White Paper

AI, HPC, Cloud, Storage, Networking

Agilex™ 7 SoC FPGAs Enable 400G-DR4-LPO Optical Modules to Significantly Reduce Power, Cost, and Latency

Altera's 116G FPGA transceivers successfully interoperate with Linear Pluggable Optics (LPO) modules. FPGA-based SmartNICs or AI NICs enable programmable acceleration and offloading for AI clusters and HPC in hyperscale cloud/data centers, storage, and networking infrastructure. Customers benefit from LPO technologies' lower power, lower cost, lower latency, and plug-and-play compatibility for high-speed interconnects to meet the increasing demands and requirements of these applications.

Authors

Dr. Mike Peng Li

Fellow and Chief Technologist of SERDES & I/O
Altera Corporation

Harrison Zou

Product Line Manager
Altera Corporation

Roshni Suresh

Transceiver Product Manager
Altera Corporation

Introduction

A new and important emerging technology for 112G electro/optical I/O interfaces is the 112G-Linear, where a linear pluggable optical module (LPO) is paired with 112G host DSP transceivers (XCVRs), providing significant power reduction for the optical module versus conventional DSP/retimed modules while maintaining performance and interoperability. In addition to power saving, the LPO and linear interface combination also offers lower cost, latency, and compatibility, making it better suited to meet the AI/ML, HPC, cloud/DC, storage, and networking application requirements compared with conventional retimed optical modules and interfaces.

To enable the LPO and Linear interface, the Optical Internetworking Forum (OIF) trade organization finalized its CEI-112G-Linear-PAM4 implementation agreement (IA)[1] specifying linear operation of 1 to N lanes of data rates up to 112Gbps-PAM4 over 20 mm to 100 mm electrical interconnect and 1 connector, with insertion loss (IL) up to 13 dB at Nyquist. To further enable the link system solution from electrical/PHY to PCS/FEC/MAC, the 100G-DR-LPO (Distance Range) specification has been developed and finalized by the LPO MSA[2]. It defines 100 G/lane optical interfaces at an 802.3 Ethernet rate of 53.125 Gbps PAM4, using single-mode fiber with a reach of up to 500 m, and with host XCVRs having DSP-based SerDes and RS (544,514) FEC. The specification leverages and builds on the OIF CEI-112G-LINEAR-PAM4 specification, 802.3 PCS/FEC/MAC logics, and optical PMD specifications. It enables Ethernet-like links with applications needing x1/2/4/8 lanes (100GE/200GE/400GE/800GE) and various optical pluggable form factors such as OSFP, QSFP, QSFP-DD.

AI/ML, HPC, cloud/DC, storage, and networking end users who have already deployed the 112G LPO pairing with host 112G DSP SERDES would like to see the power, cost, and latency benefits scale up for their next-generation deployments. Responding to this demand, the OIF has formally started the CEI-224G-Linear project to develop the CEI-224G-Linear-PAM4 specification, in parallel with CEI-224G-LR-PAM4 (Long Reach) specification development, to leverage the commonality or synergy in SERDES/DSP capability. Correspondingly, the LPO MSA has also started its 200G-DR-LPO project. This new LPO project defines 200 G/lane optical interfaces at the 802.3dj Ethernet rate of 106.25 Gbps PAM4, using single-mode fiber with up to 500 m reach, combined with host XCVRs having DSP-based SerDes having associated advanced MLSD (Maximum Likelihood Sequence Detection) signal detection, and RS (544,514) FEC. The specification will continue leveraging and building on the OIF CEI-224G-LINEAR-PAM4 specification, and 802.3dj PCS/FEC/MAC logic specifications, as well as optical PMD specifications. It enables Ethernet-like links with x1/2/4/8 lanes (200GE/400GE/800GE/1600GE) applications.

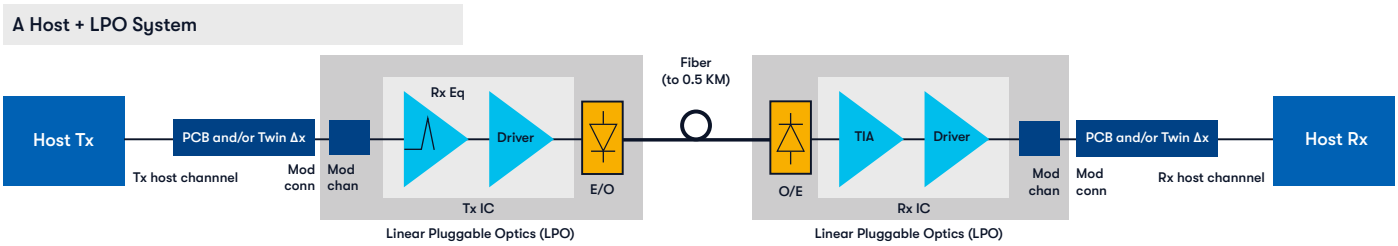

Figure 1. LPO link system diagram (source: lpo-msa.org)

How LPO Works

An LPO link system is shown in Figure 1.

Compared with a retimed optical module, LPO modules do not include any DSP or SERDES circuitry, which enables them to achieve lower power, lower cost, and lower latency. The end-to-end link with an LPO system (Figure 1) consists of the following components:

- Host Tx package (e.g. FPGA XCVR I/O) C4 bump to bump
- Host PCB or cable
- Host/optical module connector
- Optical module PCB and package
- Optical module TX including CTLE, driver, modulator, or E/O converter, optical connector
- Optical fiber
- Optical module RX including O/E converter, TIA, and driver
- Host Rx package (e.g. FPGA XCVR I/O) C4 bump to bump
- Host PCB or cable

The link diagram shows a linear channel, and all the signal impairments, such as jitter, ISI, x-talk, reflections, dispersion, noise, and interference, that need to be compensated by the host TX and RX.

An important reason why linear LPO can work at 112 Gbps PAM4 is that most, if not all, 112G CEI and Ethernet compliant SERDES have DAC/ADC, which enable the advanced DSP and associated signal compensation/processing, such as CTLE+FFE+DFE equalizers in the host RX, and FIR in the host TX. ADC/DAC/DSP-based SERDES scales better with Moore's law or process technology compared with analog/mixed signal-based SERDES architecture, enabling the LPO to be extended and scaled to its next generation of 224 Gbps PAM4.

LPO compliant host SERDES needs to meet the host output (TPA1) and input (TP4A) specifications, and the LPO compliant module needs to meet module output (TP2) and module input (TP3) specifications defined in the OIF specification[1]. In addition, the LPO and host link need to meet the link-level specifications defined in Table 1 when Ethernet test data is transmitted and received.

Parameter	Typ.
Measured BER[1] / pre FEC BER[1]	10^{-6}
t-count[2] / FEC bin count	7

[1] Measured for 3 seconds
[2] Measured for 100 seconds

Table 1. LPO Link Performance with Ethernet test (source: lpo-msa.org)

Industry-Leading Programmable Solutions: Altera Agilex™ 7 FPGAs and SOCs with High Performance 116G FHT Transceiver I/O

The Altera Agilex 7 FPGAs and SoCs device family includes the highest performance FPGAs and SoCs in the industry, 2x higher performance per watt versus competitors' 7nm FPGAs. To address the different requirements of multiple end markets and applications, the Agilex 7 family is offered in a range of densities and features grouped in the F-Series, I-Series, and M-Series variants.

- Offered within all members of the I-Series and M-Series devices are advanced transceiver I/Os with the highest data rate in the industry—up to 116 Gbps, integrated Hard IPs (400G MAC/PCS/FEC, 200G PCS/FEC, and PCIe), based on Altera's F-Tile chiplet technology.
- Selected variants based on Altera's R-Tile chiplet technology support the industry's first PCIe5.0 and Compute Express Link (CXL) 1.1/2.0 compliant devices. Agilex 7 has also demonstrated PCIe 6.0 interoperability.
- Selected variants within the M-Series devices offer in-package HBM2E memory, delivering the highest memory bandwidth in the industry—over 1 terabyte per second (TBps)

Agilex 7 devices are manufactured using a heterogeneous System-in-package (SiP) chiplet architecture, providing flexibility and time-to-market (TTM) advantages. Agilex 7 devices utilize two generations of transceiver tiles, allowing Altera to mix-and-match general-purpose and processor attach tiles to meet a variety of use cases and market needs.

F-Tile is Altera's newest generation and highest performance multi-purpose transceiver tile, comprising the following features:

- The highest data rate is 116 Gbps PAM4 / 58 Gbps NRZ (labeled as 'FHT' transceivers) alongside the general-purpose 1- 58 Gbps (labeled as 'FGT' transceivers) I/Os.
- Fully compliant with industrial specifications (e.g. IEEE and OIF-CEI, PCIe, CPRI/ECPRI, JESD204, FC, OTN/OTU).
- Integrates 400G MAC/PCS/FEC hard IP cores, 200G MAC/PCS/FEC hard IP cores, and PCIe 4.0 hard IP cores. Hard IP reduces the need to use FPGA logic for protocol interfacing while still allowing for customization options.

F-Tile has excellent flexibility and capabilities: tuning, different configurations, mappings, clocking, PLLs, dynamic reconfigurations, which make it possible to support future protocol implementations in these rapidly changing applications, such as scaling of AI/ML as well as next-generation HPC, cloud/DC, and networking/storage.

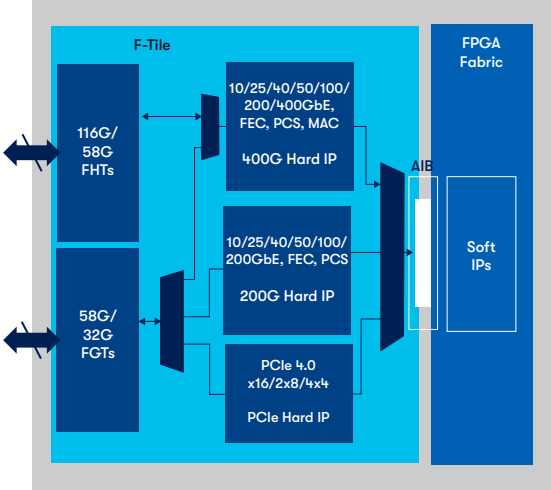

Figure 2. Altera Agilex™ 7 F-Tile Architecture

Agilex™ 7 400G-DR4-LPO Solutions

The Agilex 7 FHT transceiver I/O block uses a leading-edge ADC/DAC/DSP architecture combined with advanced equalization capability, including CTLE+FFE+DFE+ floating FFE[3]. Using the setup as shown in Figure 3, and when the Agilex FPGA host is paired with a 100G-DR-LPO-compliant LPO module, they have achieved performance results exceeding the LPO specification.

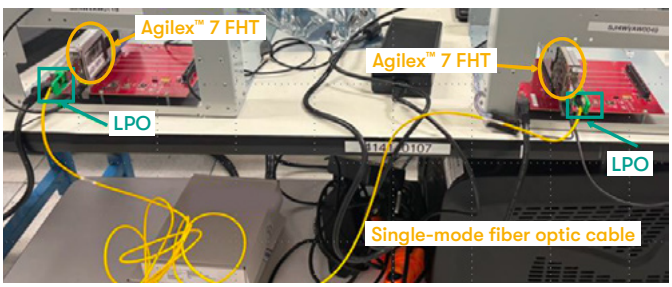

Figure 3. LPO link system demonstration setup using Agilex™ 7 FHT

The test setup operates in full-duplex mode, transmitting and receiving 400GE traffic using a 4×100G lane configuration, interoperating with the 100G-DR-LPO module. Multiple experiments have achieved the following pre-FEC BER and FEC bin count across a 100-second measurement time:

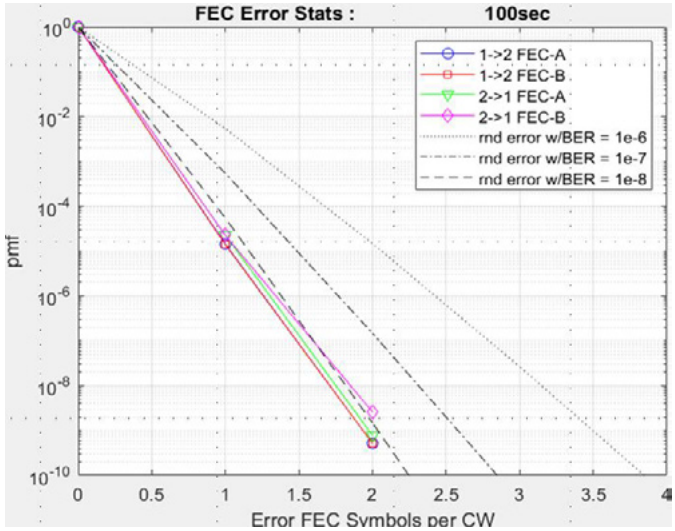

Figure 4. Measured FEC statistics vs pre-FEC BER limits with random error and 100 seconds of measurement time.[4]

Parameter	100G-DR-LPO spec	Altera Test Results
Measured BER/pre-FEC BER for 3 seconds	1E-06	1E-08
T-count/FEC bin count for 100 seconds	7	2

Table 2. Pre-FE BER and FEC bin count measurement results vs 100G-DR-LPO specification, with 100 seconds of measurement time

Over a 100-second measurement interval, the system achieved a pre-FEC BER of 1E-08, which is two orders of magnitude better than the 100G-DR-LPO specification limit of 1E-06. Additionally, the FEC bin count was measured at 2, providing a margin of 5 bins below the specification limit of 7.

To validate the long-term stability of the link, extended measurements were conducted with durations ranging from 36× and 1200× longer than the original 100-second and 3-second tests. The results, shown in Figure 5 and Table 3, confirm that the system continues to meet or exceed the 100G-DR-LPO specification across extended durations.

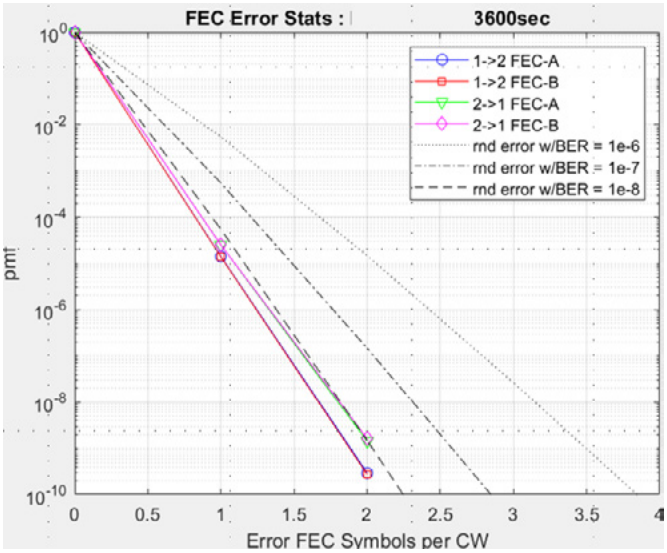

Figure 5. Measured FEC statistics vs pre-FEC BER limits with random error and 3600s measurement time[4]

Parameter	100G-DR-LPO spec	Altera Test Results (3,600-seconds)
Measured BER/pre-FEC BER for 3 seconds	1E-06	1E-08
T-count/FEC bin count for 3,600 seconds	7	2

Table 3. Pre-FE BER and FEC bin count measurement results vs 100G-DR-LPO specification, with 3,600 seconds of measurement time

With 36x and 1200x measurement time, the pre-FEC BER and FEC bin count do not get worse. This means that the error statistics are largely random, or there is no burst error source in the link, and the FEC bin count will likely follow the pre-FEC BER of 1e-8 limit as the measurement time increases. It is anticipated that at a probability of 1E-15, the FEC-bin or corrected FEC symbol length will likely be <~ 3, which is way below the RS (544, 514) correctable FEC symbol length of 15. In other words, there will probably be no uncorrectable FEC symbol with this LPO end-to-end link when the host XCVR is implemented by an Agilex 7 FHT FPGA XCVR. It is also probable that the 400GE traffic will operate at the BER << 1e-13 at the MAC, which exceeds the 400GE Ethernet specification.

Summary

Agilex 7 FPGAs successfully interoperate at 400G with LPO modules, and the results exceed LPO-MSA version 1.0 specifications. FPGA-based NICs, combined with LPO modules, can significantly improve the Total Cost of Ownership (TCO) for operators building new infrastructure to address AI scale-out/scale-up, storage, or HPC clusters. FPGAs are uniquely positioned to allow OEMs to add novel and differentiated features without needing new hardware, functions such as AI pre-processing or computing/networking offload.

Learn More

For more details about Altera's products and solutions, go to:

- Agilex 7 FPGAs and SoCs: https://www.altera.com/products/fpga/agilex/7
- Solutions per Industry Segment: https://www.altera.com (select the Solutions tab)

For specific details about the 116G FHT transceiver and LPO interoperability, contact Altera to request support: https://www.altera.com/contact

References

[1] https://www.oiforum.com/technical-work/implementation-agreements-ias/#Electrical

[2] https://www.lpo-msa.org/home/specifications-and-white-papers.html

[3] Refer to the **F-Tile Architecture and PMA and FEC Direct PHY IP User Guide**: https://www.intel.com/content/www/us/en/docs/programmable/683872/25-3/f-tile-overview.html

[4] Pmf = Probability Mass Function

WP-01347-1.0