

OIF has collaborated with the developers of an open source software tool called StatEye to provide an implementation of the mathematics specified in the CEI Implementation Agreement [12]. Certain releases of this tool include templates for the various CEI variants. StatEye performs the necessary statistical signal integrity analysis using these templates.

5.3 Ethernet Protocols

The Institute of Electrical and Electronics Engineers (IEEE) commissioned a committee to develop open network standards. Since this committee started work on this effort in February, 1980, the documents produced by this project were designated using the nomenclature IEEE 802.x. Several different network architectures were ultimately standardized, resulting in various ".x" extensions. The "Ethernet" system originally developed by Xerox was standardized and published as "IEEE 802.3 Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications" in 1985. Although the IEEE document does not use the term, the "Ethernet" name continues to be commonly associated with this standard. While Ethernet was once one of several network standards with significant deployment (for example: the 802.5 Token Ring), over time it has emerged as the dominant network standard. The IEEE 802.3 document has gone through several revisions, adding additional clauses to support additional speeds, physical layers, etc., since its original publication.

The historical network topology for an Ethernet system consisted of multiple workstations connected to a common Ethernet bus using coaxial cable. Whenever one of the workstations wanted to send data, the network device would first listen to the bus to determine whether another workstation was using it. If the bus was in use, the device would wait. If the bus was not in use, the network device would transmit a data packet, while continuing to monitor the bus for any collisions resulting from two devices starting transmission at the same time. If a collision occurred, the device would abort the transmission and wait some amount of time before trying again. The CSMA/CD portion of the title for the IEEE 802.3 document results from this historical network access method.

Beginning in the early 1990s, "star-connected" network topologies became the network configuration of choice. In a "star-connected" topology, each workstation connects to a central network unit (either a network switch or a network hub), and each connection is a point-to-point link implemented using twisted-pair wire or optical fiber. Such topologies allowed full-duplex connections such that a network device can be both transmitting and receiving at the same time. Since each link had only one transmitter, collisions were no longer possible. Nevertheless, the CSMA/CD access methods were retained in the IEEE 802.3 standard through the 10 Mbps, 100 Mbps, and 1 Gbps generations. Clauses pertaining to 10 Gbps recognize that practical implementations are point-to-point links, and no longer support CSMA/CD access methods.

5.3.1 Physical Layer Reference Model

Figure 5.14 illustrates the Media Access Control (MAC) layer and the various components (with the associated Ethernet terminology) of the Physical Layer associated with 10-Gbps Ethernet. The Reconciliation Layer remaps signals between the 32-bit MAC/PLS service interface definition (clause 36) and the 10-Gbps Ethernet Media-Independent Interface (XGMII) definition. The MAC layer (via the Reconciliation Layer) and the Physical Layer are interconnected using the XGMII. The Physical Layer consists of the Physical Coding Sublayer (PCS), Physical Medium Attachment (PMA) sublayer, and the Physical Medium Dependent (PMD) sublayer. There are multiple clauses of the IEEE 802.3 document which apply to these sublayers, depending on the type of medium being used. HSS cores are used in the high-speed electrical implementation of these layers.

Table 5.10 describes the relationship between 10-Gbps Ethernet variants and normative clauses of the IEEE 802.3 document for the physical sublayers. The published clauses relevant to 10-Gb Ethernet are summarized below:

Clause 44. Introduction to the 10Gbps Baseband Network: This clause is a normative summary of applicable requirements for the 10-Gbps LAN/WAN variants described in Table 5.10. This clause references other 10Gbps clauses.

Clause 46. Reconciliation Sublayer (RS) and 10 Gigabit Media Independent Interface (XGMII): This clause defines both the Reconciliation Sublayer (RS) and the XGMII.

Clause 47. XGMII Extender Sublayer (XGXS) and 10 Gigabit Attachment Unit Interface (XAUI): This clause defines a method of extending the physical distance between the MAC layer device and the PCS/PMA layer device. An XGXS layer device is located near the MAC layer device (DTE XGXS), and a similar XGXS layer device is located near the PCS/PMA layer device (PHY XGXS). XAUI is the interface definition which interconnects the two XGXS layer devices.

Clause 48. Physical Coding Sublayer (PCS) and Physical Medium Attachment (PMA) sublayer, type 10GBASE-X: This clause defines the PCS and PMA layers for connecting an XGMII to a PMD using a 4 lanes by 3.125Gbps electrical interface. Block encoding/decoding (8B/10B), alignment, deskew, and clock rate compensation are components of this clause.

Clause 49. Physical Coding Sublayer (PCS) for 64b/66b, type 10GBASE-R: This clause defines the PCS layer for connecting an XGMII to a 10-Gbps serial PMA layer. Block encoding/decoding (64B/66B), alignment, and clock rate compensation are components of this clause.

Clause 50. WAN Interface Sublayer (WIS), type 10GBASE-W: This clause defines a sublayer which exists between the PCS and the PMA layer for Wide Area Network (WAN) variants. This sublayer maps Ethernet packets into the Synchronous Payload Envelope (SPE) of an STS-192c SONET frame.

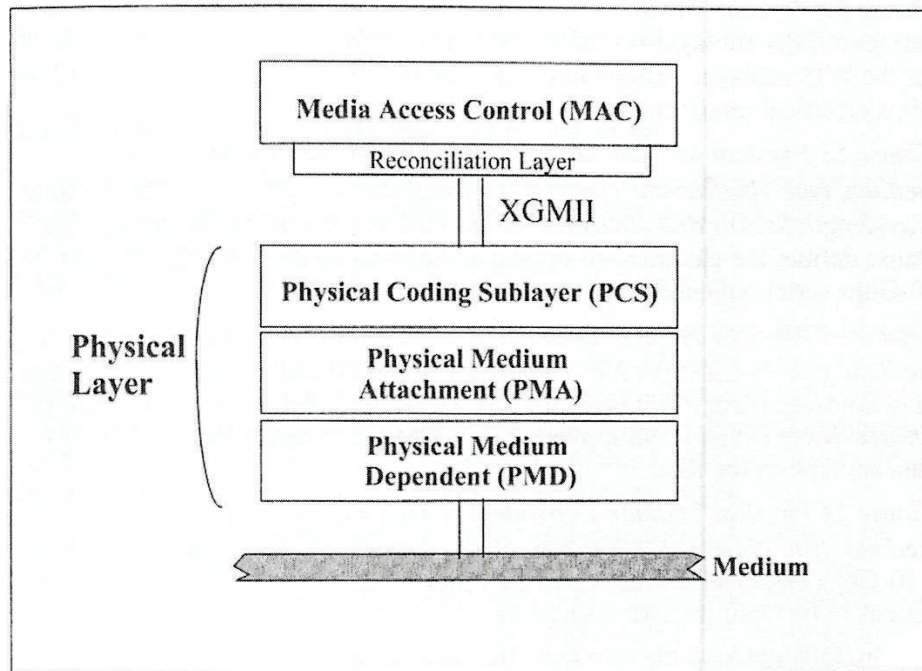


Fig. 5.14 LAN PHY/WAN PHY sublayers

Table 5.10 Relevant IEEE 802.3 clauses

Nomenclature	Type	PMD description	PCS layer clause	PMA layer clause	PMD layer clause	WIS (Clause 50)	Auto-Neg. (Clause 73)	FEC (Clause 74)
10GBASE-SR	LAN	850-nm serial optics	49	51	52			
10GBASE-LR	LAN	1,310-nm serial optics	49	51	52			
10GBASE-ER	LAN	1,550-nm serial optics	49	51	52			
10GBASE-LX4	LAN	4 x 1,310-nm CWDM optics	48	48	53			
10GBASE-CX4	LAN	4 x 3.125 Gbps electrical i/f	48	48	54			
10GBASE-SW	WAN	850-nm serial optics	49	51	52	X		
10GBASE-LW	WAN	1,310-nm serial optics	49	51	52	X		
10GBASE-EW	WAN	1,550-nm serial optics	49	51	52	X		
10GBASE-KX4	Backplane	4 x 3.125 Gbps electrical	48	48	71		X	
10GBASE-KR	Backplane	10 x Gbps serial electrical	49	51	72		X	O

Key: X = Mandatory, O = Optional

Clause 51 Physical Medium Attachment (PMA) sublayer, serial: This clause defines a PMA sublayer for interfacing the PCS sublayer defined in clause 49 (or the WIS sublayer defined in clause 50) to a PMD using a serial 10.3125-Gbps electrical interface.

Clause 52 Physical Medium Dependent (PMD) sublayer and baseband medium, type 10GBASE-S (Short Wavelength Serial) and 10GBASE-L (Long Wavelength Serial) and 10GBASE-E (Extra Long Wavelength Serial): This clause defines the electrical-to-optical and optical-to-electrical PMD layer for 10-Gbps serial optical devices.

Clause 53 Physical Medium Dependent (PMD) sublayer and baseband medium, type 10GBASE-LX4: This clause defines the electrical-to-optical and optical-to-electrical PMD layer for a 10-Gbps optical device which employs Course Wave Division Multiplexing (CWDM) to transmit four 3.125 Gbps data streams on the fiber.

Clause 54 Physical Medium Dependent (PMD) sublayer and baseband medium, type 10GBASE-CX4: This clause defines the electrical PMD layer for a 10-Gbps electrical device which uses four 3.125-Gbps electrical differential signals to transmit data across a cable.

The Ethernet variants also exist for communication across backplanes within host systems. The clauses relevant to 10-Gb Backplane Ethernet are summarized below:

Clause 69. Introduction to Ethernet operation over electrical backplanes: This clause is a normative summary of applicable requirements for the 1 Gbps and 10 Gbps backplane variants described in Table 5.10. This clause references other clauses applicable to these variants.

Clause 71. Physical Medium Dependent Sublayer and Baseband Medium, Type 10GBASE-KX4: This clause defines the electrical PMD layer for a 10-Gbps electrical device which uses four 3.125-Gbps electrical differential signals to transmit data across a backplane.

Clause 72. Physical Medium Dependent Sublayer and Baseband Medium, Type 10GBASE-KR: This clause defines the electrical PMD layer for a 10-Gbps serial electrical device which uses a 10.3125-Gbps electrical differential signal to transmit data across a backplane. This clause additionally defines a training frame structure and training state machine. This training protocol monitors the signal metrics at the receiver during link initialization and updates FFE coefficients in the transmitter in order to optimize performance of the link.

Clause 73. Auto-Negotiation for Backplane Ethernet: This clause defines autonegotiation protocols which are mandatory for PCS/PMA layer implementations for Backplane Ethernet. This protocol allows devices to negotiate which Backplane Ethernet variant is to be used, whether FEC is to be used, and other parameters.

Clause 74. Forward Error Correction (FEC) sublayer for 10GBASE-R: This clause defines an optional FEC sublayer for enhancing the BER performance of 10GBASE-KR.

5.3.2 Media Access Control (MAC) Layer

In this section the basic format of an Ethernet frame is discussed, as well as the XGMII interface definition. Other details of MAC layer function are beyond the scope of this text.

5.3.2.1 Ethernet Packet Format

Ethernet is a packet delivery system. The basic format of an Ethernet packet is shown in Fig. 5.15. Each packet consists of a preamble and start frame delimiter (SFD) to indicate the beginning of a packet, destination and source addresses to identify the recipient and the sender, a packet length field indicating the number of data bytes in the packet, a variable length data field, and a frame check sequence (FCS) field to facilitate error detection.

Each Ethernet network interface card (NIC) is assigned a unique MAC address by its manufacturer. The first 24 bits of the 48-bit MAC address identify the manufacturer and are assigned by the IEEE Registration Authority. The remaining bits are assigned by the manufacturer and are generally programmed into the hardware such that they cannot be changed. This MAC address uniquely identifies the NIC, even if the hardware moves to another location.

The data portion of the MAC frame generally carries data for a higher layer protocol, which also requires some control information to be transmitted. The data portion of the frame in Fig. 5.15 often consists of a logical link control (LLC) layer header followed by payload data. If there are fewer than 46 bytes of payload data, then the payload data is padded to achieve this minimum length.

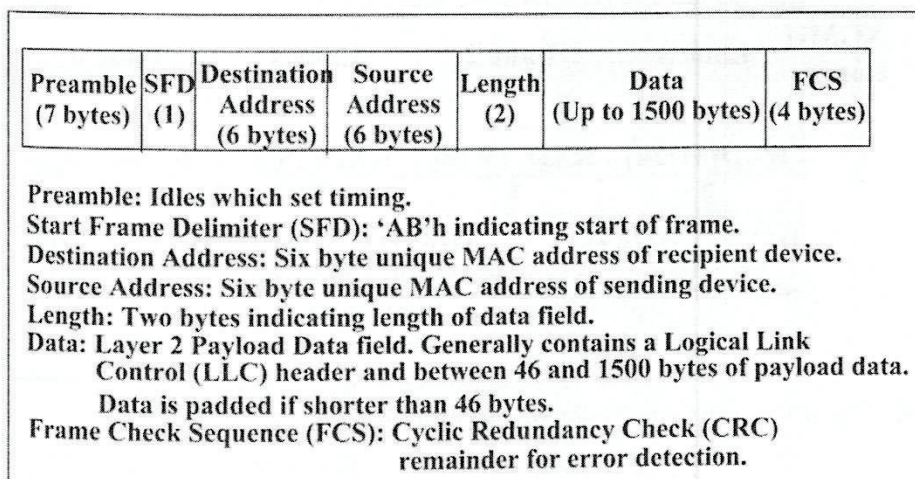


Fig. 5.15 IEEE 802.3 ethernet frame

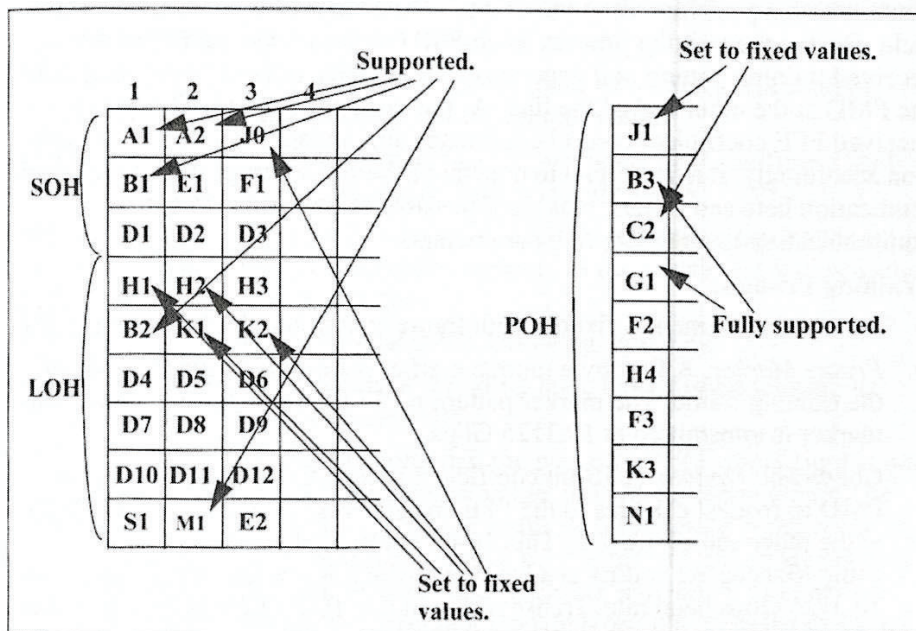


Fig. 5.20 WIS support for SONET/SDH overhead

Overhead bytes not listed above are not supported. They are set to fixed values and are not checked.

One last note on the WIS sublayer is in regard to the orientation of most significant bit and least significant bits on the interface between this layer and other sublayers. Ethernet bits are transmitted least significant bit first, while SONET/SDH octets are transmitted most significant bit first. For this reason, bit/octet ordering is interposed between the PCS and the PMA sublayers by the WIS sublayer.

5.3.5 Backplane Ethernet

There are three significant features associated with Backplane Ethernet which are not used in other Ethernet variants. These features are:

- Training protocol (defined in Clause 72 for 10GBASE-KR)
- Autonegotiation (defined in Clause 73 for all Backplane Ethernet variants)
- Forward error correction (FEC) (defined in Clause 74 for optional use with 10GBASE-KR)

5.3.5.1 Training Protocol

The training protocol is included in the IEEE 802.3 clause 72 specification for the 10GBASE-KR variant and must be executed upon start-up to initialize the PMD for this variant. This protocol continuously exchanges fixed-length training frames between the PMD and the link partner. The training frames

contain both a pseudorandom training pattern and an FFE coefficient control field. During the training process, each PMD monitors the quality of the received training pattern and generates FFE coefficient control commands to the PMD at the other end of the link. At the same time the PMD monitors received FFE coefficient control commands and updates the HSS configuration accordingly. It is important to note that Ethernet assumes full-duplex communication between PMDs, making it possible to perform two-way communication to negotiate link parameters.

Training Frames:

The training frame is a fixed length frame containing the following fields:

- *Frame Marker.* A four byte unique marker pattern denoting the start of the training frame. The marker pattern is "FFFF0000" hexadecimal. This marker is transmitted at 10.3125 Gbps.
- *Coefficient Update.* A 16-bit coefficient update field which allows the PMD to request changes to the FFE coefficients being used by the PMD at the other end of the link. This 16-bit coefficient field is transmitted using Manchester coding at a baud rate which is one-eighth of the normal 10.3125 Gbps baud rate. Transmission of this 16-bit field is the equivalent of transmitting 16 bytes at 10.3125 Gbps.
- *Status Report.* A 16-bit status field which indicates whether the PMD has finished its training process, as well as the status of any requested FFE coefficient updates. This 16-bit coefficient field is transmitted using Manchester coding at a baud rate which is one-eighth of the normal 10.3125 Gbps baud rate. Transmission of this 16-bit field is the equivalent of transmitting 16 bytes at 10.3125 Gbps.
- *Training Pattern.* Pseudorandom data generated with the polynomial:

$$G(x) = 1 + x^9 + x^{11}$$
 This field is 512 bytes in length (4,094 bits PRBS bits followed by two zero bits), transmitted at 10.3125 Gbps.

The purpose of using low-speed Manchester encoding for the *Coefficient Update* and *Status Report* fields is to ensure reliable information is communicated even when the link performance is not optimized. These fields can be implemented by transmitting the parallel data byte "00000000" or "11111111" to represent a Manchester encoded "0" bit, and transmitting the parallel data byte "00001111" or "11110000" to represent a Manchester encoded "1" bit. (The byte pattern used to denote a Manchester encoded "0" or "1" is selected to ensure that a transition always occurs between bits.)

The *Coefficient Update* field contains the following subfields:

- *Preset.* Forces all FFE coefficients to be set to a state where equalization is turned off (i.e., coefficients for the main FFE tap is set to its maximum value, and coefficients for precursor and postcursor taps are set to zero).
- *Initialize.* Forces FFE to be turned on and coefficients to be set to predefined initialization values.

- *Coefficient (+1) Update*. Directs that the postcursor FFE coefficient shall be incremented, decremented, or held at current value. .
- *Coefficient (0) Update*. Directs that the main FFE coefficient shall be incremented, decremented, or held at current value.
- *Coefficient (-1) Update*. Directs that the precursor FFE coefficient shall be incremented, decremented, or held at current value.

Update requests to increment or decrement coefficients should only be sent if the received *Status Report* field corresponding to the coefficient indicates the coefficient is in *not updated* state.

The *Status Report* field contains the following subfields:

- *Receiver Ready*. Indicates that reporting node has finished training its receiver and is ready to receive data.
- *Coefficient (+1) Status*. Reports that the postcursor FFE coefficient is *not updated*, is *updated*, is at *maximum* value, or is at *minimum* value. The PMD initially reports all coefficient status as *not updated*. After receiving a Coefficient Update request to update the coefficient, the PMD updates the coefficient and then reports either *updated*, *maximum*, or *minimum* status. After receiving a Coefficient Update request to hold the previous value, the PMD returns status to *not updated*.
- *Coefficient (0) Status*. Reports that the main FFE coefficient is *not updated*, is *updated*, is at *maximum* value, or is at *minimum* value. Status transitions are similar to the description for the postcursor FFE coefficient.
- *Coefficient (-1) Status*. Reports that the precursor FFE coefficient is *not updated*, is *updated*, is at *maximum* value, or is at *minimum* value. Status transitions are similar to the description for the postcursor FFE coefficient.

The *coefficient status* and *coefficient update* fields associated with each FFE coefficient operate according to a simple two-way handshake as described above. Each PMD assesses the quality of the received signal (through analysis of the training pattern), and then instructs its partner node to update FFE coefficients in an attempt to improve the quality of the signal.

Training State Machine:

The training state machine is also defined by clause 72. When a node detects a signal on the link, it starts sending training frames and waits for reception of valid training frames from its link partner. Once training frames are being received, the node allows the receiver equalization to train, and determines whether updates are required to the transmitter equalization. FFE coefficient updates are requested via command bits in the transmitted training frame; the link partner's acknowledgement of these requests are communicated via the status bits in the received training frame. Simultaneously, the node monitors command bits in the received training frame to determine whether local FFE coefficient updates are being requested

by the link partner, and if applicable executes these requests and returns an acknowledgement via the transmitted training frame.

It should be noted that clause 72 does *not* specify the algorithms which should be employed to determine optimal equalizer settings. These algorithms are left as an exercise to the designer, and some implementations may even choose to use preset FFE coefficient values. (Although such implementations may not work well in some system configurations.)

The methods of determining signal quality and criteria for exiting training are also *not* specified by clause 72. These methods and criteria may be based on the PRBS pattern in the training frame, or may use other proprietary features of the HSS core.

HSS Features:

Some HSS cores may implement features which aid in signal quality assessment and allow faster updates to FFE coefficient values. Features such as the following are helpful to the implementation of the clause 72 Training Protocol:

- FFE coefficient negotiation requires assessment of the eye quality of the signal being received by the receive PMD layer. This signal quality is not directly observable by measurements at higher sublayers. The HSS EX10 digital eye feature assesses eye quality and provides training state machines with quick access to these measurements. This is useful for implementing the training protocol.
- FFE coefficient updates by a partner link during the training process should be made in a timely manner. HSS core features to quickly reset, load, increment, or decrement FFE coefficients speed up training.

5.3.5.2 Autonegotiation

The autonegotiation protocol is defined in IEEE 802.3 clause 73. Backplane Ethernet PHYs are required to implement the autonegotiation protocol, however, use of the protocol is under the control of the management interface and is optional. The protocol allows a device to advertise the modes of operation it supports to another device at the remote end of a backplane Ethernet link, and to detect the corresponding operational modes being advertised by the other device. The objective is to allow the devices to agree to a configuration that maximizes the performance of the link. The protocol additionally ensures that the PHY device is attached to its link partner at the remote end of the link, and is not responding to a crosstalk signal.

Autonegotiation is performed using low-speed Manchester encoding of the information to be exchanged. Clause 73 describes a number of Differential Manchester Encoding (DME) pages which are exchanged by the protocol. Autonegotiation supports the following partial list of functions:

- Advertise which backplane Ethernet variants are supported and negotiate which variant should be selected.
- Advertise whether the interface supports FEC and negotiate whether to enable FEC.

High Speed Serdes Devices and Applications



David R. Stauffer, Jeanne Trinko Mechler, Michael Sorna, Kent Dramstad,
Clarence R. Ogilvie, Amanullah Mohammad and James Rockrohr

High Speed Serdes Devices and Applications provides a broad understanding of High Speed Serdes (HSS) device usage. The material focuses on HSS devices, and the consolidation of related topics into a single text. Chip and system designers using HSS devices must have detailed knowledge of both the features and functions of the HSS device, and the applications in which they are used. Also designers must have a working knowledge of related subjects, including: reference clock architectures, signal integrity, power dissipation, and test features and functions. The authors consolidate these topics with a specific focus on HSS devices. This approach provides the chip designer sufficient background information for using HSS devices on their chips.

The chapters can be viewed as four distinct sections. The first section relates to the features, functions, and design of HSS devices. Second are chapters that describe the features and functions of protocol logic used to implement various network protocol interface standards. The third section covers specialized topics related to HSS cores. Finally, characteristics of typical design kit models to facilitate integration within the chip design are discussed.

High Speed Serdes Devices and Applications is a useful resource for chip designers using HSS devices in their chip design.

springer.com



