CPO JDF

Co-Packaged Optical Module Discussion Document

Table of Contents

Table of Figures

Table of Tables

1. Introduction

1.1. Collaboration Agreement

Before providing feedback to the CPO JDF, please ensure that a representative from your company has reviewed, signed and returned the collaboration agreement document.

1.2. Scope

This document provides guidance on the requirements for co-packaged optical modules designed for high-radix, network switch applications with 100Gb/s electrical interfaces. It is intended to help to establish a framework for discussions on CPO technology and is not intended as a final specification. The figures and examples contained are not meant to represent preferred designs, but rather are sample embodiments to facilitate discussion.

1.3. Introduction

The CPO JDF plans to release three documents focused on different elements of Co-Packaged Optics (CPO): the optical module (covered here), the External Light Source (ELS), and the CPO assembly.

This collection of documents is intended to foster open discussion with vendors pursuing Co-Packaged Optics (CPO). The first revisions are intended to facilitate structured conversations about the different elements of CPO. The documents will be revised periodically as input are collected from vendors.

1.4. Common Terms

The following terms and acronyms are used in this document.

- CPO – Co-Packaged Optics
- 400GE – 400 Gigabit Ethernet
- 400G FR4 – 400GE optical standard utilizing 4 wavelengths on a 1310nm CWDM grid with each wavelength transmitting 106.25Gb/s using 53.13Gbaud PAM4 modulation
- 400G DR4 – 400GE optical standard utilizing ribbon fiber with 4x106.25Gb/s transmit and 4x106.25Gb/s receive lanes/fibers using 53.13Gbaud PAM4 modulation. Single wavelength range, centered at 1310nm, is specified for all lanes
- 200G FR4 - 200GE optical standard utilizing 4 wavelengths on a 1310nm CWDM grid with each wavelength transmitting 53.13Gb/s using 26.56GBaud PAM4 modulation
- CWDM 4 – 100Gb/s optical standard utilizing 4 wavelengths on a 1310nm CWDM grid with each wavelength transmitting 25.78Gb/s using NRZ modulation
- 100G PSM4 MSA – 100GE optical standard utilizing ribbon fiber with 4x25.78Gb/s transmit and 4x25.78Gb/s receive lanes/fibers using 25.78Gb/s NRZ modulation. Single wavelength range, centered at 1310nm, is specified for all lanes
- QSFP-DD –Quad Small Form-factor Pluggable Double Density used for 400GE
- BER – Bit Error Rate
- CEI-112G-XSR – Common Electrical Interface specification operating at a serial data rate of 112Gb/s over short reaches common in multi-chip modules
- CMIS – Common Management Interface Specification
- ESD – Electro-Static Discharge
- SPI - Serial Peripheral Interface
- LD – Laser Diode used as a CW source for the co-packaged optical module
- ELS – External Laser Source

1.5. References

The following specifications are referenced in this document.

- IEEE 802.3bs – Media Access Control Parameters, Physical Layers and Management Parameters for 200 Gb/s and 400 Gb/s Operation
- IEEE 802.3cu – Media Access Control Parameters, Physical Layers and Management Parameters for 100 Gb/s and 400 Gb/s over SMF at 100 Gb/s per Wavelength (currently in draft)
- 100G CWDM4 MSA Technical Specifications Rev 1.1
- 100G PSM4 MSA Specification
- OIF-CEI-x.x (CEI-112G-XSR currently in draft)
- CMIS revision 4.0 – Common Management Interface Specification
- QSFP-DD MSA QSFP-DD Hardware Specification, Rev 5.0

2. Absolute Maximum Ratings

The values in Table 1 below are for guidance and discussion and subject to change.

Parameter	Symbol	Min	Typical	Max	Unit	Note
Storage temperature	T_st	-40		85	°C	
Relative humidity, storage and transportation	RH	5		95	%	Noncondensing
ESD, electrical PINs	ESD	-1		1	kV	Human body model
Power supply voltage	Vcc	-0.3		3.6	V	
Receiver average optical power damage threshold, each lane				10	dBm	

Table 1: Absolute maximum ratings

3. Operating Conditions

See Table 2 below for a description of the operating conditions for the CPO optical module.

Parameter	Symbol	Min	Typical	Max	Unit	Note
Power supply voltage[1]	Vcc1		TBD		V	Host will supply multiple voltage rails to optical module. Min/max +/- 5%
	Vcc2		TBD			
	Vcc3		TBD			
	Vcc4		TBD			
	Vcc5		TBD			
Operating case temperature	Tcase	20		70	°C	With integrated LDs
		20		80		With external LDs
Power consumption	P			8	W	Per 400G, integrated LDs
				7.5		Per 400G, external LDs
Power consumption in Low Power Mode	P_lpm			0.8	W	Per 400G. See section 5.7
Steady state current	Icc1			TBD	A	Current for each voltage supply rail
	Icc2			TBD		
	Icc3			TBD		
	Icc4			TBD		
	Icc5			TBD		
Relative humidity	RH	5		85	%	Non-condensing

Table 2: Operating conditions

Notes:

1. Number of voltage supply rails and rail voltages shown are TBD pending input from optical module vendors. Intent is to provide multiple voltage rails to minimize the need for voltage regulation on the optical module.

4. Optical Requirements

4.1. Background

The first-generation CPO modules must interoperate with other equipment within the data center. Alternate optical approaches that may prove to be more ideally suited for CPO, but these approaches are not considered here due to the interoperation mandate. The JDF welcomes discussions on next-generation optical schemes such as narrower-spaced WDM or low-power coherent modulation schemes with enhanced link budgets.

4.2. Optical Specification

CPO modules will need to support different optical specifications depending on the application. The most common optical standards for first-generation CPO modules will be 400Gb/s as described in as described in IEEE 400GBASE-FR4 (IEEE 802.3cu in draft) and 400GBASE-DR4 (IEEE 802.3-2018, Clause 124).

Note that some CPO assemblies will require a high-speed electrical interface as well as high-speed optical interfaces. Electrical interface guidance is covered in the CPO assembly document.

4.3. Backward Compatibility

The optical modules should support backward compatibility and mixed data-rate applications. Backward compatibility means a single module will support its default 400Gb/s mode and legacy data rates (multi data-rate operation).

For 400GBASE-FR4 CPO assemblies, the optical module should support legacy operation modes that are compatible with 200Gb/s (IEEE 200GBASE-FR4) and 100Gb/s (100G CWDM4 MSA). For 400GBASE-DR4 CPO assemblies, the optical module should support a legacy operation mode that is compatible with 200Gb/s (200GBASE-DR4) and 100Gb/s PSM4 (100G PSM4 MSA).

A single optical module or CPO assembly may operate in mixed-data rate mode, where lanes can be operating at different data rates (e.g. some lanes in default 400G mode and some at legacy data rates). See Section 5.6 of this document for guidance on mixed data-rate operation.

5. Electrical Specifications

5.1. Electrical Connector

Two options are considered for the electrical connection of the optical module to the CPO substrate. The first is direct solder attach of the optical module substrate to the CPO substrate. This is the most compact, straight-forward approach, but does not easily accommodate rework during assembly or test, and soldering processes may not be compatible with epoxies used in some optical assemblies. The second approach is the use of a high-density RF/DC connector. An example is Land Grid Array (LGA) interposers which have been used for high-speed ICs and optical modules. The common approach used with LGAs is to clamp the optical module to the PCB, using the heat sink as the top plate. See Figure 1 for an example a high-speed LGA interposer used with an optical module.

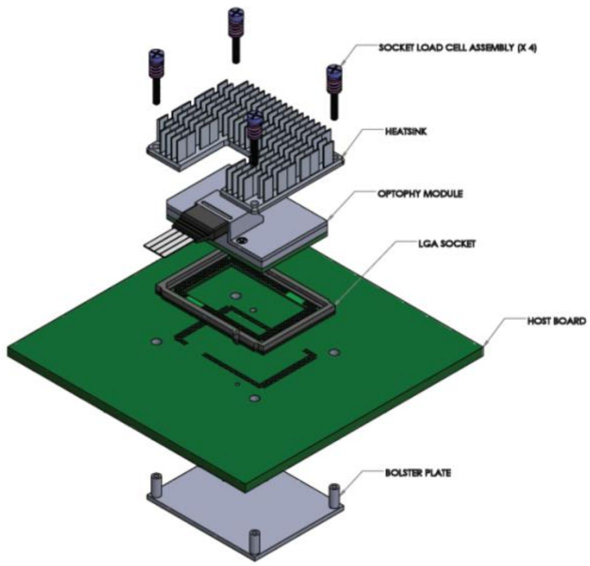

Figure 1: Optical module assembly using an LGA interposer for PCB attach (courtesy of Cisco)

5.2. Low Speed Pins and Signaling

Low speed control and sense signal electrical requirements are as described in the QSFP-DD MSA, section 4.1, including timing for soft control and status functions. ModSelL pin will not be required as it will be included in the SPI bus (SPI chip select or SPI_CS). InitL pin will be required for each CMIS instance on the module, so a 3.2Tb/s module would require a single InitL pin and a 6.4Tb/s module would require two InitL pins. All other low speed control and sense pins (ResetL, LPMode & ModPresL) will require only a single pin per optical module.

5.3. Power Supply Sequencing and Inrush Current

Power supplies will be provided to the optical module from the switch PCB. Due to the module size constraints, it may not be possible to incorporate inrush-current-limiting/enable IC's to each voltage rail. In cases where the optical module cannot incorporate such IC's, the switch will need to enable the power rails with the proper sequence and timing and the module will need provisions to tell the switch when to enable each rail or group of rails. One alternative is for the optical module to include digital I-O pins to tell the switch board when to enable a given supply or group of supplies. Another alternative is that the optical module will set flags in the management interface to inform the switch when to enable a given supply or group of supplies.

Consider an example where six voltage rails are supplied from the switch (0.7, 1.0, 1.2, 1.8, 2.5 & 3.3V) and the module will signal the switch board via I-O pins. These six supplies are arranged into three groups that need to be brought up in sequence by the switch board; group 1 = 3.3V, group 2 = 1.0, 1.8 and 2.5V, and group 3 = 0.7 and 1.2V. Once the switch is powered up it will enable group 1 supplies. After group 1 is brought up, the switch board will wait until all the optical modules on the CPO assembly have indicated through their respective I-O pins that group 2 can be enabled. The switch will then enable the group 2 supplies and this sequence will be repeated for group 3.

5.4. High Speed Electrical Specification

The high-speed electrical interface between the switch IC and the optical modules will be CEI-112G-XSR. This project is still active within the OIF. The expectation is that CEI-112G-XSR will support PCB losses of approximately 8dB at 28GHz.

5.5. Reference Clock

A reference clock signal can be provided to the optical module. If a copy of the reference clock is required, the host will provide copies to facilitate miniaturization of the optical module. The optical module vendor should attempt to minimize the number of ref clock inputs. Preliminary guidance is a maximum of one reference clock per 800Gb/s of capacity (e.g. maximum of four reference clock inputs for a 3.2Tb/s optical module). Table 3 below proposes reference clock specifications to solicit input on requirements.

Parameter	Min	Max	Units	Notes
Frequency	156.25	312.5	MHz	Supports two frequencies shown
Frequency tolerance	-	100	ppm	
Differential swing	800	1600	mV	Peak to peak
Single-ended swing	400	800	mV	Peak to peak
AC coupling	1	100	nF	
RMS jitter	200	-	fs	12 kHz to 1 MHz
	250	-		1 MHz to 20 MHz

	360	-		12 kHz – 20 MHz

Table 3: Proposed reference clock specifications

5.6. Mixed Data Rate Operation

Some applications require that the switch ports operate at different data rates. For example, a fabric switch may be aggregating 100Gb/s streams from TOR switches up to spine switches at 400Gb/s. In such a case, up ports would be configured in 400Gb/s mode and down ports would be configure in 100Gb/s.

Within the optical module, the ability to select the data rate on a per port basis is preferred, but not required. The minimum granularity required for data-rate selection is four ports.

5.7. Low Power Mode or Power Shutdown

In some applications, not all of the optical lanes will be used. Ideally, these unused lanes can be shut down such that they consume no (power shutdown) or little (low-power-mode) power. Within the optical module, the ability to select shutdown/low-power-mode on a per port basis is preferred, but not required. The minimum granularity required for shutdown/low-power-mode is four ports. Max power consumption in low-power-mode is provided in Table 2.

6. Management Interface

The management interface proposed is the same specification used for 400G pluggable modules, QSFP-DD Common Management Interface Specification (CMIS). The current CMIS covers 16 lanes and there is a desire to extend this to 32 for applications like CPO. Optical modules with >32 lanes will require multiple instances of CMIS. For the electrical interface for CMIS on CPO, we are proposing one Serial Peripheral Interface (SPI) for each CMIS instance. A preliminary SPI access protocol proposal is shown below Table 4 below. SPI interface should operate with a clock frequency ≥ 20MHz.

Byte#	Description
0	Command [7]: Transfer direction 0 = Read 1 = Write [6:2]: Reserved [1:0]: Bank select
1	[7]: Reserved [6:0]: # of bytes to transfer – 1 (0=1 byte, 0x7f = 128 bytes)
2	Page number, ignored if register address number < 0x80
3	Register address number
4	Reserved for read set up time
5	Reserved for read set up time
6	Data byte 1
7..133	Data byte 2… (up to 128)

Table 4: Proposed SPI access protocol

6.1. Timing Requirements for Control and Status

The starting point for optical module interface timing requirements will be as specified in the QSFP-DD MSA QSFP-DD Hardware Specification, Section 7.1. Modifications will be required for the CPO application, but this should serve as initial guidance on timing requirement targets. Feedback on specific requirements that cannot be achieved in CPO is sought so we can incorporate this feedback into future revisions of this document.

The management interface timing requirements in Section 6.1 do not apply to CPO, as we are recommending SPI rather than I2C interface.

7. Mechanical Considerations

7.1. Mechanical Overview

The CPO assembly consists of a high-density organic substrate, switch IC and optical modules arrayed around the perimeter. An example of a CPO assembly is shown in Figure 2 for reference; it is not intended to represent the preferred embodiment.

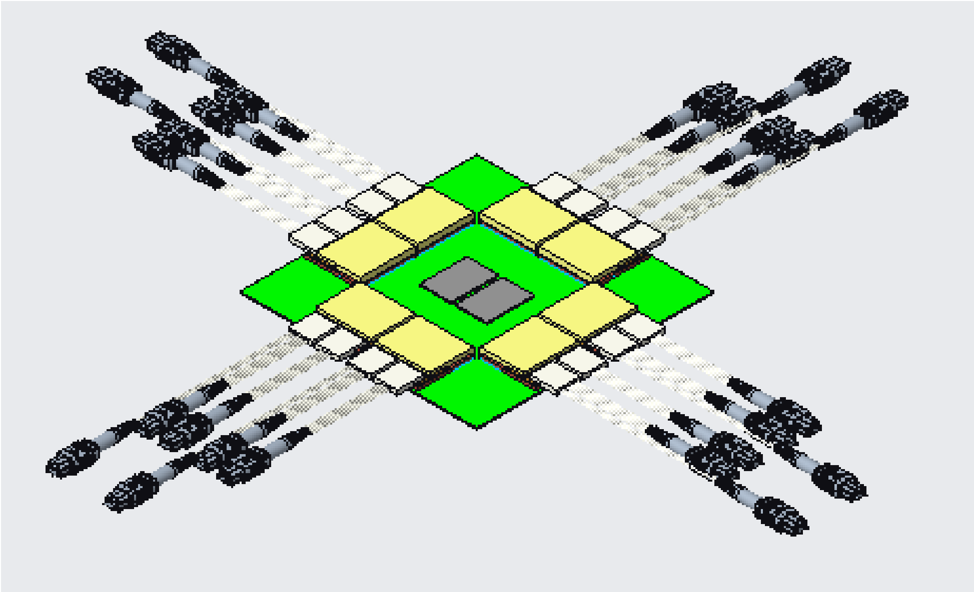

Figure 2: Example embodiment of a 51Tb/s CPO assembly

The maximum size of the organic substrates available will evolve over time; we anticipate a range of 100 to 150mm per side for first-generation CPOs. One of the main intents of CPO is to minimize the length of the switch to optical module electrical interface to save power as SERDES speed and radix increase.

The size of the optical modules will ultimately be determined by the switch IC size, max trace length the electrical interface will support, number of optical modules used (e.g. 16x 3.2Tb/s or 8x 6.4Tb/s or 4x 12.8Tb/s) and maximum size of the substrate. The intent of the guidance document is not to define a mechanical form-factor, but rather provide insight into the over-arching considerations. In the example shown in Figure 2, eight 6.4Tb/s optical modules, with a top surface (in yellow) measuring ~30x35mm each are shown arrayed around the perimeter of the switch IC attached to 150x150mm substrate.

7.2. Optical Connectors

The optical connectors provide a means to connect the optical module to the front plate of the host switch. The optical module may be pigtailed or have an integrated optical connector.

The pig-tailed version would have SMF ribbons for the Tx and Rx terminated in a high-density optical connector. For example, a 16x 400GBASE-FR4 (6.4Tb/s) module would have 16x fibers for the Tx and 16x fibers for the Rx functions. Here two 1x16 parallel fiber arrays exit the module (one Tx one Rx) and each is terminated in a 1x16 MPO connector. The MPO connector would be attached to a bulk-head adaptor on the switch faceplate. The type of high-density connector used on the switch faceplate will vary by switch design.

For a module with an integrated optical connector, an optical patch cord would be used to connect the optical module to the faceplate of the switch. High-density optical connectors available today (e.g. MPO) are not suitable for CPO applications. It is likely that new optical connector scheme will need to be developed to make this approach practical. The JDF is interested in learning about candidate designs for further investigation.

For optical modules with high lane counts using a single-fiber/lane optical standard, there will be a large number of fibers attached to the optical module. For example, a 16x 400GBASE-DR4 (6.4Tb/s) module would have 128 fibers attached for Rx and Tx functions. For high-fiber count applications, dense fiber connection schemes (e.g. MCF or 3D waveguides) are desired and the CPO JDF is interested in proposals on how to address such cases.

7.2.1. Fiber Connections for External Laser Source

Some CPO designs will include an External Laser Source (ELS) to provide a CW light source to the optical modules. In such cases, each optical module will require a minimum of one fiber per wavelength connected to the ELS module. The number of ELS fibers required will depend on the optical module requirements and the power/fiber launched from the ELS.

For example, a CPO design with eight 6.4Tb/s optical modules (using 400GBASE-FR4) would require ELS's with 32 fibers, with four fibers routed to each optical module. In the example shown in Figures 2 & 3, the Tx side of the optical module has a 24-fiber array, in which eight fibers (two per wavelength) are routed to the ELS module and the remaining 16 are used for the Tx function.

7.3. Thermal Considerations

The power density (W/cm^2) and temperature for CPO optical modules and switch IC will set benchmarks for photonic assemblies and may necessitate liquid cooling. It is desired that a single cooling solution (e.g. heatsink or cold plate) be used to cool the optical modules and the switch IC. To interface with the cooling element, the optical modules will incorporate a heat spreader on the top surface. A Thermal Interface Material (TIM) will be used between the cooling element and the module's heat spreader. To maximize cooling efficiency and minimize module case temperature, the optical module's heat spreader height should be as uniform as possible across the CPO assembly. In the example shown in Figures 2 & 3, a heat spreader of 3.5x2.8 cm acts as the thermal interface, with a power density of ~13W/cm^2.

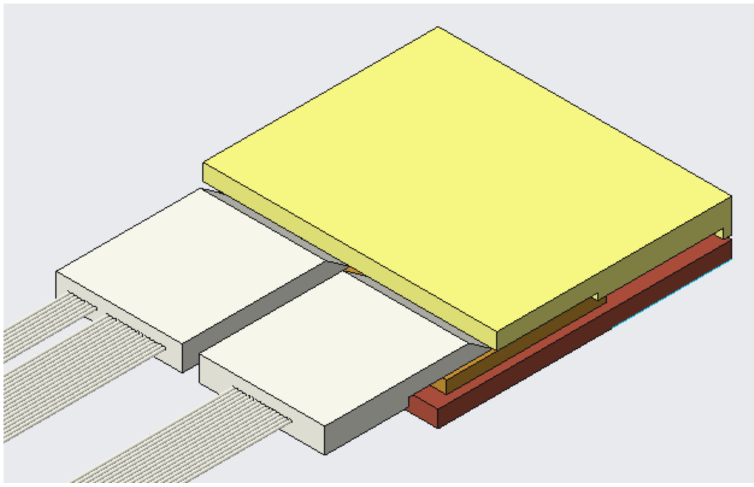

Figure 3: Example of an optical module for CPO applications

8. Environmental Conditions

The values in Table 5 below are for general guidance and discussion and subject to change.

Parameter	Symbol	Min	Typical	Max	Unit	Note
Altitude, during operation, to sea level		-50		1800	m	
Acoustic, during operation		55			dBA	
Vibration, during operation				TBD		Note1
Altitude, storage, to sea level				5	km	
Altitude, transportation, to sea level				12	km	
Gaseous contamination						Note2
Note: 1. Refer to IEC 68-2-36, IEC 68-2-6 2. Conform to Severity Level G1 per ANSI/ISA 71.04-1985.						

Table 5: Environmental conditions

9. Quality and Reliability

Detailed quality and reliability requirements will be tailored by customer and program. High level guidance for optical modules used in CPO applications is provided below.

9.1. Environmental Stress Testing

As described in Telcordia GR-468-CORE section 4.2 for Integrated Modules.

9.2. Service Life

Maximum service life will vary between five and seven years.

9.3. Operational Excellence

Specifications for reliability and operations excellence must be designed for and demonstrable to work in the cloud environment.

An operational model of a field replaceable unit that can be replaced with minimal training and inclusive of designs that can survive failures with minimal scope of impact.

10. Revision History

The revision history of this document is captured in Table 6 below.

Revision	Date	Description of Changes
1.0	September 17, 2019	Initial release

Table 6: Revision history