WHITE PAPER

PAM4 AN-LT

Autonegotiation and Link Training on new PAM4 based 400GE 200GE, 100GE and 50GE interfaces

OVERVIEW

Autonegotiation is a tried and tested process where a device is connected to a network through an Ethernet interface, and the end points of a link share information on various capabilities relevant for the communication. The autonegotiation process determines the highest common denominator among the end point capabilities, and these will be used for the communication.

Link Training is another process that can take place when a device is connected to a high-speed Ethernet port through a copper cable or backplane. In this case it is important that the characteristics of the transmitted signal are tuned to be optimally carried over the copper cable.

Xena's 7-speed dual-media test module Thor-400G-7S-1P supports the new PAM4 based 400GE, 200GE, 100GE and 50GE speeds defined in the IEEE 802.3bs and 802.3cd standards, in addition to new and legacy NRZ based versions of the 100GE, 50GE, 40GE, 25GE and 10GE standards. Installed in a ValkyrieBay or ValkyrieCompact chassis, it provides advanced Ethernet test capabilities up to Layer 3.

This White Paper looks at Autonegotiation and Link Training and explains how Thor-400G-7S-1P supports these functions for the new PAM4 based 400GE 200GE, 100GE and 50GE data rates.

"This White Paper explains how the Thor-400G-7S-1P test module supports Autonegotiation and Link Training on the new PAM4 based 400GE, 200GE, 100GE and 50GE interfaces"

PAM4 AN-LT

Contents

WHITE PAPER

INTRODUCTION

Modern communication services provide a wide range of applications like social media networking, e-commerce, on-line banking, Anything-as-a-Service (XaaS), streaming of HD video and many more. These are made available as cloud services, implemented with an increasing number of Data centers creating a continuous need for higher bandwidth for data center communication. In the new standards 802.3bs and 802.3cd IEEE has defined 400GE and 200GE data rates to support the need for higher bandwidth. 802.3cd also defines new versions of 100GE and 50GE interfaces.

Before the communication can start, the link end points must be set up with matching configurations. This can be a fixed setting; however in some cases the end points “negotiate” with each other to find the best possible configuration that both ends support, commonly referred to as “Auto-negotiation”.

For some application, like intra-rack communication in data centers, copper cables – Direct-Attach Cables (DAC) – are used being a cost effective solution for short range high speed communication. When DACs are used it is important that the characteristics of the transmitted signal is tuned to be best possibly carried through the link. To achieve this the end points communicate to perform “Link Training”.

The PAM4 Line Code

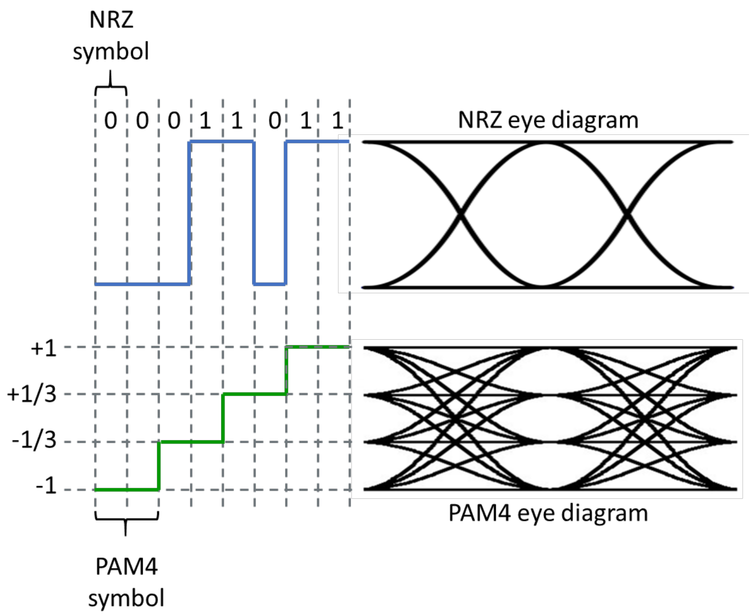


Figure 1: PAM4 and NRZ line codes and eye diagrams

Many Ethernet connections have been based on the Non-Return-to-Zero (NRZ) line code, which transfers 1 bit per sent symbol. So far this has been used for SerDes speeds up to 25 Gbps. A 100 Gbps Ethernet defined with current standards is sent on four 25 Gbps SerDes – older standards define 100 Gbps Ethernet sent on ten 10 Gbps SerDes. Higher rates could be achieved by sending signals as a higher multiple of more 25 Gbps SerDes but a desire to reduce the spectral bandwidth and number of SerDes used for the signal has driven a new line code for the high-speed signals: PAM4 (Pulse Amplitude Modulation), which encodes two bits in a single symbol by using 4 signal levels. This provides 53.125 Gbps SerDes with a symbol rate (or baud rate) of 26.5625 GBaud. For simplicity these are normally referred to as 50 Gbps or 25 GBaud SerDes. Figure 1 illustrates the difference between the NRZ and PAM line codes.

Using the PAM4 line code 400GE interfaces can be implemented with eight 50 Gbps SerDes and 200GE interfaces can be implemented with four 50 Gbps SerDes. New versions of 100GE interfaces with two 50 Gbps SerDes and 50GE interfaces with one 50 Gbps SerDes are also defined. Figure 3 shows the interfaces that are defined in the new IEEE standards 802.3bs and 802.3cd. Most are based on PAM4 line code; there is however also one 400GE interface defined that uses 16 25 Gbps SerDes with NRZ line code.

Transmitting 50 Gbps of information using the PAM4 line code requires approximately the same bandwidth as 25 Gbps sent with the NRZ line code. However as indicated in figure 1 the PAM4 signal levels are one-third of the NRZ signal levels. This means that the PAM4 Signal-to-Noise Ratio (SNR) is significantly reduced, which would lead to more errors than an NRZ coded signal. To compensate for this a strong Forward Error Correction (FEC) is mandatory in standards that specifies the use of the PAM4 line code. The FEC that must be used with the PAM4 line code is the Reed-Solomon RS(5440, 5140) "KP" RS-FEC.

Standard	New 50G	New 100G	200G	400G
CR (3m/5m) DAC - electrical	802.3cd 50GBASE-CR 50G PAM4	802.3cd 100GBASE-CR2 2x50G PAM4	802.3cd 200GBASE-CR4 4x50G PAM4	TBA
SR (100m)	802.3cd 50GBASE-SR 50G PAM4 MMF	802.3cd 100GBASE-SR2 2x50G PAM4 MMF	802.3cd 200GBASE-SR4 4x50G PAM4 MMF	802.3bs 400GBASE-SR16 16x25G NRZ MMF
DR (500m)	TBD	802.3cd 100GBASE-DR 1x100G PAM4 SMF	802.3bs 200GBASE-DR4 4x50G PAM4 SMF	802.3bs 400GBASE-DR4 4x100G PAM4 SMF
FR/LR (2km/10km)	802.3cd 50GBASE-FR/LR 1x50G PAM4 SMF	TBD	802.3bs 200GBASE-FR/LR4 4x50G PAM4 WDM	802.3bs 400GBASE-FR/LR8 8x50G PAM4 WDM

Figure 2: Overview of the main IEEE 802.3bs and IEEE 802.3cd interface technologies.

Autonegotiation

Autonegotiation was originally designed for Ethernet over twisted pair up to 1G. With this the devices at the end points of a link can negotiate common transmission parameters capabilities like speed and duplex mode. At higher speeds also choice of FEC may be relevant. For PAM 4 based 200GE, 100GE and 50GE autonegotiation is only defined for copper cables and at these speeds only the speed can be autonegotiated, as the KP RS-FEC is mandatory and duplex mode is not relevant at these speeds. Autonegotiation is not defined for 400GE because copper (DAC) interfaces are not officially defined for this rate.

During autonegotiation the end points of a link share their capabilities and choose the highest performance transmission mode they both support as illustrated in figure 3.

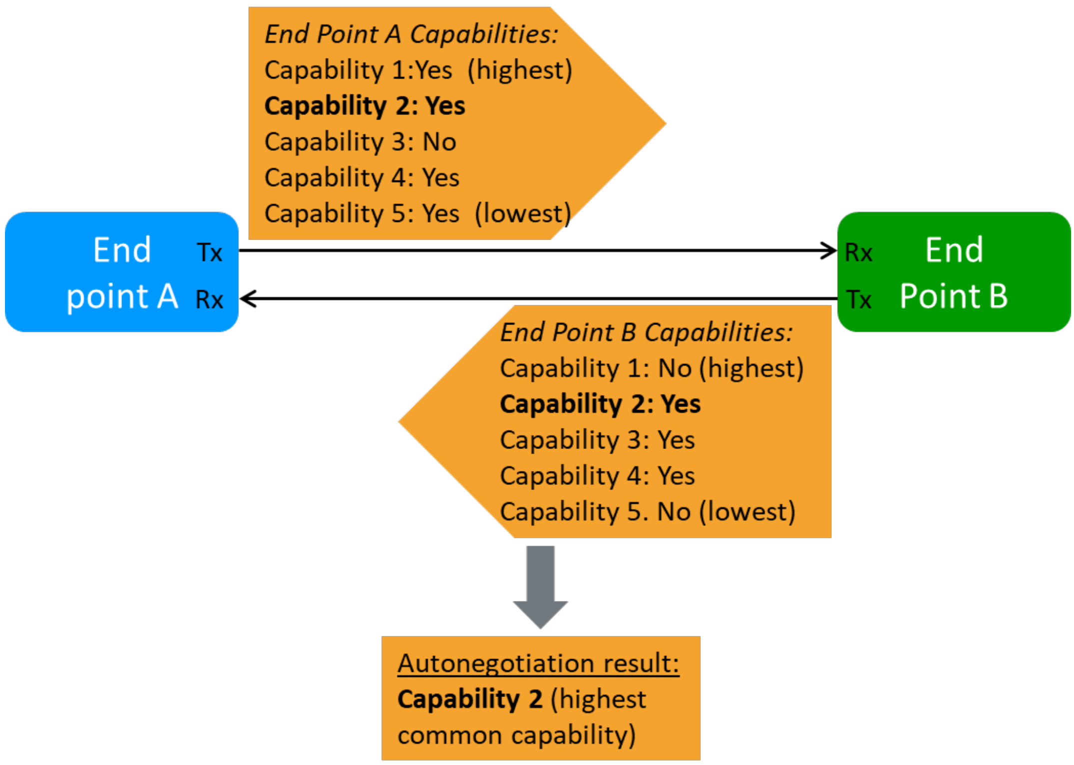


Figure 3: Autonegotiation process

Autonegotiation is defined in IEEE 802.3 Clause 73, which is updated in IEEE 802.3cd. It adds the following PAM4 based rates to those that can be autonegotiated:

- 200 G, 4 lanes: 200GBASE-KR4 or 200GBASE-CR4 (highest priority)
- 100 G, 2 lanes: 100GBASE-KR2 or 100GBASE-CR2
- 50 G, 1 lane: 50GBASE-KR or 50GBASE-CR

As indicated above autonegotiation is not defined for optical interfaces; it is only defined for copper cables (-CR) and electrical backplanes (-KR). This applies for all rates defined in IEEE 802.3 Clause 73 (ranging from 1G to 200G). At 400G only optical interfaces are currently specified, therefore autonegotiation is not defined for this rate.

Link Training

High-speed transmission of PAM4 symbols through a copper (electrical) cable is quite challenging as the symbols will affect each other when passing the cable. This can be partly compensated through a Tx equalizer, whereby the amplitude of a sent symbol (Main) is adjusted based on both immediately preceding (Pre) and following (Post) symbols. The "Main" setting will control the overall amplification/amplitude of the entire signal. Any pre/post settings come on top of that.

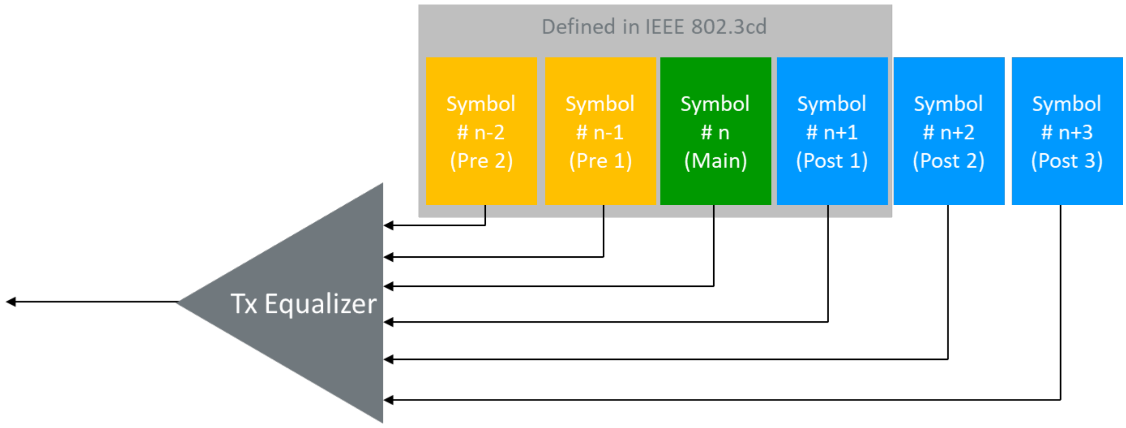


Figure 4: Tuning of transmitted signal

The two end points of the line can automatically do the adjustment of “Tx Equalizer” in a process that is known as “Link Training”. With Link Training the transmitter and receiver on the link communicate to tune the “Equalizer” settings. This is defined in IEEE 802.3cd clause 136.8.11 for 200GE, 100GE and 50GE PAM4 based signals.

Figure 5: Link Training signal

During the link training the two end points of the link will exchange signals as depicted in figure 6. It contains:

- Frame Marker: Allows the receiver to identify the start of the Link Training signal
- Coefficient Update: Suggestions for new "Equalizer" settings for the transmitter at the other end of the link (the link partner)
- Status Report: Current "Equalizer" settings for the transmitter that sends the signal and hand shaking
- Training Pattern: A test pattern that the receiver will use to determine if the updated settings improve the transmission quality

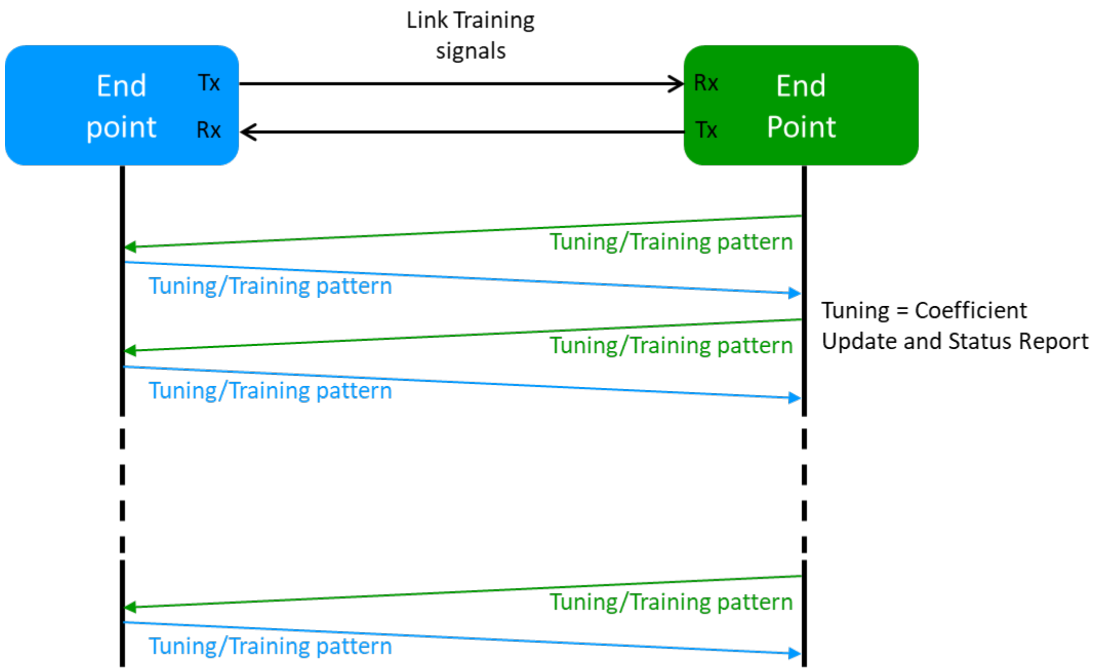


Figure 6: Link Training process

PAM4 link training will run for a couple of seconds. The receivers will continue to tune their setting while the link is in normal operation. Following the Link Training, the receivers will during normal operation continuously adjust their receiver characteristics in a non-disturbing way.

No Autonegotiation, No Link Training

In many cases Autonegotiation and Link Training are not required to establish a communication path: High speed optical transceivers and interfaces typically only run at one speed, so there is no need the negotiate this. Link Training is only required for electrical interfaces – in some cases (e.g. when short cables are used) an electrical interface may become operational just using default settings of the terminal equipment in the communication path.

No Autonegotiation, Link Training

While Link Training can be essential to make some electrical interfaces work, Autonegotiation may not be required is the link speed is fixed or if it can be manually set at both end points of a link.

Autonegotiation and Link Training

Autonegotiation and Link Training are in principle two independent processes. However if it is relevant to do both, Autonegotiation must be done first to determine the overall mode for a link and then perform the Link Training. Hereby you get the sequence shown in figure 7.

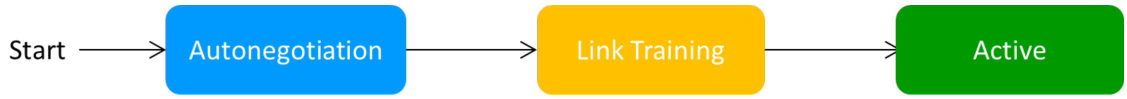


Figure 7: Autonegotiation and Link Training sequence

Thor-400G-7S-1P

Xena's 7-speed dual-media test module Thor-400G-7S-1P supports the new PAM4 based 400GE, 200GE, 100GE and 50GE speeds defined in the IEEE 802.3bs and 802.3cd standards in addition to new and legacy NRZ based versions of the 100GE, 50GE, 40GE, 25GE and 10GE standards. Installed in a ValkyrieBay or ValkyrieCompact chassis, it provides advanced Ethernet test capabilities up to Layer 3.

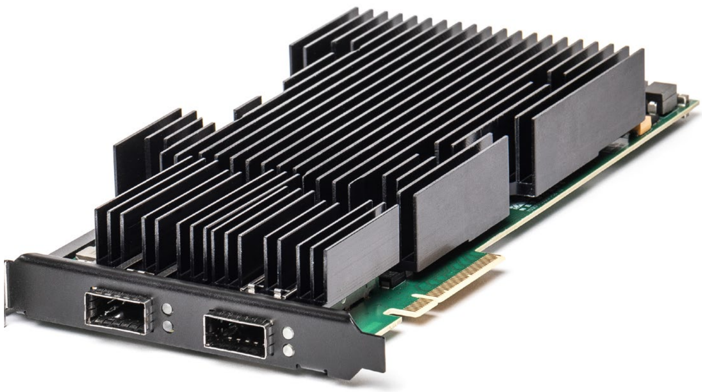


Figure 8: The versatile and powerful Xena Networks 7-speed dual-media test module Thor-400G-7S-1P

Autonegotiation and Link Training are very important features for the new electrical PAM4 based 200GE, 100GE and 50GE speeds. Link Training will also be very important for electrical PAM4 based 400GE. The Thor-400G-7S-1P will support Autonegotiation and Link Training for these data rates as indicated in table 1. This will ensure that a link can be established in cases where

Autonegotiation and Link Training is a pre-requisite for running the actual packet traffic e.g. if the DUT cannot initiate packet traffic unless Autonegotiation and/or Link Training has been performed. This will apply when copper cables (DAC cables) are used on PAM4 based 200GE, 100GE and 50GE links between the Thor-400G-7S-1P and a DUT. Thor-400G-7S-1P also supports Link Training at 400GE even though use of copper cables at this speed is currently not specified in the standards.

	400G	200G	100G	50G
FEC Type	RS-KP	RS-KP	RS-KP	RS-KP
No Autonegotiation, No Link Training	√	√	√	√
Link Training	√	√	√	√
Autonegotiation	✳≡	✳	✳	✳
Autonegotiation and Link Training in sequence	✳≡	✳	✳	✳

√ Supported

✳ Planned

✳≡ Not defined in the standards; implementation of de facto solution is planned

Table 1: Thor-400G-7S-1P Autonegotiation and Link Training capabilities at PAM4 based data rates

Manual Tuning of Transmitter Characteristics

Before a test is done using a DAC (copper) cable on a PAM4 based data rate, the Thor-400G-7S-1P transmitter characteristics can be manually tuned. For PAM4 based 200GE, 100GE and 50GE this can be done by connecting the DAC cable between the two cages of the Thor-400G-7S-1P. At 400GE, two Thor-400G-7S-1P modules will be required.

With the ValkyrieManager the user can manually adjust the “Tx-Equalizer” that defines the pre-, main- and post amplification of the symbols sent from the Thor-400G-7S-1P. This can be done for each of the SerDes used for the transmission. Alternatively all SerDes can be given the same settings by checking the “Common Parameters” box (see figure 9) – this is typically sufficient because the parallel channels exhibit very similar characteristics.

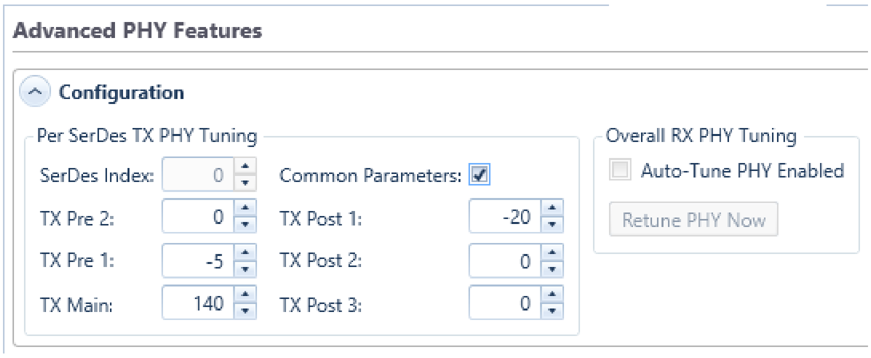
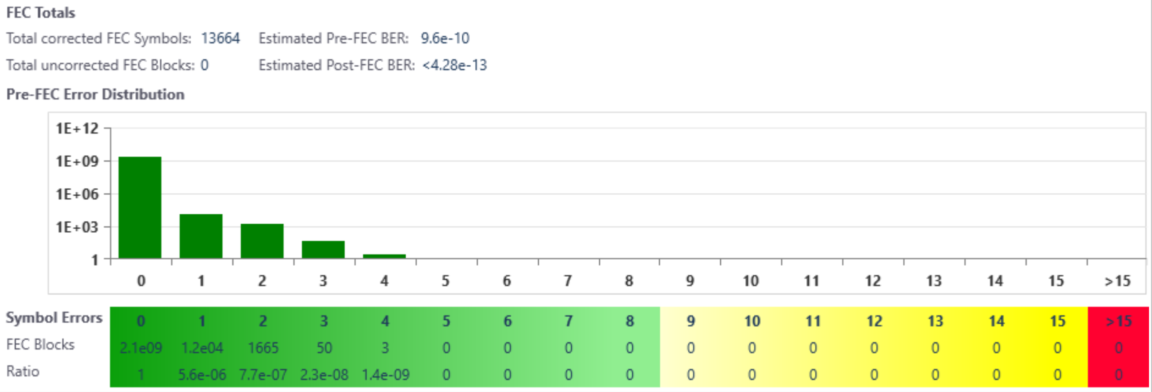


Figure 9: Manual equalizer setting (Pre 2, Pre 1, Main and Post1 are defined in IEEE 802.3cd)

A way to see how to adjust the "Tx equalizer" is to use the Pre-FEC Error Distribution graph in ValkyrieManager. It shows in how many FEC blocks a given number of symbol errors (up to 15) have been corrected. For each change of equalizer settings the Pre-FEC Error Distribution graph can be checked to see when a minimum number of symbol errors has been corrected. On many 1-3 meter DAC cables Main = 140, Pre 1 = -5 and Post 1 = -20 has empirically improved the BER.



Symbol Errors	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	>15
FEC Blocks	2.1e09	1.2e04	1665	50	3	0	0	0	0	0	0	0	0	0	0	0	0
Ratio	1	5.6e-06	7.7e-07	2.3e-08	1.4e-09	0	0	0	0	0	0	0	0	0	0	0	0

Figure 10: Pre-FEC Error Distribution graph

CONCLUSION

Autonegotiation is a process that for many years has been used when a device is connected to a network through an Ethernet interface. With this the end points of a link share information on various capabilities relevant for the communication. The autonegotiation process determines the highest common denominator among the end point capabilities, and these will be used for the communication.

Link Training is another process that may take place when a device is connected to a high speed Ethernet port through a copper cable or backplane. In this case it is important that the characteristics of the transmitted signal is tuned to be optimally carried over the copper cable.

In the new standards 802.3bs and 802.3cd IEEE has defined PAM4 based 400GE and 200GE data rates to support the need for higher bandwidth. 802.3cd also defines new versions of PAM 4 based 100GE and 50GE interfaces.

Xena's 7-speed dual-media test module Thor-400G-7S-1P supports the new PAM4 based 400GE, 200GE, 100GE and 50GE speeds defined in the IEEE 802.3bs and 802.3cd standards in addition to new and legacy NRZ based versions of the 100GE, 50GE, 40GE, 25GE and 10GE standards. Installed in a ValkyrieBay or ValkyrieCompact chassis, it provides advanced Ethernet test capabilities up to Layer 3. The Thor-400G-7S-1P supports PAM4 based communication where neither Autonegotiation nor Link Training is used and also PAM4 systems requiring Link training. In addition, support of Autonegotiation for PAM4 based 400GE, 200GE, 100GE and 50GE speeds is planned for the Thor-400G-7S-1P.