As the end of the Moore's law era approaches, increases in computational performance can no longer rely on the historical scaling of integrated circuit feature sizes that the industry has depended on for the last half century. This reduction in scaling trends, combined with manufacturing and packaging limits to maximum die size, will constrain single-die computational performance. Therefore, high-performance computers of the future will need to rely more heavily on architectural and system-level improvements. One way to provide these improvements is through the use of advanced packaging techniques, such as multichip modules (MCMs), where systems too large for a single die consist of multiple semiconductor dice, connected as shown in Figure 1, using on- and off-package high-speed, short-reach links. These short-reach links are essential to scaling computational performance in systems where several processing units must communicate at data rates approaching on-chip total bandwidth while consuming a small fraction of total device power. Another key component to scaling aggregate performance in large computational systems is the use of optical links for long-haul communication. In these systems, processors and networking switches use short-reach links to connect to special purpose electrical/optical (E/O) interface chips [1]. Even memory interfaces (e.g., high-bandwidth memory) have begun the shift toward energy-efficient short-reach links to meet market demands [2]. Overall, our goal is to enable massively scalable parallelism, where it is crucial to understand that the energy-cost of communication is a tax on computational performance. In other words, every Joule saved in communication is a Joule that can be used for computation.

The key metrics of a short-reach signaling system are bandwidth density and energy efficiency. Bandwidth density describes how much data (i.e., information) can be funneled through a given geometry and is measured in both bits/s/mm² of chip area and bits/s/mm of chip periphery. Energy efficiency describes how much energy is required to move information between two points and is measured in Joules per bit. MCMs are great example systems that require a short-reach...
signaling solution to maximize the amount of information that can be communicated within both spatial and energy constraints. Given these constraints, single-ended (SE) signaling offers the best solution when simultaneously optimizing across both metrics because it requires only one pin per signal compared to two pins per signal with differential signaling, which is used for most high-speed chip-to-chip links. It is important to emphasize that, while SE signaling sacrifices the advantages of differential signaling, in limited usage scenarios such as short-reach links, its more efficient metrics provide advantages over differential signaling. However, as discussed in the following section, densely routed SE signaling at several tens of gigabits per second presents many technical challenges. In response to this, we developed low-swing SE-signaling techniques that provide high energy efficiency when combined with simple methods that carefully control channel signal integrity.

In this article, we discuss how ground-referenced signaling (GRS) avoids many of the challenges associated with SE signaling, thanks to a unique transmitter (Tx) topology, to produce a link with high reliability, high information density, and high energy efficiency for short-reach applications. GRS-based short-reach links also use a simple but robust clock-forwarding scheme to cancel jitter without the need for a clock recovery system at the receiver (Rx), thereby, saving significant power. This approach also provides high-bandwidth jitter tracking, which we take to the limit by matching circuit delay and delay sensitivity for clock and data lanes at both ends of the link. Codesign of the link circuitry and channel takes full advantage of the ability to match the insertion delays of multiple data/clock lanes within the short-reach interconnect. The low attenuation that can be achieved in a codesigned channel is leveraged to achieve a simple, robust, and low-power design. The most recent GRS-based link introduces several innovations: 1) a dynamic voltage scaling (DVS) power supply regulator, which implements the inverse of dynamic frequency scaling and holds link performance constant over process and temperature variation; 2) link calibration circuitry, which, once set at arbitrary supply voltage and temperature, removes process variation and tracks subsequent temperature changes without the need for periodic recalibration; and 3) a fast entry/exit pause mode that reduces link standby power to about 25% of active power, while holding entry/exit times to <5 ns.

Challenges and Solutions for SE Signaling

The fastest conventional SE systems are those found in graphics memory interfaces [2], [3] employing the signaling and termination scheme shown in Figure 2(a). This approach uses a voltage-mode driver, with series resistance ($R_{\text{drive}+}$ and $R_{\text{drive}-}$), to improve back-match to the channel and a termination resistor ($R_{\text{term}}$) connected to the supply ($V_{\text{dd}}$) at the Rx. When first introduced, this approach eased the transition from the signaling used in main-memory interfaces by providing a higher common-mode (CM) voltage at the Rx for the slow transistors found in a dynamic RAM process, which is especially important because the NMOS transistor is notably faster than the PMOS transistor. In addition, this termination scheme also allowed for low-weight bus encoding to be used to save power and reduce simultaneous switching output (SSO) noise [2]. (Low-weight bus encoding guarantees that no more than 50% of the encoded output signals can conduct current simultaneously.) However, all of the challenges associated with SE signaling remain in these systems.

The first challenge is that conventional SE systems require the generation of a reference voltage ($V_{\text{ref}}$), as shown in Figure 2(b), nominally halfway between the HI and LO signal levels at the input to the Rx. There are several ways to generate the reference voltage, and all produce some amount of error (i.e., voltage reference uncertainty). This error is a fixed noise source, and conventional SE links must use sufficient signal amplitude to overcome the uncertainty. The second challenge is that conventional SE systems have a very complex and frequency-dependent return path, where their large signaling currents flow through the on-chip power distribution network (PDN) [3], as shown in Figure 2(c). The third challenge, known as SSO noise, creates the single

**FIGURE 1:** The very short reach on-package (MCM) and short-reach off-package (PCB) link examples. DRAM: dynamic RAM; GPU: graphics processing unit.
largest proportional form of noise in SE systems and ties into the use of the PDN as the return path (i.e., the second challenge). SSO noise arises because the signaling current is data dependent and has broadband frequency content. This broadband noise current flows through the PDN, where it can easily excite any resonance frequency of the power supply. This causes the on-chip power supply voltage to expand and collapse at the resonant frequency, which makes it difficult to maintain predictable circuit operation. In a conventional graphics memory interface, the signaling current ranges from 0 mA, for a logic 1, to about 15 mA/lane, for a logic 0, as illustrated in Figure 2(d).

The problem is made worse by the fact that the aggregate current from hundreds of lanes shares the same PDN at both ends of the link. In addition to modulating the power supply voltage, the noise from the large switching current couples through the PDN and onto the transmitted/received

**FIGURE 2:** (a) An example of a conventional SE link, (b) the voltage reference generation, (c) the complex signal return path, and (d) the SSO noise due to data-dependent current. GND: ground.
signals, degrading both the voltage and timing margins. To make matters worse, self-generated noise (i.e., SSO noise) is proportional to signal amplitude, and attempting to overcome it with even larger signal amplitude is futile. This large self-generated noise degrades link reliability because it alters the operating behavior of the circuitry in a nonlinear way. Another proportional noise source, which is a problem for all densely routed SE links operating above a few gigabytes per second, is crosstalk. Overall, it is the combination of these issues that creates a signaling environment where conventional SE links exhibit poor energy efficiency.

GRS, shown in Figure 3(a), uses a current-mode Tx with shunt resistance ($R_{\text{Tx}}$) to provide a back-match to the channel and a termination resistor ($R_{\text{TERM}}$) connected to ground at the Rx [4]. GRS avoids many of the challenges associated with SE signaling, beginning with the reference voltage needed at the Rx to recover data from the line signal. Because the ground network is typically the lowest impedance network in a system and the only voltage that everyone agrees on, it is used as the GRS voltage reference, as illustrated in Figure 3(b). This eliminates the need for generating a precisely matched reference at the Rx that tracks the implied reference at the Tx. In addition, the low impedance of the ground network limits the reference voltage difference between the ends of the link because the ground network leverages more than half of the metal resources dedicated to power distribution in a typical system and often as much as 60–70% of PDN resources. Because ground is used as the signal reference, the GRS Tx drives currents into the line to create a voltage that toggles symmetrically.

**FIGURE 3:** (a) The ground-referenced SE link, (b) the ground as the reference voltage, (c) the ground as the only signal return path, and (d) the constant current to eliminate SSO noise.
GRS takes a unique approach for generating the line signal by employing a pair of charge pumps to drive current into the line. The GRS Tx combines the functions of a two-to-one output multiplexer, bipolar line driver, and step-down voltage converter and even provides partial line termination. The primary advantage of this approach is that it consumes nearly constant current regardless of data polarity. There is a small ripple that occurs at the bit rate, which is two to three decades above the resonance frequency of the PDN, but it is easily filtered by on-chip power supply bypass capacitance.

Figure 4 shows a schematic of the GRS Tx, made up of the main charge pump-based line driver and an auxiliary equalizing Tx. In the Tx, a pair of even and odd half-bit-rate data streams (dat0 and dat1) are two-to-one multiplexed by the charge pumps (Pump0 and Pump1) and equalizer stages. The line driver has two identical charge pump circuits that operate on opposite phases of the half-bit-rate clock. These phases are referred to as the precharge phase and line drive phase. During the precharge phase, Pump0 charges a storage capacitor ($C_S$) to the supply rail while the clock is HI; then, when the clock goes LO, the charge stored in the capacitor is driven into the line, developing a positive line voltage if even data $dat0 = \text{HI}$ and a negative voltage if even data $dat0 = \text{LO}$. Meanwhile, the lower charge pump performs the same operation but on the opposite clock phase, driving the odd data (dat1) onto the line. This will be covered in detail in the discussion of Figure 5.

The transmit equalizer performs additive edge boosting through an ac-coupled voltage-mode driver that drives a current impulse into the line during edge transitions to enhance the speed and amplitude of the output signal transition. An edge detector within the equalizer drives a 1 unit interval (UI) pulse generator whenever the current data bit differs from the previous one. This is performed by an XOR of even and odd data bits [6], thereby ensuring that the equalizer output is driven only during edge transitions and is left in a high-Z state during idle periods to prevent modulation of the Tx return impedance. Cross-coupled inverters are added to the inboard node of the coupling capacitor ($C_{eq}$) to operate as high-impedance keeper cells that weakly hold the node voltage at either supply level during long periods of inactivity. The equalizer is divided into four equally weighted, digitally enabled segments, so that the amount of overdrive can be adjusted at run time to compensate for channel attenuation. Because the charge pump Tx and transmit equalizer are built using CMOS switches, it is easy to match the delays of their data paths at design time. This approach, combined with the DVS regulation scheme, which flattens circuit delay across process, voltage, and temperature (PVT), ensures matched delays. The Tx output is coupled to the line through a T-coil (not shown) that compensates for the parasitic capacitance of the transceiver circuits, electrostatic discharge (ESD) protection clamp,
input–output (I/O) pad, and redistribution layer (RDL) routing. It should be noted that this form of equalization reintroduces some SSO noise, but its relatively low amplitude is easily filtered by on-chip power supply bypass capacitance.

Figure 5 illustrates the operation of the charge pump line driver during precharge and when driving the line HI and LO, using switches to represent MOSFETs M0–M5. During precharge, the H-bridge transistors M0–M3 are all off, while M4 and M5 turn on to charge the storage capacitor $C_S$ to $V_{DD}$. When driving HI, the precharge transistors M4 and M5 are turned off, while M1 and M2 are enabled. Current flows from the capacitor into the line to generate a positive voltage pulse. Alternatively, when driving LO, transistors M0 and M3 are turned on, so $C_S$ is connected to draw current from the line, creating a negative voltage pulse. There are several asymmetries in these phases that tend to make positive and negative drive voltages differ. First, there are parasitic capacitors that assist during the positive drive but oppose during the negative drive [4], [5]. This effect is typically <10% of the drive voltage and can be compensated for by the appropriate sizing of M0–M3. Second, driving LO drives not only the line output below ground but also the bottom terminal of the storage capacitor $C_S$. The gates of M1, M2, and M5 are at 0 V, while the sources of these transistors see a negative voltage excursion, causing the transistors to turn on. For the 100–200-mV amplitudes that can practically be achieved in a GRS Tx of this type, transistors M1, M2, and M5 enter the weak inversion region of operation, where leakage is large enough to introduce drive asymmetry. This leakage diverts current away from the line during the negative drive, further exacerbating drive asymmetry. The leakage problem ultimately limits the drive amplitude of a GRS Tx, but, for short-reach links, 100–200 mV of drive amplitude is sufficient. The problem can be mitigated by choosing the appropriate MOSFET device types for the charge pump transistors. The best choice is standard-threshold devices for M0–M3 (so the line drive transistors track across global corners), a low-threshold PMOS for M4 (to speed up precharge), and a high-threshold device for M5 (because this device is the largest and leakiest NMOS in the pump).

Finally, nonoverlap between precharge and drive phases must be guaranteed. For the negative drive, M3 is bootstrapped by its negative-going source voltage, so the time it takes to fully turn on is slightly delayed and ensures nonoverlap between phases. For positive drive, the gate-drive voltages on M1 and M2 must be slightly delayed to ensure that precharge is completely extinguished before initiating line drive. Otherwise, the trailing edge of the precharge current aids in driving a positive current into the line, creating further asymmetry. Referring to Figure 4, the and gates that logically combine clock (expressed as $clk$) and even/odd data (dat0 and dat1) are modified so that the enable clock is delayed by one inverter delay, while disable is immediate; this mechanism is sufficient to overcome the overlap problem. Asymmetric drive leads to offset in the Tx output signal, which is indistinguishable from the Rx input offset in the calibration scheme we are using. While both forms of offset can be compensated for by tuning the Rx’s input-referred offset voltage, if the Tx output offset varies with temperature and supply voltage, it will be impossible to track output offset without recalibration. Therefore, it is critical to remove Tx output offset to the highest-possible degree at design time. While the DVS regulation scheme removes much of the PVT-dependent output offset, the

![Figure 5: GRS charge pump operation: (a) the precharge phase, (b) and (c) the line drive phase.](image-url)
use of thin-oxide MOS capacitors as the charge pump storage capacitors introduces unacceptable temperature variation in the Tx output offset. Our initial Tx design in 16-nm FinFET CMOS used a significant portion of the MOS capacitors for \( C_s \), which resulted in 20 mV of output offset variation across temperature. In the refined design described here, nearly all of the storage capacitor \( (C_s) \) is implemented as metal–oxide–metal capacitors, which reduced the Tx’s output offset variation to 3.7 mV across temperature when combined with the nonoverlap drive timing.

A charge pump–style Tx is quite general purpose. It is easy, for example, to perform higher multiplexing ratios than two-to-one to attain higher speeds. Tx elements may be combined to perform equalization for de-emphasis or to generate PAM-4 modulated outputs. When driving the line, a charge pump Tx is floating (i.e., not referred to either power supply); this feature may be useful in situations where it is desirable to allow the Rx to set the CM voltage in a differential signaling system. The downside of a charge pump Tx is that the range of operating speeds is somewhat limited, because lower-bit rate pumps require larger values of \( C_s \) storage capacitors to generate the same line voltage.

**Termination Scheme**

The return impedance of the charge pump Tx is the sum of the switch resistance and the effective resistance of the switched-capacitor circuit in the pump, given by \( 1/C_s f \). With \( C_s = 400 \) fF and \( f = 25 \) GHz (there are two pumps in parallel), this impedance is about 100 \( \Omega \). Including switches, the total return impedance is about 140 \( \Omega \). A detailed circuit simulation of the Tx gives an average equivalent impedance very close to this estimate. While the Tx’s average impedance is well controlled over PVT, thanks to the DVS regulation scheme, the switch resistance varies considerably during a drive event, as the transistors pass through saturation and triode modes of operation. In any case, both the average and instantaneous charge pump impedances are higher than the typical channel impedance, so the Tx can be viewed as a current source. In cases where the channel response is smooth, with no significant impedance discontinuities (e.g., an on-package channel), no back-match is needed, and the Tx can drive the line from its own source impedance to provide the largest possible signal into the channel. For cases where the line must be back-matched to absorb reflections from discontinuities in a more complicated channel (e.g., a printed circuit board (PCB) channel), an adjustable parallel termination resistor is required to improve return loss. This same termination resistor is repurposed as the incident wave termination at the Rx input because this GRS link is designed as a transceiver.

To provide an example, consider back-matching a 40-\( \Omega \) line. The Tx drives the line through an on-chip interconnect with a total series resistance of 4.6 \( \Omega \). Setting the termination resistor to 47 \( \Omega \) (in parallel with the Tx’s 140 \( \Omega \)) provides the required match. Because most of the termination conductance is provided by the adjustable terminator, variations in the charge pump’s output impedance cause only small changes in the overall termination impedance. In effect, the adjustable termination resistor’s large conductance linearizes the Tx’s impedance match to the channel.

**GRS Rx**

Figure 6 shows the schematic for the GRS Rx, which consists of an input amplifier followed by two subsequent CMOS inverter stages. The Rx input amplifier performs several functions: 1) level-converts the input signal from symmetric-about-ground to symmetric-about-the-inverter switching threshold, nominally \( V_{DD}/2 \) 2) provides modest voltage gain to reduce the input-referred offset and noise of the Rx 3) includes high-frequency signal peaking to overcome its own output pole and help equalize the channel 4) converts the SE input signal into a pseudo-differential output.
The subsequent CMOS inverter stage provides most of the voltage gain (~5×), while the final inverter stage acts as a limiting amplifier by providing sufficient gain to amplify the signal to CMOS levels. We use the same Rx for the clock and data paths. For the forwarded clock lane, we use the full-swing differential output signals to drive the root of the Rx clock network (RxClik).

The GRS input amplifier is based on the common-gate (CG) topology because this configuration does not suffer from bandwidth limitations associated with the Miller effect and allows for very low input CM levels. In the link reported here, the amplifier creates a complementary signal at the output of the right-hand replica branch to realize a pseudo-differential output signal. The input stage consists of two amplifier branches biased using reference current mirrors with cross-coupled source connections. The branch on the left behaves as a CG amplifier, while the branch on the right operates as a common-source (CS) amplifier, providing the necessary signal inversion for pseudo-differential output. To provide signal peaking, the amplifier load uses adjustable Gm–C active inductors, which can be tuned by changing the form factor of the Gm devices (M6 and M7) in addition to the resistance–capacitance (RC) feedback networks.

A small cross-coupled PMOS pair (M8 and M9), in parallel with the active inductors, adds a small amount of positive feedback to implement a negative equivalent resistance, thereby improving the effective quality factor. The RC feedback networks at the gates of the diode-connected current mirrors (M2 and M3) form Gm–C active inductors that equalize the current mirror nodes, which reduces the additional delay incurred through the CS amplifier path from 4 to 0.4 ps. To ensure the maximum voltage gain in the second-stage inverters, a CM feedback (CMFB) loop serves the input-stage bias currents so the differential signals at the inverter inputs are centered on the inverter’s switching threshold.

In practice, the high-frequency peaking in the input amplifier is needed to overcome the output pole of the second-stage inverter amplifiers, so most of the channel attenuation is compensated for in the Tx equalizer. Additional Rx equalization could easily be provided at the expense of increased power; however, the sizing and power consumption of the Rx in this design were chosen to reduce overall link energy.

The Rx input amplifier has gone through a progression of designs as described in [5], from a simple resistor-loaded CG amplifier to a complementary CG amplifier to, most recently, the pseudo-differential amplifier described here. The objective of these refinements was to reduce the temperature variation in the Rx’s input offset. The amplifier in Figure 6 provides two tunable offset cancellation mechanisms: 1) the Rbot resistor that approximates the parallel combination of line impedance and terminator resistor and 2) the differential current source that drives the main CG amplifier stage. The CM setting of the mirrored current sources is controlled by a CMFB loop, as described previously. Temperature drift is nearly eliminated by a synergistic combination of the DVS supply regulation, pseudo-differential input stage design, and continuous CMFB in the amplifier. Another key point to make about the Rx is that the input stage maintains the essential topology of an inverter, which is the core circuit of the reference ring oscillator (RO) used in the DVS regulation scheme. This is true even in the input stage, where, in all branches, current flows from the power supply through one PMOS and then one NMOS transistor, except for the relatively small resistors Rterm and Rbot.

Comments on Using the Inverter as a Continuous-Time Amplifier
For many circuit designers, the thought of using an inverter as a continuous-time amplifier is unnerving. However, used properly, this simple circuit is very effective. An important consideration when using an inverter as a continuous-time amplifier is that its input must be biased at the switching threshold. This is the point on the voltage transfer characteristic where there is the best chance to maintain a symmetrical output swing, which is why we include the CMFB loop shown in Figure 6.

The other consideration is leveling inverter performance across PVT variations. Most continuous-time amplifiers are biased from a reference current source to provide controlled behavior, but that is not an option for an inverter that must produce a high-bandwidth and rail-to-rail output swing. Therefore, we level the bandwidth and current consumption of the inverter stages through a DVS power supply regulation scheme, as presented in a following section. By adapting the power supply voltage and servoing the CM of the input signal, we create an operating point that enables the inverter to efficiently provide high gain and bandwidth. When an inverter is used in this manner, it is actually a current-recycling, coupled-complementary CS amplifier, which allows one to obtain ~2× the transconductance (in 16-nm FinFET CMOS) using the same current, because the NMOS and PMOS transistors work together.

Delay-Matched Clock-Forwarded Link Architecture
The overall GRS clocking architecture is shown in Figure 7, where a forwarded clock is transmitted in quadrature with eight bundled data lanes (only one data lane is shown for simplicity). The 12.5-GHz in-phase (ICLK) and quadrature (QCLK) clocks from the phase-locked loop (PLL) are used only at the transmit end of the
link. The data lanes use I_CLK, while the forwarded clock lane uses Q_CLK to achieve a quadrature data-to-clock timing relationship. The I_CLK and Q_CLK distribution buffers are designed to have identical propagation delays and power supply delay sensitivities. To preserve the quadrature data-to-clock timing at the Rx input, all data and clock channels are routed with near identical lengths. This ensures that the channels are delay matched to a tolerance much smaller than one bit period, or UI. At the receive end of the link, the data and clock signals are amplified to CMOS levels by identical amplifiers. The clock Rx amplifier drives the root of the Rx clock distribution buffer (RxClk) to provide sampling clocks at each data Rx lane. In the data lanes, adjustable full-swing CMOS delay elements are inserted between Rx amplifiers and samplers with an insertion delay (δ) designed to match that of the Rx clock distribution buffer signal (RxClk) at the midpoint of the adjustment range. Then, on a lane-by-lane basis, each tunable delay element is adjusted by trimming the internal fan-out of the cascaded CMOS stages, so that the insertion delay (δ) tracks that of the Rx clock distribution buffer. Duty factor is trimmed by adjusting the relative P-to-N transistor sizes in the delay stages. The tunable delay line provides ±10 ps of delay adjustment in 1.5-ps steps to each lane, which may be used to compensate for ±1.5-mm trace-length mismatch between each data lane and the forwarded clock lane. The delay range and step size do not vary significantly with PVT, thanks to the DVS supply regulation scheme.

The use of near-identical circuitry for transmitting and receiving both the forwarded clock and data signals, in addition to delay-length matching of all channels, ensures that the data and clock insertion delays are matched across the entire link and track across variations in PVT. This simple mechanism cancels nearly all sources of jitter, including power-supply-induced jitter, over a frequency range approaching the link bandwidth. Additionally, the jitter requirements of the PLL are greatly relaxed because any timing variations are common between the relative timing of the data and clock signals.

**DVS Regulation Scheme**

Our latest version of the GRS link employs the DVS power supply regulation scheme shown in Figure 8(a), which uses a novel regulator described in detail in [6]. At the expense of regulator losses, this arrangement varies the internal I/O supply voltage to flatten circuit speed across PVT and reduce current consumption variation.
across process corners, as shown in Figure 8(b) and (c), thereby saving considerable power that would otherwise be dissipated when providing for a sufficient design margin across corner cases. For example, at 110°C, we see only ±4.1% variation in the RO current when compared to the typical–typical process corner, which tracks across nearly all circuit blocks in the GRS link. The reduction in the current consumption variation across process corners also aids in satisfying electromigration requirements, which are usually the most challenging for the fast–fast process corner when subject to maximum voltage and temperature. What’s interesting is that this regulation scheme creates a scenario where the most challenging case is the slow–slow process corner because, in contrast to situations using a fixed power supply voltage, it consumes slightly more current than other process corners. This is true for the DVS scheme because, at the slow–slow corner, circuits always operate from a higher supply voltage.

In this DVS scheme, we use a PLL to lock a CMOS RO at 25-GHz to a 1.56-GHz reference clock; neither frequency is critical because we forward the clock with the data from the transmit end of the link and use that same clock at the receive end of the link to capture incoming data. The full-rate RO output is divided by two to provide in-phase and quadrature-phase clocks (ICLK/QCLK). The RO control input (Vreg_PLL) is set by a PMOS pass element that regulates the RO supply voltage down from the external supply (Vdd_IO = 0.95 V). The I/O circuitry operates from a second regulated supply (Vreg_IO), whose regulator uses Vreg_PLL as a reference voltage. This voltage is set in the PLL so that a reference CMOS circuit (the RO) operates at a fixed rate independent of PVT. Therefore, the I/O circuitry, which operates on a supply voltage nominally equal to Vreg_PLL, also operates at a fixed rate independent of PVT. This regulation scheme is employed at both ends of the link, where the PLL at the receive end is used solely to set the I/O supply voltage, since the PLL’s generated ICLK and QCLK signals are not used in the Rx. Also, since GRS signals about ground, there is no need for each end of the link to use the same voltage for Vreg_IO.

**Link Calibration**

An important design objective was eliminating the need for periodic link recalibration to compensate for temperature and power supply voltage variation. To make this work correctly, it must be possible to perform one, and only one, calibration at the beginning of operation, including signal offset cancellation, clock delay matching, and duty factor cancellation. Subsequently, the operating parameters must track across supply voltage and temperature variation with sufficient accuracy to maintain link margin. The PLL-referenced power supply regulation scheme described earlier, along with the described circuit design techniques used in the Tx and Rx, supports this calibrate-once policy.

Most calibration steps make use of a special operational mode in the Rx data samplers, in which the Rx sampling clock (RxClk) is swapped with a locally generated random clock that is uncorrelated with the data. A data pattern with an established ratio of ones and zeroes is transmitted to the Rx, sampled by the random clock, and accumulated in counters. Finally, the ratio of ones to zeroes is measured, and the results are used to drive the calibration loop. To remove the Rx input offset, for example, a continuous duty-factor-free 11001100 pattern is transmitted, the randomly sampled data are accumulated, and the calibration loop drives the Rx offset tuning mechanism to obtain a one-to-one ratio of ones to zeroes. Next, a repeating 1010 pattern is transmitted, and the calibration loop drives the duty-factor offset tuning for the received clock signal. A similar technique could be used to tune equalizer settings, although in very short-range links with small channel attenuation, the equalization (EQ) adjustment may be determined and set at design time.

**Fast Entry/Exit Pause Mode**

While energy-efficient operation is a requirement for a low-power link, it is not sufficient in cases where the traffic carried by the link is bursty. To deal with this problem, the link enters a pause mode whenever transmission of data is not needed. In pause, the circuitry consumes very little power while providing very fast entry and exit times between pause and active. To enter the pause mode, the forwarded clock is held Hi after transmission of the last data bit and remains in this state for the duration of the pause event. The internal ICLK signal stops toggling and is held Hi, which drops the clock distribution and data Tx power to zero, except for the forwarded clock circuitry, which must continue to drive its line HI. Because the forwarded clock is held HI, the recovered clock signal at the receive end of the link (RxClk) stops toggling, which reduces receive-side clock distribution power to zero.

A simple analog CMOS circuit, with asymmetrically skewed pull-up and pull-down strengths, detects that RxClk has paused and signals all current-consuming circuits in each Rx to shut down. The remaining active circuitry includes the PLLs at both ends of the link, which keep the Qclk at the Tx toggling and both regulators operating to maintain the internal Vreg_IO voltages as well as the forwarded-clock Tx to hold the forwarded clock signal HI. To exit the pause mode, the forwarded clock begins toggling, which causes RxClk to resume at the receive end of the link, and the CMOS pause-mode detector powers up the Rx circuitry. After a small and programmable number of parallel clock times, live/active data transmission can resume. When paused, this mechanism reduces link power by 75%, while entry and exit times are held to fewer than 5 ns. Figure 9 illustrates the signal operation of the fast entry/exit pause mode.
Crosstalk Mitigation and Compact System Floor Plan for Clock-Forwarded Links

It is important to briefly discuss another challenge for densely routed SE signaling at 25 Gb/s, which is crosstalk. To minimize the effects of crosstalk, we route all signals as striplines between ground planes in the package and the PCB. Theoretically, this eliminates far-end crosstalk (FEXT) and greatly reduces near-end crosstalk (NEXT), and it is a requirement for any densely routed SE link operating at 25 Gb/s. In addition, we use a checkerboard pattern of signals and ground returns for the vertical transitions of the C4 bumps, blind/buried vias in the package, and plated-through-hole (PTH) vias within the package core and PCB. This helps to mitigate crosstalk by creating a well-defined and low-impedance return path through the ground network. Figure 10 illustrates the floorplan of the most recent GRS demonstration vehicle, which was fabricated in 16-nm FinFET CMOS. The C4 bump pitch used was 150 μm, and the bumps have been labeled and color-coded as either S = signal, C = clock, G = ground, or P = power. For systems where the routing is limited to on-package, it is possible to remove some of the ground returns and reduce the area requirement, but this greatly reduces timing margin. However, for systems that include package and PCB routing, this checkerboard pattern is critical; without it, link reliability will be compromised due to insufficient timing margin.

A compact floor plan is preferred because all lanes use a common timing reference (i.e., the forwarded clock). This also helps to minimize clock distribution power [7], while limiting the skew below 500 fs. Length-matched routing of the RDL (270 μm) adds 60 fF of capacitance to the I/O of each lane. Ultimately, the total area used depends on the minimum C4 bump pitch, which is set by the thermomechanical limits of a reticle-limited die attached to a particular packaging technology as well as the number of ground return bumps needed to mitigate crosstalk. Any remaining area is used for power supply bypass capacitance.

Experimental Results

The latest GRS link has been demonstrated in two applications: 1) a very-short-reach link operating over a high-density interconnect (HDI) between chips mounted on the same organic package substrate and 2) a short-reach link between packaged chips operating over a short PCB channel. Before presenting the experimental results, we briefly discuss the physical configuration of each system. For the on-package system, shown in Figure 11(a), the chips are mounted on a conventional organic package with a 4-2-4 metal stack-up. The on-package signal routes are restricted to the top-side build-up HDI layers so that blind/buried vias may be used for all vertical connections, thereby eliminating via stubs. The channels are routed as striplines in the second layer of HDI, with ground planes above and below the signal traces to reduce crosstalk between adjacent lanes, and the trace lengths of all lanes are matched to within ±150 μm (±1 ps).

Figure 11 also shows the electrical model [Figure 11(b)] and frequency characteristics [Figure 11(c)] of the 10-mm channel, where the frequency response of the entire channel was modeled using a 3D field solver. Production-level ESD clamps are included at both ends of the link to ensure reliability. The extracted capacitances at either end of the link are shown in Figure 11(f), which includes the contributions of the I/O pads, ESD protection devices, and circuit parasitics. T-coils are used to compensate for these capacitance contributions at each end, thereby improving back-match and minimizing intersymbol interference. At Nyquist, the channel has ~4 dB of insertion loss (IL), and the power-sum FEXT (PSFEXT) from all eight lanes onto the clock is held at 38.5 dB below IL.

The use of T-coils is beneficial for short-reach links because the channel attenuation is so low. If the link did not include T-coils, the reflections due to the ~500 fF of capacitance would degrade the link margin. To illustrate this, simulated eye diagrams are plotted in Figure 11 for the on-package channel model without [Figure 11(d)] and with [Figure 11(e)] T-coils. Figure 11(f) shows the measured bit error rate (BER) bathtub curve for simultaneous 25-Gb/s operation of all lanes in the 10-mm package link, with unique seeds of a pseudo-random bit sequence.
(PRBS-31) data pattern applied to each lane. An on-chip phase interpolator (included only for characterization purposes) provides 1.33-ps steps for the BER bathtub plot. At a BER of $10^{-15}$, the measured aggregate eye opening is 30.7 ps (0.77 UI) with +4.6 dB of Tx EQ.

Figure 12(a) shows the physical arrangement of the package-to-package link, where two chips were mounted on conventional organic packages to signal across a conventional low-cost PCB. The signal routes in the PCB were restricted to the bottom-most stripline routing layer to minimize stub lengths associated with the PTH vias required for vertical transitions within the PCB. To further reduce crosstalk, grounded shields were inserted between adjacent signal routes on the PCB. As shown in Figure 12(b), the channel comprises 13 mm of on-package HDI traces at both ends of the link, vertical paths through the package and PCB vias, and 54-mm stripline routes on the bottom of the PCB. In this channel, the delays of the clock lane and data lanes are matched to within approximately ±0.6 mm in length (±4 ps).

The entire interconnect, including package and PCB traces, bump and ball connections, and all vias, was modeled using a 3D field solver to produce the channel frequency response in Figure 12(c). Crosstalk is shown as the power-sum of the FEXT signals from all eight aggressors. The channel has –8.5 dB of IL at Nyquist, while the PSFEXT is kept 23.1 dB below IL at Nyquist. The use of T-coils is critical for short-reach links that route through multiple impedance discontinuities; without them, the reflections due to the ~500 fF of capacitance and impedance discontinuities would make the link inoperable. To illustrate this, simulated eye diagrams are plotted in Figure 12 for the PCB channel model without [Figure 12(d)] and with [Figure 12(e)] T-coils. Figure 12(f) shows the measured BER bathtub curves for simultaneous operation of all lanes at 25 Gb/s. The data pattern is PRBS-31 with different seeds for each data lane. The measured aggregate eye opening at BER = $10^{-15}$ is 16.6 ps (0.42 UI) with +5.8 dB of transmit EQ.

The link margin is measured across temperature variation to verify that one-time calibration under arbitrary conditions of supply voltage and temperature is sufficient to guarantee subsequent link timing margin under
voltage and temperature variation. Figure 13(a) shows the bathtub curves for the 10-mm on-package link operating at temperature extremes of 10°C and 110°C, after calibration at 10°C; Figure 13(b) shows the bathtub curves for the same temperature extremes after calibration at 110°C. Data patterns were PRBS-31 with different seeds on each lane: after observing low temperature sensitivity, the measurements were concluded at BER = 10⁻¹⁰ for brevity. The bathtub plots represent the aggregate of all eight data lanes.

Figure 14 shows the measured BER bathtub curves for the on-package link and the PCB link without the additional parallel Tx termination used to match the channel impedance. For the case of the on-package link, there was no measurable change in the timing margin across all eight data lanes to a BER = 10⁻¹³. Simulations support this and also indicate that the 1.33-ps step size in our phase-interpolator, used for margining, is too coarse to resolve any notable differences in timing margin. This is due to the exceptional signal integrity of the on-package channel, which has no discontinuities. In simple terms, we have made this channel trivial to communicate over. However, when the additional parallel Tx termination is not used in the PCB link, there is an increase in reflections and crosstalk that degrades the timing margin of all lanes. Four of the eight PCB lanes are more sensitive to the absence of a good back-match, due to differences in routing and shielding, and thus fail before a BER = 10⁻³ can be achieved. Both the on-package and PCB link data patterns were PRBS-31 with different seeds on each lane.

Figure 15 summarizes the per-bit energy breakdown for the GRS link operating at 25 Gb/s/pin across the PCB channel. For the entire link, an energy efficiency of 1.17 pJ/bit is achieved for eight data lanes and one clock lane when operating from a 0.95-V external supply. All overhead from on-chip regulators, clock generation and distribution, serialization and deserialization, signal generation and detection, and the movement of low-speed data back to the interface with the core voltage domain is included. The respective energy cost of the voltage drop across the regulator’s PMOS pass element is included in the energy cost for each circuit subblock. Total regulator energy cost (0.031 pJ/bit) accounts for the regulator’s current from the 0.95-V external supply.

**FIGURE 12:** An 80-mm-long PCB link: (a) a system cross section, (b) the electrical model, (c) the frequency response, (d) an eye diagram without T-coils, (e) an eye diagram with T-coils, and (f) the measured BER for all eight data lanes.
supply that does not flow through the PMOS pass-element. For the on-package link, the energy efficiency is slightly better at 1.13 pJ/bit because less Tx equalization is needed.

Going forward, there are several things that could be done to improve link energy efficiency. Three options are as follows:

1) Amortizing the energy-cost of the PLL across a pair of bytes would improve link efficiency to ~1.12 pJ/bit.
2) Moving the DVS regulation scheme off-chip, where a high-efficiency dc–dc converter would be controlled in a manner similar to the on-chip regulator, would provide an estimated efficiency of ~0.86 pJ/bit, when combined with 1).
3) Using resonant clock distribution would provide an estimated efficiency of ~0.59 pJ/bit, when combined with 1) and 2).

**Closing Thoughts**

Eventually, poor signal integrity will prevent moderate-reach differential links from meeting communication needs while leaving sufficient energy for computation. At that time, markedly different interconnect/packaging and signaling solutions will replace many of the energy-inefficient communication methods used today. The short-reach electrical link presented in this article can be used for high-speed interconnects within MCMs, E/O interface chips, and memory interconnects. The bandwidth per-pin achieved by this link is comparable to recent 56-Gb/s experimental differential links but at roughly one-half

---

**FIGURE 13:** The BER bathtub curves demonstrate the temperature resilience of the on-package link calibrated (a) at 10 °C and (b) at 110 °C.

**FIGURE 14:** The BER bathtub curves measured without additional parallel Tx termination: (a) the on-package link is unaffected, and (b) the PCB link has four of eight lanes that fail.
the energy per bit [1], which leaves more energy to support higher system computational performance.

It is important to emphasize that a collection of methods helps to produce an environment where densely routed SE signaling is viable for 25 Gb/s and beyond. These methods span several domains and create a spiral-of-goodness, where each method complements or supports another. By doing so, GRS provides the required bandwidth density and energy efficiency to support computational scaling for the high-performance computing systems of the future.

Acknowledgments

We would like to thank Brian Zimmer, Xi Chen, Sudhir Kudva, Sanquan Song, Stephen Tell, Nikola Nedovic, Wenxu Zhao, Sunil Sudhakaran, Trey Greer, Tom Gray, and Bill Dally for their design expertise. In addition, this article would not have been possible without the assistance of the signal integrity, package design, and PCB design teams at NVIDIA.

References

About the Authors

John M. Wilson (jowilson@nvidia.com) received his B.S., M.S., and Ph.D. degrees in electrical engineering from North Carolina State University (NCSU), Raleigh. He is a senior research scientist with the Circuits Research Group, NVIDIA Inc., Durham, North Carolina. From 2003 to 2006, he was a research professor with the Department of Computer Science, UNCHC, where he was a principal contributor to the design and construction of the PixelPlanes and PixelFlow computer graphics systems and designed custom beamforming chips for the first commercial 3D medical ultrasound scanner. From 1999 to 2003, he was chief engineer with Velio Communications, and, from 2003 to 2009, he was a technical director with Rambus Inc., Chapel Hill, North Carolina. He is currently a senior distinguished research scientist with NVIDIA Inc., Durham, North Carolina, where he is working on low-energy on- and off-chip signaling.

Walker J. Turner (wturner@nvidia.com) received his B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Florida, Gainesville, in 2009, 2012, and 2015, respectively. In 2014, he was a contractor with the U.S. Army Research Laboratory, Adelphi, Maryland, developing wirelessly powered systems and integrated low-noise amplifiers for piezoelectric E-field sensors. Since 2015, he has been a senior research scientist with the Circuits Research Group, NVIDIA Inc., Durham, North Carolina. His current research interests include low-power integrated circuit design for mixed-signal and high-speed signaling applications. He is a Member of the IEEE.