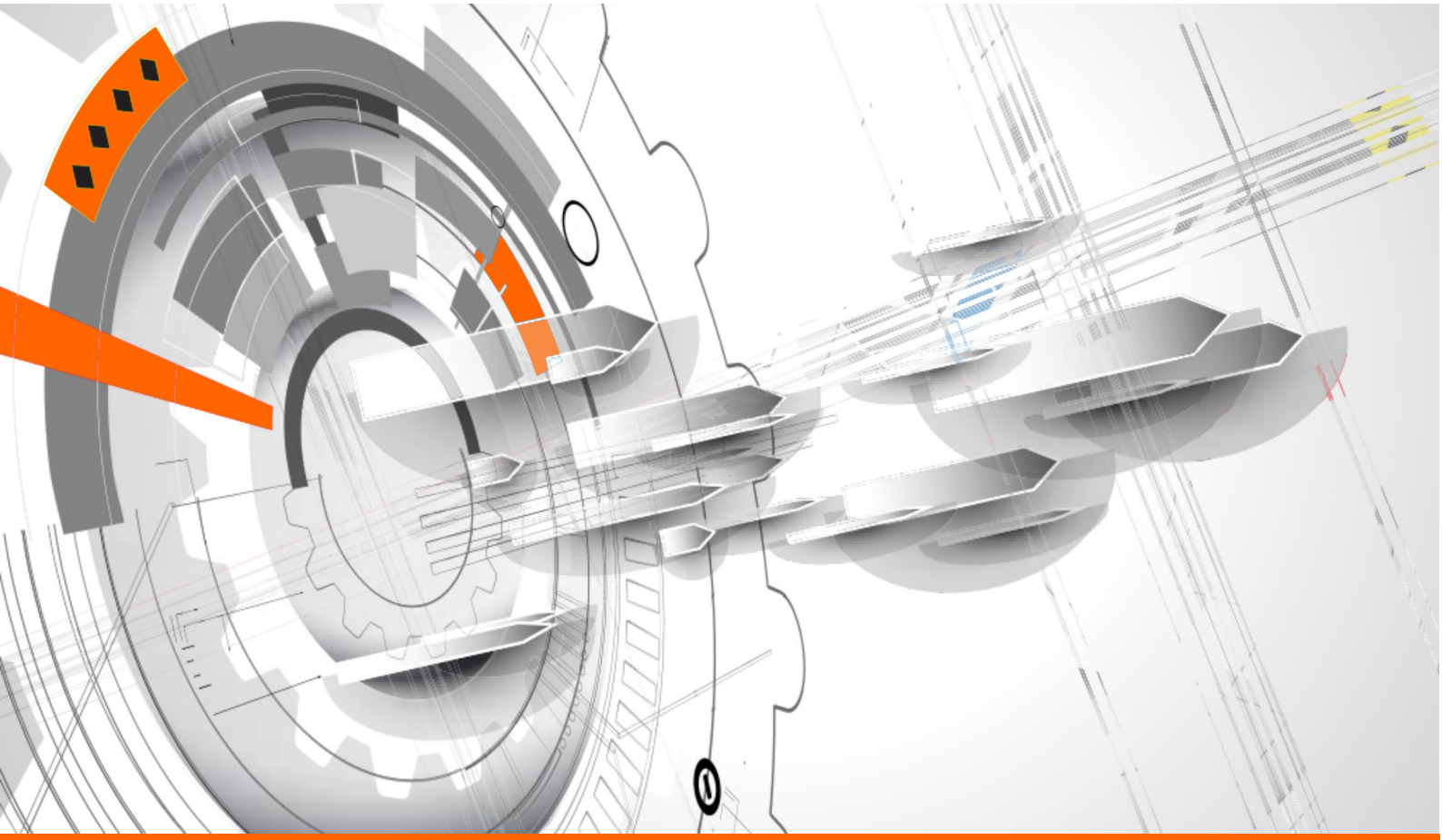
XENA™
Test. Improve. Repeat.

Red Helix

WHITE PAPER

Introducing Xena PHY

for Layer 1 Testing of 112G SerDes



WHITE PAPER

A new way to test Layer 1 features when using 112G SerDes PAM4

Supplied and supported in the UK & Ireland by **Red Helix**
Tel: 01296 397711 | Email: info@redhelix.co.uk | Web: www.redhelix.co.uk

CONTENTS

INTRODUCTION

The ever-increasing demand for more bandwidth is challenging the rack density, power consumption and cost of data centers. Currently, the best compromise is to install 400 Gbps connectivity using 100 Gbps per lane over short electrical cables wherever possible.

However, transmitting 100 Gbps using advanced modulation formats, such as 4-level Pulse Amplitude Modulation (PAM-4), over even a few meters of electrical cable causes signal integrity impairments such as loss, inter-symbol interference and cross talk. Therefore, the actual implementation and testing of the Physical layer 1 is critical.

It is common in network equipment to implement the Layer 1 features using a 3rd party PHY connected to a core ASIC/FPGA that handles the Layers above Layer 1 (see Figure 1a). The key benefit of separating the core ASIC/FPGA and the 3rd party PHY is that the core device does not have to support the advanced, high-speed analogue and digital functions because the PHY takes care of that.

This approach is well suited for line-cards and switches operating in normal transmission mode, but it is not ideal for some types of Ethernet test equipment. In addition to the regular transmission mode, Ethernet test equipment needs to offer advanced features for testing and debugging at Layer 1 going beyond the requirements defined in the standard. Such advanced functions are often not possible to implement using a 3rd party PHY with a fixed architecture.

Furthermore, standards are evolving rapidly, and test equipment must "push the envelope" to help developers of silicon and network equipment verify the performance of their products before they reach the market. Thus, test equipment needs to be future proof with the flexibility to add new test and management features as standards and new use cases evolve.

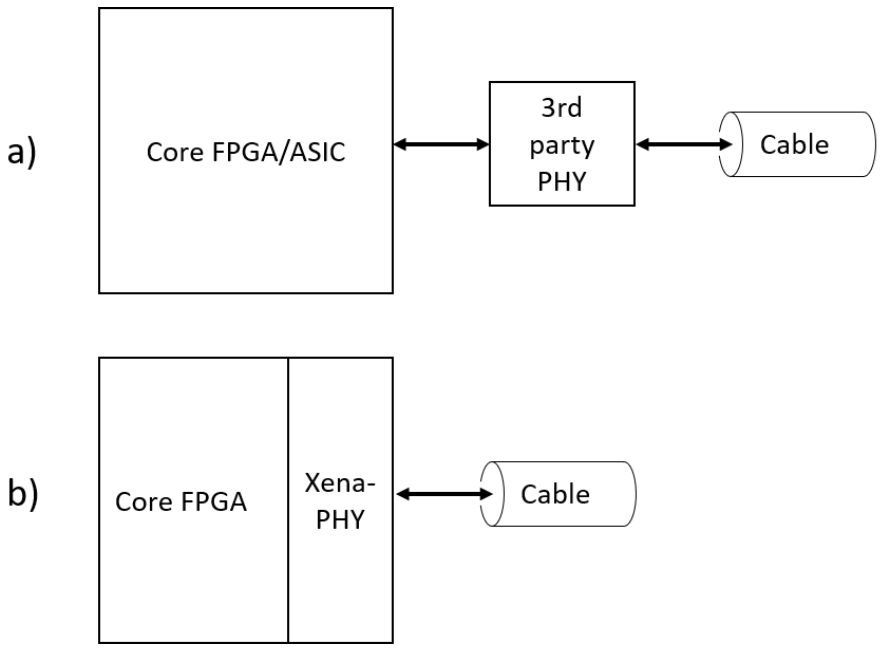


Figure 1: Implementation of PHY using a) 3rd party PHY and b) Xena PHY.

Xena Networks chose to develop and implement a new proprietary architecture for all Layer 1 functions for our latest Freya packet test module for 100, 200, 400 and 800 Gbps Ethernet using 112 Gbps SerDes. This new Xena PHY (see Figure 1b) lets us offer customers the flexibility to rapidly test new features and standards. Furthermore, the Xena PHY has proven to provide superior signal integrity when used with electrical cables.

This White Paper outlines the functional blocks of a typical Ethernet PHY using a 112 Gbps SerDes to illustrate the complexity at Layer 1 of the latest Ethernet standard.

It also outlines some critical test scenarios that can be implemented only through direct access to all Layer 1 functions like in the Xena PHY.

ETHERNET PHY

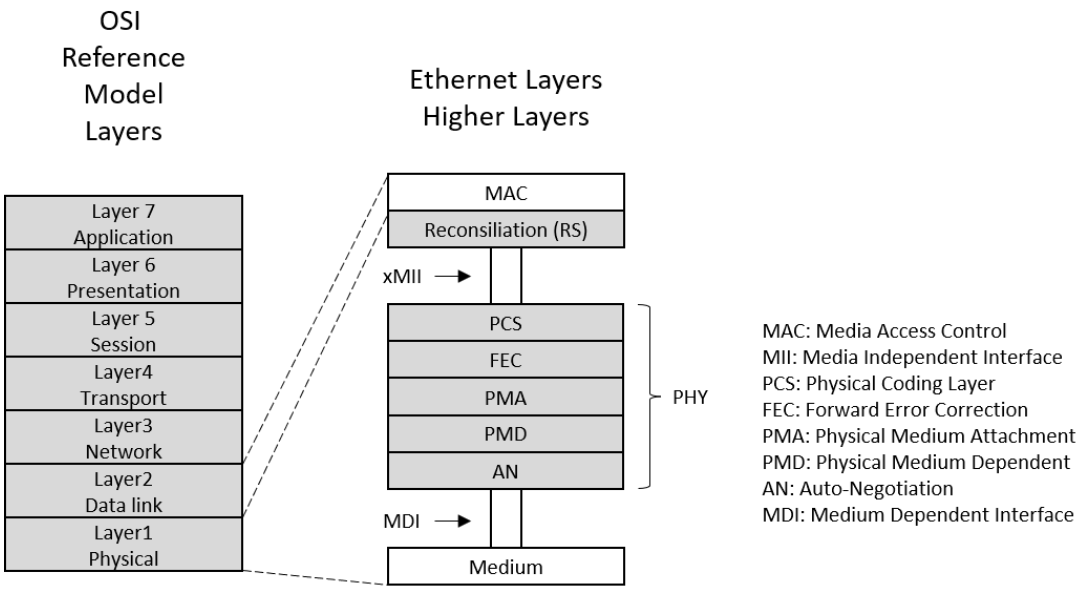


Figure 2: Details of Layer 1 and 2 of the OSI stack model.

Figure 2 shows the Open Systems Interconnect (OSI) model with the bottom Layer 1 (PHYsical or PHY) detailed for a typical high-speed Ethernet implementation. The three major sublayers in the PHY are the Physical Coding Sublayer (PCS), the Physical Medium Attachment (PMA) sublayer and the Physical Medium Dependent (PMD) sublayer. The Auto-Negotiation (AN) and Link Training (LT) are function used mainly during link bring up before real traffic is being sent. In the following we will describe the different sublayers of the PHY in more detail.

Physical Coding Sublayer (PCS)

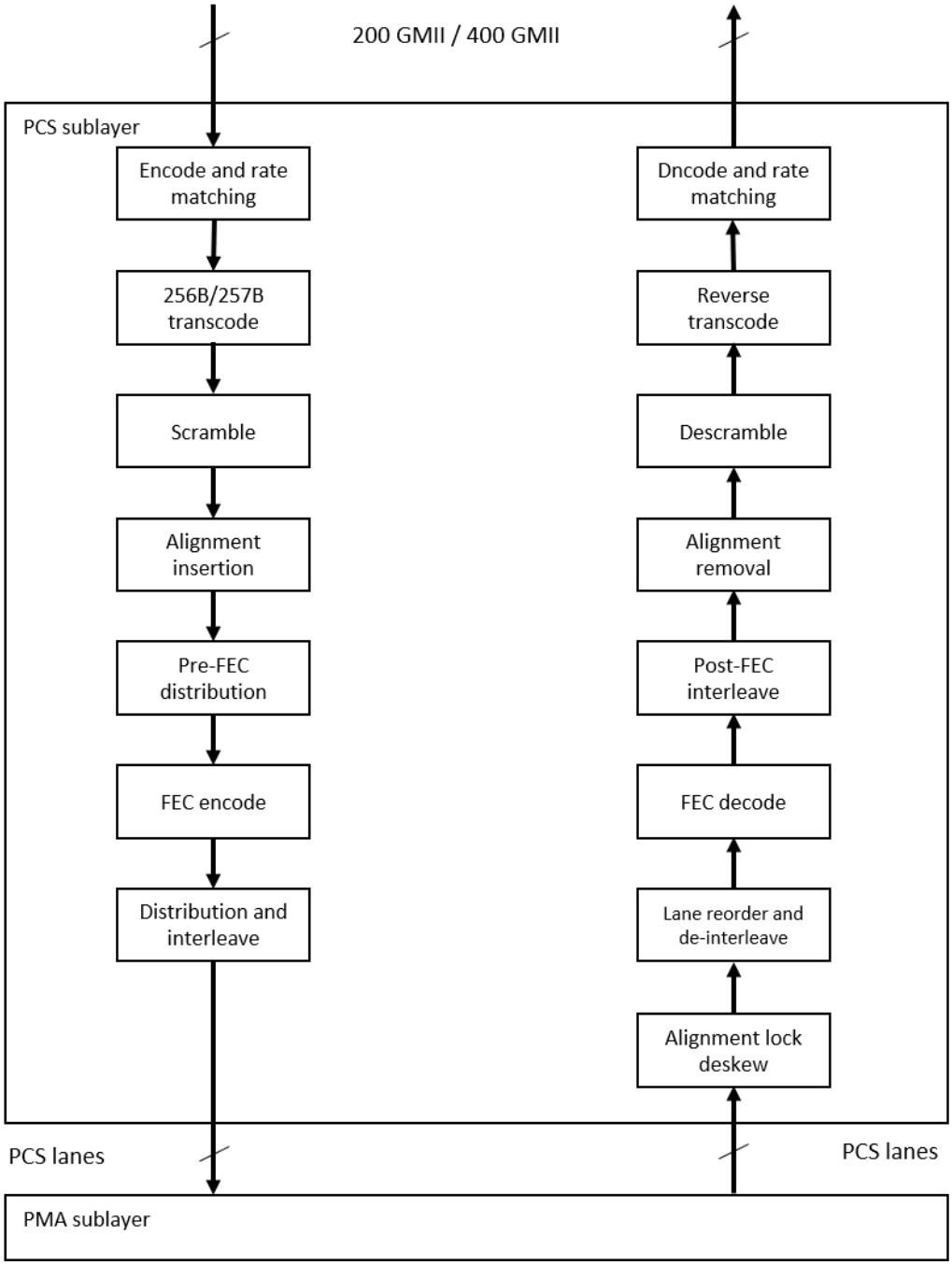


Figure 3: Functional block diagram of PCS.

The Physical Coding Sublayer (PCS) is defined in IEEE802.3 clause 119 and includes the functional blocks shown on Figure 3. The PCS receives/transmits Ethernet packets from/to the Media Access Layer (MAC) via the Media Independent Interface (MII), encodes/decodes the data and

transmits/receives to/from the PMA sublayer. IEEE802.3 has defined many different versions of the MII interface depending on the speed but, for 200 Gbps and 400 Gbps Ethernet over 112 or 56 Gbps SerDes the MII is either 200GMII or 400GMII.

In the transmit direction the PCS sub-layer receives from the MII blocks of 64 + 8 bits where the 64 bits are payload (TXD) and 8 bits (TXC) are control data. If TXC is all 0's (0x00) the 64 bits (TXD) are normal packet data whereas if TXC is all 1's (0xFF) the 64 bits (TXD) contain control words or error codes.

The 64 + 8 bits are encoded into blocks of 66 bits by the PCS sub-layer. The encoding depends on whether the data from the MII is normal packet data or control data. In case that normal packet data is received from the MII, the control bits (TXC=0x00) are simple removed and a 2 bits header with binary value "01" is added as shown on Figure 4. In case the data received from the MII is control data the type of control signal gets encoded into the 64 bits and a 2 bits header with the binary value of "10" is added. The 2 bits headers are used by the receiver to identify block boundaries.

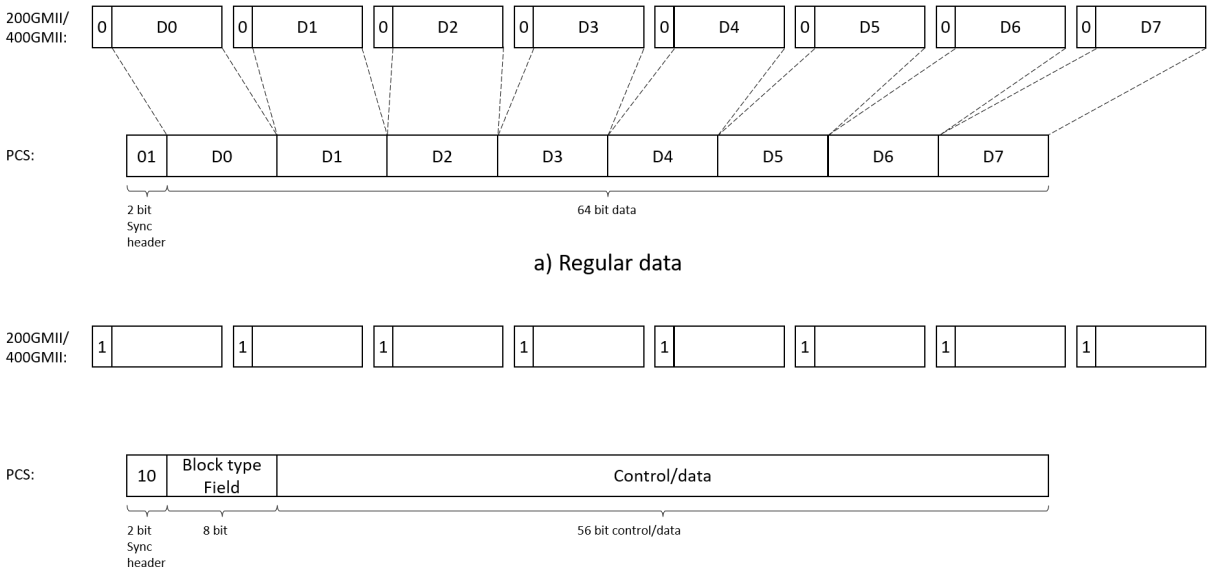


Figure 4: 64B/66B encoding of regular packet data and control data from the MII.

The 66 bits blocks are distributed one by one using round robin across a number of parallel PCS lanes (virtual lanes) with a lane rate of 26.5625 Gbps each. This means that 200 Gbps Ethernet will require 8 PCS lanes whereas 400 Gbps Ethernet will require 16 lanes. The 66B blocks are then further 256B/257B transcoded to free up overhead for a strong Forward Error Correction (FEC). The 257 bits blocks are scrambled to ensure frequent signal transitions to enable clock recovery at the remote receiver and DC-balance in the signal.

Alignment markers are added to each PCS lane to support de-skew and reordering of the individual PCS lanes at the receiver.

Forward Error Correction (FEC) is used to improve the Bit Error Rate and various versions of FECs have been part of the IEEE802.3 standard for some time. As the speed increases more and more complex FEC schemes must be used and for 100 Gbps Ethernet and above two interleaved FEC blocks are actually used. Two blocks of 5,140 bits (10,280 bits in total) are interleaved using a 10-bit round robin method. The two 5,140 bits long FEC codewords are then encoded to form two 5440 bits FEC blocks using the IEEE802.3 clause 134 RS(544,514) FEC with a symbol length of 10 bits. Finally, the FEC encoded blocks are send to the PMA sublayer.

In the receive direction, the virtual PCS lanes are re-constructed from the data received from the PMA sublayer and the alignment markers are used to deskew, re-order and de-interleave the lanes.

The FEC is decoded and the FEC codewords are de-interleaved. After removal of the alignment markers, the data is descrambled and decoded and the PCS virtual lanes are combined into one logical data stream on the MII.

Physical Medium Attachment (PMA) sublayer

The Physical Medium Attachment (PMA) sublayer described in IEEE802.3 clause 120 is responsible for bit multiplexing from a number of input lanes to a number of output lanes. Often, the PMA would multiplex the PCL lanes from the PCS to a number of PMD lanes to the PMD as shown on Figure 4. The standard however, allows multiple consecutive PMA layers multiplexing between different numbers of input and output parallel lanes. In the following we will assume one single PMA sublayer for simplicity.

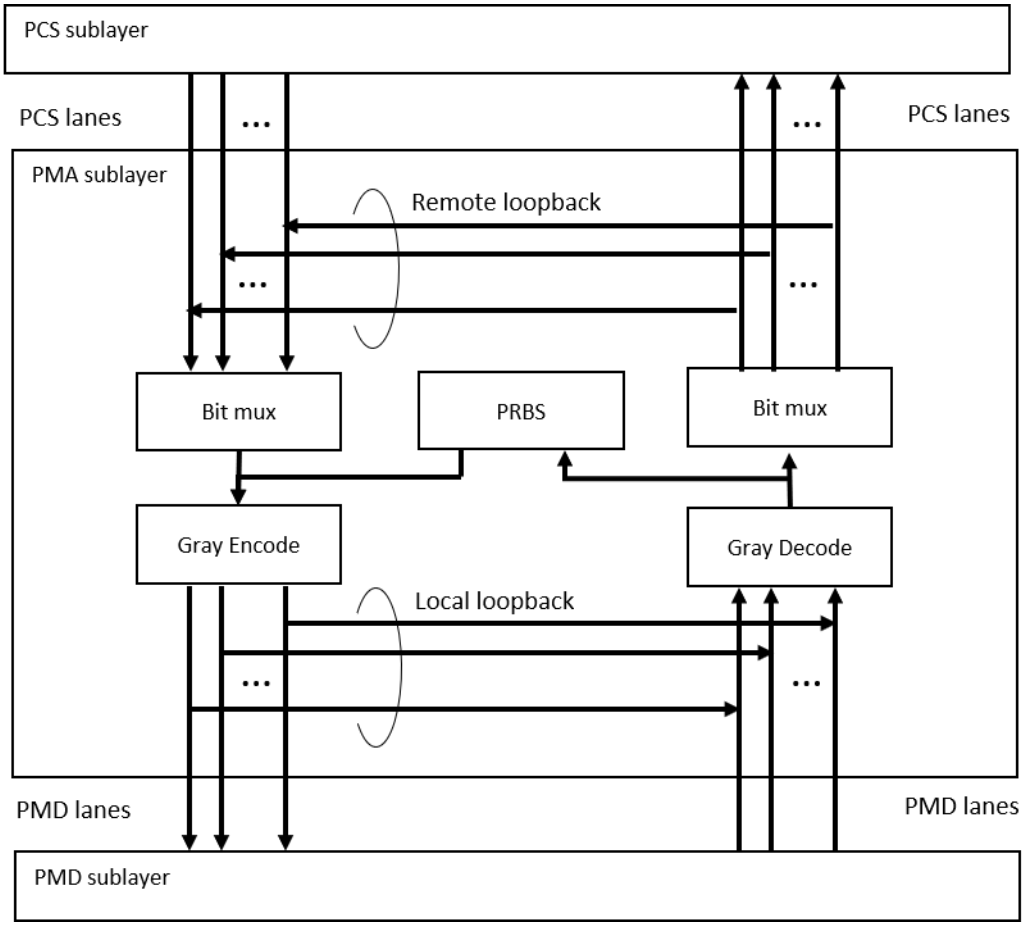


Figure 5: Block diagram of PMA sublayer.

The PAM-4 modulation format combines two bits into one symbol with the possible binary values "00", "01", "10" or "11". As described in the next section, the PMD sublayer will convert the PAM-4 symbols into four different voltage levels but, the PMA sublayer actually includes a Gray coding block that converts the symbols as shown on Figure 5. The purpose of the Gray coding is to facilitate error correction by ensuring that a change between two neighboring voltage levels leads to a change of one of the bits in the symbol only.

Binary symbol value		Gray coded symbol value
00	->	00
01	->	01
10	->	11
11	->	10

Figure 6: Gray coding of PAM-4 binary symbols

For troubleshooting and testing, the PMA sub-layer includes several blocks. The local loopback function tests the PCS and PMA sublayers of the local link partner and the remote loopback also tests the PMA sublayer functions at the remote link partner.

Several types of Pseudo Random Binary Sequence (PRBS) can be inserted per PMD lane in the transmit direction instead of the data from the PCS in order to test the performance of each lane.

Physical Medium Dependent (PMD) sublayer

The Physical Medium Dependent (PMD) sublayer provides the physical interface to the transmission medium (or Medium Dependent Interface – MDI). The physical implementation is therefore dependent on the specific medium. For instance, if the medium is an optical fiber, the transceiver includes a laser that converts the logical data to modulation of light. If the medium is an electrical cable the logical data is converted into voltage levels.

Each Gray-coded PAM4 symbol is transmitted and received as a voltage level with a symbol rate of 53.125 Gbaud per lane as shown on Figure 7. Since a symbol holds two bits, each lane transmits bits at 106.25 Gbps. 100 Gbps Ethernet thus requires a single physical lane, 200 Gbps two lanes and 400 Gbps four lanes.

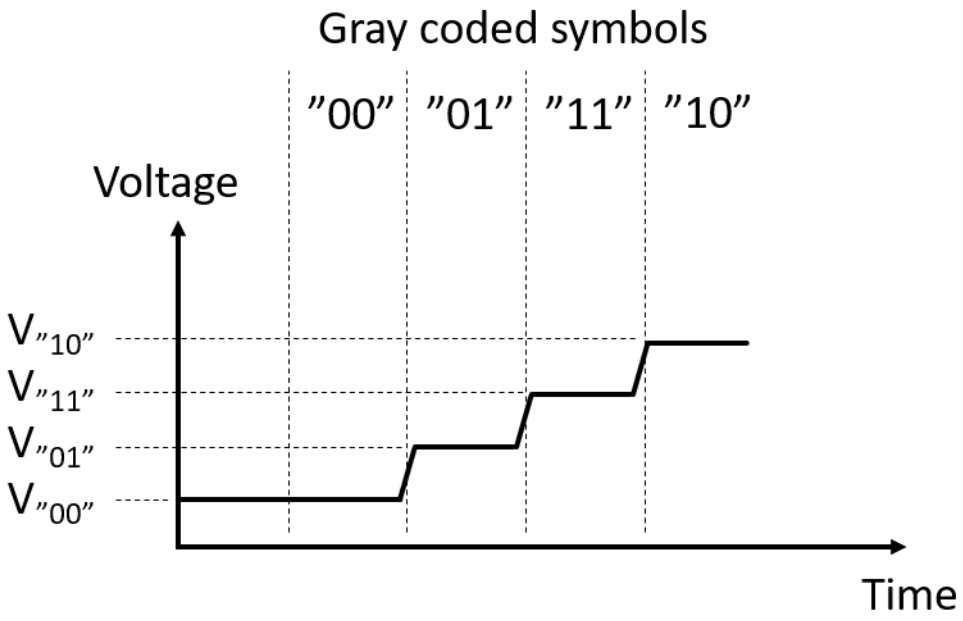


Figure 7: Conversion of Gray-coded PAM4 symbols to voltage levels.

The PMD sublayer receivers include the Clock and Data Recovery (CDR), receive equalizers as well as a transmit equalizer. The exact implementation is not defined by the standard but, it is common to parallelize the data sampling to reduce the clock frequency of the sampling circuitry.

Starting from the left-hand side on Figure 8 the received signal is first equalized using a Continuous Time Linear Equalizer (CTLE) and amplified using an Automatic Gain Controller (AGC). Both these functions are purely analog.

In a traditional (lower speed) receiver the Clock and Data Recovery (CDR) would simply be done by sampling the signal at the center of the symbol period. However, at 100 Gbps this approach is not practical and a time-interleaved Analogue-to-Digital-Converter (ADC) based method is generally used instead. To illustrate how that works, let us assume we use 4 ADCs in parallel. Then each ADC will only have to sample at ¼ of the 56 Gbaud PAM4 symbol rate i.e., 14 Gsamples/s. The ADC's sampling time is phase shifted by 0, 1, 2, and 3 full speed clock periods (0.18 nsec for 56 Gbaud), respectively. Effectively, this means that ADC1 samples symbols 0, 4, 8, …, ADC2 samples symbols 1, 5, 9, …, ADC3 samples symbols 2, 6, 10, … and ADC4 samples symbols 3, 7, 11, … . This means the receivers are a mixture of analogue and digital electronics.

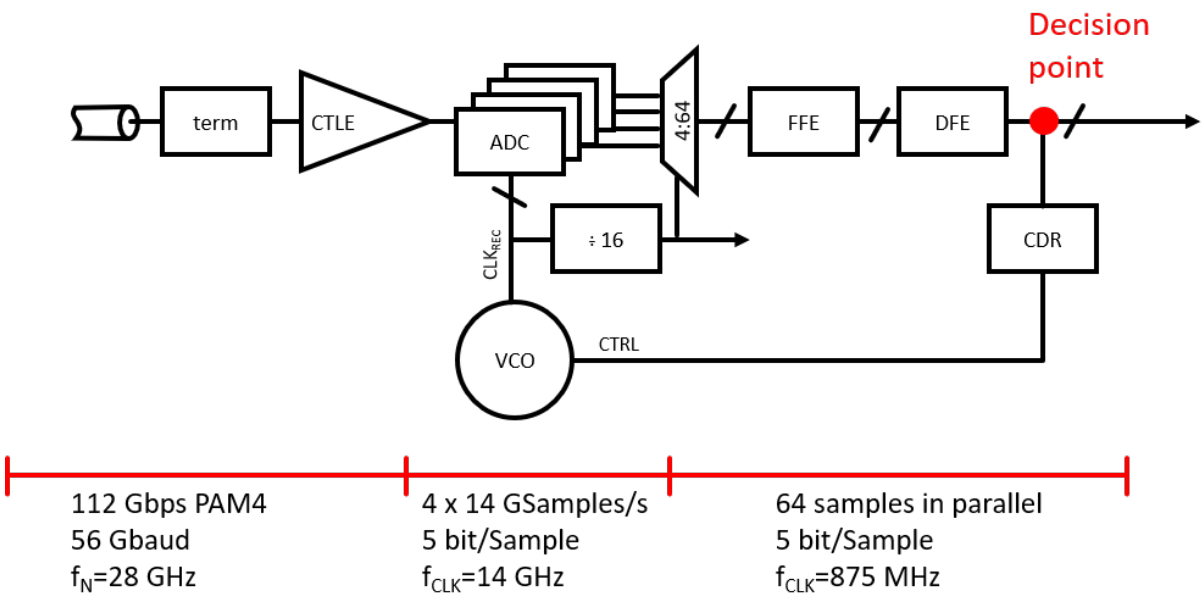


Figure 8: ADC-based SerDes receiver.

Especially for transmission of high-speed Ethernet over electrical cables, the equalizers must be tuned to optimize the signal quality for each lane and the specific transmission medium used. Link Training is an IEEE802.3 standardized protocol residing inside the PMD sublayer, whereby two devices can work together to tune their transmit equalizer settings while monitoring the error rate. In IEEE802.3ck the transmit equalizers are of the Finite Impulse Response (FIR) type with 5 taps. It is the gain of each of the five taps that is being adjusted during LT. If you would like to learn more about how Link Training works for 112 Gbps SerDes based Ethernet please contact Red Helix at info@redhelix.co.uk

Finally, the PMD sublayer includes the Auto Negotiation (AN) function. AN is a low-speed protocol whereby two devices connected over electrical cable can communicate and agree on a common set of transmission parameters before actual transmission of packet data at full speed begins. This means that all functions but the AN block in the PHY are disabled during this process. Also, AN is only active on lane 0 of a multi-lane connection.

THE XENA PHY

In the Xena PHY architecture, all sublayers of the PHY (PCS, PMA and PMD) are implemented and controlled by Xena inside the core FPGA as shown on Figure 5.

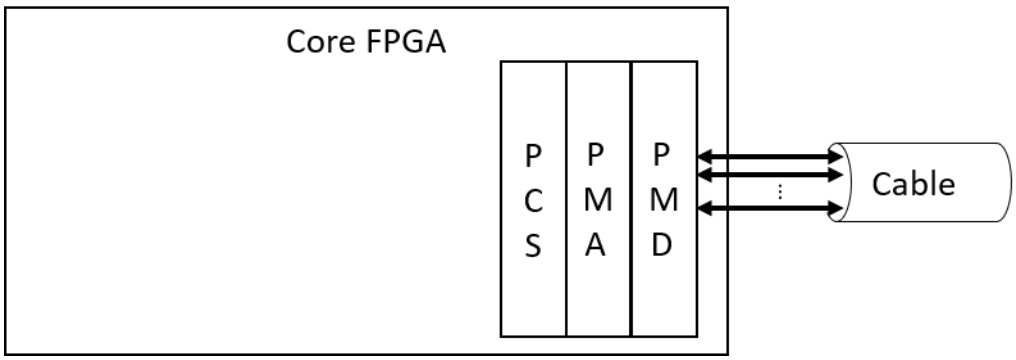


Figure 9: Implementation of the PHY sublayers in the Xena PHY module

The Xena Network's Freya series of modules whereof the Freya-800G-4S-1P module shown on Figure 8 is an example of an implementation using the Xena PHY.

Figure 10: Freya 800G-4S-1P module.

Freya supports testing of 112Gbps SerDes PAM4 based Ethernet at data rates of 100G / 200G / 400G / 800G and is designed for 800G switch, transceiver and PHY design validation and Quality Assurance.

Freya can be used for developers of Ethernet equipment for validation and testing during new product development. Another use case for Freya is installation of new equipment in for instance data centers to verify the performance of individual Ethernet ports before they are placed in actual service. This can save a lot of trouble during installation and commission as well as in daily operation.

The key L1 features supported by Freya are:

- 4-speeds: 800GE, 400GE, 200GE, and 100GE
- Dual media: QSFP-DD800 & QSFP112
- Supports 112G SerDes (PAM4 112G)
- Test with optics and DACs
- Auto-Negotiation & Link Training (AN/LT)
- Advanced signal integrity view
- RS-FEC (544,514, t=15)
- FEC error correction chart
- PRBS31Q
- Equalization controls

Use case examples

Figure 10 shows a typical use case of Freya connected to the port of a line card to test the performance of that port. For validation purposes it is of course essential to test that the line card is standards compliant.

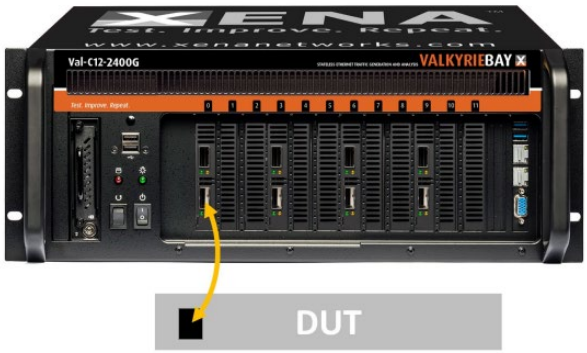


Figure 11: Test set-up with a connection between Freya and a Device Under Test (DUT).

In a real system, signal impairments like signal loss, time delays and jitter will vary over time and unit to unit. It is therefore important to stress test the device to ensure that it will work within certain parameter ranges. Freya enables you to stress test the device through deliberate insertion of time delays and bit errors at Layer 1 for instance.

Ethernet has built-in functions by which a device automatically adapts to a given transmission medium and the remote device´s port ability. We denote the normal, standardized operation of these adaptation functions as “mission mode”. However, to test that a DUT´s adaptation is working correctly, an ethernet tester needs the ability to disable the test equipment’s own adaptation functions. In this manner, only the adaption functions of the DUT are tested and not the combined DUT and tester. When the tester is in this mode, we denote it “tester mode”.

PCS lane testing

One of the key functions of the PCS sublayer is, as described previously, to handle the parallel PCS lanes. Upon reception, the PCS lanes may be delayed (or skewed) relative to each other

Because the Xena PHY enables control over the PCS sublayer, Xena has been able to add extra test features that allows manipulation of the PCS lanes. For instance, you are able to insert user-defined skew per Tx PCS lane to test that the DUT’s receiver is able to handle skew variations appropriately within the limits defined by the standard.

To test that the DUT’s receiver can identify alignment markers and reorder the PCS lanes correctly the Xena PHY includes a user defined PCS lane to SerDes mapping inside PCS sublayer of the transmitter.

FEC testing

The FEC is an important part of the PCS sublayer that needs to be tested as well. As mentioned previously, the PMA sublayer includes a PRBS generator that can be used to verify the BER performance. The Xena PHY allows you to extract the BER both pre- and post-FEC.

However, to really test the FEC you need to insert errors in a controlled manner in the PRBS bit stream. Such a function would be impossible to implement with-out the direct access to the PMA sublayer in the Xena PHY.

Equalizer testing

For connections over a few meters of electrical cables the Link Training protocol is very important. As mentioned previously, the LT protocol enables two devices to communicate about the transmit equalizer settings. Because both devices adapt their equalizers it can be hard to distinguish between the performane of the DUT and the tester. For this reason, Freya incorporates an option to “freeze” it’s equalizer such that changes occur only in the DUT.

Debugging mode

In some cases, it may be necessary to debug the AN and LT protocols in the DUT by stepping through individual protocol steps one by one. For this purpose, Freya includes an option to continuously send the same protocol information and disable the automatic generation of responses. Once it has been verified that the DUT responds correctly to one protocol information one can manually move to a new type of protocol information and test that.

Signal integrity view

A final example that really illustrates how the Xena PHY enables access into to the inner parts of the PHY functions as shown on Figure 9, which displays a signal integrity view of the sampled data in the

time-interleaved ADCs. The four levels correspond to the PAM4 levels and is constructed by plotting the sampled voltage levels of all ADCs as a function of time. If the four levels are clearly separated the signal integrity is good. By analyzing the distribution within each level, it is even possible to estimate signal-to-noise ratio and potentially other signal integrity parameters.

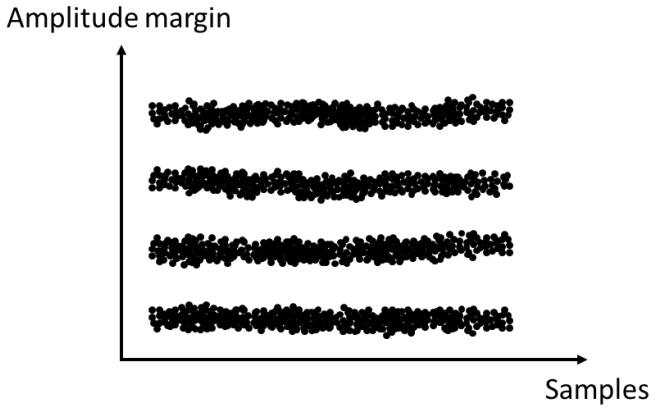


Figure 12: Sampled eye diagram for PAM4 modulated signal.

SUMMARY

Testing 100 Gbps, 200 Gbps, 400 Gbps and 800 Gbps using 112 Gbps SerDes requires focus on the physical layer (Layer 1) because the signals are heavily degraded even after short transmission lengths.

In this White Paper we have shown how Xena Networks' new architecture – the Xena PHY – provides control over all PHY sublayers. The clear benefits of this approach compared to implementations using a 3rd party PHY are the flexibility to add new features and fast access to new standard features as they evolve.

Xena's new Freya series of test modules is an example of an implementation using the Xena PHY with 112 Gbps SerDes PAM4 based Ethernet at data rates of 100G / 200G / 400G / 800G. Freya has demonstrated superior BER performance over electrical cables at a lane speed of 100 Gbps due to the specific implementation of the Layer 1 features in the Xena PHY architecture.

The Xena PHY also enables us to provide extra test features that are not possible when using a 3rd party PHY for Layer 1. Important examples of such test features are, PCS lane deskew and re-ordering, FEC testing through controlled error insertion, equalizer testing and debugging of the AN and LT protocols.

RELATED PRODUCT INFORMATION

Freya – Xena's 112Gbps SerDes Test Solution

Xena's Valkyrie TGA platform now includes Freya, a series of test modules with the functionality needed to test high-speed 112Gbps SerDes Ethernet devices and links at both Layer 1, Layer 2 and Layer 3.

For Layer 1 the relevant signal integrity parameters to test for 112 Gbps SerDes based systems includes Auto Negotiation, Link Training (for electrical cables), advanced PHY equalizer tuning, PRBS testing, pre-FEC error distribution and the signal integrity view.

For information on Xena's Freya test modules, please contact Red Helix at info@redhelix.co.uk

FreyaCompact AN/LT Test Appliance

As clearly established in this White Paper, Auto Negotiation and Link Training (AN/LT) are critical Ethernet features for high-speed, short distance connections over electrical cables because they help establish a link between two devices prior to sending real L2/L3 traffic.

To help companies test and debug AN/LT protocols for 56/112Gbps SerDes based Ethernet, Xena offers a stand-alone product called the FreyaCompact AN/LT Test Appliance.

Unlike Xena's other Freya test modules, the FreyaCompact AN/LT Test Appliance does not include the Traffic Generation & Analysis (TGA) capability.

Instead, the FreyaCompact AN/LT Test Appliance lets developers step through the protocols one message at a time enabling them to quickly and easily identify possible errors or timing issues in the implementation.

The FreyaCompact AN/LT Test Appliance can also test the true equalizer performance of a remote unit by disabling automatic equalizer adaptations in the tester unit.

These functions are available via two SW applications (provided with the hardware) which are based on Xena OpenAutomation (XOA), Xena's new open-source scripting and test automation platform.

1. The **AN/LT Protocol Test Utility** is an interactive command-line interface that makes it easy for AN/LT protocol development teams to single-step through AN/LT configurations.

2. The **AN/LT Protocol Test Suite** is aimed at labs that want to do performance and compliance testing of AN/LT, and therefore need configurable test scripts and test reporting.

There are two hardware versions of the FreyaCompact AN/LT Test Appliance – one for QSFP-DD800 and the other for OSFP.