100G and 200G per Lane Linear Drive Optics for Data Center Applications

Elaine S. Chou, Yishen Huang, Siamak Amiralizadeh, Jeffrey Rahn, J. K. Doylend, Qing Wang, Janet Chen, and Darron Young

Meta, 1 Hacker Way, Menlo Park, CA 94025, USA

eschou@meta.com

Abstract: 100G/lane linear-drive pluggable optics demonstrate interoperability with over 3 dB link margin. Simulations suggest that 200G/lane linear drive requires bump-to-bump losses below 22 dB, but transmit-side retimers increase loss tolerance beyond 34 dB.

1. Introduction

As data center networks scale in bandwidth and size, the total cost and power of pluggable optical transceivers have grown significantly. In response, the industry is exploring optical transceivers without integrated retimers or digital signal processers (DSPs) that rely instead upon the switch application-specific integrated circuit (ASIC) DSPs; these linear drive optics have been demonstrated at 100G/lane in linear-drive pluggable optics (LPO) and co-packaged optics (CPO) form factors [1, 2]. Fig. 1 compares the building blocks of a transmission link using retimed and linear drive optics. Regeneration of the signal at the optics line transmit (Tx) and host receive (Rx) sides in the retimed optical transceiver DSP segments the signal path cleanly between the electrical and optical domains. One major challenge of linear drive optics is the elimination of this regeneration between link segments, which introduces loss impairments from the electrical channel at either end of the optical link. LPO and CPO links primarily differ in the length and therefore loss of the electrical path between switch ASIC and optical transceiver.

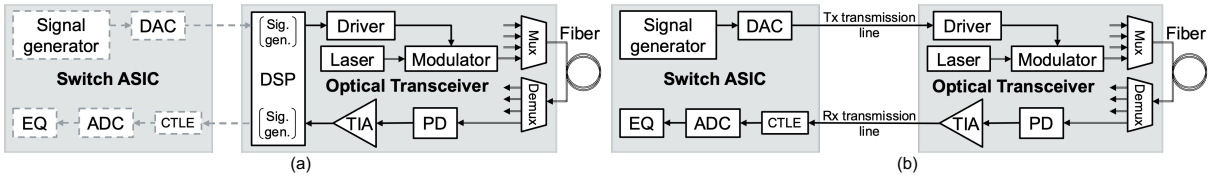

Fig. 1. Block diagrams of (a) retimed and (b) linear drive WDM optics links. With retimed optics, the link between switch ASIC and optical transceiver (dashed) can be excluded in optical link analysis.

This paper shares insights regarding 200G/lane pulse-amplitude modulation (PAM)-4 linear drive optics simulations that are benchmarked against 100G/lane LPO measurements, providing a link margin comparison between retimed optics, LPO, and CPO; further extending analysis to include retiming at the Tx side only.

2. 100G per lane measurements

Unretimed links are transparent to noise and signal distortion accumulated along the entire signal path, potentially requiring link optimization across switch ASIC, host board, and optical transceivers from multiple vendors, creating design and interoperability challenges. To address these areas of concern, we tested two preliminary 100G/lane LPO module designs in a 51.2T switch with host ASIC to optical module electronic integrated circuit (EIC) ball-to-ball loss up to 15 dB at 26 GHz. The two designs, labeled LPO1 and LPO2, have tradeoffs in bandwidth and efficiency. With Tx filter tuning for each port optimizing the switch output signal at test point (TP) 1a, both LPO module designs successfully close the link across all ports without any optical transceiver tuning [3]. Fig. 2a shows an open eye diagram measured at the optics output (TP2) on a 14-dB-loss port using LPO2. The measured LPO2 Tx response in Fig. 2b shows almost 3 dB peaking from the driver that compensates for some of the channel insertion loss. For a given LPO design, TP2 performance, shown in Fig. 2c, is consistent across all ports. Most ports achieve Tx dispersion eye closure quaternary (TDECQ) below 3.5 dB with both LPO designs. The LPO1 Tx signal has 2 dB higher extinction ratio (ER) but 1 dB higher TDECQ, consistent with the Tx design differences.

Interoperability testing between the two LPO module designs and retimed modules shows small differences in Rx sensitivity, as shown in the bit error ratio (BER) vs. optical modulation amplitude (OMA) curves in Fig. 2d.

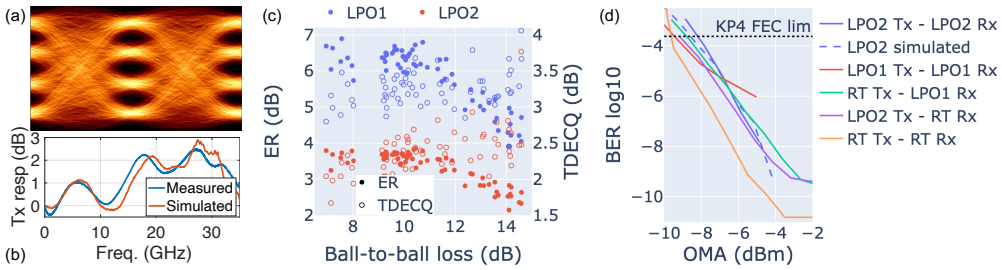

Fig. 2. 100G/lane LPO measurements: (a) LPO2 Tx eye for a port with 14 dB ball-to-ball loss at 26 GHz. (b) LPO2 Tx response, measured vs. simulated. (c) ER and TDECQ vs. loss at 26 GHz. (d) BER vs. OMA in the highest loss port for a subset of LPO and retimed (RT) combinations.

The sensitivity margin from −4.6 dBm OMA ranges from 3 dB to 5 dB across all ports, with up to 1.5 dB penalty compared to the fully retimed link. A BER floor of 10^{-9} is achieved with most optical module combinations; notably however, the LPO1 design demonstrates a degraded BER floor up to 10^{-7} attributed to Tx electrical crosstalk or reflections. The BER simulations demonstrate the accuracy of our modeling vs. the measured 100G data, and so form a baseline for the 200G/lane performance projections.

Measurements of 100G/lane pluggable optics show a 37% power reduction from 20 pJ/bit using retimed optics to 12.5 pJ/bit with LPO. CPO can achieve further improved power efficiencies of 5-10 pJ/bit [2].

3. 200G per lane projections

To forecast linear drive optics performance at 200G/lane, we ran time-domain link simulations from the switch ASIC Tx, through the optical link, to the remote switch ASIC Rx, simulating three retiming cases: linear drive, Tx retimed, and fully retimed. A Tx retimer may improve performance but also cuts LPO power savings roughly in half to 18%, assuming retimed optics and LPO power consumption scales by the same factor from 100G to 200G.

3.1. Link model

The simulation passes a psuedorandom binary sequence quaternary (PRBSQ) signal through the components of the transmission link shown in Fig. 1. Link components, with key parameters summarized in Table 1, include a signal generator with Tx pre-emphasis, digital-to-analog converter (DAC), transmission line, driver with 5 dB peaking, Mach-Zehnder modulator (MZM), single-mode fiber (SMF), photodiode (PD), transimpedance amplifier (TIA), continuous-time linear equalizer (CTLE) assumed to enable perfect clock and data recovery, analog-to-digital converter (ADC), and Rx equalization (EQ). For retimed cases, the electrical link segments on the retimed ends are not simulated because they are isolated from the optical link by signal regeneration and assumed not to affect overall link performance. For a 128-radix cable-routed switch transmitting at 200G/lane, worst-case bump-to-bump loss is estimated to be 30 dB at 56 GHz, including 11 dB of switch ASIC and optical module package loss [4], whereas CPO may achieve less than 12 dB loss. To minimize fiber count, we transmit a 4-wavelength coarse wavelength division multiplexed (CWDM4) signal with 20 nm spacing in the O-band over 500 m of SMF in line with the majority of our data center fiber infrastructure. For 200G/lane transmission, a concatenated KP4 and Hamming(128,120) code with a forward error correction (FEC) threshold of 4.85×10^{-3} [5] has been proposed to relax BER requirements of the link compared to the 2.4×10^{-4} KP4 FEC threshold widely used at 100G/lane, at the cost of 6% additional overhead that increases the 200G symbol rate from 106 Gbaud to 113 Gbaud.

Table 1. 200G/lane simulation parameters

Component	Key parameters	Component	Key parameters
DAC	41 dB SNDR, 56 GHz BW	Demux	1.5 dB loss
Driver	20 dB max gain, 56 GHz BW	PD	0.6 A/W responsivity, 65 GHz BW
Laser	−145 dB/Hz RIN, 1 MHz linewidth	TIA	18 pA/Hz$^{1/2}$ IRN, 56 GHz BW
MZM	5.5 V_π, 60 GHz BW	EQ	15 dB CTLE, 32-tap FFE, 2-tap DFE
Fiber	500 m, −5.9 to 3.3 ps/nm/km CD	FEC	KP4 / KP4 + Hamming(128,120)

SNDR: signal-to-noise-and-distortion ratio, BW: bandwidth, RIN: relative intensity noise, CD: chromatic dispersion, IRN: input referred noise, FFE: feed-foward equalizer, DFE: decision-feedback equalizer

3.2. *Simulation results*

The Tx finite impulse response (FIR) filter in the switch ASIC can compensate for channel loss at the cost of reduced signal voltage swing. If the Tx FIR is optimized to open the 113 Gbaud Tx eye, TDECQ is below 3.9 dB up to 25 dB bump-to-bump loss, but ER reduces drastically (Fig. 3a, red). If instead the Tx FIR is optimized with a constraint on the switch output signal swing, ER stays above 3 dB, but TDECQ degrades to 5 dB at 25 dB loss (Fig. 3a, blue). Fig. 3b shows eye diagrams of the swing-optimized Tx output for different loss values. To achieve 3.5 dB ER and 3.9 dB TDECQ [6], the shaded green areas in Fig. 3a, loss must be lower than 23 dB.

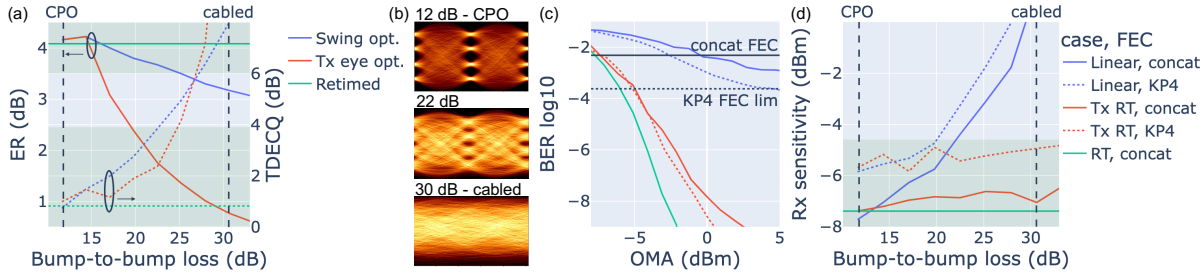

Fig. 3. Simulated 200G/lane linear drive performance for a 1264.5 nm PAM-4 signal transmitted 500 m: (a) 113 Gbaud ER / TDECQ tradeoff for two Tx FIR optimizations, compared with a retimed Tx, noting CPO and cabled switch losses at 56 GHz. All-PCB routing adds ∼3 dB to cabled loss. (b) Swing-optimized 113 Gbaud Tx eyes. (c) BER vs. OMA in a cabled switch, and (d) Rx sensitivity vs. 56 GHz loss of linear drive, retimed Tx, and fully retimed optics with concatenated or KP4 FEC.

Fig. 3c compares BER vs. OMA performance of different retiming and FEC schemes for a cabled switch system using pluggable optics. Although concatenated FEC increases the symbol rate and thus BER, due to its higher FEC threshold, Rx sensitivity improves from 5 dBm Rx OMA using KP4 FEC to 0 dBm using concatenated FEC. But ultimately, both cases lack the margin required for reliable data transmission in data center links. A Tx retimer significantly improves performance, achieving less than 1 dB sensitivity penalty compared to fully retimed optics. Rx sensitivity vs. bump-to-bump loss plotted in Fig. 3d shows that for LPO, 22 dB loss with concatenated FEC is required to meet −4.6 dBm Rx sensitivity [6]. Concatenated FEC improves Rx sensitivity and loss tolerance over KP4 FEC by approximately 1.5 dB and 2 dB, respectively. The loss penalty with a retimed Tx is within 1 dB well beyond the worst-case cabled switch loss. For less than 15 dB loss, which includes the CPO case, linear drive Tx metrics and Rx sensitivity are comparable to retimed optics.

4. Conclusion

Early 100G/lane LPO samples demonstrate open Tx eyes and over 3 dB margin in Rx sensitivity with optical transceiver-agnostic switch Tx calibration. This architecture supports interoperation between different optics designs and retiming schemes and satisfies important requirements to enable a scalable solution and open ecosystem. When we consider future 200G/lane generation LPO architectures, a stronger concatenated FEC shows significant benefit over KP4 FEC, but bump-to-bump losses below 22 dB are still needed to close an unretimed link. We therefore conclude that a Tx retimer is likely needed to close the 200G/lane LPO link with Meta's planned switch architectures. However, LPO performance can potentially be improved with new optical technologies such as thin film Lithium Niobate (TFLN) or modified switch architectures achieving lower channel insertion loss.

Acknowledgment

The authors thank Srinivas Venkataraman, Halil Cirit, Drew Alduino, Absar Ulhassan, Xuan He, Chris Berry, Rongchun Zhou, Xu Wang, and Olaf Moeller.

References

1. H. Chaouch, "Linear-drive pluggable optics (LPO) for power-efficient AI/ML optical interconnects," ECOC 2023 Exhibition, Glasgow, Scotland, Oct. 2023.
2. K. Muth, *et al.*, "High density integration technologies for SiPh based optical I/Os," ECTC 2023, Orlando, FL, 2023.
3. Y. Huang, Q. Wang, J. Chen, "Methods for integration and evaluation of LPO", OCP 2023, San Jose, CA, Oct. 2023.
4. S. Venkataraman *et al.*, "224G C2M VSR channel analysis," OIF PLL working group, oif2022.465.01, Nov. 2022.
5. L. Patra, "Updates on concatenated FEC proposal for 200G/Lane PMD," IEEE 802.3df meeting, July 2022.
6. "800G-FR4 Technical Specification Draft 1.0," 800G Pluggable MSA, June 2021.