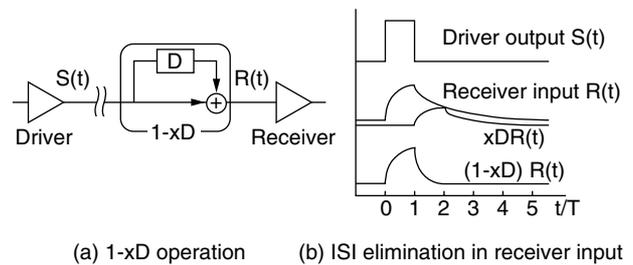
Figure 5
Inter-symbol interference (ISI) elimination.

(a) Pre-charge operation in $(n{-}1)_{th}$ bit time   (b) Data decision operation in n_{th} bit time
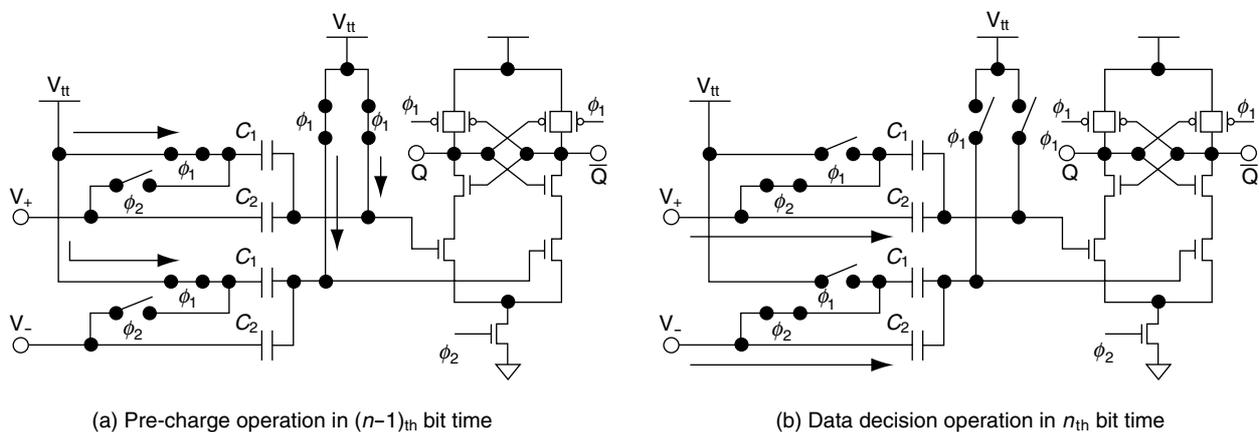
Figure 6
DPRD circuit and its 1-xD operation.

signal voltage and the reference voltage is performed in the latch amplifier input node (Figure 6 (b)). This can be expressed as:

$$V_{in} = V_n + \frac{C_1}{C_1 + C_2}(V_{tt} - V_{n-1}).$$

This 1-xD operation eliminates the ISI in the input voltage level of the latch amplifier, and the interleaving receiver operation reduces the external latency to zero. The data-clock skew tolerance is estimated to be 650 ps at 1.25 Gb/s according to SPICE simulations.
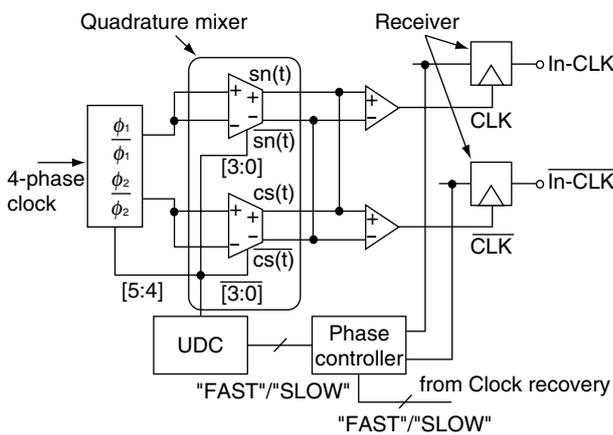
Figure 7
Phase interpolator.

3.4 Phase interpolator

The receiver clock is generated by a phase interpolator which generates an incoming data sampling clock in the DPRD receiver. The phase of the incoming data sampling clock is adjusted over the range from 0 to 2π with a 6-bit resolution.[7] The phase interpolator consists of a phase controller, a 6-bit binary up/down counter (UDC), quadrature mixers, and differential comparators (**Figure 7**). Four-phase, 625 MHz clocks with a phase difference of $\pi/2$ are sent to the mixer through the clock selector. The 2-phase current clocks are mixed with a weight controlled by the UDC code and applied to differential comparators. The receivers sample the incoming clock signal at the rising edge of the comparator output clock, and the UDC code is increased or decreased according to the sampling data so that the phase interpolator clock is adjusted to the incoming clock.

The quadrare mixer employs a differential current driver and a pMOS current-source DA converter (**Figure 8**). To guarantee a monotonous DAC output current, we employ 1-bit binary and 7-level thermometer codes during circuit implementation. The thermometer code is used as a
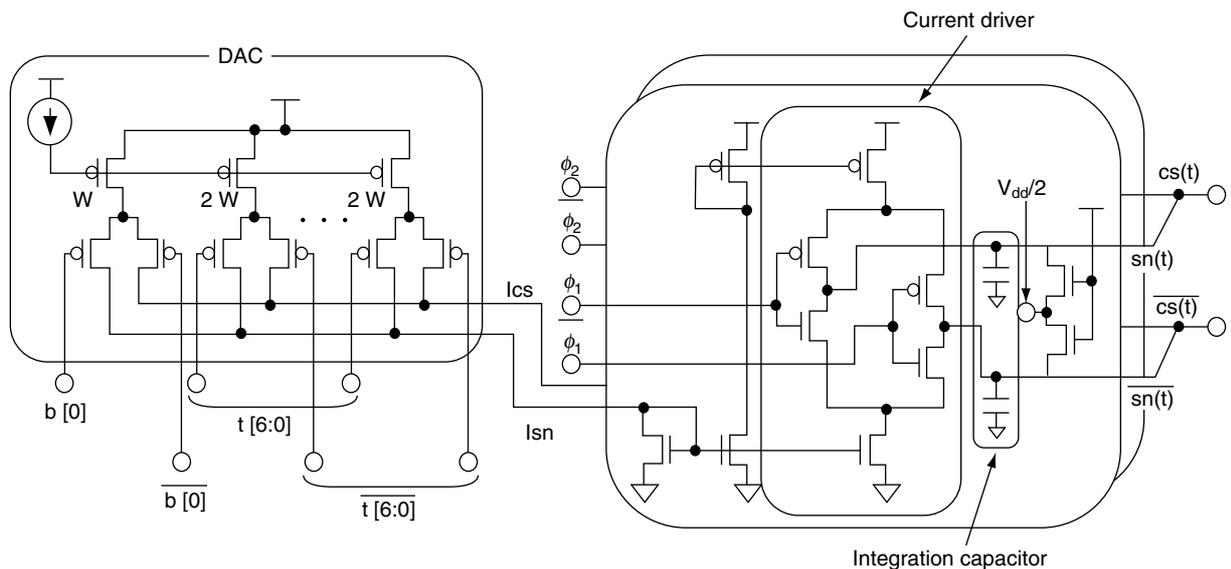
Figure 8
Quadrature mixer circuit in phase interpolator.
Circuit performs 2-phase waveform mixture controlled by DAC currents.

value compatible with the upper 3-bit value of the 4-bit binary value. The differential clock is sent to the current drivers, which in turn send a differential and square wave current clock to the integration capacitors. NMOS clamp transistors on the output node of the mixer compensate for the voltage shift resulting from conductance variations between the pMOS and nMOS current drive transistors and also compensate for the associated clock phase shift. The capacitor integrates the clock current and generates a triangular voltage waveform. The two-phase driver outputs are mixed with the weighted sum controlled by the DAC output currents, Isn and Ics.

The resulting differential voltage across the integration capacitors can be expressed as a com-bination of $(1-y)$ times sn(t) and y times cs(t), where y is a phase mixing factor ranging from -1 to 1 (**Figure 9 (a)**). By changing the y value, the comparators produce differential internal clock signals with a phase-adjustment range of 2π. The amplitude of y and the associated phase in the $\pi/2$ range is defined by the lower 4-bit value of the UDC code, while the upper 2-bit value selects the quadrant (**Figure 9 (b)**). This 6-bit code enables a phase-adjustment step of 25 ps. SPICE simulation shows that the phase is increased or decreased in steps of 25 ps ± 5 ps (**Figure 10**). Skew fluctuation of the phase interpolator out-clock resulting from ±10% variations of the 2.5 V V_{dd} was estimated to be 164 ps, while PLL jitter was estimated to be 128 ps. The total skew fluctuation is smaller than the 650 ps data-to-clock skew tolerance of the DPRD receiver.
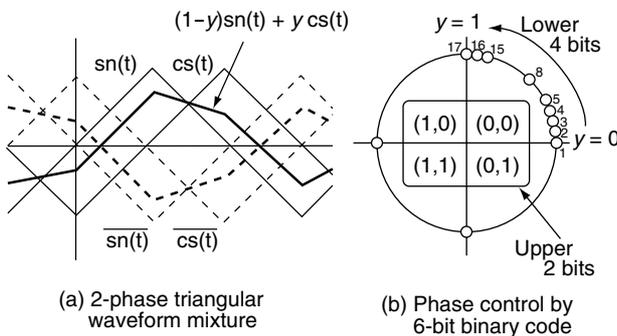
Figure 9
Phase control operation in phase interpolator.

Figure 10
Skew adjustment step versus phase control code in phase interpolator.

(a) Clock recovery loop scheme
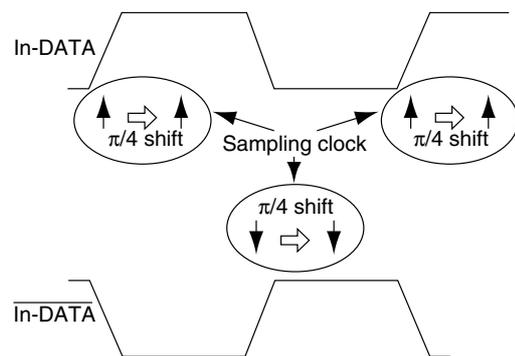
(b) Phase detection and data receiving scheme using $\pi/2$ shift

Figure 11
Clock recovery loop.

3.5 Clock recovery

The clock recovery loop compensates for the phase error between the data sampling and incoming clocks, which are derived from different crystal oscillators (**Figure 11 (a)**). In the power-on initialization, clock signals are applied to each data bit receiver and the internal clock phase is shifted by increasing or decreasing the UDC control code for the phase interpolator until the 0-to-1 boundary in the incoming clock is found. After completing the adjustment, the clock phase of the data bits is shifted by $\pi/2$ so as to sample the data at the center of the data eye, thereby compensating for the data-to-clock skew (**Figure 11 (b)**).

During the normal data-receiving state, the interpolator in the clock bit tracks the incoming clock by the 0-to-1 boundary detection and outputs a phase code to all data bits. The UDC values in the data bits are decreased or increased uniformly so that the phase interpolator out-clock can be locked to the incoming clock. Because the phase comparison between the internal and incoming clocks is performed at 8-clock intervals, the maximum rate of the frequency tracking range is $25\ \text{ps}/(1.6\ \text{ns} \times 8) = 2 \times 10^{-3}$, which is much larger than the 100 ppm frequency variation of commercially available crystal oscillators.

3.6 Retiming and deskew

The phase of the receiver clock is different for each receiver unit because each clock skew is adjusted to the sample data at the center of the data eye. The retiming circuits align the received data skew to a single internal 625 MHz retiming clock which is generated by the clock recovery loop. This alignment is achieved by sampling the receiver output using serial-connected registers. Effectively, the retiming circuit adds a delay of $\text{dT} + nT$, where $0 < \text{dT} < T$, $n = 0$ or 1, and T is the bit time.
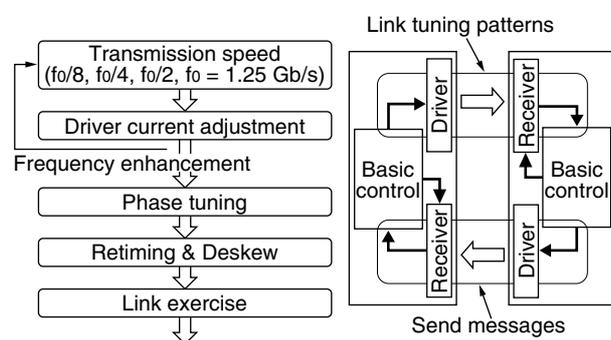
The retiming circuit aligns a bit-to-bit cable skew within a 1-bit time period to a single common clock, while a skew exceeding a 1-bit time period is aligned by the deskew circuit. The deskew circuit also uses a serial-connected D flip-flop clocked by a 312.5 MHz clock with a π phase separation. The output of the first stage is subject to 4-bit multiple-integer time delays because of the D flip-flop chains. The deskew circuit adds an additional delay of $2mT$, where $m = 0, 1, 2,$ or 3, and performs 1-to-2 demultiplexing. This results in data alignment to the internal clock with an adjustable delay of up to $8T$ (i.e., 6.4 ns at 1.25 Gb/s) as well as 1-to-4 demultiplexing.

4. Link initialization sequence

The link configuration and associated I/O interface parameter, link speed, driver output level, and receiver phase are defined by a logic sequencer called the "Basic control" (**Figure 12**). The I/O parameter tunings are performed in the power-on initialization sequence via OK/NG handshaking across the link between the I/O ports. Tuning patterns are sent to the main sequencer across the link through the cable connection. The Basic control defines appropriate receiver parameters and sends messages, for example, the driver output level, across the link according to the reception tuning patterns. When the tuning sequences are completed, the link exerciser runs a continuous random test pattern to confirm that the link is established.

This design performs a link at the fastest possible speed and the lowest possible power level to ensure reliable data transmission. It can

Figure 12
I/O link tuning sequence in power-on initialization.

88

FUJITSU Sci. Tech. J.,**36**, 1,(June 2000)

adapt to a wiring environment ranging from on-board PCB traces to 20 m twisted-pair cables without external adjustment.

5. Latency

Figure 13 shows an estimated I/O interface latency in the driver and receiver units. In the driver circuit, the data from the core logic is latched by the ϕ_0 of a 4-phase 312.5 MHz clock with a data-to-Q latency of within 800 ps. The data is 4-to-1 multiplexed and is output as 1.25 Gb/s data with the clock-to-Q latency, which was estimated to be 500 ps by SPICE simulations. The resultant maximum latency in the driver, ΔT_d, is estimated to be 1.3 ns.

In the receiver unit, incoming data is sampled by a 625 MHz receiver clock and is 1-to-4 demultiplexed through the interleaving receiver, retiming, and deskew circuits. The total of the data-to-clock latency and the clock-to-data delay

is estimated to be 5.04 ns, while the 1-to-4 unfolding latency is 3 times the bit time period. The total latency of the I/O interface, excluding the cable delay, is estimated to be 7.44 ns.

6. Chip design

We designed an interconnect test chip using a 0.25 µm CMOS technology (**Figure 14**). The chip consists of 2-port I/O interfaces having 21-bit driver and receiver arrays, the Basic control, the PLL, and SRAM. The 21-bit driver and receiver arrays are $3300 \times 1940\ \mu m^2$ and $3300 \times 900\ \mu m^2$, respectively. Each driver and receiver unit consumes 0.11 W and 0.07 W, respectively, at a supply voltage of 2.5 V. Therefore, the total power consumption of the 21-bit driver and receiver arrays is estimated to be 3.78 W (**Table 1**). The PLL and the Basic control consume 0.09 W and 0.70 W, respectively, and are shared by the ports of the I/O interface in the router chip design.

Figure 15 shows waveforms measured dur-
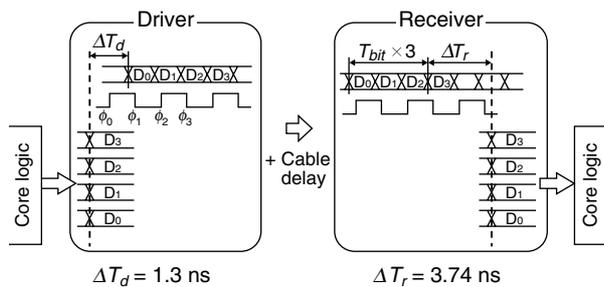
Figure 13
Estimated latency of driver and receiver units of I/O interface.

Figure 14
Parallel interconnect and self-configured link test chip.

Table 1
Estimated power consumption of I/O link.

Circuit	Power consumption (W)
Driver unit	0.11 ⎫
Receiver unit	0.07 ⎭ × 21 = 3.78 (21-bit array)
PLL	0.09
Basic control	0.70
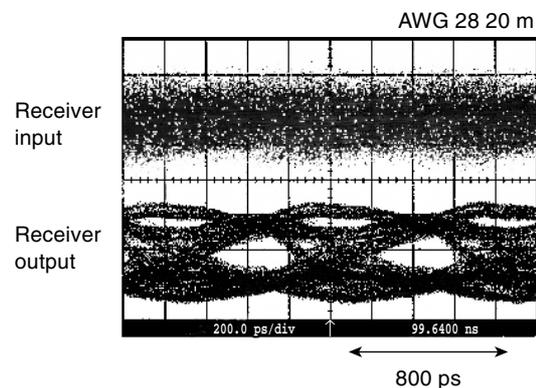
Figure 15
Measured waveforms of 1.25 Gb/s data transmission over a 20 m AWG 28 twisted-pair cable.

ing signal transmission testing. The upper wave-form shows the receiver input over a 20 m AWG 28 twisted-pair cable, while the lower waveform shows the DPRD receiver output. The figure shows that the chip provides a clear eye opening by ISI elimination for frequency-dependent cable loss. Using our I/O interface design, we achieved a reliable 1.25 Gb/s signal transmission over a 20 m cable.

7. Conclusion

We have developed a 2-btye parallel-intercon-nect I/O interface for scalable multiprocessors that provides a 1.25 Gb/s bandwidth in one signal line and a 7.4 ns latency. The DPRD receiver en-sures a low-latency equalization scheme that compensates for the frequency-dependent cable loss of cables up to 20 m in length. The phase-interpolator-based clocking scheme has a 25 ps ± 5 ps skew adjustment step and performs plesio-chronous clocking. The I/O link tuning is performed by a logic sequencer, which maximize the data rate and minimizes the power consump-tion without external manual adjustments. We designed a test chip for parallel-link interconnec-tion using a 0.25 μm CMOS process and confirmed that it was capable of 1.25 Gb/s signal transmis-sion over a 20 m AWG 28 twisted-pair cable.

References

1) R. Rettberg, W. Dally, and D. Culler: *IEEE Micro*, **18**, 1, pp.10-11 (Jan.-Feb. 1999).
2) W. Weber et al.: The Mercury Interconnect Architecture: A Cost-effective Infrastructure for High-performance Server. Proc. of the 24th International Symposium on Comput-er Architecture, 1997.
3) Charleswoth: Extending the SMP Envelope. A, *IEEE Micro*, **18**, 1, pp.39-49 (Jan.-Feb. 1999).
4) W. Dally and J. Poulton: A Tracking Clock Recovery for 4-Gbps Signaling. A, *IEEE Micro*, **18**, 1, pp.25-27 (Jan.-Feb. 1999).
5) R. Gu, J. Tran, H. Lin, A. Yee, and M. Izzard:
A 0.5-3.5Gb/s Low Power Low Jitter Serial Data CMOS Tranceiver. ISSCC Digest of Technical Papers, February 1999, pp.352-353.
6) H. Tamura et al.: Partial Response Detection Technique for Driver Power Reduction in High-Speed Memory-to-Processor Communi-cations. ISSCC Digest of Technical Papers, February 1997, pp.342-343.
7) T. Lee et al.: A 2.5 V CMOS delay-locked loop for an 18 Mbit, 500 MB/s DRAM. *IEEE J. Solid-State Circuits*, **29**, pp.1491-1496 (Dec. 1994).

Kohtaroh Gotoh received the B. S. and M. S. degrees in Electrical Engineering from Waseda University, Tokyo, Japan, in 1986 and 1988, respectively. In 1988, he joined Fujitsu Laboratories Ltd., Kawasaki, Japan, where he was en-gaged in research of Josephson devic-es and circuit design. Since 1995, he has been working on CMOS circuit design. His current research interests include high-speed I/O interface design and chip-to-chip communication.

Hideki Takauchi received the B. S. and M. S. degrees in Electrical Engineering from Waseda University, Tokyo, Japan, in 1988 and 1990, respectively. In 1990, he joined Fujitsu Laboratories Ltd., Kawasaki, Japan, where he was en-gaged in research of superconducting devices. Since 1996, he has been working on research and development of CMOS circuit design. His current re-search interests include high-speed in-terconnection circuits.

Hirotaka Tamura received the B. S., M. S., and Ph. D. degrees in Electrical Engineering from the University of Tokyo, Tokyo, Japan, in 1977, 1979, and 1982, respectively. In 1982, he joined Fujitsu Laboratories Ltd., Kawasaki, Japan, where he was engaged in re-search of Josephson devices and experimental superconducting devices. Since 1995, he has been working on research and development of CMOS circuit design. His current research interests include high-speed interconnection circuits.

90

FUJITSU Sci. Tech. J.,**36**, 1,(June 2000)