Development of Smallest-in-Class 400G Network Tester

Toru Miyauchi, Takaya Abe, Yasuji Ishizuka, Akinobu Tsuyuki

[Summary] 400G Ethernet is a key technology for implementing high-speed, large-capacity communications following the widespread rollout of 5G services and increasing use of Cloud computing applications. Installation and mainte-nance (I&M) of these networks is driving high demand for a compact, portable, 400G tester supporting easy onsite measurement. We have developed the B5-size, smallest-in-class, MT1040A network tester to lighten the burden of efficient network onsite I&M.

1 Introduction

Communications traffic is increasing continuously with the spread of 5G and Cloud computing services. Furthermore, the sudden increase in remote working, online learning, etc., is causing bandwidth issues for network operators. 400 Ethernet[1)] is a key technology for strengthening communications network bandwidth, which is driving demand for a compact, portable 400G Ethernet tester targeted at onsite network installation and maintenance (I&M).

Unlike earlier Ethernet technologies transferring two data bits per unit time, 400G Ethernet uses 4 Pulse Amplitude Modulation (PAM4) technology transferring four data bits (0 to 3) per unit time. Although PAM4 can double modulation speeds at low cost, it is easily adversely affected by noise. Consequently, PAM4 uses a built-in Forward Error Correction (FEC) function as standard. Additionally, as shown in Figure 1, PAM4 is used not only by Ethernet networks but is also used by Common Public Radio Interface (eCPRI) mobile networks, Optical Transport Networks (OTN) and Synchronous Optical Network/Synchronous Digital Hierarchy (SDH) core networks, as well as other Fibre Channel (FC) networks used by data centers and video streaming services.

Under the circumstances, we have developed the all-in-one Network Master Pro MT1040A and 400G Multirate module 400G MU104014A/15A (QSFP-DD/OSFP) options to support evaluation of each of these communication systems. The tester features are outlined below.

Compact:

- The MT1040A is the world's smallest-in-class tester for speeds from 10 Mbit/s to 400 Gbit/s accommodating direct installation of QSFP-DD and OSFP optical transceivers.

Versatile:

- Freely stackable module combinations (Figure 2) support the optimum configuration for both onsite and laboratory benchtop testing.
- Each port supports independent measurement of multiple communications technologies, including Ethernet (<100G), OTN, eCPRI/RoE, FC, etc.
- Combination with the optional OTDR module supports optical fiber testing.

Powerful:

- The 400G Ethernet FEC analysis function is built-in as standard.
- There is an easy to use optical transceiver check function.
- Support tools for automated one-button testing help increase job productivity[2)].

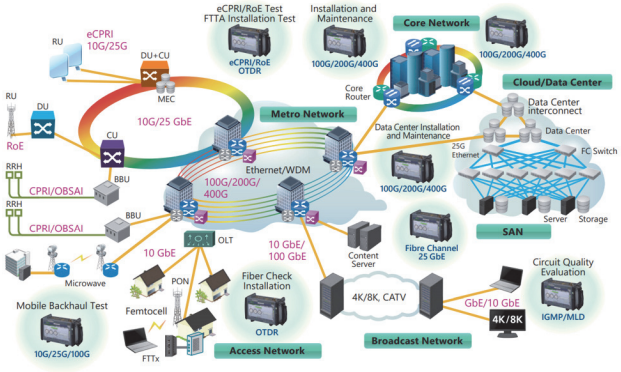

Figure 1 Network Communications Technologies and Speeds

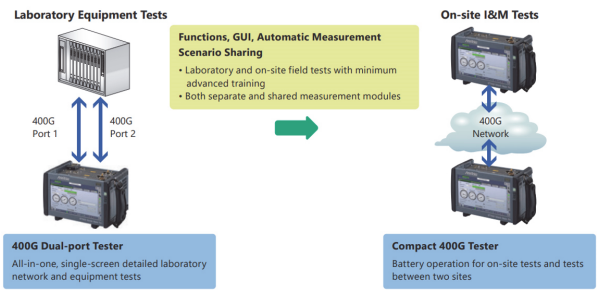

Figure 2 Flexible Test Equipment Configuration

2 400G Ethernet

This section introduces the layer-1 and layer-2 (PAM4 and FEC) technologies used by 400G Ethernet.

2.1 PAM4

400G Ethernet optical transceivers use PAM4 rather than the earlier Non Return to Zero (NRZ) technology. As shown in Figure 3, PAM4 has a higher transmission capacity than conventional amplitude modulation because 4-bit signals are transferred using four voltage levels (0, 1, 2, 3). Since NRZ transfers two bits of data using two voltage levels (0, 1), PAM4 has the merit of doubling the data rate compared to NRZ.

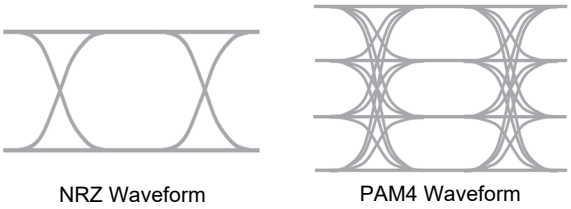

Figure 3 NRZ and PAM4 Waveforms

However, since PAM4 has an Eye opening amplitude of 1/3rd that of NRZ, the signal to noise ratio (SN) easily degraded by noise.

2.2 FEC

Since the 400G Ethernet PAM4 technology is susceptible to bit errors caused by noise, network errors must be corrected using FEC to recover the original signal condition.

The FEC-applied communications data is divided and sent in units called Codewords and errors are corrected in Codeword units when the data is received.

400G Ethernet uses RS-FEC (544,514) encoding defined by IEEE 802.3 (Figure 4). Each Codeword is composed of a total of 544 symbols formed from 514 data symbols and 30 parity symbols and each symbol has 10 bits.

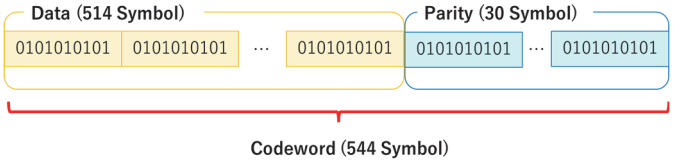

Figure 4 FEC Frame Structure

If the Symbol Error count is 15 or less, all errors in the encoded signal can be corrected. Error correction fails if there are 16 or more errors, and data in Codeword units–including symbols without errors–are abandoned. Error-corrected Codewords are called Correctable Codewords, and Codewords that could not be error-corrected are called Uncorrectable Codewords.

3 Network Master Pro MT1040A

3.1 MT1040A Outline

The MT1040A is a portable tester for evaluating the communications quality of networks operating at speeds of 10 Mbit/s to 400 Gbit/s. It accommodates a maximum of two MU104014A/15A modules for evaluating 400G Ethernet networks. In addition, for portability it supports battery operation with one MU104014A installed, offering the world smallest-in-class tester with the same B5-size as the MT1000A. Moreover, it is fully compatible with the 100G Multirate Module MU100011A and the High Performance GNSS Disciplined Oscillator MU100090B modules for the MT1000A. Figure 5 shows the external view with the MU104014A and MU104015A modules installed.

Figure 5 External View of MT1040A

3.2 Large Power Supply

Since up to two MU104014A/15A modules can be installed, the design supports supply of more power to the last-stage module in comparison to the earlier MT1000A. Furthermore, the portable design features listed below support module hot- swapping without operation errors.

- Use of push-on type connectors physically prevents loss of connector contact.
- Hardware detection of contact connection/disconnection prevents abnormal heating due to increased contact resistance.
- Users are informed about an operation abnormality when the load is instantaneously reduced if the measured PSU voltage drops when the current changes before and after installing the MU104014A/15A.

Solving these technical challenges has facilitated Ethernet testing from low 10-Mbit/s speeds to high-power-consumption 400-Gbit/s speeds.

4 400G Multirate Module MU104014A/15A

4.1 MU104014A/15A Outline

The MU104014A/15A are compact, lightweight modules with multiple interfaces supporting speeds up to 400 Gbit/s. They have electrical interfaces for various SFP28, QSFP28, QSFP-DD, and OSFP-compliant transceivers as well as an RJ45 interface, and are designed for protocol tests of 400G/100G/40G/25G/10G/1G/1000M/100M/10M Ethernet networks, and OTN, FC, eCPRI, etc., transport networks. The MU104014A has a 400G QSFP-DD port and the MU104015A has a 400G OSFP port. Figure 6 shows the external appearance of the MU104014A.

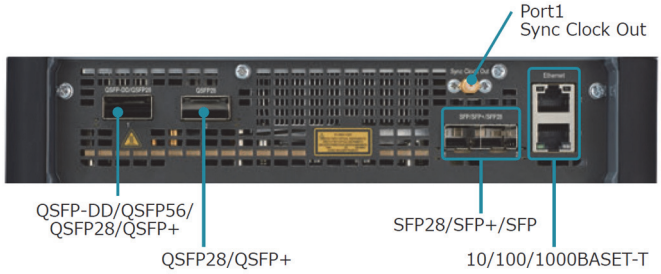

Figure 6 MU104014A Interfaces

In addition to the measurement interfaces, there is also a Sync Clock Out connector used as a trigger signal for an external sampling oscilloscope to monitor optical waveforms output from optical transceivers connected to each interface.

4.2 56-Gbit/s Electrical Interface Specifications

Each lane of the QSFP-DD/OSFP electrical interfaces supports transmission speeds of 56 Gbit/s. The 56-Gbit/s electrical interface is electrically and mechanically compliant with the Multi Source Agreement (MSA)[3] and Small Form Factor (SFF) optical transceiver standards as well as with Optical Internet working Forum (OIF) CEI-56G-VSR[4] recommendations.

4.3 Compact Structure and Exclusive Configuration of OSFP/QSFP-DD

The MU104014A/15A are composed of the main PC board such as the FPGA, and an interface board for inserting OSFP and QSFP-DD optical transceivers; the measurement module configuration can be switched simply by changing the interface board. As a result, this design has the advantage of supporting new future electrical interfaces simply by changing the interface board. Figure 7 shows the block diagram. The main and interface board power and logic signals are connected by a stacking connector while high-speed signals are connected via a multipole coaxial cable. However, inconsistent impedance occurring at the cable-connector lands conventionally seriously affects 56-Gbit/s PAM4 signal waveforms. Moreover, overcoming the limited available space caused by the small design constraints required increasing the PC board wiring density, which in turn requires consideration of crosstalk effects. As a result, 3D analysis tools were used to design a PC board with the following features to minimize degraded waveforms.

- Shorter wiring
- Removing unnecessary internal wiring
- Optimizing Printed Circuit Board via format using skip vias and back drilling
- Optimizing connector land shapes

Figure 8 shows the designed PC board wiring pattern to the multipole coaxial connector from the QSFP-DD electrical connector, and Figure 9 shows the wiring 3D analysis results (S-parameter). Since 56-Gbit/s PAM4 signals transfer 2 data bits per one clock, the transfer rate is 28 Gbaud. As a result, although the board frequency is 14 GHz, it satisfies the CEI-56G-VSR-specified system loss of not more than 10 dB.

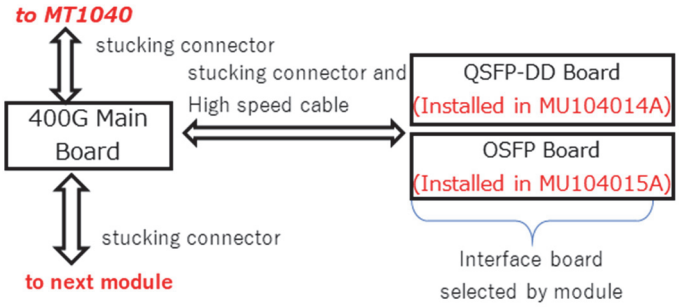

Figure 7 MU104014A/15A Block Diagram

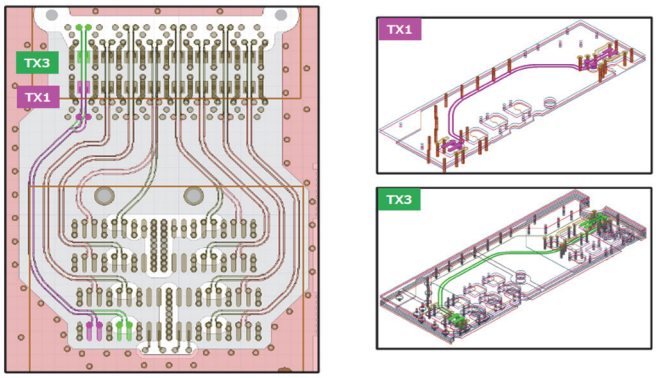

Figure 8 3D Analysis Model

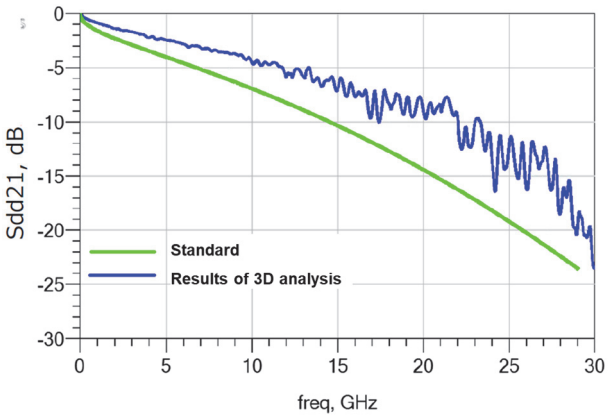

Figure 9 Analysis Results

4.4 Cooling

Dissipating heat and cooling hot QSFP-DD and QSFP28 optical transceivers and the FPGA has been a challenge. It was impossible to use a large cooling fan with heat pipes and copper heat sink due to the compact design size. As a result, we adopted the following heat dissipation and cooling design concepts using flow simulations for the structural design.

- Use small fan moving large air volume
- Create efficient air flows divided between optical module and FPGA sections
- Design air flows to pass through optical-module heat sink
- Use lightweight heatsink with low thermal resistance

Figure 10 shows the simulation results indicating that the final design suppresses internal temperature increases. In addition, we confirmed that the temperature increase in actual use is held below the upper temperature limit.

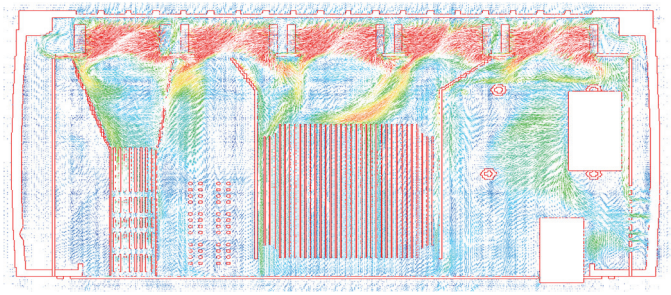

Figure 10 Thermograph Analysis Results

4.5 400G Ethernet Frame Processing

To support large data capacity, the Ethernet frame processing uses a Multi-Mac block structure deploying four 100GbE MAC blocks. This structure was adopted to avoid complex frame processing due to expansion of the internal bus width, making future expansion to 800G easier. The details of Single MAC and Multi-MAC block processing are described below.

(1) Single MAC Block Processing

This method performs serial processing even if the Tx and Rx frame sequence is lost. However, with 400G Ethernet, since the internal bus is 2048-bits wide (operation frequency of about 200 MHz), processing is more complex because there are up to four frame headers in one clock.

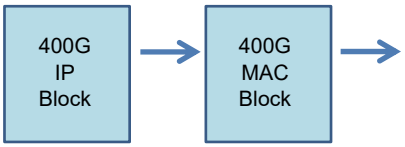

Figure 11 Single MAC Block Processing

(2) Multi-MAC Block Processing

This method suppresses design costs by using parallel processing of 100GbE MAC blocks and a downstream MUX/DEMUX circuit. However, a disadvantage of the parallel processing circuit is the inability to assure continuity of Tx frames across each MAC block.

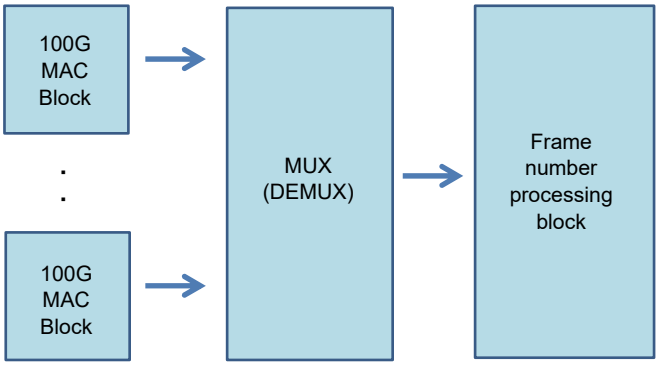

Figure 12 Multi-MAC Block Processing

To guarantee Tx frame continuity, the test frame number appended to test frames is added after the MUX operation, allowing frames to be sent according to the frame number sequence.

4.6 Partial Reconfiguration Technology

Although the 100G maximum bit rate design (100GbE/10GbE, OTU4/OTU2, etc.) supports simultaneous operation at two ports per module, partial reconfiguration technology was implemented at the FPGA to enable independent operation at each port. This technology divides the FPGA into multiple regions to configure only the specified region by requiring a FPGA configuration file provided for each region.

One partial reconfiguration partition each is used for Port 1 and Port 2, respectively. (The orange part in Figure 13 is the partial reconfiguration area.)

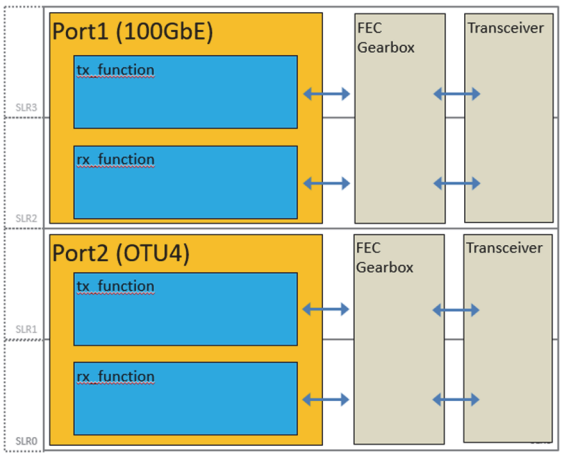

Figure 13 Partial Reconfiguration

By using this partial reconfiguration, applications can be set and switched at respective ports without affecting measurement and settings at other ports for easy signal monitoring at DUT insertion tests and in the passthrough mode.

For 200GbE and 400GbE designs, only one port needs to be supported, and partial configurations are not applied to reduce gate utilization efficiency.

4.7 400G Communications Network Measurement

The MT1040A supports the 400G Ethernet standard with BERT and Jitter/Latency measurements, as well as Frame-Capture and RFC2544 tests. In addition, since 400G PAM4 corrects communication errors using FEC, it is important to know both the BER and the level of error correction by FEC.

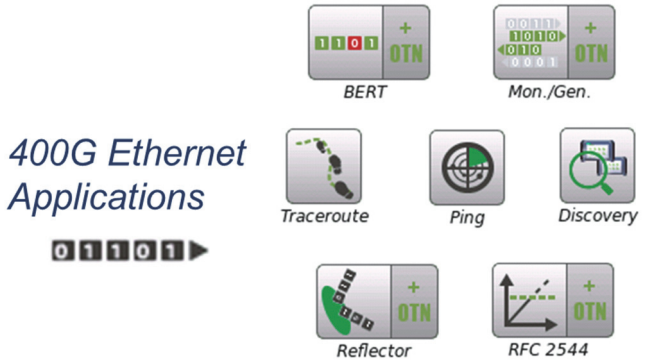

Figure 14 400G Ethernet Applications

4.8 Measuring and Visualizing FEC Communications Quality

For networks with assumed FEC, it is important that the number of errors is within the range of the error correction ability. Therefore, the MT1040A has an added function for measuring the number of Symbol Errors per Codeword as an index of the network quality.

Up to 15 Symbol Errors per Codeword can be corrected using RS-FEC (544,514). Consequently, when three Symbol Errors occur, due to the high margin there is a low possibility that error correction will fail. However, when there are 15 Symbol Errors, there is a high possibility that error correction will fail due to the lack of any margin. As a result, a function has been added to the MT1040A for setting and outputting a warning about the maximum Symbol Error count threshold for evaluations when the error correction margin is low. Consequently, it is possible to detect networks with potential high Uncorrectable Codeword counts.

Moreover, the Symbol Error occurrence rate can be plotted as a histogram and the FEC quality can be confirmed visually using evaluation results based on the threshold setting. The histogram not only makes it easy to understand the maximum Symbol Error count and error distribution, but also simplifies pass/fail evaluation using color-coded results based on the threshold setting.

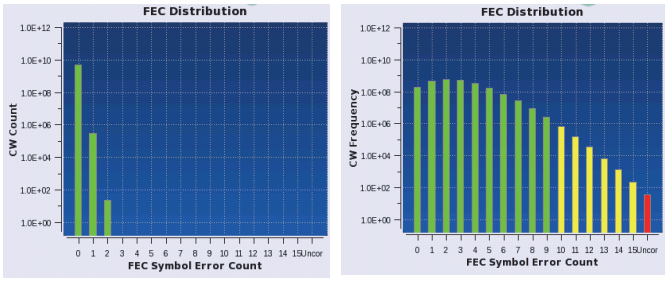

Figure 15 Symbol Error Rate Distribution

The y-axis of the histogram plots the occurrence frequency, and the x-axis plots the FEC-corrected Symbol Errors per Codeword. The histogram on the left in Figure 15 shows a good-quality network with some Codewords with a few Symbol Errors. In contrast, the histogram on the right shows a poor-quality network with many Codewords with many Symbol Errors. These color-coded histograms make it easy to visualize network quality.

5 Conclusion

400G Ethernet is a key technology for implementing high-speed, large-capacity communications following the widespread rollout of 5G services and increasing use of Cloud computing applications. Installation and maintenance (I&M) of these networks is driving high demand for a

compact, portable, 400G tester supporting easy onsite measurement. Under these circumstances, we have developed the compact, portable, battery-operated MT1040A for I&M testing and evaluating 400G Ethernet networks as well as various other communication methods and bit rates.

We hope this tester will play a key role in improving the quality of future high-speed, complex optical networks.

References

1) IEEE: "IEEE802.3bs-2017 - IEEE Standard for Ethernet Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation," Jan, 2017.
2) Mitsuhiro Usuba, Osamu Sugiyama, Yasuji Ishizuka, Atsushi Furuki, "Development of Network Tester for Precision Synchronous Network Verification", Anritsu Technical Review No.25 (2017.9)
3) QSFPDD MSA - "QSFP-DD Hardware Specification for QSFP DOUBLE DENSITY 8X PLUGGABLE TRANSCEIVER Rev. 5.0"
4) OIF: "Implementation Agreement OIF-CEI-04.0 CEI-56G-VSR-PAM4" Dec, 2017

Authors

Toru Miyauchi
Product Development Dept.
Service Infrastructure Solutions Division
Test & Measurement Company

Takaya Abe
Product Development Dept.
Service Infrastructure Solutions Division
Test & Measurement Company

Yasuji Ishizuka
Product Development Dept.
Service Infrastructure Solutions Division
Test & Measurement Company

Akinobu Tsuyuki
Product Development Dept.
Service Infrastructure Solutions Division
Test & Measurement Company

Publicly available