A 0.297-pJ/Bit 50.4-Gb/s/Wire Inverter-Based Short-Reach Simultaneous Bi-Directional Transceiver for Die-to-Die Interface in 5-nm CMOS

Yoshinori Nishi, *Member, IEEE*, John W. Poulton, *Fellow, IEEE*, Walker J. Turner, *Member, IEEE*, Xi Chen, *Member, IEEE*, Sanquan Song, *Member, IEEE*, Brian Zimmer, *Member, IEEE*, Stephen G. Tell, *Member, IEEE*, Nikola Nedovic, *Member, IEEE*, John M. Wilson, *Senior Member, IEEE*, William J. Dally, *Fellow, IEEE*, and C. Thomas Gray, *Senior Member, IEEE*

***Abstract*— This article presents a clock-forwarded, inverter-based short-reach simultaneous bi-directional (ISR-SBD) physical layer (PHY) targeted for die-to-die communication over silicon interposers or similar high-density interconnect. Short-reach links of this type are increasingly important to support larger systems built with chiplets and multiple die and to facilitate the shift to medium- and long-range optical communication based on silicon photonics. This project explores the advantages of simultaneous bi-directional signaling (SBD) over other bandwidth-doubling techniques (e.g., PAM4). Fabricated in a 5-nm standard CMOS process, the ISR-SBD PHY demonstrates 50.4 Gb/s/wire (25.2 Gb/s each direction) and 0.297 pJ/bit on a 750-mV supply over a 1.2-mm on-chip channel.**

***Index Terms*— Chip-to-chip, chiplet and interposer, chip on wafer on substrate (CoWoS), die-to-die, integrated fan-out (InFO), ISR, simultaneous bi-directional signaling (SBD), short-reach, simultaneous bi-directional, ultra-short-reach (USR), very-short-reach (VSR).**

I. Introduction

THE doubling of transistor density every 18 months (Moore's law) for the past 50 years or so has led to the rise of high-performance computing (HPC) that supports applications such as artificial intelligence and deep learning, cloud computing, real-time photo-realistic computer graphics for gaming, film production, and scientific visualization. But the scaling of CMOS technology that has powered this revolution has dramatically slowed. To stay on the development curve we have grown accustomed to, manufacturers have initiated a renaissance in packaging. Today's advanced systems feature multiple chips and chiplets on interposers, using technologies such as chip on wafer on substrate (CoWoS) and integrated fan-out (InFO) to interconnect them. These "2.5D" structures will soon be augmented by true 3-D interconnect.

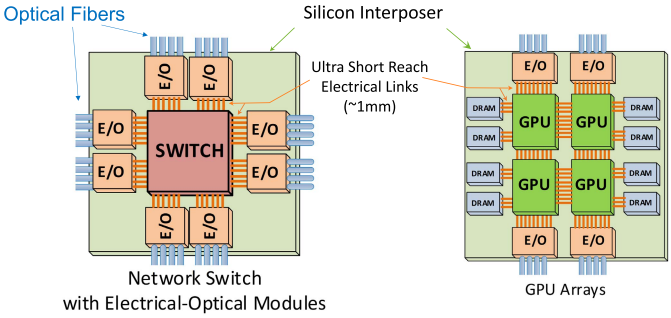


Fig. 1. Chiplet applications requiring high-bandwidth ultra-short-reach interconnects.

A more troubling aspect of slower scaling is that transistor speed stopped improving at about the 10-nm node. That, together with interconnect physics, has brought electrical signaling to the end of the line. The future of short- and medium-distance interconnect is optical, based on silicon photonics. For the next era in computing, ultra-short-reach (USR) electrical links will be a critical technology to facilitate increases in computational performance of HPC modules and to support the optical signaling needed to interconnect them. A GPU server, which is typical of today's HPC systems, consists of many GPU pods where multiple GPUs, each implemented on an interposer along with DRAM memory, are interconnected on a PCB together with network switches that drive optical cables to interconnect multiple pods. In the near future, as shown in Fig. 1, these GPU clusters, as well as network switches with their electro-optical interfaces, will be integrated onto silicon interposers, interconnected with optical fibers. Memory interconnect, interchip communication, and optical interconnect will all require area- and power-efficient USR electrical links.

Network switch bandwidth doubles approximately every two years [1], expecting to reach 800 Tb/s in each direction by 2030.

To achieve this bandwidth on a reticle-limited die (25 × 32 mm), USR links require edge bandwidth densities

Manuscript received 29 August 2022; revised 28 October 2022 and 16 December 2022; accepted 17 December 2022. Date of publication 9 January 2023; date of current version 28 March 2023. This article was approved by Associate Editor Borivoje Nikoli. *(Corresponding author: Yoshinori Nishi.)*

Yoshinori Nishi, Xi Chen, Sanquan Song, Brian Zimmer, Nikola Nedovic, and William J. Dally are with NVIDIA Inc., Santa Clara, CA 95051 USA (e-mail: ynishi@nvidia.com).

John W. Poulton, Walker J. Turner, Stephen G. Tell, John M. Wilson, and C. Thomas Gray are with NVIDIA Inc., Durham, NC 27713 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2022.3232024.

Digital Object Identifier 10.1109/JSSC.2022.3232024

above 15 Tb/s/mm, translating to about 50 Gb/s/wire at a 40-μm micro-bump pitch. USR links cannot take advantage of the many benefits of differential signaling, as pin density limitations require single-ended signaling. To stay within total device power (TDP) limits, signaling energy will need to be driven down to a few 100 fJ/bit.

How will we get to this speed, area, and power efficiency, given that the transistor speed is no longer improving? The optimal signaling speed for energy efficiency is, and will remain, in the range of 20–30 Gbaud [2]. The 20-Gb/s/wire NRZ-based signaling at 460 fJ/bit has been reported [3], but any effort to push this speed up into the 40–50 Gb/s range will incur severe power penalties because of high clock distribution power and the need for advanced equalization. PAM4 modulation would allow doubling the bit rate at the same baud rate, but it suffers from SNR degradation, decreased timing margins, and energy overhead of multiple samplers and linear signal amplifiers. In Section II, we propose simultaneous bi-directional signaling (SBD) as a method for overcoming these difficulties.

II. Signaling Scheme

A. Modulation Scheme Comparison

In this section, we compare three modulation schemes to find the best solution for 50 Gb/s/wire on interposer channels.

NRZ is the simplest and best-established solution. It is straightforward to use delay-matched clock-forwarding to achieve excellent delay tracking between data and clock and ensure robust supply noise cancellation. However, in addition to the large power penalty to drive and detect NRZ signals at 50 Gb/s and to generate and distribute 25-GHz (or equivalent 12.5-GHz multi-phase) clocks, the timing margin for a 20-ps UI is extremely small in the face of crosstalk and reflections in a 1–2-mm channel.

PAM4 is often used as a bandwidth-doubler over NRZ, but it presents many difficulties. First, the available signal amplitude is divided by 3, reducing the SNR by 9.5 dB. Second, PAM4 introduces an inherent form of inter-symbol interference (ISI) resulting from the differing signal trajectories between the four signal levels that reduces the timing margin significantly. Furthermore, to achieve good level separation mismatch ratio (RLM) for PAM4, drive impedance must be very linear across all the four signal levels at a cost in area and power, an issue that is avoided in NRZ signaling. If the received PAM4 signal is to be amplified before the samplers, this amplifier likewise must maintain good linearity. Finally, in a clock-forwarding scheme, it is difficult to maintain data clock delay tracking since the two-level clock path, which is amplified to a rail-to-rail signal, must match the delay of the four-level data path, which is analog.

SBD doubles the per-wire bandwidth for a given baud rate. SBD was explored in the 1990s [4], [5], [6], [7] and has been revisited since for high-speed wireline applications [8], [9], [10], [11]. In this method, NRZ signals are driven in opposite directions from both the ends of a channel. The signal on the line is the sum of two independent NRZ data streams and superficially looks like a three-level PAM3 signal (with

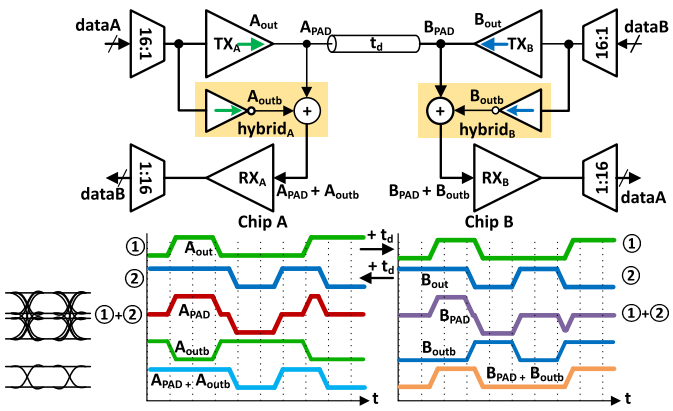


Fig. 2. Simultaneous bi-directional signaling.

channel IR drop, there are four levels). Receivers at each end are equipped with a duplexer or "hybrid" (to borrow the language of SBD telephony) that subtracts the outgoing signal from the signal on the line to recover the incoming received signal. Because SBD involves linearly added NRZ signals, driver and receiver linearity are less critical than in PAM4, and in a clock-forwarding design, the excellent clock/data tracking of NRZ can be retained. The method's downsides include sensitivity to reflections and near-end crosstalk, problems that can be largely addressed by careful channel design. Fortunately, the channels presented by USR interposer interconnects are fairly benign in this regard, and these effects can be well-controlled at the target baud rates below 30 Gbaud. Signal subtraction in the receiver's hybrid is never perfect, and these imperfections introduce leakage, in which the outgoing signal interferes with the received signal, a form of crosstalk unique to SBD. Leakage is unavoidable, but with careful design, the jitter it introduces is manageable and generally much smaller than the inherent ISI of PAM4 signaling. Based on these advantages, we have chosen SBD as a promising technique for developing a 50-Gb/s single-ended all-CMOS short-reach signaling solution.

B. Simultaneous Bi-Directional Signaling

Fig. 2 shows how SBD signaling works. Chip A and Chip B are connected through a wire, and both are the transmitting signals independently, shown as A_{out} and B_{out}. Signals pass through the channel and reach the other sides after channel delay t_d. Signals ① and ② represent how A_{out} and B_{out}, respectively, look at each pad. They are summed (① + ②) to become tri-level signals at each pad (A_{PAD} and B_{PAD}). The receiver hybrid, a signal separator, is shown in a highlighted box. It subtracts the outbound signal (A_{out} or B_{out}) from the signals at each pad (A_{PAD} or B_{PAD}) to recover the data from the far end of the line (dataB or dataA). The hybrid works simply by adding an inverted copy of the outbound signal (A_{outb} or B_{outb}) to the line signal (A_{PAD} or B_{PAD}) to obtain a clean NRZ signal into the RX data path.

C. SBD Hybrid

Fig. 3 reviews previous methods for implementing an SBD hybrid. The most common technique is to use a replica

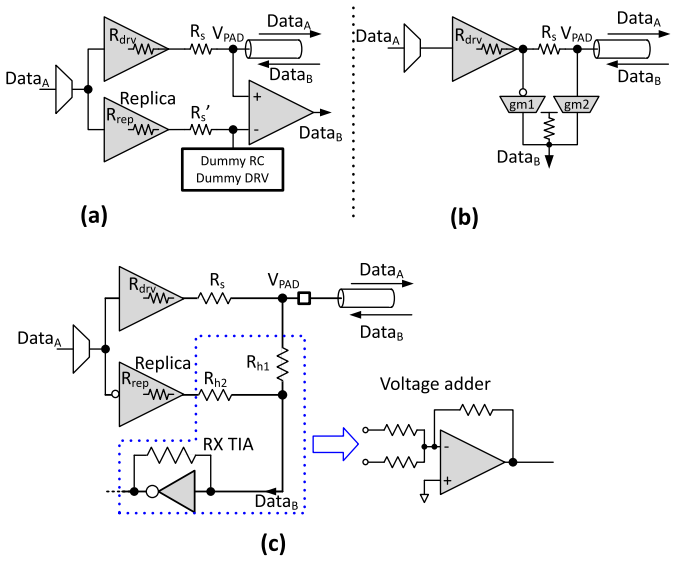


Fig. 3. Hybrid topologies. (a) Replica-based hybrid. (b) R-gm hybrid. (c) Proposed hybrid.

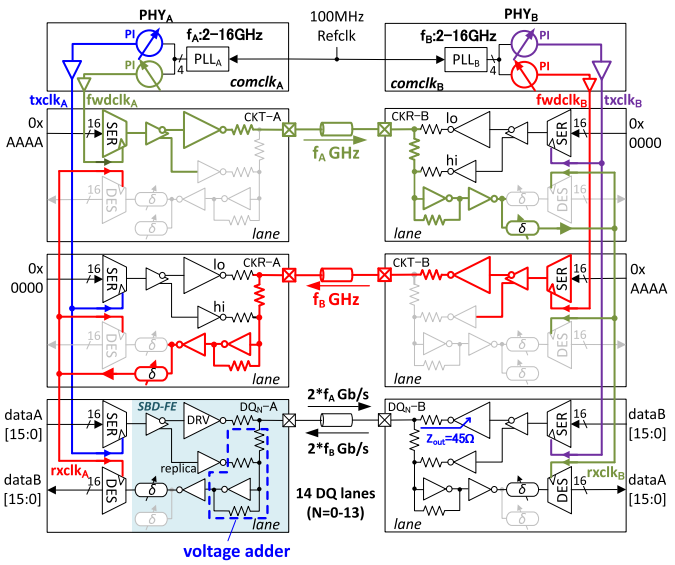


Fig. 4. ISR-SBD PHY top level.

driver and a difference amplifier as in Fig. 3(a) to subtract the outbound signal. This requires a high-speed differential amplifier that tends to be large and high power. Also, since the replica path is internal, it requires not only a dummy *RC* to match the effects of the line shunt capacitances and losses but also active circuits to mimic the driver at the far end of the line that acts as a termination to a common-mode voltage. Another scheme is the R-Gm hybrid [Fig. 3(b)], which does not require a replica [10]. This scheme is advantageous because both the inputs to the Gm cells are exposed to the line's imperfections, thereby requiring neither dummy *RC* network nor far-end driver emulation. However, one of the Gm cells must invert the data, while the other Gm is non-inverting. Data inversion is trivial for differential I/Os, but difficult for single-ended CMOS, because the asymmetry between non-inverting and inverting Gms can introduce nonlinearity and timing skews.

To overcome these challenges, we propose a single-ended SBD hybrid that uses only inverters and linear resistors, as shown in Fig. 3(c). The hybrid consists of resistors R_{h1} and R_{h2} that form a voltage adder along with the receive amplifier, to cancel the outbound signal. The receiver's first stage is a transimpedance amplifier (TIA), comprising a CMOS inverter with resistor feedback. We will cover the hybrid design in greater detail in Section IV.

III. ISR-SBD Link Design

Fig. 4 shows the top level of our inverter-based short-reach SBD (ISR-SBD) link, which implements a delay-matched clock-forwarding architecture. Each physical layer (PHY) has one PLL and two phase interpolators (PIs), one clock transmitter lane (CKT), one clock receiver lane (CKR), and 14 data lanes (DQ_N). All the 16 lanes are identical.

A. Clocking

To avoid bi-directional jitter from hybrid leakage, clock lanes are configured to work in uni-directional mode, with the opposite direction constantly sending all 1 s or all 0 s. (It is possible to forward clocks bi-directionally on a single wire, though the additional jitter from leakage would need to be allowed for in the timing budget.) Each PLL generates four equally spaced clock phases that go through two identical PIs, one for DQ lanes (txclk) and the other for the CKT lane (fwdclk). Dummy capacitive loads are included within the clock buffers to equalize the delay and supply voltage sensitivities of the two paths since the DQ clock txclk has a much larger load than the CKT clock fwdclk. The clocks are routed to each serializer (SER) on the thick upper metal layers to limit lane-to-lane skew to 1 ps or less. SER in the CKT lanes generates clock patterns, while SER in the DQ lanes sends out serialized NRZ data. Clock and data go through identical line drivers, channels, and receiver front-ends to reach the deserializers (DESs) at the other end, where the use of matched circuitry ensures equal delays in both the data and clock end-to-end paths, as we detailed in [12]. Details of delay matching are also discussed in Section III-C. It is worth noting that two PHYs can operate asynchronously ($f_A \neq f_B$), mesochronously ($f_A = f_B$ with unknown phase), or synchronously ($f_A = f_B$ with known phase relationship).

B. TXRX Lanes

Fig. 5 shows the entire TXRX lane design. Both the TX and RX use a 16-bit interface to the core logic. SER receives the global half-rate clock (txclk) into clock buffers in each TX, which then drive dividers and a 2:1 MUX in each SER. The duty cycle of the clock at the 2:1 MUX is corrected by digital control bits. Serialized data are converted into complementary CMOS form in a pre-driver (PREDRV) stage, whose delays can be adjusted to assure the symmetry between the two outputs to the line and replica drivers. The main line driver (DRV) drives the outbound signal, while the inbound signal also comes in on the same line. As we discussed in the previous section, the resistor network of the hybrid extracts

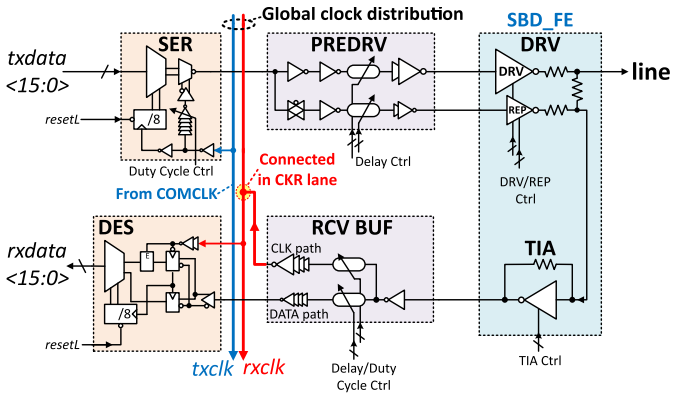


Fig. 5. TXRX lane diagram.

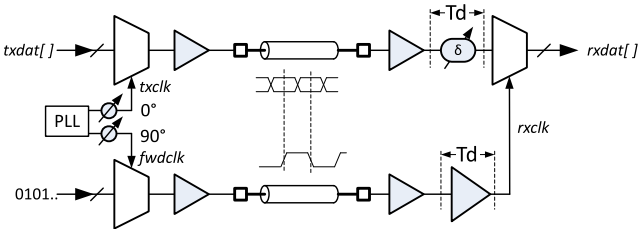


Fig. 6. Data clock delay tracking.

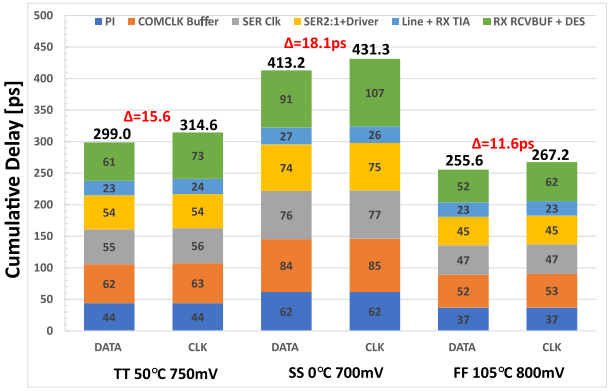


Fig. 7. Data clock delay tracking across PVT.

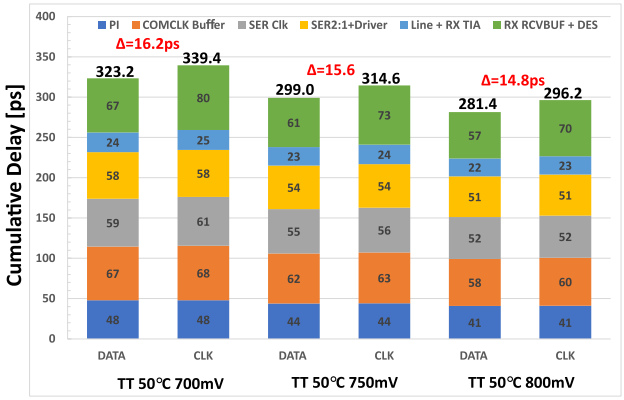


Fig. 8. Data clock delay power supply sensitivity.

the received NRZ data for the RX. In a DQ lane, the received signal goes through the DATA path of RCVBUF to DES. In the forwarded-CKR lane, the CLK path of RCVBUF is used to drive the global rxclk that drives the DES's clock input buffers in all the lanes. Those clock buffers then drive samplers and dividers in each DES.

C. Delay-Matched Clock-Forwarded Architecture

Typical clock forwarded links rely on additional circuitry (PLL, DLL, and so on) to dial-in the clock-to-data phase relationship at the receiver, making them sensitive to supply noise in the low-to-middle frequency range. Our clock forwarding architecture does not rely on such circuitry. Instead, we implement a delay-matched clock-forwarded architecture where the total insertion delays and delay sensitivities of the forwarded clock and data signals match throughout the entire signal paths, from the transmitter to the receiver, to effectively cancel out supply-induced jitter. This is achieved using near-identical circuits within both the signal paths, where we match the number of stages and the rise/fall times within each corresponding stage to ensure delay sensitivities to voltage variations are similar for both the clock and data.

Referring to Fig. 6, and for the moment considering each of the two directions of data flow as completely independent, first note that the data clock txclk and the forwarded clock parent clock fwdclk pass through identical phase interpolators. Logically one interpolator would suffice for eye centering and margining, but the dual-PI arrangement ensures delay and supply sensitivity matching. The forwarded clock is generated in an SER and driver that are identical to those used in the data lanes, again assuring delay and supply sensitivity matching. Data and clock then pass through identical channels whose delays are as closely matched as package/interposer constraints allow and are received on identical receiver structures. At the receive end of the link, the received forwarded clock passes directly from the RX amplifier into a clock buffer (CLK path in RCVBUF) that drives the sampler clock rxclk to all the data receivers. To maintain delay and delay sensitivity matching, the data also pass through a programmable delay element (δ), adjusted during calibration so that its delay Td matches that of the rxclk buffer. With every element of the data and clock paths matched for both the delay and delay variation with PVT, the relative timing tracks even if absolute delays vary significantly. This style of clock-forwarding cancels the majority of the supply-induced jitter and the common random jitter of the clock source, leaving only random jitter of the clock buffers, clock duty cycle, and ISI on data lanes to deal with in the timing budget. Note that the PLL need not be a high-performance, low-jitter element; indeed, it need not even operate at a fixed frequency but could be allowed to vary with PVT conditions as long as the receiver's clock circuitry can be guaranteed to operate at the highest frequency. To see how well this works in practice, we show the end-to-end delays for data and clock at nominal and extreme cases from simulation in Fig. 7.

For nominal conditions (TT/50 °C/750 mV), the total delay for the data is 299 ps. Clock delays are very similar for each

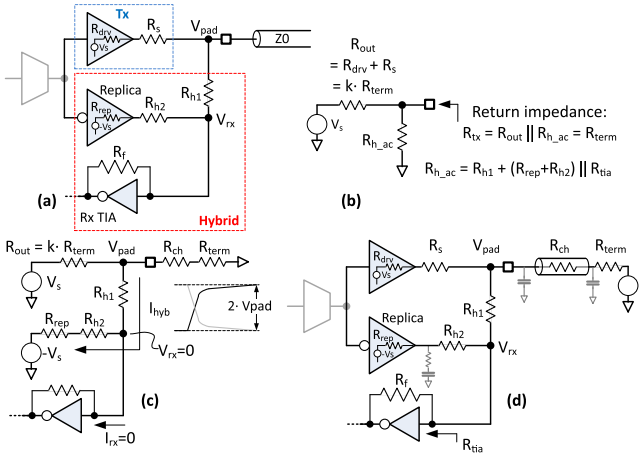


Fig. 9. Modeling of SBD hybrid. (a) Main driver and hybrid diagram. (b) Modeling for driver impedance calculation. (c) Modeling for $V_{\rm pad}$ calculation. (d) Modeling for $V_{\rm rx}$ calculation.

component except for RX RCVBUF and DES where the CLK delay is 12.5 ps larger compared with DATA. As shown in Fig. 5, the data path and the clock path split in RCVBUF because the CLK path inside CKR lane must drive the global rxclk, which has capacitance as high as 250 fF while DQ lanes use the DATA path to drive the DES only, whose input cap is less than 10 fF. The number of stages is matched and dummy loads are added to the data path to minimize the difference as much as possible. The remaining relative delay difference is limited to 11.6–18.1 ps across PVT although the absolute delays change by more than 150 ps. This implies that even if we use the same PI codes across PVT, the timing error would be limited to 6.5 ps. The majority of this variation comes from the process itself, and this can be removed by one-time calibration of the phase interpolator code. The impact of the supply voltage is shown in Fig. 8, where 750 ± 50 mV shows only 1.4-ps relative delay variation while absolute delays change by more than 40 ps, thus demonstrating the high supply noise rejection of the clock-forwarding architecture. The temperature effect is even more limited, less than 1 ps from 0 °C to 105 °C.

IV. SBD Hybrid Circuit Design

In Section II-C, we briefly reviewed a single-ended SBD hybrid that uses only linear resistors and CMOS inverters as shown in Fig. 9(a). With the inverter threshold defined as "ground" (mid-swing), the transmitter's source voltage $V_s(= \pm{\rm VDD}/2)$ drives current into the line through series resistors $R_{\rm drv}+R_s$, and a portion of this current flows into the hybrid. The replica transmitter, whose input is inverted with respect to the transmitter, drives $-V_s$ into the other end of the resistive hybrid, such that the transmitter's output current is canceled at the receiver input ($V_{\rm rx}$), leaving only the current received from the far-end transceiver as input to the receiver.

The impedance of the hybrid is arbitrary. The values of R_{h2} and $R_{\rm rep}$ are determined by whatever impedance is used for R_{h1}, so as to cancel the TX signal at $V_{\rm rx}$. However, to calculate R_{h2} and $R_{\rm rep}$ from R_{h1}, we need to know the voltage at the pad $V_{\rm pad}$, and that in turn depends on the termination impedance looking back into the transmitter [Fig. 9(b)] $R_{\rm out}||R_{h\_ac}$, where $R_{h\_ac} = R_{h1} + (R_{h2} + R_{\rm rep})||R_{\rm tia}$. To avoid a complex calculation for $V_{\rm pad}$, we use an iterative method to find the value of $R_{\rm out}$ that provides the correct termination impedance. As shown in Fig. 9(b), we estimate the ac impedance of the hybrid R_{h_ac}, and then estimate the impedance of the driver $k\cdot R_{\rm term}$ ($k > 1$) that provides the correct termination. As shown in Fig. 9(c), we calculate $V_{\rm pad}$, and thus we can write $I_{\rm hyb}$ as

$$I_{\rm hyb} = \frac{V_{\rm pad}}{R_{h1}} = V_s \frac{R_{\rm ch}+R_{\rm term}}{R_{h1}[R_{\rm ch}+(1+k)R_{\rm term}]+kR_{\rm term}(R_{\rm ch}+R_{\rm term})}. \quad (1)$$

With the far-end driver assumed to be at "ground" (mid-swing), we require the voltage $V_{\rm rx}=0$, so no current flows into the RX TIA, and therefore, $I_{\rm hyb}$ also flows through $R_{\rm rep}+R_{h2}$ into $-V_s$. Using these observations, we write

$$R_{\rm rep}+R_{h2} = R_{h1}\left[1+{\rm kR_{term}}\left(\frac{1}{R_{\rm ch}+R_{\rm term}}+\frac{1}{R_{h1}}\right)\right]. \quad (2)$$

With all the hybrid resistors calculated, we can find R_{h_ac} and $R_{\rm tx}$. If $R_{\rm tx}$ is higher (lower) than the target $R_{\rm term}$, then we can reduce (increase) k and re-run the calculation. These steps are rapidly implemented in a spreadsheet.

Finally, referring to Fig. 9(d), we calculate the signal voltage that arrives at the input of the receiver TIA from the far end of the link. Typical interposer channels will have significant losses, even at dc, represented by the channel resistance $R_{\rm ch}$. If both the ends are terminated at $R_{\rm term}$, the received dc signal at $V_{\rm pad}$ is attenuated by $R_{\rm term}/(2R_{\rm term}+R_{\rm ch})$. The signal is further attenuated as it passes through the hybrid to the receiver input.

We are free to choose R_{h1} (or R_{h2}) somewhat arbitrarily; larger R_{h1} allows a larger outbound signal into the line, but it cannot be too large, since the inbound signal into the TIA is attenuated by R_{h1} and $R_{\rm tia}||(R_{\rm rep}+R_{h2})$. Once R_{h1} is chosen, we can calculate $R_{\rm rep}+R_{h2}$ from (2), as well as calculate the required driver impedance $R_{\rm out}=R_s+R_{\rm drv}=k\cdot R_{\rm term}$.

Here, however, we run into technology-related problems. First, the three resistors R_s, R_{h1}, and R_{h2} are optimally implemented as fixed linear resistors (TiN metal resistors in the target technology). However, it is impractical to make these resistors adjustable. Adjustable resistors would require series CMOS passgates that not only present unacceptable non-linearity near the inverter threshold but also are severely constrained by ESD protection rules. Layout constraints also favor certain discrete values, particularly for R_{h1} and R_s, since they are part of the ESD protection circuitry. Therefore, the resistors are set at design time, and driver and replica impedances $R_{\rm drv}$ and $R_{\rm rep}$ are the only tuning parameters. The fixed resistors have relatively small VT variation, and a ± 15% variation for process. $R_{\rm drv}$ and $R_{\rm rep}$, however, vary not only with the process but also have fairly strong VT variation. We can deal with the process and voltage variation through calibration, and since the termination impedance, particularly for a transmission line with resistive losses, is not critical, we can live with the temperature variation. But we will have to tune $R_{\rm drv}$ and $R_{\rm rep}$ separately: consider the case

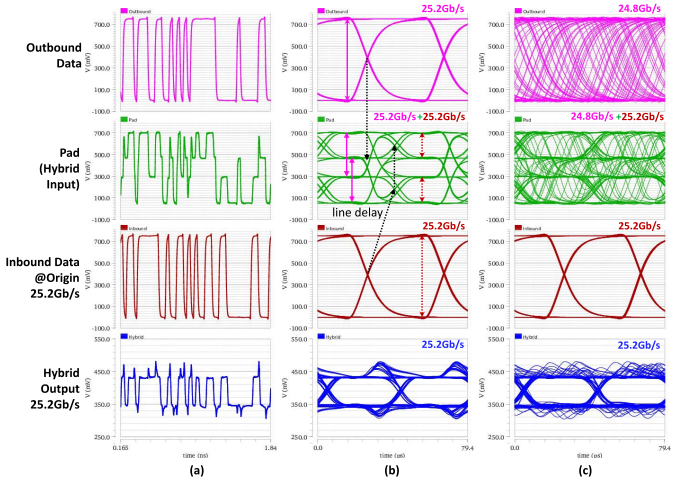


Fig. 10. Simulation waveforms for hybrid operation. (a) Voltage versus time. (b) Synchronous SBD mode. (c) Asynchronous SBD mode.

where the TiN sheet resistance is low, then R_{drv} will need to be adjusted upward to meet the termination condition on R_{out}. The replica resistance R_{rep} must be adjusted downward, however, to compensate for the lower values of R_{h1} and R_{h2}. The method for calibrating the link, including setting R_{drv} and R_{rep}, is described in Section V below.

By way of example, in the test chip of this article, we chose $R_{h1} = 190\,\Omega$ and $R_{h2} = 250\,\Omega$, leaving R_{rep} as a tuning parameter. The TIA's feedback resistor R_f is $1200\,\Omega$, and the open-loop voltage gain is about 5, so the input impedance Rtia $\sim 200\,\Omega$. To achieve reasonable linearity in the main driver, we made $R_{\text{drv}} = R_s = 22.5\,\Omega$ at the typical corner. The channel characteristic impedance target is about $40\,\Omega$, the channel resistance $R_{\text{ch}} = 21.5\,\Omega$, and the driver impedance $R_{\text{out}} = 45\,\Omega = 1.125 \cdot R_{\text{term}}$. From 2, we calculate $R_{\text{rep}} + R_{h2} = 374\,\Omega$, leaving $R_{\text{rep}} = 124\,\Omega$ and $R_{\text{tx}} = 39.5\,\Omega$, very close to the target.

As an example, the simulation waveforms are shown in Fig. 10 for simultaneous bidirectional operation in synchronous (25.2 Gb/s in both the directions) and asynchronous (25.2 and 24.8 Gb/s) modes. The output signal swing of the far-end transmitter into the line is about 0.51· VDD. When it reaches the receiver, it is further attenuated by the channel resistance to about 0.33· VDD. V_{rx} sees $R_{\text{tia}} \sim 200\,\Omega$, so that the inverter threshold is closely tracked across PVT. With this input impedance, the voltage swing into the TIA is about 100 mV (±50 mV signal). From noise simulations, we estimate the input referred noise at the TIA of 0.25 mV_{rms}, so for an estimated BER of 1E-25, we only require about 3 mV of voltage margin, readily achievable with a 50-mV signal.

V. Calibration

In Section III-C, we discussed data clock tracking that minimizes PVT variations in the receive sampler clock timing. However, link performance is also affected by factors such as clock nonidealities caused by random device mismatches and SBD leakage from resistor variations as described in Section IV. While a link may run error-free with carefully chosen default settings, performance optimization through calibration is crucial to improve the robustness of the link. Here, we explain link calibration consisting of six major parts.

A. Clock Duty Cycle

For high-speed clock paths in deep submicrometer CMOS, it is necessary to have clock duty cycle correction (DCC) circuits at multiple locations to compensate for high random device mismatches. Furthermore, because we are operating clock buffer chains near their bandwidth limits, these chains can introduce and amplify duty-factor offset with strong temperature dependence. Therefore, DCC is a distributed problem, so we have implemented DCC circuits at multiple points: at each phase interpolator output, at the final 2:1 MUX in each TX SER, and at the clock driver inside the CKR lane. During calibration, each DQ lane is configured to send clock patterns so that the DCC at the output of the txclk phase interpolator can be calibrated based on the average of all the DQ lanes.

At the receive end of the link, we use a free-running ring oscillator (3–6 GHz) implemented inside common clock (COMCLK) to sample the incoming pattern asynchronously at each DES, and then compare the number of sampled 1 and 0 s to determine duty cycle. Similar schemes have been discussed in [13], [14], and [15]. While the txclk duty cycle is calibrated by means of centering the average as described above, TX SER calibration is done per lane by minimizing the duty cycle difference between 0 × AAAA pattern and 0 × 5555 pattern, which depends on the duty cycle of the clock driving the final 2:1 MUX inside SER. The fwdclk phase interpolator and the associated TX SER are calibrated using the same scheme, but there is a difference as the forwarded clock goes into the CKR where the CLK path is used instead of the data path to the sampler as shown in Fig. 5. This is unique to the CKR lane, and thus, a different calibration mechanism is required. The duty cycle of rxclk is therefore optimized by maximizing the aggregated horizontal eye opening of the entire receiver lanes during the phase interpolator sweep described next.

B. Clock Timing

As described in Section III-C, the data and clock timing closely track across voltage and temperature. We only need to adjust the clock timing once at link initialization by sweeping the phase interpolator code to find the horizontal eye opening of each receiver lane, and then we set the PI Code to be at the center of the horizontal opening. Note that during this step, rxclk duty cycle is optimized as well.

C. Lane to Lane Skew

Deterministic lane-to-lane skews for die-to-die USR links are limited because the data channels are only a few mm in length and physically matched. In reality, data skews are dominated by random mismatches in the active circuitry. Skew compensations are done by a delay trimmer implemented in each lane ahead of DES. These trimmers have a tuning range of 5 ps.

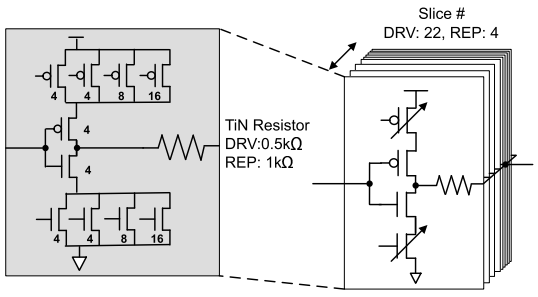


Fig. 11. Driver and replica driver schematic.

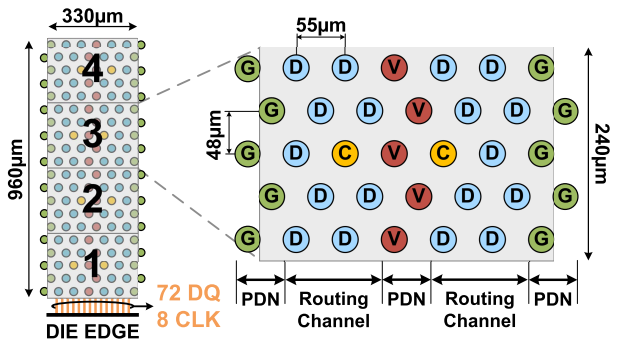


Fig. 12. Targeted four-rank interposer bump array.

D. Driver Impedance

As mentioned earlier, it is necessary to calibrate the driver impedance to closely match the line impedance and to track the TiN process variation in R_s, R_{h1}, and R_{h2}. As shown in Fig. 11, we have implemented switchable header and footer devices within the line driver so that $R_{drv} + R_s \simeq 45\,\Omega$ can be satisfied across PVT. Although not a part of the ISR-SBD PHY, a process monitor circuit connected to an external resistor can be used to determine TiN resistor deviation from the nominal value. Such a monitor can be shared across the entire die since TiN variation within the same die is much less than 1% in general.

E. SBD Hybrid

R_{rep} is the only adjustable parameter for the hybrid. It is adjusted to maintain $R_{\text{rep}} + R_{h2} \propto R_{h1}$ to compensate for passive resistor variation. As we have shown, this compensation requires R_{rep} and R_{drv} to be adjusted in opposite directions, and thus, the tuning codes cannot be shared between the main driver and the replica. We can adjust R_{rep} in the CKR lane by minimizing the difference in duty cycle of the incoming clock between two different outbound patterns, all 1s and all 0s. Since driver transistors are large and suffer negligible random mismatch, R_{rep} calibration results from CKR lane can be applied to all DQ lanes.

F. Receiver Offset

Per-lane receiver offset calibration can be performed by an offset trimmer comprising adjustable header and footer devices in the TIA inverter, which has a similar topology as the line driver in Fig. 11. However, the input-referred offset resulting from local variation in the TIA is negligible, thanks to large device sizes (24x of the minimum inverter) and negative feedback. For this reason, a common setting is applied for all the lanes. The impact of process skew (i.e., when the two communicating chips are at different process corners) is limited for this bi-directional link because the common-mode voltage that appears on the line is the average of the two sides. The most extreme scenario is N-Slow P-Fast (SF) on one end and N-Fast P-Slow (FS) on the other end, each side having common-mode voltages skewed in opposite directions. However, such a case results in the common-mode voltage on the line being near VDD/2. This sufficiently reduces the impact of skewed corners where the remaining receiver offset can be handled by the same offset trimmer.

All the calibrations listed above are done during the initialization of the link through an external python loop, which would be replaced by controller hardware in production chips. While there is very limited performance degradation from the temperature drift of the clock (calibrated through a, b, and c) owing to our design approach, the driver impedance and hybrid calibrations need more serious attention. R_{drv} could present a temperature drift of up to 20%, but the effect is reduced to half owing to 1:1 ratio between R_{drv} and R_s, where R_s consists of passive resistors with very small temperature dependence. The resulting 10% impedance mismatch is acceptable given the resistive loss of the channel. On the other hand, R_{rep} temperature variation can directly affect the accuracy of the outbound signal subtraction in the SBD hybrid. R_{h2} is therefore set higher than R_{rep} to reduce temperature sensitivity.

VI. Physical Design

The ISR-SBD link is intended for very-short-reach (VSR) communication between two chips mounted on a conventional silicon interposer package to enable high edge bandwidth density. However, to expedite the fabrication process, the link was implemented as a self-contained test site on a 5-nm chip, with limited access to external micro-bump connections. Thus, an on-chip channel was designed as a proxy for interposer traces since the fine feature sizes of both the technologies result in lossy channels with RC-dominated characteristics. This section details the designs of the targeted interposer channel, the on-chip proxy channel, and the 5-nm test site.

A. Target Four-Rank Interposer Channel Design

Fig. 12 represents a targeted interposer bump array, and Fig. 13(a) shows an interposer cross section with four copper routing layers and a thick aluminum RDL grid for power distribution. In addition to spatially offsetting signal traces every other layer, shielding wires (green), connected to either VDD or GND, are inserted between signal traces (orange) to reduce crosstalk. Maximum trace pitch is set by the micro-bump pitch, bump array configuration, PHY depth, on-package power-distribution, and the number of interposer routing layers, as shown in Fig. 12. Targeting a four-rank system with four interposer routing layers (one layer for every PHY

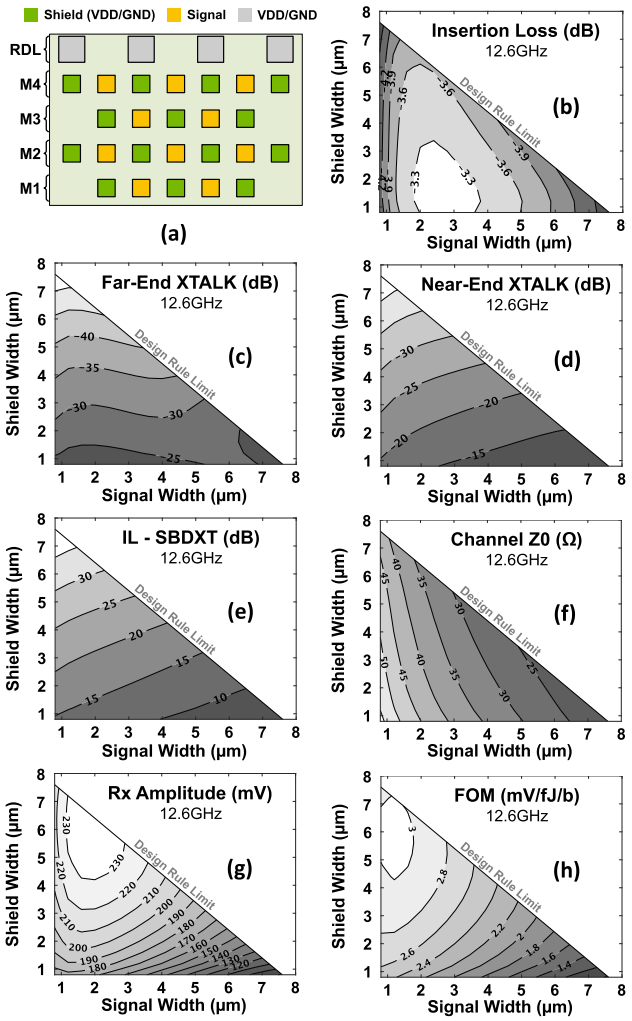


Fig. 13. Four-rank interposer channel. (a) Cross section and channel optimization contour plots for 12.14 μm signal-to-signal pitch and 1.2 mm lengths, (b) insertion loss, (c) power-sum far-end crosstalk, (d) power-sum near-end crosstalk, (e) difference in insertion loss and power-sum SBD crosstalk, (f) channel characteristic impedance, (g) received signal amplitude for 750-mV supply, and (h) ratio of received amplitude to line energy.

deep), a PHY bump array with 20 signals (18 bi-directional data + 2 uni-directional clocks) and 55-μm interstitial bump pitch (55 μm X-pitch and 48 μm Y-pitch) would result in a maximum trace pitch of 6.07 μm with 1.2-mm trace lengths.

Fig. 13(b)–(h) shows contour plots of various channel properties across signal and shield trace widths (x-dimension and y-dimension, respectively) for a fixed 6.07-μm trace pitch (12.14-μm signal-to-signal pitch) and 1.2-mm trace lengths. The channel was modeled using a 2-D-field solver with 45 Ω drive and termination resistances, 200-fF TX/RX parasitic capacitances, and 315 Ω hybrid impedances. Fig. 13(b)–(d) shows contour plots of channel insertion loss (IL), power-sum far-end crosstalk (PSFEXT), and power-sum near-end crosstalk (PSNEXT). Typical interposer links are only subject to PSFEXT, and thus use equal trace widths and spacings to balance signal attenuation with accumulated crosstalk. For the SBD link described in this work, simultaneous bi-directional crosstalk (SBDXT) is a major concern since

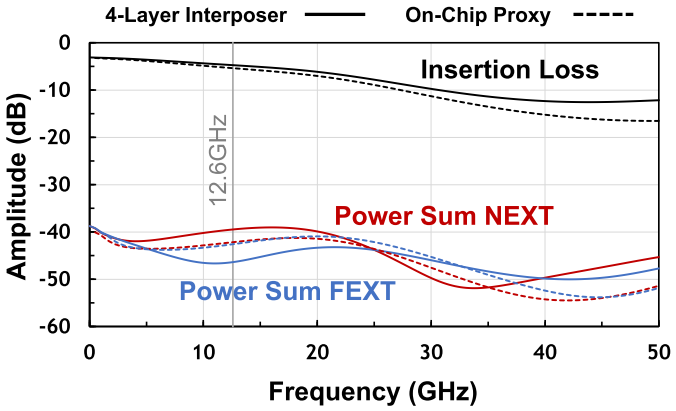


Fig. 14. On-chip and targeted four-rank silicon interposer channel frequency responses.

full-swing aggressors operate continuously on both ends of the link. This results in a victim line being subject to both PSFEXT and PSNEXT simultaneously such that SBDXT = $20 \cdot \log[10^{\mathrm{PSFEXT}/20} + 10^{\mathrm{PSNEXT}/20}]$.

Even with fully shielded traces, SBDXT dominates channel signal integrity (SI) as shown in the IL-SBDXT contour of Fig. 13(e). With trace pitch fixed, signal and shield trace widths become an important design constraint in optimizing channel SI. The optimal solution space exists in the upper-left region where crosstalk reduction through wide shield traces becomes more important than minimizing signal losses within the channel.

While maximum shield widths reduce crosstalk to >39 dB below IL, it lowers the channel characteristic impedance (Z_0) as shown in Fig. 13(f), requiring increased line energy. In addition, the voltage transfer function dependency on IL starts to become more significant as crosstalk is >25 dB below IL, as seen in the received signal amplitude in Fig. 13(g).

An optimal solution becomes apparent when looking at the ratio of received signal amplitude to line energy, Fig. 13(h), where the figure of merit ratio is maximum near signal/shield trace widths of 0.8/6.4 μm. This configuration suppresses crosstalk through wide non-maximum shield widths, which, in conjunction with minimum signal widths, achieves $Z_0 = 45\ \Omega$ to achieve sufficient eye margin with moderate line energy.

B. On-Chip Proxy Channel Design

Because there was limited access to micro-bumps on the test site, an on-chip channel was designed to match the channel characteristics of the interposer channel for a four-rank system from Section VI-A. The on-chip proxy was implemented in the upper metal layers using 1.2-mm-long traces, sized to balance insertion loss and channel impedance.

Fig. 14 shows the simulated frequency response of the targeted interposer and on-chip channels. The 3-D modeling of the on-chip channel showed total IL, PSFEXT, and PSNEXT of −5.4, −42.4, and −42.2 dB, respectively, at a 12.6-GHz Nyquist. This is in comparison to the interposer channel with 50-μm on-chip RDL routes on both ends,

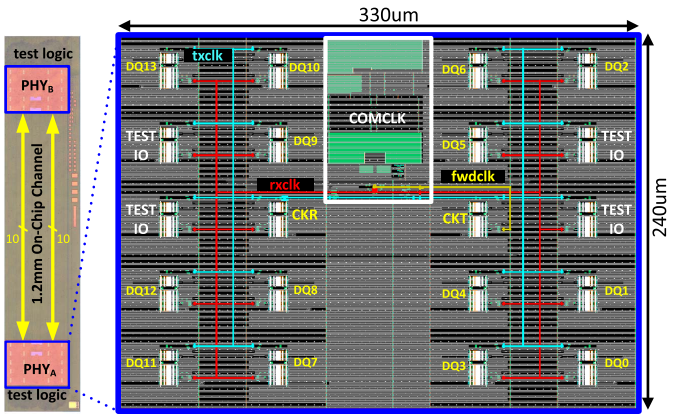


Fig. 15. Test site micrograph and ISR-SBD PHY layout.

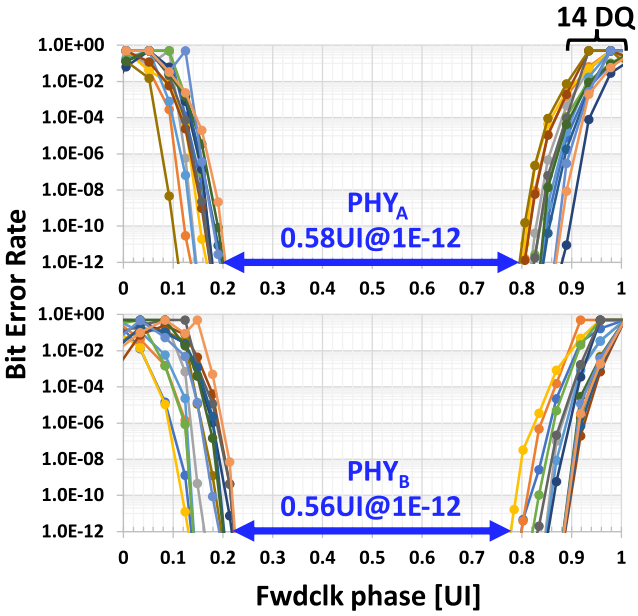


Fig. 16. Bit error rate of 14 lanes/PHY at 50.4 Gb/s/wire.

which exhibited IL = −4.8 dB, PSFEXT = −46.4 dB, and PSNEXT = −39.5 dB. The on-chip channel has slightly worse IL (−0.6 dB) and a lower Z_0 of 42 Ω, making it more difficult from an attenuation standpoint.

The on-chip channel also exhibits similar channel impedances of R = 17.7 Ω/mm, L = 339 pH/mm, and C = 210 fF/mm at 12.6 GHz, where additional upper metal routes connecting the PHY circuitry and on-chip channel help mimic losses from ESD capacitance and RDL traces in the targeted interposer channel.

C. Test Site Implementation

A chip in a 5-nm finFET technology node was the host for the ISR-SBD test-site, which shares the 750-mV processor core logic supply with the host. A die micrograph is shown in Fig. 15, where two PHY instances are placed 1.2 mm apart and connected through the on-chip channel proxy (Section VI-B), which mimics the channel SI of a four-rank interposer system. Each PHY measures 330 × 240 μm and consists of 20 I/O lanes comprising 14 SBD data lanes (DQ 0-13), two uni-directional forwarded clock lanes (CKT/CKR), and four lanes (TEST IO) for unrelated experiments. The clock distribution network is designed to drive all 20 lanes. More than 50% of the PHY area is occupied by decoupling capacitors.

The COMCLK block, located in the center of a PHY, consists of the PLL, two phase interpolators, and CMOS drivers for global clock distribution. Only the IO circuitry that requires ESD-tolerant layout and passive resistors were laid out manually; otherwise, the design uses a CMOS high-speed logic library built and characterized for automatic place & route (P&R). We use a manually guided P&R flow in which designers pre-define the placement of critical cells and nets so high-speed performance is not compromised. Static timing analysis has been performed for SER, DES, and the entire lane. The entire COMCLK, including a VCO and PI, was also built using the same P&R flow.

VII. Measurement Results

Since the test site described in Section VI-C has no external I/O access, all the measurements were performed by built-in self-test using internal PRBS generators and checkers. A common 100-MHz reference clock was inputted to both PLL instances and the multiplication factor ($2N$, where N is an integer) was set to 126 ($N = 63$) to generate 12.6-GHz clock in each PHY. All DQ lanes simultaneously transmit and receive PRBS31 patterns at 25.2 Gb/s with independent seeds between lanes. The power supply was set to 750 mV, and the calibrations described in Section V were performed except for step (d) as a process monitor circuit connected to an external resistor was unavailable on this chip.

A. Bit Error Rate

We measured the bit error rate of all 14 DQ lanes at each end of the link running simultaneously at 25.2 Gb/s/direction or 50.4 Gb/s/wire, as shown in Fig. 16. Fwdclk phase interpolator was rotated for eye scanning with nominal resolution of (1/30)th of a UI. PHY_A showed a horizontal eye-opening of 0.58UI at 1E-12. Concurrent measurement for PHY_B showed an opening of 0.56UI. Using the slope of each curve, the effective Rj was extracted as 0.75 ps_{rms} which allows us to calculate the bit error rate to the level that is not actually measurable. The eye margin at 1E-25 was extrapolated to be 0.45UI for PHY_A and 0.43UI for PHY_B. We have also measured the horizontal eye margin in an asynchronous case where A to B was running at 25.2 Gb/s ($N = 63$) and B to A at 24.8 Gb/s ($N = 62$). The results were identical to Fig. 16. This was expected since our design does not rely on fixed frequency or phase relationships between the two PLLs.

B. Power Supply Rejection

One of the most attractive characteristics of a delay-matched clock-forwarded link is its high supply rejection. Although ISR-SBD eye-opening characterization and the supply noise measurement could not be performed on the same die because

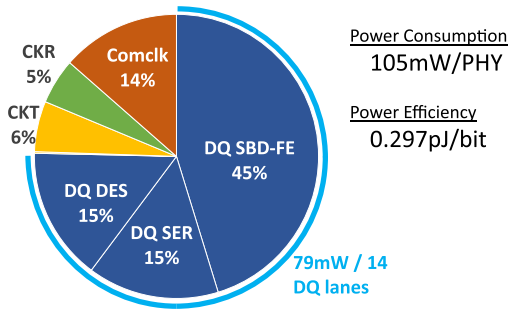


Fig. 17. Power consumption at 50.4 Gb/s/wire.

of limited access to power/ground probe pads, the dc supply voltage and temperature are kept the same between the two experiments such that the noise profiles of them remain similar. Between idle and active states of the processor core logic, the ISR-SBD's eye-opening degraded by only one PI step (∼1.3 ps nominal). To verify the amount of supply noise, power and ground probe points were exposed on another chip with the same design. The observed pk-pk noise was 65 mV when the processor core logic was idle. It jumped to 120 mV when the core activity was maximized. These high noise numbers and limited performance degradation of ISR-SBD re-verify the simulation results shown in Fig. 8.

C. Power Consumption

The measured power consumption is shown in Fig. 17 along with the break-down based on simulations. A PHY with one CKR lane, one CKT lane, 14 DQ lanes, and one COMCLK lane consumes 105 mW at 750 mV. Three-quarters of the power is consumed in the DQ lanes, where the SBD front-end contributes 45%, and the SER and DES use 15% each. The rest, which is about a quarter, comes from clocking, including one clock transmitter lane, one CKR lane, one PLL, and two PIs. The energy efficiency for the entire link at 50.4 Gb/s/wire is 105 [mW/PHY]*2[PHY]*(14[wire]*50.4[Gb/s/wire])$^{-1}$ = 0.297 pJ/bit. As shown in Fig. 15, we have four extra lanes (TEST IO lanes) for experiments unrelated to SBD. They can be easily replaced with SBD lanes to compose an 18-DQ lane PHY, which is the maximum number of DQ lanes supported by the clock distribution network in our design. Since we know the power consumption of 14 DQ lanes, we can calculate the 18-DQ PHY power efficiency by ((18/14)*79[mW] + 26 [mW])*2*(18*50.4 Gb/s)$^{-1}$ which is 0.281 pJ/bit.

VIII. Comparison

Comparison with recent work is shown in Table I. Although we have used the most advanced node, the area of the PHY did not benefit from smaller transistors because it is dictated by the target micro-bump pitch of 55 μm, most commonly used for HBM. Our test site channel is 1.2 mm, which is not the longest among the ones listed, but it can comfortably cover the majority of use cases on interposers. A single supply of 750 mV is the lowest, which allowed us to simply share the supply with processor core logic. The per-wire data rate is more than doubled, thanks to the use of SBD. It also has the best energy efficiency. It is important to note that our case counts all the contributions from every clocking component including clock generation, distribution, and forwarding.

TABLE I
Comparison to Other Published Interposer-Based Links

	Our work	VLSI21 [3]	JSSC20 [16]	JSSC16 [17]
Technology	5nm	7nm	7nm	28nm
Interposer Channel Length [mm]	1.2	1.0	0.5	2.5
Bump Pitch [μm]	55	40	40	100
Supply [mV]	750	800	800, 300	N/A
Data Rate/Wire [Gb/s]	50.4	20	8	20
Energy Efficiency [pJ/bit]	0.297 0.281 [a]	0.46	0.56	0.3 [d]
Edge Density [Tb/s/mm]	2.14 11.0 [a,b]	5.31	0.67	N/A
Areal Density [Tb/s/mm^2]	4.45 5.73 [a,c]	2.25	0.8	N/A

[a] 18-DQ PHY [b] Equivalent 72-DQ 4-rank interposer system
[c] $18DQ * 25.2Gb/s/(0.33mm * 0.24mm)$ as defined in [3]
[d] Clocking power not included.

The edge density achieved in our 14-DQ test site is 2.14 Tb/s/mm (14 DQ*50.4 Gb/s/0.33 mm), which translates to 11 Tb/s/mm for an equivalent 72-DQ four-rank interposer system using a 55-μm bump pitch, as modeled in our on-chip proxy channel. As the technology advances, scaling to a 40-μm bump pitch will enable 15 Tb/s/mm. The final row of the table lists the areal density. Our 20-lane test site shows 4.45 Tb/s/mm^2 (14 data, two clocks, and four unrelated experiments) which nearly doubles the number most recently reported in [3]. If all 20 available lanes are used for SBD (18 data + two clocks), the density further improves to 5.73 Tb/s/mm^2.

IX. Conclusion

This article demonstrated an inverter-based short-reach simultaneous bi-directional (ISR-SBD) link running at 50.4 Gb/s/wire over 1.2-mm traces. A unique single-ended SBD hybrid with CMOS inverters and passive resistors has been introduced, which significantly simplifies the design. The PHY with best-in-class power efficiency and areal density meets the demand of 5–8 years into the future, by offering 11 Tb/s/mm edge density with 55-μm bump pitch. In addition, predictable scaling of micro-bump density will take us to 15 Tb/s/mm edge density with 40-μm bump pitch.

Acknowledgment

The authors would like to thank Jamie Yan and the Silicon Solution Group at NVIDIA for their assistance in supply noise measurements.

REFERENCES

[1] W. Turner, "High-speed interconnect challenges within systems leveraging advanced packaging techniques," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 529–532.

[2] B. Casper, "Energy efficient multi-Gb/s I/O: Circuit and system design techniques," in *Proc. IEEE Workshop Microelectron. Electron Devices*, vol. 22, Apr. 2011, p. 1.

[3] Y.-Y. Hsu, P.-C. Kuo, C.-L. Chuang, P.-H. Chang, H.-H. Shen, and C.-F. Chiang, "A 7 nm 0.46pJ/bit 20Gbps with BER 1E-25 die-to-die link using minimum intrinsic auto alignment and noise-immunity encode," in *Proc. Symp. VLSI Circuits*, Jun. 2021, pp. 1–2.

[4] L. Dennison and W. W. J. Lee Dally, "High performance bidirectional signalling in VLSI systems," in *Proc. Res. Integr. Syst., Symp.*, 1993, pp. 1–20.

[5] R. Mooney, C. Dike, and S. Borkar, "A 900 Mb/s bidirectional signaling scheme," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1538–1543, Dec. 1995.

[6] M. Haycock et al., "A 2.5Gb/s bidirectional signaling technology," in *Proc. Hot Interconnects Symp. Rec.*, Aug. 1997, pp. 149–156.

[7] H. Wilson and M. Haycock, "A six-port 30-GB/s nonblocking router component using point-to-point simultaneous bidirectional signaling for high-bandwidth interconnects," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 1954–1963, Dec. 2001.

[8] R. Drost and B. Wooley, "An 8-Gb/s/pin simultaneously bidirectional transceiver in 0.35-/spl μ/m CMOS," *IEEE J. Solid-State Circuits*, vol. 39, no. 11, pp. 1894–1908, Nov. 2004.

[9] J.-H. Kim et al., "A 4-Gb/s/pin low-power memory I/O interface using 4-level simultaneous bi-directional signaling," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 89–101, Jan. 2005.

[10] Y. Tomita et al., "A 20-Gb/s simultaneous bidirectional transceiver using a resistor-transconductor hybrid in 0.11-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 627–636, Mar. 2007.

[11] Y.-H. Fan et al., "A 32-Gb/s simultaneous bidirectional source-synchronous transceiver with adaptive echo cancellation techniques," *IEEE J. Solid-State Circuits*, vol. 55, no. 2, pp. 439–451, Feb. 2020.

[12] J. W. Poulton et al., "A 1.17-pJ/b, 25-Gb/s/pin ground-referenced single-ended serial link for off- and on-package communication using a process- and temperature-adaptive voltage regulator," *IEEE J. Solid-State Circuits*, vol. 54, no. 1, pp. 43–54, Jan. 2019.

[13] T.-C. Hsueh et al., "26.4 A 25.6Gb/s differential and DDR4/GDDR5 dual-mode transmitter with digital clock calibration in 22 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 444–445.

[14] M. Mansuri, B. Casper, and F. O'Mahony, "An on-die all-digital delay measurement circuit with 250fs accuracy," in *Proc. Symp. VLSI Circuits (VLSIC)*, Jun. 2012, pp. 98–99.

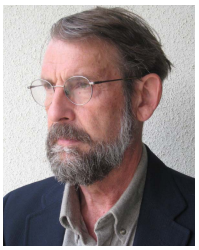
[15] R. Z. Bhatti, M. Denneau, and J. Draper, "Duty cycle measurement and correction using a random sampling technique," in *Proc. 48th Midwest Symp. Circuits Syst.*, vol. 2, Aug. 2005, pp. 1043–1046.

[16] M.-S. Lin et al., "A 7-nm 4-GHz Arm[1]-core-based CoWoS[1] chiplet design for high-performance computing," *IEEE J. Solid-State Circuits*, vol. 55, no. 4, pp. 956–966, Apr. 2020.

[17] B. Dehlaghi and A. C. Carusone, "A 20 Gb/s 0.3 pJ/b single-ended die-to-die transceiver in 28 nm-SOI CMOS," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Sep. 2015, pp. 1–4.

Yoshinori Nishi (Member, IEEE) received the B.S. and M.S. degrees in low-temperature physics from Waseda University, Tokyo, Japan, in 1997 and 1999, respectively.

From 1999 to 2003, he was with NTT Electronics Inc., Atsugi, Japan, as a Member of the Ultrahigh-Speed Device Development Group, where he was a Chief Designer for the 50Gb/s InP HEMT logic family, first 50Gb/s product in the market in 2001. In 2003, he joined Kawasaki Microelectronics Inc., Chiba, Japan, Research and Development Division, where he led the development of 10 Gb/s burst-mode CDR for 10G-EPON application as a Chief Architect, also first in the market in 2009. He joined NVIDIA Corporation, Santa Clara, CA, USA, in 2011, and led one of the NVLINK physical layer (PHY) design teams for eight years before he joined NVIDIA Research in 2020. His current research focuses on ultralow power and high-density interconnect for short-reach applications.



John W. Poulton (Fellow, IEEE) received the B.S. degree from the Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 1967, the M.S. degree from the State University of New York, Stony Brook, NY, USA, in 1969, and the Ph.D. degree from the University of North Carolina at Chapel Hill (UNCCH), Chapel Hill, NC, USA, in 1980, all in physics.

From 1981 to 1999, he was a Researcher with the Department of Computer Science, UNCCH, where he has been a Research Professor since 1995. He performed research on VLSI-based architectures for graphics and imaging and was a principal contributor to the design and construction of several generations of the pixel-planes and pixel-flow graphics systems, the latter a commercial product produced briefly by Hewlett-Packard. While at UNC, he designed the custom CMOS circuits that implemented the beam-former in one of the first commercial 3-D medical ultrasound systems. He also collaborated with Prof. William Dally to produce the textbook "Digital Systems Engineering" (1998) and to demonstrate one of the first CMOS chip-to-chip serial data links that used transmitter equalization to overcome channel-induced inter-symbol interference (ISI). From 1999 to 2003, he was the Chief Engineer with Velio Communications, where with Dally and others he developed several multi-gigabit chip-to-chip signaling systems and products based on them, including the first chip to demonstrate 1 terrabit/s of off-chip bandwidth. From 2003 to 2009, he was the Technical Director with Rambus, Inc., Chapel Hill, NC, USA, where he led an effort to build power-efficient multi-gigabit I/O systems. This effort culminated in a demonstration of a complete serial link operating at 4 pJ/bit, the lowest energy per bit demonstrated up to that time (2006–2007), and a highly energy- and pin-efficient interface for mobile DRAM (2010). Presently, he is at NVIDIA Inc., Durham, NC, USA, where he continues work on energy-efficient on- and off-chip communications circuits.

Walker J. Turner (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2009, 2012, and 2015, respectively.

He is a Senior Research Scientist at NVIDIA, Durham, NC, USA. In 2014, he worked as a Contractor for the U.S. Army Research Laboratory, Adelphi, MD, USA, developing wirelessly powered systems and integrated low-noise amplifiers for piezoelectric E-field sensors. He joined the Circuits Research Group, NVIDIA, in 2015, where he works on energy-efficient, high-speed signaling systems for on- and off-chip communication. He also served as a Teaching Assistant Professor at North Carolina State University (NCSU), Raleigh, NC, USA, from 2019 to 2020, where he instructed an undergraduate micro-electronics engineering course.

Xi Chen (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2003 and 2006, respectively, and the Ph.D. degree in electrical engineering from North Carolina State University (NCSU), Raleigh, NC, USA, in 2011.

From 2006 to 2007, he was an Analog IC Design Engineer with Accel Semiconductor, Shanghai, China. From 2007 to 2011, as a Research Assistant at Electrical and Computer Engineering Department, NCSU, he performed research on 3-D IC design methods and variation-adaptive mixed-signal ICs. He joined NVIDIA, Inc., Santa Clara, CA, USA, in January 2012 and is now a Senior Research Scientist. His current research interests include high-speed low-power I/O circuit design, energy-efficient clocking, and signaling for advanced packaging.

Dr. Chen served as a Reviewer and TPC Member for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPER, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: BRIEF PAPER, and CICC.

Sanquan Song (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2011.

From 2010 to 2013, he was with Intel Corporation, Santa Clara, CA, USA. From 2013 to 2015, he was with the Samsung Display Research and Development Laboratory, San Jose, CA, USA. In 2015, he joined the Circuits Research Group, NVIDIA Inc., Santa Clara focusing on high-speed links. His current research interests include electrical SerDes, security circuits, and silicon photonic links.

Brian Zimmer (Member, IEEE) received the B.S. degree in electrical engineering from the University of California, Davis, CA, USA, in 2010, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 2012 and 2015, respectively.

He is currently a Senior Research Scientist with the Circuits Research Group, NVIDIA Inc., Santa Clara, CA, USA. His research interests include soft error resilience, chip-to-chip interconnects, energy-efficient digital design, low-voltage SRAM design, machine learning accelerators, productive design methodologies, and variation tolerance.

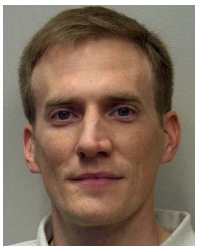
Dr. Zimmer has served for the CICC and VLSI Symposium technical program committees.

Stephen G. Tell (Member, IEEE) received the B.S.E. degree in electrical engineering from Duke University, Durham, NC, USA, in 1989, and the M.S. degree in computer science from the University of North Carolina, Chapel Hill, NC, USA, in 1991.

He worked on parallel graphics systems and high-speed signaling as a Senior Research Associate at UNC/Chapel Hill, Chapel Hill, from 1991 to 1999, and in 1999, he joined Chip2Chip Inc. (later renamed Velio Inc.), to develop circuits and control systems for high-speed SerDes products. This work continued at Rambus, where he designed the logic for a SerDes with the lowest energy per bit demonstrated up to that time. He joined NVIDIA Inc., in 2009, as a Founding Member of the Circuits Research Group, where he works as a Senior Research Scientist. He has been awarded more than 20 U.S. patents. His current research interests include custom circuit design and the surrounding test and control logic for intra- and inter-chip communication.

Nikola Nedovic (Member, IEEE) received the Dipl.Ing. degree in electrical engineering from the University of Belgrade, Belgrade, Serbia, in 1998, and the Ph.D. degree from the University of California at Davis, Davis, CA, USA, in 2003.

He is a Research Scientist at NVIDIA Corporation, Santa Clara, CA, USA. In 2001, he joined Fujitsu Laboratories of America Inc., Sunnyvale, CA, USA, where he worked on high-speed communications and high-performance and low-power circuits for electrical and optical communications. In 2016, he joined NVIDIA Research, where he works on system and circuit design for low-power high-speed links. His research interests include a range of aspects of high-speed electrical and optical wireline communications, from devices and SI to adaptive filtering and system design and modeling.



John M. Wilson (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from North Carolina State University (NCSU), Raleigh, NC, USA, in 1993, 1995, and 2003, respectively.

He is a Principal Research Scientist at the Circuits Research Group, NVIDIA. His work focuses on pathfinding for high-bandwidth density data-movement solutions and the transfer of developed technology from research to production. From 2015 to 2017, he led a team in the design of a 25-Gb/s single-ended ground referenced signaling link, in 16-nm FinFET CMOS, enabling Tera-Byte per-second communication for MCM and PCB channels at 1.17 pJ/bit. His interests include high-speed I/O circuit design, on-chip signaling, SI, advanced packaging, and chip/package co-design. From 2003 to 2006, he was a Research Professor at NCSU leading projects in advanced packaging, low-power capacitive and inductive coupled transceivers for 3-D-ICs, and circuits for on-chip global signaling. From 2006 to 2012, while with Rambus Inc., Chapel Hill, NC, USA, he worked on high-speed I/O circuit design and methods to mitigate signal and power integrity problems in memory interfaces. He has more than 72 publications and has 47 granted patents.

William J. Dally (Fellow, IEEE) is a Chief Scientist and a Senior Vice President of Research at NVIDIA Corporation, Santa Clara, CA, USA, and an Adjunct Professor and a Former Chair of computer science at Stanford University, Stanford, CA, USA. He is currently working on developing hardware and software to accelerate demanding applications including machine learning, bioinformatics, and logical inference. He has a history of designing innovative and efficient experimental computing systems. While at Bell Laboratories, he contributed to the BELLMAC32 microprocessor and designed the MARS hardware accelerator. At Caltech, he designed the MOSSIM Simulation Engine and the Torus Routing Chip which pioneered wormhole routing and virtual-channel flow control. At the Massachusetts Institute of Technology, Cambridge, MA, USA, his group built the J-Machine and the M-Machine, experimental parallel computer systems that pioneered the separation of mechanisms from programming models and demonstrated very low overhead synchronization and communication mechanisms. At Stanford University, his group developed the Imagine processor, which introduced the concepts of stream processing and partitioned register organizations, the Merrimac supercomputer, which led to GPU computing, and the ELM low-power processor. He currently leads projects on computer architecture, network architecture, circuit design, and programming systems. He has published over 250 papers in these areas, holds over 160 issued patents, and is an author of the textbooks, *Digital Design: A Systems Approach, Digital Systems Engineering, and Principles and Practices of Interconnection Networks*.

Dr. Dally is a member of the National Academy of Engineering and a fellow of the ACM and the American Academy of Arts and Sciences. He received the ACM Eckert-Mauchly Award, the IEEE Seymour Cray Award, the ACM Maurice Wilkes Award, the IEEE-CS Charles Babbage Award, the IPSJ FUNAI Achievement Award, the Caltech Distinguished Alumni Award, and the Stanford Tau-Beta-Pi Teaching Award. He is a member of President Biden's Council of Advisors on Science and Technology (PCAST).

C. Thomas Gray (Senior Member, IEEE) received the B.S. degree in computer science and mathematics from Mississippi College, Clinton, MS, USA, in 1988, and the M.S. and Ph.D. degrees in computer engineering from North Carolina State University (NCSU), Raleigh, NC, USA, in 1990 and 1993, respectively.

From 1993 to 1998, he was an Advisory Engineer with IBM, Research Triangle Park, NC, USA, working on transceiver design for communication systems. From 1998 to 2004, he was a Senior Staff Design Engineer with the Analog/Mixed Signal Design Group, Cadence Design Systems, working on SerDes system architecture. From 2004 to 2010, he was a Consultant Design Engineer with Artisan/ARM and a Technical Lead of SerDes architecture and design. In 2010, he joined Nethra Imaging as a System Architect. His work experience includes digital signal processing design and CMOS implementation of DSP blocks and high-speed serial link communication systems, architectures, and implementation. In 2011, he joined NVIDIA Inc., Durham, NC, USA, where he is currently the Senior Director of Circuit Research, leading activities related to high-speed signaling, photonics, security circuits, low-energy and resilient memories, circuits for machine learning, and variation-tolerant clocking and power delivery.