3.5 A 16-to-40Gb/s Quarter-Rate NRZ/PAM4 Dual-Mode Transmitter in 14nm CMOS

Jihwan Kim, Ajay Balankutty, Amr Elshazly, Yan-Yu Huang, Hang Song, Kai Yu, Frank O'Mahony

Intel, Hillsboro, OR

Emerging standards in wireline communication are defining a path to data-rates of 40Gb/s and beyond. Most previous standards for these networking applications use NRZ signaling. However, practical signal integrity constraints have led to a renewed interest in also supporting PAM4 for some applications and loss profiles [1-2]. Recently, several transmitters have been reported that operate between 28 and 60Gb/s using NRZ or PAM4 modulation exclusively [2-4]. However, high-speed SerDes building blocks that support both a wide frequency range and multiple forms of modulation provide more compatibility between components and avoid the development of multiple IPs. In addition, these blocks must continue to scale into the next-generation of CMOS process technologies to lower the cost by reducing area and power consumption. This paper presents a dual-mode transmitter (TX) implemented in 14nm CMOS that supports both NRZ and PAM4 modulations and operates from 16 to 40Gb/s. The TX incorporates a 4-tap NRZ FIR filter that is reconfigurable to drive PAM4 levels, quarter-rate clocking with a high-bandwidth 4:1 serializer, a duty-cycle and quadrature-error correction circuit with statistical phase error detection, and compact, multi-layer T-coils for pad capacitance (C_{pad}) reduction.

The TX block diagram is presented in Fig. 3.5.1. An on-chip pattern generator sends 8b parallel data to a mode selector through a 16:8 MUX. The data is further serialized for 4-tap FFE in NRZ mode or split into four 2b bundles (MSB and LSB) in PAM4 mode, and fed into the output stage that consists of 36 segments to cover equalization/modulation as well as PVT variation of the termination resistance. Each segment contains a retimer, an equalization (EQ) selector, a single-ended-to-differential (S2D) converter, a 4:1 serializer, a pre-driver, and a source-series terminated (SST) driver with a shared resistor. Figure 3.5.1 also illustrates the output stage segmentation for both NRZ and PAM4 modes. In NRZ mode, the segments are divided into three parts: cursor/1st pre-cursor, cursor/2nd pre-cursor, and cursor/post-cursor which contain ¼, ¼, and ½ of the total number of segments, respectively. The configurability of the NRZ TX taps supports a wide range of channel profiles while minimizing the number of segments. In PAM4 mode, the MSB is fed to ⅔ of the segments and the LSB is fed to ⅓ of the segments to generate four different voltage levels without providing equalization. The TX maintains quarter-rate clocking and operates at half the symbol rate of NRZ mode to maintain the same data and clock paths.

While half-rate and quarter-rate architectures ideally consume the same CV^2f power, a quarter-rate serialization scheme can be an attractive solution at very high data-rates because it relaxes the speed and timing requirements in the serializer. Figure 3.5.2 (a) shows the 4:1 serializer using back-to-back pass-gates. Compared to a 2:1 serializer, the 4:1 serializer relaxes the critical timing path by 2UI that is set by $t_{ck\text{-}q}+t_d+t_{setup}<3UI$, where $t_{ck\text{-}q}$, t_d, and t_{setup} are ck-to-q delay of the retimer, data buffer delay, and the setup time of the serializer, respectively. However, this topology suffers from poor ISI at the output because of charge-sharing effect at the intermediate node between two pass-gates. The junction and interconnect parasitic capacitance adds non-negligible ISI because large pass-gates are used to minimize the series resistance. When one of the four branches is transmitting data (e.g., CK4_90 and CK4_180 are high and d1 is transmitted), one of the pass-gates in adjacent branch is inevitably turned on as well (by CK4_90), enabling a path to another parasitic capacitance (at node X), which increases ISI at the output. To resolve this issue, a feed-forward charge-injection technique is used as shown in Fig. 3.5.2 (b). Auxiliary branches with smaller pass-gates forward the same data to the internal node of previous bit's branch to charge up and reduce the output ISI. Simulations at 40Gb/s show ISI is reduced by 77% and eye height is increased by 30% at the worst-case condition, adding only 10% to the serializer clock buffer power.

The TX driver is a voltage-mode SST type with a shared resistor. This architecture is primarily chosen to minimize the effective C_{pad} at the driver output and reduce power consumption compared to a CML driver. Within the enabled segments, the effective C_{pad} associated with the switch transistor diffusion and routing parasitics to the resistor is significantly reduced due to the resistive division between the switch and the resister (1:5 in this design). The shared resistors along with driver junction parasitic diodes provide the secondary ESD protection and effectively reduce the required size of ESD diodes. Furthermore, sharing the resistor reduces overall C_{pad} by halving the amount of the resistor used for termination and avoids un-driven internal terminals compared with an unshared resistor SST driver [4]. Even with these optimizations, further reduction of the total C_{pad} is still required to achieve the adequate bandwidth for 40Gb/s operation, and a compact multi-layer on-chip T-coil is used. The T-coil, shown in Fig. 3.5.3, splits the total inductance asymmetrically to absorb the different values of the driver and ESD diodes parasitic capacitance. It is implemented using two adjacent metal layers to maximize the inductance per unit area. The measured T-coil inductance is 750pH, self-resonant frequency (SRF) is 38GHz, and size is 35×45μm^2. The T-coil ESR of about 5Ω at DC is used as part of the driver output resistance. In simulation, the T-coil reduces the effective C_{pad} by over 60% across all frequencies and extends the 3dB bandwidth by a 2.3× (from 10 to 23GHz).

The quarter-rate TX requires I/Q clocks with minimal duty-cycle and quadrature error. Figure 3.5.4 shows the duty-cycle/quadrature-error detection (DCD/QED) and correction (DCC/QEC) blocks. The clock distribution is implemented in full-swing CMOS to achieve better area scaling than CML. Fanouts are chosen to provide adequate bandwidth and minimize jitter amplification. The DCD/QED senses I/Q clocks generated by a digital DLL [5] at the output of the clock buffer in the TX. The duty-cycle error is measured by comparing the accumulated number of ones when sampling the P/N clocks with the asynchronous VCO ($f_{osc} \approx 4$GHz) [6]. The quadrature-error is then detected, in a similar way to [7], by measuring the overlap between I/Q clocks. An FSM controls the DCC that is composed of programmable delay cells with rising/falling edges being independently adjusted by pull-up/pull-down controls and the QEC that contains variable capacitive loads for I/Q error correction.

Two TX lanes are fabricated in a 14nm tri-gate CMOS process (Fig. 3.5.7). Each TX (excluding PLL and LDO) occupies 130×215μm^2. A quarter-rate clock from a pair of LC-PLLs drives the TX, and measured eye diagrams with PRBS7 including the effect of 12 to 14dB channel loss are shown in Fig. 3.5.5. The TX achieves RJ of 300/380fs$_{rms}$ for 14/20GHz clock patterns at the output. At 28/40Gb/s, TJ is 10.7/12.9ps$_{pp}$ and ISI is 4.13/4.46ps$_{pp}$ in NRZ mode. In PAM4 mode, a software-based CTLE is applied at the scope (A_{dc}=0.5, z_1=2.5GHz, p_1=4.5GHz, p_2=16GHz for 33.6Gb/s and 20GHz for 40Gb/s, and gain peaking of less than 4dB) to emulate the effect of a practical RX CTLE and equalize for the channel loss. The PAM4 eye widths and heights are 0.87/0.65UI and 102/61mV, respectively, at 33.6/40Gb/s. The TX consumes 195/518mW at 28/40Gb/s in NRZ mode and 141/168mW at 33.6/40Gb/s in PAM4 mode. The power consumption at 40Gb/s in NRZ mode is significantly higher than designed because the TX is fabricated in a low-power logic version of the process technology that requires an elevated supply voltage only for NRZ mode at 40Gb/s. Nonetheless, the data indicates a clear power benefit for PAM4 operation from 28 to 40Gb/s. The performance of the TX is summarized in Fig. 3.5.6, and the power breakdown for NRZ operation at 28Gb/s is shown in Fig. 3.5.7.

Acknowledgements:
The authors thank W. Y. Li, H. S. Kim, K. Chandrashekar, A. Mezhiba, D. Shi, K. Nguyen, D. John, R. Singhal, B. D. Grossnickle, J. Bondie, D. Baker, K. Rasmussen, C. Walker, and Z. Qian for valuable contributions to this work.

References:
[1] Optical Interconnect Forum (OIF) [Online]. Available: http://www.oiforum.com/public/currentprojects.html
[2] P.-H. Chiang, et al., "60Gb/s NRZ and PAM4 Transmitters for 400GbE in 65nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 42-43, Feb. 2014.
[3] A. A. Hafez, et al., "A 32-to-48Gb/s Serializing Transmitter Using Multiphase Sampling in 65nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 38-39, Feb. 2013.
[4] C. Menolfi, et al., "A 28Gb/s Source-Series Terminated TX in 32nm CMOS SOI," *ISSCC Dig. Tech. Papers*, pp. 334-335, Feb. 2012.
[5] A. Elshazly, et al., "A 2GHz-to-7.5GHz Quadrature Clock-Generator Using Digital Delay Locked Loops for Multi-Standard I/Os in 14nm CMOS," *IEEE Symp. VLSI Circuits*, pp. 52-53, June 2014.
[6] T. -C. Hsueh, et al., "A 25.6Gb/s Differential and DDR4/GDDR5 Dual-Mode Transmitter with Digital Clock Calibration in 22nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 444-445, Feb. 2014.
[7] M. Mansuri, el al., "An On-Die All-Digital Delay Measurement Circuit with 250fs Accuracy," *IEEE Symp. VLSI Circuits*, pp. 98-99, June 2012.

 978-1-4799-6224-2/15/$31.00

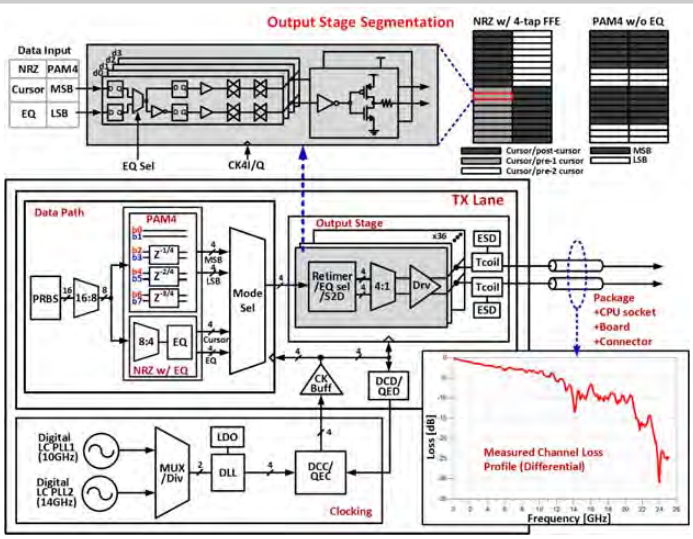

Figure 3.5.1: Block diagram of the dual-mode TX, channel loss profile, and segmentation for NRZ/PAM4 modes.

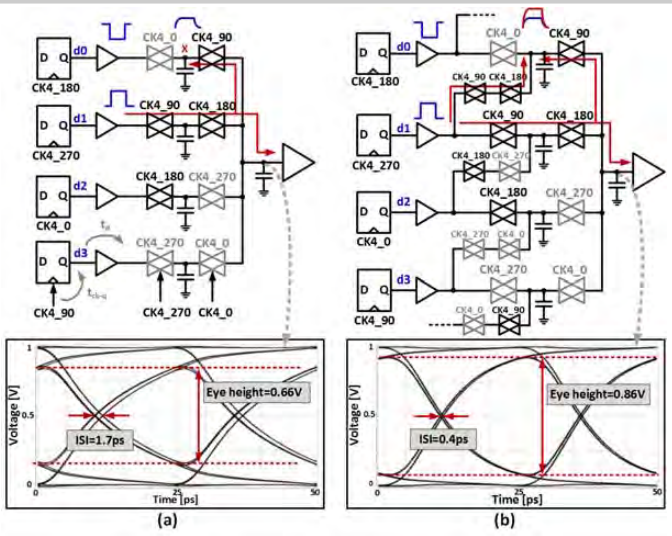

Figure 3.5.2: (a) 4:1 serializer using back-to-back pass-gates, and (b) proposed feed-forward charge injected 4:1 serializer.

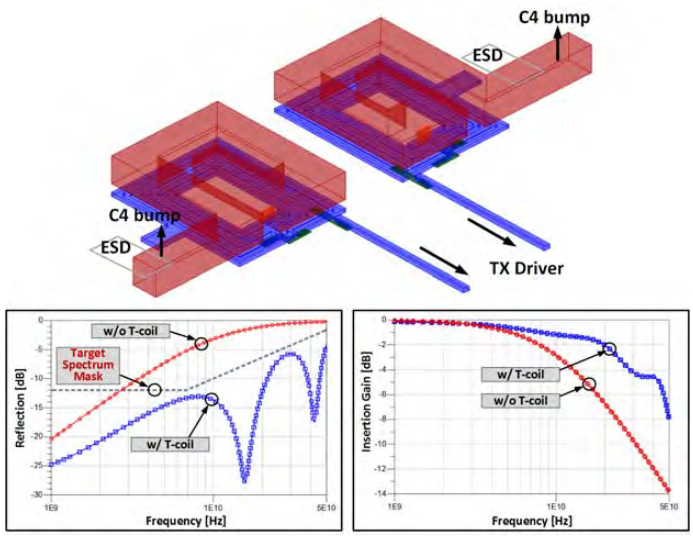

Figure 3.5.3: The multi-layer T-coil and simulated reflection/bandwidth of TX driver.

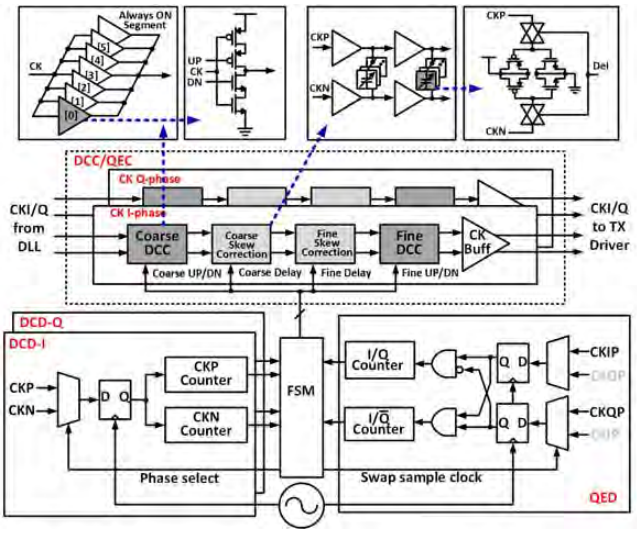

Figure 3.5.4: Block diagram of the duty-cycle/quadrature-error detection/correction circuits.

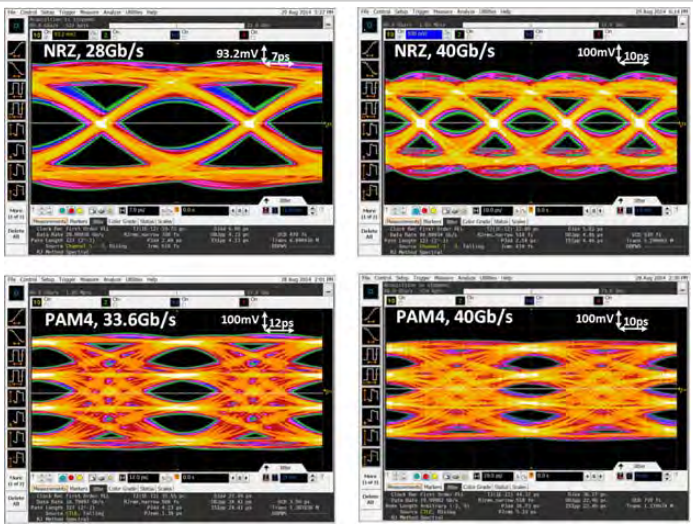

Figure 3.5.5: Measured TX eyes including test channel (PAM4 eyes include software-based CTLE at scope).

Reference	Chiang '14 [2]		Hafez '13 [3]	Menolfi '12 [4]	This work			
Technology	65nm	65nm	65nm	32nm	14nm			
Architecture	1/2 rate	1/2 rate	1/4 rate	1/2 rate	1/4 rate			
Driver Topology	CML	CML	CML	SST	SST			
TX FFE	2-tap	3-tap	No EQ	4-tap	4-tap		No EQ	
Modulation	NRZ	PAM4	NRZ	NRZ	NRZ		PAM4	
Data-rate Range (Gb/s)	60	60	32-48	25-32	16-40			
Data-rate (Gb/s)	60	60	48	28	28	40	33.6	40
Loss @ Nyquist (dB)	4.6	N/A	N/A	N/A	13	12.5	3	4
Power (mW)*	374.85	261	59.4**	217	195	518	140.7	167.5
Energy Efficiency (pJ/b)*	6.25	4.35	1.24	7.75	6.95	12.95	4.18	4.19
RJ (ps_{rms})	0.461	0.508	0.251	N/A	0.33	0.51	0.58	0.55
TJ (BER=10^{-12}) (ps_{pp})	5.331	N/A	N/A	6***	10.72	12.89	35.55	44.37
Area (mm^2)	2.1	1.14	0.4	0.0363	0.0279			

* PLL power was excluded
** TX is not fully featured
*** Measurements were done by on-wafer probing

Figure 3.5.6: TX performance comparison.

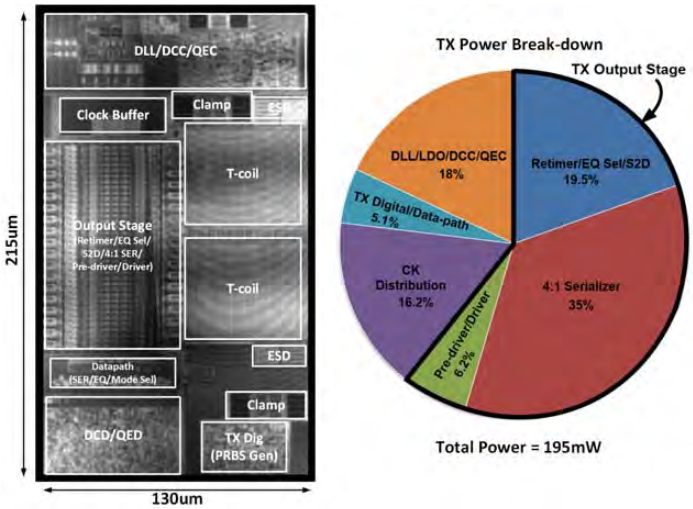

Figure 3.5.7: TX die micrograph (top metal layer is not shown) and power breakdown for 28Gb/s NRZ operation.

978-1-4799-6224-2/15/$31.00