Open Eye MSA

Systems White Paper

Tom Palkert Macom, Mark Kimber Semtech, Greg Le Cheminant Keysight, Arlon Martin Samtec

3/4/2020

1 Table of Contents

2 Summary

The Open Eye MSA was formed with the goal of relaxing the industry standard optical specifications to enable reduced power, latency, size and cost of high performance optical modules. For example, eliminating TDECQ testing reduces cost and design complexity and still provides an IEEE compatible link budget.

3 Intro to Open Eye MSA

The Open Eye MSA is an industry group formed to define optical module specifications that provide the optimum port bandwidth, power, latency and density for next generation optical switches. The specifications developed by the MSA will leverages the industry's existing test methodology for measurement of eye diagrams and BER. Modules that comply with the Eye Opening specifications will interoperate across multiple venders. Participants in the MSA include Clock and Date Recovery (CDR) IC suppliers, optical module suppliers and system developers. The MSA specifications will support both existing module types (SFP, QSFP) and higher density emerging module types (SFP-DD, DSFP, QSFP-DD, OSFP, Co-packaged Optics)

The MSA work was initiated in response to large data center requirements for higher speed, higher density, low latency optical module solutions. The existing optical specifications for these next generation module designs hampers the adoption of the latest technology by requiring high power components and complex test methodologies that are typically utilized in long haul telecommunications and/or enterprise applications.

3.1 System Requirements

Next generation data center requires a doubling of both speed and density to achieve the throughput required by Web 2.0 applications. This goal must be achieved within the limitations of low latency, cost and thermal constraints.

Figure 3-1 and Figure 3-2 show a data center architecture utilizing Middle/End of Row switches (MoR/EoR). This architecture requires the use of optical connections between the switch and server. These optical links replace the previous use of copper cable interconnects between a server and Top of Rack (ToR) switch. Because the passive copper cables dissipated no power and had very low latency, it is important that the replacement optical modules have the lowest module power and latency possible. The Open Eye MSA will achieve these goals, enabling the desired transition to MoR and EoR data center architectures.

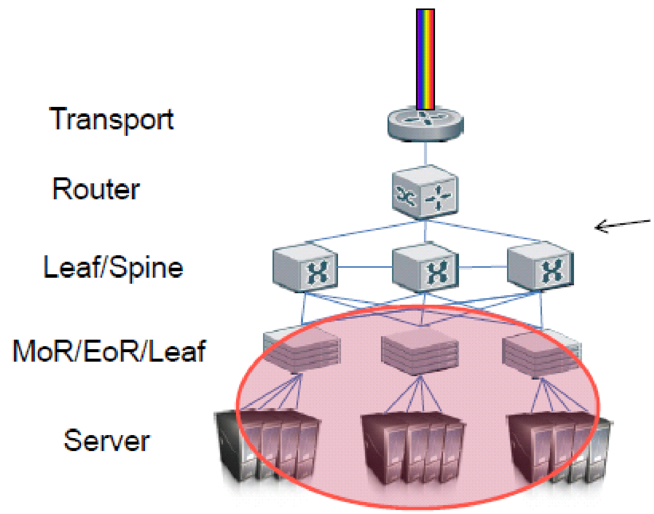

Figure 3-1: Data center server/MoR/EoR/Leaf interconnects

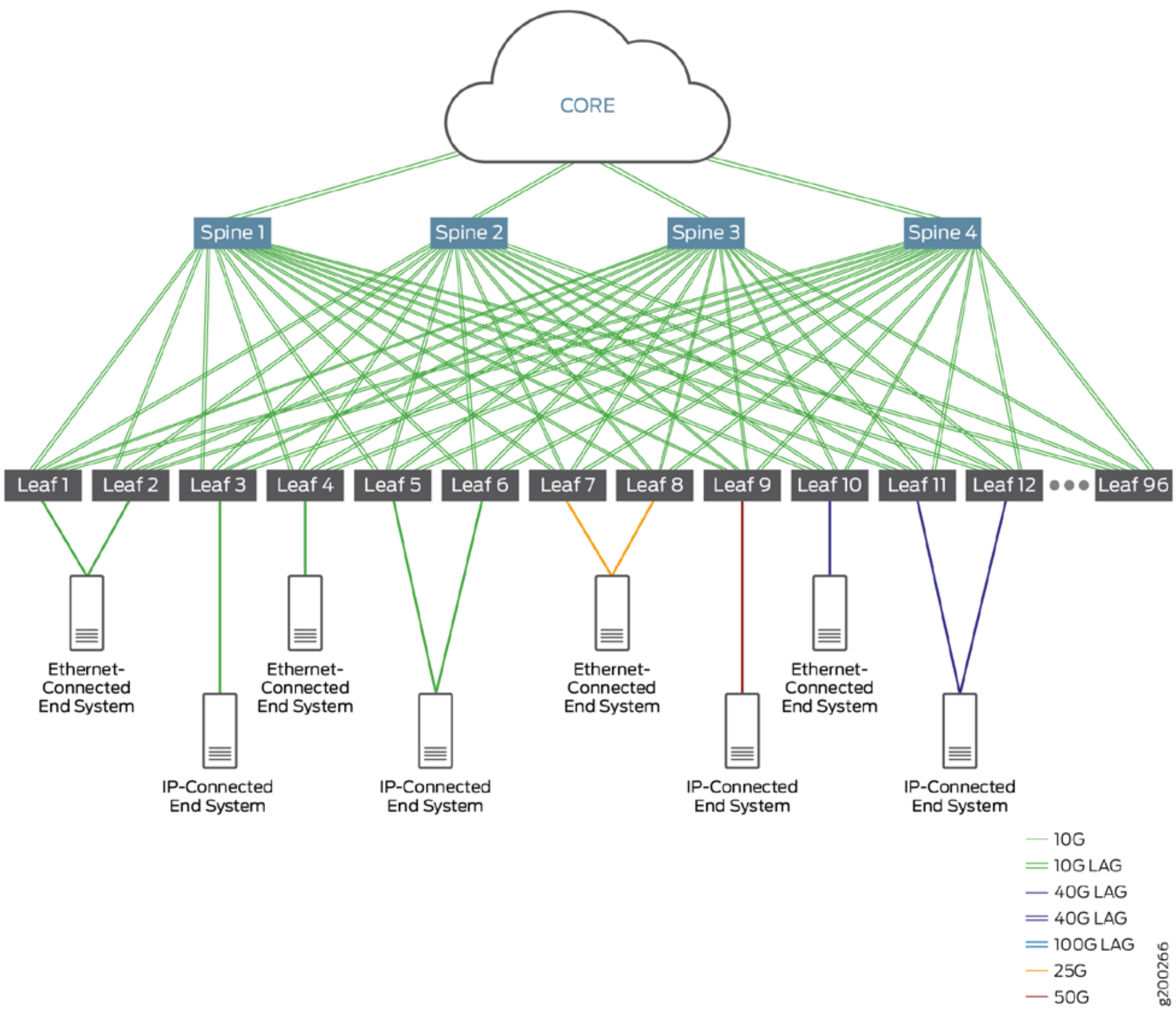

Figure 3-2: Data Center Leaf and Spine interconnects (Courtesy Juniper Networks)

3.2 Thermal requirements

Thermal requirements for data centers force the reduction of power of higher bandwidth modules in legacy form factors such as SFP and QSFP in addition to enabling higher density modules such as QSFP-DD and OSFP. The Open Eye MSA specifications will allow optical module makers to achieve lower power in next generation designs. (For details on thermal requirements see the Open Eye MSA thermal white paper) Figure 3-4 shows the layout of a typical data center network switch. The use of front panel pluggable modules places a large amount of power dissipation into the faceplate of the switch. This use of stacked pluggable module cages places additional thermal constraints on the module. Any reduction in module power dissipation provides large system benefits when applied across a 36-48 port switch. The Open Eye MSA provides a large reduction in module power dissipation that will enable large port count next generation network switches using simpler cooling techniques. This provides network switch providers a large cost savings by eliminating the requirement for advanced cooling features such as liquid cooling.

Detailed thermal analysis can be found in the Open Eye Thermal white paper. A summary calculation of the power savings provided by the Open Eye MSA 200G-FR4 module is shown in Table 1.

Figure 3-3: Block diagram of Open Eye module design vs. TDECQ module design showing power savings

Table 3-1: Module Power Comparison

Module Type	Module power (Watts)	Optical CDR type
100GBase-CWDM4	3.0-3.5	Analog
200GBase-FR4 Open Eye	3.5-5.5	Analog
200GBase-FR4 IEEE	5.5-7.5	DSP

Table 3-2: Data Center Power Savings with Open Eye Modules

Data Center size (sq ft)	Average annual Cost per KW*	# of Modules/data center	MSA module power savings (W)	Total annual cost savings with Open Eye modules
500-5000	$26,495	5,000	5,000	$132,475
5,000-10,000	$13,662	15,000	15,000	$204,330
10,000-25,000	$8,464	50,000	50,000	$423,200
25,000-50,000	$6,734	125,000	125,000	$841,750
>50,000	$5,467	175,000	175,000	$956,725
Hyper scale	$4,500	250,000	250,000	$1,125,000

* Cost to Support Compute Capacity; Poneman Institute Research Report; August 2016

IEEE compliant 200GBASE-FR4 modules using the DSP-based receivers are estimated to dissipate 7W. This includes the DSP, Optical EML 4 lane driver, 4 lane TIA and additional circuitry.

The Open Eye MSA compliant 200G-FR4 module is estimated to dissipate 4.5W. This includes the analog retimer ICs, Optical EML 4 lane driver, 4 lane TIA and additional circuitry. The power limit of the QSFP module is 5W. This maximum power limit is based on the use of traditional heat sink and fans providing air flow cooling. The 7W module will require additional thermal features to handle the additional power dissipation. A high bandwidth switch with 6.4 Tb/s of bandwidth requires 32 QSFP ports with a total power dissipation of 4.5W x 32 ports = 144W. The identical switch design populated with IEEE 200GBASE-FR4 modules dissipates 7W x 32 ports = 224W.

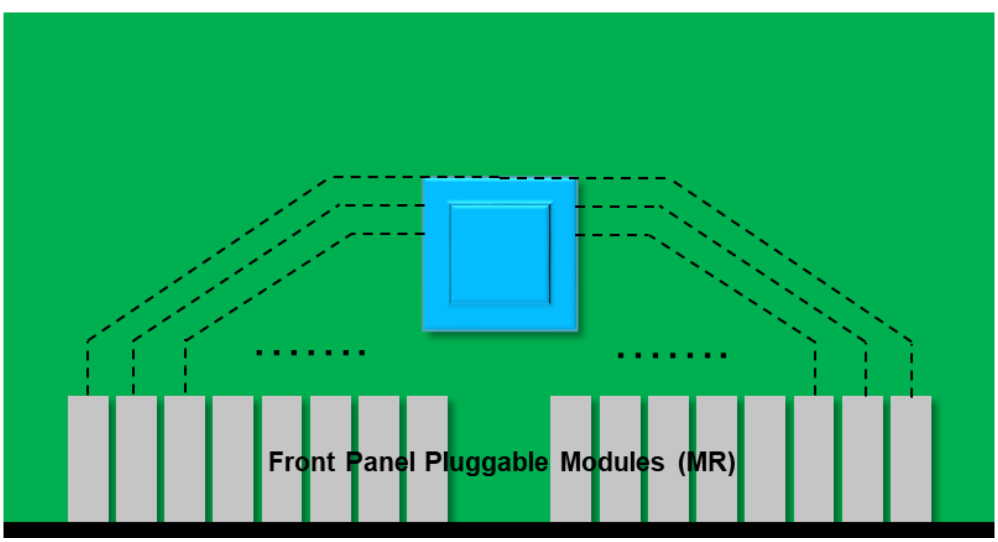

Figure 3-4: Data Center switch port density

3.3 Latency requirements

The latency of a data center connection affects the overall performance of the network. The use of DSP technology in an optical module can add up to .36 us of latency to every link. As optical links replace copper cable connections to the server, the latency of the optical link will become more important and the use of the Open Eye MSA optical module will provide a large latency improvement compared to DSP based modules.

3.4 Open Eye MSA variants

The initial specifications will define interoperability at 50Gbps PAM4 data rates over single-mode fiber (SMF) for Leaf/Spine Interconnects. Future work will include multimode fiber (MMF) connections for Server/MoR/EoR interconnects. Higher speed specifications will be developed to support 100G serial interconnects. Examples of specifications include:

- 1x50Gbps, 4x50Gbps and 8x50Gbps 2 km and 10km over duplex SMF
- 4x50Gbps over parallel single-mode (PSM)
- 1x100Gbps, 4x100Gbps and 8x100Gbps 2 km and 10km over duplex SMF
- 1x50Gbps and 4x50Gbps 100m over MMF

Note: Complete Open Eye Roadmap can be found at openeye-msa.org

3.5 Open Eye Specification Overview

The Open Eye specification covers the following items:

a) Optical Tx interfaces including Tx launch power, jitter, noise and linearity.

b) Optical Rx interfaces including Rx input power and stressed receiver sensitivity

c) Channel characteristics (Link budget)

d) Fiber specifications

Note: Detailed Open Eye specifications can be found at openeye-msa.org

A reference diagram showing the design of an Open Eye Optical module and the connections to a host system is shown in Figure 3-5. Retimers are used in both directions of data traffic.

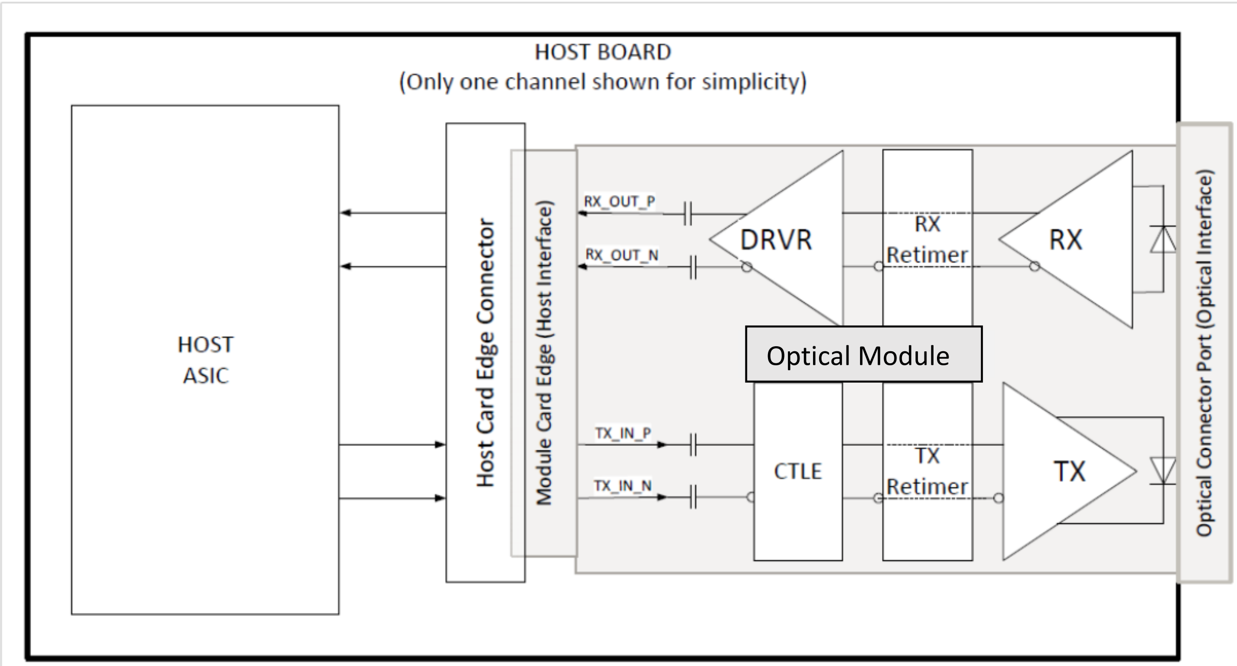

Figure 3-5: Optical Module Reference Diagram

3.6 Compatibility with Electrical and Management Specifications

The Open Eye MSA specifications will be compatible with existing optical module Electrical and Management Specifications. Supported specifications include:

- 50GAUI, 100GAUI-2
- 200GAUI-4
- 2x200GAUI-4
- 100GAUI

The electrical interface design is shown in Figure 3-6. This shows a host switch or NIC IC connected to an Open Eye MSA compliant module over a route of 4-8 inches of printed circuit board or up to 1 meter of cabled host. This electrical interface is compliant with IEEE, OIF and Fibre Channel standards.

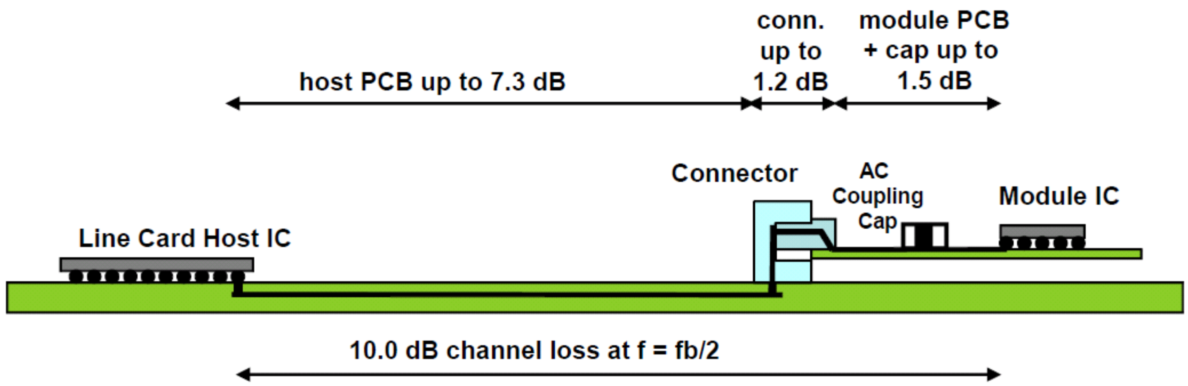

Figure 3-6: VSR/C2M chip to module electrical interface diagram

Management interface support includes:

- SFF-8472 (SFP module types)
- SFF-8636 (QSFP module types)
- CMIS (QSFP-DD, OSFP module types)

3.7 Test/TDECQ issues

Test procedures specified in IEEE documents for PAM4 based optical transmitters have enabled the needed interoperability across multiple module vendors. However, the majority of the burden of system equalization is placed on the module receiver in order to accommodate the widest variety of transmitter performance. This variety of transmitters would include low bandwidth types with minimal pre-emphasis and others having large over/undershoots and other eye impairments. A complex DSP-based receiver is required to interoperate with this wide range of transmitter performance. The result is high power, increased latency and higher cost optical modules. This higher power, latency and cost has limited the adoption of next generation optical modules in large data center applications.

This has also resulted in a complex test methodology where a transmitter is observed with a virtual equalizing receiver that must be optimized for lowest system level power penalty. If transmitter performance can be restricted to higher quality 'open eyes', both the receiver architecture and transmitter test methods can be simplified. The Open Eye MSA was initiated in response to these data center requirements.

Transmitter dispersion and eye closure quaternary (TDECQ) is a test method developed in the IEEE 802.3bs project to assess the effective power penalty of a transmitter due to inherent eye closure and channel dispersion. Noting the use of a DSP-based receiver as mentioned above, the eye closure is observed using an oscilloscope after passing the signal through a virtual 5-tap feed-forward equalizer. While the TDECQ measurement can be easily executed, optimization of the equalizer tap settings to minimize the TDECQ penalty is considered complex and adds to overall test time. Figure 3-7 shows that all 50G SMF modules tested with the TDECQ methodology result in a value of Ceq (equalizer noise gain) unity or near 0 dB, indicating equalization is not strong. This implies that the module Rx complexity and test complexity required by the IEEE TDECQ reference receiver is not required for interoperability in the vast majority of cases.

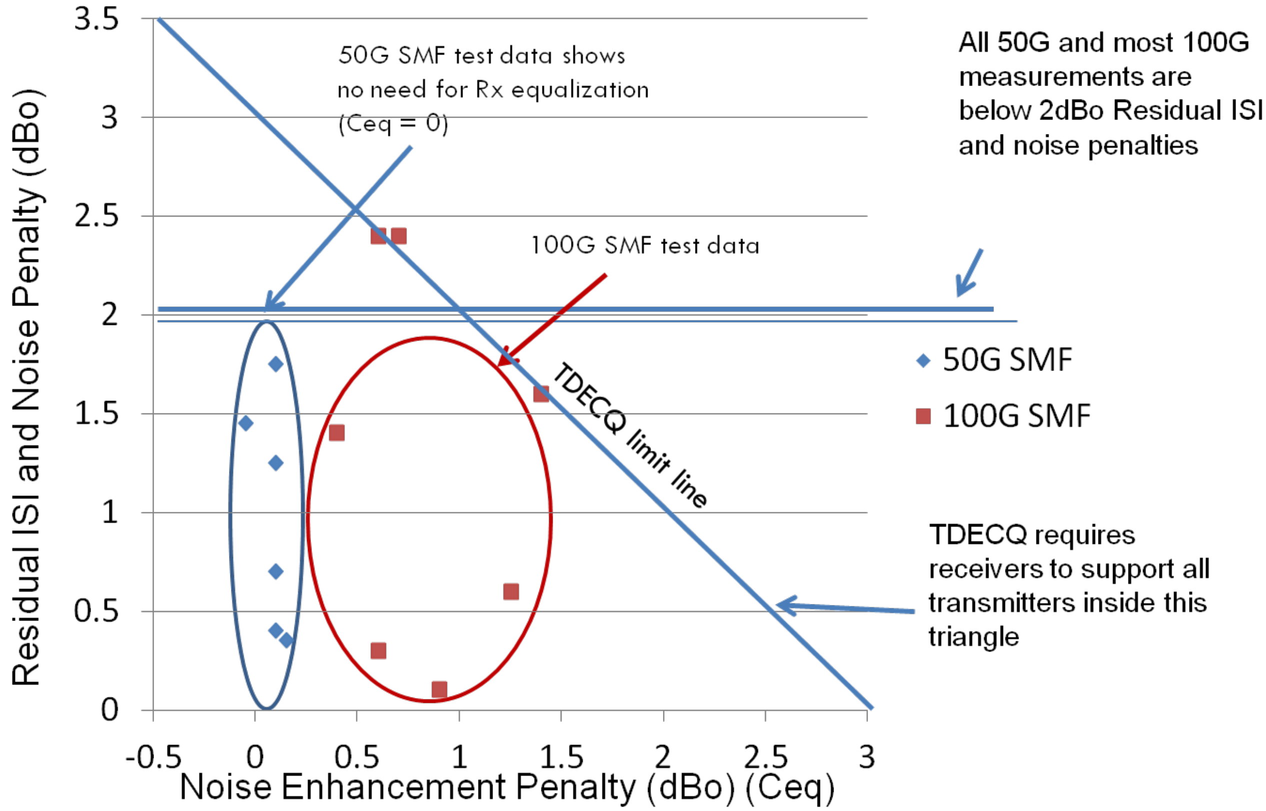

Figure 3-7: TDECQ measurement test results

3.7.1 TDECQ reference receiver

The TDECQ reference receiver complexity, if required in real receivers, results in higher component cost and higher power consumption. In addition it limits implementations using lower cost process technology choices.

Fundamentally, the TDECQ test methodology was specified to allow the use of low bandwidth Optical Tx components and assumes at least 5-tap FFE equalization in the Rx. The use of DSP technology was assumed to be readily available with low cost and low power designs enabled by advances in IC process technology. These advances have not been realized and although DSP enabled systems are being deployed for Coherent optical modules, the cost and power constraints of large data centers have not been achieved.

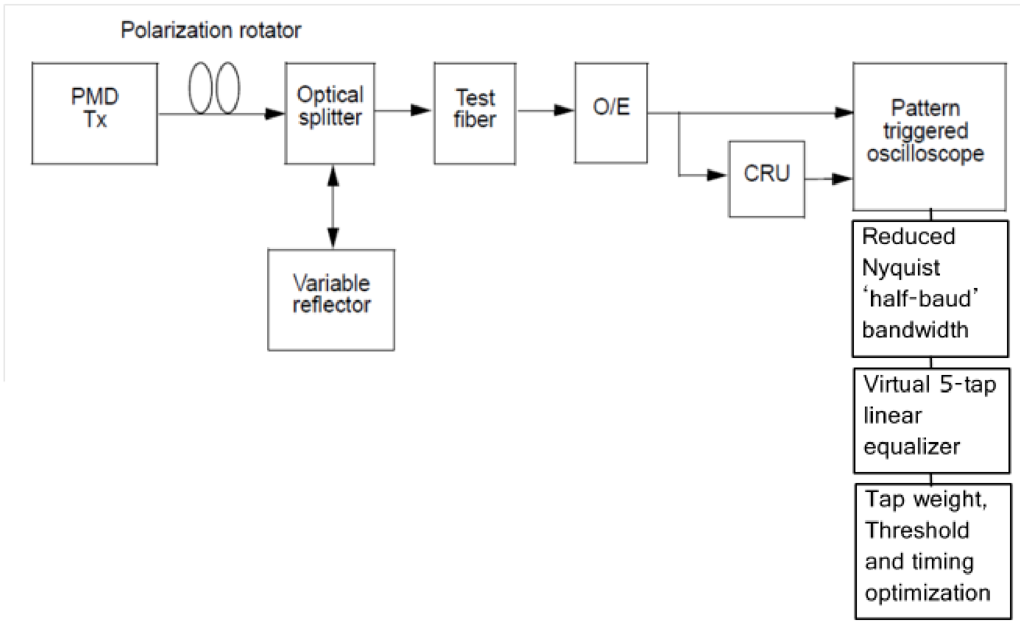

Figure 3-8: TDECQ measurement test setup

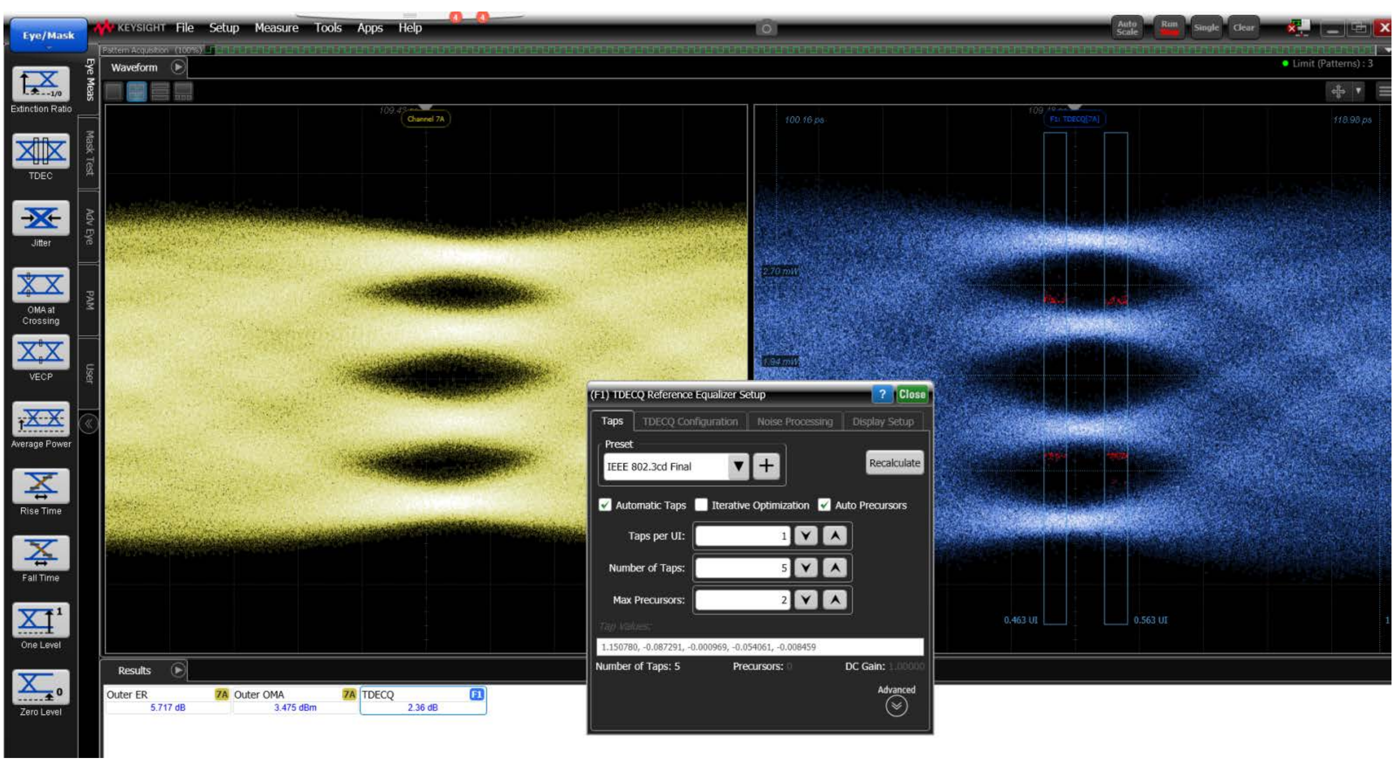

Figure 3-9: TDECQ requires optimization of a 5-tap FFE prior to eye analysis. The transmitter signal (yellow) passes through a virtual equalizer and the equalizer output waveform (blue) is measured

3.7.2 Open Eye reference Rx

The Open Eye MSA reference receiver is designed for ease of testing and design. Figure 3-9 shows a block diagram of the test setup for the Open Eye Tx output testing and does not include virtual equalization required for TDECQ.

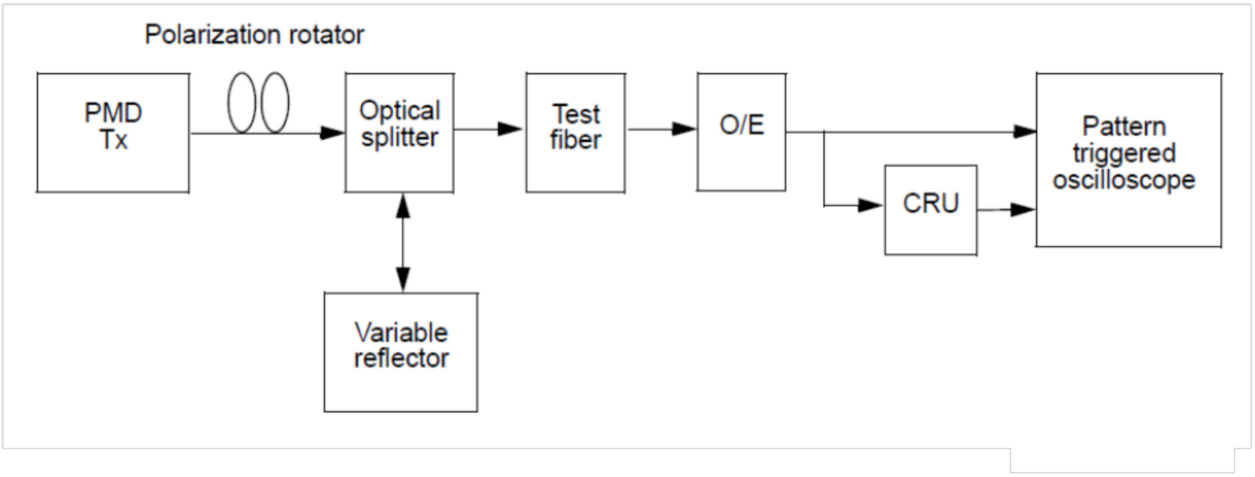

Figure 3-10: Open Eye MSA Tx test set up

The Open Eye approach is based on distributing overall link equalization in a balanced manner between transmitters and receivers allowing for lower power analog implementations. An Open Eye transmitter is required to meet its eye mask without equalization and a receiver is tested based on the worst case transmitter that can achieve the eye mask. A receiver is only required to overcome its own internal impairments and is not expected to compensate for transmitter eye closure and thus can be a simpler, low cost, low power design.

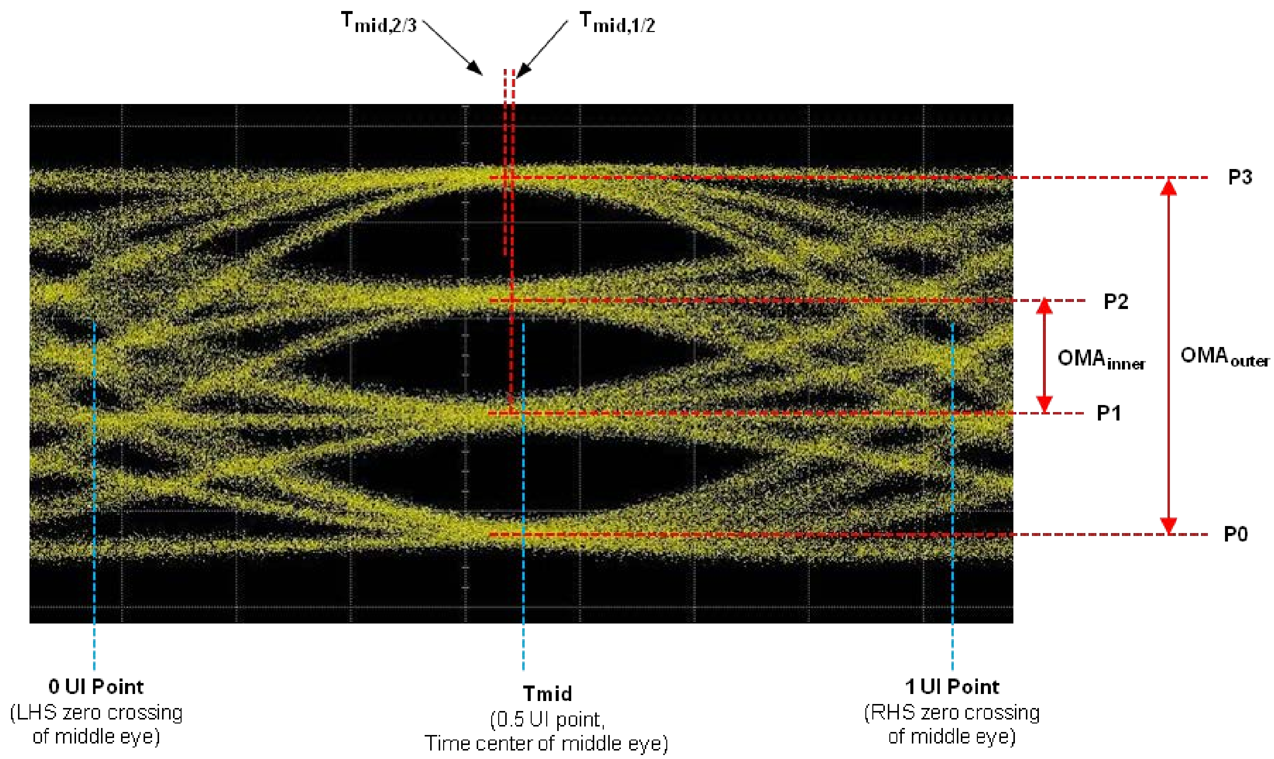

Figure 3-11: Eye Opening test method

4 System design

The design of a system supporting the Open Eye MSA for 50G 2km application is shown in Figure 4-1. A feature of the Open Eye MSA is that it will interoperate with the existing IEEE compliant modules if the Tx output is tested to the Open Eye specification.

4.1 Compatibility with IEEE link budget

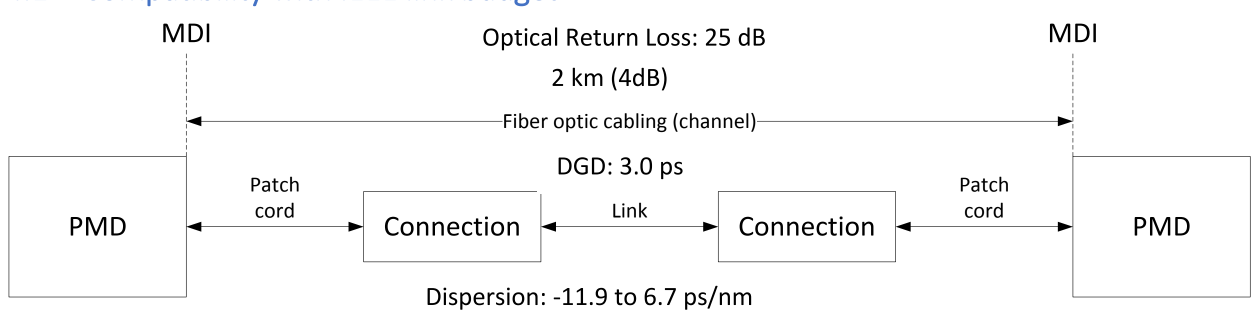

Figure 4-1: Example using IEEE 200G-FR4 link budget

5 Conclusions

The Open Eye MSA allows the most flexibility in module and system design. This is achieved by specifying the optimum separation of equalization between transmitter and receiver. In addition, the MSA enables a wide selection of technological approaches for transmitters and receivers leveraging future advances in technology including Silicon Photonics and higher bandwidth EML/DML lasers.

The Open Eye MSA will reduce power, latency, size and cost in high performance optical modules used in large data center applications. This is achieved by avoiding the complex optimization algorithm for the 5 tap FFE reference equalizer during transmitter testing.

Modules that comply with the Eye Opening specifications will interoperate across multiple venders. The MSA specifications will support both existing module types (SFP, QSFP) and higher density emerging module types (SFP-DD, DSFP, QSFP-DD, OSFP, OBO and Co-packaged optics,)
In summary the Open Eye MSA provides the following advantages to system designs:

- Better interoperability as the transmit waveform is well defined
- Lower latency and lower latency variation
- Compatible with installed Fiber Infrastructure
- Compatible with industry electrical interfaces