Low-Power Techniques for Wireline Systems

Behzad Razavi
Electrical and Computer Engineering Department
University of California, Los Angeles
razavi@ee.ucla.edu

***Abstract*—As wireline transceivers continue to push for higher speeds, their power consumption becomes increasingly more critical. This paper describes methods that can reduce the power drawn by drivers and multiplexers on the transmit side and equalizers and clock and data recovery circuits on the receive side. These concepts have been incorporated in designs ranging from 40 Gb/s to 80 Gb/s in 45-nm and 28-nm process nodes.**

I. Introduction

150x in 30 years

This is how fast the data rate in wireline systems has climbed. And yet the users' appetite for greater speeds remains unabated. From data centers to mobile devices, the general design tenet is that "fast is never fast enough." Several questions naturally present themselves. (1) How has the power consumption of serial links scaled? (2) Should we be concerned with the power normalized to the bit rate ("power efficiency") or the absolute power per link? (3) How can we ease the power-speed trade-offs? In answering these questions, we begin to see the daunting challenges that lie ahead. A great deal of work has advanced the state of the art in this domain [1]-[25].

This paper presents a number of circuit and architecture techniques that have been introduced for lowering the power consumption of wireline transceivers, leading to transmitters (TXs) operating at 40 to 80 Gb/s and drawing 32 mW to 44 mW [18], [17] and receivers (RXs) running at 40 Gb/s to 56 Gb/s and consuming 14 mW to 50 mW [13], [5]. Trade-offs between speed and power suggest that methods of improving the former can be exploited to reduce the latter.

Section II provides a high-level view of serial links and the trends governing them. Section III proposes methods of benchmarking CMOS technology nodes for wirelines systems. Sections IV and V delve into low-power TX and RX design, respectively.

II. General Considerations

In order to appreciate the power consumption problem, we consider a typical network hierarchy, illustrated in Fig. 1. In the most granular layer, "extra-short-reach" (XSR) links permit die-to-die communication over low-loss media. At the next level, "medium-reach" (MR) links provide chip-to-chip data transport, dealing with a moderate amount of loss, e.g., 5 to 10 dB. In the third layer, "long-reach" (LR) links connect boards or racks and face losses as high as 30 to 40 dB.

Three points emerge here. First, the number of XSR links in the lowest layer may be one or two orders of magnitude larger than that of the highest layer, demanding commensurately lower power consumption per XSR link. Second, as we aggregate the power consumptions of the dice, the chips, and the boards, we recognize that it is the absolute power - rather than the efficiency - that determines the cost of packaging, thermal management, and heat removal. Third, the high losses in the LR media play a central role in their choice of signaling and transceiver architectures.

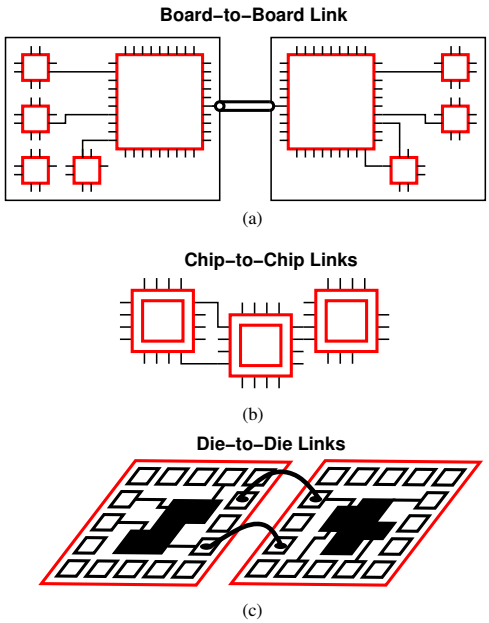


Fig. 1. Typical network hierachy: (a) board-to-board link, (b) chip-to-chip links, and (c) die-to-die links.

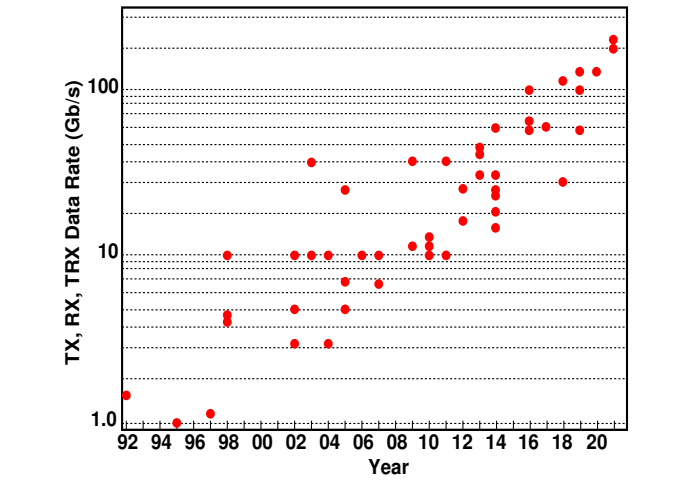
Figure 2(a) plots the speed, r_b, as a function of time, revealing a dramatic rise. Figure 2(b) suggests that the power efficiency, P/r_b, has also improved considerably, but we should remark that this figure of merit appears to have bottomed out in the past five years.

III. Technology Benchmarking

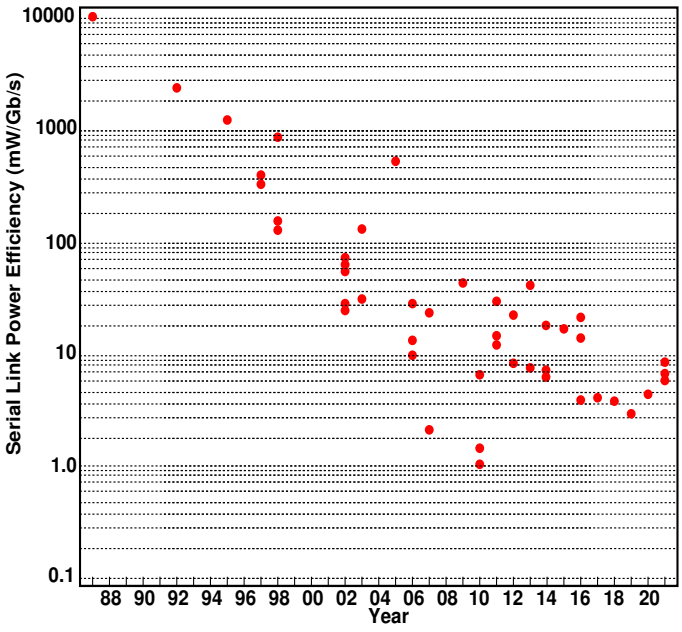
In order to quantify the power-speed trade-offs in wireline transceiver design, we wish to develop simple, easily reproducible benchmarks. Unfortunately, these attributes are not reflected in such technology figures of merit as f_T, f_{max}, and gate delay.

When optimizing a complex transceiver for power, we must observe that, for a given function and a given logic style, the power-speed trade-off is linear up to some frequency and nonlinear beyond. Conceptually illustrated in Fig. 3, this behavior manifests itself in CMOS stages as the data rate and/or the

978-1-6654-3751-6/21/$31.00 ©2021 IEEE



(a)



(b)

Fig. 2. (a) Speed, and (b) power efficiency of serial links vs. time.

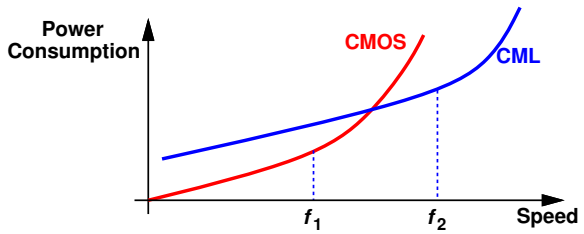


Fig. 3. Conceptual power-speed trade-offs for CMOS and CML realizations of a given function.

clock frequency reach f_1, above which the transistors must be chosen excessively wide so as to improve the speed by a small amount. Current-mode logic (CML), on the other hand, experiences such diminishing returns at a higher frequency, f_2. (We assume that CML stages are custom-designed for each frequency.) These trends indicate that the implementation of the TX and RX blocks and the transition from CMOS logic to CML must be selected according to f_1 and f_2.

The general behavior depicted in Fig. 3 also implies that the data rate per lane should be increased up to f_1 or f_2 before the *number* of lanes is raised.

A simple technology benchmark that proves useful in wireline design is the power-speed trade-off of $\div 2$ circuits. Plotted in Fig. 4 are such trade-offs for TSPC [26] and CML

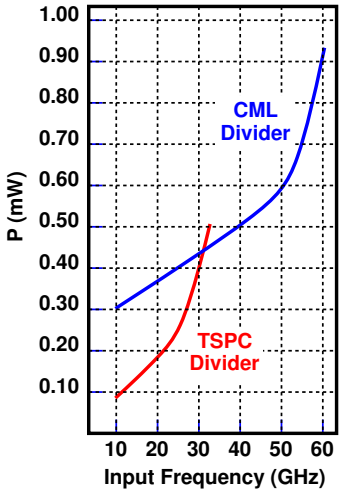


Fig. 4. Power-speed trade-offs for TSPC and CML $\div 2$ circuits.

realizations as predicted by simulations in 28-nm technology. We observe that the former follows a linear regime up to about 20 GHz, beyond which its clocked devices must be made increasingly wider so as to reach higher speeds. The CML structure, on the other hand, traverses a slower growth up to 50 GHz and then exhibits a nonlinear climb.

Another useful benchmark are flipflops (FFs) acting as retimers since they appear in many TX and RX functions. We note that the FFs in the foregoing $\div 2$ stages sense waveforms in their data paths having a rate equal to half of the clock frequency. We thus expect flipflops to follow the trends in Fig. 4, except that f_1 and f_2 are roughly halved.

In addition to CMOS and CML circuit styles, one can also consider "charge steering," a design paradigm that offers greater speeds than the former and lower power than the latter. Illustrated in Fig. 5, the basic latch structure consists of a

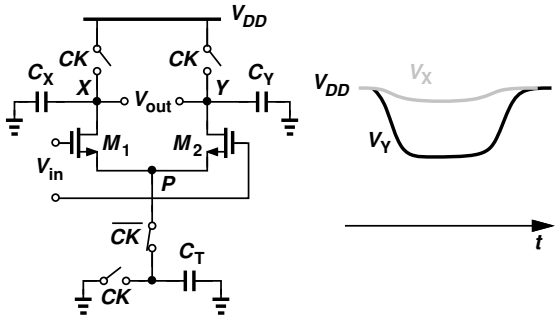


Fig. 5. Basic charge-steering stage.

differential pair, a tail capacitor, C_T, parasitic capacitances, C_X and C_Y, and some switches [14]. In the precharge mode, nodes X and Y are reset to V_{DD} and C_T is tied to the ground. In the evaluation mode, C_T is switched to node P, pulling current from C_X and C_Y. A differential voltage thus develops at the output until C_T charges up and the currents cease. That is, the circuit automatically stops the current flow, a point of contrast to integrating stages. By virtue of moderate swings, about 400 mV, charge-steering stages run faster than rail-to-rail CMOS logic. Moreover, for a given data rate, their power consumption is roughly a factor of 1.4π lower than that of CML. One drawback is that $V_X - V_Y$ exhibits a return-to-zero (RZ) behavior, requiring that the circuits be properly architected [14]. Charge steering has been applied to various TX and RX functions from 25 Gb/s to 80 Gb/s in 65-nm, 45-nm, and 28-nm nodes [14], [18], [17], [5].

IV. Transmitter Design

We begin our TX studies by describing the design environment. Shown in Fig. 6(a) is a generic transmitter architecture

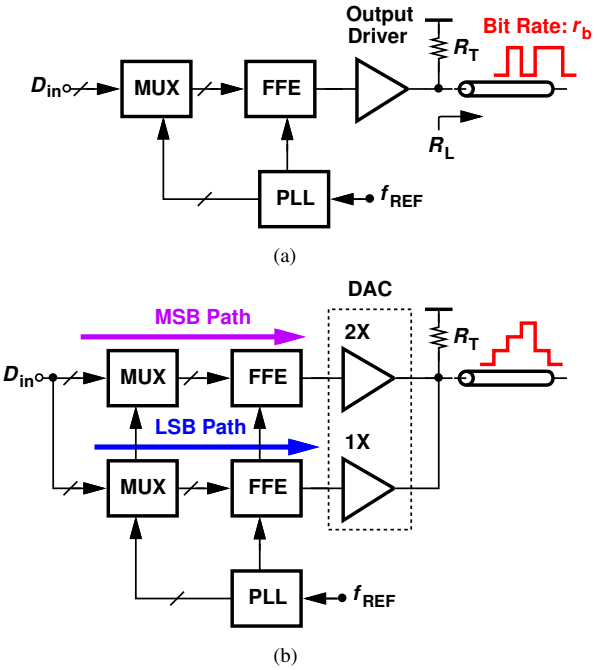


Fig. 6. Generic (a) NRZ, and (b) PAM4 transmitter architectures.

for non-return-to-zero (NRZ) data. Its signal path consists of a multiplexer (MUX), a feedforward equalizer (FFE), and an output driver. The clocks necessary for the MUX and the FFE are provided by a phase-locked loop (PLL).

The situation is more complex for PAM4 transmitters; as depicted in Fig. 6(b), the MSB and LSB are generated by two MUX chains and subsequently delivered to a 2-bit digital-to-analog converter (DAC). The MUX power therefore becomes more problematic here.

A. Driver Design

The line drivers in wireline transmitters require a certain minimum power consumption as they must launch a reasonable voltage swing across a 100-Ω differential characteristic impedance. For a single-ended peak-to-peak amplitude of 0.5 V, a current-mode topology must deliver 20 mA to a 50-Ω back-termination resistor and a 50-Ω line. The circuit therefore employs wide transistors ($W \approx 40$ μm in 28-nm technology), exhibiting a large input (and output) capacitance.

We should remark that, while the relatively unscalable 20-mA current implies a power consumption of 20 mW with $V_{DD} = 1$ V, high-speed transmitters typically draw more than 150 mW [21], [22], [23]. Thus, the other TX building blocks are equally power-hungry.

An effective approach to saving power in transmitters is to avoid nodes carrying full-rate data—except for the final output, of course. This is possible if the driver acts as a multiplexer as well. Figure 7 illustrates this trend. Considering

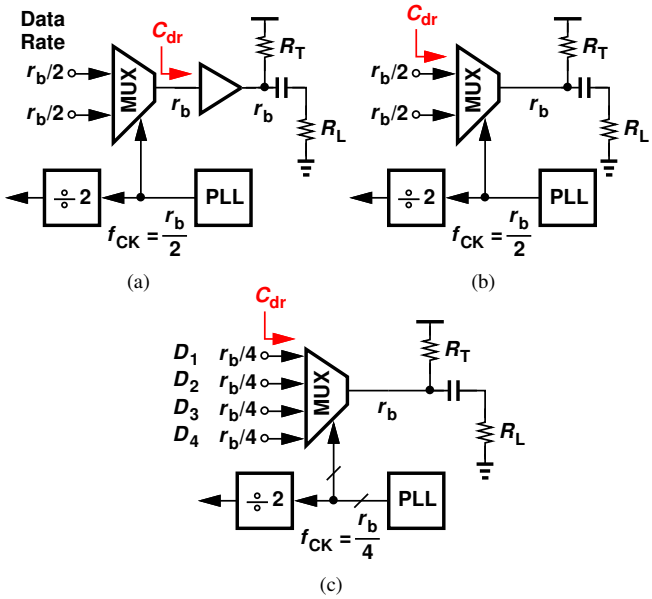


Fig. 7. TX with (a) separate MUX and driver stages, (b) a merged MUX and driver, and (c) a direct 4-to-1 MUX/driver.

the conventional chain shown in Fig. 7(a), we note that the last MUX generates full-rate data. In this power-hungry solution, the MUX proves particularly challenging as it must interface with the large input capacitance of the driver, C_{dr}. The $\div 2$ stage and the voltage-controlled oscillator (VCO) within the PLL operate at half rate. The MUX mismatches must be managed to ensure a small deterministic jitter at the output.

The next natural step is to merge the driver with the MUX [27], as in Fig. 7(b), recognizing that this TX contains only one full-rate port. The MUX incorporates transistors as wide as those in the driver of Fig. 7(a), thereby presenting a capacitance equal to C_{dr} at each of its inputs. Nonetheless, driving two ports at $r_b/2$ is less power-hungry than one port at r_b if the former falls in the linear regime of Fig. 3 and the latter, in the nonlinear regime.

The notion of utilizing the MUX as the driver can be extended to greater multiplexing factors. Shown in Fig. 7(c) is an example employing a direct 4-to-1 MUX and a maximum

clock frequency of $f_{CK} = r_b/4$. Here, four input capacitances equal to C_{dr} are driven at rate of $r_b/4$. We observe that the choice of the multiplexing factor, n, depends on f_1 and f_2 in Fig. 3. In principle, n should be increased to the point where the stages preceding the MUX enter the linear power regime. However, this trend also increases the capacitance at the *output* of the MUX proportionally. Thus, $n = 4$ generally provides a reasonable compromise.

In the architecture of Fig. 7(c), the 4-to-1 MUX requires four clock phases. We can generate overlapping quadrature clocks and perform logical operations within the MUX so as to create a nonoverlapping effect. Depicted in Fig. 8(a) is a

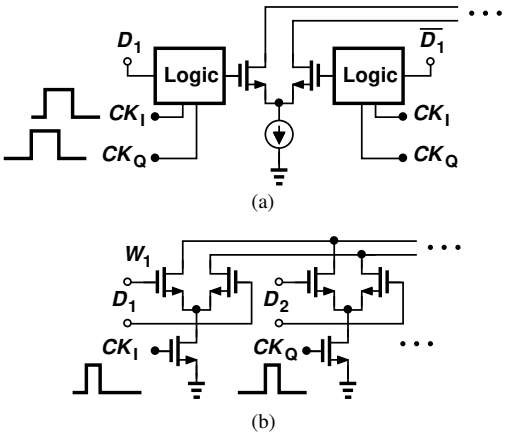


Fig. 8. Direct 4-to-1 MUX with (a) overlapping quadrature clocks, and (b) nonoverlapping quadrature clocks.

CML example [15] in 10-nm technology, where D_1, CK_I, and CK_Q are processed and then applied to the differential pair. This approach is well-suited to advanced nodes. For slower processes or higher speeds, we can opt for the design shown in Fig. 8(b), with single tail transistors driven by nonoverlapping clocks [18]. Avoiding logical operations in the data path, this method relies on direct generation of nonoverlapping quadrature phases by a particular $\div 2$ circuit topology [17].

The issue of phase mismatches in the direct 4-to-1 MUXes of Fig. 8 has been studied in [17]. A useful byproduct of the large transistors necessary here is that they exhibit small mismatches. Shown in Fig. 9 is the output spectrum of a 40-Gb/s NRZ (80-Gb/s PAM4) TX whose input data pattern is so chosen as to deliver a periodic output at 20 GHz. The mismatches introduce spurs at 10 GHz and 30 GHz; with a level of about -40 dBc, they translate to an rms phase mismatch of approximately 100 fs.

The foregoing methods can also be applied to voltage-mode circuits, also called series-source termination (SST) drivers [28]. Shown in Fig. 10(a), an SST stage provides back termination by ensuring that the on-resistance of M_1 and M_2 is equal to the desired value, e.g., $R_T = 50\ \Omega = R_L$.[1] From the differential equivalent circuit in Fig. 10(b), we observe

[1]One can alternatively select the on-resistance to be small and place a resistor in series with the drain of each transistor, but at the cost of a much higher gate capacitance.

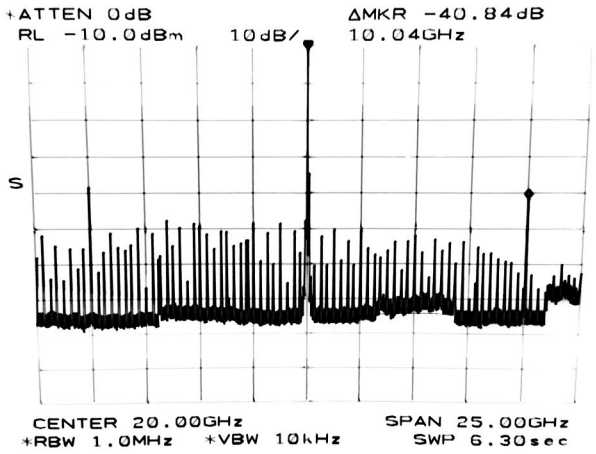


Fig. 9. Output spectrum of a TX with periodic data showing spurs due to MUX mismatches.

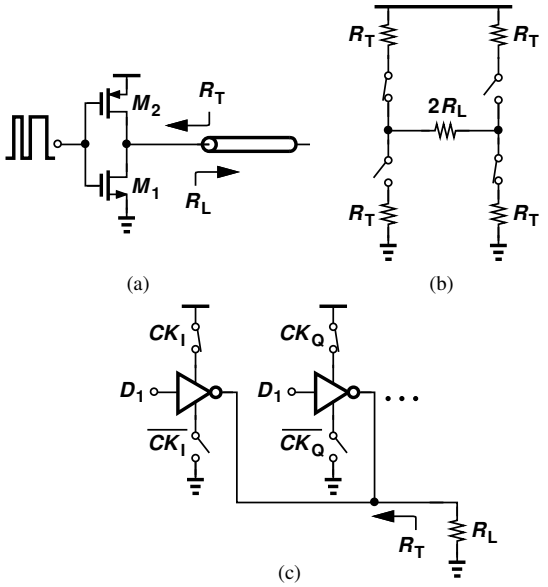


Fig. 10. (a) Basic voltage-mode driver, (b) equivalent differential model, and (c) extension to MUX/driver combination.

that SST benefits from class-D operation, drawing an average power of $V_{DD}^2/(2R_T{+}2R_L)$. This amount is one-fourth of that of current-mode implementations for a single-ended output swing of $V_{DD}/2$. Since M_1 and M_2 receive rail-to-rail inputs for complete switching, their total width is typically around half of W_1 in Fig. 8(a). As explained below, this advantage diminishes if SST embraces multiplexing.

In the spirit of Fig. 8(b), Fig. 10(c) depicts how SST stages can form a MUX as well. Unfortunately, however, the inverter and clocked transistors in this arrangement must be twice as wide as the corresponding devices in Fig. 10(a) so as to allow proper back termination. Requiring rail-to-rail gate voltage swings, the transistors therefore demand a high power in the *preceding* stages.

B. MUX Design

The MUX chain in a wireline transmitter typically incorporates a large number of latches and selectors as it must serialize data from hundreds of megabits per second to tens of gigabits per second. This issue proves particularly serious in the two-path PAM4 system of Fig. 6(b), which would demand hundreds of such stages. The very large number of transistors and the complex signal and clock distribution networks draw substantial power.

With the direct 4-to-1 MUX of Fig. 7(d) in mind, we must seek a serializer that generates D_1-D_4 from low-speed data. For example, we wish to sense 128 inputs at 312 Mb/s and produce four outputs at 20 Gb/s. The PAM4 TX of Fig. 6(b) requires two such serializers.

In its simplest form, a 2-to-1 MUX cell need only employ one latch and one selector [Fig. 11(a)], provided that the transitions of D_1 and D_2 are reasonably aligned. The latch ensures that the two selector inputs do not change simultaneously, a condition necessary for avoiding glitches in D_{out}. Suppose the PAM4 TX of Fig. 6(b) must serialize the data from, say, 312 Mb/s to 20 Gb/s in each path. Then, two binary trees of 2-to-1 MUXes and hence a total of 256 latches and selectors are required [Fig. 11(b)]. In this architecture, the number of 2-to-1 MUXes drops by a factor of 2 from one rank to the next, but the increase in speed at least doubles the power consumed by the MUX cells.

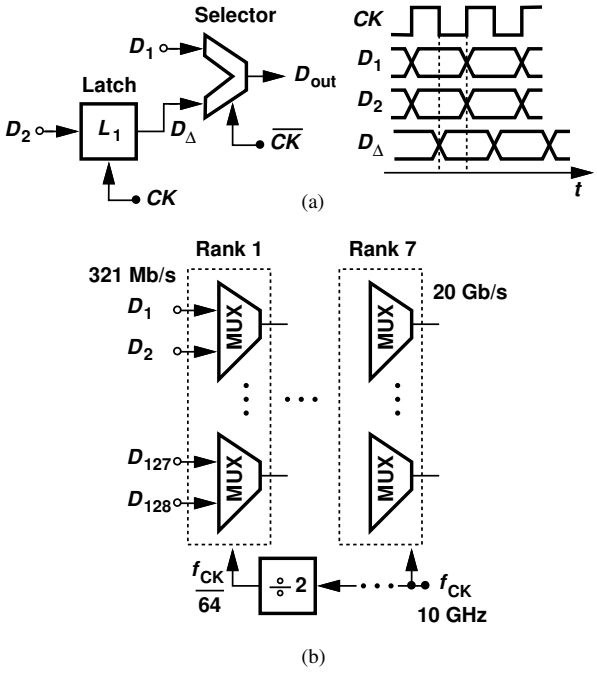


Fig. 11. (a) Basic MUX cell, and (b) binary MUX chain for generating four 20-Gb/s outputs from 128 312-Mb/s inputs.

The foregoing observations along with the trends illustrated in Fig. 3, suggest that, for minimal power consumption, the lower ranks in a serializer should employ CMOS (rail-to-rail) logic and the higher ranks, current-mode logic. The boundary between the two is given by f_1 in Fig. 3. Nevertheless, a 20-Gb/s serializer based on such a methodology would still draw tens of milliwatts in 28-nm technology. We describe two methods that reduce the power.

We begin by noting that the $\div 2$ stages in Fig. 11(b) naturally provide quadrature clocks. With such phases available, it is possible to architect the serializer such that it employs *no latches* [17]. The key point here is to guarantee that the two selector inputs in Fig. 11(a) do not change at the same time. Shown in Fig. 12(a), such a structure drives each two selector cells in a rank by quadrature clock phases, CK_a and CK_b. As a result, D_a changes only on the edges of CK_a and D_b on the edges of CK_b, enabling selector S_a to multiplex its inputs without producing glitches. This three-cell topology can be repeated to form a complete serializer. Note that the clocking action controlling the selectors does not allow device or timing mismatches to accumulate through the ranks. A 128-to-8 serializer prototype using this approach delivers eight outputs at 5 Gb/s with a total power consumption of less than 400 μW [17].

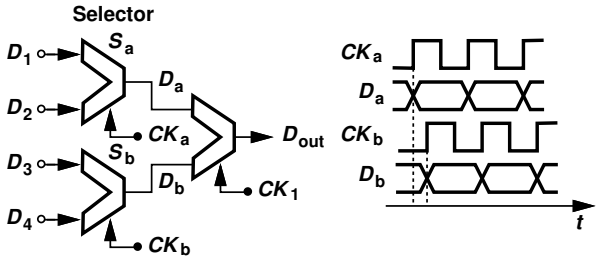


Fig. 12. MUX chain design using no latches.

The second low-power multiplexing method is based on charge steering. Figure 13 depicts a 2-to-1 selector incorporating this idea. When CK is low, nodes X and Y are precharged to V_{DD}, and C_T is discharged to zero. After CK goes high, C_T begins to draw a current from M_1-M_2 or M_3-M_4, discharging the output capacitances and creating a copy of V_{in1} or V_{in2} between X and Y. Figure 14 shows the output eye diagram of a 40-Gb/s NRZ transmitter that employs some of the techniques described above [18]. It includes a PLL and draws 32 mW in 45-nm technology.

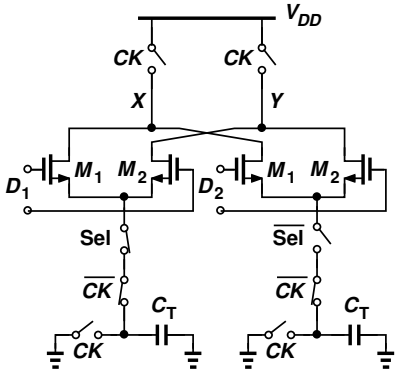


Fig. 13. Charge-steering MUX design.

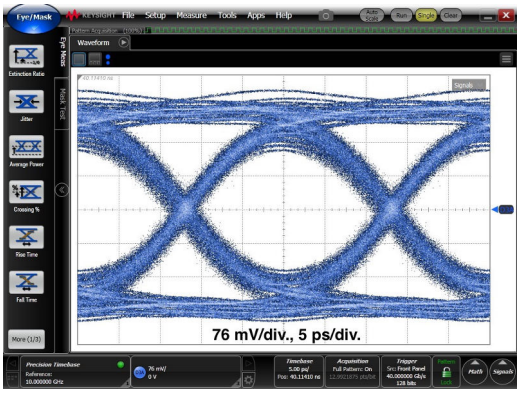


Fig. 14. Output eye diagram of a 40-Gb/s 32-mW TX.

V. Receiver Design

Wireline receivers are architected according to the type of signaling and the loss of the channel. The generally-accepted practice at present is to pursue "analog" implementations for NRZ data and ADC-based solutions for PAM4 signals.

Shown in Fig. 15(a) is a generic analog receiver. It consists

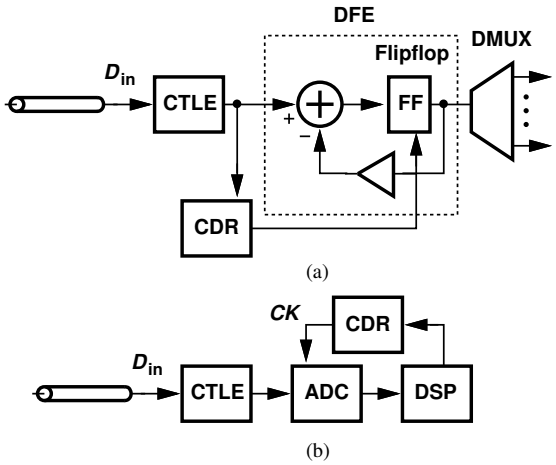


Fig. 15. Generic receiver architectures for (a) NRZ, and (b) PAM4 data.

of a continuous-time linear equalizer (CTLE), a decision-feedback equalizer (DFE), a clock and data recovery (CDR) circuit, and a demultiplexer (DMUX). The CTLE provides boost at high frequencies, partially canceling the loss of the channel, while the DFE compensates for both the loss and the effect of impedance discontinuities. The CDR circuit can sense the CTLE output if the data eye is somewhat open at this port; otherwise, the CDR must be tied to the DFE's summer output. Most of the RX power is consumed by the CTLE, the DFE, and the CDR.

Figure 15(b) depicts a generic ADC-based receiver. A CTLE boosts the high-frequency content of the data, relaxing the resolution required of the ADC. The latter's output is processed by a CDR loop so as to generate the sampling clock, CK [1], [2]. In this case, the ADC and DSP consume most of the power.

A. Linear Equalizers

Linear equalization can be implemented by continuous-time topologies, i.e., CTLEs, or discrete-time circuits, naturally called discrete-time linear equalizers (DTLEs) [13]. We study the former here.

The design of CTLEs is governed by trade-offs among the boost factor, the bandwidth, and the power consumption. With a greater number of stages in the CTLE, the first rises and the last two suffer. Inductive peaking can alleviate the bandwidth-power trade-off, but it cannot provide boosting as its complex zeros would introduce ringing.

With nanometer transistors and supply voltages below 1 V, it becomes difficult to design resistively-loaded differential pairs for a voltage gain of more than 2 or 3. Upon adding RC degeneration to realize boosting [Fig. 16(a)], we observe that the low-frequency gain drops further. In fact, for a gain of about unity per stage, we have

$$g_m R_D = 1 + \frac{g_m R_S}{2}, \tag{1}$$

noting that the left-hand side is the undegenerated gain and the right-hand side the boost factor. The latter is therefore limited to 6-10 dB, and, with channel-length modulation, it falls to about 5 dB per stage. The situation becomes more severe for PAM4 data due to the additional voltage headroom necessary for linearity.

With inductive peaking inevitably present in high-speed CTLEs, it is possible to ease the foregoing trade-offs through the use of feedforward. We first recognize that the boost in a CTLE generally requires a high-pass response and need not be implemented by only RC degeneration. Illustrated in Fig. 16(b) [5], feedforward creates an additional path through G_{mf1} to the inductor, yielding $V_P = G_{mf1}(L_D s)V_{in}$. The response of this path, shown in red in Fig. 16(c), is so chosen as to rise above the original response as the latter begins to fall due to the load capacitance. The overall response then offers both a greater boost factor *and* a wider bandwidth, a point of contrast to the reduced bandwidth in *cascaded* stages.

Figure 16(d) depicts the overall CTLE with three feedforward paths [5]. The performance is quantified by examining the response of the channel (having a loss of 25 dB at 28 GHz) plus the CTLE [Fig. 16(e)]. We observe the relatively flat profile as a result of feedforward. The 56-Gb/s CTLE draws 9 mW in 28-nm technology [5].

B. DFEs

The power-speed trade-offs encountered in DFE design can be ameliorated by a number of techniques. First, up to a certain data rate in a given process node (e.g., 40-Gb/s half-rate operation in 28-nm technology), the charge steering concept described in Section III proves viable.

Second, we can add *high-pass* branches to the DFE so as to sharpen the edges at the summing junction and hence afford a lower power for a given speed. Depicted in Fig. 17 [5], the idea is to generate

$$D_S = (1 + \alpha s)D_{in} - (1 + \beta s)K_1 D_{out}. \tag{2}$$

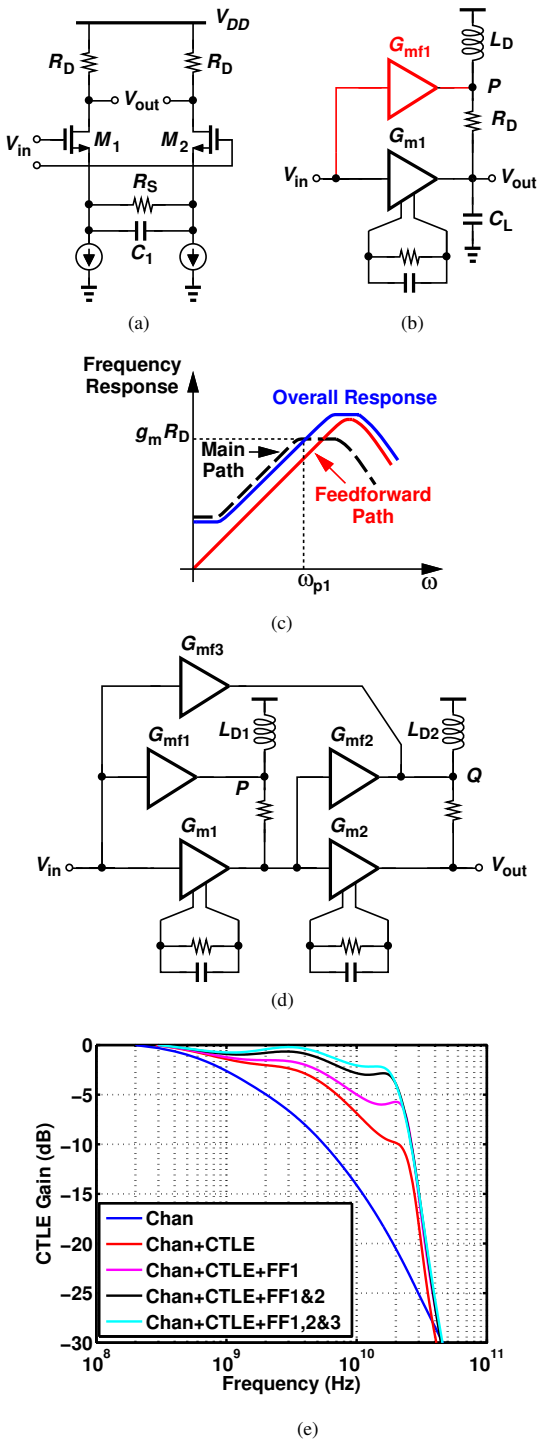


Fig. 16. (a) Basic CTLE stage, (b) use of feedforward around a CTLE, (c) effect of feedforward on the response, (d) complete CTLE, and (e) simulated behavior of the channel/CTLE cascade (FF_j corresponds to G_{mfj}.

As the waveforms suggest, $K_1 D_{out} + \beta dD_{out}/dt$ exhibits faster transitions. The high-pass feedforward branch, αs, creates the same effect for D_{in}. As a result, the edges in D_S become sharper. A half-rate version of this topology operates at 56 Gb/s while drawing 6 mW [5].

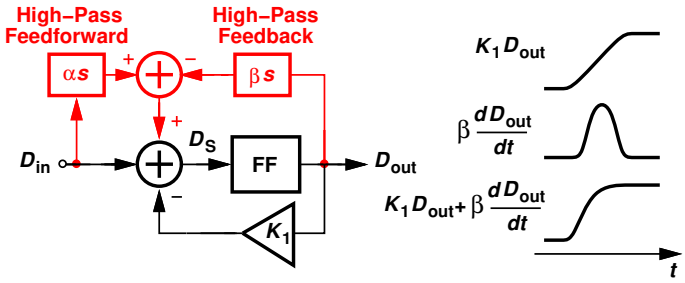


Fig. 17. DFE with high-pass branches.

The foregoing CTLE and DFE techniques have been incorporated in a 56-Gb/s NRZ receiver [5]. Figure 18 shows the measured bathtub curves in two cases: (1) a loss of 25 dB at 28 GHz (channel A), and (2) a loss of 30 dB at 28 GHz (channel B) with an FFE function of the form $-0.2 + 0.8z^{-1}$ applied to the data by the bit error rate tester (to emulate the TX FFE). The horizontal eye openings are 0.4 UI and 0.33 UI, respectively.

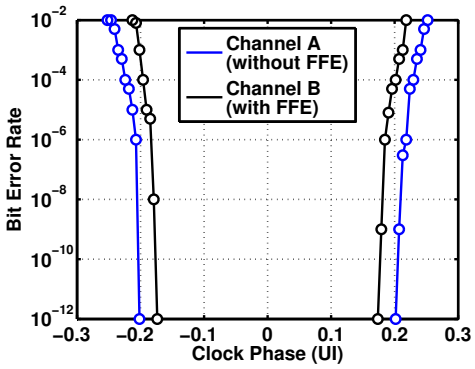
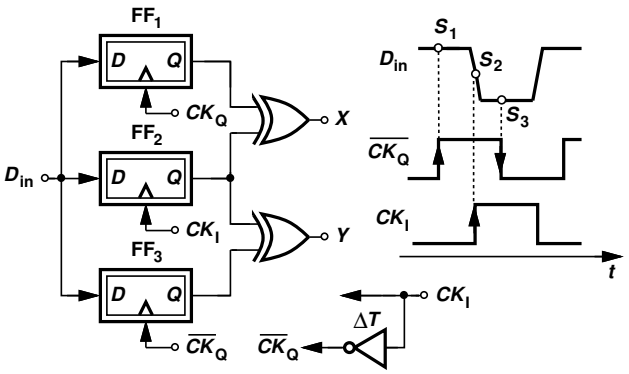


Fig. 18. Measured bathtub curves for a 56-Gb/s NRZ receiver.

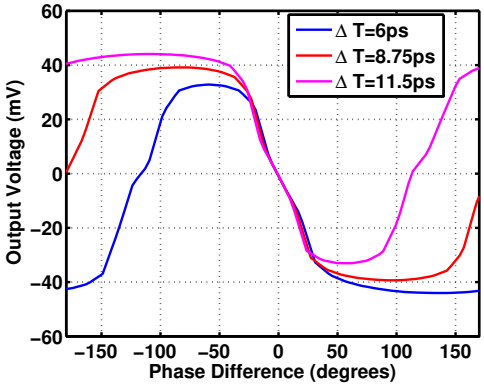
C. CDR Circuits

Power-hungry blocks in CDR circuits include the phase detector, the VCO , and the clock distribution network. The last two make half-rate architectures attractive, but the first would typically require quadrature phases in such a case. We study one method of dealing with this issue.

We turn to a half-rate Alexander PD, shown in Fig. 19. Here, the three flipflops take three consecutive samples of D_{in}, and the XOR gates use these samples to determine whether the clock is early or late. The necessary quadrature phase, CK_Q, can be generated by an inverter if we recognize that PVT variations simply shift samples S_1 and S_3 but not S_2 [5]. As can be seen from the simulated 56-Gb/s PD characteristics in Fig. 19(b), for inverter delay variations as large as $\pm 30\%$ around the nominal value of 8.75 ps, the PD gain remains relatively constant. This PD draws 11 mW at 56 Gb/s [5]. Note that, in the RX architecture of Fig. 15(a), the CDR circuit need not retime the data, as the task is delegated to the DFE. For this reason, the PD shown in Fig. 15(a) is allowed to incur occasional errors so long as they do not affect its gain.



(a)



(b)

Fig. 19. (a) Half-rate Alexander PD using an inverter for quadrature generation, and (b) resulting PD characteristics for inverter delay variations.

VI. Conclusion

A number of circuit and architecture concepts have been described that reduce the power consumption in wireline transceivers. These techniques have led to transmitters operating at 40 to 80 Gb/s and drawing 32 mW to 44 mW and receivers running at 40 Gb/s to 56 Gb/s and consuming 14 mW to 50 mW.

References

[1] P. Upadhyaya et al., "A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET," *ISSCC Dig. Tech. papers,* pp. 108-110, Feb. 2018.

[2] T. Ali et al., "A 180mW 56Gb/s DSP-Based Transceiver for High-Density IOs in Data Center Switches in 7nm FinFET Technology," *ISSCC Dig. Tech. papers,* pp. 118-120, Feb. 2019.

[3] J. Im et al., "A 112Gb/s PAM-4 Long-Reach Wireline Transceiver Using a 36-Way Time-Interleaved SAR-ADC and Inverter-Based RX Analog Front-End in 7nm FinFET," *ISSCC Dig. Tech. papers,* pp. 116-118, Feb. 2020.

[4] T. Ali et al., "A 460mW 112Gbps DSP-Based Transceiver with 38dB Loss Compensation for Next Generation Data Centers in 7nm FinFET technology," *ISSCC Dig. Tech. papers,* pp. 118-120, Feb. 2020.

[5] A. Atharav and B. Razavi, "A 56Gb/s 50mW NRZ Receiver in 28nm CMOS," *ISSCC Dig. Tech. papers,* pp. 192-194, Feb. 2021.

[6] T. Shibasaki *et al.*, "A 56Gb/s NRZ-Electrical 247mW/lane Serial-Link Transceiver in 28nm CMOS," *ISSCC Dig. Tech. papers,* pp. 64-66, Feb. 2016.

[7] P. J. Peng et al, "A 56Gb/s PAM-4/NRZ transceiver in 40nm CMOS," *ISSCC Dig. Tech. papers,* pp. 110-111, Feb. 2017.

[8] B. Dehlaghi et al., "A 1.41-pJ/b 56-Gb/s PAM-4 Receiver Using Enhanced Transition Utilization CDR and Genetic Adaptation Algorithms in 7-nm CMOS," *IEEE Solid-State Circuits Letters,* vol. 2, no. 11, pp. 248-251, Nov. 2019.

[9] J. Han et al, "Design techniques for a 60-Gb/s 288-mW NRZ transceiver with adaptive equalization and baud-rate clock and data recovery in 65-nm CMOS technology," *IEEE J. Solid-State Circuits,* vol. 52, pp. 3474-3485, Dec. 2017.

[10] E. Depaoli *et al.*, "A 4.9pJ/b 16-to-64Gb/s PAM-4 VSR transceiver in 28nm FDSOI CMOS," *ISSCC Dig. Tech. papers,* pp. 112-114, Feb. 2018.

[11] A. Cevrero *et al.*, "A 100Gb/s 1.1pJ/b PAM-4 RX with Dual-Mode 1-Tap PAM-4 / 3-Tap NRZ Speculative DFE in 14nm CMOS FinFET," *ISSCC Dig. Tech. papers,* pp. 112-114, Feb. 2019.

[12] A. Roshan-Zamir *et al.*, "A 56-Gb/s PAM4 Receiver With Low-Overhead Techniques for Threshold and Edge-Based DFE FIR- and IIR-Tap Adaptation in 65-nm CMOS," *IEEE J. Solid-State Circuits,* vol. 54, pp. 672-684, Mar. 2019.

[13] A. Manian and B. Razavi, "A 40-Gb/s 14-mW CMOS Wireline Receiver," *IEEE J. Solid-State Circuits,* vol. 52, pp. 2407-2421, Sep. 2017.

[14] J. W. Jung and B. Razavi, "A 25 Gb/s 5.8 mW CMOS Equalizer," *IEEE J. Solid-State Circuits,* vol. 50, pp. 515-526, Feb. 2015.

[15] J. Kim et al, "A 112-Gb/s PAM4 Transmitter with 3-Tap FFE in 10-nm CMOS," *ISSCC Dig. Tech. papers,* pp. 102-103, Feb. 2018.

[16] M. Hossain, Aurangozeb and N. Nguyen, "DDJ-Adaptive SAR TDC-Based Timing Recovery for Multilevel Signaling," *IEEE Journal of Solid-State Circuits,* vol. 54, no. 10, pp. 2833-2844, Oct. 2019.

[17] Y. Chang et al,"An 80-Gb/s 40-mW Wireline PAM4 Transmitter" *IEEE J. of Solid-State Circuits*, vol. 53, pp. 2214-2226, Aug. 2018.

[18] Y. Chang et al, "A 32-mW 40-Gb/s CMOS NRZ Transmitter," *CICC Proc.*, Apr. 2018.

[19] B.-J. Yoo et al, "A 56Gb/s 7.7mW/Gb/s PAM-4 Wireline Transceiver in 10nm FinFET Using MM-CDR-Based ADC Timing Skew Control and Low-Power DSP with Approximate Multiplier,"*ISSCC Dig. Tech. papers,* pp. 122-123, Feb. 2020.

[20] J. Im et al, "A 112Gb/s PAM-4 Long-Reach Wireline Transceiver Using a 36-Way Time-Interleaved SAR-ADC and Inverter-Based RX Analog Front-End in 7nm FinFET," *ISSCC Dig. Tech. papers,* pp. 116-117, Feb. 2020.

[21] Groen et al, "A 10-to-112Gb/s DSP-DAC-Based Transmitter with 1.2Vppd Output Swing in 7nm FinFET," *ISSCC Dig. Tech. papers,* pp. 120-121, Feb. 2020.

[22] J. Kim et al, "A 224Gb/s DAC-Based PAM-4 Transmitter with 8-Tap FFE in 10nm CMOS," *ISSCC Dig. Tech. papers,* pp. 126-127, Feb. 2021.

[23] M. Choi et al, "An Output-Bandwidth-Optimized 200Gb/s PAM-4 100Gb/s NRZ Transmitter with 5-Tap FFE in 28nm CMOS," *ISSCC Dig. Tech. papers,* pp. 128-129, Feb. 2021.

[24] M. Kossel et al, "An 8b DAC-Based SST TX Using Metal Gate Resistors with 1.4pJ/b Efficiency at 112Gb/s PAM-4 and 8-Tap FFE in 7nm CMOS," *ISSCC Dig. Tech. papers,* pp. 130-131, Feb. 2021.

[25] P. Mishra et al, "A 112Gb/s ADC-DSP-Based PAM-4 Transceiver for Long-Reach Applications with ¿40dB Channel Loss in 7nm FinFET," *ISSCC Dig. Tech. papers,* pp. 138-139, Feb. 2021.

[26] J. Yuan and C. Svensson, "High-Speed CMOS Circuit Technique," *IEEE J. of Solid-State Circuits,* vol. 24, pp. 62-70, Feb. 1989.

[27] C. K. Yang et al, R. Farjad-Rad, and M. H. Horowitz, "A 0.5-um CMOS 4.0-Gb/s Serial Link Transceiver with Data Recovery Using Oversampling," *IEEE J. of Solid-State Circuits*, vol. 33, pp. 713-722, May 1998.

[28] M. Kossel et al, "A T-Coil Enhanced 8.5-Gb/s High-Swing SST Transmitter in 65-nm Bulk CMOS with ¡-16 dB Return Loss Over 10-GHz Bandwidth," *IEEE J. Solid-State Circuits*, vol. 43, pp. 2905-2920, Dec. 2008.