Information technology—
Telecommunications and information exchange between systems—
Local and metropolitan area networks—Specific requirements—

Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) access method and physical layer specifications

SECTION THREE: This section includes Clauses 34 through 43 and Annexes 36A through 43C.

34. Introduction to 1000 Mb/s baseband network

34.1 Overview

Gigabit Ethernet couples an extended version of the ISO/IEC 8802-3 (CSMA/CD MAC) to a family of 1000 Mb/s Physical Layers. The relationships among Gigabit Ethernet, the extended ISO/IEC 8802-3 (CSMA/CD MAC), and the ISO/IEC Open System Interconnection (OSI) reference model are shown in Figure 34–1.

Gigabit Ethernet uses the extended ISO/IEC 8802-3 MAC layer interface, connected through a Gigabit Media Independent Interface layer to Physical Layer entities (PHY sublayers) such as 1000BASE-LX, 1000BASE-SX, and 1000BASE-CX, and 1000BASE-T.

Gigabit Ethernet extends the ISO/IEC 8802-3 MAC beyond 100 Mb/s to 1000 Mb/s. The bit rate is faster, and the bit times are shorter—both in proportion to the change in bandwidth. In full duplex mode, the minimum packet transmission time has been reduced by a factor of ten. Achievable topologies for 1000 Mb/s full duplex operation are comparable to those found in 100BASE-T full duplex mode. In half duplex mode, the minimum packet transmission time has been reduced, but not by a factor of ten. Cable delay budgets are similar to those in 100BASE-T. The resulting achievable topologies for the half duplex 1000 Mb/s CSMA/CD MAC are similar to those found in half duplex 100BASE-T.

34.1.1 Reconciliation Sublayer (RS) and Gigabit Media Independent Interface (GMII)

The Gigabit Media Independent Interface (Clause 35) provides an interconnection between the Media Access Control (MAC) sublayer and Physical Layer entities (PHY) and between PHY Layer and Station Management (STA) entities. This GMII supports 1000 Mb/s operation through its eight bit wide (octet wide) transmit and receive paths. The Reconciliation sublayer provides a mapping between the signals provided at the GMII and the MAC/PLS service definition.

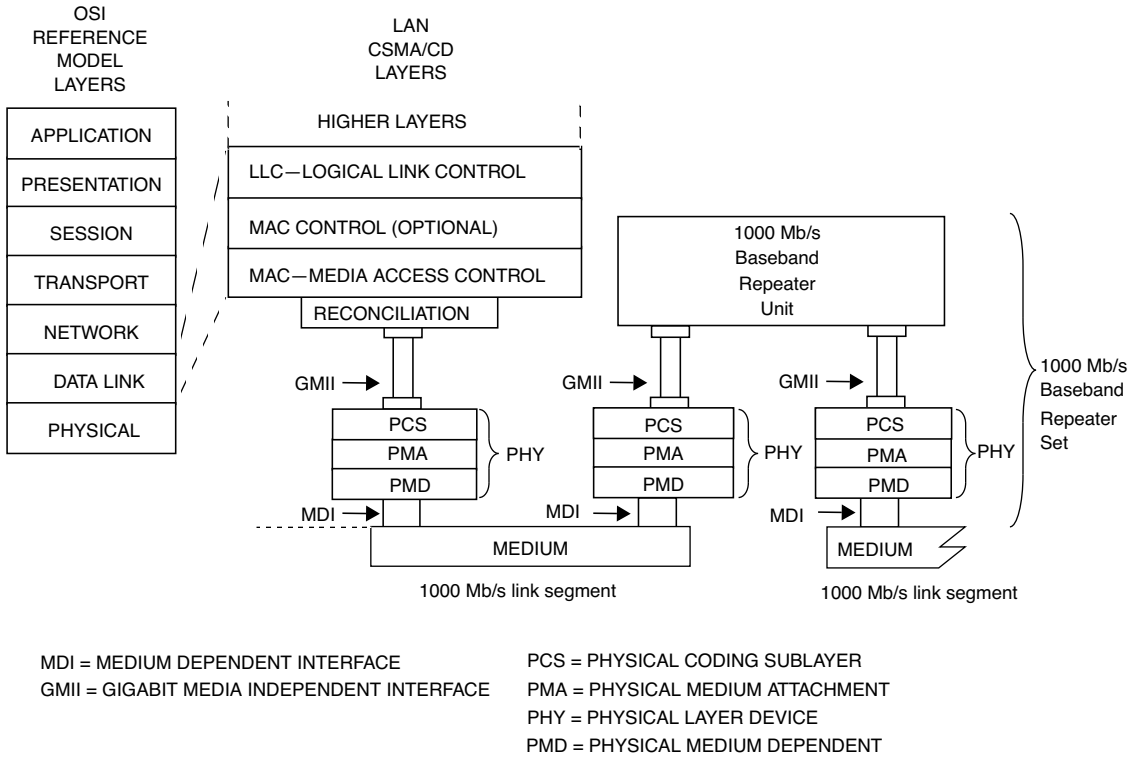


Figure 34–1—Architectural positioning of Gigabit Ethernet (1000 Mb/s operation)

34.1.2 Physical Layer signaling systems

This standard specifies a family of Physical Layer implementations. The generic term 1000 Mb/s MAC refers to any use of the 1000 Mb/s ISO/IEC 8802-3 CSMA/CD MAC (the Gigabit Ethernet MAC) coupled with any physical layer implementation.

The term 1000BASE-X refers to a specific family of physical layer implementations specified in Clauses 36–39. The 1000BASE-X family of physical layer standards has been adapted from the ANSI X3.230-1994 [B20] (Fibre Channel) FC-0 and FC-1 physical layer specifications and the associated 8B/10B data coding method. The 1000BASE-X family of physical layer implementations is composed of 1000BASE-SX, 1000BASE-LX, and 1000BASE-CX.

All 1000BASE-X PHY devices share the use of common PCS, PMA, and Auto-Negotiation specifications (see Clauses 36 and 37). The 1000BASE-T PHY (Clause 40) uses four pairs of Category 5 balanced copper cabling. Clause 40 defines its own PCS, which does not use 8B/10B coding.

Specifications unique to the physical operation of each physical layer device are shown in the following table:

1000BASE-SX Short Wave Length Optical	Duplex multimode fibers	Clause 38
1000BASE-LX Long Wave Length Optical	Duplex single-mode fibers or Duplex multimode fibers	Clause 38
1000BASE-CX Shielded Jumper Cable	Two pairs of specialized balanced cabling	Clause 39
1000BASE-T Category 5 UTP	Advanced multilevel signaling over four pairs of Category 5 balanced copper cabling.	Clause 40

34.1.3 Repeater

A repeater set (Clause 41) is an integral part of any Gigabit Ethernet network with more than two DTEs in a collision domain. A repeater set extends the physical system topology by coupling two or more segments. Only one repeater is permitted within a single collision domain.

34.1.4 Auto-Negotiation, type 1000BASE-X

Auto-Negotiation (Clause 37) provides a 1000BASE-X device with the capability to detect the abilities (modes of operation) supported by the device at the other end of a link segment, determine common abilities, and configure for joint operation. Auto-Negotiation is performed upon link startup through the use of a special sequence of reserved link code words. Clause 37 adopts the basic architecture and algorithms from Clause 28, but not the use of fast link pulses.

34.1.5 Physical Layer line signaling for 10 Mb/s and 100 Mb/s Auto-Negotiation on twisted pair

Auto-Negotiation (Clause 28) is used by 1000BASE-T devices to detect the abilities (modes of operation) supported by the device at the other end of a link segment, determine common abilities, and configure for joint operation. Auto-Negotiation is performed upon link startup through the use of a special sequence of fast link pulses.

34.1.6 Management

Managed objects, attributes, and actions are defined for all Gigabit Ethernet components (Clause 30). That clause consolidates all IEEE 802.3® management specifications so that 10/100/1000 Mb/s agents can be managed by existing network management stations with little or no modification to the agent code.

34.2 State diagrams

State machine diagrams take precedence over text.

The conventions of 1.2 are adopted, along with the extensions listed in 21.5.

34.3 Protocol Implementation Conformance Statement (PICS) proforma

The supplier of a protocol implementation that is claimed to conform to any part of IEEE 802.3®, Clauses 35 through 41, shall complete a Protocol Implementation Conformance Statement (PICS) proforma.

A completed PICS proforma is the PICS for the implementation in question. The PICS is a statement of which capabilities and options of the protocol have been implemented. A PICS is included at the end of each clause as appropriate. Each of the Gigabit Ethernet PICS conforms to the same notation and conventions used in 100BASE-T (see 21.6).

34.4 Relation of Gigabit Ethernet to other standards

Suitable entries for Table G1 of ISO/IEC 11801: 1995, annex G, would be as follows:

a) Within the section Optical Link:
CSMA/CD 1000BASE-SX ISO/IEC 8802-3/ PDAM 26
b) Within the section Optical Link:
CSMA/CD 1000BASE-LX ISO/IEC 8802-3/PDAM 26
c) Within the section Balanced Cabling Link Class D (defined up to 100MHz):
CSMA/CD 1000BASE-T* ISO/IEC8802-3/DAD 1995

*To support 1000BASE-T applications, Class D links shall meet the requirements for return loss, ELFEXT and MDELFEXT specified in 40.7.

A suitable entry for Table G5 of ISO/IEC 11801: 1995, Annex G, would be as follows:

Table 34–1—Table G5 of ISO/IEC 11801

	Fibre			**Optical link per 6**								
	per Clauses 5, 7, and 8			**Horizontal**			**Building backbone**			**Campus backbone**		
	62.5/125 μm MMF	**50/125 μm MMF**	**10/125 μm SMF**	**62.5/125 μm MMF**	**50/125 μm MMF**	**10/125 μm SMF**	**62.5/125 μm MMF**	**50/125 μm MMF**	**10/125 μm SMF**	**62.5/125 μm MMF**	**50/125 μm MMF**	**10/125 μm SMF**
8802-3: 1000BASE-SX	I	I		N	N		I	N		I	I	
8802-3: 1000BASE-LX	I	I	I	N	N	N	N	N	N	I	I	N
NOTE—"N" denotes normative support of the media in the standard. "I" denotes that there is information in the International Standard regarding operation on this media.												

Suitable entries for table G4 of ISO/IEC 11801:1995 Annex G would be:

Table 34–2—Table G4 of ISO/IEC 11801:1995

	Balanced cabling							**Performance based cabling per 6**												
	per Clauses 5, 7, and 8							**Class A**			**Class B**			**Class C**			**Class D**			
	Cat3 100Ω	**Cat4 100Ω**	**Cat5 100Ω**	**Cat3 120Ω**	**Cat4 120Ω**	**Cat5 120Ω**	**150Ω**	**100Ω**	**120Ω**	**150Ω**	**100Ω**	**120Ω**	**150Ω**	**100Ω**	**120Ω**	**150Ω**	**100Ω**	**120Ω**	**150Ω**
8802-3: 1000BASE-T			I[a]														I[a]		
NOTE—"I" denotes that there is information in the International Standard regarding operation on this media.																				

[a]8802-3 imposes additional requirements on return loss, ELFEXT and MDELFEXT.

35. Reconciliation Sublayer (RS) and Gigabit Media Independent Interface (GMII)

35.1 Overview

This clause defines the logical and electrical characteristics for the Reconciliation Sublayer (RS) and Gigabit Media Independent Interface (GMII) between CSMA/CD media access controllers and various PHYs. Figure 35–1 shows the relationship of the Reconciliation sublayer and GMII to the ISO/IEC OSI reference model.

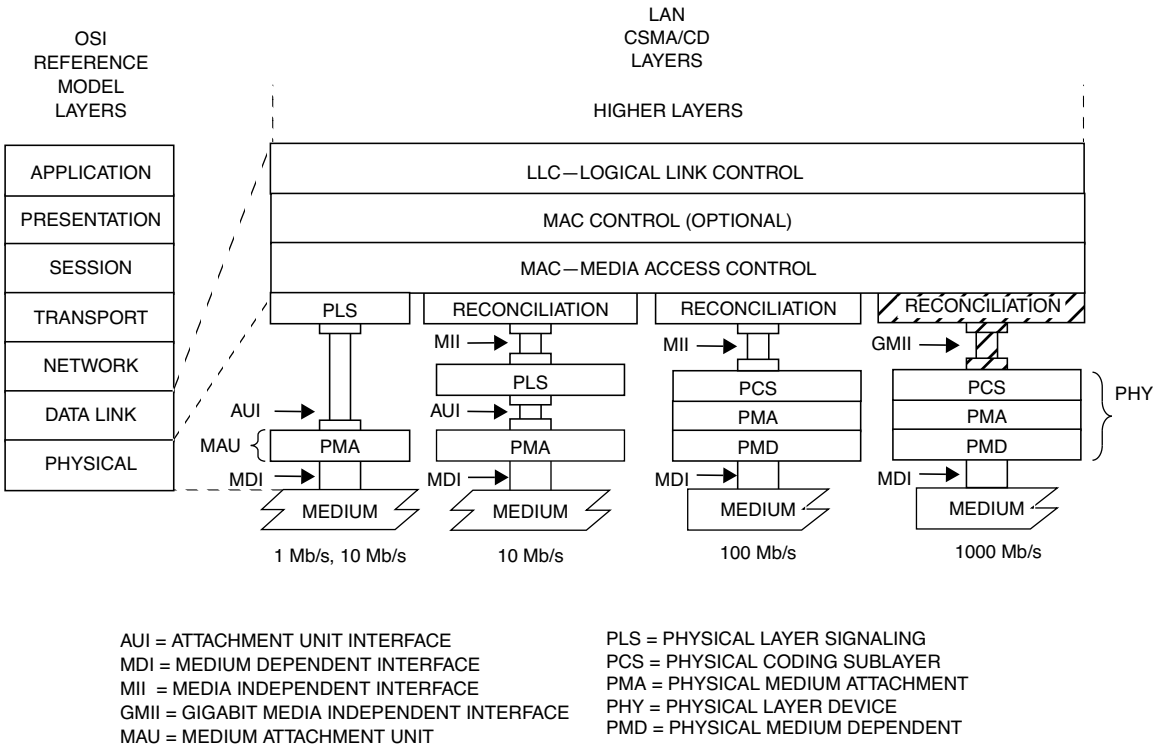


Figure 35–1—GMII location in the OSI protocol stack

The purpose of this interface is to provide a simple, inexpensive, and easy-to-implement interconnection between Media Access Control (MAC) sublayer and PHYs, and between PHYs and Station Management (STA) entities.

This interface has the following characteristics:

a) It is capable of supporting 1000 Mb/s operation.
b) Data and delimiters are synchronous to clock references.
c) It provides independent eight-bit-wide transmit and receive data paths.
d) It provides a simple management interface.
e) It uses signal levels, compatible with common CMOS digital ASIC processes and some bipolar processes.
f) It provides for full duplex operation.

35.1.1 Summary of major concepts

a) The GMII is based on the MII defined in Clause 22.
b) Each direction of data transfer is serviced by Data (an eight-bit bundle), Delimiter, Error, and Clock signals.
c) Two media status signals are provided. One indicates the presence of carrier, and the other indicates the occurrence of a collision.
d) The GMII uses the MII management interface composed of two signals that provide access to management parameters and services as specified in Clause 22.
e) MII signal names have been retained and the functions of most signals are the same, but additional valid combinations of signals have been defined for 1000 Mb/s operation.
f) The Reconciliation sublayer maps the signal set provided at the GMII to the PLS service primitives provided to the MAC.
g) GMII signals are defined such that an implementation may multiplex most GMII signals with the similar PMA service interface defined in Clause 36.

35.1.2 Application

This clause applies to the interface between the MAC and PHYs, and between PHYs and Station Management entities. The implementation of the interface is primarily intended as a chip-to-chip (integrated circuit to integrated circuit) interface implemented with traces on a printed circuit board. A motherboard-to-daughterboard interface between two or more printed circuit boards is not precluded.

This interface is used to provide media independence so that an identical media access controller may be used with any of the copper and optical PHY types.

35.1.3 Rate of operation

The GMII supports only 1000 Mb/s operation and is defined within this clause. Operation at 10 Mb/s and 100 Mb/s is supported by the MII defined in Clause 22.

PHYs that provide a GMII shall support 1000 Mb/s operation, and may support additional rates using other interfaces (e.g., MII). PHYs must report the rates at which they are capable of operating via the management interface, as described in 22.2.4. Reconciliation sublayers that provide a GMII shall support 1000 Mb/s and may support additional rates using other interfaces.

35.1.4 Allocation of functions

The allocation of functions at the GMII balances the need for media independence with the need for a simple and cost-effective interface.

While the Attachment Unit Interface (AUI) was defined to exist between the Physical Signaling (PLS) and Physical Medium Attachment (PMA) sublayers for 10 Mb/s DTEs, the GMII (like the Clause 22 MII) maximizes media independence by cleanly separating the Data Link and Physical Layers of the ISO/IEC seven-layer reference model. This allocation also recognizes that implementations can benefit from a close coupling between the PLS or PCS sublayer and the PMA sublayer.

35.2 Functional specifications

The GMII is designed to make the differences among the various media transparent to the MAC sublayer. The selection of logical control signals and the functional procedures are all designed to this end.

35.2.1 Mapping of GMII signals to PLS service primitives and Station Management

The Reconciliation sublayer maps the signals provided at the GMII to the PLS service primitives defined in Clause 6. The PLS service primitives provided by the Reconciliation sublayer, and described here, behave in exactly the same manner as defined in Clause 6.

Figure 35–2 depicts a schematic view of the Reconciliation sublayer inputs and outputs, and demonstrates that the GMII management interface is controlled by the Station Management entity (STA).

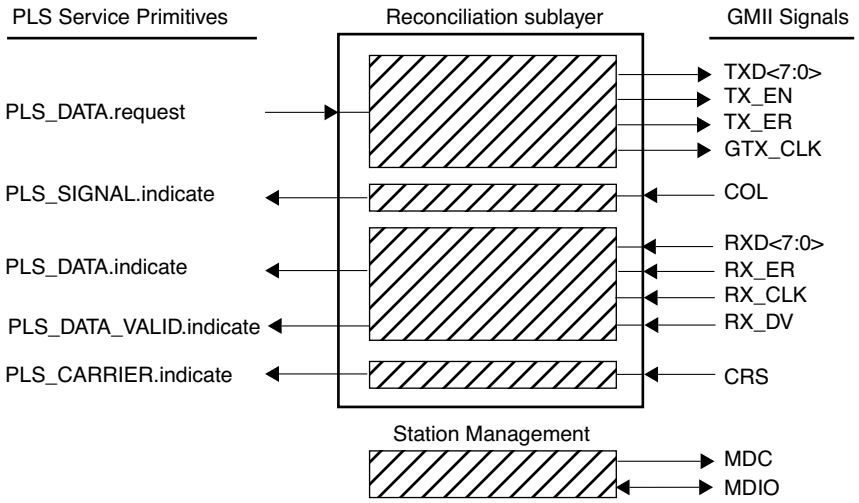


Figure 35–2—Reconciliation Sublayer (RS) inputs and outputs and STA connections to GMII

35.2.1.1 Mapping of PLS_DATA.request

35.2.1.1.1 Function

Map the primitive PLS_DATA.request to the GMII signals TXD<7:0>, TX_EN, TX_ER and GTX_CLK.

35.2.1.1.2 Semantics of the service primitive

PLS_DATA.request (OUTPUT_UNIT)

The OUTPUT_UNIT parameter can take one of five values: ONE, ZERO, TRANSMIT_COMPLETE, EXTEND or EXTEND_ERROR. It represents or is equivalent to a single data bit. These values are conveyed by the signals TX_EN, TX_ER, TXD<7>, TXD<6>, TXD<5>, TXD<4>, TXD<3>, TXD<2>, TXD<1> and TXD<0>.

Each of the eight TXD signals conveys either a ONE or ZERO of data while TX_EN is asserted. Eight data bit equivalents of EXTEND or EXTEND_ERROR are conveyed by a specific encoding of the TXD<7:0> signals when TX_EN is not asserted, and TX_ER is asserted, see Table 35–1. Synchronization between the Reconciliation sublayer and the PHY is achieved by way of the GTX_CLK signal. The value TRANSMIT_COMPLETE is conveyed by the de-assertion of either TX_EN or TX_ER at the end of a MAC's transmission.

35.2.1.1.3 When generated

The GTX_CLK signal is generated by the Reconciliation sublayer. The TXD<7:0>, TX_EN and TX_ER signals are generated by the Reconciliation sublayer after every group of eight PLS_DATA.request transactions from the MAC sublayer to request the transmission of eight data bits on the physical medium, to extend the carrier event the equivalent of eight bits, or to stop transmission.

35.2.1.2 Mapping of PLS_DATA.indicate

35.2.1.2.1 Function

Map the primitive PLS_DATA.indicate to the GMII signals RXD<7:0>, RX_DV, RX_ER, and RX_CLK.

35.2.1.2.2 Semantics of the service primitive

PLS_DATA.indicate (INPUT_UNIT)

The INPUT_UNIT parameter can take one of three values: ONE, ZERO or EXTEND. It represents or is equivalent to a single data bit. These values are derived from the signals RX_DV, RX_ER, RXD<7>, RXD<6>, RXD<5>, RXD<4>, RXD<3>, RXD<2>, RXD<1>, and RXD<0>. The value of the data transferred to the MAC is controlled by GMII error indications, see 35.2.1.5.

Each of the eight RXD signals conveys either a ONE or ZERO of data while RX_DV is asserted. Eight data bit equivalents of EXTEND are conveyed by a specific encoding of the RXD<7:0> signals when RX_DV is not asserted, and RX_ER is asserted; see Table 35–2. Synchronization between the Reconciliation sublayer and the PHY is achieved by way of the RX_CLK signal.

35.2.1.2.3 When generated

This primitive is generated to all MAC sublayer entities in the network after a PLS_DATA.request is issued. Each octet transferred on RXD<7:0> will result in the generation of eight PLS_DATA.indicate transactions.

35.2.1.3 Mapping of PLS_CARRIER.indicate

35.2.1.3.1 Function

Map the primitive PLS_CARRIER.indicate to the GMII signal CRS.

35.2.1.3.2 Semantics of the service primitive

PLS_CARRIER.indicate (CARRIER_STATUS)

The CARRIER_STATUS parameter can take one of two values: CARRIER_ON or CARRIER_OFF. CARRIER_STATUS assumes the value CARRIER_ON when the GMII signal CRS is asserted and assumes the value CARRIER_OFF when CRS is de-asserted.

35.2.1.3.3 When generated

The PLS_CARRIER.indicate service primitive is generated by the Reconciliation sublayer whenever the CARRIER_STATUS parameter changes from CARRIER_ON to CARRIER_OFF or vice versa.

35.2.1.4 Mapping of PLS_SIGNAL.indicate

35.2.1.4.1 Function

Map the primitive PLS_SIGNAL.indicate to the GMII signal COL.

35.2.1.4.2 Semantics of the service primitive

PLS_SIGNAL.indicate (SIGNAL_STATUS)

The SIGNAL_STATUS parameter can take one of two values: SIGNAL_ERROR or NO_SIGNAL_ERROR. SIGNAL_STATUS assumes the value SIGNAL_ERROR when the GMII signal COL is asserted, and assumes the value NO_SIGNAL_ERROR when COL is de-asserted.

35.2.1.4.3 When generated

The PLS_SIGNAL.indicate service primitive is generated whenever SIGNAL_STATUS makes a transition from SIGNAL_ERROR to NO_SIGNAL_ERROR or vice versa.

35.2.1.5 Response to error indications from GMII

If, during frame reception, both RX_DV and RX_ER are asserted, the Reconciliation sublayer shall ensure that the MAC will detect a FrameCheckError in that frame.

Carrier is extended when RX_DV is not asserted and RX_ER is asserted with a proper encoding of RXD<7:0>. When a Carrier Extend Error is received during the extension, the Reconciliation sublayer shall send PLS_DATA.indicate values of ONE or ZERO and ensure that the MAC will detect a FrameCheckError in the sequence.

These requirements may be met by incorporating a function in the Reconciliation sublayer that produces a received frame data sequence delivered to the MAC sublayer that is guaranteed to not yield a valid CRC result, as specified by the algorithm in 3.2.8. This data sequence may be produced by substituting data delivered to the MAC.

Other techniques may be employed to respond to Data Reception Error or Carrier Extend Error provided that the result is that the MAC sublayer behaves as though a FrameCheckError occurred in the received frame.

35.2.1.6 Conditions for generation of TX_ER

If, during the process of transmitting a frame, it is necessary to request that the PHY deliberately corrupt the contents of the frame in such a manner that a receiver will detect the corruption with the highest degree of probability, then Transmit Error Propagation shall be asserted by the appropriate encoding of TX_ER, and TX_EN. Similarly, if during the process of transmitting carrier extension to a frame, it is necessary to request that the PHY deliberately corrupt the contents of the carrier extension in such a manner that a receiver will detect the corruption with the highest degree of probability, then Carrier Extend Error shall be signalled by the appropriate encoding of TXD<7:0>.

This capability has additional use within a repeater. For example, a repeater that detects an RX_ER during frame reception on an input port may propagate that error indication to its output ports by asserting TX_ER during the process of transmitting that frame.

35.2.1.7 Mapping of PLS_DATA_VALID.indicate

35.2.1.7.1 Function

Map the primitive PLS_DATA_VALID.indicate to the GMII signals RX_DV, RX_ER and RXD<7:0>.

35.2.1.7.2 Semantics of the service primitive

PLS_DATA_VALID.indicate (DATA_VALID_STATUS)

The DATA_VALID_STATUS parameter can take one of two values: DATA_VALID or DATA_NOT_VALID. DATA_VALID_STATUS assumes the value DATA_VALID when the GMII signal RX_DV is asserted, or when RX_DV is not asserted, RX_ER is asserted and the values of RXD<7:0> indicate Carrier Extend or Carrier Extend Error. DATA_VALID_STATUS assumes the value DATA_NOT_VALID at all other times.

35.2.1.7.3 When generated

The PLS_DATA_VALID.indicate service primitive is generated by the Reconciliation sublayer whenever DATA_VALID_STATUS parameter changes from DATA_VALID to DATA_NOT_VALID or vice versa.

35.2.2 GMII signal functional specifications

35.2.2.1 GTX_CLK (1000 Mb/s transmit clock)

GTX_CLK is a continuous clock used for operation at 1000 Mb/s. GTX_CLK provides the timing reference for the transfer of the TX_EN, TX_ER, and TXD signals from the Reconciliation sublayer to the PHY. The values of TX_EN, TX_ER, and TXD are sampled by the PHY on the rising edge of GTX_CLK. GTX_CLK is sourced by the Reconciliation sublayer.

The GTX_CLK frequency is nominally 125 MHz, one-eighth of the transmit data rate.

35.2.2.2 RX_CLK (receive clock)

RX_CLK is a continuous clock that provides the timing reference for the transfer of the RX_DV, RX_ER and RXD signals from the PHY to the Reconciliation sublayer. RX_DV, RX_ER and RXD are sampled by the Reconciliation sublayer on the rising edge of RX_CLK. RX_CLK is sourced by the PHY.

The PHY may recover the RX_CLK from the received data or it may derive the RX_CLK reference from a local clock (e.g., GTX_CLK). When derived from the received data, RX_CLK shall have a frequency equal to one-eighth of the data rate of the received signal, and when derived from a local clock a nominal frequency of 125 MHz.

When the signal received from the medium is continuous and the PHY can recover the RX_CLK reference and supply the RX_CLK on a continuous basis, there is no need to transition between the recovered clock reference and a local clock reference on a frame-by-frame basis. If loss of received signal from the medium causes a PHY to lose the recovered RX_CLK reference, the PHY shall source the RX_CLK from a local clock reference.

Transitions from local clock to recovered clock or from recovered clock to local clock shall be made only while RX_DV and RX_ER are de-asserted. During the interval between the assertion of CRS and the assertion of RX_DV at the beginning of a frame, the PHY may extend a cycle of RX_CLK by holding it in either the high or low condition until the PHY has successfully locked onto the recovered clock. Following the de-assertion of RX_DV at the end of a frame, or the de-assertion of RX_ER at the end of carrier extension, the PHY may extend a cycle of RX_CLK by holding it in either the high or low condition for an interval that shall not exceed twice the nominal clock period.

NOTE—This standard neither requires nor assumes a guaranteed phase relationship between the RX_CLK and GTX_CLK signals. See additional information in 35.4.

35.2.2.3 TX_EN (transmit enable)

TX_EN in combination with TX_ER indicates the Reconciliation sublayer is presenting data on the GMII for transmission. It shall be asserted by the Reconciliation sublayer synchronously with the first octet of the preamble and shall remain asserted while all octets to be transmitted are presented to the GMII. TX_EN shall be negated prior to the first rising edge of GTX_CLK following the final data octet of a frame. TX_EN is driven by the Reconciliation sublayer and shall transition synchronously with respect to the GTX_CLK.

Figure 35–3 depicts TX_EN behavior during a frame transmission with no collisions and without carrier extension or errors.

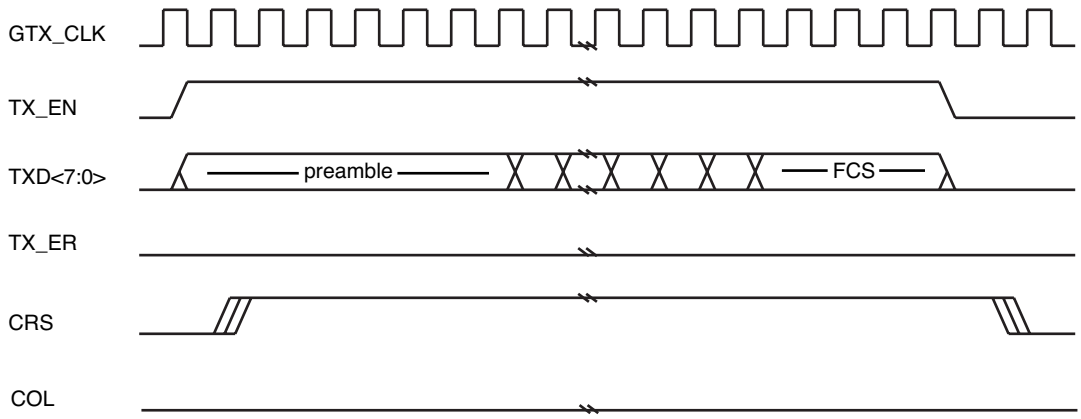


Figure 35–3—Basic frame transmission

35.2.2.4 TXD (transmit data)

TXD is a bundle of eight data signals (TXD<7:0>) that are driven by the Reconciliation sublayer. TXD<7:0> shall transition synchronously with respect to the GTX_CLK. For each GTX_CLK period in which TX_EN is asserted and TX_ER is de-asserted, data are presented on TXD<7:0> to the PHY for transmission. TXD<0> is the least significant bit. While TX_EN and TX_ER are both de-asserted, TXD<7:0> shall have no effect upon the PHY.

While TX_EN is de-asserted and TX_ER is asserted, TXD<7:0> are used to request the PHY to generate Carrier Extend or Carrier Extend Error code-groups. The use of TXD<7:0> during the transmission of a frame with carrier extension is described in 35.2.2.5. Carrier extension shall only be signalled immediately following the data portion of a frame.

Table 35–1 specifies the permissible encodings of TXD<7:0>, TX_EN, and TX_ER.

Table 35–1—Permissible encodings of TXD<7:0>, TX_EN, and TX_ER

TX_EN	TX_ER	TXD<7:0>	Description	PLS_DATA.request parameter
0	0	00 through FF	Normal inter-frame	TRANSMIT_COMPLETE
0	1	00 through 0E	Reserved	—
0	1	0F	Carrier Extend	EXTEND (eight bits)
0	1	10 through 1E	Reserved	—
0	1	1F	Carrier Extend Error	EXTEND_ERROR (eight bits)
0	1	20 through FF	Reserved	—
1	0	00 through FF	Normal data transmission	ZERO, ONE (eight bits)
1	1	00 through FF	Transmit error propagation	No applicable parameter
NOTE—Values in TXD<7:0> column are in hexadecimal.				

35.2.2.5 TX_ER (transmit coding error)

TX_ER is driven by the Reconciliation Sublayer and shall transition synchronously with respect to the GTX_CLK. When TX_ER is asserted for one or more TX_CLK periods while TX_EN is also asserted, the PHY shall emit one or more code-groups that are not part of the valid data or delimiter set somewhere in the frame being transmitted. The relative position of the error within the frame need not be preserved. Figure 35–4 shows the behavior of TX_ER during the transmission of a frame propagating an error.

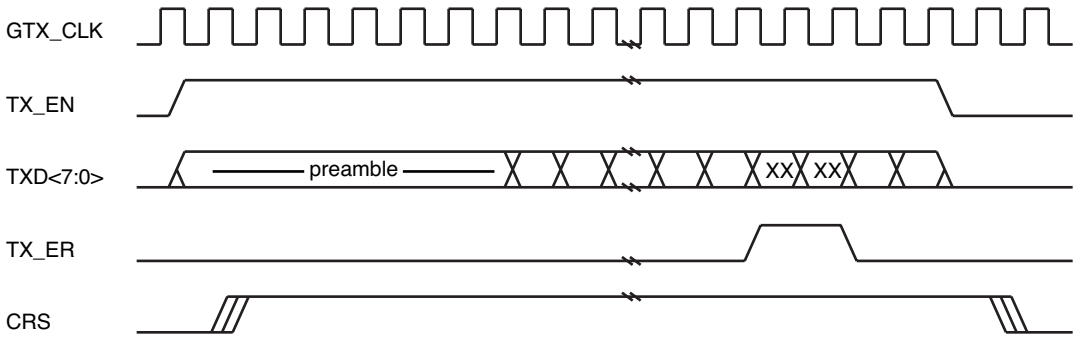


Figure 35–4—Propagating an error within a frame

Assertion of appropriate TXD values when TX_EN is de-asserted and TX_ER is asserted will cause the PHY to generate either Carrier Extend or Carrier Extend Error code-groups. The transition from TX_EN asserted and TX_ER de-asserted to TX_EN de-asserted and TX_ER asserted with TXD specifying Carrier Extend shall result in the PHY transmitting an end-of-packet delimiter as the initial code-groups of the carrier extension. Figures 35–5 and 35–6 show the behavior of TX_ER during the transmission of carrier extension. The propagation of an error in carrier extension is requested by holding TX_EN de-asserted and TX_ER asserted along with the appropriate value of TXD<7:0>.

Burst transmission of frames also uses carrier extension between frames of the burst. Figure 35–7 shows the behavior of TX_ER and TX_EN during burst transmission.

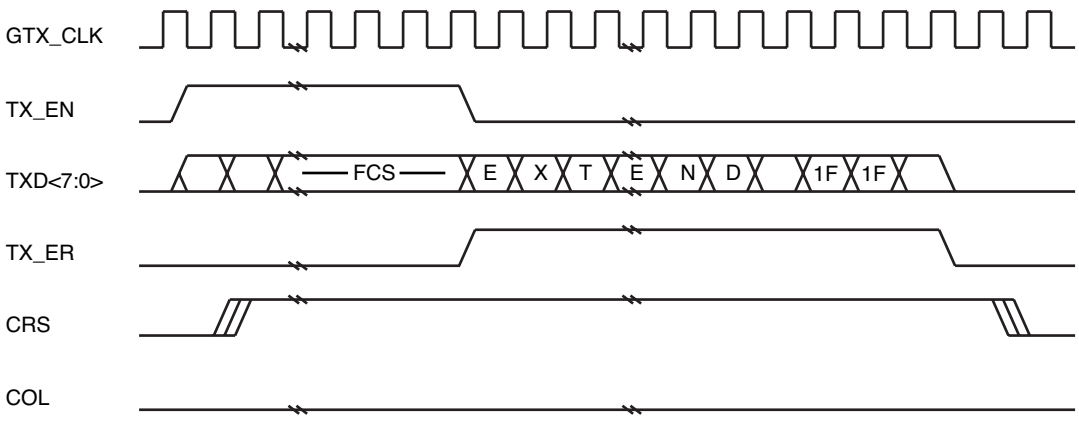


Figure 35–5—Propagating an error within carrier extension

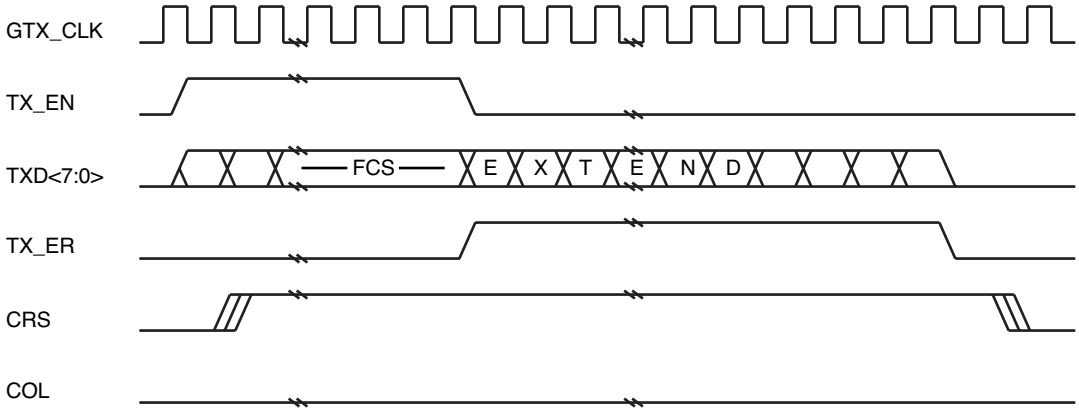


Figure 35–6—Transmission with carrier extension

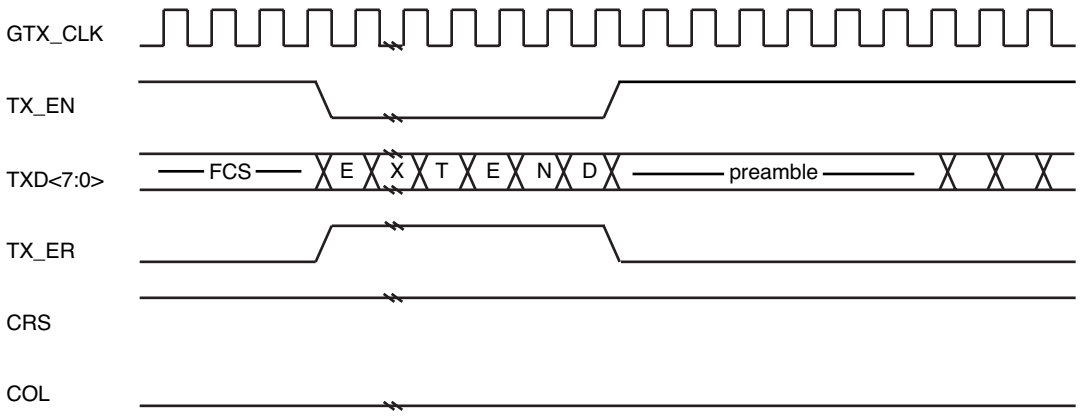


Figure 35–7—Burst transmission

35.2.2.6 RX_DV (receive data valid)

RX_DV is driven by the PHY to indicate that the PHY is presenting recovered and decoded data on the RXD<7:0> bundle. RX_DV shall transition synchronously with respect to the RX_CLK. RX_DV shall be asserted continuously from the first recovered octet of the frame through the final recovered octet and shall be negated prior to the first rising edge of RX_CLK that follows the final octet. In order for a received frame to be correctly interpreted by the Reconciliation sublayer and the MAC sublayer, RX_DV must encompass the frame, starting no later than the Start Frame Delimiter (SFD) and excluding any End-of-Frame delimiter.

Figure 35–8 shows the behavior of RX_DV during frame reception with no errors or carrier extension.

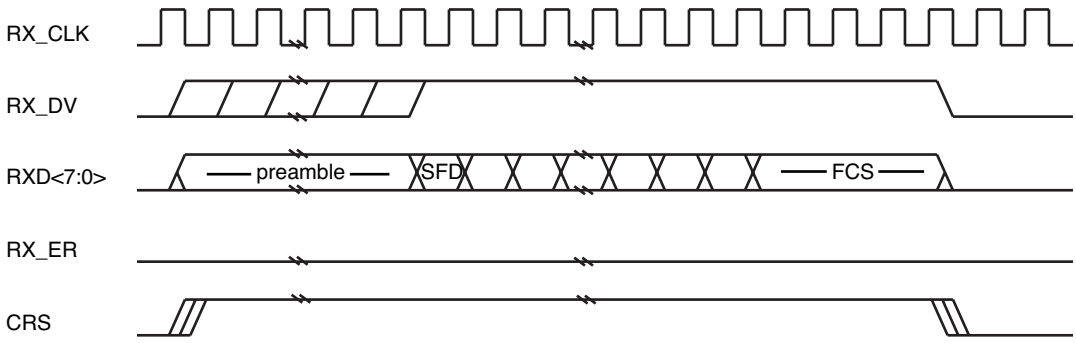


Figure 35–8—Basic frame reception

35.2.2.7 RXD (receive data)

RXD is a bundle of eight data signals (RXD<7:0>) that are driven by the PHY. RXD<7:0> shall transition synchronously with respect to RX_CLK. For each RX_CLK period in which RX_DV is asserted, RXD<7:0> transfer eight bits of recovered data from the PHY to the Reconciliation sublayer. RXD<0> is the least significant bit. Figure 35–8 shows the behavior of RXD<7:0> during frame reception.

While RX_DV is de-asserted, the PHY may provide a False Carrier indication by asserting the RX_ER signal while driving the specific value listed in Table 35–2 onto RXD<7:0>. See 36.2.5.2.3 for a description of the conditions under which a PHY will provide a False Carrier indication.

In order for a frame to be correctly interpreted by the MAC sublayer, a completely formed SFD must be passed across the GMII.

In a DTE operating in half duplex mode, a PHY is not required to loop data transmitted on TXD<7:0> back to RXD<7:0> unless the loopback mode of operation is selected as defined in 22.2.4.1.2. In a DTE operating in full duplex mode, data transmitted on TXD <7:0> shall not be looped back to RXD <7:0> unless the loopback mode of operation is selected.

While RX_DV is de-asserted and RX_ER is asserted, a specific RXD<7:0> value is used to transfer recovered Carrier Extend from the PHY to the Reconciliation sublayer. A Carrier Extend Error is indicated by another specific value of RXD<7:0>. Figure 35–9 shows the behavior of RX_DV during frame reception with carrier extension. Carrier extension shall only be signalled immediately following frame reception.

Burst transmission of frames also uses carrier extension between frames of the burst. Figure 35–10 shows the behavior of RX_ER and RX_DV during burst reception.

Table 35–2 specifies the permissible encoding of RXD<7:0>, RX_ER, and RX_DV, along with the specific indication provided by each code.

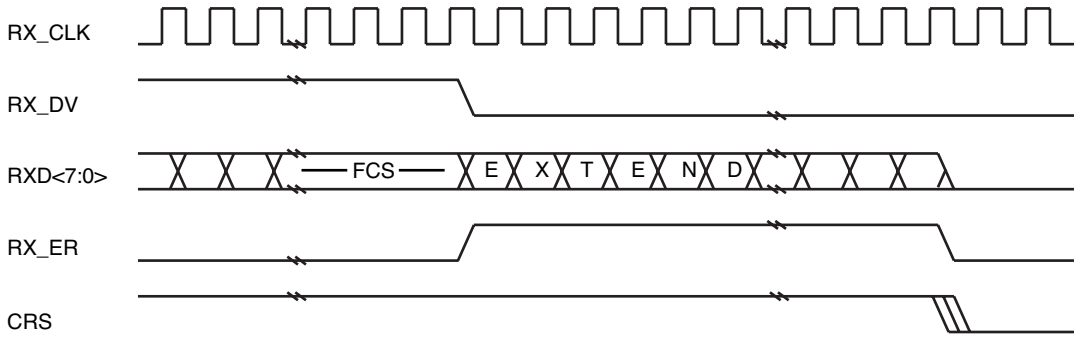


Figure 35–9—Frame reception with carrier extension

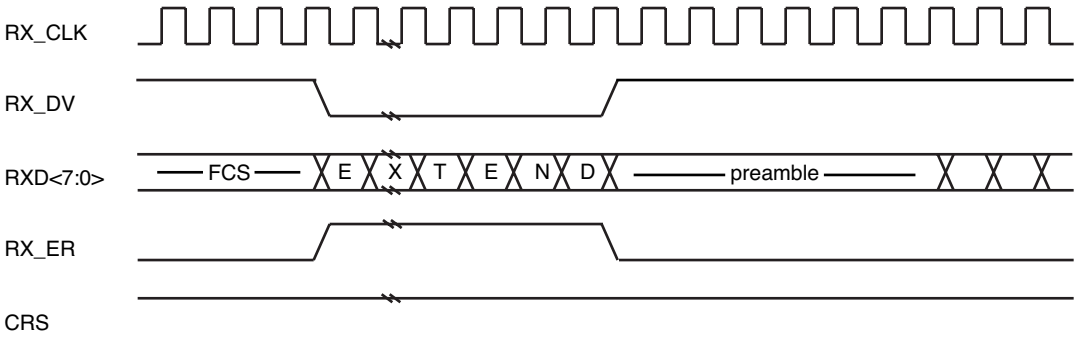


Figure 35–10—Burst reception

Table 35–2—Permissible encoding of RXD<7:0>, RX_ER, and RX_DV

RX_DV	RX_ER	RXD<7:0>	Description	PLS_DATA.indicate parameter
0	0	00 through FF	Normal inter-frame	No applicable parameter
0	1	00	Normal inter-frame	No applicable parameter
0	1	01 through 0D	Reserved	—
0	1	0E	False Carrier indication	No applicable parameter
0	1	0F	Carrier Extend	EXTEND (eight bits)
0	1	10 through 1E	Reserved	—
0	1	1F	Carrier Extend Error	ZERO, ONE (eight bits)
0	1	20 through FF	Reserved	—
1	0	00 through FF	Normal data reception	ZERO, ONE (eight bits)
1	1	00 through FF	Data reception error	ZERO, ONE (eight bits)
NOTE—Values in RXD<7:0> column are in hexadecimal.				

35.2.2.8 RX_ER (receive error)

RX_ER is driven by the PHY and shall transition synchronously with respect to RX_CLK. When RX_DV is asserted, RX_ER shall be asserted for one or more RX_CLK periods to indicate to the Reconciliation sublayer that an error (e.g., a coding error, or another error that the PHY is capable of detecting that may otherwise be undetectable at the MAC sublayer) was detected somewhere in the frame presently being transferred from the PHY to the Reconciliation sublayer.

The effect of RX_ER on the Reconciliation sublayer is defined in 35.2.1.5. Figure 35–11 shows the behavior of RX_ER during the reception of a frame with errors. Two independent error cases are illustrated. When RX_DV is asserted, assertion of RX_ER indicates an error within the data octets of a frame. An error within carrier extension is indicated by driving the appropriate value on RXD<7:0> while keeping RX_ER asserted.

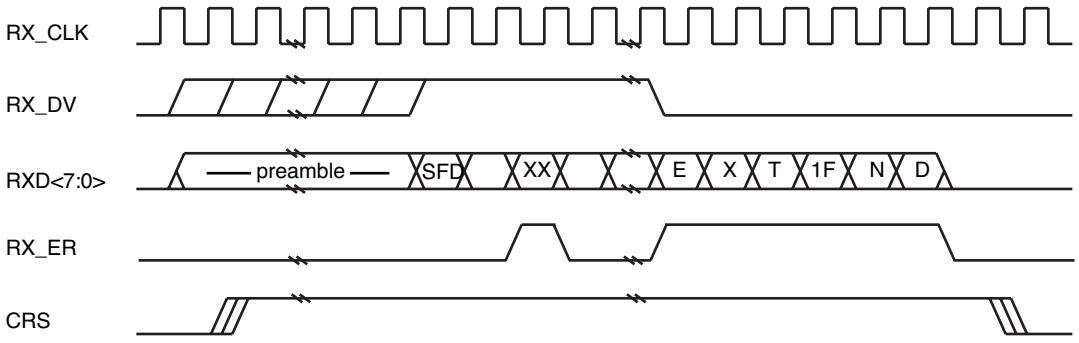


Figure 35–11—Two examples of reception with error

Assertion of RX_ER when RX_DV is de-asserted with specific RXD values indicates the decode of carrier extension by the PHY. The transition from RX_DV asserted and RX_ER de-asserted to RX_DV de-asserted and RX_ER asserted with RXD specifying Carrier Extend shall result in the Reconciliation sublayer indicating EXTEND INPUT_UNITs to the MAC. Figure 35–9 shows the behavior of RX_DV and RX_ER during frame reception with carrier extension.

While RX_DV is de-asserted, the PHY may provide a False Carrier indication by asserting the RX_ER signal for at least one cycle of the RX_CLK while driving the appropriate value onto RXD<7:0>, as defined in Table 35–2. See 36.2.5.2.3 for a description of the conditions under which a PHY will provide a False Carrier indication. Figure 35–12 shows the behavior of RX_ER, RX_DV and RXD<7:0> during a False Carrier indication.

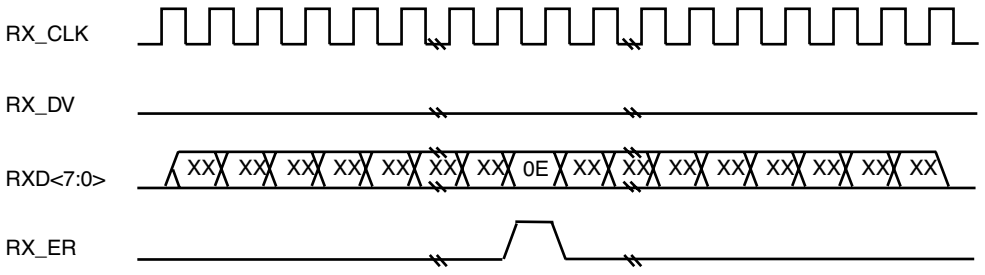


Figure 35–12—False Carrier indication

35.2.2.9 CRS (carrier sense)

CRS is driven by the PHY. Except when used in a repeater, a PHY in half duplex mode shall assert CRS when either the transmit or receive medium is non-idle and shall de-assert CRS when both the transmit and receive media are idle. The PHY shall ensure that CRS remains asserted throughout the duration of a collision condition.

When used in a repeater, a PHY shall assert CRS when the receive medium is non-idle and shall de-assert CRS when the receive medium is idle.

CRS is not required to transition synchronously with respect to either the GTX_CLK or the RX_CLK.

The behavior of CRS is unspecified when the PHY is in full duplex mode.

Figure 35–3 and Figure 35–5 show the behavior of CRS during a frame transmission without a collision, while Figure 35–13 and Figure 35–14 show the behavior of CRS during a frame transmission with a collision.

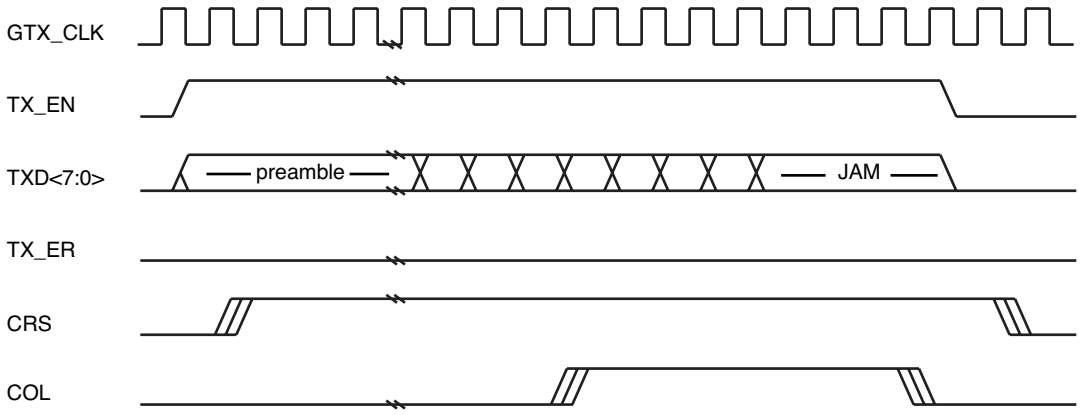


Figure 35–13—Transmission with collision

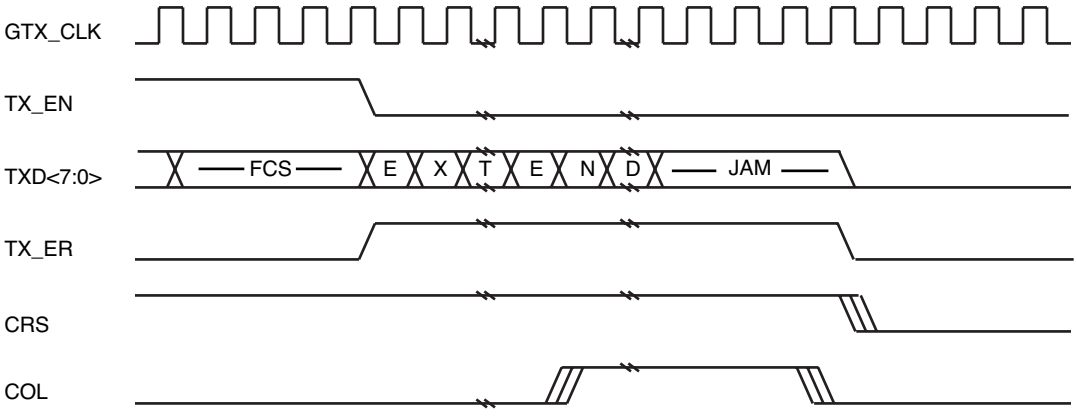


Figure 35–14—Transmission with collision in carrier extension

35.2.2.10 COL (collision detected)

COL is driven by the PHY and shall be asserted upon detection of a collision on the medium, and shall remain asserted while the collision condition persists.

COL is not required to transition synchronously with respect to either the GTX_CLK or the RX_CLK.

The behavior of the COL signal is unspecified when the PHY is in full duplex mode.

Figure 35–13 and Figure 35–14 show the behavior of COL during a frame transmission with a collision.

35.2.2.11 MDC (management data clock)

MDC is specified in 22.2.2.11.

35.2.2.12 MDIO (management data input/output)

MDIO is specified in 22.2.2.12.

35.2.3 GMII data stream

Data frames transmitted through the GMII shall be transferred within the data stream shown in Figure 35–15.

<inter-frame><preamble><sfd><data><efd><extend>

Figure 35–15—GMII data stream

For the GMII, transmission and reception of each octet of data shall be as shown in Figure 35–16.

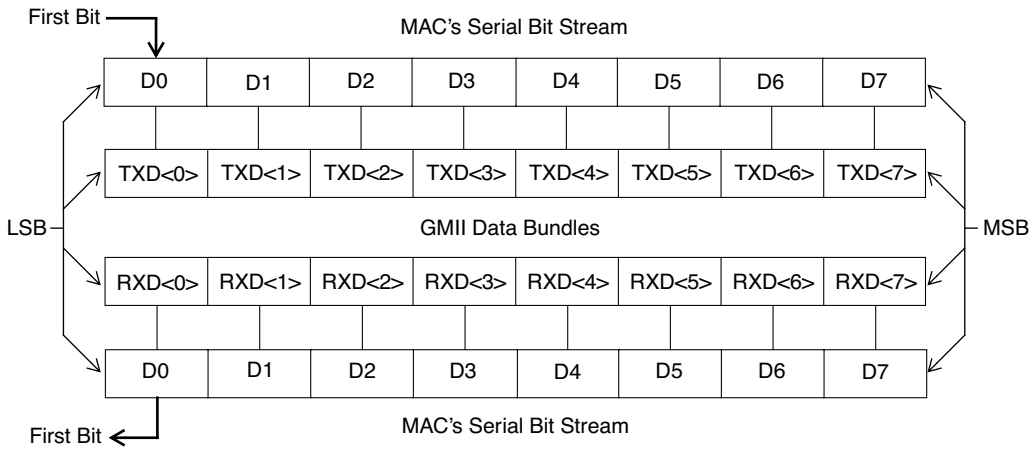


Figure 35–16—Relationship of data bundles to MAC serial bit stream

35.2.3.1 Inter-frame <inter-frame>

The inter-frame <inter-frame> period on a GMII transmit or receive path is an interval during which no data activity occurs on the path. Between bursts or single frame transmissions, the absence of data activity on the receive path is indicated by the de-assertion of both RX_DV and RX_ER or the de-assertion of the RX_DV signal with an RXD<7:0> value of 00 hexadecimal. On the transmit path the absence of data activity is indicated by the de-assertion of both TX_EN and TX_ER.

Between frames within a burst, the inter-frame period is signalled as Carrier Extend on the GMII. As shown in Figure 35–7, this is done by asserting TX_ER with the appropriate encoding of TXD<7:0> simultaneous with the de-assertion of TX_EN on the transmit path; and, as shown in Figure 35–10, by asserting RX_ER with the appropriate encoding of RXD<7:0> simultaneous with the de-assertion of RX_DV on the receive path.

Within a burst, the MAC interFrameSpacing parameter defined in Clause 4, is measured from the de-assertion of the TX_EN signal to the assertion of the TX_EN signal, and between bursts measured from the de-assertion of the CRS signal to the assertion of the CRS signal.

35.2.3.2 Preamble <preamble> and start of frame delimiter <sfd>

35.2.3.2.1 Transmit case

The preamble <preamble> begins a frame transmission. The bit value of the preamble field at the GMII is unchanged from that specified in 4.2.5 and when generated by a MAC shall consist of 7 octets with the following bit values:

10101010 10101010 10101010 10101010 10101010 10101010 10101010

The SFD (Start Frame Delimiter) <sfd> indicates the start of a frame and immediately follows the preamble. The bit value of the SFD at the GMII is unchanged from that specified in 4.2.6, and is the following bit sequence:

10101011

The preamble and SFD are shown above with their bits ordered for serial transmission from left to right. As shown, the leftmost bit of each octet is the LSB of the octet and the rightmost bit of each octet is the MSB of the octet.

The preamble and SFD shall be transmitted through the GMII as octets starting from the assertion of TX_EN.

35.2.3.2.2 Receive case

The conditions for assertion of RX_DV are defined in 35.2.2.6. The operation of 1000 Mb/s PHYs can result in shrinkage of the preamble between transmission at the source GMII and reception at the destination GMII. Table 35–3 depicts the case where no preamble bytes are conveyed across the GMII. This case may not be possible with a specific PHY, but illustrates the minimum preamble with which MAC shall be able to operate. Table 35–4 depicts the case where the entire preamble is conveyed across the GMII.

Table 35–3—Start of receive with no preamble preceding SFD

Signal	Bit values of octets received through GMII[a]			
RXD0	X	X	1[b]	D0[c]
RXD1	X	X	0	D1
RXD2	X	X	1	D2
RXD3	X	X	0	D3
RXD4	X	X	1	D4
RXD5	X	X	0	D5
RXD6	X	X	1	D6
RXD7	X	X	1	D7
RX_DV	0	0	1	1

[a]Leftmost octet is the first received.
[b]Start Frame Delimiter octet.
[c]D0 through D7 is the first octet of the PDU (first octet of the Destination Address).

Table 35–4—Start of receive with entire preamble preceding SFD

Signal	Bit values of octets received through GMII[a]									
RXD0	X	1	1	1	1	1	1	1	1[b]	D0[c]
RXD1	X	0	0	0	0	0	0	0	0	D1
RXD2	X	1	1	1	1	1	1	1	1	D2
RXD3	X	0	0	0	0	0	0	0	0	D3
RXD4	X	1	1	1	1	1	1	1	1	D4
RXD5	X	0	0	0	0	0	0	0	0	D5
RXD6	X	1	1	1	1	1	1	1	1	D6
RXD7	X	0	0	0	0	0	0	0	1	D7
RX_DV	0	1	1	1	1	1	1	1	1	1

[a]Leftmost octet is the first received.
[b]Start Frame Delimiter octet.
[c]D0 through D7 is the first octet of the PDU (first octet of the Destination Address).

35.2.3.3 Data <data>

The data <data> in a well-formed frame shall consist of a set of data octets.

35.2.3.4 End-of-Frame delimiter <efd>

De-assertion of the TX_EN signal constitutes an End-of-Frame delimiter <efd> for data conveyed on TXD<7:0>, and de-assertion of RX_DV constitutes an End-of-Frame delimiter for data conveyed on RXD<7:0>.

35.2.3.5 Carrier extension <extend>

The Reconciliation sublayer signals carrier extension <extend> on the transmit path by the assertion of the TX_ER signal with the appropriate value of TXD<7:0> simultaneous with the de-assertion of the TX_EN signal. Carrier extension is signaled on the receive path by the assertion of the RX_ER signal with the appropriate encoding on RXD<7:0> simultaneous with the de-assertion of RX_DV. Carrier extension may not be present on all frames.

35.2.3.6 Definition of Start of Packet and End of Packet Delimiters

For the purposes of Clause 30 layer management, the Start of Packet delimiter is defined as the rising edge of RX_DV; and the End of Packet delimiter is defined as the falling edge of RX_DV. (See Clause 30.2.2.2.2.)

35.2.4 MAC delay constraints (with GMII)

A Gigabit Ethernet MAC with a GMII shall comply with the delay constraints in Table 35–5.

Table 35–5—MAC delay constraints (with GMII)

Sublayer measurement points	Event	Min (bits)	Max (bits)	Input timing reference	Output timing reference
MAC ⇔ GMII	MAC transmit start to TX_EN = 1 sampled		48		GTX_CLK rising
	CRS assert to MAC detect[a]	0	48		
	CRS de-assert to MAC detect[a]	0	48		
	CRS assert to TX_EN = 1 sampled (worst-case nondeferred transmit)		112		GTX_CLK rising
	COL assert to MAC detect	0	48		
	COL de-assert to MAC detect	0	48		
	COL assert to TXD = Jam sampled (worst-case collision response)		112		GTX_CLK rising; first octet of jam

[a]For any given implementation: Max de-assert – Min. assert ≤ 16 bits.

35.2.5 Management functions

The GMII shall use the MII management register set specified in 22.2.4. The detailed description of some management registers are dependent on the PHY type and are specified in either 28.2.4 or 37.2.5.

35.3 Signal mapping

The GMII is specified such that implementors may share pins for implementation of the GMII, the MII specified in Clause 22 and the TBI specified in Clause 36. A recommended mapping of the signals for the GMII, MII, and TBI is shown in Table 35–6. Implementors using this recommended mapping are to comply with the GMII electrical characteristics in 35.4, MII electrical characteristics in 22.3, and the TBI electrical characteristics in 36.3 as appropriate for the implemented interfaces.

In an implementation supporting the MII and GMII, some signal pins are not used in both interfaces. For example, the TXD and RXD data bundles are four bits wide for the MII and eight bits wide for the GMII. Also, the GTX_CLK is only used when operating as a GMII while TX_CLK is used when operating as an MII.

Similarly, an implementation supporting both the GMII and TBI interfaces will map TBI signals onto the GMII control signal pins of TX_ER, TX_EN, RX_ER, and RX_DV. The COL and CRS signals of the GMII have no corollary in the TBI.

It is recommended that unused signal pins be driven to a valid logic state.

Table 35–6—Signal mapping

GMII	MII	TBI
TX_ER	TX_ER	TX<9>
TX_EN	TX_EN	TX<8>
TXD<7>		TX<7>
TXD<6>		TX<6>
TXD<5>		TX<5>
TXD<4>		TX<4>
TXD<3>	TXD<3>	TX<3>
TXD<2>	TXD<2>	TX<2>
TXD<1>	TXD<1>	TX<1>
TXD<0>	TXD<0>	TX<0>
COL	COL	

GMII	MII	TBI
RX_ER	RX_ER	RX<9>
RX_DV	RX_DV	RX<8>
RXD<7>		RX<7>
RXD<6>		RX<6>
RXD<5>		RX<5>
RXD<4>		RX<4>
RXD<3>	RXD<3>	RX<3>
RXD<2>	RXD<2>	RX<2>
RXD<1>	RXD<1>	RX<1>
RXD<0>	RXD<0>	RX<0>
CRS	CRS	

35.4 Electrical characteristics

The electrical characteristics of the GMII are specified such that the GMII can be applied within a variety of 1000 Mb/s equipment types. The electrical specifications are optimized for an integrated circuit to integrated circuit application environment. This includes applications where a number of PHY integrated circuits may be connected to a single integrated circuit as may be found in a repeater. Though specified for use on a single circuit board, applications to a motherboard-to-daughterboard interconnection are not precluded.

The electrical characteristics specified in this clause apply to all GMII signals except MDIO and MDC. The electrical characteristics for MDIO and MDC are specified in 22.3.4.

35.4.1 DC characteristics

All GMII drivers and receivers shall comply with the dc parametric attributes specified in Table 35–7.

The potential applied to the input of a GMII receiver may exceed the potential of the receiver's power supply (i.e., a GMII driver powered from a 3.6 V supply driving V_{OH} into a GMII receiver powered from a 2.5 V supply). Tolerance for dissimilar GMII driver and receiver supply potentials is implicit in these specifications.

35.4.2 AC characteristics

The GMII ac electrical characteristics are specified in a manner that allows the implementor flexibility in selecting the GMII topologies its devices support and the techniques used to achieve the specified characteristics.

Table 35–7—DC specifications

Symbol	Parameter	Conditions		Min	Max	Units
V_{OH}	Output High Voltage	I_{OH} = -1.0 mA	V_{CC} = Min	2.10	3.60	V
V_{OL}	Output Low Voltage	I_{OL} = 1.0 mA	V_{CC} = Min	GND	0.50	V
V_{IH}	Input High Voltage			1.70	—	V
V_{IL}	Input Low Voltage			—	0.90	V
I_{IH}	Input High Current	V_{CC} = Max	V_{IN} = 2.1 V	—	40	μA
I_{IL}	Input Low Current	V_{CC} = Max	V_{IN} = 0.5 V	–600	—	μA

All GMII devices are required to support point-to-point links. The electrical length of the circuit board traces used to implement these links can be long enough to exhibit transmission line effects and require some form of termination. The implementor is allowed the flexibility to select the driver output characteristics and the termination technique and components to be used with its drivers for point-to-point links.

Implementors may elect to support other GMII topologies in addition to the point-to-point topology and may specify different termination techniques and components for each supported topology.

Since the output characteristics and output voltage waveforms of GMII drivers depend on the termination technique and the location of the termination components, the ac output characteristics of GMII drivers are not explicitly specified. Rather, the ac characteristics of the signal delivered to a GMII receiver are specified. These characteristics are independent of the topology and termination technique and apply uniformly to all GMII applications.

35.4.2.1 Signal Timing measurements

All GMII ac timing measurements are made at the GMII receiver input and are specified relative to the V_{IL_AC}(max) and V_{IH_AC}(min) thresholds.

The GTX_CLK and RX_CLK parameters t_{PERIOD}, t_{HIGH}, and t_{LOW} are defined in Figure 35–17. The GTX_CLK and RX_CLK parameters t_R and t_F and other transient performance specifications are defined in Figure 35–18. These parameters and the GTX_CLK and RX_CLK rising and falling slew rates are measured using the "GMII Point-to-Point Test Circuit" shown in Figure 35–20.

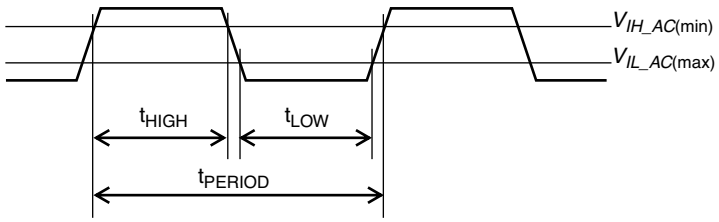
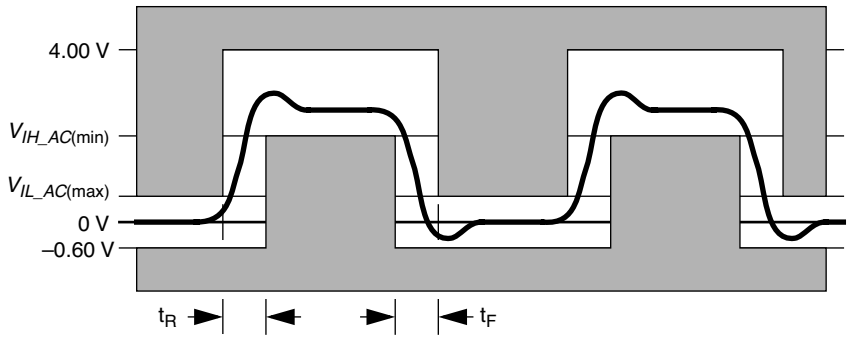


Figure 35–17—GTX_CLK and RX_CLK timing parameters at receiver input

The t_{SETUP} and t_{HOLD} parameters are defined in Figure 35–19. These parameters are measured using the "GMII Setup and Hold Time Test Circuit" shown in Figure 35–21.



NOTE—As measured at input measurement point

Figure 35–18—GMII receiver input potential template

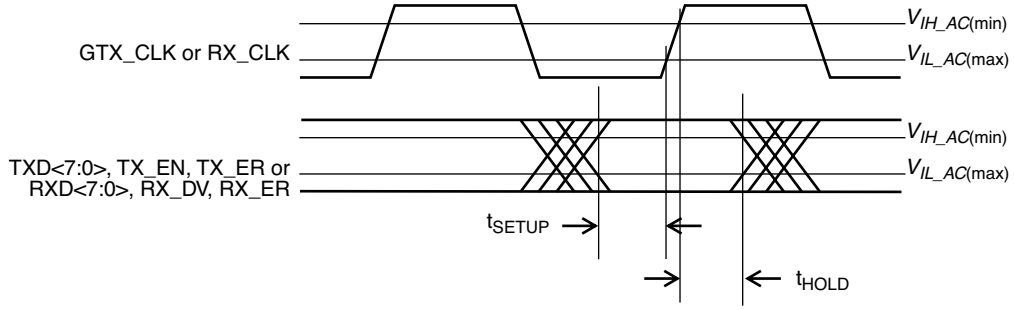


Figure 35–19—GMII signal timing at receiver input

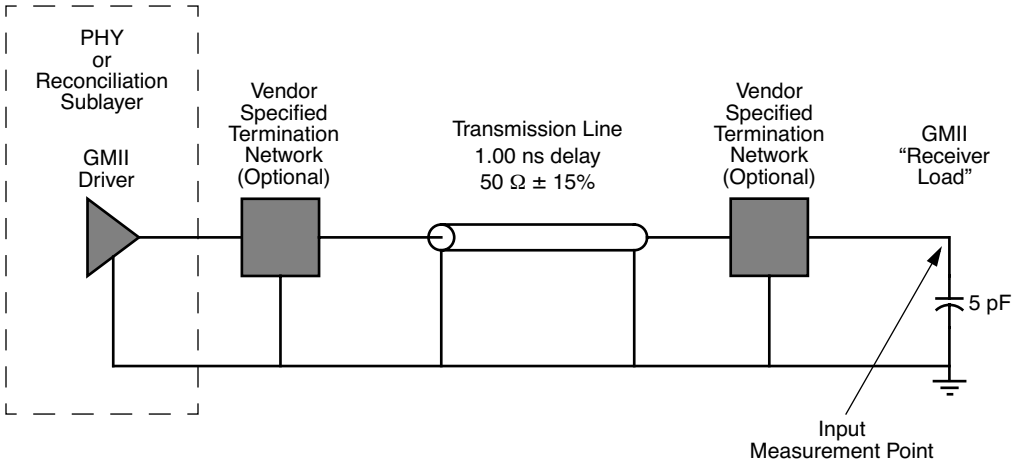


Figure 35–20—GMII point-to-point test circuit

35.4.2.2 GMII test circuit topology

The "GMII Point-to-Point Test Circuit" is defined in Figure 35–20. All parameter measurements made with this circuit are made at the "Input Measurement Point" defined in Figure 35–20. The 5 pF capacitor is included to approximate the input load of a GMII receiver. The termination networks used to implement the "GMII Point-to-Point Test Circuit" shall be those specified by the implementor of the GMII driver for 50 Ω ± 15%

impedance transmission line point-to-point links. One or both of the termination networks specified by the implementor of the GMII driver may be straight-through connections if the networks are not needed to comply with the GMII ac and transient performance specifications.

The "GMII Point-to-Point Test Circuit" specifies a 1 ns transmission line. In a GMII implementation, the circuit board traces between the PHY and Reconciliation sublayer are not restricted to a delay of 1 ns.

The "GMII Setup and Hold Time Test Circuit" is defined in Figure 35–21. The circuit is comprised of the source of the synchronous GMII signal under test and its clock (the Reconciliation Layer or the PHY) and two "GMII Point-to-Point Test Circuits." One of the test circuits includes the GMII driver for the signal under test, the other test circuit includes the GMII driver for the clock that provides timing for the signal under test. The signal under test is measured at the "Signal Measurement Point" relative to its clock, which is measured at the "Clock Measurement Point" as defined in Figure 35–21.

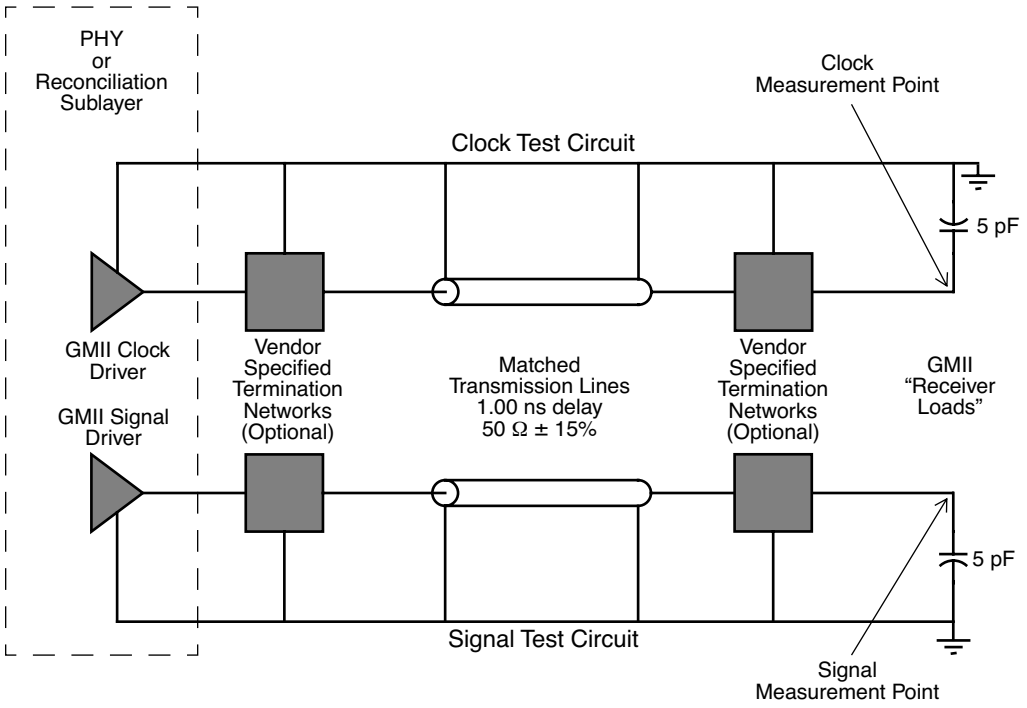


Figure 35–21—GMII setup and hold time test circuit

35.4.2.3 GMII ac specifications

A GMII driver, when used in combination with the termination networks specified by the implementor of the driver for a specific GMII topology, shall produce a potential at the input pin of any GMII receiver in that topology that complies with the input potential template shown in Figure 35–18. This requirement applies for all GMII signals and any GMII topology.

To ensure that all GMII devices support point-to-point links, a GMII driver, when driving the "GMII Point-to-Point Test Circuit" shown in Figure 35–20, shall produce a potential at the "Input Measurement Point" of the "GMII Point-to-Point Test Circuit" that complies with the input potential template shown in Figure 35–18.

All GMII signal sources, including the GMII drivers, GMII receivers and GMII signals shall comply with the ac specifications in Table 35–8.

Table 35–8—AC specifications

Symbol	Parameter	Conditions	Min	Max	Units
V_{IL_AC}	Input Low Voltage ac	–	–	0.70	V
V_{IH_AC}	Input High Voltage ac	–	1.90	–	V
f_{FREQ}	GTX_CLK Frequency	–	125 – 100 ppm	125 + 100 ppm	MHz
t_{PERIOD}	GTX_CLK Period	–	7.50	8.50	ns
t_{PERIOD}	RX_CLK Period	–	7.50	–	ns
t_{HIGH}	GTX_CLK, RX_CLK Time High	–	2.50	–	ns
t_{LOW}	GTX_CLK, RX_CLK Time Low	–	2.50	–	ns
t_R	GTX_CLK, RX_CLK Rise Time	V_{IL_AC}(max) to V_{IH_AC}(min)	–	1.00	ns
t_F	GTX_CLK, RX_CLK Fall Time	V_{IH_AC}(min) to V_{IL_AC}(max)	–	1.00	ns
–	Magnitude of GTX_CLK, RX_CLK Slew Rate (rising)[a]	V_{IL_AC}(max) to V_{IH_AC}(min)	0.6	–	V/ns
–	Magnitude of GTX_CLK, RX_CLK Slew Rate (falling)[a]	V_{IH_AC}(min) to V_{IL_AC}(max)	0.6	–	V/ns
t_{SETUP}	TXD, TX_EN, TX_ER Setup to ↑ GTX_CLK and RXD, RX_DV, RX_ER Setup to ↑ RX_CLK	–	2.50	–	ns
t_{HOLD}	TXD, TX_EN, TX_ER Hold from ↑ GTX_CLK and RXD, RX_DV, RX_ER Hold from ↑ RX_CLK	–	0.50	–	ns
t_{SETUP} (RCVR)	TXD, TX_EN, TX_ER Setup to ↑ GTX_CLK and RXD, RX_DV, RX_ER Setup to ↑ RX_CLK	–	2.00	–	ns
t_{HOLD} (RCVR)	TXD, TX_EN, TX_ER Hold from ↑ GTX_CLK and RXD, RX_DV, RX_ER Hold from ↑ RX_CLK	–	0.00	–	ns

[a]Clock Skew rate is the instantaneous rate of change of the clock potential with respect to time (dV/dt), not an average value over the entire rise or fall time interval. Conformance with this specification guarantees that the clock signals will rise and fall monotonically through the switching region.

Two sets of setup and hold time parameters are specified in Table 35–8. The first set, t_{SETUP} and t_{HOLD}, applies to the source of a synchronous GMII signal and its clock and is measured using the "GMII Setup and Hold Time Test Circuit," which has transmission lines with matched propagation delays in the "clock" and "signal" paths. The second set, $t_{SETUP(RCVR)}$ and $t_{HOLD(RCVR)}$, applies to the GMII receiver and specifies the minimum setup and hold times available to the GMII receiver at its input pins. The difference between the two sets of setup and hold time parameters provides margin for a small amount of mismatch in the propagation delays of the "clock" path and the "signal" paths in GMII applications.

The GMII ac specifications in Table 35–8 and the transient performance specifications in Figure 35–18 shall be met under all combination of worst-case GMII driver process and supply potential variation, ambient temperature, transmission line impedance variation, and termination network component impedance variation.

Designers of components containing GMII receivers should note that there is no upper bound specified on the magnitude of the slew rate of signals that may be applied to the input of a GMII receiver. The high-

frequency energy in a high slew rate (short rise time) signal can excite the parasitic reactances of the receiver package and input pad to such a degree that the signal at the receiver input pin and the signal at the input pad differ significantly. This is particularly true for GTX_CLK and RX_CLK, which transition at twice the rate of other signals in the interface.

35.5 Protocol Implementation Conformance Statement (PICS) proforma for Clause 35, Reconciliation Sublayer (RS) and Gigabit Media Independent Interface (GMII)[1]

35.5.1 Introduction

The supplier of a protocol implementation that is claimed to conform to Clause 35, Reconciliation Sublayer (RS) and Gigabit Media Independent Interface (GMII), shall complete the following Protocol Implementation Conformance Statement (PICS) proforma.

A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

35.5.2 Identification

35.5.2.1 Implementation identification

Supplier	
Contact point for enquiries about the PICS	
Implementation Name(s) and Version(s)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Names(s)	
NOTE 1—Only the first three items are required for all implementations; other information may be completed as appropriate in meeting the requirements for the identification. NOTE 2—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

[1]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so that it can be used for its intended purpose and may further publish the completed PICS.

35.5.2.2 Protocol summary

Identification of protocol standard	IEEE Std 802.3-2002®, Clause 35, Reconciliation Sublayer (RS) and Gigabit Media Independent Interface (GMII)
Identification of amendments and corrigenda to this PICS proforma that have been completed as part of this PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21; the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®.)	

Date of Statement	

35.5.2.3 Major capabilities/options

Item	Feature	Subclause	Value/Comment	Status	Support
*EL	GMII electrical interface	35.4		O	Yes [ ] No [ ]
*G1	PHY support of GMII	35.1.3		O	Yes [ ] No [ ]
G2	Reconciliation sublayer support of GMII	35.1.3		O	Yes [ ] No [ ]
*HD	Half duplex capability	35.2.2.6		O	Yes [ ] No [ ]

35.5.3 PICS proforma tables for reconciliation sublayer and Gigabit Media Independent Interface

35.5.3.1 Mapping of PLS service primitives

Item	Feature	Subclause	Value/Comment	Status	Support
PL1	Response to error in frame	35.2.1.5	Must produce FrameCheckError when RX_DV and RX_ER are asserted	M	Yes [ ]
PL2	Response to error in extension	35.2.1.5	Must produce FrameCheckError on received Carrier Extend Error	M	Yes [ ]
PL2a	Propagation of errors in frame	35.2.1.6	Assert TX_ER while TX_EN asserted	O	Yes [ ]
PL3	Propagation of errors in extension	35.2.1.6	Must send ONE or ZERO and assert Carrier Extend Error to propagate error within carrier extension	O	Yes [ ] No [ ]

35.5.3.2 GMII signal functional specifications

Item	Feature	Subclause	Value/Comment	Status	Support
SF1					
SF2	RX_CLK frequency	35.2.2.2	One eighth of received data rate or nominal 125 MHz.	M	Yes [ ]
SF3	RX_CLK source on loss of signal	35.2.2.2	Nominal clock reference (e.g., GTX_CLK)	M	Yes [ ]
SF4	RX_CLK transitions between recovered and nominal clock	35.2.2.2	While RX_DV de-asserted	M	Yes [ ]
SF5	RX_CLK max high/low time following de-assertion of RX_DV	35.2.2.2	Maximum 2 times the nominal period	M	Yes [ ]
SF6	TX_EN assertion	35.2.2.3	On first octet of preamble	M	Yes [ ]
SF7	TX_EN remains asserted	35.2.2.3	Stay asserted while all octets are transmitted over GMII	M	Yes [ ]
SF8	TX_EN negation	35.2.2.3	Before first GTX_CLK after final octet of frame	M	Yes [ ]
SF9	TX_EN transitions	35.2.2.3	Synchronous with GTX_CLK	M	Yes [ ]
SF10	TXD <7:0> transitions	35.2.2.4	Synchronous with GTX_CLK	M	Yes [ ]
SF11	TXD <7:0> effect on PHY while TX_EN and TX_ER are de-asserted	35.2.2.4	No effect	M	Yes [ ]
SF12	Signalling carrier extension	35.2.2.4	Only immediately following frame	M	Yes [ ]
SF13	TX_ER transitions	35.2.2.5	Synchronous with GTX_CLK	M	Yes [ ]
SF14	TX_ER effect on PHY while TX_EN is asserted	35.2.2.5	Cause PHY to emit invalid code-group	M	Yes [ ]
SF15	Transmission of end-of-packet delimiter	35.2.2.5	On de-assertion of TX_EN and simultaneous assertion of TX_ER	M	Yes [ ]
SF16	TX_ER implementation	35.2.2.5	At GMII of PHY	M	Yes [ ]
SF17	TX_ER implementation	35.2.2.5	Implemented if half duplex operation supported.	HD:M	Yes [ ] N/A [ ]
SF18	TX_ER driven	35.2.2.5	To valid state even if constant	M	Yes [ ]
SF19	RX_DV transitions	35.2.2.6	Synchronous with RX_CLK	M	Yes [ ]
SF20	RX_DV assertion	35.2.2.6	From first recovered octet to final octet of a frame	M	Yes [ ]
SF21	RX_DV negation	35.2.2.6	Before the first RX_CLK following the final octet of the frame	M	Yes [ ]
SF22	RXD <7:0> transitions	35.2.2.7	Synchronous with RX_CLK	M	Yes [ ]
SF22a	RXD loopback	35.2.2.7	No loopback unless loopback mode selected	M	Yes [ ]

35.5.3.2 GMII signal functional specifications *(continued)*

Item	Feature	Subclause	Value/Comment	Status	Support
SF23	Signalling carrier extension	35.2.2.7	Only immediately following frame	M	Yes [ ]
SF24	RX_ER transitions	35.2.2.8	Synchronous with RX_CLK	M	Yes [ ]
SF25	RX_ER assertion	35.2.2.8	By PHY to indicate error	M	Yes [ ]
SF26	Generation of EXTEND	35.2.2.8	In response to simultaneous de-assertion of RX_DV and assertion of RX_ER by PHY	M	Yes [ ]
SF27	CRS assertion	35.2.2.9	By PHY when either transmit or receive is NON-IDLE	M	Yes [ ]
SF28	CRS de-assertion	35.2.2.9	By PHY when both transmit and receive are IDLE	M	Yes [ ]
SF29	CRS assertion during collision	35.2.2.9	Remain asserted throughout	M	Yes [ ]
SF30	CRS assertion—repeater	35.2.2.9	By repeater when receive is NON-IDLE	M	Yes [ ]
SF31	CRS de-assertion—repeater	35.2.2.9	By repeater when medium is IDLE	M	Yes [ ]
SF32	COL assertion	35.2.2.10	By PHY upon collision on medium	M	Yes [ ]
SF33	COL remains asserted while collision persists	35.2.2.10		M	Yes [ ]

35.5.3.3 Data stream structure

Item	Feature	Subclause	Value/Comment	Status	Support
DS1	Format of transmitted data stream	35.2.3	Per Figure 35–15	M	Yes [ ]
DS2	Transmission order	35.2.3	Per Figure 35–16	M	Yes [ ]
DS3	Preamble 7 octets long	35.2.3.2	10101010 10101010 10101010 10101010 10101010 10101010 10101010	M	Yes [ ]
DS4	Preamble and SFD transmission	35.2.3.2	Starting at assertion of TX_EN	M	Yes [ ]
DS5	Minimum preamble	35.2.3.2	MAC operates with minimum preamble	M	Yes [ ]
DS6	Data length	35.2.3.3	Set of octets	M	Yes [ ]

35.5.3.4 Delay constraints

Item	Feature	Subclause	Value/Comment	Status	Support
DC1	MAC delay	35.2.4	Comply with Table 35–5	M	Yes [ ]

35.5.3.5 Management functions

Item	Feature	Subclause	Value/Comment	Status	Support
MF1	Management registers	35.2.5	GMII base registers as defined in 22.4	M	Yes [ ]

35.5.3.6 Electrical characteristics

Item	Feature	Subclause	Value/Comment	Status	Support
EC1	DC specifications	35.4.1	All drivers and receivers per Table 35–7	EL:M	Yes [ ] N/A [ ]
EC3	AC and transient specifications	35.4.2.3	Under all combinations of worst case parameters	EL:M	Yes [ ] N/A [ ]
EC4	Topology input potential	35.4.2.3	Complies with Figure 35–18 at each receiver of topology	EL:M	Yes [ ] N/A [ ]
EC5	Tested driver input potential	35.4.2.3	Complies with Figure 35–18 as tested per Figure 35–20	EL:M	Yes [ ] N/A [ ]
EC6	Test circuit termination	35.4.2.2	As specified by GMII driver implementor	EL:M	Yes [ ] N/A [ ]
EC7	AC specifications	35.4.2.3	Per Table 35–8	EL:M	Yes [ ] N/A [ ]

36. Physical Coding Sublayer (PCS) and Physical Medium Attachment (PMA) sublayer, type 1000BASE-X

36.1 Overview

36.1.1 Scope

This clause specifies the Physical Coding Sublayer (PCS) and the Physical Medium Attachment (PMA) sublayer that are common to a family of 1000 Mb/s Physical Layer implementations, collectively known as 1000BASE-X. There are currently three embodiments within this family: 1000BASE-CX, 1000BASE-LX, and 1000BASE-SX. The 1000BASE-CX embodiment specifies operation over a single copper media: two pairs of 150 Ω balanced copper cabling. 1000BASE-LX specifies operation over a pair of optical fibers using long-wavelength optical transmission. 1000BASE-SX specifies operation over a pair of optical fibers using short-wavelength optical transmission. The term 1000BASE-X is used when referring to issues common to any of the subvariants.

1000BASE-X is based on the Physical Layer standards developed by ANSI X3.230-1994 (Fibre Channel Physical and Signaling Interface). In particular, this standard uses the same 8B/10B coding as Fibre Channel, a PMA sublayer compatible with speed-enhanced versions of the ANSI 10-bit serializer chip, and similar optical and electrical specifications.

1000BASE-X PCS and PMA sublayers map the interface characteristics of the PMD sublayer (including MDI) to the services expected by the Reconciliation sublayer. 1000BASE-X can be extended to support any other full duplex medium requiring only that the medium be compliant at the PMD level.

36.1.2 Objectives

The following are the objectives of 1000BASE-X:

a) To support the CSMA/CD MAC;
b) To support the 1000 Mb/s repeater;
c) To provide for Auto-Negotiation among like 1000 Mb/s PMDs;
d) To provide 1000 Mb/s data rate at the GMII;
e) To support cable plants using 150 Ω balanced copper cabling, or optical fiber compliant with ISO/IEC 11801: 1995;
f) To allow for a nominal network extent of up to 3 km, including
    1) 150 Ω balanced links of 25 m span;
    2) one-repeater networks of 50 m span (using all 150 Ω balanced copper cabling);
    3) one-repeater networks of 200 m span (using fiber); and
    4) DTE/DTE links of 3000 m (using fiber);
g) To preserve full duplex behavior of underlying PMD channels;
h) To support a BER objective of 10^{-12}.

36.1.3 Relationship of 1000BASE-X to other standards

Figure 36–1 depicts the relationships among the 1000BASE-X sublayers (shown shaded), the CSMA/CD MAC and reconciliation layers, and the ISO/IEC 8802-2 LLC.

36.1.4 Summary of 1000BASE-X sublayers

The following provides an overview of the 1000BASE-X sublayers.[2]

36.1.4.1 Physical Coding Sublayer (PCS)

The PCS interface is the Gigabit Media Independent Interface (GMII) that provides a uniform interface to the Reconciliation sublayer for all 1000 Mb/s PHY implementations (e.g., not only 1000BASE-X but also other possible types of gigabit PHY entities). 1000BASE-X provides services to the GMII in a manner analogous to how 100BASE-X provides services to the 100 Mb/s MII.

The 1000BASE-X PCS provides all services required by the GMII, including

a) Encoding (decoding) of GMII data octets to (from) ten-bit code-groups (8B/10B) for communication with the underlying PMA;
b) Generating Carrier Sense and Collision Detect indications for use by PHY's half duplex clients;
c) Managing the Auto-Negotiation process, and informing the management entity via the GMII when the PHY is ready for use.

36.1.4.2 Physical Medium Attachment (PMA) sublayer

The PMA provides a medium-independent means for the PCS to support the use of a range of serial-bit-oriented physical media. The 1000BASE-X PMA performs the following functions:

a) Mapping of transmit and receive code-groups between the PCS and PMA via the PMA Service Interface;
b) Serialization (deserialization) of code-groups for transmission (reception) on the underlying serial PMD;
c) Recovery of clock from the 8B/10B-coded data supplied by the PMD;
d) Mapping of transmit and receive bits between the PMA and PMD via the PMD Service Interface;
e) Data loopback at the PMD Service Interface.

36.1.4.3 Physical Medium Dependent (PMD) sublayer

1000BASE-X physical layer signaling for fiber and copper media is adapted from ANSI X3.230-1994 (FC-PH), Clauses 6 and 7 respectively. These clauses define 1062.5 Mb/s, full duplex signaling systems that accommodate single-mode optical fiber, multimode optical fiber, and 150 Ω balanced copper cabling. 1000BASE-X adapts these basic physical layer specifications for use with the PMD sublayer and mediums specified in Clause 38 and Clause 39.

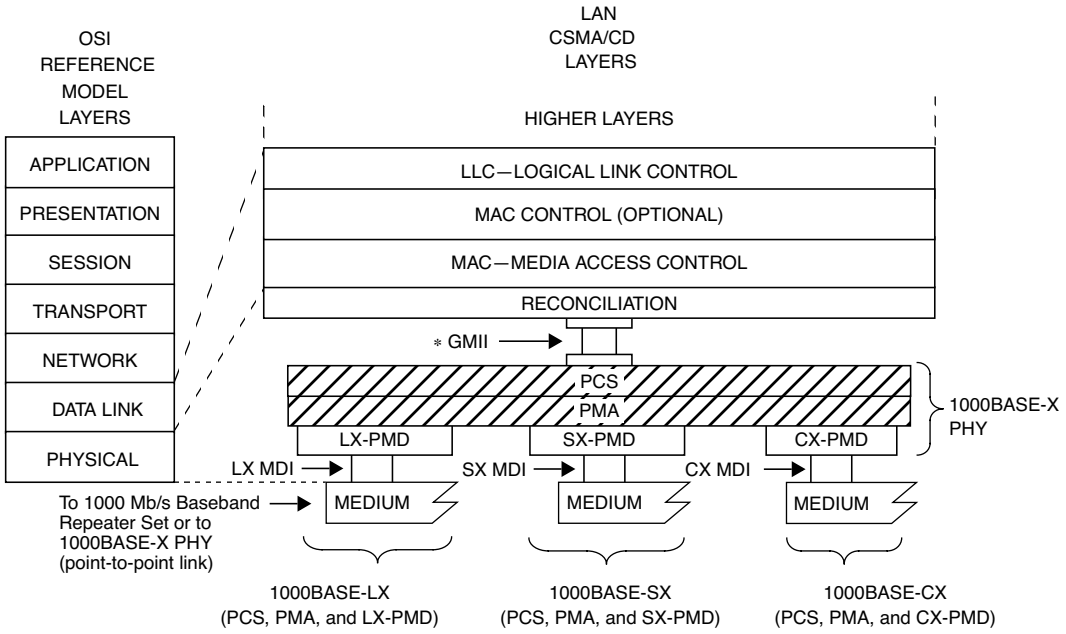
The MDI, logically subsumed within each PMD subclause, is the actual medium attachment, including connectors, for the various supported media.

Figure 36–1 depicts the relationship between 1000BASE-X and its associated PMD sublayers.

36.1.5 Inter-sublayer interfaces

There are a number of interfaces employed by 1000BASE-X. Some (such as the PMA Service Interface) use an abstract service model to define the operation of the interface. An optional physical instantiation of the PCS Interface has been defined. It is called the GMII (Gigabit Media Independent Interface). An optional physical instantiation of the PMA Service Interface has also been defined (see 36.3.3). It is adapted from

[2]The 1000BASE-X PHY consists of that portion of the Physical Layer between the MDI and GMII consisting of the PCS, PMA, and PMD sublayers. The 1000BASE-X PHY is roughly analogous to the 100BASE-X PHY.



MDI=MEDIUM DEPENDENT INTERFACE
GMII=GIGABIT MEDIA INDEPENDENT INTERFACE
PCS=PHYSICAL CODING SUBLAYER
PMA=PHYSICAL MEDIUM ATTACHMENT
PHY=PHYSICAL LAYER DEVICE
PMD=PHYSICAL MEDIUM DEPENDENT
LX-PMD=PMD FOR FIBER—LONG WAVELENGTH, Clause 38
SX-PMD=PMD FOR FIBER—SHORT WAVELENGTH, Clause 38
CX-PMD=PMD FOR 150 Ω BALANCED COPPER CABLING, Clause 39

NOTE—The PMD sublayers are mutually independent.

* GMII is optional.

Figure 36–1—Relationship of 1000BASE-X and the PMDs

ANSI Technical Report TR/X3.18-1997 (Fibre Channel—10-bit Interface). Figure 36–2 depicts the relationship and mapping of the services provided by all of the interfaces relevant to 1000BASE-X.

It is important to note that, while this specification defines interfaces in terms of bits, octets, and code-groups, implementors may choose other data path widths for implementation convenience. The only exceptions are a) the GMII, which, when implemented at an observable interconnection port, uses an octet-wide data path as specified in Clause 35, b) the PMA Service Interface, which, when physically implemented as the TBI (Ten-Bit Interface) at an observable interconnection port, uses a 10-bit wide data path as specified in 36.3.3, and c) the MDI, which uses a serial, physical interface.

36.1.6 Functional block diagram

Figure 36–2 provides a functional block diagram of the 1000BASE-X PHY.

36.1.7 State diagram conventions

The body of this standard is comprised of state diagrams, including the associated definitions of variables, constants, and functions. Should there be a discrepancy between a state diagram and descriptive text, the state diagram prevails.

The notation used in the state diagrams follows the conventions of 21.5. State diagram timers follow the conventions of 14.2.3.2.

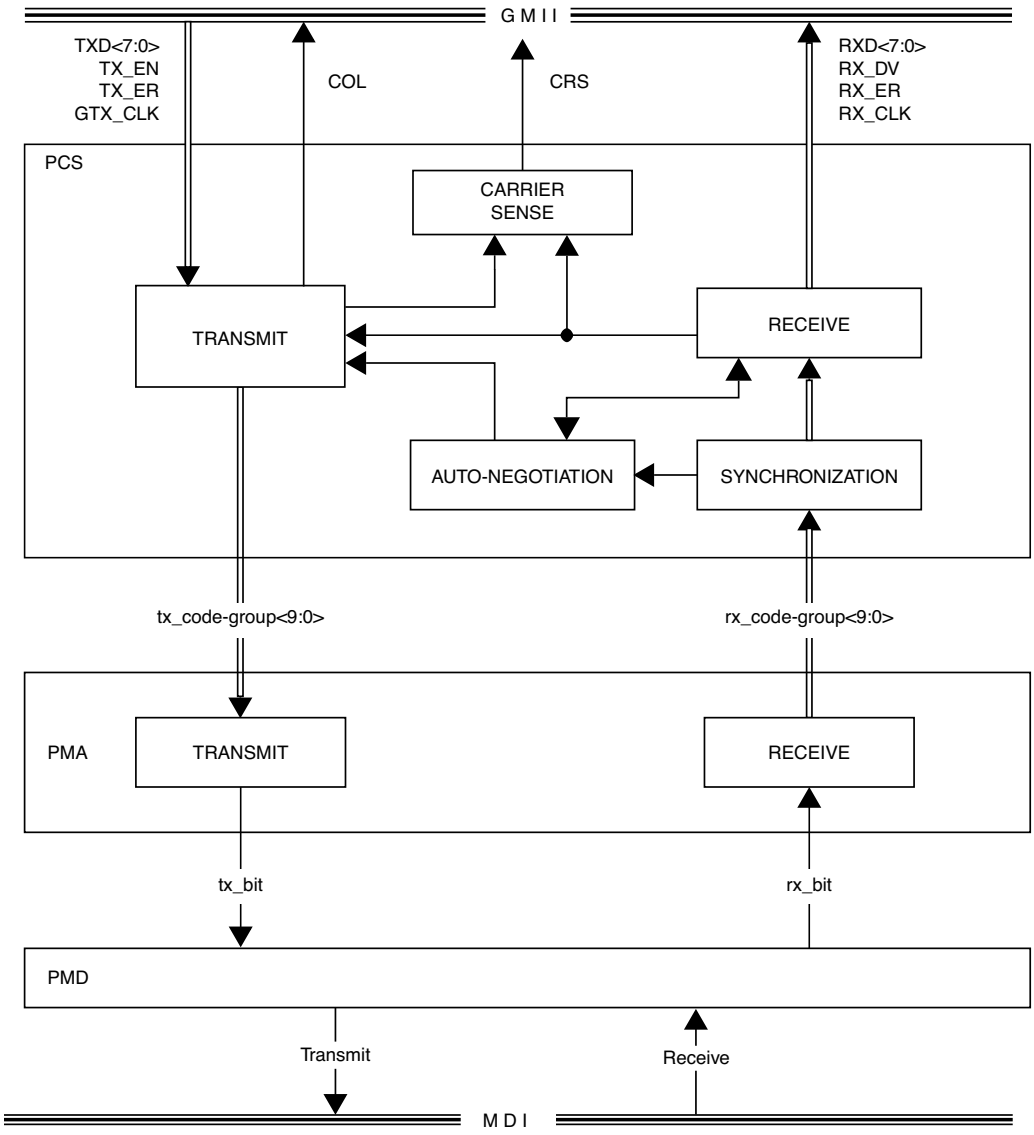


Figure 36–2—Functional block diagram

36.2 Physical Coding Sublayer (PCS)

36.2.1 PCS Interface (GMII)

The PCS Service Interface allows the 1000BASE-X PCS to transfer information to and from a PCS client. PCS clients include the MAC (via the Reconciliation sublayer) and repeater. The PCS Interface is precisely defined as the Gigabit Media Independent Interface (GMII) in Clause 35.

In this clause, the setting of GMII variables to TRUE or FALSE is equivalent, respectively, to “asserting” or “de-asserting” them as specified in Clause 35.

36.2.2 Functions within the PCS

The PCS comprises the PCS Transmit, Carrier Sense, Synchronization, PCS Receive, and Auto-Negotiation processes for 1000BASE-X. The PCS shields the Reconciliation sublayer (and MAC) from the specific nature of the underlying channel. When communicating with the GMII, the PCS uses an octet-wide, synchronous data path, with packet delimiting being provided by separate transmit control signals (TX_EN and TX_ER) and receive control signals (RX_DV and RX_ER). When communicating with the PMA, the PCS uses a ten-bit wide, synchronous data path, which conveys ten-bit code-groups. At the PMA Service Interface, code-group alignment and MAC packet delimiting are made possible by embedding special non-data code-groups in the transmitted code-group stream. The PCS provides the functions necessary to map packets between the GMII format and the PMA Service Interface format.

The PCS Transmit process continuously generates code-groups based upon the TXD <7:0>, TX_EN, and TX_ER signals on the GMII, sending them immediately to the PMA Service Interface via the PMA_UNITDATA.request primitive. The PCS Transmit process generates the GMII signal COL based on whether a reception is occurring simultaneously with transmission. Additionally, it generates the internal flag, transmitting, for use by the Carrier Sense process. The PCS Transmit process monitors the Auto-Negotiation process xmit flag to determine whether to transmit data or reconfigure the link.

The Carrier Sense process controls the GMII signal CRS (see Figure 36–8).

The PCS Synchronization process continuously accepts code-groups via the PMA_UNITDATA.indicate primitive and conveys received code-groups to the PCS Receive process via the SYNC_UNITDATA.indicate primitive. The PCS Synchronization process sets the sync_status flag to indicate whether the PMA is functioning dependably (as well as can be determined without exhaustive error-rate analysis).

The PCS Receive process continuously accepts code-groups via the SYNC_UNITDATA.indicate primitive. The PCS Receive process monitors these code-groups and generates RXD <7:0>, RX_DV, and RX_ER on the GMII, and the internal flag, receiving, used by the Carrier Sense and Transmit processes.

The PCS Auto-Negotiation process sets the xmit flag to inform the PCS Transmit process to either transmit normal idles interspersed with packets as requested by the GMII or to reconfigure the link. The PCS Auto-Negotiation process is specified in Clause 37.

36.2.3 Use of code-groups

The PCS maps GMII signals into ten-bit code groups, and vice versa, using an 8B/10B block coding scheme. Implicit in the definition of a code-group is an establishment of code-group boundaries by a PMA code-group alignment function as specified in 36.3.2.4. Code-groups are unobservable and have no meaning outside the PCS. The PCS functions ENCODE and DECODE generate, manipulate, and interpret code-groups as provided by the rules in 36.2.4.

36.2.4 8B/10B transmission code

The PCS uses a transmission code to improve the transmission characteristics of information to be transferred across the link. The encodings defined by the transmission code ensure that sufficient transitions are present in the PHY bit stream to make clock recovery possible at the receiver. Such encoding also greatly increases the likelihood of detecting any single or multiple bit errors that may occur during transmission and reception of information. In addition, some of the special code-groups of the transmission code contain a distinct and easily recognizable bit pattern that assists a receiver in achieving code-group alignment on the incoming PHY bit stream. The 8B/10B transmission code specified for use in this standard has a high transition density, is a run-length-limited code, and is dc-balanced. The transition density of the 8B/10B symbols ranges from 3 to 8 transitions per symbol.

The definition of the 8B/10B transmission code in this standard is identical to that specified in ANSI X3.230-1994 (FC-PH), Clause 11. The relationship of code-group bit positions to PMA and other PCS constructs is illustrated in Figure 36–3.

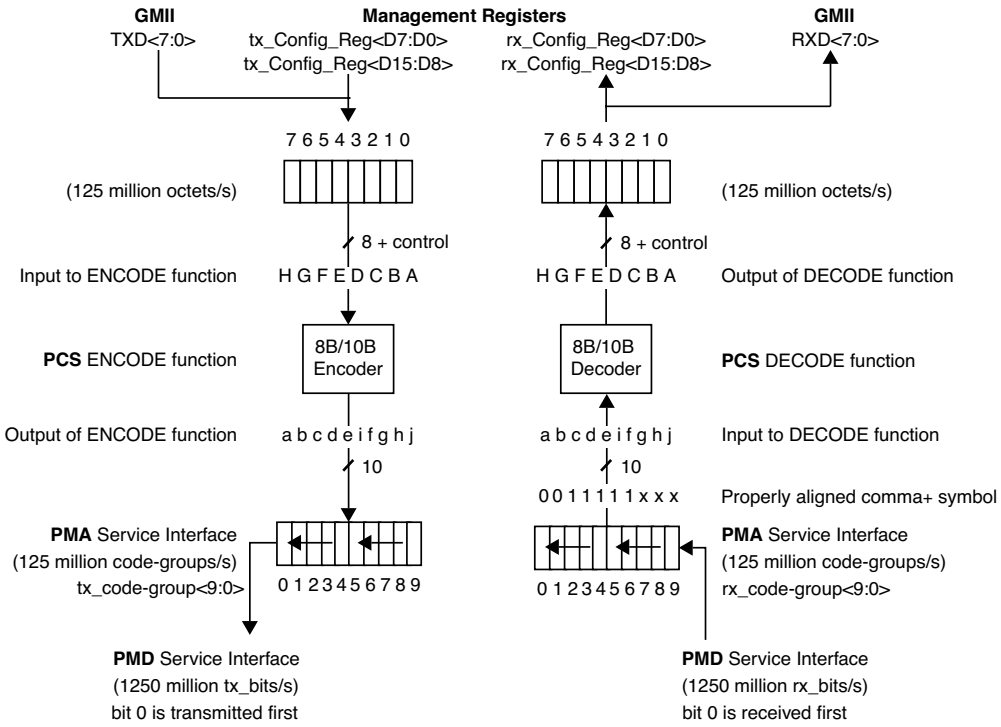


Figure 36–3—PCS reference diagram

36.2.4.1 Notation conventions

8B/10B transmission code uses letter notation for describing the bits of an unencoded information octet and a single control variable. Each bit of the unencoded information octet contains either a binary zero or a binary one. A control variable, Z, has either the value D or the value K. When the control variable associated with an unencoded information octet contains the value D, the associated encoded code-group is referred to as a data code-group. When the control variable associated with an unencoded information octet contains the value K, the associated encoded code-group is referred to as a special code-group.

The bit notation of A,B,C,D,E,F,G,H for an unencoded information octet is used in the description of the 8B/10B transmission code. The bits A,B,C,D,E,F,G,H are translated to bits a,b,c,d,e,i,f,g,h,j of 10-bit transmission code-groups. 8B/10B code-group bit assignments are illustrated in Figure 36–3. Each valid code-group has been given a name using the following convention: /Dx.y/ for the 256 valid data code-groups, and /Kx.y/ for special control code-groups, where x is the decimal value of bits EDCBA, and y is the decimal value of bits HGF.

36.2.4.2 Transmission order

Code-group bit transmission order is illustrated in Figure 36–3.

Code-groups within multi-code-group ordered_sets (as specified in Table 36–3) are transmitted sequentially beginning with the special code-group used to distinguish the ordered_set (e.g., /K28.5/) and proceeding code-group by code-group from left to right within the definition of the ordered_set until all code-groups of the ordered_set are transmitted.

The first code-group of every multi-code-group ordered_set is transmitted in an even-numbered code-group position counting from the first code-group after a reset or power-on. Subsequent code-groups continuously alternate as odd and even-numbered code-groups.

The contents of a packet are transmitted sequentially beginning with the ordered_set used to denote the Start_of_Packet (the SPD delimiter) and proceeding code-group by code-group from left to right within the definition of the packet until the ordered_set used to denote the End_of_Packet (the EPD delimiter) is transmitted.

36.2.4.3 Valid and invalid code-groups

Table 36–1a defines the valid data code-groups (D code-groups) of the 8B/10B transmission code. Table 36–2 defines the valid special code-groups (K code-groups) of the code. The tables are used for both generating valid code-groups (encoding) and checking the validity of received code-groups (decoding). In the tables, each octet entry has two columns that represent two (not necessarily different) code-groups. The two columns correspond to the valid code-group based on the current value of the running disparity (Current RD – or Current RD +). Running disparity is a binary parameter with either the value negative (–) or the value positive (+). Annex 36B provides several 8B/10B transmission code running disparity calculation examples.

36.2.4.4 Running disparity rules

After powering on or exiting a test mode, the transmitter shall assume the negative value for its initial running disparity. Upon transmission of any code-group, the transmitter shall calculate a new value for its running disparity based on the contents of the transmitted code-group.

After powering on or exiting a test mode, the receiver should assume either the positive or negative value for its initial running disparity. Upon the reception of any code-group, the receiver determines whether the code-group is valid or invalid and calculates a new value for its running disparity based on the contents of the received code-group.

The following rules for running disparity shall be used to calculate the new running disparity value for code-groups that have been transmitted (transmitter's running disparity) and that have been received (receiver's running disparity).

Running disparity for a code-group is calculated on the basis of sub-blocks, where the first six bits (abcdei) form one sub-block (six-bit sub-block) and the second four bits (fghj) form the other sub-block (four-bit sub-block). Running disparity at the beginning of the six-bit sub-block is the running disparity at the end of the last code-group. Running disparity at the beginning of the four-bit sub-block is the running disparity at the end of the six-bit sub-block. Running disparity at the end of the code-group is the running disparity at the end of the four-bit sub-block.

Running disparity for the sub-blocks is calculated as follows:

a) Running disparity at the end of any sub-block is positive if the sub-block contains more ones than zeros. It is also positive at the end of the six-bit sub-block if the six-bit sub-block is 000111, and it is positive at the end of the four-bit sub-block if the four-bit sub-block is 0011;
b) Running disparity at the end of any sub-block is negative if the sub-block contains more zeros than ones. It is also negative at the end of the six-bit sub-block if the six-bit sub-block is 111000, and it is negative at the end of the four-bit sub-block if the four-bit sub-block is 1100;
c) Otherwise, running disparity at the end of the sub-block is the same as at the beginning of the sub-block.

NOTE—All sub-blocks with equal numbers of zeros and ones are disparity neutral. In order to limit the run length of 0's or 1's between sub-blocks, the 8B/10B transmission code rules specify that sub-blocks encoded as 000111 or 0011 are generated only when the running disparity at the beginning of the sub-block is positive; thus, running disparity at the end of these sub-blocks is also positive. Likewise, sub-blocks containing 111000 or 1100 are generated only when the running disparity at the beginning of the sub-block is negative; thus, running disparity at the end of these sub-blocks is also negative.

36.2.4.5 Generating code-groups

The appropriate entry in either Table 36–1a or Table 36–2 is found for each octet for which a code-group is to be generated (encoded). The current value of the transmitter's running disparity shall be used to select the code-group from its corresponding column. For each code-group transmitted, a new value of the running disparity is calculated. This new value is used as the transmitter's current running disparity for the next octet to be encoded and transmitted.

36.2.4.6 Checking the validity of received code-groups

The following rules shall be used to determine the validity of received code groups:

a) The column in Tables 36–1a and 36–2 corresponding to the current value of the receiver's running disparity is searched for the received code-group;
b) If the received code-group is found in the proper column, according to the current running disparity, then the code-group is considered valid and, for data code-groups, the associated data octet determined (decoded);
c) If the received code-group is not found in that column, then the code-group is considered invalid;
d) Independent of the code-group's validity, the received code-group is used to calculate a new value of running disparity. The new value is used as the receiver's current running disparity for the next received code-group.

Detection of an invalid code-group does not necessarily indicate that the code-group in which the invalid code-group was detected is in error. Invalid code-groups may result from a prior error which altered the running disparity of the PHY bit stream but which did not result in a detectable error at the code-group in which the error occurred.

The number of invalid code-groups detected is proportional to the bit-error-rate (BER) of the link. Link error monitoring may be performed by counting invalid code-groups.

36.2.4.7 Ordered_sets

Eight ordered_sets, consisting of a single special code-group or combinations of special and data code-groups are specifically defined. Ordered_sets which include /K28.5/ provide the ability to obtain bit and code-group synchronization and establish ordered_set alignment (see 36.2.4.9 and 36.3.2.4). Ordered_sets provide for the delineation of a packet and synchronization between the transmitter and receiver circuits at opposite ends of a link. Table 36–3 lists the defined ordered_sets.

36.2.4.7.1 Ordered_set rules

Ordered_sets are specified according to the following rules:

a) Ordered_sets consist of either one, two, or four code-groups;
b) The first code-group of all ordered_sets is always a special code-group;
c) The second code-group of all multi-code-group ordered_sets is always a data code-group. The second code-group is used to distinguish the ordered set from all other ordered sets. The second code-group provides a high bit transition density.

Table 36–3 lists the defined ordered_sets.

Table 36–1a—Valid data code-groups

Code Group Name	Octet Value	Octet Bits HGF EDCBA	Current RD –	Current RD +
			abcdei fghj	abcdei fghj
D0.0	00	000 00000	100111 0100	011000 1011
D1.0	01	000 00001	011101 0100	100010 1011
D2.0	02	000 00010	101101 0100	010010 1011
D3.0	03	000 00011	110001 1011	110001 0100
D4.0	04	000 00100	110101 0100	001010 1011
D5.0	05	000 00101	101001 1011	101001 0100
D6.0	06	000 00110	011001 1011	011001 0100
D7.0	07	000 00111	111000 1011	000111 0100
D8.0	08	000 01000	111001 0100	000110 1011
D9.0	09	000 01001	100101 1011	100101 0100
D10.0	0A	000 01010	010101 1011	010101 0100
D11.0	0B	000 01011	110100 1011	110100 0100
D12.0	0C	000 01100	001101 1011	001101 0100
D13.0	0D	000 01101	101100 1011	101100 0100
D14.0	0E	000 01110	011100 1011	011100 0100
D15.0	0F	000 01111	010111 0100	101000 1011
D16.0	10	000 10000	011011 0100	100100 1011
D17.0	11	000 10001	100011 1011	100011 0100
D18.0	12	000 10010	010011 1011	010011 0100
D19.0	13	000 10011	110010 1011	110010 0100
D20.0	14	000 10100	001011 1011	001011 0100
D21.0	15	000 10101	101010 1011	101010 0100
D22.0	16	000 10110	011010 1011	011010 0100
D23.0	17	000 10111	111010 0100	000101 1011
D24.0	18	000 11000	110011 0100	001100 1011
D25.0	19	000 11001	100110 1011	100110 0100
D26.0	1A	000 11010	010110 1011	010110 0100
D27.0	1B	000 11011	110110 0100	001001 1011
D28.0	1C	000 11100	001110 1011	001110 0100
D29.0	1D	000 11101	101110 0100	010001 1011
D30.0	1E	000 11110	011110 0100	100001 1011
D31.0	1F	000 11111	101011 0100	010100 1011
D0.1	20	001 00000	100111 1001	011000 1001
D1.1	21	001 00001	011101 1001	100010 1001
D2.1	22	001 00010	101101 1001	010010 1001
D3.1	23	001 00011	110001 1001	110001 1001
D4.1	24	001 00100	110101 1001	001010 1001
D5.1	25	001 00101	101001 1001	101001 1001
D6.1	26	001 00110	011001 1001	011001 1001
D7.1	27	001 00111	111000 1001	000111 1001
D8.1	28	001 01000	111001 1001	000110 1001
D9.1	29	001 01001	100101 1001	100101 1001
D10.1	2A	001 01010	010101 1001	010101 1001
D11.1	2B	001 01011	110100 1001	110100 1001
D12.1	2C	001 01100	001101 1001	001101 1001
D13.1	2D	001 01101	101100 1001	101100 1001
D14.1	2E	001 01110	011100 1001	011100 1001
D15.1	2F	001 01111	010111 1001	101000 1001
D16.1	30	001 10000	011011 1001	100100 1001
D17.1	31	001 10001	100011 1001	100011 1001
D18.1	32	001 10010	010011 1001	010011 1001
D19.1	33	001 10011	110010 1001	110010 1001
D20.1	34	001 10100	001011 1001	001011 1001
D21.1	35	001 10101	101010 1001	101010 1001
D22.1	36	001 10110	011010 1001	011010 1001
D23.1	37	001 10111	111010 1001	000101 1001
D24.1	38	001 11000	110011 1001	001100 1001
D25.1	39	001 11001	100110 1001	100110 1001
D26.1	3A	001 11010	010110 1001	010110 1001
D27.1	3B	001 11011	110110 1001	001001 1001

(continued)

Table 36–1b—Valid data code-groups

Code Group Name	Octet Value	Octet Bits HGF EDCBA	Current RD –	Current RD +
			abcdei fghj	abcdei fghj
D28.1	3C	001 11100	001110 1001	001110 1001
D29.1	3D	001 11101	101110 1001	010001 1001
D30.1	3E	001 11110	011110 1001	100001 1001
D31.1	3F	001 11111	101011 1001	010100 1001
D0.2	40	010 00000	100111 0101	011000 0101
D1.2	41	010 00001	011101 0101	100010 0101
D2.2	42	010 00010	101101 0101	010010 0101
D3.2	43	010 00011	110001 0101	110001 0101
D4.2	44	010 00100	110101 0101	001010 0101
D5.2	45	010 00101	101001 0101	101001 0101
D6.2	46	010 00110	011001 0101	011001 0101
D7.2	47	010 00111	111000 0101	000111 0101
D8.2	48	010 01000	111001 0101	000110 0101
D9.2	49	010 01001	100101 0101	100101 0101
D10.2	4A	010 01010	010101 0101	010101 0101
D11.2	4B	010 01011	110100 0101	110100 0101
D12.2	4C	010 01100	001101 0101	001101 0101
D13.2	4D	010 01101	101100 0101	101100 0101
D14.2	4E	010 01110	011100 0101	011100 0101
D15.2	4F	010 01111	010111 0101	101000 0101
D16.2	50	010 10000	011011 0101	100100 0101
D17.2	51	010 10001	100011 0101	100011 0101
D18.2	52	010 10010	010011 0101	010011 0101
D19.2	53	010 10011	110010 0101	110010 0101
D20.2	54	010 10100	001011 0101	001011 0101
D21.2	55	010 10101	101010 0101	101010 0101
D22.2	56	010 10110	011010 0101	011010 0101
D23.2	57	010 10111	111010 0101	000101 0101
D24.2	58	010 11000	110011 0101	001100 0101
D25.2	59	010 11001	100110 0101	100110 0101
D26.2	5A	010 11010	010110 0101	010110 0101
D27.2	5B	010 11011	110110 0101	001001 0101
D28.2	5C	010 11100	001110 0101	001110 0101
D29.2	5D	010 11101	101110 0101	010001 0101
D30.2	5E	010 11110	011110 0101	100001 0101
D31.2	5F	010 11111	101011 0101	010100 0101
D0.3	60	011 00000	100111 0011	011000 1100
D1.3	61	011 00001	011101 0011	100010 1100
D2.3	62	011 00010	101101 0011	010010 1100
D3.3	63	011 00011	110001 1100	110001 0011
D4.3	64	011 00100	110101 0011	001010 1100
D5.3	65	011 00101	101001 1100	101001 0011
D6.3	66	011 00110	011001 1100	011001 0011
D7.3	67	011 00111	111000 1100	000111 0011
D8.3	68	011 01000	111001 0011	000110 1100
D9.3	69	011 01001	100101 1100	100101 0011
D10.3	6A	011 01010	010101 1100	010101 0011
D11.3	6B	011 01011	110100 1100	110100 0011
D12.3	6C	011 01100	001101 1100	001101 0011
D13.3	6D	011 01101	101100 1100	101100 0011
D14.3	6E	011 01110	011100 1100	011100 0011
D15.3	6F	011 01111	010111 0011	101000 1100
D16.3	70	011 10000	011011 0011	100100 1100
D17.3	71	011 10001	100011 1100	100011 0011
D18.3	72	011 10010	010011 1100	010011 0011
D19.3	73	011 10011	110010 1100	110010 0011
D20.3	74	011 10100	001011 1100	001011 0011
D21.3	75	011 10101	101010 1100	101010 0011
D22.3	76	011 10110	011010 1100	011010 0011
D23.3	77	011 10111	111010 0011	000101 1100

(continued)

Table 36–1c—Valid data code-groups

Code Group Name	Octet Value	Octet Bits HGF EDCBA	Current RD –	Current RD +
			abcdei fghj	abcdei fghj
D24.3	78	011 11000	110011 0011	001100 1100
D25.3	79	011 11001	100110 1100	100110 0011
D26.3	7A	011 11010	010110 1100	010110 0011
D27.3	7B	011 11011	110110 0011	001001 1100
D28.3	7C	011 11100	001110 1100	001110 0011
D29.3	7D	011 11101	101110 0011	010001 1100
D30.3	7E	011 11110	011110 0011	100001 1100
D31.3	7F	011 11111	101011 0011	010100 1100
D0.4	80	100 00000	100111 0010	011000 1101
D1.4	81	100 00001	011101 0010	100010 1101
D2.4	82	100 00010	101101 0010	010010 1101
D3.4	83	100 00011	110001 1101	110001 0010
D4.4	84	100 00100	110101 0010	001010 1101
D5.4	85	100 00101	101001 1101	101001 0010
D6.4	86	100 00110	011001 1101	011001 0010
D7.4	87	100 00111	111000 1101	000111 0010
D8.4	88	100 01000	111001 0010	000110 1101
D9.4	89	100 01001	100101 1101	100101 0010
D10.4	8A	100 01010	010101 1101	010101 0010
D11.4	8B	100 01011	110100 1101	110100 0010
D12.4	8C	100 01100	001101 1101	001101 0010
D13.4	8D	100 01101	101100 1101	101100 0010
D14.4	8E	100 01110	011100 1101	011100 0010
D15.4	8F	100 01111	010111 0010	101000 1101
D16.4	90	100 10000	011011 0010	100100 1101
D17.4	91	100 10001	100011 1101	100011 0010
D18.4	92	100 10010	010011 1101	010011 0010
D19.4	93	100 10011	110010 1101	110010 0010
D20.4	94	100 10100	001011 1101	001011 0010
D21.4	95	100 10101	101010 1101	101010 0010
D22.4	96	100 10110	011010 1101	011010 0010
D23.4	97	100 10111	111010 0010	000101 1101
D24.4	98	100 11000	110011 0010	001100 1101
D25.4	99	100 11001	100110 1101	100110 0010
D26.4	9A	100 11010	010110 1101	010110 0010
D27.4	9B	100 11011	110110 0010	001001 1101
D28.4	9C	100 11100	001110 1101	001110 0010
D29.4	9D	100 11101	101110 0010	010001 1101
D30.4	9E	100 11110	011110 0010	100001 1101
D31.4	9F	100 11111	101011 0010	010100 1101
D0.5	A0	101 00000	100111 1010	011000 1010
D1.5	A1	101 00001	011101 1010	100010 1010
D2.5	A2	101 00010	101101 1010	010010 1010
D3.5	A3	101 00011	110001 1010	110001 1010
D4.5	A4	101 00100	110101 1010	001010 1010
D5.5	A5	101 00101	101001 1010	101001 1010
D6.5	A6	101 00110	011001 1010	011001 1010
D7.5	A7	101 00111	111000 1010	000111 1010
D8.5	A8	101 01000	111001 1010	000110 1010
D9.5	A9	101 01001	100101 1010	100101 1010
D10.5	AA	101 01010	010101 1010	010101 1010
D11.5	AB	101 01011	110100 1010	110100 1010
D12.5	AC	101 01100	001101 1010	001101 1010
D13.5	AD	101 01101	101100 1010	101100 1010
D14.5	AE	101 01110	011100 1010	011100 1010
D15.5	AF	101 01111	010111 1010	101000 1010
D16.5	B0	101 10000	011011 1010	100100 1010
D17.5	B1	101 10001	100011 1010	100011 1010
D18.5	B2	101 10010	010011 1010	010011 1010
D19.5	B3	101 10011	110010 1010	110010 1010

(continued)

Table 36–1d—Valid data code-groups

Code Group Name	Octet Value	Octet Bits HGF EDCBA	Current RD –	Current RD +
			abcdei fghj	abcdei fghj
D20.5	B4	101 10100	001011 1010	001011 1010
D21.5	B5	101 10101	101010 1010	101010 1010
D22.5	B6	101 10110	011010 1010	011010 1010
D23.5	B7	101 10111	111010 1010	000101 1010
D24.5	B8	101 11000	110011 1010	001100 1010
D25.5	B9	101 11001	100110 1010	100110 1010
D26.5	BA	101 11010	010110 1010	010110 1010
D27.5	BB	101 11011	110110 1010	001001 1010
D28.5	BC	101 11100	001110 1010	001110 1010
D29.5	BD	101 11101	101110 1010	010001 1010
D30.5	BE	101 11110	011110 1010	100001 1010
D31.5	BF	101 11111	101011 1010	010100 1010
D0.6	C0	110 00000	100111 0110	011000 0110
D1.6	C1	110 00001	011101 0110	100010 0110
D2.6	C2	110 00010	101101 0110	010010 0110
D3.6	C3	110 00011	110001 0110	110001 0110
D4.6	C4	110 00100	110101 0110	001010 0110
D5.6	C5	110 00101	101001 0110	101001 0110
D6.6	C6	110 00110	011001 0110	011001 0110
D7.6	C7	110 00111	111000 0110	000111 0110
D8.6	C8	110 01000	111001 0110	000110 0110
D9.6	C9	110 01001	100101 0110	100101 0110
D10.6	CA	110 01010	010101 0110	010101 0110
D11.6	CB	110 01011	110100 0110	110100 0110
D12.6	CC	110 01100	001101 0110	001101 0110
D13.6	CD	110 01101	101100 0110	101100 0110
D14.6	CE	110 01110	011100 0110	011100 0110
D15.6	CF	110 01111	010111 0110	101000 0110
D16.6	D0	110 10000	011011 0110	100100 0110
D17.6	D1	110 10001	100011 0110	100011 0110
D18.6	D2	110 10010	010011 0110	010011 0110
D19.6	D3	110 10011	110010 0110	110010 0110
D20.6	D4	110 10100	001011 0110	001011 0110
D21.6	D5	110 10101	101010 0110	101010 0110
D22.6	D6	110 10110	011010 0110	011010 0110
D23.6	D7	110 10111	111010 0110	000101 0110
D24.6	D8	110 11000	110011 0110	001100 0110
D25.6	D9	110 11001	100110 0110	100110 0110
D26.6	DA	110 11010	010110 0110	010110 0110
D27.6	DB	110 11011	110110 0110	001001 0110
D28.6	DC	110 11100	001110 0110	001110 0110
D29.6	DD	110 11101	101110 0110	010001 0110
D30.6	DE	110 11110	011110 0110	100001 0110
D31.6	DF	110 11111	101011 0110	010100 0110
D0.7	E0	111 00000	100111 0001	011000 1110
D1.7	E1	111 00001	011101 0001	100010 1110
D2.7	E2	111 00010	101101 0001	010010 1110
D3.7	E3	111 00011	110001 1110	110001 0001
D4.7	E4	111 00100	110101 0001	001010 1110
D5.7	E5	111 00101	101001 1110	101001 0001
D6.7	E6	111 00110	011001 1110	011001 0001
D7.7	E7	111 00111	111000 1110	000111 0001
D8.7	E8	111 01000	111001 0001	000110 1110
D9.7	E9	111 01001	100101 1110	100101 0001
D10.7	EA	111 01010	010101 1110	010101 0001
D11.7	EB	111 01011	110100 1110	110100 1000
D12.7	EC	111 01100	001101 1110	001101 0001
D13.7	ED	111 01101	101100 1110	101100 1000
D14.7	EE	111 01110	011100 1110	011100 1000
D15.7	EF	111 01111	010111 0001	101000 1110

(continued)

Table 36–1e—Valid data code-groups

Code Group Name	Octet Value	Octet Bits HGF EDCBA	Current RD –	Current RD +
			abcdei fghj	abcdei fghj
D16.7	F0	111 10000	011011 0001	100100 1110
D17.7	F1	111 10001	100011 0111	100011 0001
D18.7	F2	111 10010	010011 0111	010011 0001
D19.7	F3	111 10011	110010 1110	110010 0001
D20.7	F4	111 10100	001011 0111	001011 0001
D21.7	F5	111 10101	101010 1110	101010 0001
D22.7	F6	111 10110	011010 1110	011010 0001
D23.7	F7	111 10111	111010 0001	000101 1110
D24.7	F8	111 11000	110011 0001	001100 1110
D25.7	F9	111 11001	100110 1110	100110 0001
D26.7	FA	111 11010	010110 1110	010110 0001
D27.7	FB	111 11011	110110 0001	001001 1110
D28.7	FC	111 11100	001110 1110	001110 0001
D29.7	FD	111 11101	101110 0001	010001 1110
D30.7	FE	111 11110	011110 0001	100001 1110
D31.7	FF	111 11111	101011 0001	010100 1110

(concluded)

Table 36–2—Valid special code-groups

Code Group Name	Octet Value	Octet Bits HGF EDCBA	Current RD –	Current RD +	Notes
			abcdei fghj	abcdei fghj	
K28.0	1C	000 11100	001111 0100	110000 1011	1
K28.1	3C	001 11100	001111 1001	110000 0110	1,2
K28.2	5C	010 11100	001111 0101	110000 1010	1
K28.3	7C	011 11100	001111 0011	110000 1100	1
K28.4	9C	100 11100	001111 0010	110000 1101	1
K28.5	BC	101 11100	001111 1010	110000 0101	2
K28.6	DC	110 11100	001111 0110	110000 1001	1
K28.7	FC	111 11100	001111 1000	110000 0111	1,2
K23.7	F7	111 10111	111010 1000	000101 0111	
K27.7	FB	111 11011	110110 1000	001001 0111	
K29.7	FD	111 11101	101110 1000	010001 0111	
K30.7	FE	111 11110	011110 1000	100001 0111	

NOTE 1—Reserved.

NOTE 2—Contains a comma.

36.2.4.8 /K28.5/ code-group considerations

The /K28.5/ special code-group is chosen as the first code-group of all ordered_sets that are signaled repeatedly and for the purpose of allowing a receiver to synchronize to the incoming bit stream (i.e., /C/ and /I/), for the following reasons:

a) Bits abcdeif make up a comma. The comma can be used to easily find and verify code-group and ordered_set boundaries of the rx_bit stream.
b) Bits ghj of the encoded code-group present the maximum number of transitions, simplifying receiver acquisition of bit synchronization.

Table 36–3—Defined ordered_sets

Code	Ordered_Set	Number of Code-Groups	Encoding
/C/	**Configuration**		Alternating /C1/ and /C2/
/C1/	Configuration 1	4	/K28.5/D21.5/Config_Reg[a]
/C2/	Configuration 2	4	/K28.5/D2.2/Config_Reg[a]
/I/	**IDLE**		Correcting /I1/, Preserving /I2/
/I1/	IDLE 1	2	/K28.5/D5.6/
/I2/	IDLE 2	2	/K28.5/D16.2/
	Encapsulation		
/R/	Carrier_Extend	1	/K23.7/
/S/	Start_of_Packet	1	/K27.7/
/T/	End_of_Packet	1	/K29.7/
/V/	Error_Propagation	1	/K30.7/

[a]Two data code-groups representing the Config_Reg value.

36.2.4.9 Comma considerations

The seven bit comma string is defined as either b'0011111' (comma+) or b'1100000' (comma-). The /I/ and /C/ ordered_sets and their associated protocols are specified to ensure that comma+ is transmitted with either equivalent or greater frequency than comma- for the duration of their transmission. This is done to ensure compatibility with common components.

The comma contained within the /K28.1/, /K28.5/, and /K28.7/ special code-groups is a singular bit pattern, which, in the absence of transmission errors, cannot appear in any other location of a code-group and cannot be generated across the boundaries of any two adjacent code-groups with the following exception:

The /K28.7/ special code-group is used by 1000BASE-X for diagnostic purposes only (see Annex 36A). This code-group, if followed by any of the following special or data code-groups: /K28.x/, /D3.x/, /D11.x/, /D12.x/, /D19.x/, /D20.x/, or /D28.x/, where x is a value in the range 0 to 7, inclusive, causes a comma to be generated across the boundaries of the two adjacent code-groups. A comma across the boundaries of any two adjacent code-groups may cause code-group realignment (see 36.3.2.4).

36.2.4.10 Configuration (/C/)

Configuration, defined as the continuous repetition of the ordered sets /C1/ and /C2/, is used to convey the 16-bit Configuration Register (Config_Reg) to the link partner. See Clause 37 for a description of the Config_Reg contents.

The ordered_sets, /C1/ and /C2/, are defined in Table 36–3. The /C1/ ordered_set is defined such that the running disparity at the end of the first two code-groups is opposite that of the beginning running disparity. The /C2/ ordered_set is defined such that the running disparity at the end of the first two code-groups is the same as the beginning running disparity. For a constant Config_Reg value, the running disparity after transmitting the sequence /C1/C2/ will be the opposite of what it was at the start of the sequence. This ensures that K28.5s containing comma+ will be sent during configuration.

36.2.4.11 Data (/D/)

A data code-group, when not used to distinguish or convey information for a defined ordered_set, conveys one octet of arbitrary data between the GMII and the PCS. The sequence of data code-groups is arbitrary, where any data code-group can be followed by any other data code-group. Data code-groups are coded and decoded but not interpreted by the PCS. Successful decoding of the data code-groups depends on proper receipt of the Start_of_Packet delimiter, as defined in 36.2.4.13 and the checking of validity, as defined in 36.2.4.6.

36.2.4.12 IDLE (/I/)

IDLE ordered_sets (/I/) are transmitted continuously and repetitively whenever the GMII is idle (TX_EN and TX_ER are both inactive). /I/ provides a continuous fill pattern to establish and maintain clock synchronization. /I/ is emitted from, and interpreted by, the PCS. /I/ consists of one or more consecutively transmitted /I1/ or /I2/ ordered_sets, as defined in Table 36–3.

The /I1/ ordered_set is defined such that the running disparity at the end of the transmitted /I1/ is opposite that of the beginning running disparity. The /I2/ ordered_set is defined such that the running disparity at the end of the transmitted /I2/ is the same as the beginning running disparity. The first /I/ following a packet or Configuration ordered_set restores the current positive or negative running disparity to a negative value. All subsequent /I/s are /I2/ to ensure negative ending running disparity.

Distinct carrier events are separated by /I/s.

Implementations of this standard may benefit from the ability to add or remove /I2/ from the code-group stream one /I2/ at a time without altering the beginning running disparity associated with the code-group subsequent to the removed /I2/.

A received ordered set which consists of two code-groups, the first of which is /K28.5/ and the second of which is a data code-group other than /D21.5/ or /D2.2/ is treated as an /I/ ordered_set.

36.2.4.13 Start_of_Packet (SPD) delimiter

A Start_of_Packet delimiter (SPD) is used to delineate the starting boundary of a data transmission sequence and to authenticate carrier events. Upon each fresh assertion of TX_EN by the GMII, and subsequent to the completion of PCS transmission of the current ordered_set, the PCS replaces the current octet of the MAC preamble with SPD. Upon initiation of packet reception, the PCS replaces the received SPD delimiter with the data octet value associated with the first preamble octet. A SPD delimiter consists of the code-group /S/, as defined in Table 36–3.

SPD follows /I/ for a single packet or the first packet in a burst.

SPD follows /R/ for the second and subsequent packets of a burst.

36.2.4.14 End_of_Packet delimiter (EPD)

An End_of_Packet delimiter (EPD) is used to delineate the ending boundary of a packet. The EPD is transmitted by the PCS following each de-assertion of TX_EN on the GMII, which follows the last data octet comprising the FCS of the MAC packet. On reception, EPD is interpreted by the PCS as terminating a packet. A EPD delimiter consists of the code-groups /T/R/R/ or /T/R/K28.5/. The code-group /T/ is defined in Table 36–3. See 36.2.4.15 for the definition of code-groups used for /R/. /K28.5/ normally occurs as the first code-group of the /I/ ordered_set. See 36.2.4.12 for the definition of code-groups used for /I/.

The receiver considers the MAC interpacket gap (IPG) to have begun two octets prior to the transmission of /I/. For example, when a packet is terminated by EPD, the /T/R/ portion of the EPD occupies part of the region considered by the MAC to be the IPG.

36.2.4.14.1 EPD rules

a) The PCS transmits a /T/R/ following the last data octet from the MAC;
b) If the MAC indicates carrier extension to the PCS, Carrier_Extend rules are in effect. See 36.2.4.15.1;
c) If the MAC does not indicate carrier extension to the PCS, perform the following:
    1) If /R/ is transmitted in an even-numbered code-group position, the PCS appends a single additional /R/ to the code-group stream to ensure that the subsequent /I/ is aligned on an even-numbered code-group boundary and EPD transmission is complete;
    2) The PCS transmits /I/.

36.2.4.15 Carrier_Extend (/R/)

Carrier_Extend (/R/) is used for the following purposes:

a) Carrier extension: Used by the MAC to extend the duration of the carrier event. When used for this purpose, carrier extension is emitted from and interpreted by the MAC and coded to and decoded from the corresponding code-group by the PCS. In order to extend carrier, the GMII must deassert TX_EN. The deassertion of TX_EN and simultaneous assertion of TX_ER causes the PCS to emit an /R/ with a two-octet delay, which gives the PCS time to complete its EPD before commencing transmissions. The number of /R/ code-groups emitted from the PCS equals the number of GMII GTX_CLK periods during which it extends carrier;
b) Packet separation: Carrier extension is used by the MAC to separate packets within a burst of packets. When used for this purpose, carrier extension is emitted from and interpreted by the MAC and coded to and decoded from the corresponding code-group by the PCS;
c) EPD2: The first /R/ following the /T/ in the End_of_Packet delimiters /T/R/I/ or /T/R/R/I/;
d) EPD3: The second /R/ following the /T/ in the End_of_Packet delimiter /T/R/R/I/. This /R/ is used, if necessary, to pad the only or last packet of a burst of packets so that the subsequent /I/ is aligned on an even-numbered code-group boundary. When used for this purpose, Carrier_Extend is emitted from, and interpreted by, the PCS. An EPD of /T/R/R/ results in one /R/ being delivered to the PCS client (see 36.2.4.14.1).

Carrier_Extend consists of one or more consecutively transmitted /R/ ordered_sets, as defined in Table 36–3.

36.2.4.15.1 Carrier_Extend rules

a) If the MAC indicates carrier extension to the PCS, the initial /T/R/ is followed by one /R/ for each octet of carrier extension received from the MAC;
b) If the last /R/ is transmitted in an even-numbered code-group position, the PCS appends a single additional /R/ to the code-group stream to ensure that the subsequent /I/ is aligned on an even-numbered code-group boundary.

36.2.4.16 Error_Propagation (/V/)

Error_Propagation (/V/) indicates that the PCS client wishes to indicate a transmission error to its peer entity. The normal use of Error_Propagation is for repeaters to propagate received errors. /V/ is emitted from the PCS, at the request of the PCS client through the use of the TX_ER signal, as described in Clause 35. Error_Propagation is emitted from, and interpreted by, the PCS. Error_Propagation consists of the ordered_set /V/, as defined in Table 36–3.

The presence of Error_Propagation or any invalid code-group on the medium denotes a collision artifact or an error condition. Invalid code-groups are not intentionally transmitted onto the medium by DTEs. The PCS processes and conditionally indicates the reception of /V/ or an invalid code-group on the GMII as false carrier, data errors, or carrier extend errors, depending on its current context.

36.2.4.17 Encapsulation

The 1000BASE-X PCS accepts packets from the MAC through the Reconciliation sublayer and GMII. Due to the continuously signaled nature of the underlying PMA, and the encoding performed by the PCS, the 1000BASE-X PCS encapsulates MAC frames into a code-group stream. The PCS decodes the code-group stream received from the PMA, extracts packets from it, and passes the packets to the MAC via the Reconciliation sublayer and GMII.

Figure 36–4 depicts the PCS encapsulation of a MAC packet based on GMII signals.

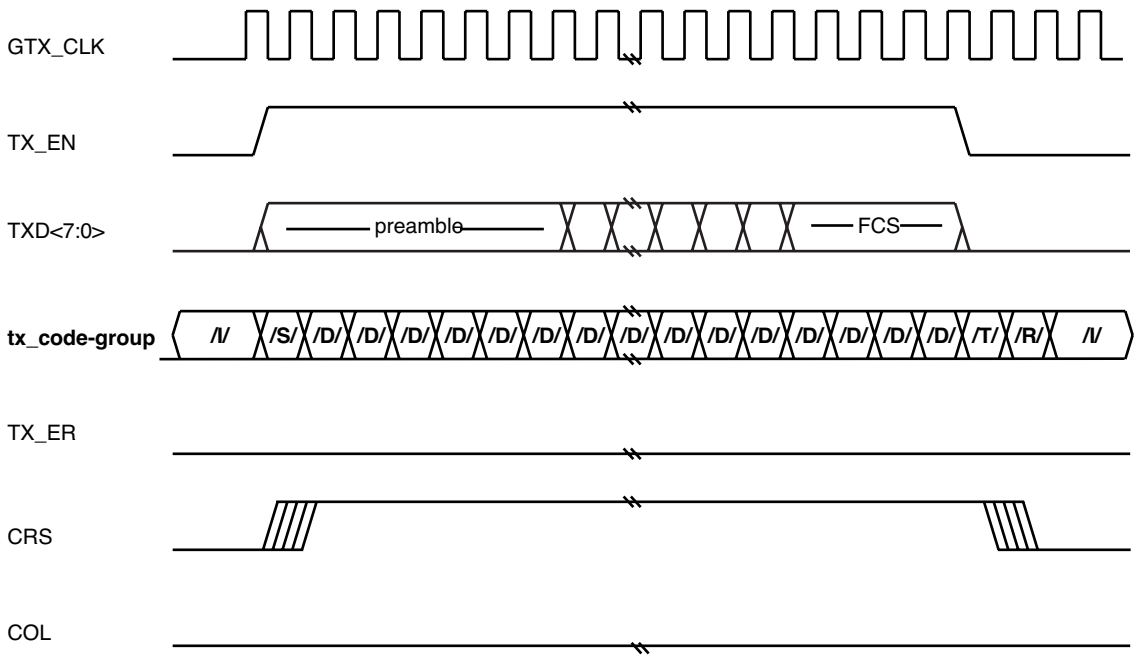


Figure 36–4—PCS encapsulation

36.2.4.18 Mapping between GMII, PCS and PMA

Figure 36–3 depicts the mapping of the octet-wide data path of the GMII to the ten-bit-wide code-groups of the PCS, and the one-bit paths of the PMA/PMD interface.

The PCS encodes an octet received from the GMII into a ten-bit code-group, according to Figure 36–3. Code-groups are serialized into a tx_bit stream by the PMA and passed to the PMD for transmission on the underlying medium, according to Figure 36–3. The first transmitted tx_bit is tx_code-group<0>, and the last tx_bit transmitted is tx_code-group<9>. There is no numerical significance ascribed to the bits within a code-group; that is, the code-group is simply a ten-bit pattern that has some predefined interpretation.

Similarly, the PMA deserializes rx_bits received from the PMD, according to Figure 36–3. The PCS Receive process converts rx_code-group<9:0>'s into GMII data octets, according to 36.2.5.2.2.

36.2.5 Detailed functions and state diagrams

The notation used in the state diagrams in this clause follow the conventions in 21.5. State diagram variables follow the conventions of 21.5.2 except when the variable has a default value. Variables in a state diagram with default values evaluate to the variable default in each state where the variable value is not explicitly set.

Timeless states are employed as an editorial convenience to facilitate the distribution of transition conditions from prior states. No actions are taken within these states. Exit conditions are evaluated for timeless states. Timeless states are as follows:

a) PCS transmit ordered_set state TX_PACKET;
b) PCS transmit code-group state GENERATE_CODE_GROUPS;
c) PCS transmit code-group state IDLE_DISPARITY_TEST;
d) PCS receive state RECEIVE;
e) PCS receive state EPD2_CHECK_END.

36.2.5.1 State variables

36.2.5.1.1 Notation conventions

/x/
: Denotes the constant code-group specified in 36.2.5.1.2 (valid code-groups must follow the rules of running disparity as per 36.2.4.5 and 36.2.4.6).

[/x/]
: Denotes the latched received value of the constant code-group (/x/) specified in 36.2.5.1.2 and conveyed by the SYNC_UNITDATA.indicate message described in 36.2.5.1.6.

36.2.5.1.2 Constants

/C/
: The Configuration ordered_set group, comprising either the /C1/ or /C2/ ordered_set, as specified in 36.2.4.10. Conveys the Config_Reg value as tx_Config_Reg<D15:D0> for the PCS Transmit process and rx_Config_Reg<D15:D0> for the PCS Receive process.

/COMMA/
: The set of special code-groups which include a comma as specified in 36.2.4.9 and listed in Table 36–2.

/D/
: The set of 256 code-groups corresponding to valid data, as specified in 36.2.4.11.

/Dx.y/
: One of the set of 256 code-groups corresponding to valid data, as specified in 36.2.4.11.

/I/
: The IDLE ordered_set group, comprising either the /I1/ or /I2/ ordered_sets, as specified in 36.2.4.12.

/INVALID/
: The set of invalid data or special code-groups, as specified in 36.2.4.6.

/Kx.y/
: One of the set of 12 code-groups corresponding to valid special code-groups, as specified in Table 36–2.

/R/
The code-group used as either: End_of_Packet delimiter part 2; End_of_Packet delimiter part 3; Carrier_Extend; and /I/ alignment.

/S/
The code-group corresponding to the Start_of_Packet delimiter (SPD) as specified in 36.2.4.13.

/T/
The code-group used for the End_of_Packet delimiter part 1.

/V/
The Error_Propagation code-group, as specified in 36.2.4.16.

36.2.5.1.3 Variables

cgbad
Alias for the following terms: ((rx_code-groupŒ/INVALID/) + (rx_code-group=/COMMA/ *rx_even=TRUE)) * PMA_UNITDATA.indicate

cggood
Alias for the following terms: !((rx_code-groupŒ/INVALID/) + (rx_code-group=/COMMA/ *rx_even=TRUE)) * PMA_UNITDATA.indicate

COL
The COL signal of the GMII as specified in Clause 35.

CRS
The CRS signal of the GMII as specified in Clause 35.

EVEN
The latched state of the rx_even variable, when rx_even=TRUE, as conveyed by the SYNC_UNITDATA.indicate message described in 36.2.5.1.6.

mr_loopback
A boolean that indicates the enabling and disabling of data being looped back through the PHY. Loopback of data through the PHY is enabled when Control register bit 0.14 is set to one.

Values: FALSE; Loopback through the PHY is disabled.
TRUE; Loopback through the PHY is enabled.

mr_main_reset
Controls the resetting of the PCS via Control Register bit 0.15.

Values: FALSE; Do not reset the PCS.
TRUE; Reset the PCS.

ODD
The latched state of the rx_even variable, when rx_even=FALSE, as conveyed by the SYNC_UNITDATA.indicate message described in 36.2.5.1.6.

power_on
Condition that is true until such time as the power supply for the device that contains the PCS has reached the operating region. The condition is also true when the device has low power mode set via Control register bit 0.11.

Values: FALSE; The device is completely powered (default).
TRUE; The device has not been completely powered.

NOTE—Power_on evaluates to its default value in each state where it is not explicitly set.

receiving

A boolean set by the PCS Receive process to indicate carrier activity. Used by the Carrier Sense process, and also interpreted by the PCS Transmit process for indicating a collision. (See also 36.2.5.1.4, carrier_detect(x).)

Values: TRUE; Carrier being received.
FALSE; Carrier not being received.

repeater_mode

A boolean used to make the assertion of Carrier Sense occur only in response to receive activity when the PCS is used in a CSMA/CD repeater. This variable is set to TRUE in a repeater application, and set to FALSE in all other applications.

Values: TRUE; Allows the assertion of CRS in response to receive activity only.
FALSE; Allows the assertion of CRS in response to either transmit or receive activity.

rx_bit

A binary parameter conveyed by the PMD_UNITDATA.indicate service primitive, as specified in 38.1.1.2, to the PMA.

Values: ZERO; Data bit is a logical zero.
ONE; Data bit is a logical one.

rx_code-group<9:0>

A 10-bit vector represented by the most recently received code-group from the PMA. The element rx_code-group<0> is the least recently received (oldest) rx_bit; rx_code-group<9> is the most recently received rx_bit (newest). When code-group alignment has been achieved, this vector contains precisely one code-group.

rx_Config_Reg<D15:D0>

A 16-bit array that contains the data bits received from a /C/ ordered_set as defined in 36.2.4.10. Conveyed by the PCS Receive process to the PCS Auto-Negotiation process. The format of the data bits is context dependent, relative to the state of the Auto-Negotiation function, and is presented in 37.2.1.1 and 37.2.4.3.1. For each element within the array:

Values: ZERO; Data bit is a logical zero.
ONE; Data bit is a logical one.

RX_DV

The RX_DV signal of the GMII as specified in Clause 35. Set by the PCS Receive process.

RX_ER

The RX_ER signal of the GMII as specified in Clause 35. Set by the PCS Receive process.

rx_even

A boolean set by the PCS Synchronization process to designate received code-groups as either even- or odd-numbered code-groups as specified in 36.2.4.2.

Values: TRUE; Even-numbered code-group being received.
FALSE; Odd-numbered code-group being received.

RXD<7:0>

The RXD<7:0> signal of the GMII as specified in Clause 35. Set by the PCS Receive process.

signal_detect

A boolean set by the PMD continuously via the PMD_SIGNAL.indicate(signal_detect) message to indicate the status of the incoming link signal.

Values: FAIL; A signal is not present on the link.
OK; A signal is present on the link.

sync_status

A parameter set by the PCS Synchronization process to reflect the status of the link as viewed by the receiver.

Values: FAIL; The receiver is not synchronized to code-group boundaries.
OK; The receiver is synchronized to code-group boundaries.

transmitting

A boolean set by the PCS Transmit process to indicate that packet transmission is in progress. Used by the Carrier Sense process and internally by the PCS Transmit process for indicating a collision.

Values: TRUE; The PCS is transmitting a packet.
FALSE; The PCS is not transmitting a packet.

tx_bit

A binary parameter used to convey data from the PMA to the PMD via the PMD_UNITDATA.request service primitive as specified in 38.1.1.1.

Values: ZERO; Data bit is a logical zero.
ONE; Data bit is a logical one.

tx_code-group<9:0>

A vector of bits representing one code-group, as specified in Tables 36–1a or 36–2, which has been prepared for transmission by the PCS Transmit process. This vector is conveyed to the PMA as the parameter of a PMD_UNITDATA.request(tx_bit) service primitive. The element tx_code-group<0> is the first tx_bit transmitted; tx_code-group<9> is the last tx_bit transmitted.

tx_Config_Reg<D15:D0>

A 16-bit array that contains the data bits to be transmitted in a /C/ ordered_set as defined in 36.2.4.10. Conveyed by the PCS Auto-Negotiation process to the PCS Transmit process. The format of the data bits is context dependent, relative to the state of the Auto-Negotiation function, and is presented in 37.2.1.1 and 37.2.4.3.1. For each element within the array:

Values: ZERO; Data bit is a logical zero.
ONE; Data bit is a logical one.

tx_disparity

A boolean set by the PCS Transmit process to indicate the running disparity at the end of code-group transmission as a binary value. Running disparity is described in 36.2.4.3.

Values: POSITIVE
NEGATIVE

TX_EN

The TX_EN signal of the GMII as specified in Clause 35.

TX_ER

The TX_ER signal of the GMII as specified in Clause 35.

tx_even

A boolean set by the PCS Transmit process to designate transmitted code-groups as either even- or odd-numbered code-groups as specified in 36.2.4.2.

Values: TRUE; Even-numbered code-group being transmitted.
FALSE; Odd-numbered code-group being transmitted.

tx_o_set

One of the following defined ordered_sets: /C/, /T/, /R/, /I/, /S/, /V/, or the code-group /D/.

TXD<7:0>

The TXD<7:0> signal of the GMII as specified in Clause 35.

xmit

Defined in 37.3.1.1.

36.2.5.1.4 Functions

carrier_detect

In the PCS Receive process, this function uses for input the latched code-group ([/x/]) and latched rx_even (EVEN/ODD) parameters of the SYNC_UNITDATA.indicate message from the PCS Synchronization process. When SYNC_UNITDATA.indicate message indicates EVEN, the carrier_detect function detects carrier when either:

a) A two or more bit difference between [/x/] and both /K28.5/ encodings exists (see Table 36–2); or

b) A two to nine bit difference between [/x/] and the expected /K28.5/ (based on current running disparity) exists.

Values: TRUE; Carrier is detected.
FALSE; Carrier is not detected.

check_end

Prescient End_of_Packet and Carrier_Extend function used by the PCS Receive process to set RX_ER and RXD<7:0> signals. The check_end function returns the current and next two code-groups in rx_code-group<9:0>.

DECODE ([/x]/)

In the PCS Receive process, this function takes as its argument the latched value of rx_code-group<9:0> ([/x/]) and the current running disparity, and returns the corresponding GMII RXD<7:0>, rx_Config_Reg<D7:D0>, or rx_Config_Reg<D15:D8> octet, per Table 36–1a–e. DECODE also updates the current running disparity per the running disparity rules outlined in 36.2.4.4.

ENCODE(x)

In the PCS Transmit process, this function takes as its argument (x), where x is a GMII TXD<7:0>, tx_Config_Reg<D7:D0>, or tx_Config_Reg<D15:D8> octet, and the current running disparity, and returns the corresponding ten-bit code-group per Table 36–1a. ENCODE also updates the current running disparity per Table 36–1a–e.

signal_detectCHANGE

In the PCS Synchronization process, this function monitors the signal_detect variable for a state change. The function is set upon state change detection.

Values: TRUE; A signal_detect variable state change has been detected.
FALSE; A signal_detect variable state change has not been detected (default).

NOTE—Signal_detectCHANGE is set by this function definition; it is not set explicitly in the state diagrams. Signal_detectCHANGE evaluates to its default value upon state entry.

VOID(x)

x Œ /D/, /T/, /R/, /K28.5/. Substitutes /V/ on a per code-group basis as requested by the GMII.

If [TX_EN=FALSE * TX_ER=TRUE * TXD≠(0000 1111)],

then return /V/;

Else if [TX_EN=TRUE * TX_ER=TRUE],

then return /V/;

Else return x.

xmitCHANGE

In the PCS Transmit process, this function monitors the xmit variable for a state change. The function is set upon state change detection.

Values: TRUE; An xmit variable state change has been detected.
FALSE; An xmit variable state change has not been detected (default).

NOTE—XmitCHANGE is set by this function definition; it is not set explicitly in the state diagrams. XmitCHANGE evaluates to its default value upon entry to state TX_TEST_XMIT.

36.2.5.1.5 Counters

good_cgs

Count of consecutive valid code-groups received.

36.2.5.1.6 Message

PMA_UNITDATA.indicate(rx_code-group<9:0>)

A signal sent by the PMA Receive process conveying the next code-group received over the medium (see 36.3.1.2).

PMA_UNITDATA.request(tx_code-group<9:0>)

A signal sent to the PMA Transmit process conveying the next code-group ready for transmission over the medium (see 36.3.1.1).

PMD_SIGNAL.indicate(signal_detect)

A signal sent by the PMD to indicate the status of the signal being received on the MDI.

PUDI

Alias for PMA_UNITDATA.indicate(rx_code-group<9:0>).

PUDR

Alias for PMA_UNITDATA.request(tx_code-group<9:0>).

RUDI

Alias for RX_UNITDATA.indicate(parameter).

RX_UNITDATA.indicate(parameter)

A signal sent by the PCS Receive process to the PCS Auto_Negotiation process conveying the following parameters:

Parameters: INVALID; indicates that an error condition has been detected while receiving /C/ or /I/ ordered_sets;

/C/; the /C/ ordered_set has been received;
/I/; the /I/ ordered_set has been received.

SUDI
Alias for SYNC_UNITDATA.indicate(parameters).

SYNC_UNITDATA.indicate(parameters)
A signal sent by the PCS Synchronization process to the PCS Receive process conveying the following parameters:

Parameters: [/x/]; the latched value of the indicated code-group (/x/);
EVEN/ODD; The latched state of the rx_even variable;

Value: EVEN; Passed when the latched state of rx_even=TRUE.
ODD; Passed when the latched state of rx_even=FALSE.

TX_OSET.indicate
A signal sent to the PCS Transmit ordered_set process from the PCS Transmit code-group process signifying the completion of transmission of one ordered_set.

36.2.5.1.7 Timer

cg_timer
A continuous free-running timer.

Values: The condition cg_timer_done becomes true upon timer expiration.

Restart when: immediately after expiration; restarting the timer resets the condition cg_timer_done.

Duration: 8 ns nominal.

If the GMII is implemented, cg_timer shall expire synchronously with the rising edge of GTX_CLK (see tolerance required for GTX_CLK in 35.4.2.3). In the absence of a GMII, cg_timer shall expire every 8 ns ± 0.01%. In the PCS transmit code-group state diagram, the message PMA_UNITDATA.request is issued concurrently with cg_timer_done.

36.2.5.2 State diagrams

36.2.5.2.1 Transmit

The PCS Transmit process is depicted in two state diagrams: PCS Transmit ordered_set and PCS Transmit code-group. The PCS shall implement its Transmit process as depicted in Figures 36–5 and 36–6, including compliance with the associated state variables as specified in 36.2.5.1.

The Transmit ordered_set process continuously sources ordered_sets to the Transmit code-group process. When initially invoked, and when the Auto-Negotiation process xmit flag indicates CONFIGURATION, the Auto-Negotiation process is invoked. When the Auto-Negotiation process xmit flag indicates IDLE, and between packets (as delimited by the GMII), /I/ is sourced. Upon the assertion of TX_EN by the GMII when the Auto-Negotiation process xmit flag indicates DATA, the SPD ordered_set is sourced. Following the SPD, /D/ code-groups are sourced until TX_EN is deasserted. Following the de-assertion of TX_EN, EPD ordered_sets are sourced. If TX_ER is asserted when TX_EN is deasserted and carrier extend error is not indicated by TXD, /R/ ordered_sets are sourced for as many GTX_CLK periods as TX_ER is asserted with a delay of two GTX_CLK periods to first source the /T/ and /R/ ordered sets. If carrier extend error is indicated by TXD during carrier extend, /V/ ordered_sets are sourced. If TX_EN and TX_ER are both de-asserted, the /R/ ordered_set may be sourced, after which the sourcing of /I/ is resumed. If, while TX_EN is

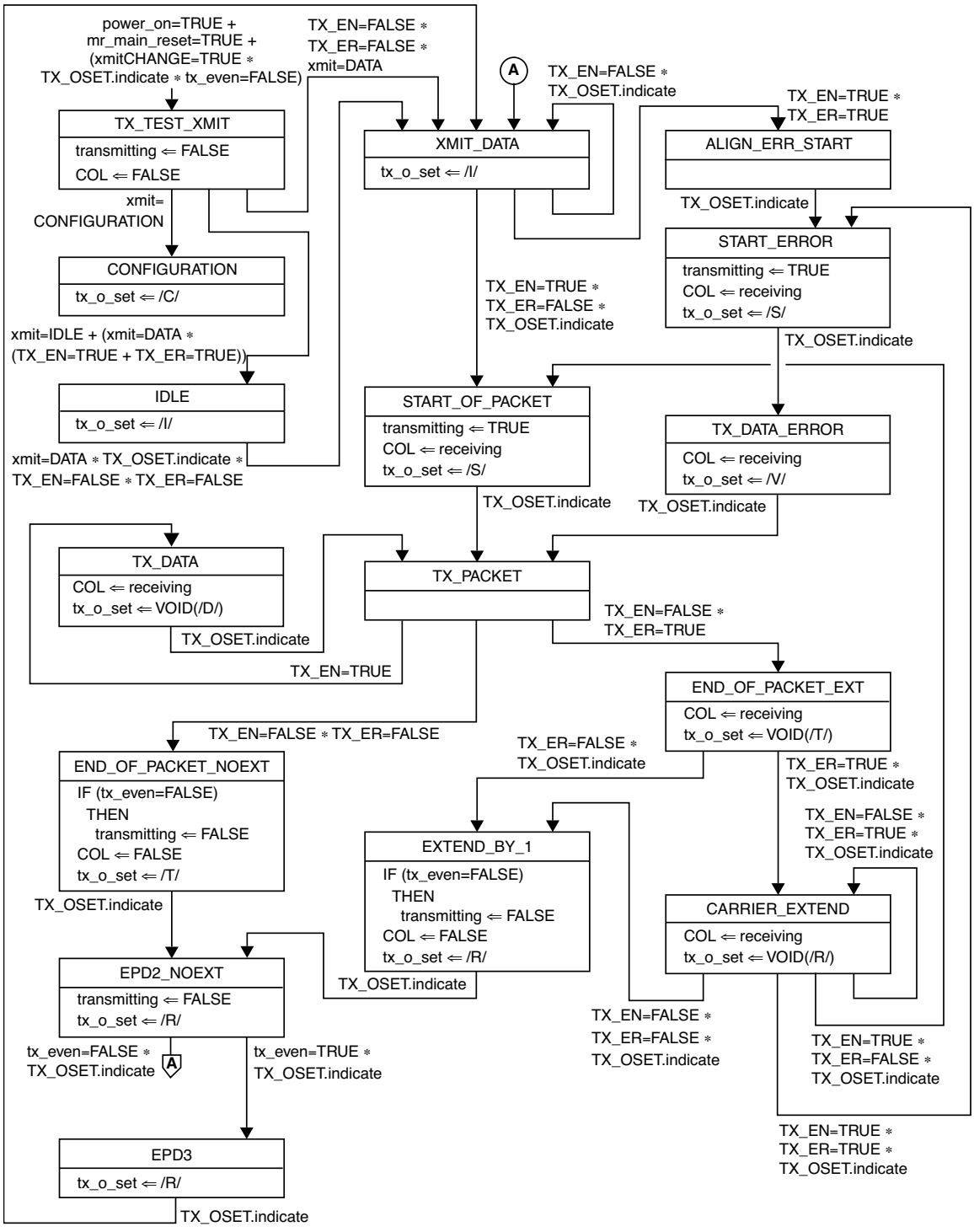


Figure 36–5—PCS transmit ordered_set state diagram

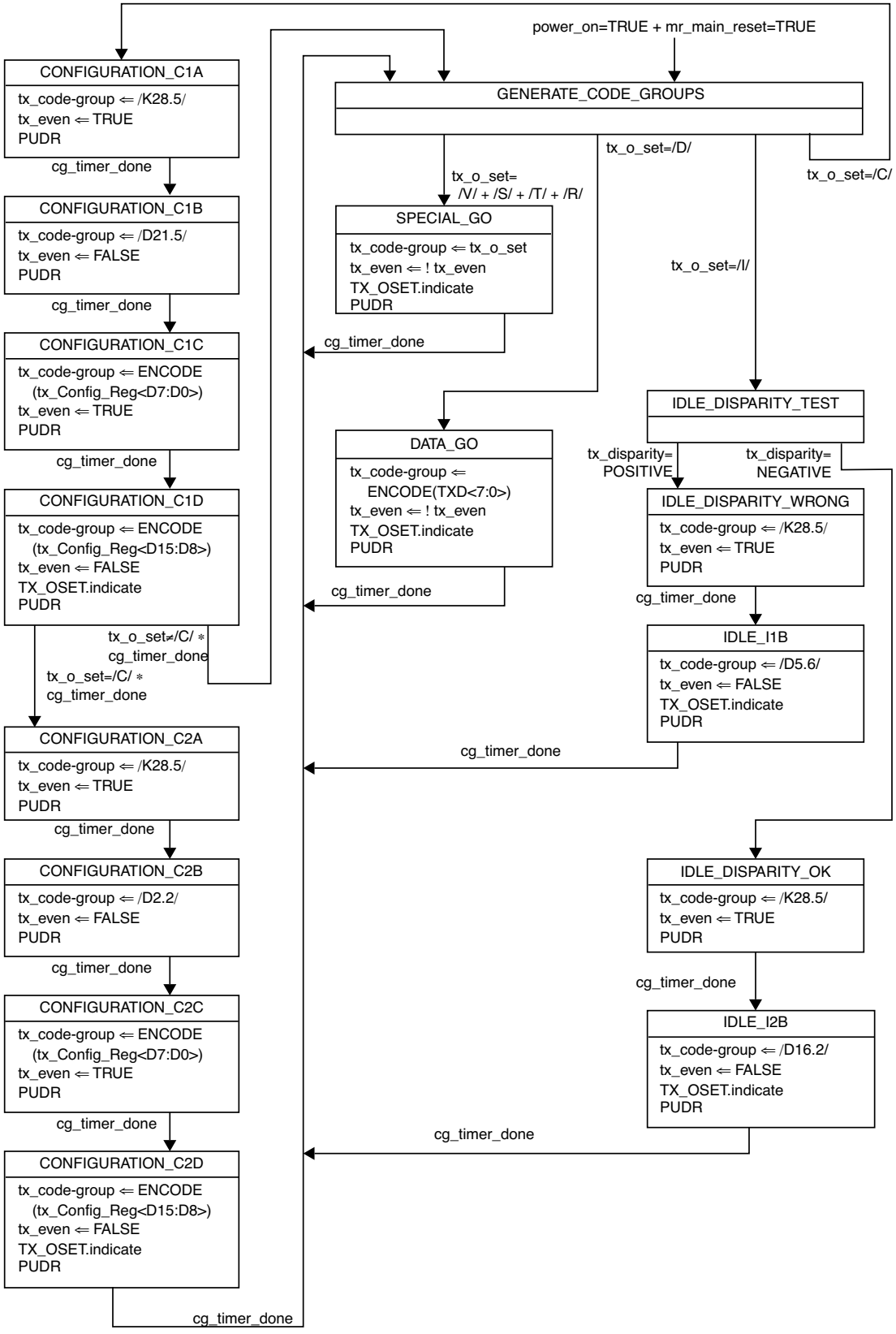
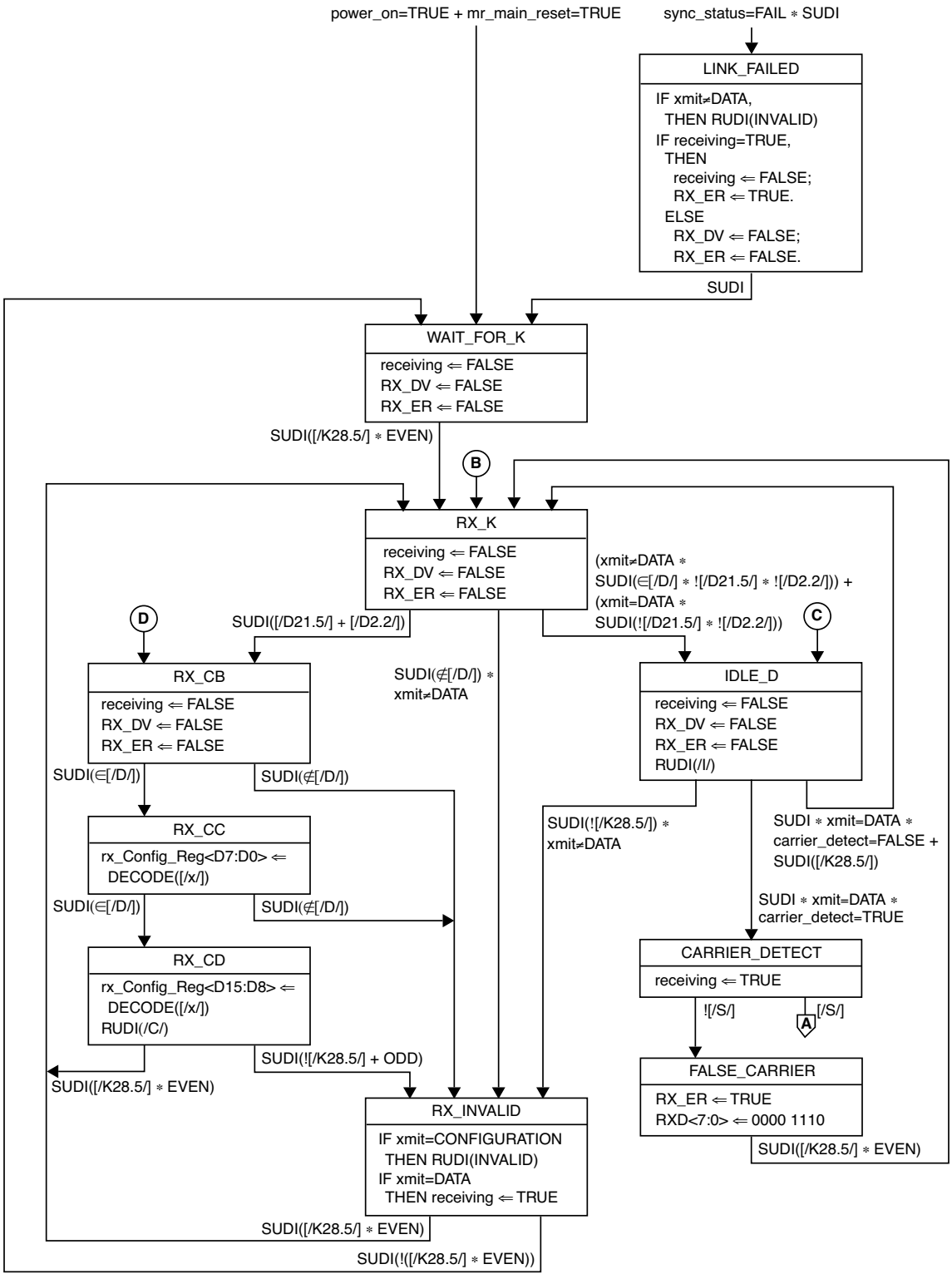
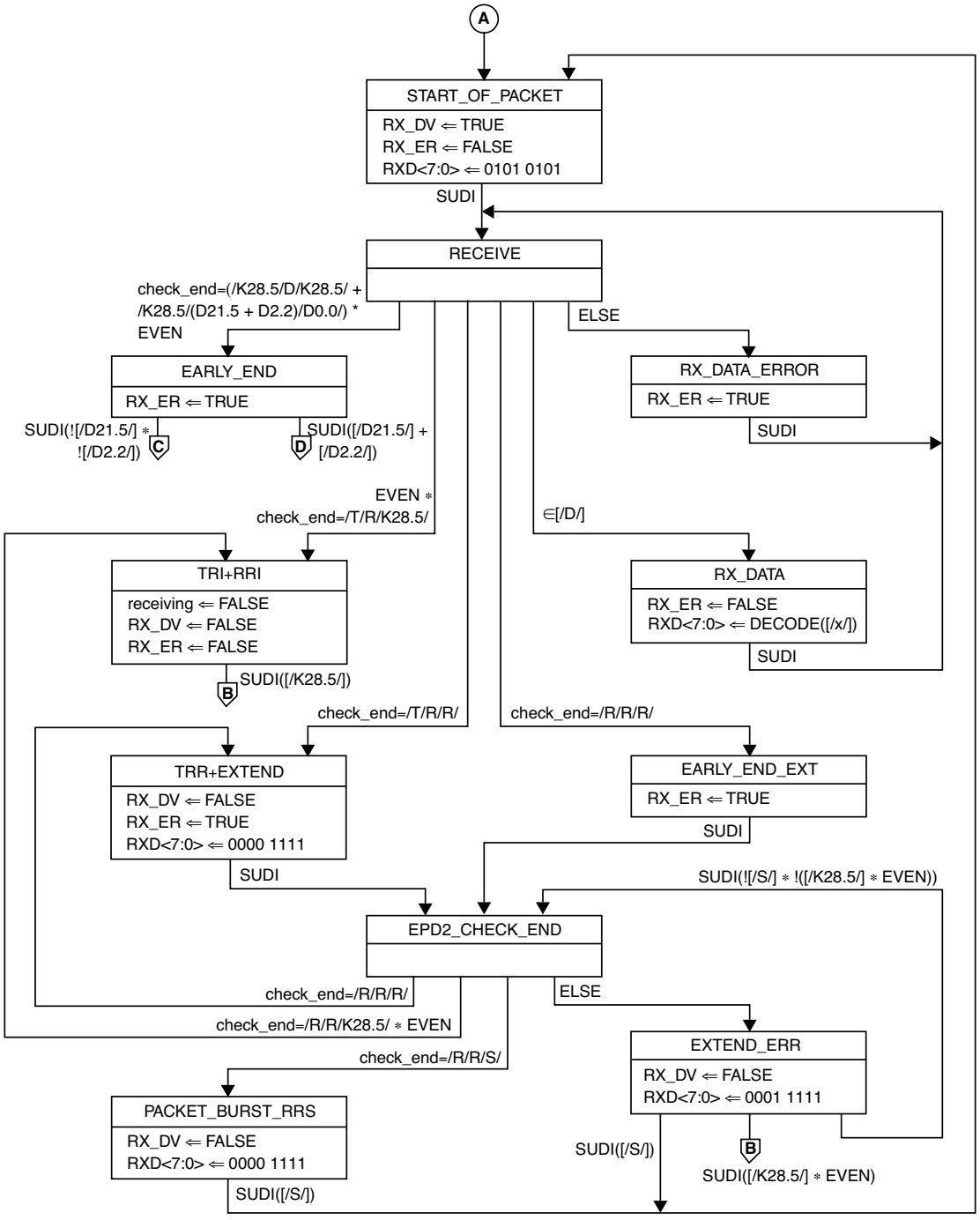


Figure 36–6—PCS transmit code-group state diagram



NOTE—Outgoing arcs leading to labeled polygons flow offpage to corresponding incoming arcs leading from labeled circles on Figure 36–7b, and vice versa.

Figure 36–7a—PCS receive state diagram, part a



NOTE—Outgoing arcs leading to labeled polygons flow offpage to corresponding incoming arcs leading from labeled circles on Figure 36–7a, and vice versa.

Figure 36–7b—PCS receive state diagram, part b

asserted, the TX_ER signal is asserted, the /V/ ordered_set is sourced except when the SPD ordered set is selected for sourcing.

Collision detection is implemented by noting the occurrence of carrier receptions during transmissions, following the models of 10BASE-T and 100BASE-X.

The Transmit code-group process continuously sources tx_code-group<9:0> to the PMA based on the ordered_sets sourced to it by the Transmit ordered_set process. The Transmit code-group process determines the proper code-group to source based on even/odd-numbered code-group alignment, running disparity requirements, and ordered_set format.

36.2.5.2.2 Receive

The PCS shall implement its Receive process as depicted in Figure 36–7a and Figure 36–7b, including compliance with the associated state variables as specified in 36.2.5.1.

The PCS Receive process continuously passes RXD<7:0> and sets the RX_DV and RX_ER signals to the GMII based on the received code-group from the PMA.

When the Auto-Negotiation process xmit flag indicates CONFIGURATION or IDLE, the PCS Receive process continuously passes /C/ and /I/ ordered sets and rx_Config_Reg<D15:D0> to the Auto-Negotiation process.

36.2.5.2.3 State variable function carrier_detect(x)

The detection of carrier on the underlying channel is used both by the MAC (via the GMII CRS signal and the Reconciliation sublayer) for deferral purposes, and by the PCS Transmit process for collision detection. A carrier event, signaled by the assertion of receiving, is indicated by the detection of a difference between the received code-group and /K28.5/ as specified in 36.2.5.1.4.

A carrier event is in error if it does not start with an SPD. The PCS Receive process performs this function by continuously monitoring incoming code-groups for specific patterns that indicate non-/I/ activity such as SPD. The detection of an SPD carrier event causes the PCS to substitute the value (01010101) for the SPD, set RXD<7:0> to this value, and assert RX_DV. The pattern substituted for the SPD is consistent with the preamble pattern expected by the MAC. The detection of a non-SPD carrier event (false carrier) causes the PCS to substitute the value (00001110) for the code-group received, set RXD<7:0> to this value, and assert RX_ER.

36.2.5.2.4 Code-group stream decoding

Subsequent to the detection of an SPD carrier event, the PCS Receive process performs the DECODE function on the incoming code-groups, passing decoded data to the GMII, including those corresponding to the remainder of the MAC preamble and SFD. The GMII signal RX_ER is asserted upon decoding any code-group following the SPD that neither is a valid /D/ code-group nor follows the EPD rules in 36.2.4.14.1.

Packets are terminated with an EPD as specified in 36.2.4.14. The PCS Receive process performs the check_end function to preserve the ability of the MAC to properly delimit the FCS at the end of a packet.

Detection of /T/R/R/ or /T/R/K28.5/ by the check_end function denotes normal (i.e. non-error) packet termination. Detection of /R/R/R/ by the check_end function denotes packet termination with error and Carrier_Extend processing. Detection of /K28.5/D/K28.5/ by the check_end function denotes packet termination with error. Detection of /K28.5/(D21.5 or D2.2)/D0.0 by the check_end function denotes packet termination with error.

36.2.5.2.5 Carrier sense

The Carrier Sense process generates the signal CRS on the GMII, which (via the Reconciliation sublayer) the MAC uses for deferral.

The PCS shall implement the Carrier Sense process as depicted in Figure 36–8 including compliance with the associated state variables as specified in 36.2.5.1.

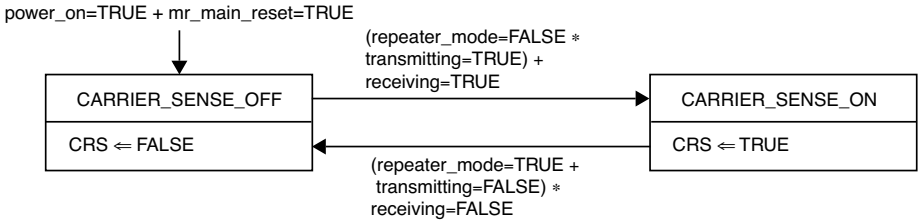


Figure 36–8—Carrier sense state diagram

36.2.5.2.6 Synchronization

The PCS shall implement the Synchronization process as depicted in Figure 36–9 including compliance with the associated state variables as specified in 36.2.5.1. The Synchronization process is responsible for determining whether the underlying receive channel is ready for operation. Failure of the underlying channel typically causes the PMA client to suspend normal actions.

A receiver that is in the LOSS_OF_SYNC state and that has acquired bit synchronization attempts to acquire code-group synchronization via the Synchronization process. Code-group synchronization is acquired by the detection of three ordered_sets containing commas in their leftmost bit positions without intervening invalid code-group errors. Upon acquisition of code-group synchronization, the receiver enters the SYNC_ACQUIRED_1 state. Acquisition of synchronization ensures the alignment of multi-code-group ordered_sets to even-numbered code-group boundaries.

Once synchronization is acquired, the Synchronization process tests received code-groups in sets of four code-groups and employs multiple sub-states, effecting hysteresis, to move between the SYNC_ACQUIRED_1 and LOSS_OF_SYNC states.

The condition sync_status=FAIL existing for ten ms or more causes the PCS Auto-Negotiation process to begin and the PCS Transmit process to begin transmission of /C/. Upon reception of three matching /C/s from the link partner, the PCS Auto-Negotiation process begins. The internal signal receiving is de-asserted in the PCS Receive process LINK_FAILED state when sync_status=FAIL and a code-group is received.

36.2.5.2.7 Auto-Negotiation process

The Auto-Negotiation process shall provide the means to exchange configuration information between two devices that share a link segment and to automatically configure both devices to take maximum advantage of their abilities. See Clause 37 for a description of the Auto-Negotiation process and Config_Reg contents.

Upon successful completion of the Auto-Negotiation process, the xmit flag is set to DATA and normal link operation is enabled. The Auto-Negotiation process utilizes the PCS Transmit and Receive processes to convey Config_Reg contents.

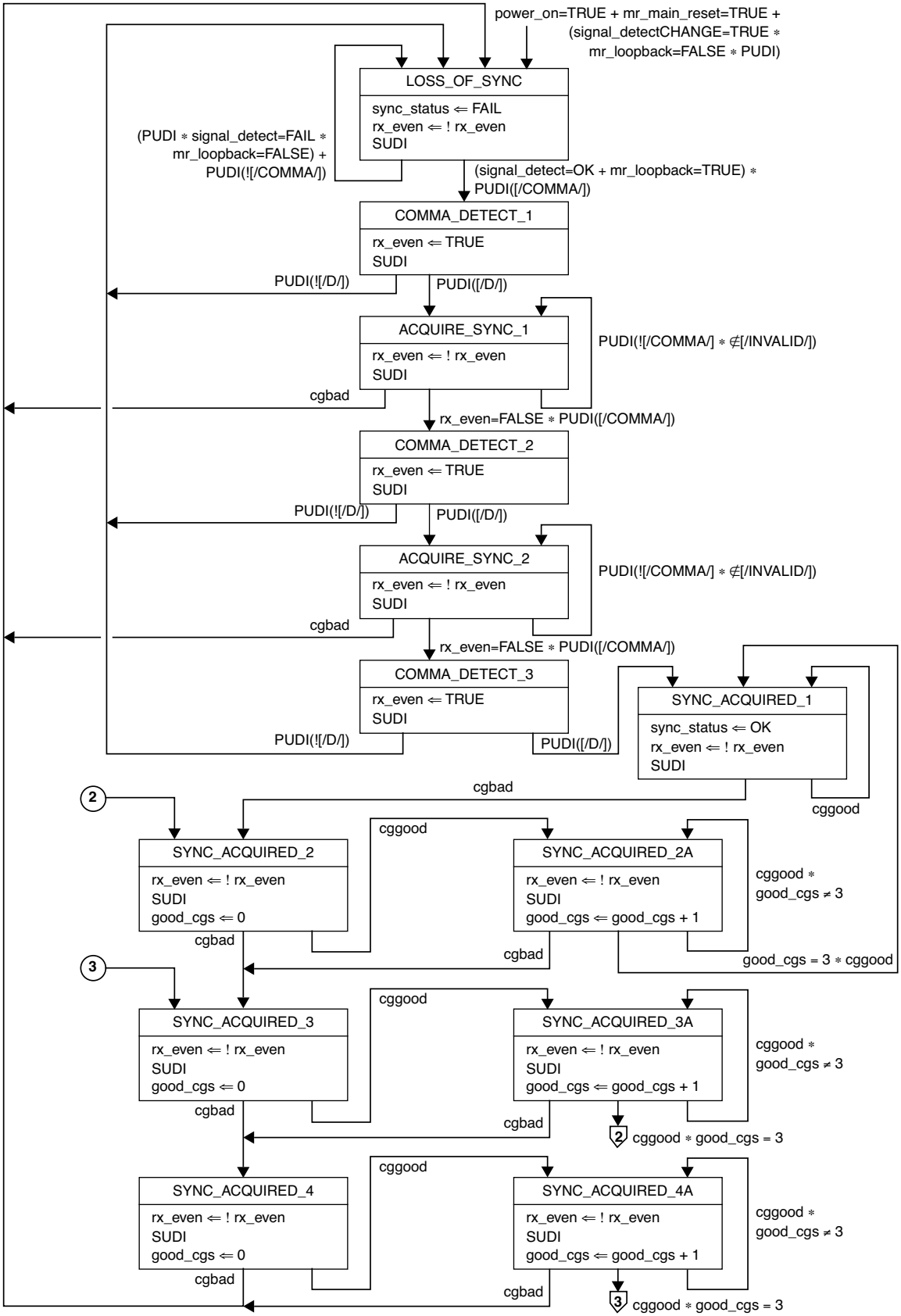


Figure 36–9—Synchronization state diagram

36.3 Physical Medium Attachment (PMA) sublayer

36.3.1 Service Interface

The PMA provides a Service Interface to the PCS. These services are described in an abstract manner and do not imply any particular implementation. The PMA Service Interface supports the exchange of code-groups between PCS entities. The PMA converts code-groups into bits and passes these to the PMD, and vice versa. It also generates an additional status indication for use by its client.

The following primitives are defined:

PMA_UNITDATA.request(tx_code-group<9:0>)

PMA_UNITDATA.indicate(rx_code-group<9:0>)

36.3.1.1 PMA_UNITDATA.request

This primitive defines the transfer of data (in the form of code-groups) from the PCS to the PMA. PMA_UNITDATA.request is generated by the PCS Transmit process.

36.3.1.1.1 Semantics of the service primitive

PMA_UNITDATA.request(tx_code-group<9:0>)

The data conveyed by PMA_UNITDATA.request is the tx_code-group<9:0> parameter defined in 36.2.5.1.3.

36.3.1.1.2 When generated

The PCS continuously sends, at a nominal rate of 125 MHz, as governed by GTX_CLK, tx_code-group<9:0> to the PMA.

36.3.1.1.3 Effect of receipt

Upon receipt of this primitive, the PMA generates a series of ten PMD_UNITDATA.request primitives, requesting transmission of the indicated tx_bit to the PMD.

36.3.1.2 PMA_UNITDATA.indicate

This primitive defines the transfer of data (in the form of code-groups) from the PMA to the PCS. PMA_UNITDATA.indicate is used by the PCS Synchronization process.

36.3.1.2.1 Semantics of the service primitive

PMA_UNITDATA.indicate(rx_code-group<9:0>)

The data conveyed by PMA_UNITDATA.indicate is the rx_code-group<9:0> parameter defined in 36.2.5.1.3.

36.3.1.2.2 When generated

The PMA continuously sends one rx_code-group<9:0> to the PCS corresponding to the receipt of each code-group aligned set of ten PMD_UNITDATA.indicate primitives received from the PMD. The nominal rate of the PMA_UNITDATA.indicate primitive is 125 MHz, as governed by the recovered bit clock.

36.3.1.2.3 Effect of receipt

The effect of receipt of this primitive by the client is unspecified by the PMA sublayer.

36.3.2 Functions within the PMA

Figure 36–3 depicts the mapping of the octet-wide data path of the GMII to the ten-bit-wide code-groups of the PMA Service Interface, and on to the serial PMD Service Interface. The PMA comprises the PMA Transmit and PMA Receive processes for 1000BASE-X.

The PMA Transmit process serializes tx_code-groups into tx_bits and passes them to the PMD for transmission on the underlying medium, according to Figure 36–3. Similarly, the PMA Receive process deserializes rx_bits received from the PMD according to Figure 36–3. The PMA continuously conveys ten-bit code-groups to the PCS, independent of code-group alignment. After code-group alignment is achieved, based on comma detection, the PCS converts code-groups into GMII data octets, according to 36.2.5.2.2.

The proper alignment of a comma used for code-group synchronization is depicted in Figure 36–3.

NOTE—Strict adherence to manufacturer-supplied guidelines for the operation and use of PMA serializer components is required to meet the jitter specifications of Clauses 38 and 39. The supplied guidelines should address the quality of power supply filtering associated with the transmit clock generator, and also the purity of the reference clock fed to the transmit clock generator.

36.3.2.1 Data delay

The PMA maps a nonaligned one-bit data path from the PMD to an aligned, ten-bit-wide data path to the PCS, on the receive side. Logically, received bits must be buffered to facilitate proper code-group alignment. These functions necessitate an internal PMA delay of at least ten bit times. In practice, code-group alignment may necessitate even longer delays of the incoming rx_bit stream.

36.3.2.2 PMA transmit function

The PMA Transmit function passes data unaltered (except for serializing) from the PCS directly to the PMD. Upon receipt of a PMA_UNITDATA.request primitive, the PMA Transmit function shall serialize the ten bits of the tx_code-group<9:0> parameter and transmit them to the PMD in the form of ten successive PMD_UNITDATA.request primitives, with tx_code-group<0> transmitted first, and tx_code-group<9> transmitted last.

36.3.2.3 PMA receive function

The PMA Receive function passes data unaltered (except for deserializing and possible code-group slipping upon code-group alignment) from the PMD directly to the PCS. Upon receipt of ten successive PMD_UNITDATA.indicate primitives, the PMA shall assemble the ten received rx_bits into a single ten-bit value and pass that value to the PCS as the rx_code-group<9:0> parameter of the primitive PMA_UNITDATA.indicate, with the first received bit installed in rx_code-group<0> and the last received bit installed in rx_code-group<9>. An exception to this operation is specified in 36.3.2.4.

36.3.2.4 Code-group alignment

In the event the PMA sublayer detects a comma+ within the incoming rx_bit stream, it may realign its current code-group boundary, if necessary, to that of the received comma+ as shown in Figure 36–3. This process is referred to in this document as code-group alignment. The code-group alignment function shall be operational when the EN_CDET signal is active (see 36.3.3.1). During the code-group alignment process, the PMA sublayer may delete or modify up to four, but shall delete or modify no more than four, ten-bit

code-groups in order to align the correct receive clock and code-group containing the comma+. This process is referred to as code-group slipping.

In addition, the PMA sublayer is permitted to realign the current code-group boundary upon receipt of a comma-pattern.

36.3.3 A physical instantiation of the PMA Service Interface

The ten-bit interface (TBI) is defined to provide compatibility among devices designed by different manufacturers. There is no requirement for a compliant device to implement or expose the TBI. A TBI implementation shall behave as described in 36.3.3 through 36.3.6.

Figure 36–10 illustrates the TBI functions and interfaces.

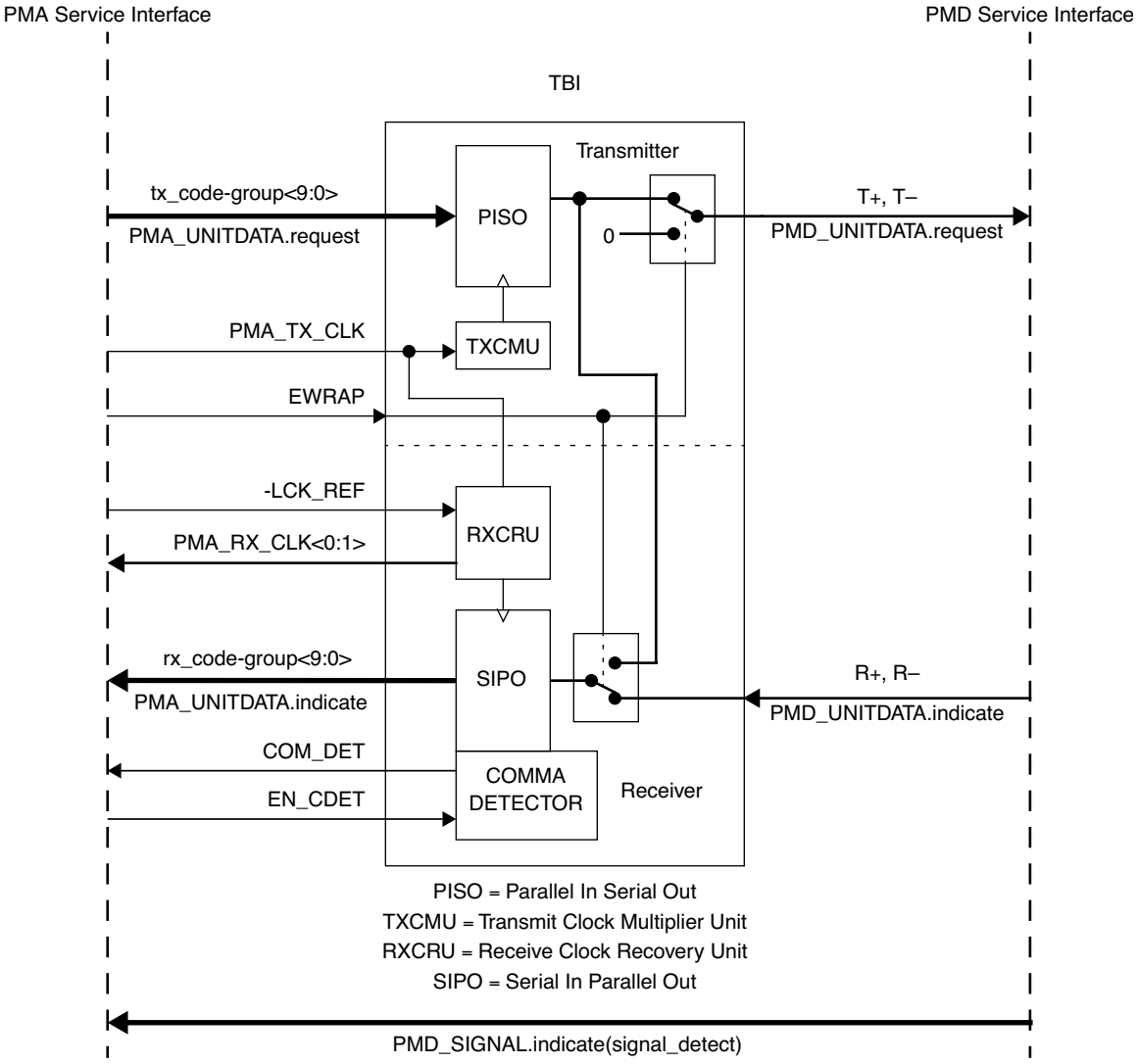


Figure 36–10—TBI reference diagram

As depicted in Figure 36–10, the TBI connects the PCS and PMD sublayers. It is equipped for full duplex transmission of code-groups at 125 MHz. The PCS provides code-groups on tx_code-group<9:0> to the

PMA transmit function, which latches the data on the rising edge of the 125 MHz PMA_TX_CLK. An internal Clock Multiplier Unit uses PMA_TX_CLK to generate the internal 1250 MHz bit clock that is used to serialize the latched data out of the PMA outputs, if EWRAP is Low, or internally loop it back to the Receive function input, if EWRAP is High.

The PMA Receive function accepts 1250 Mb/s serial data from either the PMD, if EWRAP is Low, or the PMA transmit function, if EWRAP is High, and extracts a bit clock and recovered data from the serial inputs in a clock recovery unit. The recovered data is deserialized and conveyed to the PCS on rx_code-group<9:0>. Two recovered clocks, PMA_RX_CLK<0> and PMA_RX_CLK<1>, which are at 1/20th the baud (62.5 MHz), and 180˚ out-of-phase with one another, are used by the PMA to latch the received 10-bit code-groups. Even and odd-numbered code-groups are latched on successive rising edges of PMA_RX_CLK<1> and PMA_RX_CLK<0>, respectively.

Code-group alignment occurs in the PMA Receive function, if enabled by EN_CDET, when a comma pattern occurs in the PHY bit stream. Upon recognition of the comma pattern, the PMA Receive function outputs the ten-bit code-group containing the comma on rx_code-group<9:0> with the alignment specified in Figure 36–3, and clocked on the rising edge of PMA_RX_CLK<1>.

This TBI provides a Lock_to_Reference_Clock (LCK_REF) input, which may be used to lock the clock recovery unit to PMA_TX_CLK rather than incoming serial data. In the absence of serial data or invalid serial data, the PMA Receive function passes many 8B/10B invalid code-groups across to the PCS. A circuit may be constructed to detect those errors and, using LCK_REF, re-center the receiver clock recovery unit to PMA_TX_CLK in preparation for reacquiring lock on the incoming PHY bit stream.

36.3.3.1 Required signals

In the event this TBI is made accessible, the signals listed in Table 36–4 are provided, with the meanings described elsewhere in this section. Note that not all of these signals are used by the PCS.

Table 36–4—TBI required signals

Symbol	Signal Name	Signal Type	Active Level
tx_code-group<9:0>	Transmit Data	Input	H
PMA_TX_CLK	Transmit Clock	Input	↑
EWRAP	Enable Wrap	Input	H
rx_code-group<9:0>	Receive Data	Output	H
PMA_RX_CLK<0>	Receive Clock 0	Output	↑
PMA_RX_CLK<1>	Receive Clock 1	Output	↑
COM_DET	Comma Detect	Output	H
-LCK_REF	Lock to Reference	Input	L
EN_CDET	Enable Comma Detect	Input	H

tx_code-group<9:0>
: The 10-bit parallel transmit data presented to the PMA for serialization and transmission onto the media. The order of transmission is tx_bit<0> first, followed by tx_bit<1> through tx_bit<9>.

PMA_TX_CLK
: The 125 MHz transmit code-group clock. This code-group clock is used to latch data into the PMA for transmission. PMA_TX_CLK is also used by the transmitter clock multiplier unit to generate the 1250 MHz bit rate clock. PMA_TX_CLK is also used by the receiver when -LCK_REF is active. PMA_TX_CLK has a ±100 ppm tolerance. PMA_TX_CLK is derived from GMII GTX_CLK.

EWRAP
: EWRAP enables the TBI to electrically loop transmit data to the receiver. The serial outputs on the transmitter are held in a static state during EWRAP operation. EWRAP may optionally be tied low (function disabled).

rx_code-group<9:0>
: Presents the 10-bit parallel receive code-group data to the PCS for further processing. When code-groups are properly aligned, any received code-group containing a comma is clocked by PMA_RX_CLK<1>.

PMA_RX_CLK<0>
: The 62.5 MHz receive clock that the protocol device uses to latch odd-numbered code-groups in the received PHY bit stream. This clock may be stretched during code-group alignment, and is not shortened.

PMA_RX_CLK<1>
: The 62.5 MHz receive clock that the protocol device uses to latch even-numbered code-groups in the received PHY bit stream. PMA_RX_CLK<1> is 180˚ out-of-phase with PMA_RX_CLK<0>. This clock may be stretched during code-group alignment, and is not shortened.

COM_DET
: An indication that the code-group associated with the current PMA_RX_CLK<1> contains a valid comma. When EN_CDET is asserted, the TBI is required to detect and code-group-align to the comma+ bit sequence. Optionally, the TBI may also detect and code-group-align to the comma- bit sequence. The TBI provides this signal as an output, but it may not be used by the PCS.

-LCK_REF
: Causes the TBI clock recovery unit to lock to PMA_TX_CLK. The TBI attains frequency lock within 500 ms. This function is not used by the PCS.

  NOTE—Implementors may find it necessary to use this signal in order to meet the clock recovery requirements of the PMA sublayer.

EN_CDET
: Enables the TBI to perform the code-group alignment function on a comma (see 36.2.4.9, 36.3.2.4). When EN_CDET is asserted the code-group alignment function is operational. This signal is optionally generated by the PMA client. The PMA sublayer may leave this function always enabled.

36.3.3.2 Summary of control signal usage

Table 36–5 lists all possible combinations of control signals on this TBI, including the valid combinations as well as the undefined combinations.

36.3.4 General electrical characteristics of the TBI

In the event this TBI is made accessible, the following subclauses specify the general electrical characteristics of the TBI.

Table 36–5—TBI combinations of control signals

EWRAP	-LCK_REF	EN_CDET	Interpretation
L	L	L	Undefined
L	L	H	Lock receiver clock recovery unit toPMA_TX_CLK
L	H	L	Normal operation; COM_DET disabled
L	H	H	Normal operation; COM_DET enabled
H	L	L	Undefined
H	L	H	Undefined
H	H	L	Loop transmit data to receiver; COM_DET disabled
H	H	H	Loop transmit data to receiver; COM_DET enabled

36.3.4.1 DC characteristics

Table 36–6 documents the required dc parametric attributes required of all inputs to the TBI and the dc parametric attributes associated with the outputs of the TBI. The inputs levels to the TBI may be greater than the power supply level (i.e., 5 V output driving V_{OH} into a 3.3 V input), tolerance to mismatched input levels is optional. TBI devices not tolerant of mismatched inputs levels that meet Table 36–6 requirements are still regarded as compliant.

Table 36–6—DC specifications

Symbol	Parameter	Conditions		Min	Typ	Max	Units
V_{OH}	Output High Voltage	I_{OH} = –400 μA	V_{CC} = Min	2.2	3.0	V_{CC}	V
V_{OL}	Output Low Voltage	I_{OL} = 1 mA	V_{CC} = Min	GND	0.25	0.6	V
V_{IH}	Input High Voltage			2.0	—	V_{CC}[a] + 10%	V
V_{IL}	Input Low Voltage			GND	—	0.8	V
I_{IH}	Input High Current	V_{CC} = Max	V_{IN} = 2.4 V	—	—	40	µA
I_{IL}	Input Low Current	V_{CC} = Max	V_{IN} = 0.4 V	—	—	600	µA
C_{IN}	Input Capacitance			—	—	4.0	pf
t_R	Clock Rise Time	0.8 V to 2.0 V		0.7	—	2.4	ns
t_F	Clock Fall Time	2.0 V to 0.8 V		0.7	—	2.4	ns
t_R	Data Rise Time	0.8 V to 2.0 V		0.7	—	—	ns
t_F	Data Fall Time	2.0 V to 0.8 V		0.7	—	—	ns

[a]Refers to the driving device power supply.

36.3.4.2 Valid signal levels

All ac measurements are made from the 1.4 V level of the clock to the valid input or output data levels as shown in Figure 36–11.

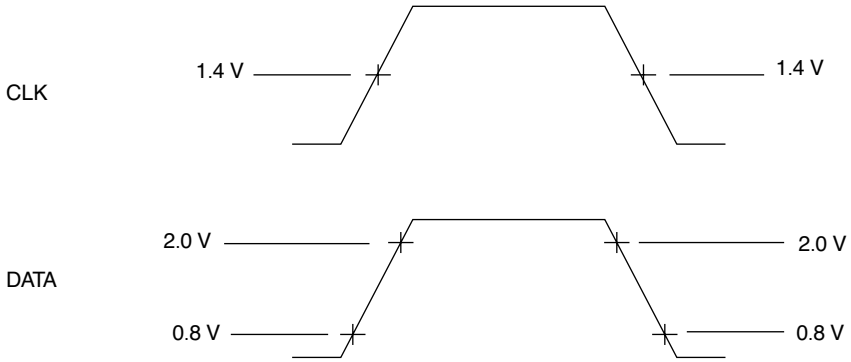


Figure 36–11—Input/output valid level for ac measurements

36.3.4.3 Rise and fall time definition

The rise and fall time definition for PMA_TX-CLK, PMA_RX_CLK<0>, PMA_RX_CLK<1>, and DATA is shown in Figure 36–12.

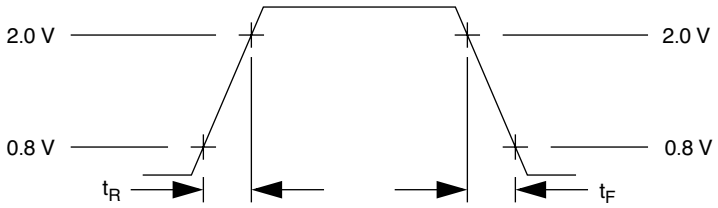


Figure 36–12—Rise and fall time definition

36.3.4.4 Output load

All ac measurements are assumed to have the output load of 10 pF.

36.3.5 TBI transmit interface electrical characteristics

In the event this TBI is made accessible, the electrical characteristics of the TBI transmit interface are specified in the following subclauses.

36.3.5.1 Transmit data (tx_code-group<9:0>)

The tx_code-group<9:0> signals carry data from the PCS to PMA to be serialized to the PMD in accordance with the transmission order shown in Figure 36–3. All tx_code-group<9:0> data conforms to valid code-groups.

36.3.5.2 TBI transmit interface timing

The TBI transmit interface timings in Table 36–7 defines the TBI input. All transitions in Figure 36–13 are specified from the PMA_TX_CLK reference level (1.4 V), to valid input signal levels.

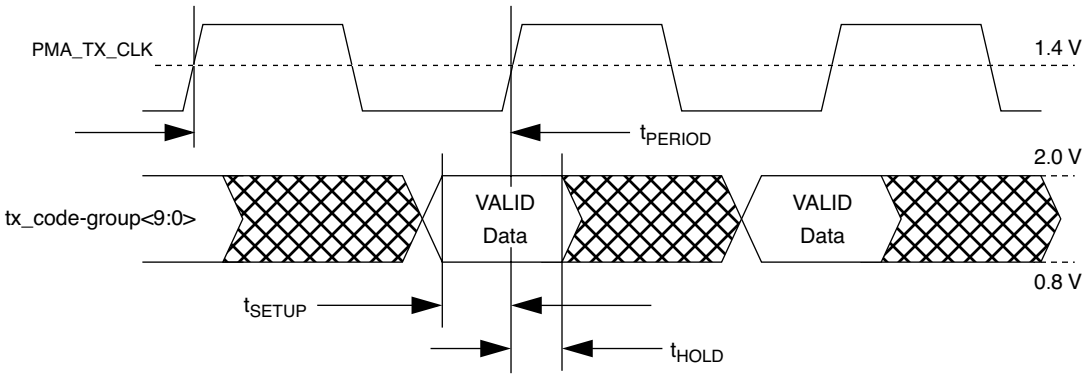


Figure 36–13—TBI transmit interface timing diagram

Table 36–7—Transmit ac specification

Parameter	Description	Min	Typ	Max	Units
t_{PERIOD}	PMA_TX_CLK Period[a]	—	8	—	ns
t_{SETUP}	Data Setup to ↑ PMA_TX_CLK	2.0	—	—	ns
t_{HOLD}	Data Hold from ↑ PMA_TX_CLK	1.0	—	—	ns
t_{DUTY}	PMA_TX_CLK Duty Cycle	40		60	%

[a] ±100 ppm tolerance on PMA_TX_CLK frequency.

36.3.6 TBI receive interface electrical characteristics

In the event this TBI is made accessible, the electrical characteristics of the TBI receive interface are specified in the following subclauses.

The TBI receive interface timings in Table 36–8 define the TBI output. All transitions in Figure 36–14 are specified from the Receive Clock reference level (1.4 V) to valid output signal levels.

36.3.6.1 Receive data (rx_code-group<9:0>)

The 10 receive data signals rx_code-group<9:0> carry parallel data from the PMA sublayer to the PCS sublayer during the rising edge of the receive clock (i.e., PMA_RX_CLK<1> transitions from Low to High). When properly locked and aligned, data transferred across this interface conforms to valid code-groups.

36.3.6.2 Receive clock (PMA_RX_CLK<0>, PMA_RX_CLK<1>)

The receive clocks supplied to the PCS and GMII are derived from the recovered bit clock. PMA_RX_CLK<0> is 180˚ out-of-phase with PMA_RX_CLK<1>.

Table 36–8 specifies a receive clock drift (t_{DRIFT}), which is applicable under all input conditions to the receiver (including invalid or absent input signals). However, the restriction does not apply when the receiver is realigning to a new code-group boundary and the receive clocks are being stretched to a new code-group boundary to avoid short pulses. During the code-group alignment process the receive clocks may slow a fixed amount, depending on the bit offset of the new comma and then return to the nominal frequency.

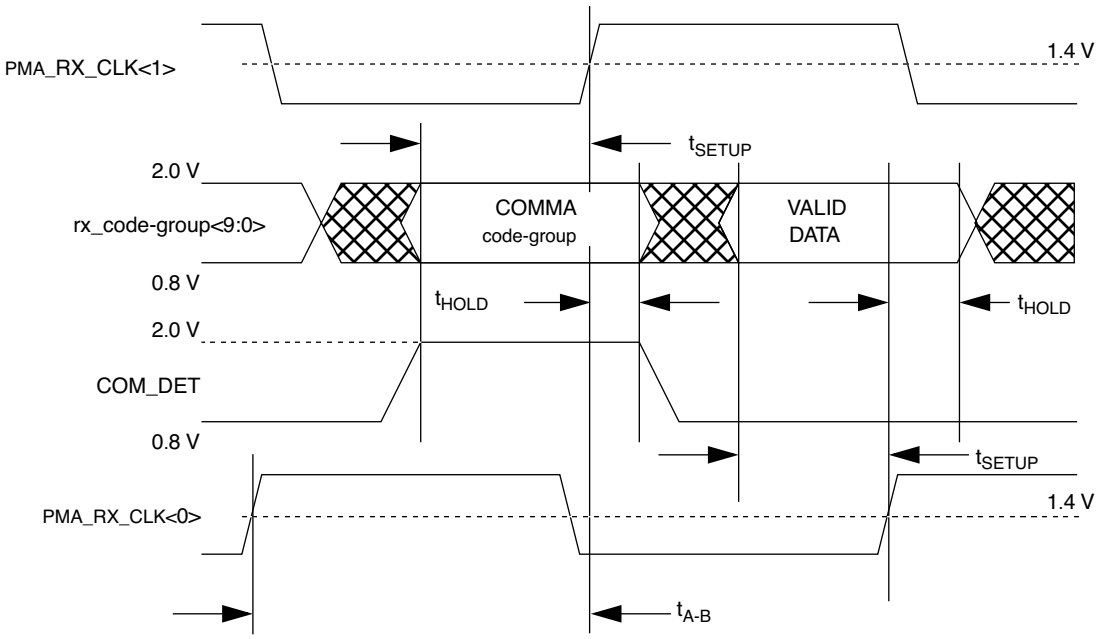


Figure 36–14—TBI receive interface timing diagram

Table 36–8—Receive bus ac specification

Parameter	Description	Min	Typ	Max	Units
t_{FREQ}	PMA_RX_CLK Frequency	—	62.5	—	MHz
t_{DRIFT}	PMA_RX_CLK Drift Rate[a]	0.2	—	—	μs/MHz
t_{SETUP}	Data Setup Before ↑PMA_RX_CLK	2.5	—	—	ns
t_{HOLD}	Data Hold After ↑PMA_RX_CLK	1.5	—	—	ns
t_{DUTY}	PMA_RX_CLK Duty Cycle	40	—	60	%
t_{A-B}	PMA_RX_CLK Skew	7.5	—	8.5	ns

[a]t_{DRIFT} is the (minimum) time for PMA_RX_CLK to drift from 63.5 MHz to 64.5 MHz or 60 MHz to 59 MHz from the PMA_RX_CLK lock value. It is applicable under all input signal conditions (except where noted in 36.3.2.4), including invalid or absent input signals, provided that the receiver clock recovery unit was previously locked to PMA_TX_CLK or to a valid input signal.

36.3.7 Loopback mode

Loopback mode shall be provided, as specified in this subclause, by the transmitter and receiver of a device as a test function to the device. When Loopback mode is selected, transmission requests passed to the transmitter are shunted directly to the receiver, overriding any signal detected by the receiver on its attached link. A device is explicitly placed in Loopback mode (i.e., Loopback mode is not the normal mode of operation of a device). The method of implementing Loopback mode is not defined by this standard.

NOTE—Loopback mode may be implemented either in the parallel or the serial circuitry of a device.

36.3.7.1 Receiver considerations

A receiver may be placed in Loopback mode. Entry into or exit from Loopback mode may result in a temporary loss of synchronization.

36.3.7.2 Transmitter considerations

A transmitter may be placed in Loopback mode. The external behavior of a transmitter (i.e., the activity of a transmitter with respect to its attached link) in Loopback mode is specified in 22.2.4.1.2.

36.3.8 Test functions

A limited set of test functions may be provided as an implementation option for testing of the transmitter function.

Some test functions that are not defined by this standard may be provided by certain implementations. Compliance with the standard is not affected by the provision or exclusion of such functions by an implementation. Random jitter test patterns for 1000BASE-X are specified in Annex 36A.

A typical test function is the ability to transmit invalid code-groups within an otherwise valid PHY bit stream. Certain invalid PHY bit streams may cause a receiver to lose word and/or bit synchronization. See ANSI X3.230-1994 [B20] (FC-PH), subclause 5.4, for a more detailed discussion of receiver and transmitter behavior under various test conditions.

36.4 Compatibility considerations

There is no requirement for a compliant device to implement or expose any of the interfaces specified for the PCS or PMA. Implementations of a GMII shall comply with the requirements as specified in Clause 35. Implementations of a TBI shall comply with the requirements as specified in 36.3.3.

36.5 Delay constraints

In half duplex mode, proper operation of a CSMA/CD LAN demands that there be an upper bound on the propagation delays through the network. This implies that MAC, PHY, and repeater implementors must conform to certain delay minima and maxima, and that network planners and administrators conform to constraints regarding the cable topology and concatenation of devices. MAC constraints are contained in 35.2.4 and Table 36–5. Topological constraints are contained in Clause 42.

In full duplex mode, predictable operation of the MAC Control PAUSE operation (Clause 31, Annex 31B) also demands that there be an upper bound on the propagation delays through the network. This implies that MAC, MAC Control sublayer, and PHY implementors must conform to certain delay maxima, and that network planners and administrators conform to constraints regarding the cable topology and concatenation of devices.

The reference point for all MDI measurements is the 50% point of the mid-cell transition corresponding to the reference bit, as measured at the MDI.

36.5.1 MDI to GMII delay constraints

Every 1000BASE-X PHY associated with a GMII shall comply with the bit time delay constraints specified in Table 36–9a for half duplex operation and Table 36–9b for full duplex operation. These figures apply for all 1000BASE-X PMDs. For any given implementation, the assertion and deassertion delays on CRS shall be equal.

Table 36–9a—MDI to GMII delay constraints (half duplex mode)

Sublayer measurement points	Event	Min (bit time)	Max (bit time)	Input timing reference	Output timing reference
GMII ⇔ MDI	TX_EN=1 sampled to MDI output	—	136	PMA_TX_CLK rising	1st bit of /S/
	MDI input to CRS assert	—	192	1st bit of /S/	
	MDI input to CRS de-assert	—	192	1st bit of /K28.5/	
	MDI input to COL assert	—	192	1st bit of /S/	
	MDI input to COL de-assert	—	192	1st bit of /K28.5/	
	TX_EN=1 sampled to CRS assert	—	16	PMA_TX_CLK rising	
	TX_EN=0 sampled to CRS de-assert	—	16	PMA_TX_CLK rising	

Table 36–9b—MDI to GMII delay constraints (full duplex mode)

Sublayer measurement points	Event	Min (bit time)	Max (bit time)	Input timing reference	Output timing reference
GMII ⇔ MDI	TX_EN=1 sampled to MDI output	—	136	PMA_TX_CLK rising	1st bit of /S/
	MDI input to RX_DV de-assert	—	192	1st bit of /T/	RX_CLK rising

36.5.2 DTE delay constraints (half duplex mode)

Every DTE with a 1000BASE-X PHY shall comply with the bit time delay constraints specified in Table 36–10 for half duplex operation. These figures apply for all 1000BASE-X PMDs.

Table 36–10—DTE delay constraints (half duplex mode)

Sublayer measurement points	Event	Min (bit time)	Max (bit time)	Input timing reference	Output timing reference
MAC ⇔ MDI	MAC transmit start to MDI output	—	184		1st bit of /S/
	MDI input to MDI output (worst-case nondeferred transmit)	—	440	1st bit of /S/	1st bit of /S/
	MDI input to collision detect	—	240	1st bit of /S/	
	MDI input to MDI output = Jam (worst-case collision response)	—	440	1st bit of /S/	1st bit of jam

36.5.3 Carrier de-assertion/assertion constraint (half duplex mode)

To ensure fair access to the network, each DTE operating in half duplex mode shall, additionally, satisfy the following:

(MAX MDI to MAC Carrier De-assert Detect) – (MIN MDI to MAC Carrier Assert Detect) < 16 bits

36.6 Environmental specifications

All equipment subject to this clause shall conform to the requirements of 14.7 and applicable sections of ISO/IEC 11801: 1995.

36.7 Protocol Implementation Conformance Statement (PICS) proforma for Clause 36, Physical Coding Sublayer (PCS) and Physical Medium Attachment (PMA) sublayer, type 1000BASE-X[3]

36.7.1 Introduction

The supplier of a protocol implementation that is claimed to conform to Clause 36, Physical Coding Sublayer (PCS) and Physical Medium Attachment (PMA) sublayer, type 1000BASE-X, shall complete the following Protocol Implementation Conformance Statement (PICS) proforma. A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

36.7.2 Identification

36.7.2.1 Implementation identification

Supplier (Note 1)	
Contact point for enquiries about the PICS (Note 1)	
Implementation Name(s) and Version(s) (Notes 1 and 3)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Names(s) (Note 2)	
NOTE 1—Required for all implementations. NOTE 2—May be completed as appropriate in meeting the requirements for the identification. NOTE 3—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

[3]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so that it can be used for its intended purpose and may further publish the completed PICS.

36.7.2.2 Protocol summary

Identification of protocol standard	IEEE Std 802.3-2002®, Clause 36, Physical Coding Sublayer (PCS) and Physical Medium Attachment (PMA) sublayer, type 1000BASE-X
Identification of amendments and corrigenda to this PICS proforma that have been completed as part of this PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21; the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®.)	

Date of Statement	

36.7.3 Major Capabilities/Options

Item	Feature	Subclause	Value/Comment	Status	Support
*PMA	Ten-bit interface (TBI)	36.4		O	Yes [ ] No [ ]
*GMII	PHY associated with GMII	36.4		O	Yes [ ] No [ ]
*DTE	DTE with PHY not associated with GMII	36.5.2		O	Yes [ ] No [ ]
*FDX	PHY supports full duplex mode	36.5		O	Yes [ ] No [ ]
*HDX	PHY supports half duplex mode	36.5		O	Yes [ ] No [ ]
NOTE—The following abbreviations are used: *HDGM: HDX and GMII *FDGM: FDX and GMII *HDTE: HDX and DTE					

36.7.4 PICS proforma tables for the PCS and PMA sublayer, type 1000BASE-X

36.7.4.1 Compatibility considerations

Item	Feature	Subclause	Value/Comment	Status	Support
CC1	Test functions Annex 36A support	36.3.8		O	Yes [ ] No [ ]
CC2	Environmental specifications	36.6		M	Yes [ ]

36.7.4.2 Code-group functions

Item	Feature	Subclause	Value/Comment	Status	Support
CG1	Transmitter initial running disparity	36.2.4.4	Transmitter initial running disparity assumes negative value	M	Yes [ ]
CG2	Transmitter running disparity calculation	36.2.4.4	Running disparity is calculated after each code-group transmitted	M	Yes [ ]
CG3	Validating received code-groups	36.2.4.6		M	Yes [ ]
CG4	Running disparity rules	36.2.4.4	Running disparity is calculated after each code-group reception	M	Yes [ ]
CG5	Transmitted code-group is chosen from the corresponding running disparity	36.2.4.5		M	Yes [ ]

36.7.4.3 State diagrams

Item	Feature	Subclause	Value/Comment	Status	Support
SD1	Transmit ordered_set	36.2.5.2.1	Meets the requirements of Figure 36–5	M	Yes [ ]
SD2	Transmit code-group	36.2.5.2.1	Meets the requirements of Figure 36–6	M	Yes [ ]
SD3	Receive	36.2.5.2.2	Meets the requirements of Figures 36–7a and 36–7b	M	Yes [ ]
SD4	Carrier sense	36.2.5.2.5	Meets the requirements of Figure 36–8	M	Yes [ ]
SD5	Synchronization	36.2.5.2.6	Meets the requirements of Figure 36–9	M	Yes [ ]
SD6	Auto-Negotiation	36.2.5.2.7	Described in Clause 37	M	Yes [ ]

36.7.4.4 PMA functions

Item	Feature	Subclause	Value/Comment	Status	Support
PMA1	Transmit function	36.3.2.2		M	Yes [ ]
PMA2	Receive function	36.3.2.3		M	Yes [ ]
PMA3	Code-group alignment	36.3.2.4	When EN_CDET is active	M	Yes [ ]
PMA4	Loopback mode	36.3.7		M	Yes [ ]

36.7.4.5 PMA transmit function

Item	Feature	Subclause	Value/Comment	Status	Support
PMT1	cg_timer expiration	36.2.5.1.7	See 35.4.2.3	GMII:M	Yes [ ] N/A [ ]
PMT2	cg_timer expiration	36.2.5.1.7	8 ns ± 0.01%	!GMII: M	Yes [ ] N/A [ ]

36.7.4.6 PMA code-group alignment function

Item	Feature	Subclause	Value/Comment	Status	Support
CDT1	Code-group alignment to comma-	36.3.2.4		O	Yes [ ] N/A [ ]
CDT2	Code-group slipping limit	36.3.2.4	Deletion or modification of no more than four code-groups	M	Yes [ ]
CDT3	Code-group alignment to comma+	36.3.2.4		O	Yes [ ] N/A [ ]

36.7.4.7 TBI

Item	Feature	Subclause	Value/Comment	Status	Support
TBI1	TBI requirements	36.3.3		PMA:M	Yes [ ] N/A [ ]

36.7.4.8 Delay constraints

Item	Feature	Subclause	Value/Comment	Status	Support
TIM1	Equal carrier de-assertion and assertion delay on CRS	36.5.1		HDGM:M	Yes [ ] N/A [ ]
TIM2	MDI to GMII delay constraints for half duplex	36.5.1	Table 36–9a	HDGM:M	Yes [ ] N/A [ ]
TIM3	MDI to GMII delay constraints for full duplex	36.5.1	Table 36–9b	FDGM:M	Yes [ ] N/A [ ]
TIM4	DTE delay constraints for half duplex	36.5.2	Table 36–10	HDTE:M	Yes [ ] N/A [ ]
TIM5	Carrier de-assertion/assertion constraints	36.5.3		HDTE:M	Yes [ ] N/A [ ]

37. Auto-Negotiation function, type 1000BASE-X

37.1 Overview

37.1.1 Scope

Clause 37 describes the 1000BASE-X Auto-Negotiation (AN) function that allows a device (local device) to advertise modes of operation it possesses to a device at the remote end of a link segment (link partner) and to detect corresponding operational modes that the link partner may be advertising.

The Auto-Negotiation function exchanges information between two devices that share a link segment and automatically configures both devices to take maximum advantage of their abilities. Auto-Negotiation is performed with /C/ and /I/ ordered_sets defined in Clause 36, such that no packet or upper layer protocol overhead is added to the network devices. Auto-Negotiation does not test the link segment characteristics (see 37.1.4).

The function allows the devices at both ends of a link segment to advertise abilities, acknowledge receipt and understanding of the common mode(s) of operation that both devices share, and to reject the use of operational modes that are not shared by both devices. Where more than one common mode exists between the two devices, a mechanism is provided to allow the devices to resolve to a single mode of operation using a predetermined priority resolution function (see 37.2.4.2). The Auto-Negotiation function allows the devices to switch between the various operational modes in an ordered fashion, permits management to disable or enable the Auto-Negotiation function, and allows management to select a specific operational mode.

The basic mechanism to achieve Auto-Negotiation is to pass information encapsulated within /C/ ordered_sets. /C/ ordered_sets are directly analogous to FLP Bursts as defined in Clause 28 that accomplish the same function. Each device issues /C/ ordered_sets at power up, on command from management, upon detection of a PHY error, or due to user interaction.

37.1.2 Application perspective/objectives

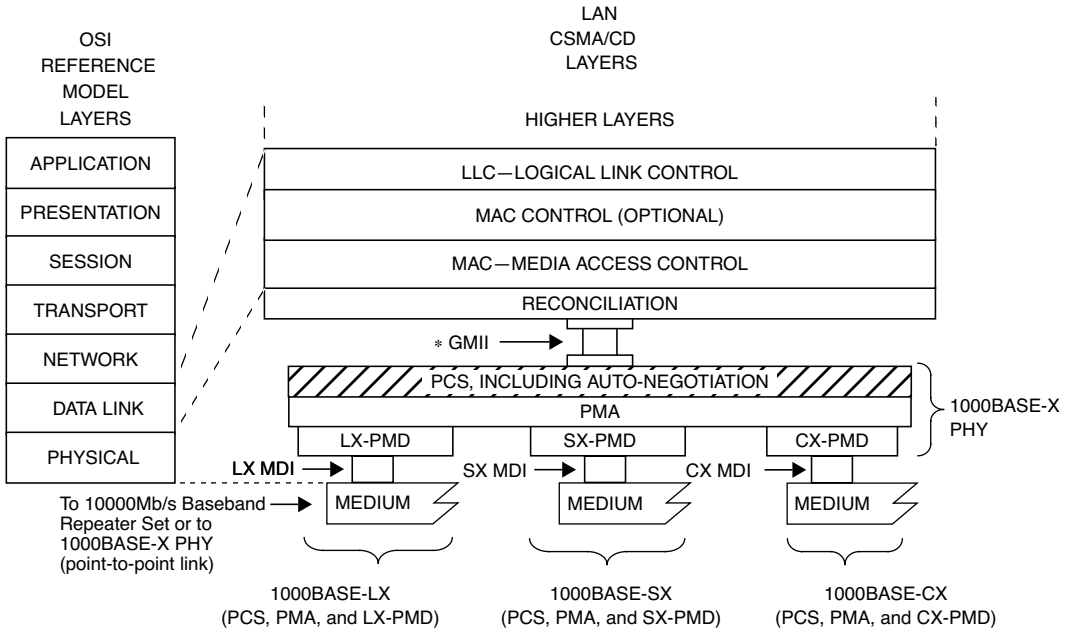
This Auto-Negotiation function is designed to be expandable and allows 1000BASE-X devices to self-configure a jointly compatible operating mode.

The following are the objectives of Auto-Negotiation:

a) To be reasonable and cost-effective to implement;
b) Must provide a sufficiently extensible code space to
    3) Meet existing and future requirements;
    4) Allow simple extension without impacting the installed base;
    5) Accommodate remote fault signals; and
    6) Accommodate link partner ability detection.
c) Must allow manual or Network Management configuration to override the Auto-Negotiation;
d) Must be capable of operation in the absence of Network Management;
e) Must allow the ability to renegotiate;
f) Must operate when
    1) The link is initially connected;
    2) A device at either end of the link is powered up, reset, or a renegotiation request is made.
g) May be enabled by automatic, manual, or Network Management intervention;
h) To complete the base page Auto-Negotiation function in a bounded time period;
i) To operate using a peer-to-peer exchange of information with no requirement for a master device (not master-slave);
j) Must be robust in the 1000BASE-X MDI cable noise environment;
k) Must not significantly impact EMI/RFI emissions.

37.1.3 Relationship to ISO/IEC 8802-3

The Auto-Negotiation function is provided at the PCS sublayer of the Physical Layer of the OSI reference model as shown in Figure 36–1. Devices that support multiple modes of operation may advertise this fact using this function. The transfer of information is observable only at the MDI or on the medium.



MDI=MEDIUM DEPENDENT INTERFACE
GMII=GIGABIT MEDIA INDEPENDENT INTERFACE
PCS=PHYSICAL CODING SUBLAYER
PMA=PHYSICAL MEDIUM ATTACHMENT
PHY=PHYSICAL LAYER DEVICE
PMD=PHYSICAL MEDIUM DEPENDENT
LX-PMD=PMD FOR FIBER—LONG WAVELENGTH, Clause 38
SX-PMD=PMD FOR FIBER—SHORT WAVELENGTH, Clause 38
CX-PMD=PMD FOR 150 Ω BALANCED COPPER CABLING, Clause 39

NOTE—The PMD sublayers are mutually independent.
* GMII is optional.

Figure 37–1—Location of the Auto-Negotiation function

37.1.4 Compatibility considerations

37.1.4.1 Auto-Negotiation

1000BASE-X devices provide the Auto-Negotiation function. Auto-Negotiation does not perform cable tests, such as cable performance measurements. Some PHYs that explicitly require use of high-performance cables, may require knowledge of the cable type, or additional robustness tests (such as monitoring invalid code-groups, CRC, or framing errors) to determine if the link segment is adequate.

37.1.4.2 Management interface

Manual or automatic invocation of Auto-Negotiation may result in frame loss. Exit from Auto-Negotiation to normal MAC frame processing may also result in frame loss as one link end may resume normal MAC frame processing ahead of its link partner.

37.1.4.2.1 GMII management interface

Auto-Negotiation signaling does not occur across the GMII. Control of the Auto-Negotiation function may be supported through the Management Interface of the GMII or equivalent. If an explicit embodiment of the GMII is supported, the Control and Status registers to support the Auto-Negotiation function shall be implemented in accordance with the definitions in Clause 22 and 37.2.5.

37.1.4.3 Interoperability between Auto-Negotiation compatible devices

An Auto-Negotiation compatible device decodes the base page from the received /C/ ordered_sets and examines the contents for the highest common ability that both devices share. Both devices acknowledge correct receipt of each other's base page by responding with base pages containing the Acknowledge Bit set. After both devices complete acknowledgment, and any desired next page exchange, both devices enable the highest common mode negotiated. The highest common mode is resolved using the priority resolution hierarchy specified in 37.2.4.2.

37.1.4.4 User Configuration with Auto-Negotiation

Rather than disabling Auto-Negotiation, the following behavior is suggested in order to improve interoperability with other Auto-Negotiation devices. When a device is configured for one specific mode of operation (e.g. 1000BASE-X Full Duplex), it is recommended to continue using Auto-Negotiation but only advertise the specifically selected ability or abilities. This can be done by the Management agent only setting the bits in the advertisement registers that correspond to the selected abilities.

37.2 Functional specifications

The Auto-Negotiation function includes the Auto-Negotiation Transmit, Receive, and Arbitration functions specified in the state diagram of Figure 37–6 and utilizes the PCS Transmit and Receive state diagrams of Clause 36.

The Auto-Negotiation function provides an optional Management function that provides a control and status mechanism. Management may provide additional control of Auto-Negotiation through the Management function, but the presence of a management agent is not required.

37.2.1 Config_Reg encoding

The Config_Reg base page, transmitted by a local device or received from a link partner, is encapsulated within a /C/ ordered_set and shall convey the encoding shown in Figure 37–2. Auto-Negotiation supports additional pages using the Next Page function. Encodings for the Config_Reg(s) used in next page exchange are defined in 37.2.4.3.1. Config_Reg bits labeled as "rsvd" are reserved and shall be set to a logic zero.

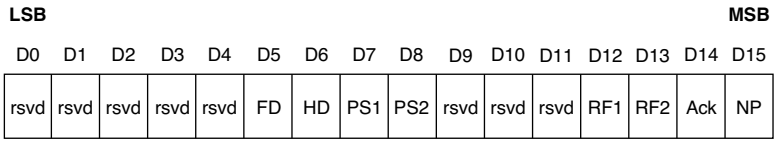


Figure 37–2—Config_Reg base page encoding

37.2.1.1 Base page to management register mapping

Several base page bits shown in Figure 37–2 indicate capabilities that are sourced from management registers. Table 37–1 describes how the management registers map to the management function interface signals.

The bit format of the rx_Config_Reg<D15:D0> and tx_Config_Reg<D15:D0> variables is context dependent, relative to the state of the Auto-Negotiation function, and is presented here and in 37.2.4.3.1.

Table 37–1—Config_Reg base page to management register mapping

Config_Reg base page bits	Management register bit
Full Duplex (FD)	4.5 Full Duplex
Half Duplex (HD)	4.6 Half Duplex
PAUSE (PS1)	4.7 PAUSE
ASM_DIR (PS2)	4.8 ASM_DIR
Remote Fault (RF2, RF1)	4.13:12 Remote Fault

37.2.1.2 Full duplex

Full Duplex (FD) is encoded in bit D5 of the base Config_Reg.

37.2.1.3 Half duplex

Half Duplex (HD) is encoded in bit D6 of the base Config_Reg.

37.2.1.4 Pause

Pause (PS1, PS2) is encoded in bits D7 and D8 of the base Config_Reg. Pause provides a pause capability exchange mechanism. Pause encoding is specified in Table 37–2.

Table 37–2—Pause encoding

PAUSE (D7)	ASM_DIR(D8)	Capability
0	0	No PAUSE
0	1	Asymmetric PAUSE toward link partner
1	0	Symmetric PAUSE
1	1	Both Symmetric PAUSE and Asymmetric PAUSE toward local device

The PAUSE bit indicates that the device is capable of providing the symmetric PAUSE functions as defined in Annex 31B. The ASM_DIR bit indicates that asymmetric PAUSE operation is supported. The value of the PAUSE bit when the ASM_DIR bit is set indicates the direction PAUSE frames are supported for flow across the link. Asymmetric PAUSE configuration results in independent enabling of the PAUSE receive and PAUSE transmit functions as defined by Annex 31B. See 37.2.4.2 for PAUSE configuration resolution.

37.2.1.5 Remote fault

Sensing of faults in a device as well as subsequent association of faults with the Remote Fault function encodings is optional. Remote Fault (RF) is encoded in bits D12 and D13 of the base page. The default value

is 0b00. Remote Fault provides a standard transport mechanism for the transmission of simple fault and error information. The Remote Fault function may indicate to the link partner that a fault or error condition has occurred. The two Remote Fault bits, RF1 and RF2, shall be encoded as specified in Table 37–3.

Table 37–3—Remote Fault encoding

RF1	RF2	Description
0	0	No error, link OK (default)
0	1	Offline
1	0	Link_Failure
1	1	Auto-Negotiation_Error

If the local device has no mechanism to detect a fault or associate a fault condition with the received Remote Fault function encodings, then it shall transmit the default Remote Fault encoding of 0b00.

A local device may indicate it has sensed a fault to its link partner by setting a nonzero Remote Fault encoding in its base page and renegotiating.

If the local device sets the Remote Fault encoding to a nonzero value, it may also use the Next Page function to specify information about the fault that has occurred. Remote Fault Message Page Codes may be specified for this purpose (see Annex 28C).

The Remote Fault encoding shall remain set until after the Auto-Negotiation process transitions into IDLE_DETECT state with the base page, at which time the Remote Fault encoding is reset to 0b00. On receipt of a base page with a nonzero Remote Fault encoding, the device shall set the Remote Fault bit in the Status register (1.4) to logic one if the GMII management function is present.

37.2.1.5.1 No error, link OK

A Remote Fault encoding of 0b00 indicates that no remote fault or error condition has been detected by the local device.

37.2.1.5.2 Offline

A Remote Fault encoding of 0b01 indicates that the local device is going Offline. A local device may indicate Offline prior to powering off, running transmitter tests, or removing the local device from the active configuration. A local device need not successfully complete the Auto-Negotiation function from the receive perspective after completing the Auto-Negotiation function indicating Offline from its transmit perspective before further action is taken (e.g., powering off, running transmitter tests, removing the local device from the active configuration, etc.).

37.2.1.5.3 Link_Failure

A Remote Fault encoding of 0b10 indicates that the local device has detected a Link_Failure condition indicated by loss of synchronization. While sync_status = FAIL, remote fault information is not signaled. When sync_status becomes OK, stored remote fault information is signaled (see 36.2.5.1.3 and 36.2.5.2.6). Another indication of a link failure condition is provided by the reception of /C/ ordered_sets having rx_Config_Reg<D15:D0> = 0 for a duration exceeding link_timer.

37.2.1.5.4 Auto-Negotiation_Error

A Remote Fault encoding of 0b11 indicates that the local device has detected a Auto-Negotiation_Error. Resolution which precludes operation between a local device and link partner shall be reflected to the link partner by the local device by indicating a Remote Fault code of Auto-Negotiation_Error.

37.2.1.6 Acknowledge

Acknowledge (Ack) is encoded in bit D14 of the base and next pages (see Figures 37–2, 37–3, and 37–4). The Ack bit is used by the Auto-Negotiation function to indicate that a device has successfully received its link partner's base or next page.

This bit is set to logic one after the device has successfully received at least three consecutive and matching rx_Config_Reg<D15:D0> values (ignoring the Acknowledge bit value), and, for next page exchanges, remains set until the next page information has been loaded into the AN next page transmit register (register 7). After the Auto-Negotiation process COMPLETE_ACKNOWLEDGE state has been entered, the tx_Config_Reg<D15:D0> value is transmitted for the link_timer duration.

37.2.1.7 Next page

The base page and subsequent next pages may set the NP bit to a logic one to request next page transmission. Subsequent next pages may set the NP bit to a logic zero in order to communicate that there is no more next page information to be sent (see 37.2.4.3). A device may implement next page ability and choose not to engage in a Next Page exchange by setting the NP bit to a logic zero.

37.2.2 Transmit function requirements

The Transmit function provides the ability to transmit /C/ ordered_sets. After Power-On, link restart, or renegotiation, the Transmit function transmits /C/ ordered_sets containing zeroes indicating the restart condition. After sending sufficient zeroes, the /C/ ordered_sets contain the Config_Reg base page value defined in 37.2.1. The local device may modify the Config_Reg value to disable an ability it possesses, but shall not transmit an ability it does not possess. This makes possible the distinction between local abilities and advertised abilities so that devices capable of multiple modes may negotiate to a mode lower in priority than the highest common local ability.

The Transmit function shall utilize the PCS Transmit process specified in 36.2.5.2.1.

37.2.2.1 Transmit function to Auto-Negotiation process interface requirements

The variable tx_Config_Reg<D15:D0> is derived from mr_adv_abilities<16:1> or mr_np_tx<16:1>. This variable is the management representation of the AN advertisement register during base page exchange and the AN next page transmit register during next page exchange.

When the xmit variable is set to CONFIGURATION by the Auto-Negotiation process, the PCS Transmit function encodes the contents of the tx_Config_Reg<D15:D0> into the appropriate /C/ ordered_set for transmission onto the MDI. When the xmit variable is set to IDLE by the Auto-Negotiation process, the PCS Transmit function transmits /I/ ordered_sets onto the MDI. When the xmit variable is set to DATA by the Auto-Negotiation process, the PCS Transmit function transmits /I/ ordered_sets interspersed with packets onto the MDI.

37.2.3 Receive function requirements

The PCS Receive function detects /C/ and /I/ ordered_sets. For received /C/, the PCS Receive function decodes the information contained within, and stores the data in rx_Config_Reg<D15:D0>.

The Receive function shall utilize the PCS Receive process specified in 36.2.5.2.2.

37.2.3.1 Receive function to Auto-Negotiation process interface requirements

The PCS Receive function provides the Auto-Negotiation process and management function with the results of rx_Config_Reg<D15:D0>. The PCS Auto-Negotiation function generates the ability_match, acknowledge_match, consistency_match, and idle_match signals.

The PCS Receive process sets the RX_UNITDATA.indicate(/C/) message when a /C/ ordered_set is received.

The PCS Receive process sets the RX_UNITDATA.indicate(/I/) message when a /I/ ordered_set is received.

The PCS Receive process sets the RX_UNITDATA.indicate(INVALID) message when an error condition is detected while not in normal receive processing (when the xmit variable is set to CONFIGURATION). The error conditions are specified in the PCS Receive state diagram of Figure 36–7a.

37.2.4 Arbitration process requirements

The Arbitration process ensures proper sequencing of the Auto-Negotiation function using the Transmit function and Receive function. The Arbitration process enables the Transmit function to advertise and acknowledge abilities. Upon completion of Auto-Negotiation information exchange, the Arbitration process determines the highest common mode using the priority resolution function and enables the appropriate functions.

37.2.4.1 Renegotiation function

A renegotiation request from any entity, such as a management agent, causes the Auto-Negotiation function to be restarted from Auto-Negotiation process state AN_ENABLE.

37.2.4.2 Priority resolution function

Since a local device and a link partner may have multiple common abilities, a mechanism to resolve which mode to configure is necessary. Auto-Negotiation shall provide the Priority Resolution function that defines the hierarchy of supported technologies.

Priority resolution is supported for pause and half/full duplex modes of operation. Full duplex shall have priority over half duplex mode. Priority resolution for pause capability shall be resolved as specified by Table 37–4. Resolution that precludes operation between a local device and link partner is reflected to the link partner by the local device by indicating a Remote Fault code of Auto-Negotiation_Error, if the remote fault function is supported (see 37.2.1.5).

37.2.4.3 Next Page function

Support for transmission and reception of additional page encodings beyond the base page (next pages) is optional. The Next Page function enables the exchange of user or application specific data. Data is carried by next pages of information, which follow the transmission and acknowledgment procedures used for the base pages. Two types of next page encodings are defined:

a) Message Pages (contain an eleven-bit formatted Message Code Field);
b) Unformatted Pages (contain an eleven-bit Unformatted Code Field).

Table 37–4—Pause priority resolution

Local Device		Link Partner		Local Resolution	Link Partner Resolution
PAUSE	ASM_DIR	PAUSE	ASM_DIR		
0	0	—	—	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
0	1	0	—	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
0	1	1	0	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
0	1	1	1	Enable PAUSE transmit, Disable PAUSE receive	Enable PAUSE receive, Disable PAUSE transmit
1	0	0	—	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
1	0	1	—	Enable PAUSE Transmit and Receive	Enable PAUSE Transmit and Receive
1	1	0	0	Disable PAUSE Transmit and Receive	Disable PAUSE Transmit and Receive
1	1	0	1	Enable PAUSE receive, Disable PAUSE transmit	Enable PAUSE transmit, Disable PAUSE receive
1	1	1	—	Enable PAUSE Transmit and Receive	Enable PAUSE Transmit and Receive

A dual acknowledgment system is used. Acknowledge (Ack) is used to acknowledge receipt of the information (see 37.2.1.6). Acknowledge 2 (Ack2) is used to indicate that the receiver is able to act on the information (or perform the task) defined in the message (see 37.2.4.3.5).

Next page operation is controlled by the same two mandatory control bits, NP and Ack, used in the base page. The Toggle bit is used to ensure proper synchronization between the local device and the link partner.

Next page exchange occurs after the base page exchange has been completed. Next page exchange consists of using the Auto-Negotiation arbitration process to send Message or Unformatted next pages. Unformatted Pages can be combined to send extended messages. Any number of next pages may be sent in any order.

Subsequent to base page exchange, a next page exchange is invoked only if both the local device and its link partner have advertised next page ability during the base page exchange.

If the Next Page function is supported by both link ends and a next page exchange has been invoked by both link ends, the next page exchange ends when both ends of a link segment set their NP bits to logic zero, indicating that neither link end has further pages to transmit. It is possible for the link partner to have more next pages to transmit than the local device. Once a local device has completed transmission of its next page information, if any, it shall transmit Message Pages with a Null Message Code (see Annex 28C) and the NP bit set to logic zero while its link partner continues to transmit valid next pages. A device shall recognize reception of Message Pages with a Null Message Code and the NP bit set to logic zero as the end of its link partner's next page information. If both the local device and its link partner advertise Next Page ability in their base pages, then both devices shall send at least one Next Page. If a device advertises Next Page ability

and has no next page information to send but is willing to receive next pages, and its link partner also advertises Next Page ability, it shall send a Message Page with a Null Message Code. The variable mr_np_loaded is set to TRUE to indicate that the local device has loaded its Auto-Negotiation next page transmit register with next page information for transmission.

A local device that requires or expects an Ack2 response from its link partner (to indicate a next page transaction has been received and can be acted upon), must terminate the next page sequence with a Null Message Code, in order to allow the link partner to transport the final Ack2 status.

37.2.4.3.1 Next page encodings

The next page shall use the encoding shown in Figure 37–3 and Figure 37–4 for the NP, Ack, MP, Ack2, and T bits. The eleven-bit field <D10:D0> is encoded as a Message Code Field if the MP bit is logic one and an Unformatted Code Field if MP is set to logic zero. The bit format of the rx_Config_Reg<D15:D0> and tx_Config_Reg<D15:D0> variables is context dependent, relative to the state of the Auto-Negotiation function, and is presented here and in 37.2.1.1.

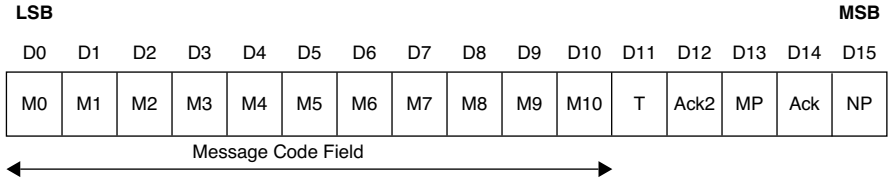


Figure 37–3—Message page encoding

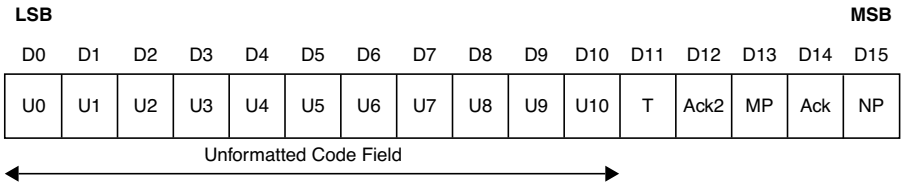


Figure 37–4—Unformatted page encoding

37.2.4.3.2 Next page

The Next Page (NP) bit is used by the Next Page function to indicate whether or not this is the last next page to be transmitted. NP shall be set as follows:

logic zero = Last page.
logic one = Additional next page(s) to follow.

37.2.4.3.3 Acknowledge

As defined in 37.2.1.6.

37.2.4.3.4 Message page

The Message Page (MP) bit is used by the Next Page function to differentiate a Message Page from an Unformatted Page. MP shall be set as follows:

logic zero = Unformatted Page.
logic one = Message Page.

37.2.4.3.5 Acknowledge 2

The Acknowledge 2 (Ack2) bit is used by the Next Page function to indicate that a device has the ability to comply with the message. Ack2 shall be set as follows:

logic zero = Cannot comply with message.
logic one = Can comply with message.

37.2.4.3.6 Toggle

The Toggle (T) bit is used by the Arbitration function to ensure synchronization with the link partner during next page exchange. This bit takes the opposite value of the Toggle bit in the previously exchanged page. The initial value of the Toggle bit in the first next page transmitted is the inverse of tx_Config_Reg<D11> in the base page that preceded the next page exchange and, therefore, may assume a value of logic one or zero. The Toggle bit is set as follows:

logic zero = Previous value of tx_Config_Reg<D11> equalled logic one.
logic one = Previous value of tx_Config_Reg<D11> equalled logic zero.

37.2.4.3.7 Message page encoding

Message Pages are formatted pages that carry a single predefined Message Code, which is enumerated in Annex 28C. There are 2048 Message Codes available. The allocation of these codes is specified in Annex 28C. If the Message Page bit is set to logic one, the bit encoding of the Config_Reg value is interpreted as a Message Page.

37.2.4.3.8 Message Code Field

Message Code Field (M<10:0>) is an eleven-bit wide field, encoding 2048 possible messages. Message Code Field definitions are shown in Annex 28C. Combinations not specified are reserved for future use. Reserved combinations of the Message Code Field shall not be transmitted.

37.2.4.3.9 Unformatted page encoding

Unformatted Pages carry the messages indicated by Message Pages. Five control bits are predefined, the remaining eleven bits are interpreted based on the preceding message page. If the Message Page bit is set to logic zero, then the bit encoding of the Config_Reg value is interpreted as an Unformatted Page.

37.2.4.3.10 Unformatted Code Field

Unformatted Code Field (U<10:0>) is an eleven-bit wide field, which may contain an arbitrary value.

37.2.4.3.11 Use of next pages

The following rules for next page usage shall be observed:

a) A next page exchange is invoked when the local device and the link partner advertise (in their base pages) that they have next page information to transmit;
b) Next page exchange continues until neither device on a link has more pages to transmit as indicated by the NP bit. A Message Page with a Null Message Code Field value is sent if the device has no other information to transmit;
c) A Message Code can carry either a specific message or information that defines how following Unformatted Page(s) should be interpreted;

d) If a Message Code references Unformatted Pages, the Unformatted Pages immediately follow the referencing Message Code in the order specified by the Message Code;
e) Unformatted Page users are responsible for controlling the format and sequencing for their Unformatted Pages.

37.2.4.3.12 Management register requirements

The AN next page transmit register defined in 37.2.5.1.6 holds the next page to be sent by Auto-Negotiation. Received next pages are stored in the AN link partner ability next page register defined in 37.2.5.1.7.

37.2.5 Management function requirements

The management interface is used to communicate Auto-Negotiation information to the management entity. Mandatory functions specified here reference bits in GMII registers 0, 1, 4, 5, 6, 7, 8, 15. Where an implementation does not use a GMII, equivalent functions to these bits must be included.

37.2.5.1 Management registers

The Auto-Negotiation function shall utilize six dedicated management registers:

a) Control register (Register 0);
b) Status register (Register 1);
c) AN advertisement register (Register 4);
d) AN link partner ability base page register (Register 5);
e) AN expansion register (Register 6);
f) Extended Status register (Register 15).

If Next Page is supported, the Auto-Negotiation function shall utilize an additional two management registers:

g) AN next page transmit register (Register 7);
h) AN link partner ability next page register (Register 8).

37.2.5.1.1 Control register (Register 0)

This register provides the mechanism to enable or disable Auto-Negotiation, restart Auto-Negotiation, and allow for manual configuration when Auto-Negotiation is not enabled. The definition for this register is provided in Clause 22.

When manual configuration is in effect at a local device, manual configuration should also be effected for the link partner to ensure predictable configuration. When manual configuration is in effect, values for the PAUSE bits (PS1, PS2) should result in a valid operational mode between the local device and the link partner.

37.2.5.1.2 Status register (Register 1)

This register includes information about all modes of operations supported by the local device and the status of Auto-Negotiation. The definition for this register is provided in Clause 22.

37.2.5.1.3 AN advertisement register (Register 4) (R/W)

This register contains the advertised ability of the local device (see Table 37–5). Before Auto-Negotiation starts, this register is configured to advertise the abilities of the local device.

Table 37–5—AN advertisement register bit definitions

Bit(s)	Name	Description	R/W
4.15	Next Page	See 37.2.1.7	R/W
4.14	Reserved	Write as zero, ignore on read	RO
4.13:12	Remote Fault	See 37.2.1.5	R/W
4.11:9	Reserved	Write as zero, ignore on read	RO
4.8:7	Pause	See 37.2.1.4	R/W
4.6	Half Duplex	See 37.2.1.3	R/W
4.5	Full Duplex	See 37.2.1.2	R/W
4.4:0	Reserved	Write as zero, ignore on read	RO

37.2.5.1.4 AN link partner ability base page register (Register 5) (RO)

All of the bits in the AN link partner ability base page register are read only. A write to the AN link partner ability base page register has no effect.

This register contains the advertised ability of the link partner (see Table 37–6). The bit definitions are a direct representation of the link partner's base page. Upon successful completion of Auto-Negotiation, the Status register Auto-Negotiation Complete bit (1.5) is set to logic one.

The values contained in this register are guaranteed to be valid either once Auto-Negotiation has successfully completed, as indicated by bit 1.5 or when the Page Received bit (6.1) is set to logic one to indicate that a new base page has been received and stored in the Auto-Negotiation link partner ability base register.

Table 37–6—AN link partner ability base page register bit definitions

Bit(s)	Name	Description	R/W
5.15	Next Page	See 37.2.1.7	RO
5.14	Acknowledge	See 37.2.1.6	RO
5.13:12	Remote Fault	See 37.2.1.5	RO
5.11:9	Reserved	Ignore on read	RO
5.8:7	Pause	See 37.2.1.4	RO
5.6	Half Duplex	See 37.2.1.3	RO
5.5	Full Duplex	See 37.2.1.2	RO
5.4:0	Reserved	Ignore on read	RO

37.2.5.1.5 AN expansion register (Register 6) (RO)

All of the bits in the Auto-Negotiation expansion register are read only; a write to the Auto-Negotiation expansion register has no effect.

Table 37–7—AN expansion register bit definitions

Bit(s)	Name	Description	R/W	Default
6.15:3	Reserved	Ignore on read	RO	0
6.2	Next Page Able	1 = Local device is next page able 0 = Local device is not next page able	RO	0
6.1	Page Received	1 = A new page has been received 0 = A new page has not been received	RO/ LH	0
6.0	Reserved	Ignore on read	RO	0

Bits 6.15:3 and 6.0 are reserved for future Auto-Negotiation expansion.

The Next Page Able bit (6.2) is set to logic one to indicate that the local device supports the Next Page function. The Next Page Able bit is set to logic zero if the Next Page function is not supported by the local device.

The Page Received bit (6.1) is set to logic one to indicate that a new page has been received and stored in the applicable AN link partner ability base or next page register.

The Page Received bit shall be reset to logic zero on a read of the AN expansion register (Register 6). Subsequent to the setting of the Page Received bit, and in order to prevent overlay of the AN link partner ability next page register, the AN link partner ability next page register should be read before the AN next page transmit register is written.

37.2.5.1.6 AN next page transmit register (Register 7)

This register contains the next page value to be transmitted, if required. The definition for this register is provided in 22.2.4.1.6.

37.2.5.1.7 AN link partner ability next page register (Register 8)

This register contains the advertised ability of the link partner's next page. The definition for this register is provided in 32.5.4.2 for changes to 28.2.4.1.4.

37.2.5.1.8 Extended status register (Register 15)

This register includes additional information about all modes of operations supported by the local device. The definition for this register is provided in Clause 22.

37.2.5.1.9 State diagram variable to management register mapping

The state diagram of Figure 37–6 generates and accepts variables of the form "mr_x," where x is an individual signal name. These variables comprise a management interface that may be connected to the GMII management function or other equivalent function. Table 37–8 describes how PCS state diagram variables in both Clauses 36 and 37 map to management register bits.

37.2.5.2 Auto-Negotiation managed object class

The Auto-Negotiation Managed Object Class is defined in Clause 30.

Table 37–8—PCS state diagram variable to management register mapping

State diagram variable	Management register bit
mr_adv_ability<16:1>	4.15:0 Auto-Negotiation advertisement register
mr_an_complete	1.5 Auto-Negotiation complete
mr_an_enable	0.12 Auto-Negotiation enable
mr_loopback	0.14 Loopback (see 36.2.5.1.3)
mr_lp_adv_ability<16:1>	5.15:0 AN link partner ability register
mr_lp_np_rx<16:1>	8.15:0 AN link partner next page ability register
mr_main_reset	0.15 Reset
mr_np_able	6.2 Next page able
mr_np_loaded	Set on write to the AN next page transmit register; cleared by Auto-Negotiation state diagram
mr_np_tx<16:1>	7.15:0 AN next page transmit register
mr_page_rx	6.1 Page received
mr_restart_an	0.9 Auto-Negotiation restart
xmit=DATA	1.2 Link status

37.2.6 Absence of management function

In the absence of any management function, the advertised abilities shall be provided through a logical equivalent of mr_adv_ability<16:1>.

37.3 Detailed functions and state diagrams

The notation used in the state diagram in Figure 37–6 follows the conventions in 21.5. State diagram variables follow the conventions of 21.5.2 except when the variable has a default value. Variables in a state diagram with default values evaluate to the variable default in each state where the variable value is not explicitly set. Variables using the "mr_x" notation do not have state diagram defaults; however, their appropriate initialization conditions when mapped to the management interface are covered in 22.2.4. The variables, timers, and counters used in the state diagrams are defined in 37.3.1.

Auto-Negotiation shall implement the Auto-Negotiation state diagram and meet the Auto-Negotiation state diagram interface requirements of the Receive and Transmit functions. Additional requirements to these state diagrams are made in the respective functional requirements sections. In the case of any ambiguity between stated requirements and the state diagrams, the state diagrams take precedence. A functional reference diagram of Auto-Negotiation is shown in Figure 37–5.

37.3.1 State diagram variables

Variables with <16:1> or <D15:D0> appended to the end of the variable name indicate arrays that can be mapped to 16-bit management registers. For these variables, "<x>" indexes an element or set of elements in the array, where "x" may be as follows:

— Any integer or set of integers.

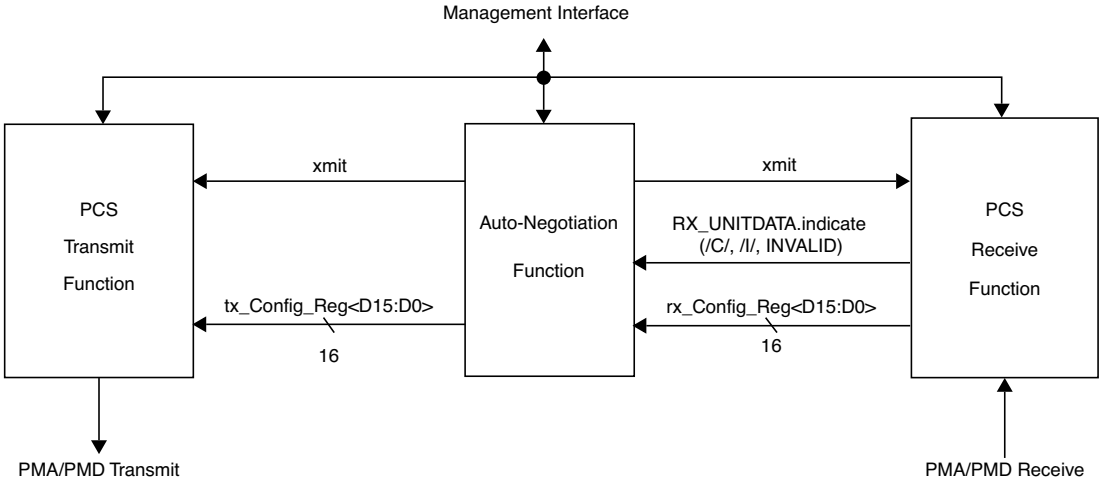


Figure 37–5—Functional reference diagram

— Any variable that takes on integer values.
— NP; represents the index of the Next Page bit.
— ACK; represents the index of the Acknowledge bit.
— RF; represents the index of the Remote Fault bits.

Variables of the form "mr_x," where x is a label, comprise a management interface that is intended to be connected to the GMII Management function. However, an implementation-specific management interface may provide the control and status function of these bits.

37.3.1.1 Variables

an_sync_status
: Qualified version of sync_status for use by Auto-Negotiation to detect a sync_status timeout condition.

  Values: OK; The variable sync_status defined in 36.2.5.1.3 is OK.
  FAIL; The variable sync_status defined in 36.2.5.1.3 is FAIL for a duration greater than or equal to the link timer.

mr_adv_ability<16:1>
: A 16-bit array that contains the advertised ability base page of the local device to be conveyed to tx_Config_Reg<D15:D0> for transmission to the link partner. For each element within the array:

  Values: ZERO; Data bit is logical zero.
  ONE; Data bit is logical one.

mr_an_complete
: Status indicating whether Auto-Negotiation has completed or not.

  Values: FALSE; Auto-Negotiation has not completed.
  TRUE; Auto-Negotiation has completed.

mr_an_enable
: Controls the enabling and disabling of the Auto-Negotiation function for 1000BASE-X. Auto-Negotiation function for 1000BASE-X is enabled when Control register bit 0.12 is set to one.

  Values: FALSE; Auto-Negotiation is disabled.
  TRUE; Auto-Negotiation is enabled.

mr_lp_adv_ability<16:1>

A 16-bit array that contains the advertised ability base page of the link partner conveyed from rx_Config_Reg<D15:D0>. For each element within the array:

Values: ZERO; Data bit is logical zero.
ONE; Data bit is logical one.

mr_lp_np_rx<16:1>

A 16-bit array that contains the advertised ability of the link partner's next page conveyed from rx_Config_Reg<D15:D0>. For each element within the array:

Values: ZERO; Data bit is logical zero.
ONE; Data bit is logical one.

mr_main_reset

Controls the resetting of the Auto-Negotiation function.

Values: FALSE; Do not reset the Auto-Negotiation function.
TRUE; Reset the Auto-Negotiation function.

mr_np_able

Status indicating whether the local device supports next page exchange.

Values: FALSE; The local device does not support next page exchange.
TRUE; The local device supports next page exchange.

mr_np_loaded

Status indicating whether a new page has been loaded into the AN next page transmit register (register 7).

Values: FALSE; A new page has not been loaded.
TRUE; A new page has been loaded.

mr_np_tx<16:1>

A 16-bit array that contains the new next page to transmit. If a next page exchange is invoked, this array is conveyed to tx_Config_Reg<D15:D0> for transmission to the link partner. For each element within the array:

Values: ZERO; Data bit is logical zero.
ONE; Data bit is logical one.

mr_page_rx

Status indicating whether a new page has been received. A new page has been successfully received when acknowledge_match=TRUE and consistency_match=TRUE and the rx_Config_Reg<D15:D0> value has been written to mr_lp_adv_ability<16:1> or mr_lp_np_rx<16:1>, depending on whether the page received was a base or next page, respectively.

Values: FALSE; A new page has not been received.
TRUE; A new page has been received.

NOTE—For the first setting of mr_page_rx, mr_lp_adv_ability is valid but need not be read as it is preserved through a next page operation. On subsequent settings of mr_page_rx, mr_lp_np_rx must be read prior to loading mr_np_tx register in order to avoid the overlay of next page information.

mr_restart_an

Controls renegotiation via management control.

Values: FALSE; Do not restart Auto-Negotiation.
TRUE; Restart Auto-Negotiation.

np_rx

Flag to hold value of rx_Config_Reg<NP> upon entry to state COMPLETE ACKNOWLEDGE.

This value is associated with the value of rx_Config_Reg<NP> when acknowledge_match was last set.

Values: ZERO; The local device np_rx bit equals logic zero.
ONE; The local device np_rx bit equals logic one.

power_on

Condition that is true until such time as the power supply for the device that contains the Auto-Negotiation function has reached the operating region. The condition is also true when the device has low power mode set via Control register bit 0.11.

Values: FALSE; The device is completely powered (default).
TRUE; The device has not been completely powered.

NOTE—Power_on evaluates to its default value in each state where it is not explicitly set.

resolve_priority

Controls the invocation of the priority resolution function specified in Table 37–4.

Values: OFF; The priority resolution function is not invoked (default).
ON; The priority resolution function is invoked.

NOTE—Resolve_priority evaluates to its default value in each state where it is not explicitly set.

rx_Config_Reg<D15:D0>

Defined in 36.2.5.1.3.

sync_status

Defined in 36.2.5.1.3.

toggle_rx

Flag to keep track of the state of the link partner Toggle bit.

Values: ZERO; The link partner Toggle bit equals logic zero.
ONE; The link partner Toggle bit equals logic one.

toggle_tx

Flag to keep track of the state of the local device Toggle bit.

Values: ZERO; The local device Toggle bit equals logic zero.
ONE; The local device Toggle bit equals logic one.

tx_Config_Reg<D15:D0>

Defined in 36.2.5.1.3. This array may be loaded from mr_adv_ability or mr_np_tx.

xmit

A parameter set by the PCS Auto-Negotiation process to reflect the source of information to the PCS Transmit process.

Values: CONFIGURATION: Tx_Config_Reg<D15:D0> information is being sourced from the PCS Auto-Negotiation process.
DATA: /I/, sourced from the PCS, is interspersed with packets sourced from the MAC.
IDLE: /I/ is being sourced from the PCS Auto-Negotiation process.

37.3.1.2 Functions

ability_match

For a stream of /C/ and /I/ ordered_sets, this function continuously indicates whether the last three consecutive rx_Config_Reg<D15,D13:D0> values match. Three consecutive rx_Config_Reg<D15,D13:D0> values are any three rx_Config_Reg<D15,D13:D0> values received one after the other, regardless of whether the rx_Config_Reg<D15,D13:D0> value has

already been used in a rx_Config_Reg<D15,D13:D0> match comparison or not. The match count is reset upon receipt of /I/. The match count is reset upon receipt of a second or third consecutive rx_Config_Reg<D15,D13:D0> value which does not match the rx_Config_Reg<D15,D13:D0> values for which the match count was set to one.

Values: FALSE; Three matching consecutive rx_Config_Reg<D15,D13:D0> values have not been received (default).
TRUE; Three matching consecutive rx_Config_Reg<D15,D13:D0> values have been received.

NOTE—Ability_match is set by this function definition; it is not set explicitly in the state diagrams. Ability_match evaluates to its default value upon state entry.

acknowledge_match

For a stream of /C/ and /I/ ordered_sets, this function continuously indicates whether the last three consecutive rx_Config_Reg<D15:D0> values match and have the Acknowledge bit set. Three consecutive rx_Config_Reg<D15:D0> values are any three rx_Config_Reg<D15:D0> values contained within three /C/ ordered_sets received one after the other, regardless of whether the rx_Config_Reg<D15:D0> value has already been used in a rx_Config_Reg<D15:D0> match comparison or not. The match count is reset upon receipt of /I/. The match count is reset upon receipt of a second or third consecutive rx_Config_Reg<D15:D0> value which does not match the rx_Config_Reg<D15:D0> values for which the match count was set to one.

Values: FALSE; Three matching and consecutive rx_Config_Reg<D15:D0> values have not been received with the Acknowledge bit set (default).
TRUE; Three matching and consecutive rx_Config_Reg<D15:D0> values have been received with the Acknowledge bit set.

NOTE—Acknowledge_match is set by this function definition; it is not set explicitly in the state diagrams. Acknowledge_match evaluates to its default value upon state entry.

an_enableCHANGE

This function monitors the mr_an_enable variable for a state change. The function is set to TRUE on state change detection.

Values: TRUE; A mr_an_enable variable state change has been detected.
FALSE; A mr_an_enable variable state change has not been detected (default).

NOTE—An_enableCHANGE is set by this function definition; it is not set explicitly in the state diagrams. An_enableCHANGE evaluates to its default value upon state entry.

consistency_match

Indicates that the rx_Config_Reg<D15,D13:D0> value that caused ability_match to be set, for the transition from states ABILITY_DETECT or NEXT_PAGE_WAIT to state ACKNOWLEDGE_DETECT, is the same as the rx_Config_Reg<D15,D13:D0> value that caused acknowledge_match to be set.

Values: FALSE; The rx_Config_Reg<D15,D13:D0> value that caused ability_match to be set is not the same as the rx_Config_Reg<D15,D13:D0> value that caused acknowledge_match to be set, ignoring the Acknowledge bit value.
TRUE; The rx_Config_Reg<D15,D13:D0> value that caused ability_match to be set is the same as the rx_Config_Reg<D15,D13:D0> value that caused acknowledge_match to be set, independent of the Acknowledge bit value.

NOTE—Consistency_match is set by this function definition; it is not set explicitly in the state diagrams.

idle_match
: For a stream of /C/ and /I/ ordered_sets, this function continuously indicates whether three consecutive /I/ ordered_sets have been received. The match count is reset upon receipt of /C/.

  Values: FALSE; Three consecutive /I/ ordered_sets have not been received (default).
  TRUE; Three consecutive /I/ ordered_sets have been received.

NOTE—Idle_match is set by this function definition; it is not set explicitly in the state diagrams. Idle_match evaluates to its default value upon state entry.

37.3.1.3 Messages

RUDI
: Alias for RX_UNITDATA.indicate(parameter). Defined in 36.2.5.1.6.

RX_UNITDATA.indicate
: Defined in 36.2.5.1.6.

37.3.1.4 Timers

All timers operate in the manner described in 14.2.3.2.

link_timer
: Timer used to ensure Auto-Negotiation protocol stability and register read/write by the management interface.

  Duration: 10 ms, tolerance +10 ms, –0 s.

37.3.1.5 State diagrams

The Auto-Negotiation state diagram is specified in Figure 37–6.

37.4 Environmental specifications

All equipment subject to this clause shall conform to the requirements of 14.7 and applicable sections of ISO/IEC 11801: 1995.

37.5 Protocol Implementation Conformance Statement (PICS) proforma for Clause 37, Auto-Negotiation function, type 1000BASE-X[4]

37.5.1 Introduction

The supplier of a protocol implementation that is claimed to conform to Clause 37, Auto-Negotiation function, type 1000BASE-X, shall complete the following Protocol Implementation Conformance Statement (PICS) proforma.

A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

37.5.2 Identification

[4]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so that it can be used for its intended purpose and may further publish the completed PICS.

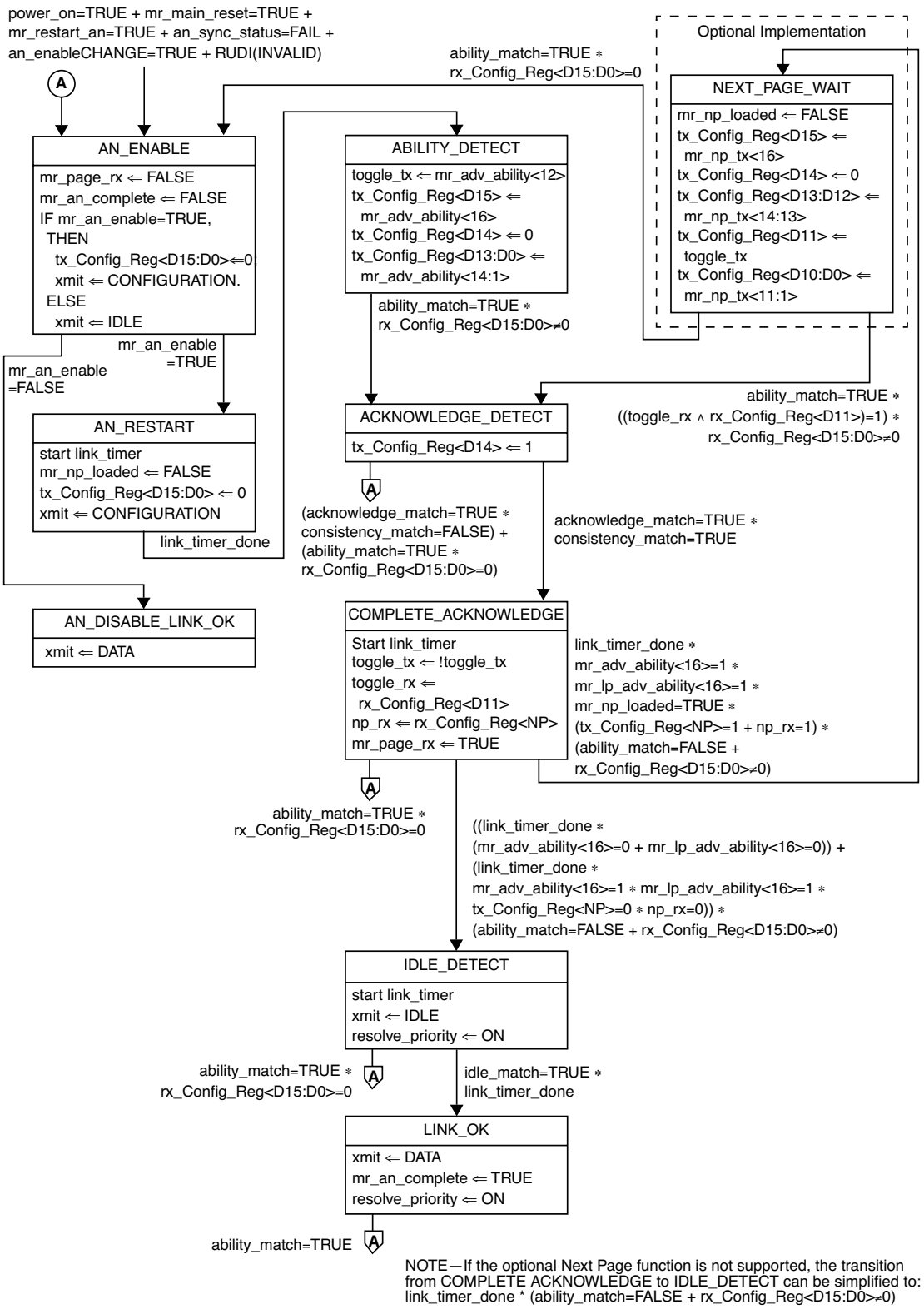


Figure 37–6—Auto-Negotiation state diagram

37.5.2.1 Implementation identification

Supplier (Note 1)	
Contact point for enquiries about the PICS (Note 1)	
Implementation Name(s) and Version(s) (Notes 1 and 3)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Names(s) (Note 2)	

NOTE 1—Required for all implementations.

NOTE 2—May be completed as appropriate in meeting the requirements for the identification.

NOTE 3—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).

37.5.2.2 Protocol summary

Identification of protocol standard	IEEE Std 802.3-2002®, Clause 37, Auto-Negotiation function, type 1000BASE-X
Identification of amendments and corrigenda to this PICS proforma that have been completed as part of this PICS	
Have any Exception items been required? (See Clause 21; the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®.)	No [ ] Yes [ ]

Date of Statement	

37.5.3 Major Capabilities/Options

Item	Feature	Subclause	Value/Comment	Status	Support
*GMII	GMII Management Interface	37.1.4.2.1		O	Yes [ ] No [ ]
*RF	Remote Fault function	37.2.1.5		O	Yes [ ] No [ ]
*NP	Next Page function	37.2.4.3		O	Yes [ ] No [ ]

In addition, the following predicate name is defined for use when different implementations from the set above have common parameters:

*NPM: GMII and NP

37.5.4 PICS proforma tables for the Auto-Negotiation function, type 1000BASE-X

37.5.4.1 Compatibility considerations

Item	Feature	Subclause	Value/Comment	Status	Support
CC1	Provision of logical equivalent of mr_adv_ability<16:1>	37.2.6	In the absence of any management function	M	Yes [ ]
CC2	Environmental specifications	37.4		M	Yes [ ]

37.5.4.2 Auto-Negotiation functions

Item	Feature	Subclause	Value/Comment	Status	Support
AN1	Config_Reg encoding	37.2.1		M	Yes [ ]
AN2	Priority Resolution function	37.2.4.2		M	Yes [ ]
AN3	Auto-Negotiation state diagram	37.3		M	Yes [ ]

37.5.4.2.1 Config_Reg

Item	Feature	Subclause	Value/Comment	Status	Support
CR1	Reserved bits	37.2.1	Set to zero	M	Yes [ ]
CR2	Default encoding of Remote Fault bits	37.2.1.5	0b00	M	Yes [ ]

37.5.4.2.2 Remote Fault functions

Item	Feature	Subclause	Value/Comment	Status	Support
RF1	Remote Fault encoding	37.2.1.5		RF:M	Yes [ ] N/A [ ]
RF2	Use of Remote Fault Message Page code	37.2.1.5	To signal additional fault information	RF:O	Yes [ ] No [ ]
RF3	Notification duration	37.2.1.5	Remains set until transition to IDLE_DETECT state, then reset to 0b00	RF:M	Yes [ ] N/A [ ]
RF4	Status Register RF bit (1.4)	37.2.1.5	Upon detection of nonzero Remote Fault encoding	RF:M	Yes [ ] N/A [ ]
RF5	Offline indication	37.2.1.5.2	0b01	RF:O	Yes [ ] No [ ]
RF6	Link_Failure indication	37.2.1.5.3	0b10	RF:O	Yes [ ] No [ ]
RF7	Auto-Negotiation_Error	37.2.1.5.4	0b11	RF:M	Yes [ ] N/A [ ]

37.5.4.2.3 AN transmit functions

Item	Feature	Subclause	Value/Comment	Status	Support
TX1	PCS Transmit function support	37.2.2		M	Yes [ ]
TX2	Transmission of non-possessive abilities	37.2.2	A device shall not transmit an ability it does not possess.	M	Yes [ ]

37.5.4.2.4 AN receive functions

Item	Feature	Subclause	Value/Comment	Status	Support
RX1	PCS Receive function support	37.2.3		M	Yes [ ]

37.5.4.2.5 Priority resolution functions

Item	Feature	Subclause	Value/Comment	Status	Support
PR1	Full duplex priority over half duplex	37.2.4.2		M	Yes [ ]
PR2	Priority resolution for pause capability	37.2.4.2	Specified in Table 37–4	M	Yes [ ]

37.5.4.2.6 Next page functions

Item	Feature	Subclause	Value/Comment	Status	Support
	Transmission of Message Pages with a Null Message Code	37.2.4.3	Upon local device completion of next page requests	NP:M	Yes [ ] N/A [ ]
	Recognition of Message Pages with a Null Message Code	37.2.4.3	Signifies the end of link partner next page information	NP:M	Yes [ ] N/A [ ]
	Initial Next Page Exchange	37.2.4.3	Upon advertisement of NP ability by both devices	NP:M	Yes [ ] N/A [ ]
	Next Page Receipt Ability	37.2.4.3	Indicated by advertising NP ability via the NP bit	NP:M	Yes [ ] N/A [ ]
	Next page Config_Reg encoding	37.2.4.3.1		NP:M	Yes [ ] N/A [ ]
	Next Page (NP) bit setting	37.2.4.3.2	For link_timer after entry into COMPLETE_ACKNOWLEDGE state	NP:M	Yes [ ] N/A [ ]
	Message Page (MP) bit setting	37.2.4.3.4		NP:M	Yes [ ] N/A [ ]
	Acknowledge 2 (Ack2) bit setting	37.2.4.3.5		NP:M	Yes [ ] N/A [ ]
	Message Code Field combinations	37.2.4.3.8	Reserved combinations shall not be transmitted.	NP:M	Yes [ ] N/A [ ]
	Next Page usage rules	37.2.4.3.11		NP:M	Yes [ ] N/A [ ]

37.5.4.2.7 Management registers

Item	Feature	Subclause	Value/Comment	Status	Support
MR1	Control and Status registers	37.1.4.2.1		GMII:M	Yes [ ] N/A [ ]
MR2	Register usage	37.2.5.1	Logical equivalent of Registers 0, 1, 4, 5, 6 and 15	GMII:M	Yes [ ] N/A [ ]
MR3	Next Page Register usage	37.2.5.1	Logical equivalent of Registers 7 and 8	NPM:M	Yes [ ] N/A [ ]
MR4	Page Received resetting	37.2.5.1.5	Reset upon read to AN expansion register	M	Yes [ ]

38. Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-LX (Long Wavelength Laser) and 1000BASE-SX (Short Wavelength Laser)

38.1 Overview

This clause specifies the 1000BASE-SX PMD and the 1000BASE-LX PMD (including MDI) and baseband medium for multimode and single-mode fiber. In order to form a complete Physical Layer, it shall be integrated with the 1000BASE-X PCS and PMA of Clause 36, and integrated with the management functions which are accessible through the Management Interface defined in Clause 35, which are hereby incorporated by reference.

38.1.1 Physical Medium Dependent (PMD) sublayer service interface

The following specifies the services provided by the 1000BASE-SX and 1000BASE-LX PMD. These PMD sublayers are described in an abstract manner and do not imply any particular implementation. It should be noted that these services are based on similar interfaces defined in ANSI X3.230-1994 [B20] (FC-PH).

The PMD Service Interface supports the exchange of encoded 8B/10B characters between PMA entities. The PMD translates the encoded 8B/10B characters to and from signals suitable for the specified medium.

The following primitives are defined:

PMD_UNITDATA.request
PMD_UNITDATA.indicate
PMD_SIGNAL.indicate

NOTE—Delay requirements from the MDI to GMII that include the PMD layer are specified in Clause 36. Of this budget, 4 ns is reserved for each of the transmit and receive functions of the PMD.

38.1.1.1 PMD_UNITDATA.request

This primitive defines the transfer of data (in the form of encoded 8B/10B characters) from the PMA to the PMD.

38.1.1.1.1 Semantics of the service primitive

PMD_UNITDATA.request (tx_bit)

The data conveyed by PMD_UNITDATA.request is a continuous sequence of encoded 8B/10B characters. The tx_bit parameter can take one of two values: ONE or ZERO.

38.1.1.1.2 When generated

The PMA continuously sends the appropriate encoded 8B/10B characters to the PMD for transmission on the medium, at a nominal 1.25 GBd signaling speed.

38.1.1.1.3 Effect of receipt

Upon receipt of this primitive, the PMD converts the specified encoded 8B/10B characters into the appropriate signals on the MDI.

38.1.1.2 PMD_UNITDATA.indicate

This primitive defines the transfer of data (in the form of encoded 8B/10B characters) from the PMD to the PMA.

38.1.1.2.1 Semantics of the service primitive

PMD_UNITDATA.indicate (rx_bit)

The data conveyed by PMD_UNITDATA.indicate is a continuous sequence of encoded 8B/10B characters. The rx_bit parameter can take one of two values: ONE or ZERO.

38.1.1.2.2 When generated

The PMD continuously sends encoded 8B/10B characters to the PMA corresponding to the signals received from the MDI.

38.1.1.2.3 Effect of receipt

The effect of receipt of this primitive by the client is unspecified by the PMD sublayer.

38.1.1.3 PMD_SIGNAL.indicate

This primitive is generated by the PMD to indicate the status of the signal being received from the MDI.

38.1.1.3.1 Semantics of the service primitive

PMD_SIGNAL.indicate(SIGNAL_DETECT)

The SIGNAL_DETECT parameter can take on one of two values: OK or FAIL, indicating whether the PMD is detecting light at the receiver (OK) or not (FAIL). When SIGNAL_DETECT = FAIL, then rx_bit is undefined, but consequent actions based on PMD_UNITDATA.indicate, where necessary, interpret rx_bit as a logic ZERO.

NOTE—SIGNAL_DETECT = OK does not guarantee that rx_bit is known good. It is possible for a poor quality link to provide sufficient light for a SIGNAL_DETECT = OK indication and still not meet the 10^{-12} BER objective.

38.1.1.3.2 When generated

The PMD generates this primitive to indicate a change in the value of SIGNAL_DETECT.

38.1.1.3.3 Effect of receipt

The effect of receipt of this primitive by the client is unspecified by the PMD sublayer.

38.1.2 Medium Dependent Interface (MDI)

The MDI, a physical interface associated with a PMD, is comprised of an electrical or optical medium connection.

38.2 PMD functional specifications

The 1000BASE-X PMDs perform the Transmit and Receive functions that convey data between the PMD service interface and the MDI.

38.2.1 PMD block diagram

For purposes of system conformance, the PMD sublayer is standardized at the following points. The optical transmit signal is defined at the output end of a patch cord (TP2), between 2 and 5 m in length, of a type consistent with the link type connected to the transmitter receptacle defined in 38.11.2. If a single-mode fiber offset-launch mode-conditioning patch cord is used, the optical transmit signal is defined at the end of this single-mode fiber offset-launch mode-conditioning patch cord at TP2. Unless specified otherwise, all transmitter measurements and tests defined in 38.6 are made at TP2. The optical receive signal is defined at the output of the fiber optic cabling (TP3) connected to the receiver receptacle defined in 38.11.2. Unless specified otherwise, all receiver measurements and tests defined in 38.6 are made at TP3.

TP1 and TP4 are standardized reference points for use by implementers to certify component conformance. The electrical specifications of the PMD service interface (TP1 and TP4) are not system compliance points (these are not readily testable in a system implementation). It is expected that in many implementations, TP1 and TP4 will be common between 1000BASE-SX, 1000BASE-LX, and 1000BASE-CX (Clause 39).

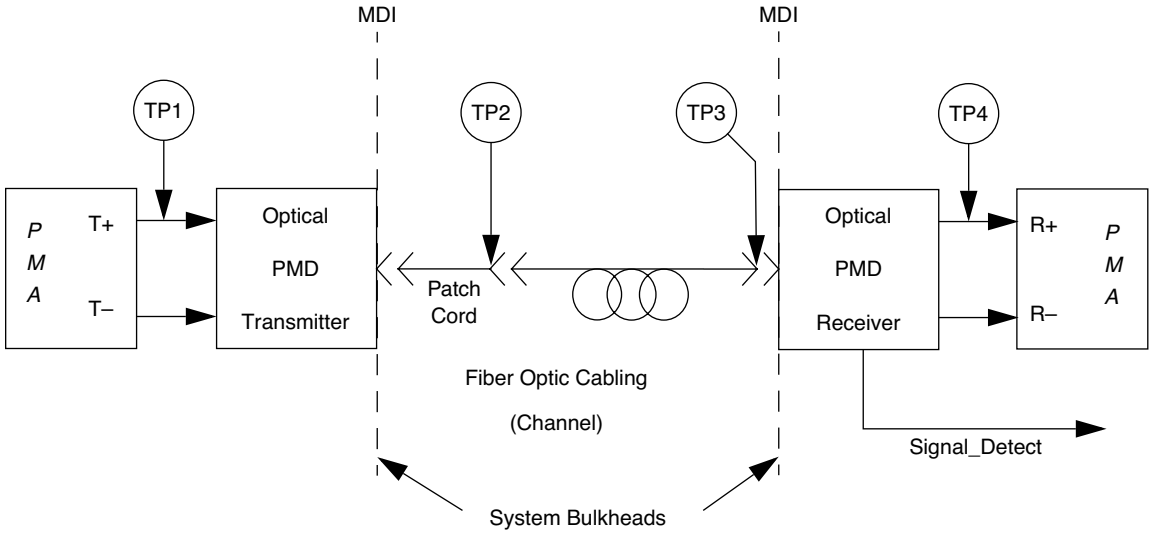


Figure 38–1—1000BASE-X block diagram

38.2.2 PMD transmit function

The PMD Transmit function shall convey the bits requested by the PMD service interface message PMD_UNITDATA.request(tx_bit) to the MDI according to the optical specifications in this clause. The higher optical power level shall correspond to tx_bit = ONE.

38.2.3 PMD receive function

The PMD Receive function shall convey the bits received from the MDI according to the optical specifications in this clause to the PMD service interface using the message PMD_UNITDATA.indicate(rx_bit). The higher optical power level shall correspond to rx_bit = ONE.

38.2.4 PMD signal detect function

The PMD Signal Detect function shall report to the PMD service interface, using the message PMD_SIGNAL.indicate(SIGNAL_DETECT), which is signaled continuously. PMD_SIGNAL.indicate is intended to be an indicator of optical signal presence.

The value of the SIGNAL_DETECT parameter shall be generated according to the conditions defined in Table 38–1. The PMD receiver is not required to verify whether a compliant 1000BASE-X signal is being received. This standard imposes no response time requirements on the generation of the SIGNAL_DETECT parameter..

Table 38–1—SIGNAL_DETECT value definition

Receive conditions	Signal detect value
Input_optical_power ≤ –30 dBm	FAIL
Input_optical_power ≥ Receive sensitivity AND compliant 1000BASE-X signal input	OK
All other conditions	Unspecified

As an unavoidable consequence of the requirements for the setting of the SIGNAL_DETECT parameter, implementations must provide adequate margin between the input optical power level at which the SIGNAL_DETECT parameter is set to OK, and the inherent noise level of the PMD due to cross talk, power supply noise, etc.

Various implementations of the Signal Detect function are permitted by this standard, including implementations which generate the SIGNAL_DETECT parameter values in response to the amplitude of the 8B/10B modulation of the optical signal and implementations that respond to the average optical power of the 8B/10B-modulated optical signal.

38.3 PMD to MDI optical specifications for 1000BASE-SX

The operating range for 1000BASE-SX is defined in Table 38–2. A 1000BASE-SX compliant transceiver supports both multimode fiber media types listed in Table 38–2 (i.e., both 50 μm and 62.5 μm multimode fiber) according to the specifications defined in 38.11. A transceiver that exceeds the operational range requirement while meeting all other optical specifications is considered compliant (e.g., a 50 μm solution operating at 600 m meets the minimum range requirement of 2 to 550 m).

Table 38–2—Operating range for 1000BASE-SX over each optical fiber type

Fiber type	Modal bandwidth @ 850 nm (min. overfilled launch) (MHz · km)	Minimum range (meters)
62.5 μm MMF	160	2 to 220
62.5 μm MMF	200	2 to 275
50 μm MMF	400	2 to 500
50 μm MMF	500	2 to 550
10 μm SMF	N/A	Not supported

38.3.1 Transmitter optical specifications

The 1000BASE-SX transmitter shall meet the specifications defined in Table 38–3 per measurement techniques defined in 38.6. It shall also meet a transmit mask of the eye measurement as defined in 38.6.5.

Table 38–3—1000BASE-SX transmit characteristics

Description	62.5 μm MMF	50 μm MMF	Unit
Transmitter type	Shortwave Laser		
Signaling speed (range)	1.25 ± 100 ppm		GBd
Wavelength (λ, range)	770 to 860		nm
T_{rise}/T_{fall} (max; 20%-80%; λ > 830 nm)	0.26		ns
T_{rise}/T_{fall} (max; 20%-80%; λ ≤ 830 nm)	0.21		ns
RMS spectral width (max)	0.85		nm
Average launch power (max)	See footnote [a]		dBm
Average launch power (min)	–9.5		dBm
Average launch power of OFF transmitter (max)[b]	–30		dBm
Extinction ratio (min)	9		dB
RIN (max)	–117		dB/Hz
Coupled Power Ratio (CPR) (min)[c]	9 < CPR		dB

[a]The 1000BASE-SX launch power shall be the lesser of the class 1 safety limit as defined by 38.7.2 or the average receive power (max) defined by Table 38–4.

[b]Examples of an OFF transmitter are: no power supplied to the PMD, laser shutdown for safety conditions, activation of a "transmit disable" or other optional module laser shut down conditions. During all conditions when the PMA is powered, the ac signal (data) into the transmit port will be valid encoded 8B/10B patterns (this is a requirement of the PCS layers) except for short durations during system power-on-reset or diagnostics when the PMA is placed in a loopback mode.

[c]Radial overfilled launches as described in 38A.2, while they may meet CPR ranges, should be avoided.

The CPR specification provides sufficient mode volume so that individual multimode fiber (MMF) modes do not dominate fiber performance. This reduces the effect of peak-to-peak differential mode delay (DMD) between the launched mode groups and diminishes the resulting pulse-splitting-induced nulls in the frequency response.

38.3.2 Receive optical specifications

The 1000BASE-SX receiver shall meet the specifications defined in Table 38–4 per measurement techniques defined in 38.6. The sampling instant is defined to occur at the eye center. The receive sensitivity includes the extinction ratio penalty.

Table 38–4—1000BASE-SX receive characteristics

Description	62.5 μm MMF	50 μm MMF	Unit
Signaling Speed (range)	1.25 ± 100 ppm		GBd
Wavelength (range)	770 to 860		nm
Average receive power (max)	0		dBm
Receive sensitivity	–17		dBm
Return loss (min)	12		dB
Stressed receive sensitivity[a, b]	–12.5	–13.5	dBm
Vertical eye-closure penalty[c]	2.60	2.20	dB
Receive electrical 3 dB upper cutoff frequency (max)	1500		MHz

[a]Measured with conformance test signal at TP3 (see 38.6.11) for BER = 10^{-12} at the eye center.
[b]Measured with a transmit signal having a 9 dB extinction ratio. If another extinction ratio is used, the stressed receive sensitivity should be corrected for the extinction ratio penalty.
[c]Vertical eye-closure penalty is a test condition for measuring stressed receive sensitivity. It is not a required characteristic of the receiver.

38.3.3 Worst-case 1000BASE-SX link power budget and penalties (informative)

The worst-case power budget and link penalties for a 1000BASE-SX channel are shown in Table 38–5.

Table 38–5—Worst-case 1000BASE-SX link power budget and penalties[a]

Parameter	62.5 μm MMF		50 μm MMF		Unit
Modal bandwidth as measured at 850 nm (minimum, overfilled launch)	160	200	400	500	MHz · km
Link power budget	7.5	7.5	7.5	7.5	dB
Operating distance	220	275	500	550	m
Channel insertion loss[b, c]	2.38	2.60	3.37	3.56	dB
Link power penalties[c]	4.27	4.29	4.07	3.57	dB
Unallocated margin in link power budget[c]	0.84	0.60	0.05	0.37	dB

[a]Link penalties are used for link budget calculations. They are not requirements and are not meant to be tested.
[b]Operating distances used to calculate the channel insertion loss (see 1.4) are the maximum values specified in Table 38–2.
[c]A wavelength of 830 nm is used to calculate channel insertion loss, link power penalties, and unallocated margin.

38.4 PMD to MDI optical specifications for 1000BASE-LX

The operating range for 1000BASE-LX is defined in Table 38–6. A 1000BASE-LX compliant transceiver supports all media types listed in Table 38–6 (i.e., 50 μm and 62.5 μm multimode fiber, and 10 μm single-mode fiber) according to the specifications defined in 38.11. A transceiver which exceeds the operational range requirement while meeting all other optical specifications is considered compliant (e.g., a single-mode solution operating at 5500 m meets the minimum range requirement of 2 to 5000 m).

Table 38–6—Operating range for 1000BASE-LX over each optical fiber type

Fiber type	Modal bandwidth @ 1300 nm (min. overfilled launch) (MHz · km)	Minimum range (meters)
62.5 μm MMF	500	2 to 550
50 μm MMF	400	2 to 550
50 μm MMF	500	2 to 550
10 μm SMF	N/A	2 to 5000

38.4.1 Transmitter optical specifications

The 1000BASE-LX transmitter shall meet the specifications defined in Table 38–7 per measurement techniques defined in 38.6. It shall also meet a transmit mask of the eye measurement as defined in 38.6.5. To ensure that the specifications of Table 38–7 are met with MMF links, the 1000BASE-LX transmitter output shall be coupled through a single-mode fiber offset-launch mode-conditioning patch cord, as defined in 38.11.4.

Table 38–7—1000BASE-LX transmit characteristics

Description	62.5 μm MMF	50 μm MMF	10 μm SMF	Unit
Transmitter type	Longwave Laser			
Signaling speed (range)	1.25 ± 100 ppm			GBd
Wavelength (range)	1270 to 1355			nm
T_{rise}/T_{fall} (max, 20-80% response time)	0.26			ns
RMS spectral width (max)	4			nm
Average launch power (max)	–3			dBm
Average launch power (min)	–11.5	–11.5	–11.0	dBm
Average launch power of OFF transmitter (max)	–30			dBm
Extinction ratio (min)	9			dB
RIN (max)	–120			dB/Hz
Coupled Power Ratio (CPR)[a]	28 < CPR< 40	12 < CPR < 20	N/A	dB

[a]Due to the dual media (single-mode and multimode) support of the LX transmitter, fulfillment of this specification requires a single-mode fiber offset-launch mode-conditioning patch cord described in 38.11.4 for MMF operation. This patch cord is not used for single-mode operation.

Conditioned launch (CL) produces sufficient mode volume so that individual multimode fiber (MMF) modes do not dominate fiber performance. This reduces the effect of peak-to-peak differential mode delay (DMD) between the launched mode groups and diminishes the resulting pulse-splitting-induced nulls in the frequency response.

A CL is produced by using a single-mode fiber offset-launch mode-conditioning patch cord, inserted at both ends of a full duplex link, between the optical PMD MDI and the remainder of the link segment. The single-mode fiber offset-launch mode-conditioning patch cord contains a fiber of the same type as the cable (i.e., 62.5 μm or 50 μm fiber) connected to the optical PMD receiver input MDI and a specialized fiber/connector assembly connected to the optical PMD transmitter output.

38.4.2 Receive optical specifications

The 1000BASE-LX receiver shall meet the specifications defined in Table 38–8 per measurement techniques defined in 38.6. The sampling instant is defined to occur at the eye center. The receive sensitivity includes the extinction ratio penalty.

Table 38–8—1000BASE-LX receive characteristics

Description	Value	Unit
Signaling speed (range)	1.25 ± 100 ppm	GBd
Wavelength (range)	1270 to 1355	nm
Average receive power (max)	–3	dBm
Receive sensitivity	–19	dBm
Return loss (min)	12	dB
Stressed receive sensitivity[a, b]	–14.4	dBm
Vertical eye-closure penalty[c]	2.60	dB
Receive electrical 3 dB upper cutoff frequency (max)	1500	MHz

[a]Measured with conformance test signal at TP3 (see 38.6.11) for BER = 10^{-12} at the eye center.
[b]Measured with a transmit signal having a 9 dB extinction ratio. If another extinction ratio is used, the stressed receive sensitivity should be corrected for the extinction ratio penalty.
[c]Vertical eye-closure penalty is a test condition for measuring stressed receive sensitivity. It is not a required characteristic of the receiver.

38.4.3 Worst-case 1000BASE-LX link power budget and penalties (informative)

The worst-case power budget and link penalties for a 1000BASE-LX channel are shown in Table 38–9.

Table 38–9—Worst-case 1000BASE-LX link power budget and penalties[a]

Parameter	62.5 μm MMF	50 μm MMF		10 μm SMF	Unit
Modal bandwidth as measured at 1300 nm (minimum, overfilled launch)	500	400	500	N/A	MHz · km
Link power budget	7.5	7.5	7.5	8.0	dB
Operating distance	550	550	550	5000	m
Channel insertion loss[b, c]	2.35	2.35	2.35	4.57	dB
Link power penalties[c]	3.48	5.08	3.96	3.27	dB
Unallocated margin in link power budget[c]	1.67	0.07	1.19	0.16	dB

[a]Link penalties are used for link budget calculations. They are not requirements and are not meant to be tested.
[b]Operating distances used to calculate the channel insertion loss (see 1.4) are the maximum values specified in Table 38–6.
[c]A wavelength of 1270 nm is used to calculate channel insertion loss, link power penalties, and unallocated margin.

38.5 Jitter specifications for 1000BASE-SX and 1000BASE-LX

Numbers in the Table 38–10 represent high-frequency jitter (above 637 kHz) and do not include low-frequency jitter or wander. Implementations shall conform to the normative values highlighted **in bold** in Table 38–10 (see measurement procedure in 38.6.8). All other values are informative.

Table 38–10—1000BASE-SX and 1000BASE-LX jitter budget

Compliance point	Total jitter[a]		Deterministic jitter	
	UI	ps	UI	ps
TP1	**0.240**	**192**	0.100	80
TP1 to TP2	0.284	227	0.100	80
TP2	**0.431**	**345**	0.200	160
TP2 to TP3	0.170	136	0.050	40
TP3	**0.510**	**408**	0.250	200
TP3 to TP4	0.332	266	0.212	170
TP4[b]	**0.749**	**599**	0.462	370

[a] Total jitter is composed of both deterministic and random components. The allowed random jitter equals the allowed total jitter minus the actual deterministic jitter at that point.
[b]Measured with a conformance test signal at TP3 (see 38.6.11) set to an average optical power 0.5 dB greater than the stressed receive sensitivity from Table 38–4 for 1000BASE-SX and Table 38–8 for 1000BASE-LX.

38.6 Optical measurement requirements

All optical measurements shall be made through a short patch cable, between 2 and 5 m in length. If a single-mode fiber offset-launch mode-conditioning patch cord is used, the optical transmit signal is defined at the output end (TP2) of the single-mode fiber offset-launch mode-conditioning patch cord.

38.6.1 Center wavelength and spectral width measurements

The center wavelength and spectral width (RMS) shall be measured using an optical spectrum analyzer per ANSI/EIA/TIA-455-127-1991 [B8]. Center wavelength and spectral width shall be measured under modulated conditions using a valid 1000BASE-X signal.

38.6.2 Optical power measurements

Optical power shall be measured using the methods specified in ANSI/EIA-455-95-1986 [B7]. This measurement may be made with the node transmitting any valid encoded 8B/10B data stream.

38.6.3 Extinction ratio measurements

Extinction ratio shall be measured using the methods specified in ANSI/TIA/EIA-526-4A-1997 [B13]. This measurement may be made with the node transmitting a data pattern defined in 36A.2. This is a repeating K28.7 data pattern. The extinction ratio is measured under fully modulated conditions with worst-case reflections.

NOTE—A repeating K28.7 data pattern generates a 125 MHz square wave.

38.6.4 Relative Intensity Noise (RIN)

RIN shall be measured according to ANSI X3.230-1994 [B20] (FC-PH), Annex A, A.5, *Relative intensity noise (RIN) measuring procedure*. Per this FC-PH annex, "This procedure describes a component test which may not be appropriate for a system level test depending on the implementation." RIN is referred to as RIN_{12} in the referenced standard. For multimode fiber measurements, the polarization rotator referenced in ANSI X3.230-1994 should be omitted, and the single-mode fiber should be replaced with a multimode fiber.

38.6.5 Transmitter optical waveform (transmit eye)

The required transmitter pulse shape characteristics are specified in the form of a mask of the transmitter eye diagram as shown in Figure 38–2. The transmit mask is not used for response time and jitter specification.

Normalized amplitudes of 0.0 and 1.0 represent the amplitudes of logic ZERO and ONE, respectively.

The eye shall be measured with respect to the mask of the eye using a fourth-order Bessel-Thomson filter having a transfer function given by

$$H(p) = \frac{105}{105 + 105y + 45y^2 + 10y^3 + y^4}$$

where

$$y = 2.114p\,; \quad p = \frac{j\omega}{\omega_r}\,; \quad \omega_r = 2\pi f_r\,; \quad f_r = 0.9375\text{GHz}$$

and where the filter response vs. frequency range for this fourth order Bessel-Thomson filter is defined in ITU-T G.957 , along with the allowed tolerances for its physical implementation.

NOTE 1—This Bessel-Thomson filter is not intended to represent the noise filter used within an optical receiver, but is intended to provide uniform measurement conditions at the transmitter.

NOTE 2—The fourth-order Bessel-Thomson filter is reactive. In order to suppress reflections, a 6 dB attenuator may be required at the filter input and/or output.

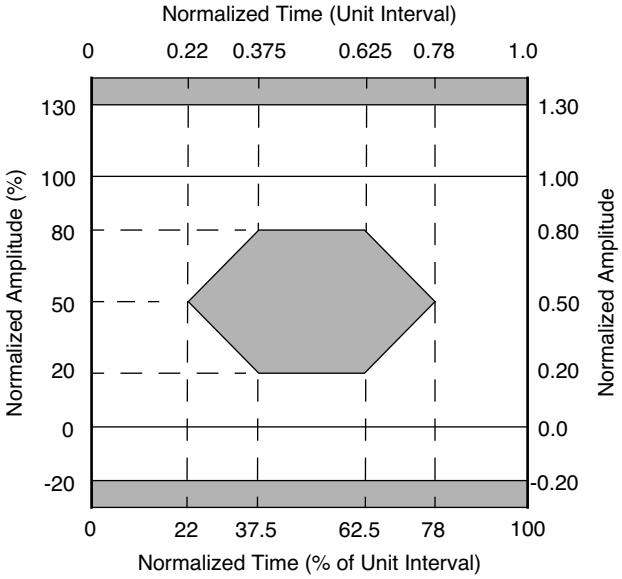


Figure 38–2—Transmitter eye mask definition

38.6.6 Transmit rise/fall characteristics

Optical response time specifications are based on unfiltered waveforms. Some lasers have overshoot and ringing on the optical waveforms which, if unfiltered, reduce the accuracy of the measured 20–80% response times. For the purpose of standardizing the measurement method, measured waveforms shall conform to the mask defined in Figure 38–2. If a filter is needed to conform to the mask, the filter response should be removed using the equation:

$$T_{rise,fall} = \sqrt{(T_{rise,fall\_measured})^2 - (T_{rise,fall\_filter})^2}$$

where the filter may be different for rise and fall. Any filter should have an impulse response equivalent to a fourth order Bessel-Thomson filter. The fourth-order Bessel-Thomson filter defined in 38.6.5 may be a convenient filter for this measurement; however, its low bandwidth adversely impacts the accuracy of the $T_{rise,fall}$ measurements.

38.6.7 Receive sensitivity measurements

The receive sensitivity shall be measured using a worst-case extinction ratio penalty while sampling at the eye center.

The stressed receive sensitivity shall be measured using the conformance test signal at TP3, as specified in 38.6.11. After correcting for the extinction ratio of the source, the stressed receive sensitivity shall meet the conditions specified in Table 38–4 for 1000BASE-SX and in Table 38–8 for 1000BASE-LX.

38.6.8 Total jitter measurements

All total jitter measurements shall be made according to the method in ANSI X3.230-1994 [B20] (FC-PH), Annex A, A.4.2, *Active output interface eye opening measurement*. Total jitter at TP2 shall be measured utilizing a BERT (Bit Error Rate Test) test set. References to use of the Bessel-Thomson filter shall substitute use of the Bessel-Thomson filter defined in this clause (see 38.6.5). The test shall utilize the mixed frequency test pattern specified in 36A.3.

Total jitter at TP4 shall be measured using the conformance test signal at TP3, as specified in 38.6.11. The optical power shall be 0.5 dB greater than (to account for eye opening penalty) the stressed receive sensitivity level in Table 38–4 for 1000BASE-SX and in Table 38–8 for 1000BASE-LX. This power level shall be corrected if the extinction ratio differs from the specified extinction ratio (min) of 9 dB. Measurements shall be taken directly at TP4 without additional Bessel-Thomson filters.

Jitter measurement may use a clock recovery unit (commonly referred to in the industry as a “golden PLL”) to remove low-frequency jitter from the measurement as shown in Figure 38–3. The clock recovery unit has a low pass filter with 20 dB/decade rolloff with –3 dB point of 637 kHz. For this measurement, the recovered clock will run at the signaling speed. The golden PLL is used to approximate the PLL in the deserializer function of the PMA. The PMA deserializer is able to track a large amount of low-frequency jitter (such as drift or wander) below its bandwidth. This low-frequency jitter would create a large measurement penalty, but does not affect operation of the link.

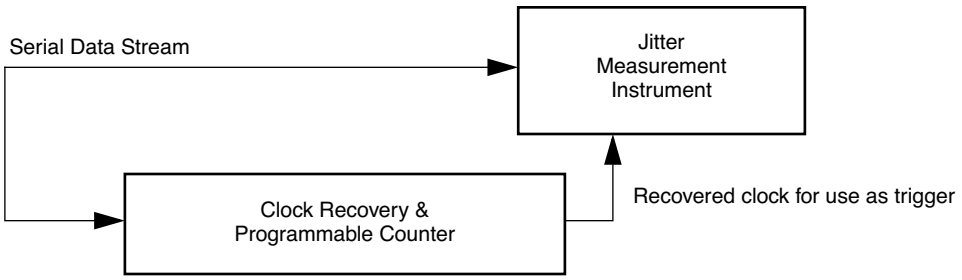


Figure 38–3—Utilization of clock recovery unit during measurement

38.6.9 Deterministic jitter measurement (informative)

Deterministic jitter should be measured according to ANSI X3.230-1994 [B20] (FC-PH), Annex A, A.4.3, *DJ Measurement*. The test utilizes the mixed frequency test pattern specified in 36A.3. This method utilizes a digital sampling scope to measure actual vs. predicted arrival of bit transitions of the 36A.3 data pattern (alternating K28.5 code-groups).

It is convenient to use the clock recovery unit described in 38.6.8 for purposes of generating a trigger for the test equipment. This recovered clock should have a frequency equivalent to 1/20th of the signaling speed.

38.6.10 Coupled Power Ratio (CPR) measurements

Coupled Power Ratio (CPR) is measured in accordance with ANSI/EIA/TIA-526-14A [B14]. Measured CPR values are time averaged to eliminate variation from speckle fluctuations. The coupled power ratio shall be measured for compliance with Table 38–3 and Table 38–7.

38.6.11 Conformance test signal at TP3 for receiver testing

Receivers being tested for conformance to the stressed receive sensitivity requirements of 38.6.7 and the total jitter requirements of 38.6.8 shall be tested using a conformance test signal at TP3 conforming to the requirements described in Figure 38–4. The conformance test signal shall be generated using the short continuous random test pattern defined in 36A.5. The conformance test signal is conditioned by applying deterministic jitter (DJ) and intersymbol interference (ISI). The conditioned conformance test signal is shown schematically in Figure 38–4. The horizontal eye closure (reduction of pulse width) caused by the duty cycle distortion (DCD) component of DJ shall be no less than 65 ps. The vertical eye-closure penalty shall be greater than or equal to the value specified in Table 38–4 for 1000BASE-SX and Table 38–8 for 1000BASE-LX. The DJ cannot be added with a simple phase modulation, which does not account for the DCD component of DJ.

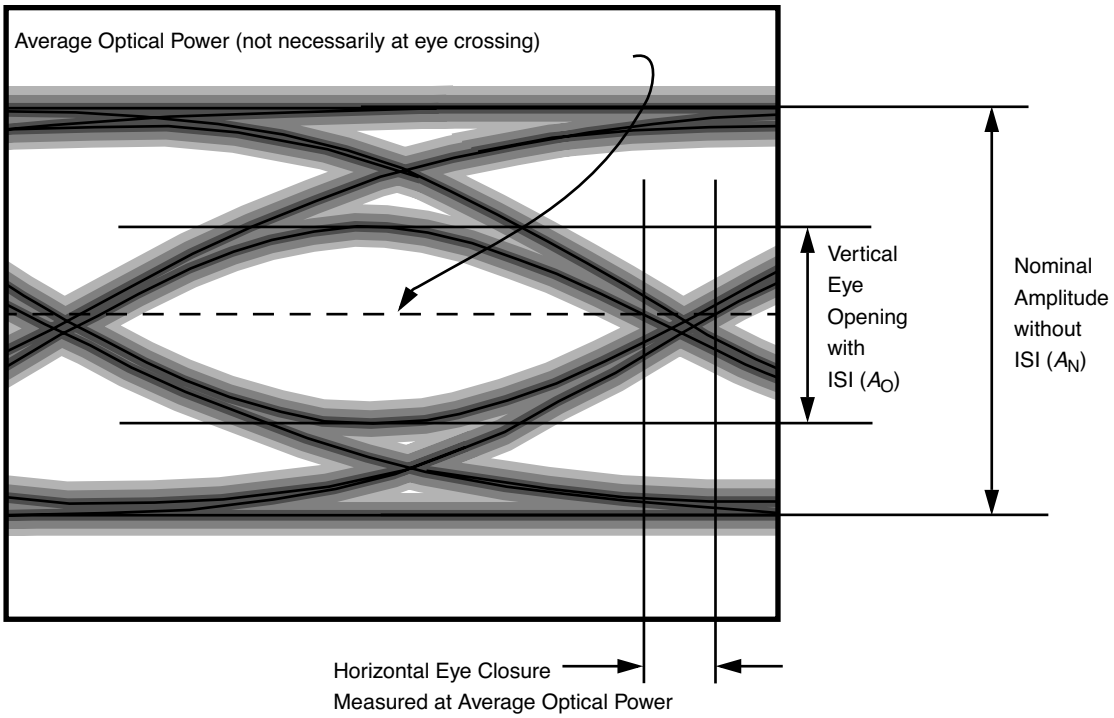


Figure 38–4—Required characteristics of the conformance test signal at TP3

The vertical eye-closure penalty is given by

$$\text{Vertical eye-closure penalty [dB]} = 10 \cdot \log \frac{A_O}{A_N}$$

where A_O is the amplitude of the eye opening, and A_N is the normal amplitude without ISI, as measured in Figure 38–4.

Figure 38–5 shows the recommended test set up for producing the conformance test signal at TP3. The coaxial cable is adjusted in length to produce the correct DCD component of DJ. Since the coaxial cable can produce the incorrect ISI, a limiting amplifier is used to restore fast rise and fall times. A Bessel-Thomson filter is selected to produce the minimum ISI induced eye closure as specified per Table 38–4 for 1000BASE-SX and Table 38–8 for 1000BASE-LX. This conditioned signal is used to drive a high bandwidth linearly modulated laser source.

The vertical and horizontal eye closures to be used for receiver conformance testing are verified using a fast photodetector and amplifier. The bandwidth of the photodetector shall be at least 2.5 GHz and be coupled through a 1.875 GHz fourth-order Bessel-Thomson filter to the oscilloscope input. Special care should be taken to ensure that all the light from the fiber is collected by the fast photodetector and that there is negligible mode selective loss, especially in the optical attenuator.

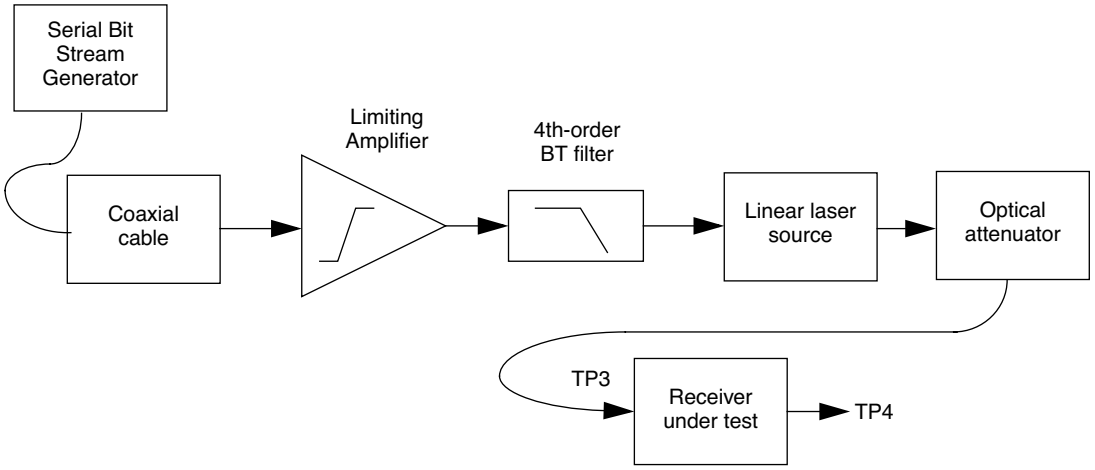


Figure 38–5—Apparatus for generating receiver conformance test signal at TP3

38.6.12 Measurement of the receiver 3 dB electrical upper cutoff frequency

The receiver 3 dB electrical upper cutoff frequency shall be measured as described below. The test setup is shown in Figure 38–6. The test is performed with a laser that is suitable for analog signal transmission. The laser is modulated by a digital data signal. In addition to the digital modulation, the laser is modulated with an analog signal. The analog and digital signals should be asynchronous. The data pattern to be used for this test is the short continuous random test pattern defined in 36A.5. The frequency response of the laser must be sufficient to allow it to respond to both the digital modulation and the analog modulation. The laser should be biased so that it remains linear when driven by the combined signals.

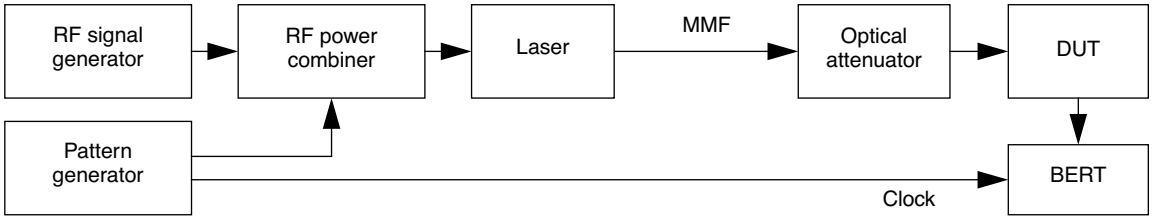


Figure 38–6—Test setup for receiver bandwidth measurement

The 3 dB upper cutoff frequency is measured using the following steps a) through e):

a) Calibrate the frequency response characteristics of the test equipment including the analog radio frequency (RF) signal generator, RF power combiner, and laser source. Measure the laser's extinction ratio according to 38.6.3. With the exception of extinction ratio, the optical source shall meet the requirements of Clause 38.
b) Configure the test equipment as shown in Figure 38–6. Take care to minimize changes to the signal path that could affect the system frequency response after the calibration in step a. Connect the laser

output with no RF modulation applied to the receiver under test through an optical attenuator and taking into account the extinction ratio of the source, set the optical power to a level that approximates the stressed receive sensitivity level in Table 38–4 for 1000BASE-SX and in Table 38–8 for 1000BASE-LX.

c) Locate the center of the eye with the BERT. Turn on the RF modulation while maintaining the same average optical power established in step b.
d) Measure the necessary RF modulation amplitude (in dBm) required to achieve a constant BER (e.g., 10^{-8}) for a number of frequencies.
e) The receiver 3 dB electrical upper cutoff frequency is that frequency where the corrected RF modulation amplitude (the measured amplitude in "d" corrected with the calibration data in "a") increases by 3 dB (electrical). If necessary, interpolate between the measured response values.

38.7 Environmental specifications

38.7.1 General safety

All equipment meeting this standard shall conform to IEC-60950: 1991.

38.7.2 Laser safety

1000BASE-X optical transceivers shall be Class 1 laser certified under any condition of operation. This includes single fault conditions whether coupled into a fiber or out of an open bore. Transceivers shall be certified to be in conformance with IEC 60825-1.

Conformance to additional laser safety standards may be required for operation within specific geographic regions.

Laser safety standards and regulations require that the manufacturer of a laser product provide information about the product's laser, safety features, labeling, use, maintenance and service. This documentation shall explicitly define requirements and usage restrictions on the host system necessary to meet these safety certifications.[5]

38.7.3 Installation

Sound installation practice, as defined by applicable local codes and regulations, shall be followed in every instance in which such practice is applicable.

38.8 Environment

Normative specifications in this clause shall be met by a system integrating a 1000BASE-X PMD over the life of the product while the product operates within the manufacturer's range of environmental, power, and other specifications.

It is recommended that manufacturers indicate in the literature associated with the PHY the operating environmental conditions to facilitate selection, installation, and maintenance.

It is recommended that manufacturers indicate, in the literature associated with the components of the optical link, the distance and operating environmental conditions over which the specifications of this clause will be met.

[5]A host system that fails to meet the manufacturers requirements and/or usage restrictions may emit laser radiation in excess of the safety limits of one or more safety standards. In such a case, the host manufacturer is required to obtain its own laser safety certification.

38.8.1 Electromagnetic emission

A system integrating a 1000BASE-X PMD shall comply with applicable local and national codes for the limitation of electromagnetic interference.

38.8.2 Temperature, humidity, and handling

The optical link is expected to operate over a reasonable range of environmental conditions related to temperature, humidity, and physical handling (such as shock and vibration). Specific requirements and values for these parameters are considered to be beyond the scope of this standard.

38.9 PMD labeling requirements

It is recommended that each PHY (and supporting documentation) be labeled in a manner visible to the user with at least the following parameters, according to the PMD-MDI type.

PMD MDI type 1000BASE-SX:

a) 1000BASE-SX
b) Applicable safety warnings

PMD MDI type 1000BASE-LX:

a) 1000BASE-LX
b) Applicable safety warnings

Labeling requirements for Class 1 lasers are given in the laser safety standards referenced in 38.7.2.

38.10 Fiber optic cabling model

The fiber optic cabling model is shown in Figure 38–7.

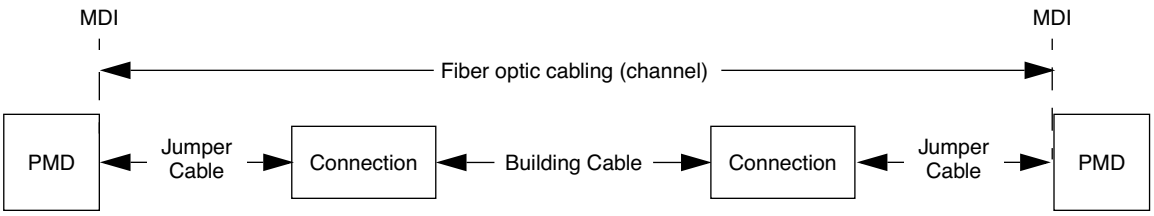


Figure 38–7—Fiber optic cabling model

The channel insertion loss is given in Table 38–11. Insertion loss measurements of installed fiber cables are made in accordance with ANSI/TIA/EIA-526-14A [B14], method B; and ANSI/TIA/EIA-526-7 [B15], method A-1. The fiber optic cabling model (channel) defined here is the same as a simplex fiber optic link segment. The term channel is used here for consistency with generic cabling standards.

Table 38–11—Channel insertion loss

Description	62.5 μm MMF			50 μm MMF			10 μm SMF	Unit
Wavelength	850	850	1300	850	850	1300	1310	nm
Modal bandwidth (min; overfilled launch)	160	200	500	400	500	400 or 500	N/A	MHz · km
Operating distance	220	275	550	500	550	550	5000	m
Channel insertion loss[a] [b]	2.33	2.53	2.32	3.25	3.43	2.32	4.5	dB

[a]These channel insertion loss numbers are based on the nominal operating wavelength.
[b]Operating distances used to calculate channel insertion loss are those listed in this table.

38.11 Characteristics of the fiber optic cabling

The 1000BASE-SX and 1000BASE-LX fiber optic cabling shall meet the specifications defined in Table 38–12. The fiber optic cabling consists of one or more sections of fiber optic cable and any intermediate connections required to connect sections together. It also includes a connector plug at each end to connect to the MDI. The fiber optic cabling spans from one MDI to another MDI, as shown in Figure 38–7.

38.11.1 Optical fiber and cable

The fiber optic cable requirements are satisfied by the fibers specified in IEC 60793-2:1992. Types A1a (50/125 μm multimode), A1b (62.5/125 μm multimode), and B1 (10/125 μm single-mode) with the exceptions noted in Table 38–12.

Table 38–12—Optical fiber and cable characteristics

Description	62.5 μm MMF		50 μm MMF		10 μm SMF	Unit
Nominal fiber specification wavelength	**850**	**1300**	**850**	**1300**	**1310**	**nm**
Fiber cable attenuation (max)	3.75[a]	1.5	3.5	1.5	0.5	dB/km
Modal Bandwidth (min; overfilled launch)	160	500	400	400	N/A	MHz · km
	200	500	500	500	N/A	MHz · km
Zero dispersion wavelength (λ_0)	$1320 \le \lambda_0 \le 1365$		$1295 \le \lambda_0 \le 1320$		$1300 \le \lambda_0 \le 1324$	nm
Dispersion slope (max) (S_0)	0.11 for $1320 \le \lambda_0 \le 1348$ and $0.001(1458 - \lambda_0)$ for $1348 \le \lambda_0 \le 1365$		0.11 for $1300 \le \lambda_0 \le 1320$ and $0.001(\lambda_0 - 1190)$ for $1295 \le \lambda_0 \le 1300$		0.093	ps / nm^2 · km

[a]This value of attenuation is a relaxation of the standard (IEC 60793-2, type A1b, category less than or equal to 3.5 dB/km).

38.11.2 Optical fiber connection

An optical fiber connection as shown in Figure 38–7 consists of a mated pair of optical connectors. The 1000BASE-SX or 1000BASE-LX PMD is coupled to the fiber optic cabling through a connector plug into the MDI optical receptacle, as shown in 38.11.3.

38.11.2.1 Connection insertion loss

The insertion loss is specified for a connection, which consists of a mated pair of optical connectors.

The maximum link distances for multimode fiber are calculated based on an allocation of 1.5 dB total connection and splice loss. For example, this allocation supports three connections with an average insertion loss equal to 0.5 dB (or less) per connection, or two connections (as shown in Figure 38–7) with a maximum insertion loss of 0.75 dB. Connections with different loss characteristics may be used provided the requirements of Table 38–11 and Table 38–12 are met.

The maximum link distances for single-mode fiber are calculated based on an allocation of 2.0 dB total connection and splice loss. For example, this allocation supports four connections with an average insertion loss per connection of 0.5 dB. Connections with different loss characteristics may be used provided the requirements of Table 38–11 and Table 38–12 are met.

38.11.2.2 Connection return loss

The return loss for multimode connections shall be greater than 20 dB.

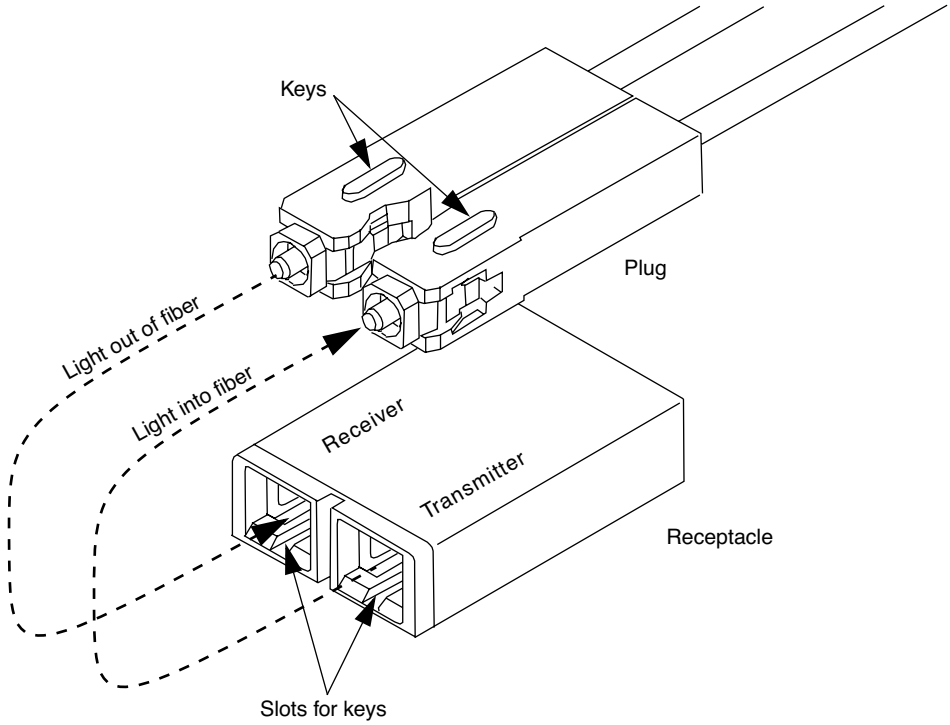
The return loss for single-mode connections shall be greater than 26 dB.

38.11.3 Medium Dependent Interface (MDI)

The 1000BASE-SX and 1000BASE-LX PMD is coupled to the fiber optic cabling through a connector plug into the MDI optical receptacle. The 1000BASE-SX and 1000BASE-LX MDI optical receptacles shall be the duplex SC, meeting the following requirements:

a) Meet the dimension and interface specifications of IEC 61754-4 [B25] and IEC 61754-4, Interface 4-2.
b) Meet the performance specifications as specified in ISO/IEC 11801.
c) Ensure that polarity is maintained.
d) The receive side of the receptacle is located on the left when viewed looking into the transceiver optical ports with the keys on the bottom surface.

A sample drawing of a duplex SC connector and receptacle is provided in Figure 38–8.



NOTE—Connector keys are used for transmit/receive polarity only. The connector keys do not differentiate between single-mode and multimode connectors.

Figure 38–8—Duplex SC connector and receptacle (informative)

38.11.4 single-mode fiber offset-launch mode-conditioning patch cord for MMF operation of 1000BASE-LX

This subclause specifies an example embodiment of a mode conditioner for 1000BASE-LX operation with MMF cabling. The MMF cabling should meet all of the specifications of 38.10. For 1000BASE-LX the mode conditioner consists of a single-mode fiber permanently coupled off-center to a graded index fiber. This example embodiment of a patch cord is not intended to exclude other physical implementations of offset-launch mode conditioners. However, any implementation of an offset-launch mode conditioner used for 1000BASE-LX shall meet the specifications of Table 38–13. The offset launch must be contained within the patch cord assembly and is not adjustable by the user.

Table 38–13—single-mode fiber offset-launch mode conditioner specifications

Description	62.5 μm MMF	50 μm MMF	Unit
Maximum insertion loss	0.5	0.5	dB
Coupled Power Ratio (CPR)	28 < CPR < 40	12 < CPR < 20	dB
Optical center offset between SMF and MMF	17 < Offset < 23	10 < Offset < 16	μm
Maximum angular offset	1	1	degree

All patch cord connecting ferrules containing the single-mode-to-multimode offset launch shall have single-mode tolerances (IEC 61754-4 [B25] grade 1 ferrule).

The single-mode fiber used in the construction of the single-mode fiber offset-launch mode conditioner shall meet the requirements of 38.11.1. The multimode fiber used in the construction of the single-mode fiber offset-launch mode conditioner shall be of the same type as the cabling over which the 1000BASE-LX link is to be operated. If the cabling is 62.5 μm MMF then the MMF used in the construction of the mode conditioner should be of type 62.5 μm MMF. If the cabling is 50 μm MMF, then the MMF used in the construction of the mode conditioner should be of type 50 μm MMF.

Figure 38–9 shows the preferred embodiment of the single-mode fiber offset-launch mode-conditioning patch cord. This patch cord consists of duplex fibers including a single-mode-to-multimode offset launch fiber connected to the transmitter MDI and a second conventional graded index MMF connected to the receiver MDI. The preferred configuration is a plug-to-plug patch cord since it maximizes the power budget margin of the 1000BASE-LX link. The single-mode end of the patch cord shall be labeled “To Equipment.” The multimode end of the patch cord shall be labeled “To Cable.” The color identifier of the single-mode fiber connector shall be blue. The color identifier of all multimode fiber connector plugs shall be beige. The patch cord assembly is labeled “Offset-launch mode-conditioning patch cord assembly.” Labelling identifies which size multimode fiber is used in the construction of the patch cord. The keying of the SC duplex optical plug ensures that the single-mode fiber end is automatically aligned to the transmitter MDI.

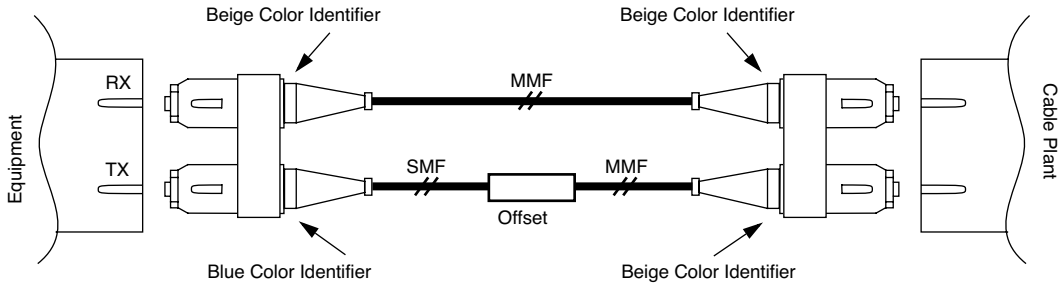


Figure 38–9—1000BASE-LX single-mode fiber offset-launch mode-conditioning patch cord assembly

38.12 Protocol Implementation Conformance Statement (PICS) proforma for Clause 38, Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-LX (Long Wavelength Laser) and 1000BASE-SX (Short Wavelength Laser)[6]

38.12.1 Introduction

The supplier of a protocol implementation that is claimed to conform to Clause 38, Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-LX (Long Wavelength Laser) and 1000BASE-SX (Short Wavelength Laser), shall complete the following Protocol Implementation Conformance Statement (PICS) proforma. A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

38.12.2 Identification

38.12.2.1 Implementation identification

Supplier	
Contact point for enquiries about the PICS	
Implementation Name(s) and Version(s)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Names(s)	
NOTE 1—Only the first three items are required for all implementations; other information may be completed as appropriate in meeting the requirements for the identification. NOTE 2—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

38.12.2.2 Protocol summary

Identification of protocol standard	IEEE Std 802.3-2002®, Clause 38, Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-LX (Long Wavelength Laser) and 1000BASE-SX (Short Wavelength Laser)
Identification of amendments and corrigenda to this PICS proforma that have been completed as part of this PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21; the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®.)	

Date of Statement	

[6]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so that it can be used for its intended purpose and may further publish the completed PICS.

38.12.3 Major capabilities/options

Item	Feature	Subclause	Value/Comment	Status	Support
*LX	1000BASE-LX PMD	38.1	Device supports long wavelength operation (1270–1355 nm).	O/1	Yes [ ] No [ ]
*SX	1000BASE-SX PMD	38.1	Device supports short wavelength operation (770–860 nm). Either this option, or option LX, must be checked.	O/1	Yes [ ] No [ ]
*INS	Installation / cable	38.10	Items marked with INS include installation practices and cable specifications not applicable to a PHY manufacturer.	O	Yes [ ] No [ ]
*OFP	Single-mode offset-launch mode-conditioning patch cord	38.11.4	Items marked with OFP include installation practices and cable specifications not applicable to a PHY manufacturer.	O	Yes [ ] No [ ]
*TP1	Standardized reference point TP1 exposed and available for testing.	38.2.1	This point may be made available for use by implementors to certify component conformance.	O	Yes [ ] No [ ]
*TP4	Standardized reference point TP4 exposed and available for testing.	38.2.1	This point may be made available for use by implementors to certify component conformance.	O	Yes [ ] No [ ]

38.12.4 PICS proforma tables for Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-LX (Long Wavelength Laser) and 1000BASE-SX (Short Wavelength Laser)

38.12.4.1 PMD functional specifications

Item	Feature	Subclause	Value/Comment	Status	Support
FN1	Integration with 1000BASE-X PCS and PMA and management functions	38.1		M	Yes [ ]
FN2	Transmit function	38.2.2	Convey bits requested by PMD_UNITDATA.request() to the MDI	M	Yes [ ]
FN3	Mapping between optical signal and logical signal for transmitter	38.2.2	Higher optical power is a logical 1.	M	Yes [ ]
FN4	Receive function	38.2.3	Convey bits received from the MDI to PMD_UNITDATA.indicate()	M	Yes [ ]
FN5	Mapping between optical signal and logical signal for receiver	38.2.3	Higher optical power is a logical 1.	M	Yes [ ]
FN6	Signal detect function	38.2.4	Report to the PMD service interface the message PMD_SIGNAL.indicate(SIGNAL_DETECT)	M	Yes [ ]
FN7	Signal detect behavior	38.2.4	Meets requirements of Table 38–1	M	Yes [ ]

38.12.4.2 PMD to MDI optical specifications for 1000BASE-SX

Item	Feature	Subclause	Value/Comment	Status	Support
PMS1	Transmitter meets specifications in Table 38–3	38.3.1	Per measurement techniques in 38.6	SX:M	Yes [ ] N/A [ ]
PMS2	Transmitter eye measurement	38.3.1	Per 38.6.5	SX:M	Yes [ ] N/A [ ]
PMS3	Launch power	38.3.1	Lesser of class 1 safety limit per 38.7.2 or maximum receive power in Table 38–4	SX:M	Yes [ ] N/A [ ]
PMS4	Receiver meets specifications in Table 38–4	38.3.2	Per measurement techniques in 38.6	SX:M	Yes [ ] N/A [ ]

38.12.4.3 PMD to MDI optical specifications for 1000BASE-LX

Item	Feature	Subclause	Value/Comment	Status	Support
PML1	Transmitter meets specifications in Table 38–7	38.4.1	Per measurement techniques in 38.6	LX:M	Yes [ ] N/A [ ]
PML2	Transmitter eye measurement	38.4.1	Per 38.6.5	LX:M	Yes [ ] N/A [ ]
PML3	Offset-launch mode-conditioning patch cord	38.4.1	Required for LX multimode operation	LX:M	Yes [ ] N/A [ ]
PML4	Receiver meets specifications in Table 38–8	38.4.2	Per measurement techniques in 38.6	LX:M	Yes [ ] N/A [ ]

38.12.4.4 Jitter specifications

Item	Feature	Subclause	Value/Comment	Status	Support
JT1	Total jitter specification at TP1	38.5	Meets specification of bold entries in Table 38–10	TP1:M	Yes [ ] N/A [ ]
JT2	Total jitter specification at TP2	38.5	Meets specification of bold entries in Table 38–10	M	Yes [ ]
JT3	Total jitter specification at TP3	38.5	Meets specification of bold entries in Table 38–10	INS:M	Yes [ ] N/A [ ]
JT4	Total jitter specification at TP4	38.5	Meets specification of bold entries in Table 38–10	TP4:M	Yes [ ] N/A [ ]

38.12.4.5 Optical measurement requirements

Item	Feature	Subclause	Value/Comment	Status	Support
OR1	Length of patch cord used for measurements	38.6	2 to 5 m	M	Yes [ ]
OR2	Center wavelength and spectral width measurement conditions	38.6.1	Using optical spectrum analyzer per ANSI/EIA/TIA-455-127-1991 [B8]	M	Yes [ ]
OR3	Center wavelength and spectral width measurement conditions	38.6.1	Under modulated conditions using a valid 1000BASE-X signal	M	Yes [ ]
OR4	Optical power measurement conditions	38.6.2	Per ANSI/EIA-455-95-1986 [B7]	M	Yes [ ]
OR5	Extinction ratio measurement conditions	38.6.3	Per ANSI/TIA/EIA-526-4A-1997 [B13] using patch cable per 38.6	M	Yes [ ]
OR6	RIN test methods	38.6.4	ANSI X3.230-1994 [B20] (FC-PH), Annex A, A.5 using patch cable per 38.6	M	Yes [ ]

38.12.4.5 Optical measurement requirements *(continued)*

Item	Feature	Subclause	Value/Comment	Status	Support
OR7	Transmit eye mask measurement conditions	38.6.5	Using fourth-order Bessel-Thomson filter per 38.6.5, using patch cable per 38.6	M	Yes [ ]
OR8	Transmit rise/fall measurement conditions	38.6.6	Waveforms conform to mask in Figure 38–2, measure from 20% to 80%, using patch cable per 38.6	M	Yes [ ]
OR9	Receive sensitivity measurement conditions	38.6.7	Worst-case extinction ratio penalty while sampling at the eye center using patch cable per 38.6	M	Yes [ ]
OR10	Stressed receive sensitivity	38.6.7	Per 38.6.11, using patch cable per 38.6	M	Yes [ ]
OR11	Stressed receive sensitivity	38.6.7	Meet Table 38–4	SX:M	Yes [ ] N/A [ ]
OR12	Stressed receive sensitivity	38.6.7	Meet Table 38–8	LX:M	Yes [ ] N/A [ ]
OR13	Total jitter measurement conditions	38.6.8	ANSI X3.230-1994 [B20] (FC-PH), Annex A, Subclause A.4.2	M	Yes [ ]
OR14	Total jitter measurement conditions at TP2	38.6.8	Using BERT	M	Yes [ ]
OR15	Total jitter measurement conditions at TP2	38.6.8	Using Bessel-Thomson filter defined in 38.6.5	M	Yes [ ]
OR16	Total jitter measurement conditions	38.6.8	Using mixed frequency pattern specified in 36A.3	M	Yes [ ]
OR17	Total jitter measurement conditions at TP4	38.6.8	Using conformance test signal at TP3 (see 38.6.11)	M	Yes [ ]
OR18	Optical power used for total jitter measurement at TP4	38.6.8	0.5 dB greater than stressed receive sensitivity given in Table 38–4 (for SX) or (for LX)	M	Yes [ ]
OR19	Optical power used for total jitter measurement at TP4	38.6.8	Corrected for extinction ratio	M	Yes [ ]
OR20	Total jitter measurement conditions at TP4	38.6.8	Measured without Bessel-Thomson filters	M	Yes [ ]
OR21	Coupled power ratio	38.6.10	Measured using TIA/EIA-526-14A [B14], meets values in Table 38–3 (for SX) or (for LX)	M	Yes [ ]
OR22	Compliance test signal at TP3	38.6.11	Meets requirements of Figure 38–4	M	Yes [ ]
OR23	Compliance test signal at TP3	38.6.11	Pattern specified in 36A.5	M	Yes [ ]
OR24	Compliance test signal at TP3	38.6.11	DJ eye closure no less than 65 ps	M	Yes [ ]

38.12.4.5 Optical measurement requirements *(continued)*

Item	Feature	Subclause	Value/Comment	Status	Support
OR25	Compliance test signal at TP3	38.6.11	Vertical eye-closure penalty meets requirements of Table 38–4	SX:M	Yes [ ] N/A [ ]
OR26	Compliance test signal at TP3	38.6.11	Vertical eye-closure penalty meets requirements of Table 38–8	LX:M	Yes [ ] N/A [ ]
OR27	Compliance test signal at TP3	38.6.11	Bandwidth of photodetector >2.5 GHz, and couple through 4th order Bessel-Thomson filter	M	Yes [ ]
OR28	Receiver electrical cutoff frequency measurement procedure	38.6.12	As described in 38.6.12	M	Yes [ ]
OR29	Optical source used for cutoff frequency measurement	38.6.12	With the exception of extinction ratio, meets requirements of Clause 38	M	Yes [ ]
OR30	Compliance with IEC 60950-1991	38.7.1		M	Yes [ ]
OR31	Laser safety compliance	38.7.2	Class 1	M	Yes [ ]
OR32	Laser safety compliance test conditions	38.7.2	IEC 60825-1	M	Yes [ ]
OR33	Documentation explicitly defines requirements and usage restrictions on the host system necessary to meet after certifications	38.7.2		M	Yes [ ]
OR34	Sound installation practices	38.7.3		INS:M	Yes [ ] N/A [ ]
OR35	Compliance with all requirements over the life of the product	38.8		M	Yes [ ]
OR36	Compliance with applicable local and national codes for the limitation of electromagnetic interference	38.8.1		M	Yes [ ]

38.12.4.6 Characteristics of the fiber optic cabling

Item	Feature	Subclause	Value/Comment	Status	Support
LI1	Fiber optic cabling	38.11	Meets specifications in Table 38–11	INS:M	Yes [ ] N/A [ ]
LI2	Return loss for multimode connections	38.11.2.2	> 20 dB	INS:M	Yes [ ] N/A [ ]
LI3	Return loss for single-mode connections	38.11.2.2	> 26 dB	INS:M	Yes [ ] N/A [ ]
LI4	MDI optical plug	38.11.3	Duplex SC meeting IEC 61754-4, IEC 61754-4: 1997 [B25] Interface 4-2, and ISO/IEC 11801, maintains polarity and ensures orientation.	INS:M	Yes [ ]
LI5	MDI optical receptacle	38.11.3	Duplex SC meeting IEC 61754-4: 1997 [B25] Interface 4-2, and ISO/IEC 11801, maintains polarity and ensures orientation.	M	Yes [ ]
LI6	Offset-launch mode-conditioning patch cord	38.11.4	Meet conditions of Table 38–13	OFP:M	Yes [ ] N/A [ ]
LI7	Single-mode ferrules in offset-launch mode-conditioning patch cords	38.11.4	IEC 61754-4: 1997 [B25] grade 1 ferrule	OFP:M	Yes [ ] N/A [ ]
LI8	Single-mode fiber in offset-launch mode-conditioning patch cords	38.11.4	Per 38.11.1	OFP:M	Yes [ ] N/A [ ]
LI9	Multimode fiber in offset-launch mode-conditioning patch cords	38.11.4	Same type as used in LX cable plant	OFP:M	Yes [ ] N/A [ ]
LI10	Label on single-mode end of offset-launch mode-conditioning patch cords	38.11.4	Labeled “To Equipment”	OFP:M	Yes [ ] N/A [ ]
LI11	Label on multimode end of offset-launch mode-conditioning patch cords	38.11.4	Labeled “To Cable”	OFP:M	Yes [ ] N/A [ ]
LI12	Color identifier of single-mode fiber connector	38.11.4	Blue	OFP:M	Yes [ ] N/A [ ]
LI13	Color identifier of multimode fiber connector	38.11.4	Beige	OFP:M	Yes [ ] N/A [ ]

39. Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX (short-haul copper)

39.1 Overview

This clause specifies the 1000BASE-CX PMD (including MDI) and baseband medium for short-haul copper. In order to form a complete 1000BASE-CX Physical Layer it shall be integrated with the 1000BASE-X PCS of Clause 36 and the PMD of Clause 38, which are hereby incorporated by reference. As such, the 1000BASE-CX PMD shall comply with the PMD service interface specified in 38.1.1.

1000BASE-CX has a minimum operating range of 0.1 to 25 m. Jumper cables, described in 39.4, are used to interconnect 1000BASE-CX PMDs. These cables shall not be concatenated to achieve longer distances. A 1000BASE-CX jumper cable assembly consists of a continuous shielded balanced cable terminated at each end with a polarized shielded plug described in 39.5.1. The jumper cable assembly provides an output signal on contacts R+/R– meeting the requirements shown in Figure 39–5 when a transmit signal compliant with Figures 39–3 and 39–4 is connected to the T+/T– contacts at the near-end MDI connector.

The links described in this clause are applied only to homogenous ground applications such as between devices within a cabinet or rack, or between cabinets interconnected by a common ground return or ground plane. This restriction minimizes safety and interference concerns caused by any voltage differences that could otherwise exist between equipment grounds.

39.2 Functional specifications

The 1000BASE-CX PMD performs three functions, Transmit, Receive, and Signal Status according to the service interface definition in 38.1.

39.2.1 PMD transmit function

The PMD Transmit function shall convey the bits requested by the PMD service interface message PMD_UNITDATA.request(tx_bit) to the MDI according electrical specifications in 39.3.1. The higher output voltage of T+ minus T– (differential voltage) shall correspond to tx_bit = ONE.

39.2.2 PMD receive function

The PMD Receive function shall convey the bits received at the MDI in accordance with the electrical specifications of 39.3.2 to the PMD service interface using the message PMD_UNITDATA.indicate(rx_bit). The higher output voltage of R+ minus R– (differential voltage) shall correspond to rx_bit = ONE.

39.2.3 PMD signal detect function

The PMD Signal Detect function shall report to the PMD service interface, using the message PMD_SIGNAL.indicate(SIGNAL_DETECT), which is signaled continuously. PMD_SIGNAL.indicate is intended to be an indicator of electrical signal presence.

The value of the SIGNAL_DETECT parameter shall be generated according to the conditions defined in Table 39–1. The PMD receiver is not required to verify whether a compliant 1000BASE-X signal is being received. This standard imposes no response time requirements on the generation of the SIGNAL_DETECT parameter.

As an unavoidable consequence of the requirements for the setting of the SIGNAL_DETECT parameter, implementations must provide adequate margin between the input signal level at which the

Table 39–1—SIGNAL_DETECT value definition

Receive Conditions	Signal Detect Value
V_{input}, Receiver < (receiver sensitivity + worst-case local system noise)[a]	FAIL
Minimum differential sensitivity ≤ $V_{input,}$ Receiver ≤ Maximum differential input AND compliant 1000BASE-X signal input	OK
All other conditions	Unspecified

[a]Worst-case local system noise includes all receiver coupled noise sources (NEXT, power supply noise, and any reflected signals). Receive sensitivity is the actual sensitivity of the specific port (as opposed to the minimum differential sensitivity).

SIGNAL_DETECT parameter is set to OK, and the inherent noise level of the PMD due to cross talk, power supply noise, etc.

Various implementations of the Signal Detect function are permitted by this standard, including implementations which generate the SIGNAL_DETECT parameter values in response to the amplitude of the 8B/10B modulation of the electrical signal

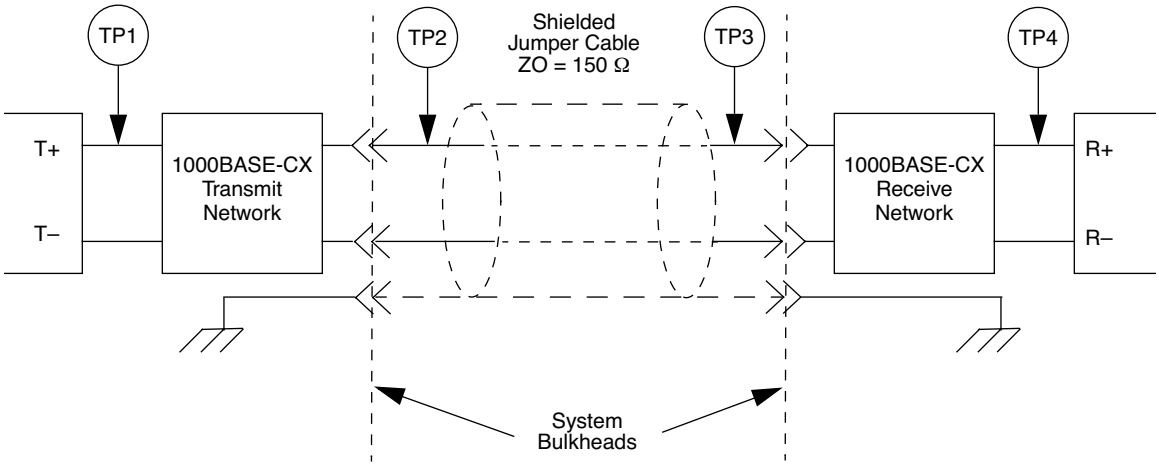
39.3 PMD to MDI electrical specifications

All interface specifications are valid only at the point of entry and exit from the equipment. These points are identified as points TP2 and TP3 as shown in Figure 39–1. The specifications assume that all measurements are made after a mated connector pair, relative to the source or destination.

TP1 and TP4 are standardized reference points for use by implementers to certify component conformance. The electrical specifications of the PMD service interface (TP1 and TP4) are not system compliance points (these are not readily testable in a system implementation). It is expected that in many implementations TP1 and TP4 will be common between 1000BASE-SX (Clause 38), 1000BASE-LX (Clause 38), and 1000BASE-CX.

PMD specifications shall be measured using the measurement techniques defined in 39.6.

The reference points for all connections are those points TP2 and TP3 where the cabinet Faraday shield transitions between the cabinet and the jumper cable shield. If sections of transmission line exist within the Faraday shield, they are considered to be part of the associated transmit or receive network, and not part of the jumper cable assembly.



NOTE—Jumper cable assembly shielding is attached to the system chassis via the connector shroud.

Figure 39–1—1000BASE-CX link (half link is shown)

Schematics in the diagrams in this clause are for illustration only and do not represent the only feasible implementation.

39.3.1 Transmitter electrical specifications

The output driver is assumed to have output levels approximating those of Emitter Coupled Logic (ECL), as measured at TP1. The transmitter shall meet the specifications in Table 39–2.

Table 39–2—Transmitter characteristics at TP2

Description	Value	Unit
Type	(P)ECL	
Data rate	1000	Mb/s
Clock tolerance	±100	ppm
Nominal signalling speed	1250	MBd
Differential amplitude (p-p)		
Max (worst case p-p)	2000	mV
Min (opening)	1100	mV
Max (OFF)[a]	170	mV
Rise/Fall time (20-80%)		
maximum	327	ps
minimum	85	ps
Differential skew (max)	25	ps

[a]Examples of an OFF transmitter are no power supplied to the PMD and PMA transmit output being driven to a static state during loopback.

For all links, the output driver shall be ac-coupled to the jumper cable assembly through a transmission network, and have output levels, measured at the input to the jumper cable assembly (TP2), meeting the eye diagram requirements of Figure 39–3 and Figure 39–4, when terminated as shown in Figure 39–2. The symbols X1 and X2 in Figure 39–3 and Figure 39–4 are defined in Table 39–3.

The normalized amplitude limits in Figure 39–3 are set to allow signal overshoot of 10% and undershoot of 20%, relative to the amplitudes determined to be a logic 1 and 0. The absolute transmitter output timing and

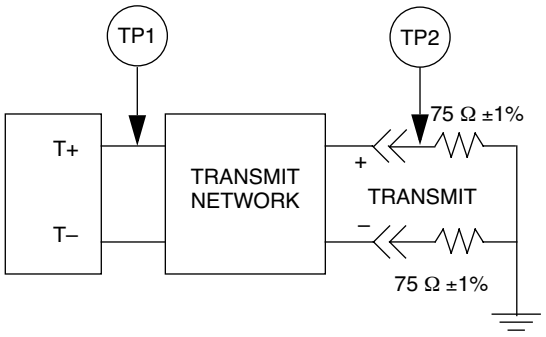


Figure 39–2—Balanced transmitter test load

amplitude requirements are specified in Table 39–2, Table 39–3, and Figure 39–4. The normalized transmitter output timing and amplitude requirements are specified in Table 39–2, Table 39–3, and Figure 39–3. The transmit masks of Figure 39–3 and Figure 39–4 are not used for response time and jitter specifications.

NOTE 1—The relationship between Figure 39–3 and Figure 39–4 can best be explained by a counter example. If a transmitter outputs a nominal 600 mV-ppd logic one level with overshoot to 900 mV-ppd, it will pass the absolute mask of Figure 39–4 but will not pass the normalized mask of Figure 39–3. Normalized, this signal would have 50% overshoot. This exceeds the 10% overshoot limit defined by the normalized eye mask.

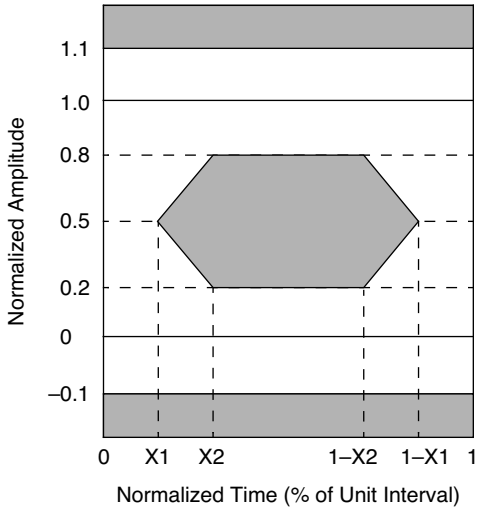


Figure 39–3—Normalized eye diagram mask at TP2

Table 39–3—Normalized time intervals for TP2

Symbol	Value	Units
X1	0.14	Unit Intervals (UI)
X2	0.34	Unit Intervals (UI)

The recommended interface to electrical transmission media is via transformer or capacitive coupling.

NOTE 2—All specifications, unless specifically listed otherwise, are based on differential measurements.

NOTE 3—All times indicated for TDR measurements are recorded times. Recorded times are twice the transit time of the TDR signal.

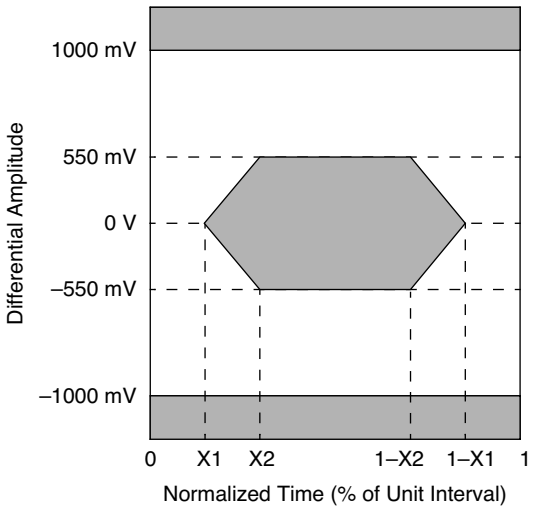


Figure 39–4—Absolute eye diagram mask at TP2

NOTE 4—The transmit differential skew is the maximum allowed time difference (on both low-to-high and high-to low transitions) as measured at TP2, between the true and complement signals. This time difference is measured at the mid-way point on the signal swing of the true and complement signals. These are single-ended measurements.

NOTE 5—The transmitter amplitude maximum specification identifies the maximum p-p signal that can be delivered into a resistive load matching that shown in Figure 39–2.

NOTE 6—The transmitter amplitude minimum specification identifies the minimum allowed p-p eye amplitude opening that can be delivered into a resistive load matching that shown in Figure 39–2.

NOTE 7—The normalized 1 is that amplitude determined to be the average amplitude when driving a logic 1. The normalized 0 is that amplitude determined to be the average amplitude when driving a logic 0.

NOTE 8—Eye diagram assumes the presence of only high-frequency jitter components that are not tracked by the clock recovery circuit. For this standard the lower cutoff frequency for jitter is 637 kHz.

39.3.2 Receiver electrical specifications

The receiver shall be ac-coupled to the media through a receive network located between TP3 and TP4 as shown in Figure 39–1. The receiver shall meet the signal requirements listed in Table 39–4.

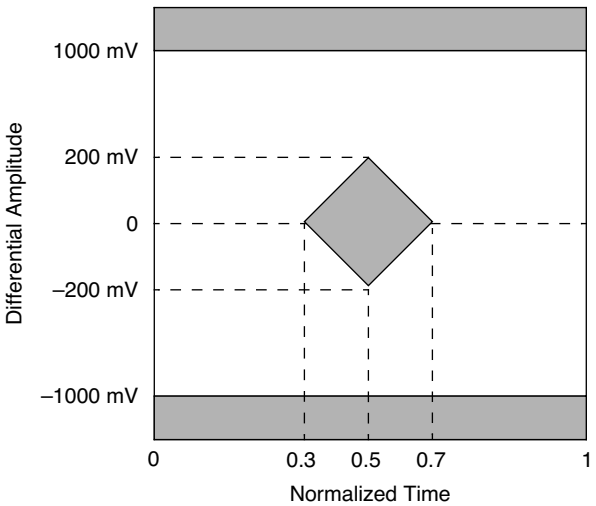


Figure 39–5—Eye diagram mask at point-TP3

Table 39–4—Receiver characteristics (TP3)

Description	Value	Units
Data rate	1000	Mb/s
Nominal signalling speed	1250	MBd
Tolerance	±100	ppm
Minimum differential sensitivity (peak-peak)	400	mV
Maximum differential input (peak-peak)	2000	mV
Input Impedance @ TP3		
TDR Rise Time	85	ps
Exception_window[a]	700	ps
Through_connection	150 ± 30	Ω
At Termination[b]	150 ± 10	Ω
Differential Skew	175	ps

[a]Within the Exception_window no single impedance excursion shall exceed the Through_Connection-impedance tolerance for a period of twice the TDR rise time specification.

[b]The input impedance at TP3, for the termination, shall be recorded 4.0 ns following the reference location determined by an open connector between TP3 and TP4.

The minimum input amplitude to the receiver listed in Table 39–4 and Figure 39–5 is a worst case specification across all environmental conditions. Restricted environments may allow operation at lower minimum differential voltages, allowing significantly longer operating distances.

NOTE 1—All specifications, unless specifically listed otherwise, are based on differential measurements.

NOTE 2—The receiver minimum differential sensitivity identifies the minimum p-p eye amplitude at TP3 to meet the BER objective.

NOTE 3—Eye diagrams assume the presence of only high-frequency jitter components that are not tracked by the clock recovery circuit. For this standard the lower cutoff frequency for jitter is 637 kHz.

NOTE 4—All times indicated for TDR measurements are recorded times. Recorded times are twice the transit time of the TDR signal.

NOTE 5—Through_Connection impedance describes the impedance tolerance through a mated connector. This tolerance is greater than the termination or cable impedance due to limits in the technology of the connectors.

39.3.3 Jitter specifications for 1000BASE-CX

The 1000BASE-CX PMD shall meet the total jitter specifications defined in Table 38–10. Normative values are highlighted **in bold.** All other values are informative. Compliance points are defined in 39.3.

Jitter shall be measured as defined in 38.6.8 with the exception that no measurement will require the use of an optical to electrical converter (O/E).

Deterministic jitter budgetary specifications are included here to assist implementers in specifying components. Measurements for DJ are described in 38.6.9 with the exception that no measurement will require the use of an O/E.

Table 39–5—1000BASE-CX jitter budget

Compliance point	Total jitter[a]		Deterministic jitter	
	UI	ps	UI	ps
TP1	**0.240**	**192**	0.120	96
TP1 to TP2	0.090	72	0.020	16
TP2	**0.279**	**223**	0.140	112
TP2 to TP3	0.480	384	0.260	208
TP3	**0.660**	**528**	0.400	320
TP3 to TP4	0.050	40	0.050	40
TP4	**0.710**	**568**	0.450	360

[a]Total jitter is composed of both deterministic and random components. The allowed random jitter equals the allowed total jitter minus the actual deterministic jitter at that point.

39.4 Jumper cable assembly characteristics

A 1000BASE-CX compliant jumper cable assembly shall consist of a continuous shielded balanced cable terminated at each end with a polarized shielded plug as described in 39.5.1. The jumper cable assembly shall provide an output signal on contacts R+/R– meeting the requirements shown in Figure 39–5 when a transmit signal compliant with Figures 39–3 and 39–4 is connected to the T+/T– contacts at the near-end MDI connector. This jumper cable assembly shall have electrical and performance characteristics as described in Table 39–6. Jumper cable assembly specifications shall be measured using the measurement techniques defined in 39.6. The jumper cable assembly may have integrated compensation networks.

NOTE 1—Jumper cable assemblies that meet the requirements for ANSI X3.230-1994 [B20] (FC-PH) may not meet the requirements of this clause.

NOTE 2—Through_Connection impedance describes the impedance tolerance through a mated connector. This tolerance is greater than the termination or cable impedance due to limits in the technology of the connectors.

To produce jumper cable assemblies capable of delivering signals compliant with the requirements of 39.4, the assemblies should generally have characteristics equal to or better than those in Table 39–7.

39.4.1 Compensation networks

A jumper cable assembly may include an equalizer network to meet the specifications and signal quality requirements (e.g., receiver eye mask at TP3) of this clause. The equalizer shall need no adjustment. All jumper cable assemblies containing such circuits shall be marked with information identifying the specific designed operational characteristics of the jumper cable assembly.

39.4.2 Shielding

The jumper cable assembly shall provide class 2 or better shielding in accordance with IEC 61196-1.

39.5 MDI specification

This clause defines the Media Dependent Interface (MDI). The 1000BASE-CX PMD of 39.3 is coupled to the jumper cable assembly by the media dependent interface (MDI).

Table 39–6—Jumper cable assembly characteristics (normative)

Description	Value	Unit
Differential skew (max)	150	ps
Link Impedance @ TP2/TP3[a] Through_connection Cable	 150 ± 30 150 ± 10	 W W
TDR rise time Exception_window[b]	85 700	ps ps
Round-trip delay (max)[c]	253 253	bit times ns

[a]The link impedance measurement identifies the impedance mismatches present in the jumper cable assembly when terminated in its characteristic impedance. This measurement includes mated connectors at both ends of the Jumper cable assembly (points TP2 and TP3). The link impedance for the jumper cable assembly, shall be recorded 4.0 ns following the reference location determined by an open connector at TP2 and TP3.

[b]Within the Exception_window no single impedance excursion shall exceed the Through_Connection-impedance tolerance for a period of twice the TDR rise time specification. The Exception_window (used with specific impedance measurements) identifies the maximum time period during which the measured impedance is allowed to exceed the listed impedance tolerance. The maximum excursion within the Exception_window at TP3 shall not exceed ±33% of the nominal cable impedance.

[c]Used in Clause 42. This delay is a budgetary requirement of the upper layers. It is easily met by the jumper cable delay characteristics in this clause.

Table 39–7—Jumper cable assembly characteristics (informative)

Description	Value	Unit
Attenuation (max.) at 625 MHz	8.8	dB
Minimum NEXT loss @ 85 ps Tr (max)	6 24.5	% dB

39.5.1 MDI connectors

Connectors meeting the requirements of 39.5.1.1 (Style-1) and 39.5.1.2 (Style-2) shall be used as the mechanical interface between the PMD of 39.3 and the jumper cable assembly of 39.4. The plug connector shall be used on the jumper cable assembly and the receptacle on the PHY. Style-1 or style-2 connectors may be used as the MDI interface. To limit possible cross-plugging with non-1000BASE-CX interfaces that make use of the Style-1 connector, it is recommended that the Style-2 connector be used as the MDI connector.

39.5.1.1 Style-1 connector specification

The style-1 balanced connector for balanced jumper cable assemblies shall be the 9-pin shielded D-subminiature connector, with the mechanical mating interface defined by IEC 60807-3, having pinouts matching those in Figure 39–6, and the signal quality and electrical requirements of this clause. The style-1 connector pin assignments are shown in Figure 39–6 and Table 39–8.

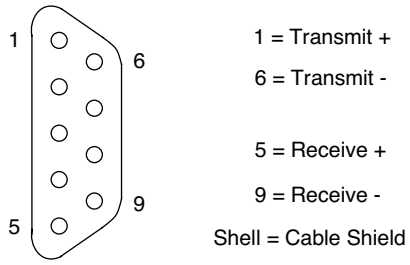


Figure 39–6—Style-1 balanced connector receptacle pin assignments

39.5.1.2 Style-2 connector specification

The style-2 balanced cable connector is the 8-pin shielded ANSI Fibre Channel style-2 connector with the mechanical mating interface defined by IEC 61076-3-103, having pinouts matching those shown in Figure 39–7, and conforming to the signal quality and electrical requirements of this clause.

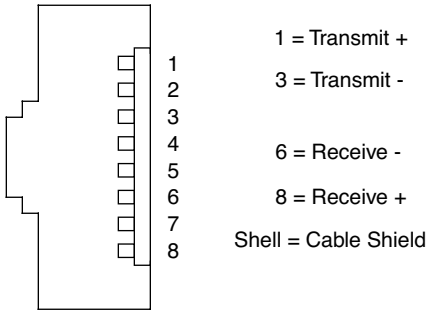


Figure 39–7—Style-2 balanced connector receptacle pin assignments

The style-1 or style-2 connector may be populated with optional contacts to support additional functions. The presence of such contacts in the connector assemblies does not imply support for these additional functions.

NOTE 1—Style-1 pins 2 and 8 (Style-2 pins 7 and 2) are reserved for applications that assign these pins to power and ground.

NOTE 2—Style-1 pin 3 (Style-2 pin 4) is reserved for applications that assign this pin to a Fault Detect function.

NOTE 3—Style-1 pin 7 (Style-2 pin 5) is reserved for applications that assign this pin to an Output Disable function.

Table 39–8—MDI contact assignments

Contact		PMD MDI signal
Style-1	Style-2	
1	1	Transmit +
2	7	Reserved
3	4	Reserved
4		Mechanical key
5	8	Receive +
6	3	Transmit –
7	5	Reserved
8	2	Reserved
9	6	Receive –

39.5.1.3 Style-2 connector example drawing (informative)

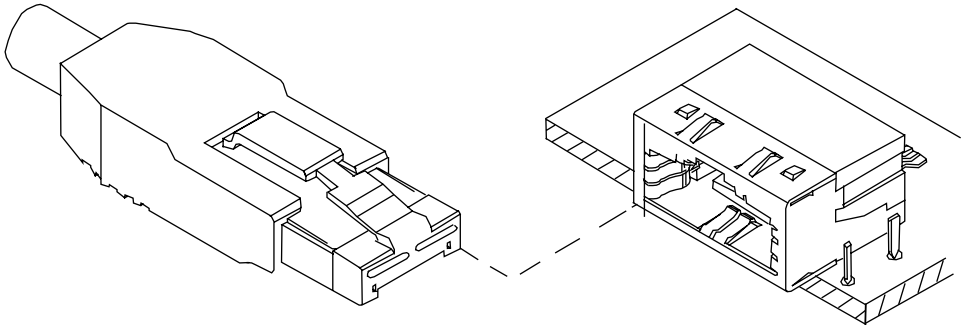


Figure 39–8—Style-2 connector, example drawing

39.5.2 Crossover function

The default jumper cable assembly shall be wired in a crossover fashion as shown in Figure 39–9, with each pair being attached to the transmitter contacts at one end and the receiver contacts at the other end.

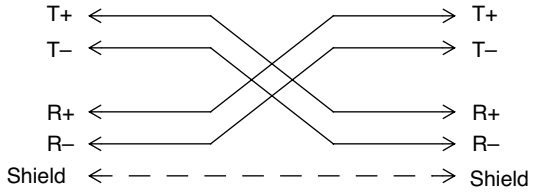


Figure 39–9—Balanced cable wiring

39.6 Electrical measurement requirements

Electrical measurements shall be performed as described in this subclause.

39.6.1 Transmit rise/fall time

Rise time is a differential measurement across the T+ and T– outputs with a load present (including test equipment) equivalent to that shown in Figure 39–2. Both rising and falling edges are measured. The 100% and 0% levels are the normalized 1 and 0 levels present when sending an alternating K28.5 character stream.

Once the normalized amplitude is determined, the data pattern is changed to a continuous D21.5 character stream. The rise time specification is the time interval between the normalized 20% and 80% amplitude levels.

39.6.2 Transmit skew measurement

The transmitter skew is the time difference between the T+ and T– outputs measured at the normalized 50% crossover point with a load present (including test equipment) equivalent to that shown in Figure 39–2. This measurement is taken using two single ended probes. Skew in the test set-up must be calibrated out.

Normalized amplitudes can be determined using the method described in 39.6.1.

A continuous D21.5 or K28.7 data pattern is transmitted by the device under test. The data is averaged using an averaging scope. An easy method to view and measure the skew between these signals is to invert one.

39.6.3 Transmit eye (normalized and absolute)

This test is made as a differential measurement at the bulkhead connector. The scope trigger must either be a recovered clock as defined in 38.6.8 or a character clock internal to the equipment. The data pattern for this is the alternating K28.5.

If a character trigger is used, the overshoot/undershoot percentages must be measured at all ten bit positions. The load for this test is that shown in Figure 39–2.

39.6.4 Through_connection impedance

This is a differential TDR or equivalent measurement that must be made through a mated connector pair or pairs. Any lead-in trace or cable to the connector that is part of the test fixture should provide a reasonable impedance match so as to not effect the actual measurement. All TDR measurements must be filtered to the TDR rise time specification. Any test fixture used with these TDR tests must be calibrated to remove the effects of the test fixture, and verified to produce accurate results.

The impedance Through_connection interval starts at the first point where the measured impedance exceeds the limits for the termination and ends at the point that the impedance returns to within the termination impedance limits and remains there.

Within this Through_connection interval, an Exception_window exists where the impedance is allowed to exceed the Through_connection impedance limits up to a maximum deviation of ±33% of the nominal link impedance. The Exception_window begins at the point where the measured impedance first exceeds the impedance tolerance limits for Through_connection.

39.6.5 Jumper cable intra-pair differential skew

The jumper cable intra-pair differential skew measurement is conducted to determine the skew, or difference in velocity, of each wire in a cable pair when driven with a differential source. This measurement requires two

mated connectors, one at the signal source and one at the opposite end of the cable. A pair of matched, complimentary signals (S+, S–) are driven into the T+ and T– contacts of the connector. These signals are time conditioned to start at the same point. This test shall be performed at both ends of the jumper cable assembly.

The jumper cable intra-pair skew is the time difference between the R+ and R– outputs of the excited pair within the jumper cable assembly measured at the normalized 50% crossover point with a load present (including test equipment) equivalent to that shown in Figure 39–2. This measurement is taken using two single ended probes. Skew in the test set-up must be calibrated out.

Normalized amplitudes can be determined using the method described in 39.6.1.

A continuous square wave is used for S+, S–. The data is averaged using an averaging scope. An easy method to view and measure the skew between these signals is to invert one. A differential TDR can provide a convenient method to time condition the input signals.

39.6.6 Receiver link signal

This differential measurement is made at the end of the jumper cable assembly, through mated connectors with a load present (including test equipment) equivalent to that shown in Figure 39–2. The signal is measured with an alternating K28.5 character stream and is tested to the mask requirements of Figure 39–5.

39.6.7 Near-End Cross Talk (NEXT)

NEXT Loss tests are conducted using a differential TDR (or equivalent) filtered to the rise time limit (near-end cross talk at a maximum T_r of 85 ps) in Table 39–6. The T+ and T– inputs of the jumper cable connector are excited to create a disturber pair while the R+ and R– contacts of the disturbed pair are measured within the same connector. The far-end R+/R– outputs of the disturber pair are terminated per Figure 39–2. The R+ and R– signals of the disturbed pair are terminated with a load (including test equipment) equivalent to that shown in Figure 39–2. The T+ and T– inputs of disturbed pair shall be terminated per Figure 39–2. This test shall be performed at both ends of the jumper cable assembly.

39.6.8 Differential time-domain reflectometry (TDR) measurement procedure

The differential TDR test setup measures the reflected waveform returned from a load when driven with a step input. It is obtained by driving the load under test with a step waveform using a driver with a specified source impedance and rise time. The reflected waveform is the difference between (a) the observed waveform at the device under test when driven with the specified test signal, and (b) the waveform that results when driving a standard test load with the same specified test signal. From this measured result we can infer the impedance of the device under test. The derivative of a time-domain reflectometry measurement is the time-domain equivalent of S_{11} parameter testing used in carrier-based systems.

For the measurement of 1000BASE-CX jumper cables, the following test conditions apply:

a) The driving waveform is sourced from a balanced, differential 150 Ω source with an 85 ps rise time (see 39.6.8.1).
b) The test setup is calibrated (see 39.6.8.2).

39.6.8.1 Driving waveform

If the natural differential output impedance of the driving waveform is not 75 Ω, it may be adjusted to within 75 ± 5 Ω by an attenuating resistive pad. When the driving point resistance is 100 Ω (as would be the case with a differential signal source having two independent, antipodal, 50 Ω sources), a good pad design is shown in Figure 39–10, where R1 = 173.2 Ω and R2 = 43.3 Ω. All resistors are surface-mount packages

soldered directly to the test fixture with no intervening leads or traces, and the whole structure is mounted on a solid ground plane (used in three places).

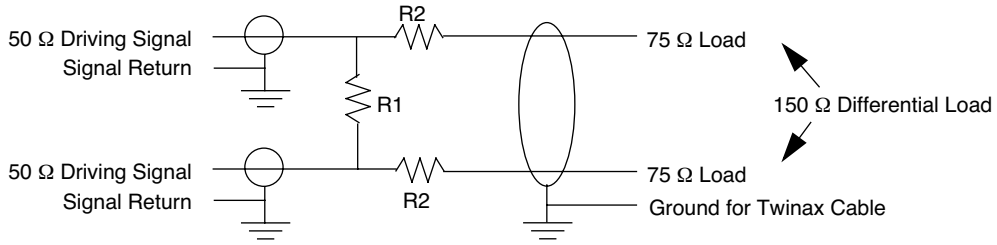


Figure 39–10—Differential TDR pad adapter

If the natural rise time of the driver is less than 85 ps, the resulting measured time-waveforms must be filtered to reduce the apparent rise time to 85 ± 10 ps.

39.6.8.2 Calibration of the test setup

Three measurements are made, with a short, and open, and a known test load. The value of the test resistance should be constant to within 1% over the frequency range dc to 6 GHz, and of known value. The value of the test resistance should be within the range 75 ± 5 Ω.

The differential voltages measured across the device-under-test terminals in these three cases are called V_{short}, V_{open}, and V_{test}, respectively. From these three measurements we will compute three intermediate quantities:

$$A = (V_{open} - V_{short}) /2$$

$$B = (V_{open} + V_{short}) /2$$

$$Z_0 = Z_{test} \cdot (V_{open} - V_{test})/(V_{test} - V_{short})$$

The value of Z_0 is the actual driving point impedance of the tester. It must be within 75 ± 5 Ω.

For any device under test, the conversion from measured voltage $V_{measured}$ to impedance is as follows:

$$\text{Measured impedance} = Z_0 \cdot (1 + V')/(1 - V')$$

where $V' = (V_{measured} - B)/A$.

39.7 Environmental specifications

All equipment subject to this clause shall conform to the requirements of 14.7 and applicable sections of ISO/IEC 11801: 1995. References to the MAU or AUI should be replaced with PHY or DTE and AUI to jumper cable assembly, as appropriate. Subclause 14.7.2.4, *Telephony voltage*, should be ignored. Should a case occur where, through a cabling error, two transmitters or receivers are directly connected, no damage shall occur to any transmitter, receiver, or other link component in the system. The link shall be able to withstand such an invalid connection without component failure or degradation for an indefinite period of time.

Systems connected with 1000BASE-CX links shall meet the bonding requirements (common ground connection) of ISO/IEC 11801: 1995, subclause 9.2, for shielded cable assemblies. Cable shield(s) shall be earthed (chassis ground) through the bulkhead connector shell(s) on both ends of the jumper cable assembly as shown in Figure 39–1.

39.8 Protocol Implementation Conformance Statement (PICS) proforma for Clause 39, Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX[7]

39.8.1 Introduction

The supplier of a protocol implementation that is claimed to conform to Clause 39, Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX, shall complete the following Protocol Implementation Conformance Statement (PICS) proforma. A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

39.8.2 Identification

39.8.2.1 Implementation identification

Supplier	
Contact point for enquiries about the PICS	
Implementation Name(s) and Version(s)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Names(s)	
NOTE 1—Only the first three items are required for all implementations; other information may be completed as appropriate in meeting the requirements for the identification. NOTE 2—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

39.8.2.2 Protocol summary

Identification of protocol standard	IEEE Std 802.3-2002®, Clause 39, Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX
Identification of amendments and corrigenda to this PICS proforma that have been completed as part of this PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21; the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®.)	

Date of Statement	

[7]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so that it can be used for its intended purpose and may further publish the completed PICS.

39.8.3 Major capabilities/options

Item	Feature	Subclause	Value/Comment	Status	Support
*INS	Installation / cable	39.4	Items marked with INS include installation practices and cable specifications not applicable to a PHY manufacturer	O	Yes [ ] No [ ]
*STY1	Style-1 MDI	39.5	Either the style-1 or the style-2 MDI must be provided	O/1	Yes [ ] No [ ]
*STY2	Style-2 MDI	39.5		O/1	Yes [ ] No [ ]
*TP1	Standardized reference point TP1 exposed and available for testing.	39.3	This point may be made available for use by implementors to certify component conformance.	O	Yes [ ] No [ ]
*TP4	Standardized reference point TP4 exposed and available for testing.	39.3	This point may be made available for use by implementors to certify component conformance.	O	Yes [ ] No [ ]

39.8.4 PICS proforma tables for Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX (short-haul copper)

39.8.4.1 PMD functional specifications

Item	Feature	Subclause	Value/Comment	Status	Support
FN1	Integration with 1000BASE-X PCS and PMA	39.1		M	Yes [ ]
FN2	Complies with PMD service interface of 38.2	39.1		M	Yes [ ]
FN3	Jumper cables not concatenated	39.1		INS:M	Yes [ ] N/A [ ]
FN5	Transmit function	39.2.1	Convey bits requested by PMD_UNITDATA.request() to the MDI	M	Yes [ ]
FN6	Transmitter logical to electrical mapping	39.2.1;	Logical one equates to electrical high	M	Yes [ ]
FN7	Receive function	39.2.2	Convey bits received from the MDI to PMD_UNITDATA.indicate()	M	Yes [ ]
FN8	Receiver logical to electrical mapping	39.2.2	Logical one equates to electrical high.	M	Yes [ ]
FN9	Signal detect function	39.2.3	Report to the PMD service interface the message PMD_SIGNAL.indicate(SIGNAL_DETECT)	M	Yes [ ]
FN10	Signal detect behavior	39.2.3	Meets requirements of Table 39–1	M	Yes [ ]

39.8.4.2 PMD to MDI electrical specifications

Item	Feature	Subclause	Value/Comment	Status	Support
PM1	Measurement requirements	39.3	Electrical measurements are made according to the tests specified in 39.6.	M	Yes [ ]
PM2	Transmitter characteristics	39.3.1	Transmitters meets requirements of Table 39–2	M	Yes [ ]
PM3	Transmitter coupling	39.3.1	AC-coupled	M	Yes [ ]
PM4	Transmitter eye diagram	39.3.1	Meets requirements of Figure 39–3 and Figure 39–4 when terminated as shown in Figure 39–2	M	Yes [ ]
PM5	Receiver coupling	39.3.2	AC-coupled	M	Yes [ ]
PM6	Receiver characteristics	39.3.2	Meet requirements of Table 39–4	M	Yes [ ]
PM7	Measurement conditions for input impedance at TP3	39.3.2	4 ns following reference location	M	Yes [ ]
PM8	Total jitter specification at TP1	39.3.3	Meets specification of bold entries in Table 38–10	TP1:M	Yes [ ] N/A [ ]
PM9	Total jitter specification at TP2	39.3.3	Meets specification of bold entries in Table 38–10	M	Yes [ ]
PM10	Total jitter specification at TP3	39.3.3	Meets specification of bold entries in Table 38–10	INS:M	Yes [ ] N/A [ ]
PM11	Total jitter specification at TP4	39.3.3	Meets specification of bold entries in Table 38–10	TP4:M	Yes [ ] N/A [ ]
PM12	Measurement conditions for jitter	39.3.3	Per 38.6.8 (with exceptions)	M	Yes [ ]

39.8.4.3 Jumper cable assembly characteristics

Item	Feature	Subclause	Value/Comment	Status	Support
LI1	Two polarized, shielded plug per 39.5.1 and shielded with electrical characteristics per Table 39–6	39.4	As defined in Table 39–6	INS:M	Yes [ ]
LI2	Delivers compliant signal when driven with worst case source signal	39.4	Transmit signal compliant with Figures 39–3 and 39–4, receive signal complaint with Figure 39–5, into a load compliant with Figure 39–2	INS:M	Yes [ ]
LI3	Measurement requirements	39.4	Electrical measurements are made according to the tests specified in 39.6.	INS:M	Yes [ ]
LI4	Maximum excursion during Exception_window of cable impedance measurement	39.4	± 33% of nominal cable impedance	INS:M	Yes [ ]
LI5	Measurement conditions for link impedance	39.4	4 ns following the reference location between TP3 and TP4	INS:M	Yes [ ]
LI6	Equalizer needs no adjustment	39.4.1		INS:M	Yes [ ] N/A [ ]
LI7	Cables containing equalizers shall be marked	39.4.1		INS:M	Yes [ ] N/A [ ]
LI8	Cable shielding	39.4.2	Class 2 or better per IEC 61196-1	INS:M	Yes [ ]

39.8.4.4 Other requirements

Item	Feature	Subclause	Value/Comment	Status	Support
OR1	Style-1 connector	39.5.1.1	9-pin shielded D-subminiature with the mechanical mating interface defined by IEC 60807-3.	STY1:M	Yes [ ] N/A [ ]
OR2	Style-2 connector	39.5.1.2	8-pin ANSI Fibre Channel style-2 connector with mechanical mating interface defined by IEC 61076-3-103.	STY2:M	Yes [ ] N/A [ ]
OR3	Default cable assembly wired in a crossover assembly	39.5.2		INS:M	Yes [ ]
OR4	Transmit rise/fall time measurement	39.6.1	Meet requirements of Table 39–2 with load equivalent to Figure 39–2	M	Yes [ ]
OR5	Transmit skew measurement	39.6.2	Meet requirements of Table 39–2 with load equivalent to Figure 39–2	M	Yes [ ]
OR6	Transmit eye measurement	39.6.3	Meet requirements of Figure 39–3 and Figure 39–4 with load equivalent to Figure 39–2	M	Yes [ ]
OR7	Through_connection impedance measurement	39.6.4	Meet requirements of Table 39–4 with load equivalent to Figure 39–2	M	Yes [ ]
OR8	Jumper cable assembly differential skew measurement	39.6.5	Meet requirements of Table 39–6 with load equivalent to Figure 39–2	M	Yes [ ]
OR9	Receiver link signal	39.6.6	Meet requirements of Figure 39–5 with load equivalent to Figure 39–2	M	Yes [ ]
OR10	NEXT Loss measurement	39.6.7	Meet requirements of Table 39–6 with load equivalent to Figure 39–2	M	Yes [ ]
OR11	Conformance to 14.7 and applicable sections of ISO/IEC 11801:1995.	39.7		M	Yes [ ]
OR12	Cabling errors shall cause no damage to transmitter, receiver, or other link components	39.7		M	Yes [ ]
OR13	Withstand invalid connection for indefinite period	39.7		M	Yes [ ]
OR14	System meets common ground requirements of ISO/IEC 11801	39.7	Per ISO/IEC 11801, subclause 9.2	INS:M	Yes [ ]
OR15	Cable shields earthed on both ends of cable	39.7		INS:M	Yes [ ]

40. Physical Coding Sublayer (PCS), Physical Medium Attachment (PMA) sublayer and baseband medium, type 1000BASE-T

40.1 Overview

The 1000BASE-T PHY is one of the Gigabit Ethernet family of high-speed CSMA/CD network specifications. The 1000BASE-T Physical Coding Sublayer (PCS), Physical Medium Attachment (PMA) and baseband medium specifications are intended for users who want 1000 Mb/s performance over Category 5 balanced twisted-pair cabling systems. 1000BASE-T signaling requires four pairs of Category 5 balanced cabling, as specified in ISO/IEC 11801:1995 and ANSI/EIA/TIA-568-A (1995) and tested for the additional performance parameters specified in 40.7 using testing procedures defined in proposed ANSI/TIA/EIA TSB95.

This clause defines the type 1000BASE-T PCS, type 1000BASE-T PMA sublayer, and type 1000BASE-T Medium Dependent Interface (MDI). Together, the PCS and the PMA sublayer comprise a 1000BASE-T Physical layer (PHY). Provided in this document are fully functional, electrical, and mechanical specifications for the type 1000BASE-T PCS, PMA, and MDI. This clause also specifies the baseband medium used with 1000BASE-T.

40.1.1 Objectives

The following are the objectives of 1000BASE-T:

a) Support the CSMA/CD MAC
b) Comply with the specifications for the GMII (Clause 35)
c) Support the 1000Mb/s repeater (Clause 41)
d) Provide line transmission that supports full and half duplex operation
e) Meet or exceed FCC Class A/CISPR or better operation
f) Support operation over 100 meters of Category 5 balanced cabling as defined in 40.7
g) Bit Error Rate of less than or equal to 10^{-10}
h) Support Auto-Negotiation (Clause 28)

40.1.2 Relationship of 1000BASE-T to other standards

Relations between the 1000BASE-T PHY, the ISO Open Systems Interconnection (OSI) Reference Model, and the IEEE 802.3® CSMA/CD LAN Model are shown in Figure 40–1. The PHY sub-layers (shown shaded) in Figure 40–1 connect one Clause 4 Media Access Control (MAC) layer to the medium.

40.1.3 Operation of 1000BASE-T

The 1000BASE-T PHY employs full duplex baseband transmission over four pairs of Category 5 balanced cabling. The aggregate data rate of 1000 Mb/s is achieved by transmission at a data rate of 250 Mb/s over each wire pair, as shown in Figure 40–2. The use of hybrids and cancellers enables full duplex transmission by allowing symbols to be transmitted and received on the same wire pairs at the same time. Baseband signaling with a modulation rate of 125 Mbaud is used on each of the wire pairs. The transmitted symbols are selected from a four-dimensional 5-level symbol constellation. Each four-dimensional symbol can be viewed as a 4-tuple (A_n, B_n, C_n, D_n) of one-dimensional quinary symbols taken from the set {2, 1, 0, –1, –2}. 1000BASE-T uses a continuous signaling system; in the absence of data, Idle symbols are transmitted. Idle mode is a subset of code-groups in that each symbol is restricted to the set {2, 0, –2}to improve synchronization. Five-level Pulse Amplitude Modulation (PAM5) is employed for transmission over each wire pair. The modulation rate of 125 MBaud matches the GMII clock rate of 125 MHz and results in a symbol period of 8 ns.

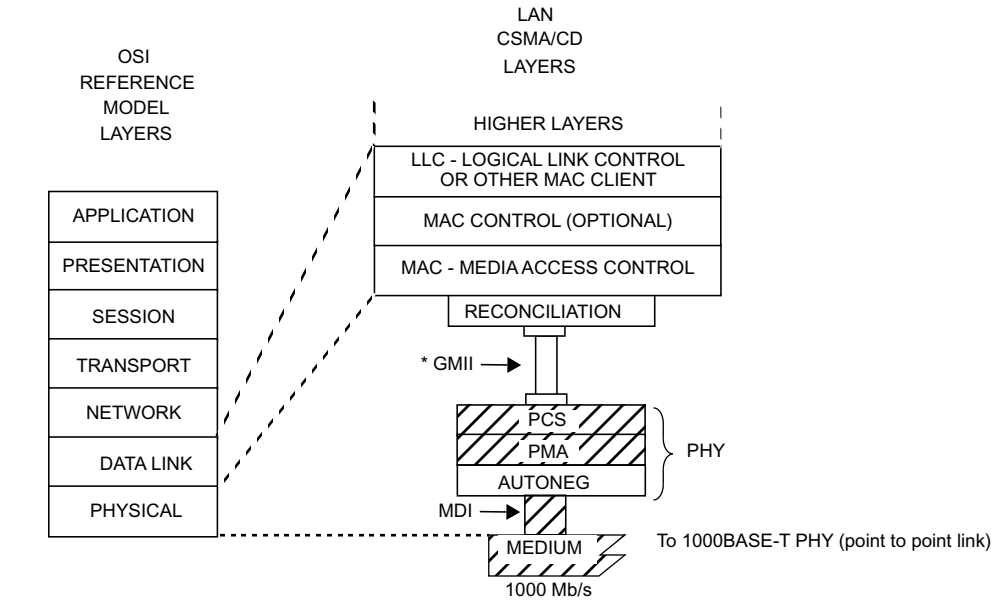


Figure 40–1—Type 1000BASE-T PHY relationship to the ISO Open Systems Interconnection (OSI) Reference Model and the IEEE 802.3® CSMA/CD LAN Model

A 1000BASE-T PHY can be configured either as a MASTER PHY or as a SLAVE PHY. The MASTER-SLAVE relationship between two stations sharing a link segment is established during Auto-Negotiation (see Clause 28, 40.5, and Annex 28C). The MASTER PHY uses a local clock to determine the timing of transmitter operations. The SLAVE PHY recovers the clock from the received signal and uses it to determine the timing of transmitter operations, i.e., it performs loop timing, as illustrated in Figure 40–3. In a multiport to single-port connection, the multiport device is typically set to be MASTER and the single-port device is set to be SLAVE.

The PCS and PMA subclauses of this document are summarized in 40.1.3.1 and 40.1.3.2. Figure 40–3 shows the functional block diagram.

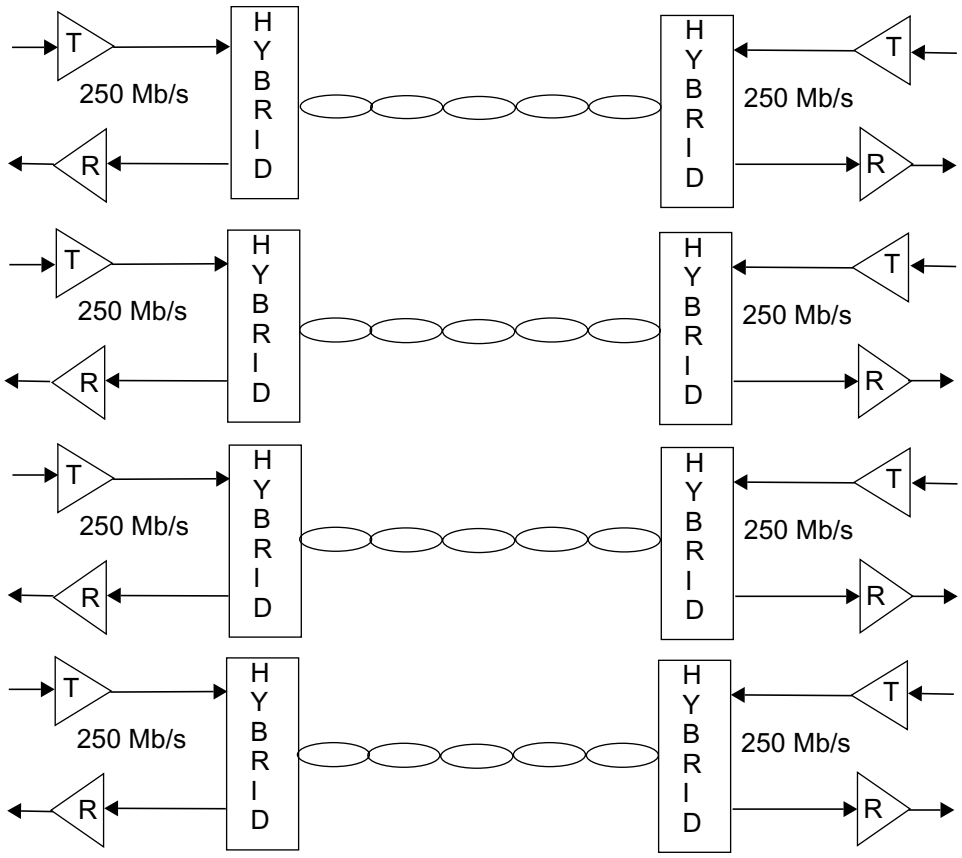
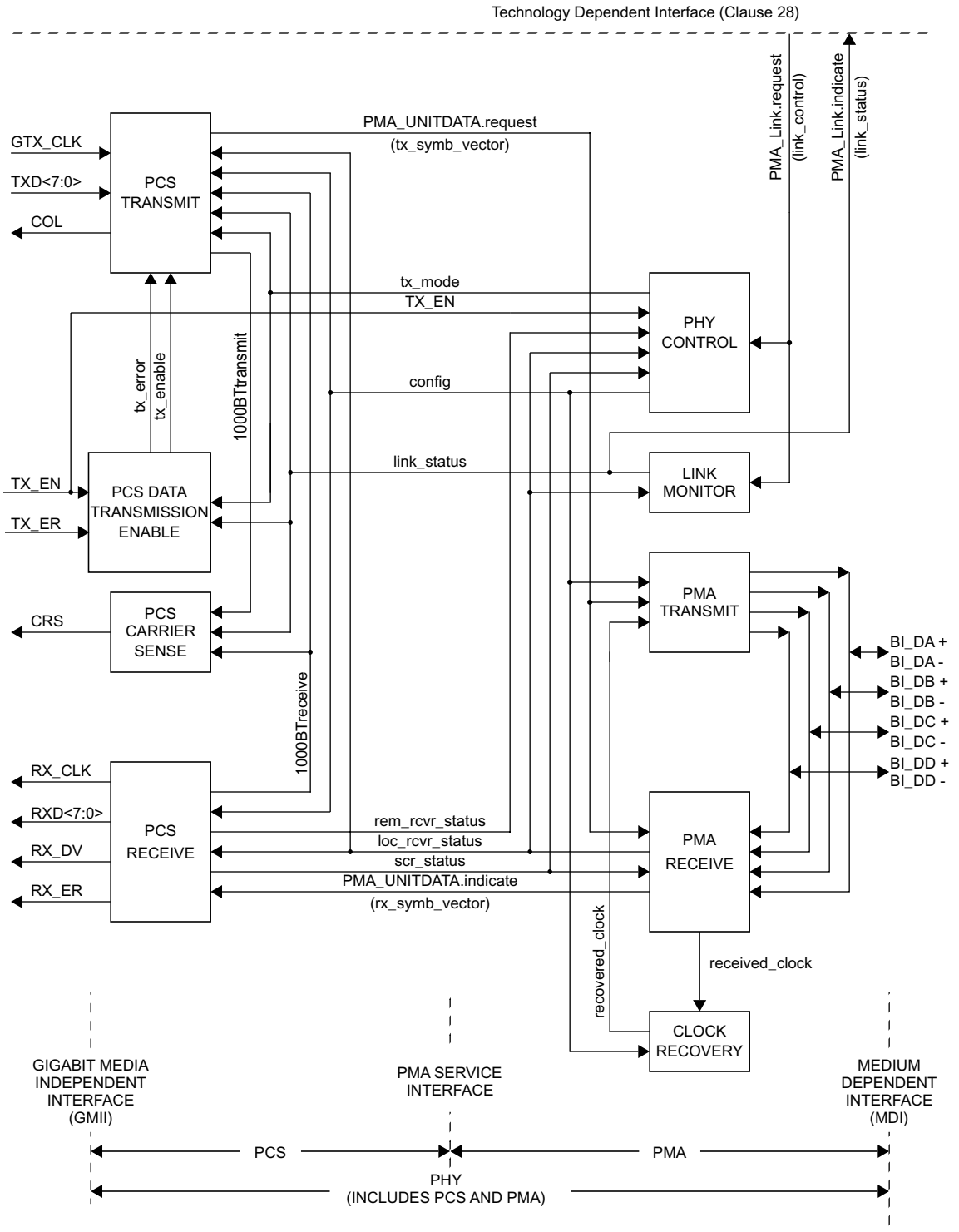


Figure 40–2—1000BASE-T topology



NOTE The recovered_clock arc is shown to indicate delivery of the received clock signal back to PMA TRANSMIT for loop timing.

Figure 40–3—Functional block diagram

40.1.3.1 Physical Coding Sublayer (PCS)

The 1000BASE-T PCS couples a Gigabit Media Independent Interface (GMII), as described in Clause 35, to a Physical Medium Attachment (PMA) sublayer.

The functions performed by the PCS comprise the generation of continuous code-groups to be transmitted over four channels and the processing of code-groups received from the remote PHY. The process of converting data bits to code-groups is called 4D-PAM5, which refers to the four-dimensional 5-level Pulse Amplitude Modulation coding technique used. Through this coding scheme, eight bits are converted to one transmission of four quinary symbols.

During the beginning of a frame's transmission, when TX_EN is asserted from the GMII, two code-groups representing the Start-of-Stream delimiter are transmitted followed by code-groups representing the octets coming from the GMII. Immediately following the data octets, the GMII sets TX_EN=FALSE, upon which the end of a frame is transmitted. The end of a frame consists of two convolutional state reset symbol periods and two End-of-Stream delimiter symbol periods. This is followed by an optional series of carrier extend symbol periods and, possibly, the start of a new frame during frame bursting. Otherwise, the end of a frame is followed by a series of symbols encoded in the idle mode. The nature of the encoding that follows the end of a frame is determined by the GMII signals TX_ER and TXD<7:0> as specified in Clause 35.

Between frames, a special subset of code-groups using only the symbols {2, 0, –2} is transmitted. This is called idle mode. Idle mode encoding takes into account the information of whether the local PHY is operating reliably or not (see 40.4.2.4) and allows this information to be conveyed to the remote station. During normal operation, idle mode is followed by a data mode that begins with a Start-of-Stream delimiter.

Further patterns are used for signaling a transmit error and other control functions during transmission of a data stream.

The PCS Receive processes code-groups provided by the PMA. The PCS Receive detects the beginning and the end of frames of data and, during the reception of data, descrambles and decodes the received code-groups into octets RXD<7:0> that are passed to the GMII. The conversion of code-groups to octets uses an 8B1Q4 data decoding technique. PCS Receive also detects errors in the received sequences and signals them to the GMII. Furthermore, the PCS contains a PCS Carrier Sense function, a PCS Collision Presence function, and a management interface.

The PCS functions and state diagrams are specified in 40.3. The signals provided by the PCS at the GMII conform to the interface requirements of Clause 35. The PCS Service Interfaces to the GMII and the PMA are abstract message-passing interfaces specified in 40.2.

40.1.3.2 Physical Medium Attachment (PMA) sublayer

The PMA couples messages from the PMA service interface onto the balanced cabling physical medium and provides the link management and PHY Control functions. The PMA provides full duplex communications at 125 MBaud over four pairs of balanced cabling up to 100 m in length.

The PMA Transmit function comprises four independent transmitters to generate five-level, pulse-amplitude modulated signals on each of the four pairs BI_DA, BI_DB, BI_DC, and BI_DD, as described in 40.4.3.1.

The PMA Receive function comprises four independent receivers for five-level pulse-amplitude modulated signals on each of the four pairs BI_DA, BI_DB, BI_DC, and BI_DD, as described in 40.4.3.2. This signal encoding technique is referred to as 4D-PAM5. The receivers are responsible for acquiring clock and providing code-groups to the PCS as defined by the PMA_UNITDATA.indicate message. The PMA also contains functions for Link Monitor.

The PMA PHY Control function generates signals that control the PCS and PMA sublayer operations. PHY Control begins following the completion of Auto-Negotiation and provides the start-up functions required for successful 1000BASE-T operation. It determines whether the PHY operates in a normal state, enabling data transmission over the link segment, or whether the PHY sends special code-groups that represent the idle mode. The latter occurs when either one or both of the PHYs that share a link segment are not operating reliably.

PMA functions and state diagrams are specified in 40.4. PMA electrical specifications are given in 40.6.

40.1.4 Signaling

1000BASE-T signaling is performed by the PCS generating continuous code-group sequences that the PMA transmits over each wire pair. The signaling scheme achieves a number of objectives including

a) Forward Error Correction (FEC) coded symbol mapping for data.
b) Algorithmic mapping and inverse mapping from octet data to a quartet of quinary symbols and back.
c) Uncorrelated symbols in the transmitted symbol stream.
d) No correlation between symbol streams traveling both directions on any pair combination.
e) No correlation between symbol streams on pairs BI_DA, BI_DB, BI_DC, and BI_DD.
f) Idle mode uses a subset of code-groups in that each symbol is restricted to the set {2, 0, –2}to ease synchronization, start-up, and retraining.
g) Ability to rapidly or immediately determine if a symbol stream represents data or idle or carrier extension.
h) Robust delimiters for Start-of-Stream delimiter (SSD), End-of-Stream delimiter (ESD), and other control signals.
i) Ability to signal the status of the local receiver to the remote PHY to indicate that the local receiver is not operating reliably and requires retraining.
j) Ability to automatically detect and correct for pair swapping and unexpected crossover connections.
k) Ability to automatically detect and correct for incorrect polarity in the connections.
l) Ability to automatically correct for differential delay variations across the wire-pairs.

The PHY operates in two basic modes, normal mode or training mode. In normal mode, PCS generates code-groups that represent data, control, or idles for transmission by the PMA. In training mode, the PCS is directed to generate only idle code-groups for transmission by the PMA, which enable the receiver at the other end to train until it is ready to operate in normal mode. (See the PCS reference diagram in 40.2.)

40.1.5 Inter-sublayer interfaces

All implementations of the balanced cabling link are compatible at the MDI. Designers are free to implement circuitry within the PCS and PMA in an application-dependent manner provided that the MDI and GMII (if the GMII is implemented) specifications are met. When the PHY is incorporated within the physical bounds of a single-port device or a multiport device, implementation of the GMII is optional. System operation from the perspective of signals at the MDI and management objects are identical whether the GMII is implemented or not.

40.1.6 Conventions in this clause

The body of this clause contains state diagrams, including definitions of variables, constants, and functions. Should there be a discrepancy between a state diagram and descriptive text, the state diagram prevails.

The notation used in the state diagrams follows the conventions of 21.5.

The values of all components in test circuits shall be accurate to within ±1% unless otherwise stated.

Default initializations, unless specifically specified, are left to the implementor.

40.2 1000BASE-T Service Primitives and Interfaces

1000BASE-T transfers data and control information across the following four service interfaces:

a) Gigabit Media Independent Interface (GMII)
b) PMA Service Interface
c) Medium Dependent Interface (MDI)
d) Technology-Dependent Interface

The GMII is specified in Clause 35; the Technology-Dependent Interface is specified in Clause 28. The PMA Service Interface is defined in 40.2.2 and the MDI is defined in 40.8.

40.2.1 Technology-Dependent Interface

1000BASE-T uses the following service primitives to exchange status indications and control signals across the Technology-Dependent Interface as specified in Clause 28:

PMA_LINK.request (link_control)

PMA_LINK.indicate (link_status)

40.2.1.1 PMA_LINK.request

This primitive allows the Auto-Negotiation algorithm to enable and disable operation of the PMA as specified in 28.2.6.2.

40.2.1.1.1 Semantics of the primitive

PMA_LINK.request (link_control)

The link_control parameter can take on one of three values: SCAN_FOR_CARRIER, DISABLE, or ENABLE.

SCAN_FOR_CARRIER	Used by the Auto-Negotiation algorithm prior to receiving any fast link pulses. During this mode the PMA reports link_status=FAIL.PHY processes are disabled.
DISABLE	Set by the Auto-Negotiation algorithm in the event fast link pulses are detected. PHY processes are disabled. This allows the Auto-Negotiation algorithm to determine how to configure the link.
ENABLE	Used by Auto-Negotiation to turn control over to the PHY for data processing functions.

40.2.1.1.2 When generated

Auto-Negotiation generates this primitive to indicate a change in link_control as described in Clause 28.

40.2.1.1.3 Effect of receipt

This primitive affects operation of the PMA Link Monitor function as defined in 40.4.2.5.

40.2.1.2 PMA_LINK.indicate

This primitive is generated by the PMA to indicate the status of the underlying medium as specified in 28.2.6.1. This primitive informs the PCS, PMA PHY Control function, and the Auto-Negotiation algorithm about the status of the underlying link.

40.2.1.2.1 Semantics of the primitive

PMA_LINK.indicate (link_status)

The link_status parameter can take on one of three values: FAIL, READY, or OK.

FAIL	No valid link established.
READY	The Link Monitor function indicates that a 1000BASE-T link is intact and ready to be established.
OK	The Link Monitor function indicates that a valid 1000BASE-T link is established. Reliable reception of signals transmitted from the remote PHY is possible.

40.2.1.2.2 When generated

The PMA generates this primitive continuously to indicate the value of link_status in compliance with the state diagram given in Figure 40–16.

40.2.1.2.3 Effect of receipt

The effect of receipt of this primitive is specified in 40.3.3.1.

40.2.2 PMA Service Interface

1000BASE-T uses the following service primitives to exchange symbol vectors, status indications, and control signals across the service interfaces:

PMA_TXMODE.indicate (tx_mode)

PMA_CONFIG.indicate (config)

PMA_UNITDATA.request (tx_symb_vector)

PMA_UNITDATA.indicate (rx_symb_vector)

PMA_SCRSTATUS.request (scr_status)

PMA_RXSTATUS.indicate (loc_rcvr_status)

PMA_REMRXSTATUS.request (rem_rcvr_status)

The use of these primitives is illustrated in Figure 40–4.

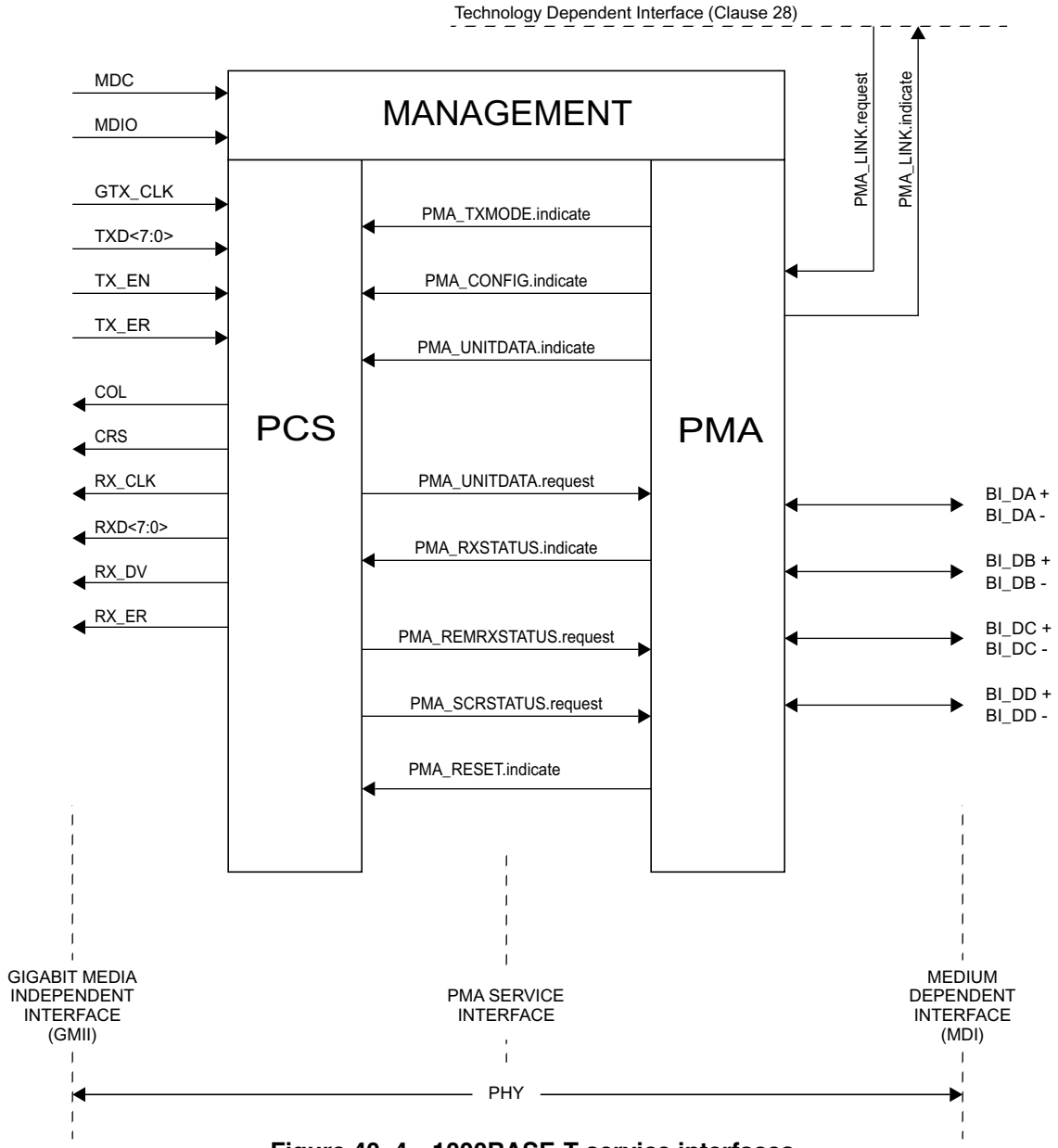


Figure 40–4—1000BASE-T service interfaces

40.2.3 PMA_TXMODE.indicate

The transmitter in a 1000BASE-T link normally sends over the four pairs, code-groups that can represent a GMII data stream, control information, or idles.

40.2.3.1 Semantics of the primitive

PMA_TXMODE.indicate (tx_mode)

PMA_TXMODE.indicate specifies to PCS Transmit via the parameter tx_mode what sequence of code-groups the PCS should be transmitting. The parameter tx_mode can take on one of the following three values of the form:

SEND_N	This value is continuously asserted when transmission of sequences of code-groups representing a GMII data stream (data mode), control mode or idle mode is to take place.
SEND_I	This value is continuously asserted in case transmission of sequences of code-groups representing the idle mode is to take place.
SEND_Z	This value is continuously asserted in case transmission of zeros is required.

40.2.3.2 When generated

The PMA PHY Control function generates PMA_TXMODE.indicate messages continuously.

40.2.3.3 Effect of receipt

Upon receipt of this primitive, the PCS performs its Transmit function as described in 40.3.1.3.

40.2.4 PMA_CONFIG.indicate

Each PHY in a 1000BASE-T link is capable of operating as a MASTER PHY and as a SLAVE PHY. MASTER-SLAVE configuration is determined during Auto-Negotiation (40.5). The result of this negotiation is provided to the PMA.

40.2.4.1 Semantics of the primitive

PMA_CONFIG.indicate (config)

PMA_CONFIG.indicate specifies to PCS and PMA Transmit via the parameter config whether the PHY must operate as a MASTER PHY or as a SLAVE PHY. The parameter config can take on one of the following two values of the form:

MASTER	This value is continuously asserted when the PHY must operate as a MASTER PHY.
SLAVE	This value is continuously asserted when the PHY must operate as a SLAVE PHY.

40.2.4.2 When generated

PMA generates PMA_CONFIG.indicate messages continuously.

40.2.4.3 Effect of receipt

PCS and PMA Clock Recovery perform their functions in MASTER or SLAVE configuration according to the value assumed by the parameter config.

40.2.5 PMA_UNITDATA.request

This primitive defines the transfer of code-groups in the form of the tx_symb_vector parameter from the PCS to the PMA. The code-groups are obtained in the PCS Transmit function using the encoding rules defined in 40.3.1.3 to represent GMII data streams, an idle mode, or other sequences.

40.2.5.1 Semantics of the primitive

PMA_UNITDATA.request (tx_symb_vector)

During transmission, the PMA_UNITDATA.request simultaneously conveys to the PMA via the parameter tx_symb_vector the value of the symbols to be sent over each of the four transmit pairs BI_DA, BI_DB, BI_DC, and BI_DD. The tx_symb_vector parameter takes on the form:

SYMB_4D	A vector of four quinary symbols, one for each of the four transmit pairs BI_DA, BI_DB, BI_DC, and BI_DD. Each quinary symbol may take on one of the values –2, –1, 0, +1, or +2.

The quinary symbols that are elements of tx_symb_vector are called, according to the pair on which each will be transmitted, tx_symb_vector[BI_DA], tx_symb_vector[BI_DB], tx_symb_vector[BI_DC], and tx_symb_vector[BI_DD].

40.2.5.2 When generated

The PCS generates PMA_UNITDATA.request (SYMB_4D) synchronously with every transmit clock cycle.

40.2.5.3 Effect of receipt

Upon receipt of this primitive the PMA transmits on the MDI the signals corresponding to the indicated quinary symbols. The parameter tx_symb_vector is also used by the PMA Receive function to process the signals received on pairs BI_DA, BI_DB, BI_DC, and BI_DD.

40.2.6 PMA_UNITDATA.indicate

This primitive defines the transfer of code-groups in the form of the rx_symb_vector parameter from the PMA to the PCS.

40.2.6.1 Semantics of the primitive

PMA_UNITDATA.indicate (rx_symb_vector)

During reception the PMA_UNITDATA.indicate simultaneously conveys to the PCS via the parameter rx_symb_vector the values of the symbols detected on each of the four receive pairs BI_DA, BI_DB, BI_DC, and BI_DD. The rx_symbol_vector parameter takes on the form:

SYMB_4D	A vector of four quinary symbols, one for each of the four receive pairs BI_DA, BI_DB, BI_DC, and BI_DD. Each quinary symbol may take on one of the values –2, –1, 0, +1, or +2.

The quinary symbols that are elements of rx_symb_vector are called, according to the pair upon which each symbol was received, rx_symbol_vector[BI_DA], rx_symbol_vector[BI_DB], rx_symbol_vector[BI_DC], and rx_symb_vector[BI_DD].

40.2.6.2 When generated

The PMA generates PMA_UNITDATA.indicate (SYMB_4D) messages synchronously with signals received at the MDI. The nominal rate of the PMA_UNITDATA.indicate primitive is 125 MHz, as governed by the recovered clock.

40.2.6.3 Effect of receipt

The effect of receipt of this primitive is unspecified.

40.2.7 PMA_SCRSTATUS.request

This primitive is generated by PCS Receive to communicate the status of the descrambler for the local PHY. The parameter scr_status conveys to the PMA Receive function the information that the descrambler has achieved synchronization.

40.2.7.1 Semantics of the primitive

PMA_SCRSTATUS.request (scr_status)

The scr_status parameter can take on one of two values of the form:

OK	The descrambler has achieved synchronization.
NOT_OK	The descrambler is not synchronized.

40.2.7.2 When generated

PCS Receive generates PMA_SCRSTATUS.request messages continuously.

40.2.7.3 Effect of receipt

The effect of receipt of this primitive is specified in 40.4.2.3, 40.4.2.4, and 40.4.6.1.

40.2.8 PMA_RXSTATUS.indicate

This primitive is generated by PMA Receive to indicate the status of the receive link at the local PHY. The parameter loc_rcvr_status conveys to the PCS Transmit, PCS Receive, PMA PHY Control function, and Link Monitor the information on whether the status of the overall receive link is satisfactory or not. Note that loc_rcvr_status is used by the PCS Receive decoding functions. The criterion for setting the parameter loc_rcvr_status is left to the implementor. It can be based, for example, on observing the mean-square error at the decision point of the receiver and detecting errors during reception of symbol streams that represent the idle mode.

40.2.8.1 Semantics of the primitive

PMA_RXSTATUS.indicate (loc_rcvr_status)

The loc_rcvr_status parameter can take on one of two values of the form:

OK	This value is asserted and remains true during reliable operation of the receive link for the local PHY.
NOT_OK	This value is asserted whenever operation of the link for the local PHY is unreliable.

40.2.8.2 When generated

PMA Receive generates PMA_RXSTATUS.indicate messages continuously on the basis of signals received at the MDI.

40.2.8.3 Effect of receipt

The effect of receipt of this primitive is specified in Figure 40–15 and in subclauses 40.2 and 40.4.6.2.

40.2.9 PMA_REMRXSTATUS.request

This primitive is generated by PCS Receive to indicate the status of the receive link at the remote PHY as communicated by the remote PHY via its encoding of its loc_rcvr_status parameter. The parameter rem_rcvr_status conveys to the PMA PHY Control function the information on whether reliable operation of the remote PHY is detected or not. The criterion for setting the parameter rem_rcvr_status is left to the implementor. It can be based, for example, on asserting rem_rcvr_status is NOT_OK until loc_rcvr_status is OK and then asserting the detected value of rem_rcvr_status after proper PCS receive decoding is achieved.

40.2.9.1 Semantics of the primitive

PMA_REMRXSTATUS.request (rem_rcvr_status)

The rem_rcvr_status parameter can take on one of two values of the form:

OK	The receive link for the remote PHY is operating reliably.
NOT_OK	Reliable operation of the receive link for the remote PHY is not detected.

40.2.9.2 When generated

The PCS generates PMA_REMRXSTATUS.request messages continuously on the basis on signals received at the MDI.

40.2.9.3 Effect of receipt

The effect of receipt of this primitive is specified in Figure 40–15.

40.2.10 PMA_RESET.indicate

This primitive is used to pass the PMA Reset function to the PCS (pcs_reset=ON) when reset is enabled.

The PMA_RESET.indicate primitive can take on one of two values:

TRUE	Reset is enabled.
FALSE	Reset is not enabled.

40.2.10.1 When generated

The PMA Reset function is executed as described in 40.4.2.1.

40.2.10.2 Effect of receipt

The effect of receipt of this primitive is specified in 40.4.2.1.

40.3 Physical Coding Sublayer (PCS)

The PCS comprises one PCS Reset function and four simultaneous and asynchronous operating functions. The PCS operating functions are: PCS Transmit Enable, PCS Transmit, PCS Receive, and PCS Carrier Sense. All operating functions start immediately after the successful completion of the PCS Reset function.

The PCS reference diagram, Figure 40–5, shows how the four operating functions relate to the messages of the PCS-PMA interface. Connections from the management interface (signals MDC and MDIO) to other

layers are pervasive, and are not shown in Figure 40–5. Management is specified in Clause 30. See also Figure 40–7, which defines the structure of frames passed from PCS to PMA.

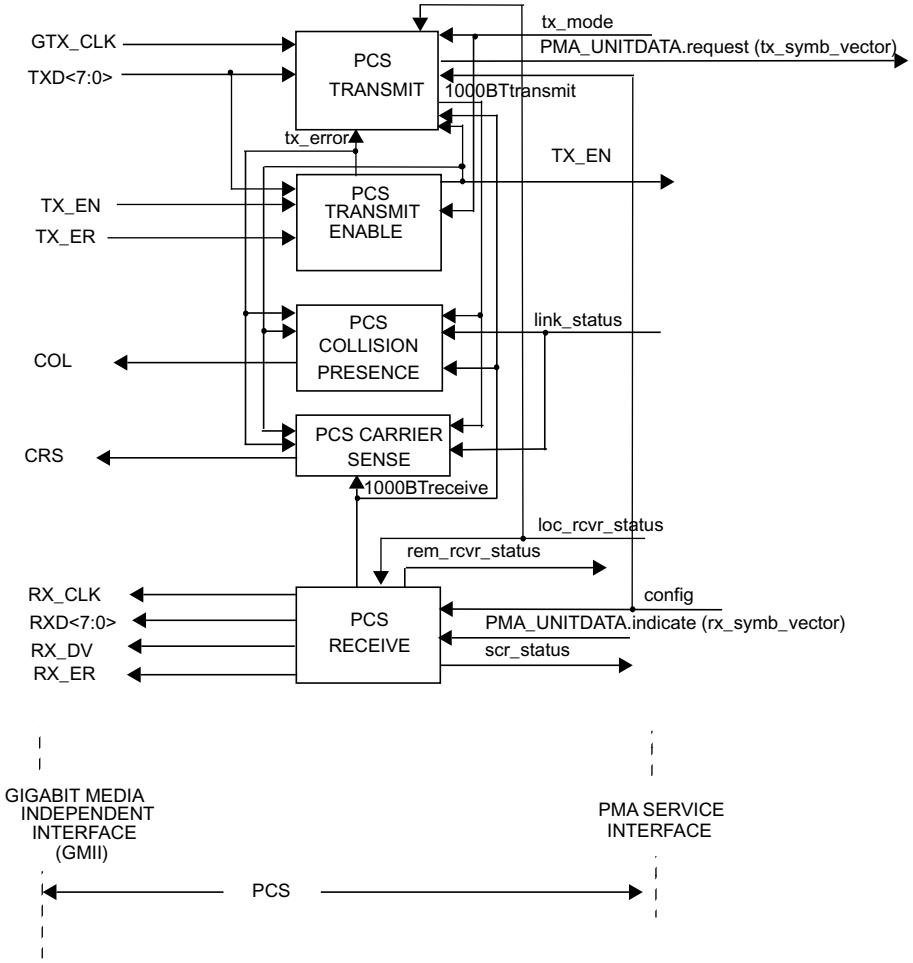


Figure 40–5—PCS reference diagram

40.3.1 PCS functions

40.3.1.1 PCS Reset function

PCS Reset initializes all PCS functions. The PCS Reset function shall be executed whenever one of the following conditions occur:

a) Power on (see 36.2.5.1.3).
b) The receipt of a request for reset from the management entity.

PCS Reset sets pcs_reset=ON while any of the above reset conditions hold true. All state diagrams take the open-ended pcs_reset branch upon execution of PCS Reset. The reference diagrams do not explicitly show the PCS Reset function.

40.3.1.2 PCS Data Transmission Enable

The PCS Data Transmission Enabling process generates the signals tx_enable and tx_error, which PCS Transmit uses for data and carrier extension encoding. The process uses logical operations on tx_mode, TX_ER, TX_EN, and TXD<7:0>. The PCS shall implement the Data Transmission Enabling process as depicted in Figure 40–8 including compliance with the associated state variables as specified in 40.3.3.

40.3.1.3 PCS Transmit function

The PCS Transmit function shall conform to the PCS Transmit state diagram in Figure 40–9.

The PCS Transmit function generates the GMII signal COL based on whether a reception is occurring simultaneously with transmission. The PCS Transmit function is not required to generate the GMII signal COL in a 1000BASE-T PHY that does not support half duplex operation.

In each symbol period, PCS Transmit generates a code-group (A_n, B_n, C_n, D_n) that is transferred to the PMA via the PMA_UNITDATA.request primitive. The PMA transmits symbols A_n, B_n, C_n, D_n over wire-pairs BI_DA, BI_DB, BI_DC, and BI_DD respectively. The integer, n, is a time index that is introduced to establish a temporal relationship between different symbol periods. A symbol period, T, is nominally equal to 8 ns. In normal mode of operation, between streams of data indicated by the parameter tx_enable, PCS Transmit generates sequences of vectors using the encoding rules defined for the idle mode. Upon assertion of tx_enable, PCS Transmit passes a SSD of two consecutive vectors of four quinary symbols to the PMA, replacing the first two preamble octets. Following the SSD, each TXD<7:0> octet is encoded using an 4D-PAM5 technique into a vector of four quinary symbols until tx_enable is de-asserted. If TX_ER is asserted while tx_enable is also asserted, then PCS Transmit passes to the PMA vectors indicating a transmit error. Note that if the signal TX_ER is asserted while SSD is being sent, the transmission of the error condition is delayed until transmission of SSD has been completed. Following the de-assertion of tx_enable, a Convolutional State Reset (CSReset) of two consecutive code-groups, followed by an ESD of two consecutive code-groups, is generated, after which the transmission of idle or control mode is resumed.

If a PMA_TXMODE.indicate message has the value SEND_Z, PCS Transmit passes a vector of zeros at each symbol period to the PMA via the PMA_UNITDATA.request primitive.

If a PMA_TXMODE.indicate message has the value SEND_I, PCS Transmit generates sequences of code-groups according to the encoding rule in training mode. Special code-groups that use only the values {+2, 0, –2} are transmitted in this case. Training mode encoding also takes into account the value of the parameter loc_rcvr_status. By this mechanism, a PHY indicates the status of its own receiver to the link partner during idle transmission.

In the normal mode of operation, the PMA_TXMODE.indicate message has the value SEND_N, and the PCS Transmit function uses an 8B1Q4 coding technique to generate at each symbol period code-groups that represent data, control or idle based on the code-groups defined in Table 40–1 and Table 40–2. During transmission of data, the TXD<7:0> bits are scrambled by the PCS using a side-stream scrambler, then encoded into a code-group of quinary symbols and transferred to the PMA. During data encoding, PCS Transmit utilizes a three-state convolutional encoder.

The transition from idle or carrier extension to data is signalled by inserting a SSD, and the end of transmission of data is signalled by an ESD. Further code-groups are reserved for signaling the assertion of TX_ER within a stream of data, carrier extension, CSReset, and other control functions. During idle and carrier extension encoding, special code-groups with symbol values restricted to the set {2, 0, –2} are used. These code-groups are also generated using the transmit side-stream scrambler. However, the encoding rules for the idle, SSD, and carrier extend code-groups are different from the encoding rules for data, CSReset, CSExtend, and ESD code-groups. During idle, SSD, and carrier extension, the PCS Transmit function reverses the sign of the transmitted symbols. This allows, at the receiver, sequences of code-groups that represent data,

CSReset, CSExtend, and ESD to be easily distinguished from sequences of code-groups that represent SSD, carrier extension, and idle.

PCS encoding involves the generation of the four-bit words $Sx_n[3:0]$, $Sy_n[3:0]$, and $Sg_n[3:0]$ from which the quinary symbols (A_n, B_n, C_n, D_n) are obtained. The four-bit words $Sx_n[3:0]$, $Sy_n[3:0]$, and $Sg_n[3:0]$ are determined (as explained in 40.3.1.3.2) from sequences of pseudorandom binary symbols derived from the transmit side-stream scrambler.

40.3.1.3.1 Side-stream scrambler polynomials

The PCS Transmit function employs side-stream scrambling. If the parameter config provided to the PCS by the PMA PHY Control function via the PMA_CONFIG.indicate message assumes the value MASTER, PCS Transmit shall employ

$$g_M(x) = 1 + x^{13} + x^{33}$$

as transmitter side-stream scrambler generator polynomial. If the PMA_CONFIG.indicate message assumes the value of SLAVE, PCS Transmit shall employ

$$g_S(x) = 1 + x^{20} + x^{33}$$

as transmitter side-stream scrambler generator polynomial. An implementation of master and slave PHY side-stream scramblers by linear-feedback shift registers is shown in Figure 40–6. The bits stored in the shift register delay line at time n are denoted by $Scr_n[32:0]$. At each symbol period, the shift register is advanced by one bit, and one new bit represented by $Scr_n[0]$ is generated. The transmitter side-stream scrambler is reset upon execution of the PCS Reset function. If PCS Reset is executed, all bits of the 33-bit vector representing the side-stream scrambler state are arbitrarily set. The initialization of the scrambler state is left to the implementor. In no case shall the scrambler state be initialized to all zeros.

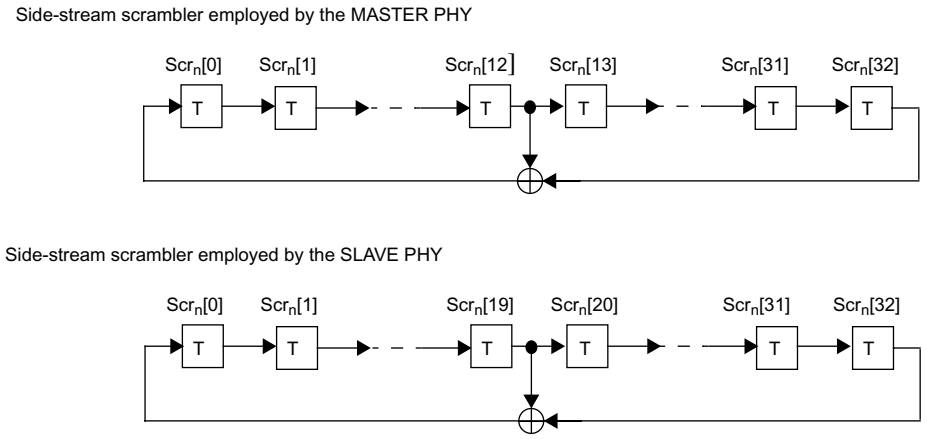


Figure 40–6—A realization of side-stream scramblers by linear feedback shift registers

40.3.1.3.2 Generation of bits $Sx_n[3:0]$, $Sy_n[3:0]$, and $Sg_n[3:0]$

PCS Transmit encoding rules are based on the generation, at time n, of the twelve bits $Sx_n[3:0]$, $Sy_n[3:0]$, and $Sg_n[3:0]$. The eight bits, $Sx_n[3:0]$ and $Sy_n[3:0]$, are used to generate the scrambler octet $Sc_n[7:0]$ for decorrelating the GMII data word TXD<7:0> during data transmission and for generating the idle and training symbols. The four bits, $Sg_n[3:0]$, are used to randomize the signs of the quinary symbols (A_n, B_n, C_n,

D_n) so that each symbol stream has no dc bias. These twelve bits are generated in a systematic fashion using three bits, X_n, Y_n, and $Scr_n[0]$, and an auxiliary generating polynomial, g(x). The two bits, X_n and Y_n, are mutually uncorrelated and also uncorrelated with the bit $Scr_n[0]$. For both master and slave PHYs, they are obtained by the same linear combinations of bits stored in the transmit scrambler shift register delay line. These two bits are derived from elements of the same maximum-length shift register sequence of length $2^{33} - 1$ as $Scr_n[0]$, but shifted in time. The associated delays are all large and different so that there is no short-term correlation among the bits $Scr_n[0]$, X_n, and Y_n. The bits X_n and Y_n are generated as follows:

$X_n = Scr_n[4] \wedge Scr_n[6]$

$Y_n = Scr_n[1] \wedge Scr_n[5]$

where ^ denotes the XOR logic operator. From the three bits X_n, Y_n, and $Scr_n[0]$, further mutually uncorrelated bit streams are obtained systematically using the generating polynomial

$g(x) = x^3 \wedge x^8$

The four bits $Sy_n[3:0]$ are generated using the bit $Scr_n[0]$ and g(x) as in the following equations:

$Sy_n[0] = Scr_n[0]$

$Sy_n[1] = g(Scr_n[0]) = Scr_n[3] \wedge Scr_n[8]$

$Sy_n[2] = g^2(Scr_n[0]) = Scr_n[6] \wedge Scr_n[16]$

$Sy_n[3] = g^3(Scr_n[0]) = Scr_n[9] \wedge Scr_n[14] \wedge Scr_n[19] \wedge Scr_n[24]$

The four bits $Sx_n[3:0]$ are generated using the bit X_n and g(x) as in the following equations:

$Sx_n[0] = X_n = Scr_n[4] \wedge Scr_n[6]$

$Sx_n[1] = g(X_n) = Scr_n[7] \wedge Scr_n[9] \wedge Scr_n[12] \wedge Scr_n[14]$

$Sx_n[2] = g^2(X_n) = Scr_n[10] \wedge Scr_n[12] \wedge Scr_n[20] \wedge Scr_n[22]$

$Sx_n[3] = g^3(X_n) = Scr_n[13] \wedge Scr_n[15] \wedge Scr_n[18] \wedge Scr_n[20] \wedge Scr_n[23] \wedge Scr_n[25] \wedge Scr_n[28] \wedge Scr_n[30]$

The four bits $Sg_n[3:0]$ are generated using the bit Y_n and g(x) as in the following equations:

$Sg_n[0] = Y_n = Scr_n[1] \wedge Scr_n[5]$

$Sg_n[1] = g(Y_n) = Scr_n[4] \wedge Scr_n[8] \wedge Scr_n[9] \wedge Scr_n[13]$

$Sg_n[2] = g^2(Y_n) = Scr_n[7] \wedge Scr_n[11] \wedge Scr_n[17] \wedge Scr_n[21]$

$Sg_n[3] = g^3(Y_n) = Scr_n[10] \wedge Scr_n[14] \wedge Scr_n[15] \wedge Scr_n[19] \wedge Scr_n[20] \wedge Scr_n[24] \wedge Scr_n[25] \wedge Scr_n[29]$

By construction, the twelve bits $Sx_n[3:0]$, $Sy_n[3:0]$, and $Sg_n[3:0]$ are derived from elements of the same maximum-length shift register sequence of length $2^{33}-1$ as $Scr_n[0]$, but shifted in time by varying delays. The associated delays are all large and different so that there is no apparent correlation among the bits.

40.3.1.3.3 Generation of bits $Sc_n[7:0]$

The bits $Sc_n[7:0]$ are used to scramble the GMII data octet TXD[7:0] and for control, idle, and training mode quartet generation. The definition of these bits is dependent upon the bits $Sx_n[3:0]$ and $Sy_n[3:0]$ that are specified in 40.3.1.3.2, the variable tx_mode that is obtained through the PMA Service Interface, the variable tx_enable_n that is defined in Figure 40–8, and the time index n.

The four bits $Sc_n[7:4]$ are defined as

$$Sc_n[7:4] = \begin{cases} Sx_n[3:0] & \text{if } (tx\_enable_{n-2} = 1) \\ [0\ 0\ 0\ 0] & \text{else} \end{cases}$$

The bits $Sc_n[3:1]$ are defined as

$$Sc_n[3:1] = \begin{cases} [0\ 0\ 0] & \text{if } (tx\_mode = SEND\_Z) \\ Sy_n[3:1] & \text{else if } (n\text{-}n_0) = 0 \pmod 2 \\ (Sy_{n-1}[3:1] \wedge [1\ 1\ 1]) & \text{else} \end{cases}$$

where n_0 denotes the time index of the last transmitter side-stream scrambler reset.

The bit $Sc_n[0]$ is defined as

$$Sc_n[0] = \begin{cases} 0 & \text{if } (tx\_mode = SEND\_Z) \\ Sy_n[0] & \text{else} \end{cases}$$

40.3.1.3.4 Generation of bits $Sd_n[8:0]$

The PCS Transmit function generates a nine-bit word $Sd_n[8:0]$ from Sc_n that represents either a convolutionally encoded stream of data, control, or idle mode code-groups. The convolutional encoder uses a three-bit word $cs_n[2:0]$, which is defined as

$$cs_n[1] = \begin{cases} Sd_n[6] \wedge cs_{n-1}[0] & \text{if } (tx\_enable_{n-2} = 1) \\ 0 & \text{else} \end{cases}$$

$$cs_n[2] = \begin{cases} Sd_n[7] \wedge cs_{n-1}[1] & \text{if } (tx\_enable_{n-2} = 1) \\ 0 & \text{else} \end{cases}$$

$$cs_n[0] = cs_{n-1}[2]$$

from which $Sd_n[8]$ is obtained as

$$Sd_n[8] = cs_n[0]$$

The convolutional encoder bits are non-zero only during the transmission of data. Upon the completion of a data frame, the convolutional encoder bits are reset using the bit $csreset_n$. The bit $csreset_n$ is defined as

$$csreset_n = (tx\_enable_{n-2}) \text{ and } (\text{not } tx\_enable_n)$$

The bits $Sd_n[7:6]$ are derived from the bits $Sc_n[7:6]$, the GMII data bits $TXD_n[7:6]$, and from the convolutional encoder bits as

$$Sd_n[7] = \begin{cases} Sc_n[7] \wedge TXD_n[7] & \text{if } (csreset_n = 0 \text{ and } tx\_enable_{n-2} = 1) \\ cs_{n-1}[1] & \text{else if } (csreset_n = 1) \\ Sc_n[7] & \text{else} \end{cases}$$

$$Sd_n[6] = \begin{cases} Sc_n[6] \wedge TXD_n[6] & \text{if } (csreset_n = 0 \text{ and } tx\_enable_{n-2} = 1) \\ cs_{n-1}[0] & \text{else if } (csreset_n = 1) \\ Sc_n[6] & \text{else} \end{cases}$$

The bits $Sd_n[5:3]$ are derived from the bits $Sc_n[5:3]$ and the GMII data bits $TXD_n[5:3]$ as

$$Sd_n[5:3] = \begin{cases} Sc_n[5:3] \wedge TXD_n[5:3] & \text{if } (tx\_enable_{n-2} = 1) \\ Sc_n[5:3] & \text{else} \end{cases}$$

The bit $Sd_n[2]$ is used to scramble the GMII data bit $TXD_n[2]$ during data mode and to encode loc_rcvr_status otherwise. It is defined as

$$Sd_n[2] = \begin{cases} Sc_n[2] \wedge TXD_n[2] & \text{if } (tx\_enable_{n-2} = 1) \\ Sc_n[2] \wedge 1 & \text{else if } (loc\_rcvr\_status = \text{OK}) \\ Sc_n[2] & \text{else} \end{cases}$$

The bits $Sd_n[1:0]$ are used to transmit carrier extension information during tx_mode=SEND_N and are thus dependent upon the bits $cext_n$ and $cext\_err_n$. These bits are dependent on the variable tx_error_n, which is defined in Figure 40–8. These bits are defined as

$$cext_n = \begin{cases} tx\_error_n & \text{if } ((tx\_enable_n = 0) \text{ and } (TXD_n[7:0] = 0\text{x}0\text{F})) \\ 0 & \text{else} \end{cases}$$

$$cext\_err_n = \begin{cases} tx\_error_n & \text{if } ((tx\_enable_n = 0) \text{ and } (TXD_n[7:0] \neq 0\text{x}0\text{F})) \\ 0 & \text{else} \end{cases}$$

$$Sd_n[1] = \begin{cases} Sc_n[1] \wedge TXD_n[1] & \text{if } (tx\_enable_{n-2} = 1) \\ Sc_n[1] \wedge cext\_err_n & \text{else} \end{cases}$$

$$Sd_n[0] = \begin{cases} Sc_n[0] \wedge TXD_n[0] \text{ if } (tx\_enable_{n-2} = 1) \\ Sc_n[0] \wedge cext_n \text{ else} \end{cases}$$

40.3.1.3.5 Generation of quinary symbols TA_n, TB_n, TC_n, TD_n

The nine-bit word $Sd_n[8:0]$ is mapped to a quartet of quinary symbols (TA_n, TB_n, TC_n, TD_n) according to Table 40–1 and Table 40–2 shown as $Sd_n[6:8] + Sd_n[5:0]$.

Encoding of error indication:

If tx_error_n=1 when the condition (tx_enable_n * tx_enable_{n-2}) = 1, error indication is signaled by means of symbol substitution. In this condition, the values of $Sd_n[5:0]$ are ignored during mapping and the symbols corresponding to the row denoted as "xmt_err" in Table 40–1 and Table 40–2 shall be used.

Encoding of Convolutional Encoder Reset:

If tx_error_n=0 when the variable $csreset_n$ = 1, the convolutional encoder reset condition is normal. This condition is indicated by means of symbol substitution, where the values of $Sd_n[5:0]$ are ignored during mapping and the symbols corresponding to the row denoted as "CSReset" in Table 40–1 and Table 40–2 shall be used.

Encoding of Carrier Extension during Convolutional Encoder Reset:

If tx_error_n=1 when the variable $csreset_n$ = 1, the convolutional encoder reset condition indicates carrier extension. In this condition, the values of $Sd_n[5:0]$ are ignored during mapping and the symbols corresponding to the row denoted as "CSExtend" in Table 40–1 and Table 40–2 shall be used when TXD_n = 0x'0F, and the row denoted as "CSExtend_Err" in Table 40–1 and Table 40–2 shall be used when $TXD_n \neq$ 0x'0F. The latter condition denotes carrier extension with error. In case carrier extension with error is indicated during the first octet of CSReset, the error condition shall be encoded during the second octet of CSReset, and during the subsequent two octets of the End-of-Stream delimiter as well. Thus, the error condition is assumed to persist during the symbol substitutions at the End-of-Stream.

Encoding of Start-of-Stream delimiter:

The Start-of-Stream delimiter (SSD) is related to the condition SSD_n, which is defined as (tx_enable_n) * (!tx_enable_{n-2}) = 1, where "*" and "!" denote the logic AND and NOT operators, respectively. For the generation of SSD, the first two octets of the preamble in a data stream are mapped to the symbols corresponding to the rows denoted as SSD1 and SSD2 respectively in Table 40–1. The symbols corresponding to the SSD1 row shall be used when the condition (tx_enable_n) * (!tx_enable_{n-1}) = 1. The symbols corresponding to the SSD2 row shall be used when the condition (tx_enable_{n-1}) * (!tx_enable_{n-2}) = 1.

Encoding of End-of-Stream delimiter:

The definition of an End-of-Stream delimiter (ESD) is related to the condition ESD_n, which is defined as (!tx_enable_{n-2}) * (tn_enable_{n-4}) = 1. This occurs during the third and fourth symbol periods after transmission of the last octet of a data stream.

If carrier extend error is indicated during ESD, the symbols corresponding to the ESD_Ext_Err row shall be used. The two conditions upon which this may occur are

(tx_error_n) * (tx_error_{n-1}) * (tx_error_{n-2}) * ($TXD_n \neq$ 0x0F) = 1, and

$(tx\_error_n) * (tx\_error_{n-1}) * (tx\_error_{n-2}) * (tx\_error_{n-3}) * (TXD_n \neq 0x0F) = 1$.

The symbols corresponding to the ESD1 row in Table 40–1 shall be used when the condition $(!tx\_enable_{n-2}) * (tx\_enable_{n-3}) = 1$, in the absence of carrier extend error indication at time n.

The symbols corresponding to the ESD2_Ext_0 row in Table 40–1 shall be used when the condition $(!tx\_enable_{n-3}) * (tx\_enable_{n-4}) * (!tx\_error_n) * (!tx\_error_{n-1}) = 1$.

The symbols corresponding to the ESD2_Ext_1 row in Table 40–1 shall be used when the condition $(!tx\_enable_{n-3}) * (tx\_enable_{n-4}) * (!tx\_error_n) * (tx\_error_{n-1}) * (tx\_error_{n-2}) * (tx\_error_{n-3}) = 1$.

The symbols corresponding to the ESD2_Ext_2 row in Table 40–1 shall be used when the condition $(!tx\_enable_{n-3}) * (tx\_enable_{n-4}) * (tx\_error_n) * (tx\_error_{n-1}) * (tx\_error_{n-2}) * (tx\_error_{n-3}) * (TXD_n = 0x0F) = 1$, in the absence of carrier extend error indication.

NOTE—The ASCII for Table 40–1 and Table 40–2 is available from http://www.ieee802.org/3/publication/index.html.[8]

Table 40–1—Bit-to-symbol mapping (even subsets)

		$Sd_n[6:8]$ = [000]	$Sd_n[6:8]$ = [010]	$Sd_n[6:8]$ = [100]	$Sd_n[6:8]$ = [110]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Normal	000000	0, 0, 0, 0	0, 0,+1,+1	0,+1,+1, 0	0,+1, 0,+1
Normal	000001	–2, 0, 0, 0	–2, 0,+1,+1	–2,+1,+1, 0	–2,+1, 0,+1
Normal	000010	0,–2, 0, 0	0,–2,+1,+1	0,–1,+1, 0	0,–1, 0,+1
Normal	000011	–2,–2, 0, 0	–2,–2,+1,+1	–2,–1,+1, 0	–2,–1, 0,+1
Normal	000100	0, 0,–2, 0	0, 0,–1,+1	0,+1,–1, 0	0,+1,–2,+1
Normal	000101	–2, 0,–2, 0	–2, 0,–1,+1	–2,+1,–1, 0	–2,+1,–2,+1
Normal	000110	0,–2,–2, 0	0,–2,–1,+1	0,–1,–1, 0	0,–1,–2,+1
Normal	000111	–2,–2,–2, 0	–2,–2,–1,+1	–2,–1,–1, 0	–2,–1,–2,+1
Normal	001000	0, 0, 0,–2	0, 0,+1,–1	0,+1,+1,–2	0,+1, 0,–1
Normal	001001	–2, 0, 0,–2	–2, 0,+1,–1	–2,+1,+1,–2	–2,+1, 0,–1
Normal	001010	0,–2, 0,–2	0,–2,+1,–1	0,–1,+1,–2	0,–1, 0,–1
Normal	001011	–2,–2, 0,–2	–2,–2,+1,–1	–2,–1,+1,–2	–2,–1, 0,–1
Normal	001100	0, 0,–2,–2	0, 0,–1,–1	0,+1,–1,–2	0,+1,–2,–1
Normal	001101	–2, 0,–2,–2	–2, 0,–1,–1	–2,+1,–1,–2	–2,+1,–2,–1
Normal	001110	0,–2,–2,–2	0,–2,–1,–1	0,–1,–1,–2	0,–1,–2,–1
Normal	001111	–2,–2,–2,–2	–2,–2,–1,–1	–2,–1,–1,–2	–2,–1,–2,–1
Normal	010000	+1,+1,+1,+1	+1,+1, 0, 0	+1, 0, 0,+1	+1, 0,+1, 0
Normal	010001	–1,+1,+1,+1	–1,+1, 0, 0	–1, 0, 0,+1	–1, 0,+1, 0
Normal	010010	+1,–1,+1,+1	+1,–1, 0, 0	+1,–2, 0,+1	+1,–2,+1, 0

[8]*Copyright release for symbol codes:* Users of this standard may freely reproduce the symbol codes in this subclause so it can be used for its intended purpose.

Table 40–1—Bit-to-symbol mapping (even subsets) *(continued)*

		$Sd_n[6:8]$ = [000]	$Sd_n[6:8]$ = [010]	$Sd_n[6:8]$ = [100]	$Sd_n[6:8]$ = [110]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Normal	010011	–1,–1,+1,+1	–1,–1, 0, 0	–1,–2, 0,+1	–1,–2,+1, 0
Normal	010100	+1,+1,–1,+1	+1,+1,–2, 0	+1, 0,–2,+1	+1, 0,–1, 0
Normal	010101	–1,+1,–1,+1	–1,+1,–2, 0	–1, 0,–2,+1	–1, 0,–1, 0
Normal	010110	+1,–1,–1,+1	+1,–1,–2, 0	+1,–2,–2,+1	+1,–2,–1, 0
Normal	010111	–1,–1,–1,+1	–1,–1,–2, 0	–1,–2,–2,+1	–1,–2,–1, 0
Normal	011000	+1,+1,+1,–1	+1,+1, 0,–2	+1, 0, 0,–1	+1, 0,+1,–2
Normal	011001	–1,+1,+1,–1	–1,+1, 0,–2	–1, 0, 0,–1	–1, 0,+1,–2
Normal	011010	+1,–1,+1,–1	+1,–1, 0,–2	+1,–2, 0,–1	+1,–2,+1,–2
Normal	011011	–1,–1,+1,–1	–1,–1, 0,–2	–1,–2, 0,–1	–1,–2,+1,–2
Normal	011100	+1,+1,–1,–1	+1,+1,–2,–2	+1, 0,–2,–1	+1, 0,–1,–2
Normal	011101	–1,+1,–1,–1	–1,+1,–2,–2	–1, 0,–2,–1	–1, 0,–1,–2
Normal	011110	+1,–1,–1,–1	+1,–1,–2,–2	+1,–2,–2,–1	+1,–2,–1,–2
Normal	011111	–1,–1,–1,–1	–1,–1,–2,–2	–1,–2,–2,–1	–1,–2,–1,–2
Normal	100000	+2, 0, 0, 0	+2, 0,+1,+1	+2,+1,+1, 0	+2,+1, 0,+1
Normal	100001	+2,–2, 0, 0	+2,–2,+1,+1	+2,–1,+1, 0	+2,–1, 0,+1
Normal	100010	+2, 0,–2, 0	+2, 0,–1,+1	+2,+1,–1, 0	+2,+1,–2,+1
Normal	100011	+2,–2,–2, 0	+2,–2,–1,+1	+2,–1,–1, 0	+2,–1,–2,+1
Normal	100100	+2, 0, 0,–2	+2, 0,+1,–1	+2,+1,+1,–2	+2,+1, 0,–1
Normal	100101	+2,–2, 0,–2	+2,–2,+1,–1	+2,–1,+1,–2	+2,–1, 0,–1
Normal	100110	+2, 0,–2,–2	+2, 0,–1,–1	+2,+1,–1,–2	+2,+1,–2,–1
Normal	100111	+2,–2,–2,–2	+2,–2,–1,–1	+2,–1,–1,–2	+2,–1,–2,–1
Normal	101000	0, 0,+2, 0	+1,+1,+2, 0	+1, 0,+2,+1	0,+1,+2,+1
Normal	101001	–2, 0,+2, 0	–1,+1,+2, 0	–1, 0,+2,+1	–2,+1,+2,+1
Normal	101010	0,–2,+2, 0	+1,–1,+2, 0	+1,–2,+2,+1	0,–1,+2,+1
Normal	101011	–2,–2,+2, 0	–1,–1,+2, 0	–1,–2,+2,+1	–2,–1,+2,+1
Normal	101100	0, 0,+2,–2	+1,+1,+2,–2	+1, 0,+2,–1	0,+1,+2,–1
Normal	101101	–2, 0,+2,–2	–1,+1,+2,–2	–1, 0,+2,–1	–2,+1,+2,–1
Normal	101110	0,–2,+2,–2	+1,–1,+2,–2	+1,–2,+2,–1	0,–1,+2,–1
Normal	101111	–2,–2,+2,–2	–1,–1,+2,–2	–1,–2,+2,–1	–2,–1,+2,–1
Normal	110000	0,+2, 0, 0	0,+2,+1,+1	+1,+2, 0,+1	+1,+2,+1, 0
Normal	110001	–2,+2, 0, 0	–2,+2,+1,+1	–1,+2, 0,+1	–1,+2,+1, 0

Table 40–1—Bit-to-symbol mapping (even subsets) *(continued)*

		$Sd_n[6:8]$ = [000]	$Sd_n[6:8]$ = [010]	$Sd_n[6:8]$ = [100]	$Sd_n[6:8]$ = [110]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Normal	110010	0,+2,–2, 0	0,+2,–1,+1	+1,+2,–2,+1	+1,+2,–1, 0
Normal	110011	–2,+2,–2, 0	–2,+2,–1,+1	–1,+2,–2,+1	–1,+2,–1, 0
Normal	110100	0,+2, 0,–2	0,+2,+1,–1	+1,+2, 0,–1	+1,+2,+1,–2
Normal	110101	–2,+2, 0,–2	–2,+2,+1,–1	–1,+2, 0,–1	–1,+2,+1,–2
Normal	110110	0,+2,–2,–2	0,+2,–1,–1	+1,+2,–2,–1	+1,+2,–1,–2
Normal	110111	–2,+2,–2,–2	–2,+2,–1,–1	–1,+2,–2,–1	–1,+2,–1,–2
Normal	111000	0, 0, 0,+2	+1,+1, 0,+2	0,+1,+1,+2	+1, 0,+1,+2
Normal	111001	–2, 0, 0,+2	–1,+1, 0,+2	–2,+1,+1,+2	–1, 0,+1,+2
Normal	111010	0,–2, 0,+2	+1,–1, 0,+2	0,–1,+1,+2	+1,–2,+1,+2
Normal	111011	–2,–2, 0,+2	–1,–1, 0,+2	–2,–1,+1,+2	–1,–2,+1,+2
Normal	111100	0, 0,–2,+2	+1,+1,–2,+2	0,+1,–1,+2	+1, 0,–1,+2
Normal	111101	–2, 0,–2,+2	–1,+1,–2,+2	–2,+1,–1,+2	–1, 0,–1,+2
Normal	111110	0,–2,–2,+2	+1,–1,–2,+2	0,–1,–1,+2	+1,–2,–1,+2
Normal	111111	–2,–2,–2,+2	–1,–1,–2,+2	–2,–1,–1,+2	–1,–2,–1,+2
xmt_err	XXXXXX	0,+2,+2,0	+1,+1,+2,+2	+2,+1,+1,+2	+2,+1,+2,+1
CSExtend_Err	XXXXXX	–2,+2,+2,–2	–1,–1,+2,+2	+2,–1,–1,+2	+2,–1,+2,–1
CSExtend	XXXXXX	+2, 0, 0,+2	+2,+2,+1,+1	+1,+2,+2,+1	+1,+2,+1,+2
CSReset	XXXXXX	+2,–2,–2,+2	+2,+2,–1,–1	–1,+2,+2,–1	–1,+2,–1,+2
SSD1	XXXXXX	+2,+2,+2,+2	—	—	—
SSD2	XXXXXX	+2,+2,+2,–2	—	—	—
ESD1	XXXXXX	+2,+2,+2,+2	—	—	—
ESD2_Ext_0	XXXXXX	+2,+2,+2,–2	—	—	—
ESD2_Ext_1	XXXXXX	+2,+2, –2,+2	—	—	—
ESD2_Ext_2	XXXXXX	+2,–2,+2,+2	—	—	—
ESD_Ext_Err	XXXXXX	–2,+2,+2,+2	—	—	—
Idle/Carrier Extension	000000	0, 0, 0, 0	—	—	—
Idle/Carrier Extension	000001	–2, 0, 0, 0	—	—	—
Idle/Carrier Extension	000010	0,–2, 0, 0	—	—	—

Table 40–1—Bit-to-symbol mapping (even subsets) *(continued)*

		$Sd_n[6:8]$ = [000]	$Sd_n[6:8]$ = [010]	$Sd_n[6:8]$ = [100]	$Sd_n[6:8]$ = [110]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Idle/Carrier Extension	000011	–2,–2, 0, 0	—	—	—
Idle/Carrier Extension	000100	0, 0,–2, 0	—	—	—
Idle/Carrier Extension	000101	–2, 0,–2, 0	—	—	—
Idle/Carrier Extension	000110	0,–2,–2, 0	—	—	—
Idle/Carrier Extension	000111	–2,–2,–2, 0	—	—	—
Idle/Carrier Extension	001000	0, 0, 0,–2	—	—	—
Idle/Carrier Extension	001001	–2, 0, 0,–2	—	—	—
Idle/Carrier Extension	001010	0,–2, 0,–2	—	—	—
Idle/Carrier Extension	001011	–2,–2, 0,–2	—	—	—
Idle/Carrier Extension	001100	0, 0,–2,–2	—	—	—
Idle/Carrier Extension	001101	–2, 0,–2,–2	—	—	—
Idle/Carrier Extension	001110	0,–2,–2,–2	—	—	—
Idle/Carrier Extension	001111	–2,–2,–2,–2	—	—	—

Table 40–2—Bit-to-symbol mapping (odd subsets)

		$Sd_n[6:8]$ = [001]	$Sd_n[6:8]$ = [011]	$Sd_n[6:8]$ = [101]	$Sd_n[6:8]$ = [111]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Normal	000000	0, 0, 0,+1	0, 0,+1, 0	0,+1,+1,+1	0,+1, 0, 0
Normal	000001	–2, 0, 0,+1	–2, 0,+1, 0	–2,+1,+1,+1	–2,+1, 0, 0
Normal	000010	0,–2, 0,+1	0,–2,+1, 0	0,–1,+1,+1	0,–1, 0, 0
Normal	000011	–2,–2, 0,+1	–2,–2,+1, 0	–2,–1,+1,+1	–2,–1, 0, 0
Normal	000100	0, 0,–2,+1	0, 0,–1, 0	0,+1,–1,+1	0,+1,–2, 0

Table 40–2—Bit-to-symbol mapping (odd subsets) *(continued)*

		$Sd_n[6:8]$ = [001]	$Sd_n[6:8]$ = [011]	$Sd_n[6:8]$ = [101]	$Sd_n[6:8]$ = [111]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Normal	000101	–2, 0,–2,+1	–2, 0,–1, 0	–2,+1,–1,+1	–2,+1,–2, 0
Normal	000110	0,–2,–2,+1	0,–2,–1, 0	0,–1,–1,+1	0,–1,–2, 0
Normal	000111	–2,–2,–2,+1	–2,–2,–1, 0	–2,–1,–1,+1	–2,–1,–2, 0
Normal	001000	0, 0, 0,–1	0, 0,+1,–2	0,+1,+1,–1	0,+1, 0,–2
Normal	001001	–2, 0, 0,–1	–2, 0,+1,–2	–2,+1,+1,–1	–2,+1, 0,–2
Normal	001010	0,–2, 0,–1	0,–2,+1,–2	0,–1,+1,–1	0,–1, 0,–2
Normal	001011	–2,–2, 0,–1	–2,–2,+1,–2	–2,–1,+1,–1	–2,–1, 0,–2
Normal	001100	0, 0,–2,–1	0, 0,–1,–2	0,+1,–1,–1	0,+1,–2,–2
Normal	001101	–2, 0,–2,–1	–2, 0,–1,–2	–2,+1,–1,–1	–2,+1,–2,–2
Normal	001110	0,–2,–2,–1	0,–2,–1,–2	0,–1,–1,–1	0,–1,–2,–2
Normal	001111	–2,–2,–2,–1	–2,–2,–1,–2	–2,–1,–1,–1	–2,–1,–2,–2
Normal	010000	+1,+1,+1, 0	+1,+1, 0,+1	+1, 0, 0, 0	+1, 0,+1,+1
Normal	010001	–1,+1,+1, 0	–1,+1, 0,+1	–1, 0, 0, 0	–1, 0,+1,+1
Normal	010010	+1,–1,+1, 0	+1,–1, 0,+1	+1,–2, 0, 0	+1,–2,+1,+1
Normal	010011	–1,–1,+1, 0	–1,–1, 0,+1	–1,–2, 0, 0	–1,–2,+1,+1
Normal	010100	+1,+1,–1, 0	+1,+1,–2,+1	+1, 0,–2, 0	+1, 0,–1,+1
Normal	010101	–1,+1,–1, 0	–1,+1,–2,+1	–1, 0,–2, 0	–1, 0,–1,+1
Normal	010110	+1,–1,–1, 0	+1,–1,–2,+1	+1,–2,–2, 0	+1,–2,–1,+1
Normal	010111	–1,–1,–1, 0	–1,–1,–2,+1	–1,–2,–2, 0	–1,–2,–1,+1
Normal	011000	+1,+1,+1,–2	+1,+1, 0,–1	+1, 0, 0,–2	+1, 0,+1,–1
Normal	011001	–1,+1,+1,–2	–1,+1, 0,–1	–1, 0, 0,–2	–1, 0,+1,–1
Normal	011010	+1,–1,+1,–2	+1,–1, 0,–1	+1,–2, 0,–2	+1,–2,+1,–1
Normal	011011	–1,–1,+1,–2	–1,–1, 0,–1	–1,–2, 0,–2	–1,–2,+1,–1
Normal	011100	+1,+1,–1,–2	+1,+1,–2,–1	+1, 0,–2,–2	+1, 0,–1,–1
Normal	011101	–1,+1,–1,–2	–1,+1,–2,–1	–1, 0,–2,–2	–1, 0,–1,–1
Normal	011110	+1,–1,–1,–2	+1,–1,–2,–1	+1,–2,–2,–2	+1,–2,–1,–1
Normal	011111	–1,–1,–1,–2	–1,–1,–2,–1	–1,–2,–2,–2	–1,–2,–1,–1
Normal	100000	+2, 0, 0,+1	+2, 0,+1, 0	+2,+1,+1,+1	+2,+1, 0, 0
Normal	100001	+2,–2, 0,+1	+2,–2,+1, 0	+2,–1,+1,+1	+2,–1, 0, 0
Normal	100010	+2, 0,–2,+1	+2, 0,–1, 0	+2,+1,–1,+1	+2,+1,–2, 0

Table 40–2—Bit-to-symbol mapping (odd subsets) *(continued)*

		$Sd_n[6:8]$ = [001]	$Sd_n[6:8]$ = [011]	$Sd_n[6:8]$ = [101]	$Sd_n[6:8]$ = [111]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
Normal	100011	+2,–2,–2,+1	+2,–2,–1, 0	+2,–1,–1,+1	+2,–1,–2, 0
Normal	100100	+2, 0, 0,–1	+2, 0,+1,–2	+2,+1,+1,–1	+2,+1, 0,–2
Normal	100101	+2,–2, 0,–1	+2,–2,+1,–2	+2,–1,+1,–1	+2,–1, 0,–2
Normal	100110	+2, 0,–2,–1	+2, 0,–1,–2	+2,+1,–1,–1	+2,+1,–2,–2
Normal	100111	+2,–2,–2,–1	+2,–2,–1,–2	+2,–1,–1,–1	+2,–1,–2,–2
Normal	101000	0, 0,+2,+1	+1,+1,+2,+1	+1, 0,+2, 0	0,+1,+2, 0
Normal	101001	–2, 0,+2,+1	–1,+1,+2,+1	–1, 0,+2, 0	–2,+1,+2, 0
Normal	101010	0,–2,+2,+1	+1,–1,+2,+1	+1,–2,+2, 0	0,–1,+2, 0
Normal	101011	–2,–2,+2,+1	–1,–1,+2,+1	–1,–2,+2, 0	–2,–1,+2, 0
Normal	101100	0, 0,+2,–1	+1,+1,+2,–1	+1, 0,+2,–2	0,+1,+2,–2
Normal	101101	–2, 0,+2,–1	–1,+1,+2,–1	–1, 0,+2,–2	–2,+1,+2,–2
Normal	101110	0,–2,+2,–1	+1,–1,+2,–1	+1,–2,+2,–2	0,–1,+2,–2
Normal	101111	–2,–2,+2,–1	–1,–1,+2,–1	–1,–2,+2,–2	–2,–1,+2,–2
Normal	110000	0,+2, 0,+1	0,+2,+1, 0	+1,+2, 0, 0	+1,+2,+1,+1
Normal	110001	–2,+2, 0,+1	–2,+2,+1, 0	–1,+2, 0, 0	–1,+2,+1,+1
Normal	110010	0,+2,–2,+1	0,+2,–1, 0	+1,+2,–2, 0	+1,+2,–1,+1
Normal	110011	–2,+2,–2,+1	–2,+2,–1, 0	–1,+2,–2, 0	–1,+2,–1,+1
Normal	110100	0,+2, 0,–1	0,+2,+1,–2	+1,+2, 0,–2	+1,+2,+1,–1
Normal	110101	–2,+2, 0,–1	–2,+2,+1,–2	–1,+2, 0,–2	–1,+2,+1,–1
Normal	110110	0,+2,–2,–1	0,+2,–1,–2	+1,+2,–2,–2	+1,+2,–1,–1
Normal	110111	–2,+2,–2,–1	–2,+2,–1,–2	–1,+2,–2,–2	–1,+2,–1,–1
Normal	111000	+1,+1,+1,+2	0, 0,+1,+2	+1, 0, 0,+2	0,+1, 0,+2
Normal	111001	–1,+1,+1,+2	–2, 0,+1,+2	–1, 0, 0,+2	–2,+1, 0,+2
Normal	111010	+1,–1,+1,+2	0,–2,+1,+2	+1,–2, 0,+2	0,–1, 0,+2
Normal	111011	–1,–1,+1,+2	–2,–2,+1,+2	–1,–2, 0,+2	–2,–1, 0,+2
Normal	111100	+1,+1,–1,+2	0, 0,–1,+2	+1, 0,–2,+2	0,+1,–2,+2
Normal	111101	–1,+1,–1,+2	–2, 0,–1,+2	–1, 0,–2,+2	–2,+1,–2,+2
Normal	111110	+1,–1,–1,+2	0,–2,–1,+2	+1,–2,–2,+2	0,–1,–2,+2
Normal	111111	–1,–1,–1,+2	–2,–2,–1,+2	–1,–2,–2,+2	–2,–1,–2,+2
xmt_err	XXXXXX	+2,+2, 0,+1	0,+2,+1,+2	+1,+2,+2, 0	+2,+1,+2, 0

Table 40–2—Bit-to-symbol mapping (odd subsets) *(continued)*

		$Sd_n[6:8]$ = [001]	$Sd_n[6:8]$ = [011]	$Sd_n[6:8]$ = [101]	$Sd_n[6:8]$ = [111]
Condition	**$Sd_n[5:0]$**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**	**TA_n,TB_n,TC_n, TD_n**
CSExtend_Err	XXXXXX	+2,+2, –2,–1	–2,+2,–1,+2	–1,+2,+2,–2	+2,–1,+2,–2
CSExtend	XXXXXX	+2, 0,+2,+1	+2, 0,+1,+2	+1, 0,+2,+2	+2,+1, 0,+2
CSReset	XXXXXX	+2,–2,+2,–1	+2,–2,–1,+2	–1,–2,+2,+2	+2,–1,–2,+2

40.3.1.3.6 Generation of A_n, B_n, C_n, D_n

The four bits $Sg_n[3:0]$ are used to randomize the signs of the quinary symbols (A_n, B_n, C_n, D_n) so that each symbol stream has no dc bias. The bits are used to generate binary symbols (SnA_n, SnB_n, SnC_n, SnD_n) that, when multiplied by the quinary symbols (TA_n, TB_n, TC_n, TD_n), result in (A_n, B_n, C_n, D_n).

PCS Transmit ensures a distinction between code-groups transmitted during idle mode plus SSD and those transmitted during other symbol periods. This distinction is accomplished by reversing the mapping of the sign bits when the condition (tx_enable_{n-2} + tx_enable_{n-4}) = 1. This sign reversal is controlled by the variable $Srev_n$ defined as

$$Srev_n = tx\_enable_{n-2} + tx\_enable_{n-4}$$

The binary symbols SnA_n, SnB_n, SnC_n, and SnD_n are defined using $Sg_n[3:0]$ as

$$SnA_n = \begin{cases} +1 \text{ if } [(Sg_n[0] \wedge Srev_n) = 0] \\ -1 \text{ else} \end{cases}$$

$$SnB_n = \begin{cases} +1 \text{ if } [(Sg_n[1] \wedge Srev_n) = 0] \\ -1 \text{ else} \end{cases}$$

$$SnC_n = \begin{cases} +1 \text{ if } [(Sg_n[2] \wedge Srev_n) = 0] \\ -1 \text{ else} \end{cases}$$

$$SnD_n = \begin{cases} +1 \text{ if } [(Sg_n[3] \wedge Srev_n) = 0] \\ -1 \text{ else} \end{cases}$$

The quinary symbols (A_n, B_n, C_n, D_n) are generated as the product of (SnA_n, SnB_n, SnC_n, SnD_n) and (TA_n, TB_n, TC_n, TD_n) respectively.

$$A_n = TA_n \times SnA_n$$

$$B_n = TB_n \times SnB_n$$

$$C_n = TC_n \times SnC_n$$

$$D_n = TD_n \times SnD_n$$

40.3.1.4 PCS Receive function

The PCS Receive function shall conform to the PCS Receive state diagram in Figure 40–10a including compliance with the associated state variables as specified in 40.3.3.

The PCS Receive function accepts received code-groups provided by the PMA Receive function via the parameter rx_symb_vector. To achieve correct operation, PCS Receive uses the knowledge of the encoding rules that are employed in the idle mode. PCS Receive generates the sequence of vectors of four quinary symbols (RA_n, RB_n, RC_n, RD_n) and indicates the reliable acquisition of the descrambler state by setting the parameter scr_status to OK. The sequence (RA_n, RB_n, RC_n, RD_n) is processed to generate the signals RXD<7:0>, RX_DV, and RX_ER, which are presented to the GMII. PCS Receive detects the transmission of a stream of data from the remote station and conveys this information to the PCS Carrier Sense and PCS Transmit functions via the parameter 1000BTreceive.

40.3.1.4.1 Decoding of code-groups

When the PMA indicates that correct receiver operation has been achieved by setting the loc_rcvr_status parameter to the value OK, the PCS Receive continuously checks that the received sequence satisfies the encoding rule used in idle mode. When a violation is detected, PCS Receive assigns the value TRUE to the parameter 1000BTreceive and, by examining the last two received vectors (RA_{n-1}, RB_{n-1}, RC_{n-1}, RD_{n-1}) and (RA_n, RB_n, RC_n, RD_n), determines whether the violation is due to reception of SSD or to a receiver error.

Upon detection of SSD, PCS Receive also assigns the value TRUE to the parameter 1000BTreceive that is provided to the PCS Carrier Sense and Collision Presence functions. During the two symbol periods corresponding to SSD, PCS Receive replaces SSD by preamble bits. Upon the detection of SSD, the signal RX_DV is asserted and each received vector is decoded into a data octet RXD<7:0> until ESD is detected.

Upon detection of a receiver error, the signal RX_ER is asserted and the parameter rxerror_status assumes the value ERROR. De-assertion of RX_ER and transition to the IDLE state (rxerror_status=NO_ERROR) takes place upon detection of four consecutive vectors satisfying the encoding rule used in idle mode.

During reception of a stream of data, PCS Receive checks that the symbols RA_n, RB_n, RC_n, RD_n follow the encoding rule defined in 40.3.1.3.5 for ESD whenever they assume values ± 2. PCS Receive processes two consecutive vectors at each time n to detect ESD. Upon detection of ESD, PCS Receive de-asserts the signal RX_DV on the GMII. If the last symbol period of ESD indicates that a carrier extension is present, PCS Receive will assert the RX_ER signal on the GMII. If no extension is indicated in the ESD2 quartet, PCS Receive assigns the value FALSE to the parameter receiving. If an extension is present, the transition to the IDLE state occurs after detection of a valid idle symbol period and the parameter receiving remains TRUE until check_idle is TRUE. If a violation of the encoding rules is detected, PCS Receive asserts the signal RX_ER for at least one symbol period.

A premature stream termination is caused by the detection of invalid symbols during the reception of a data stream. Then, PCS Receive waits for the reception of four consecutive vectors satisfying the encoding rule used in idle mode prior to de-asserting the error indication. Note that RX_DV remains asserted during the symbol periods corresponding to the first three idle vectors, while RX_ER=TRUE is signaled on the GMII. The signal RX_ER is also asserted in the LINK FAILED state, which ensures that RX_ER remains asserted for at least one symbol period.

40.3.1.4.2 Receiver descrambler polynomials

The PHY shall descramble the data stream and return the proper sequence of code-groups to the decoding process for generation of RXD<7:0> to the GMII. For side-stream descrambling, the MASTER PHY shall employ the receiver descrambler generator polynomial $g'_M(x) = 1 + x^{20} + x^{33}$ and the SLAVE PHY shall employ the receiver descrambler generator polynomial $g'_S(x) = 1 + x^{13} + x^{33}$.

40.3.1.5 PCS Carrier Sense function

The PCS Carrier Sense function generates the GMII signal CRS, which the MAC uses for deferral in half duplex mode. The PCS shall conform to the Carrier Sense state diagram as depicted in Figure 40–11 including compliance with the associated state variables as specified in 40.3.3. The PCS Carrier Sense function is not required in a 1000BASE-T PHY that does not support half duplex operation.

40.3.2 Stream structure

The tx_symb_vector and rx-symb_vector structure is shown in Figure 40–7.

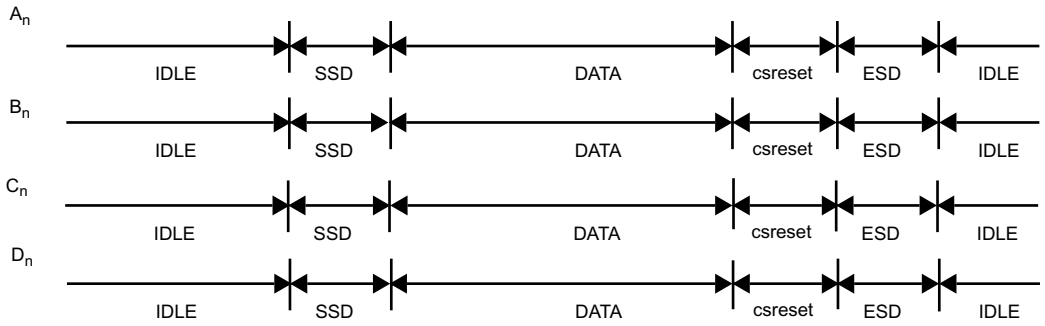


Figure 40–7—The tx_symb_vector and rx-symb_vector structure

40.3.3 State variables

40.3.3.1 Variables

CEXT

A vector of four quinary symbols corresponding to the code-group generated in idle mode to denote carrier extension, as specified in 40.3.1.3.

CEXT_Err

A vector of four quinary symbols corresponding to the code-group generated in idle mode to denote carrier extension with error indication, as specified in 40.3.1.3.

COL

The COL signal of the GMII as specified in 35.2.2.10.

config

The config parameter set by PMA and passed to the PCS via the PMA_CONFIG.indicate primitive.
Values: MASTER, SLAVE.

CRS

The CRS signal of the GMII as specified in 35.2.2.9.

CSExtend

A vector of four quinary symbols corresponding to the code-group indicating convolutional encoder reset condition during carrier extension, as specified in 40.3.1.3.

CSExtend_Err

A vector of four quinary symbols corresponding to the code-group indicating convolutional encoder reset condition during carrier extension with error indication, as specified in 40.3.1.3.

CSReset

A vector of four quinary symbols corresponding to the code-group indicating convolutional encoder reset condition in the absence of carrier extension, as specified in 40.3.1.3.

DATA

A vector of four quinary symbols corresponding to the code-group indicating valid data, as specified in 40.3.1.3.

ESD1

A vector of four quinary symbols corresponding to the first code-group of End-of-Stream delimiter, as specified in 40.3.1.3.

ESD2_Ext_0

A vector of four quinary symbols corresponding to the second code-group of End-of-Stream delimiter in the absence of carrier extension over the two ESD symbol periods, as specified in 40.3.1.3.

ESD2_Ext_1

A vector of four quinary symbols corresponding to the second code-group of End-of-Stream delimiter when carrier extension is indicated during the first symbol period of the End-of-Stream delimiter, but not during the second symbol period, as specified in 40.3.1.3.

ESD2_Ext_2

A vector of four quinary symbols corresponding to the second code-group of End-of-Stream delimiter when carrier extension is indicated during the two symbol periods of the End-of-Stream delimiter, as specified in 40.3.1.3.

ESD_Ext_Err

A vector of four quinary symbols corresponding to either the first or second code-group of End-of-Stream delimiter when carrier extension with error is indicated during the End-of-Stream delimiter, as specified in 40.3.1.3.

IDLE

A sequence of vectors of four quinary symbols representing the special code-group generated in idle mode in the absence of carrier extension or carrier extension with error indication, as specified in 40.3.1.3.

link_status

The link_status parameter set by PMA Link Monitor and passed to the PCS via the PMA_LINK.indicate primitive.
Values: OK or FAIL

loc_rcvr_status
: The loc_rcvr_status parameter set by the PMA Receive function and passed to the PCS via the PMA_RXSTATUS.indicate primitive.
Values: OK or NOT_OK

pcs_reset
: The pcs_reset parameter set by the PCS Reset function.
Values: ON or OFF

(RA_n, RB_n, RC_n, RD_n)
: The vector of the four correctly aligned most recently received quinary symbols generated by PCS Receive at time n.

1000BTreceive
: The receiving parameter generated by the PCS Receive function.
Values: TRUE or FALSE

rem_rcvr_status
: The rem_rcvr_status parameter generated by PCS Receive.
Values: OK or NOT_OK

repeater_mode
: See 36.2.5.1.3

Rx_n
: Alias for rx_symb_vector (a vector RA_n, RB_n, RC_n, RD_n) at time n.

rxerror_status
: The rxerror_status parameter set by the PCS Receive function.
Values: ERROR or NO_ERROR

RX_DV
: The RX_DV signal of the GMII as specified in 35.2.2.6.

RX_ER
: The RX_ER signal of the GMII as specified in 35.2.2.8.

rx_symb_vector
: A vector of four quinary symbols received by the PMA and passed to the PCS via the PMA_UNITDATA.indicate primitive.
Value: SYMB_4D

RXD[7:0]
: The RXD<7:0> signal of the GMII as specified in 35.2.2.7.

SSD1
: A vector of four quinary symbols corresponding to the first code-group of the Start-of-Stream delimiter, as specified in 40.3.1.3.5.

SSD2
: A vector of four quinary symbols corresponding to the second code-group of the Start-of-Stream delimiter, as specified in 40.3.1.3.5.

1000BTtransmit
: A boolean used by the PCS Transmit Process to indicate whether a frame transmission is in progress. Used by Carrier Sense process.
Values: TRUE: The PCS is transmitting a stream
FALSE: The PCS is not transmitting a stream

TXD[7:0]
: The TXD<7:0> signal of the GMII as specified in 35.2.2.4.

tx_enable
: The tx_enable parameter generated by PCS Transmit as specified in Figure 40–8.
Values: TRUE or FALSE

tx_error
: The tx_error parameter generated by PCS Transmit as specified in Figure 40–8.
Values: TRUE or FALSE

TX_EN
: The TX_EN signal of the GMII as specified in 35.2.2.3.

TX_ER
: The TX_ER signal of the GMII as specified in 35.2.2.5.

tx_mode
: The tx_mode parameter set by the PMA PHY Control function and passed to the PCS via the PMA_TXMODE.indicate primitive.
Values: SEND_Z, SEND_N, or SEND_I

Tx_n
: Alias for tx_symb_vector at time n.

tx_symb_vector
: A vector of four quinary symbols generated by the PCS Transmit function and passed to the PMA via the PMA_UNITDATA.request primitive.
Value: SYMB_4D

xmt_err
: A vector of four quinary symbols corresponding to a transmit error indication during normal data transmission or reception, as specified in 40.3.1.3.

40.3.3.2 Functions

check_end
: A function used by the PCS Receive process to detect the reception of valid ESD symbols. The check_end function operates on the next two rx_symb_vectors, (Rx_{n+1}) and (Rx_{n+2}), available via PMA_UNITDATA.indicate, and returns a boolean value indicating whether these two consecutive vectors contain symbols corresponding to a legal ESD encoding or not, as specified in 40.3.1.3.

check_idle
: A function used by the PCS Receive process to detect the reception of valid idle code-groups after an error condition during the process. The check_idle function operates on the current rx_symb_vector and the next three rx_symb_vectors available via PMA_UNITDATA.indicate and returns a boolean value indicating whether the four consecutive vectors contain symbols corresponding to the idle mode encoding or not, as specified in 40.3.1.3.

DECODE
: In the PCS Receive process, this function takes as its argument the value of rx_symb_vector and returns the corresponding GMII RXD<7:0> octet. DECODE follows the rules outlined in 40.2.6.1.

ENCODE
: In the PCS Transmit process, this function takes as its argument GMII TXD <7:0> and returns the corresponding tx_symb_vector. ENCODE follows the rules outlined in 40.2.5.1.

40.3.3.3 Timer

symb_timer
: Continuous timer: The condition symb_timer_done becomes true upon timer expiration.

  Restart time: Immediately after expiration; timer restart resets the condition symb_timer_done.

  Duration: 8 ns nominal. (See clock tolerance in 40.6.1.2.6.)

  Symb-timer shall be generated synchronously with TX_TCLK. In the PCS Transmit state diagram, the message PMA_UNITDATA.request is issued concurrently with symb_timer_done.

40.3.3.4 Messages

PMA_UNITDATA.indicate (rx_symb_vector)
: A signal sent by PMA Receive indicating that a vector of four quinary symbols is available in rx_symb_vector. (See 40.2.6.)

PMA_UNITDATA.request (tx_symb_vector)
: A signal sent to PMA Transmit indicating that a vector of four quinary symbols is available in tx_symb_vector. (See 40.2.5.)

PUDI
: Alias for PMA_UNITDATA.indicate (rx_symb_vector).

PUDR
: Alias for PMA_UNITDATA.request (tx_symb_vector).

STD
: Alias for symb_timer_done.

40.3.4 State diagrams

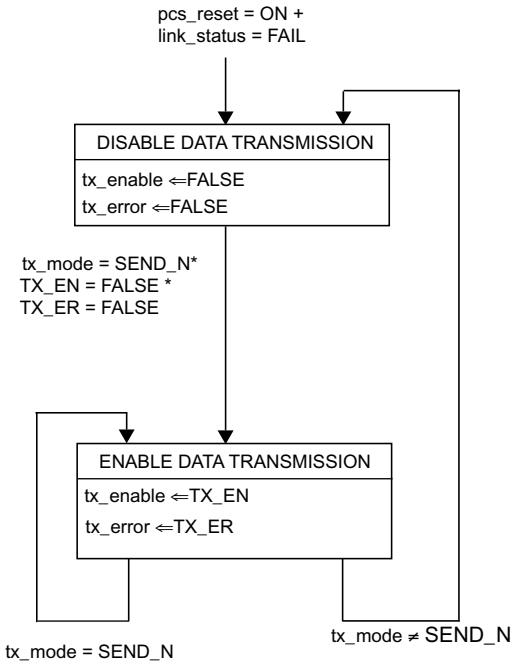


Figure 40–8—PCS Data Transmission Enabling state diagram

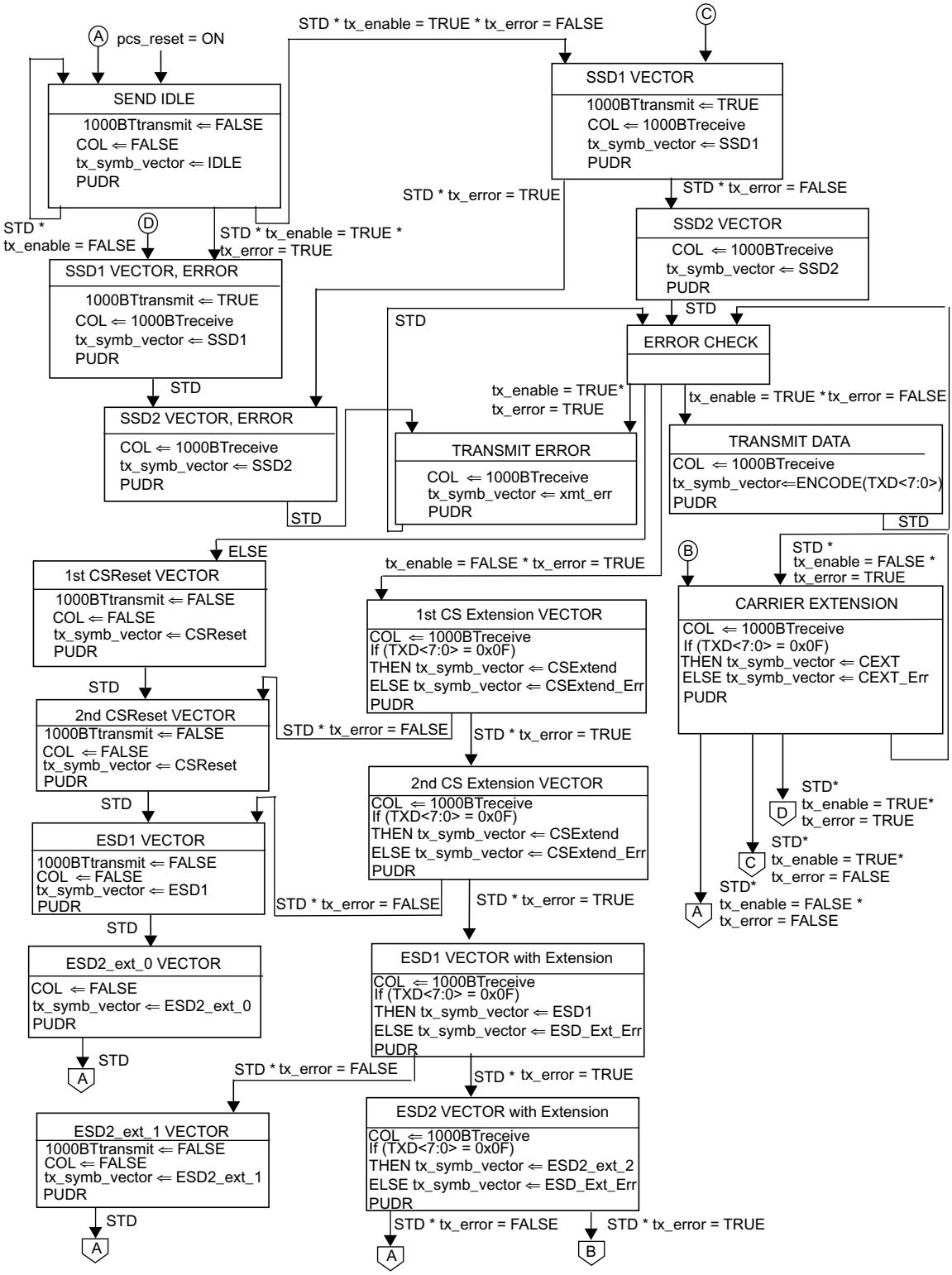


Figure 40–9—PCS Transmit state diagram

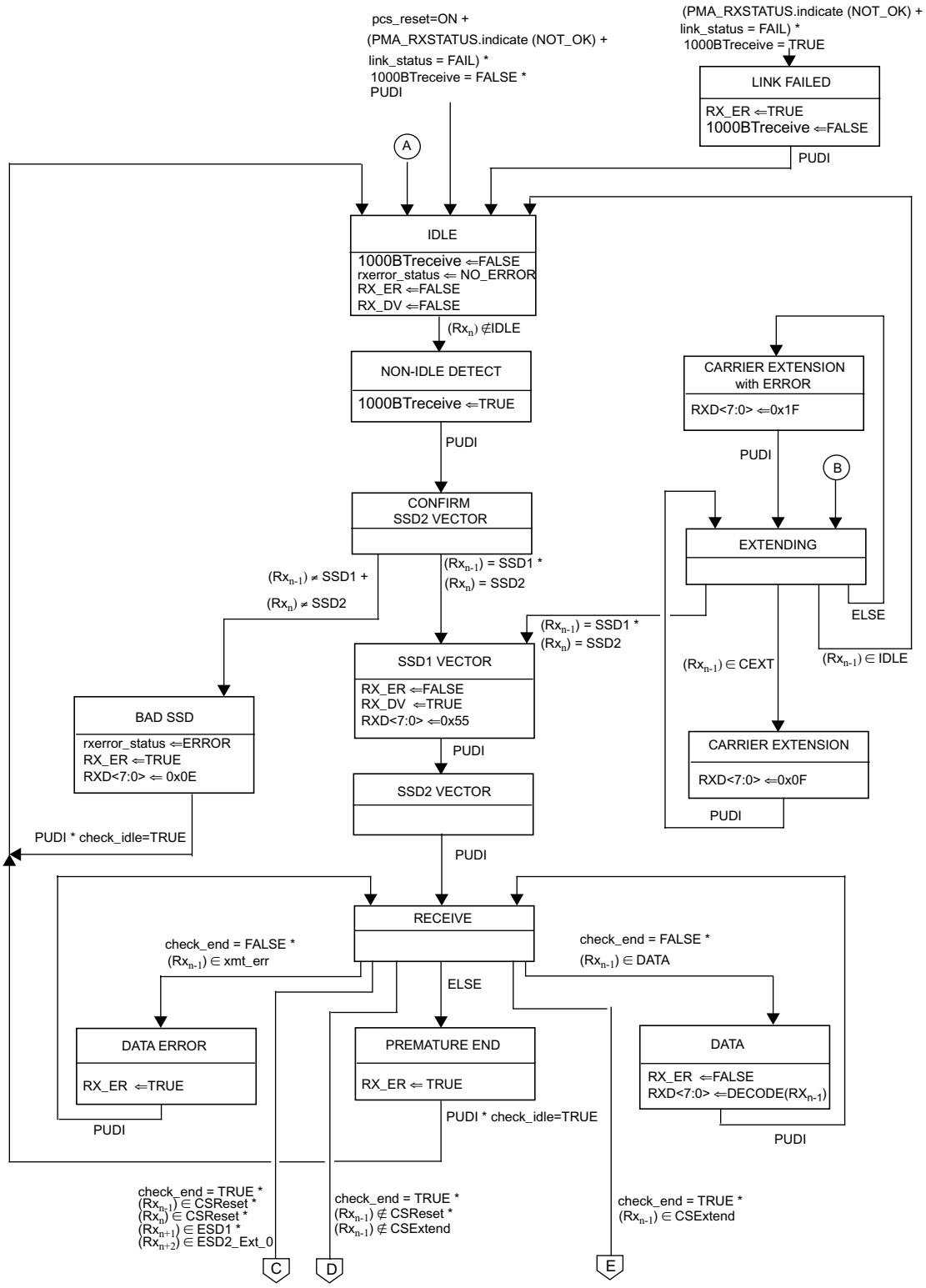


Figure 40–10a—PCS Receive state diagram, part a

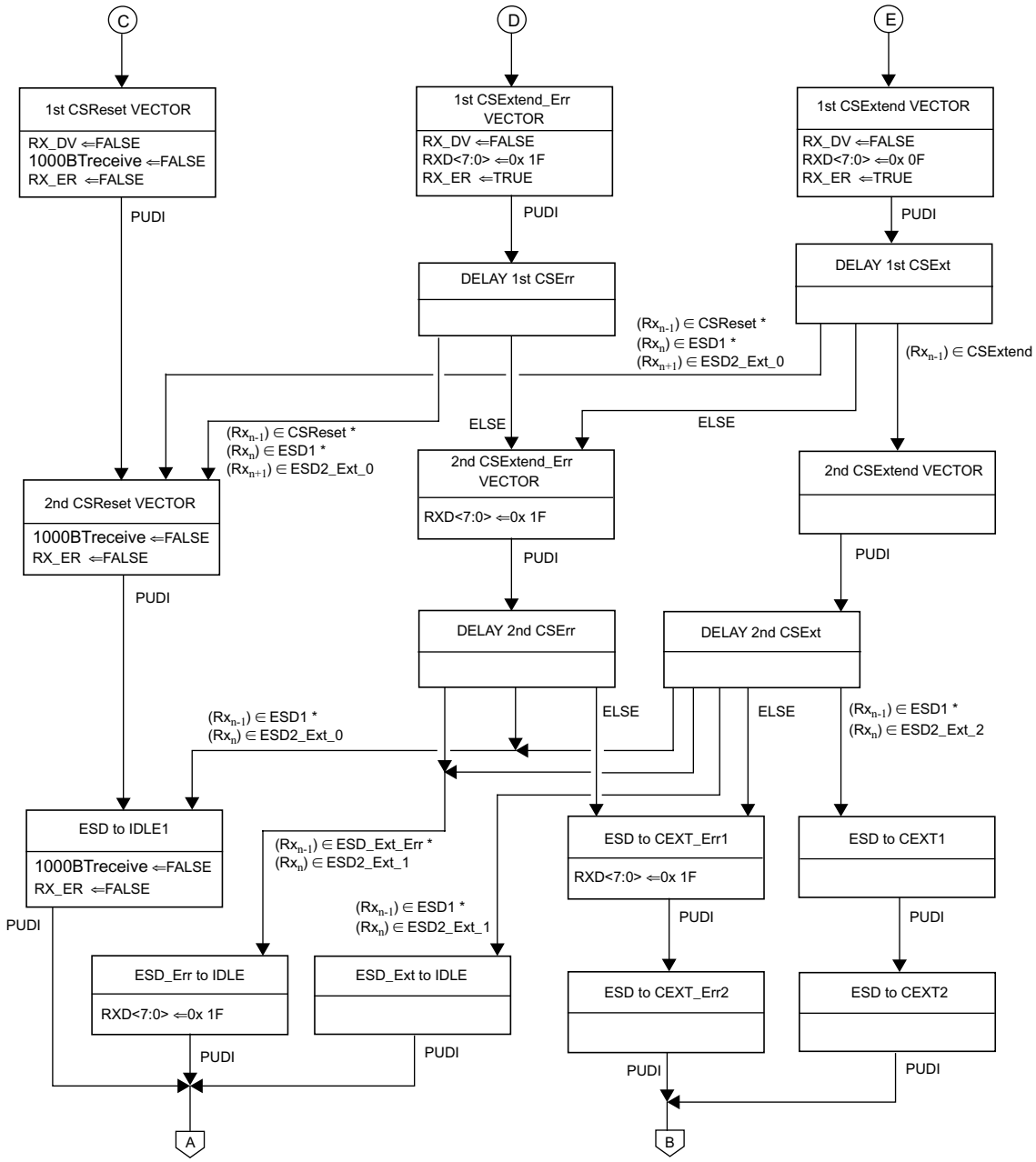


Figure 40–10b—PCS Receive state diagram, part b

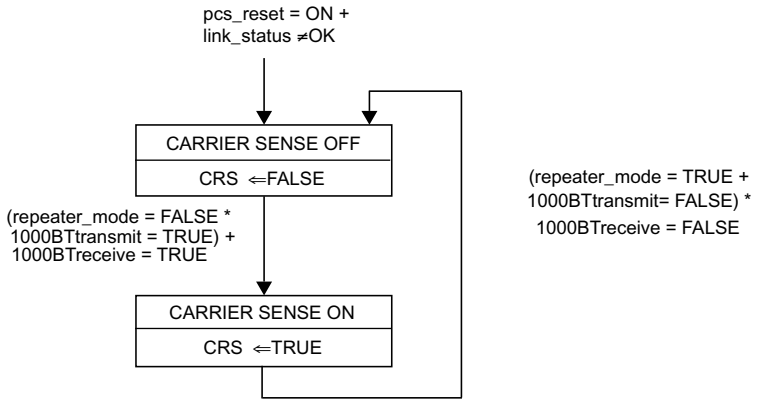


Figure 40–11—PCS Carrier Sense state diagram

40.3.4.1 Supplement to state diagram

Figure 40–12 reiterates the information shown in Figure 40–9 in timing diagram format. It is informative only. Time proceeds from left to right in the figure.

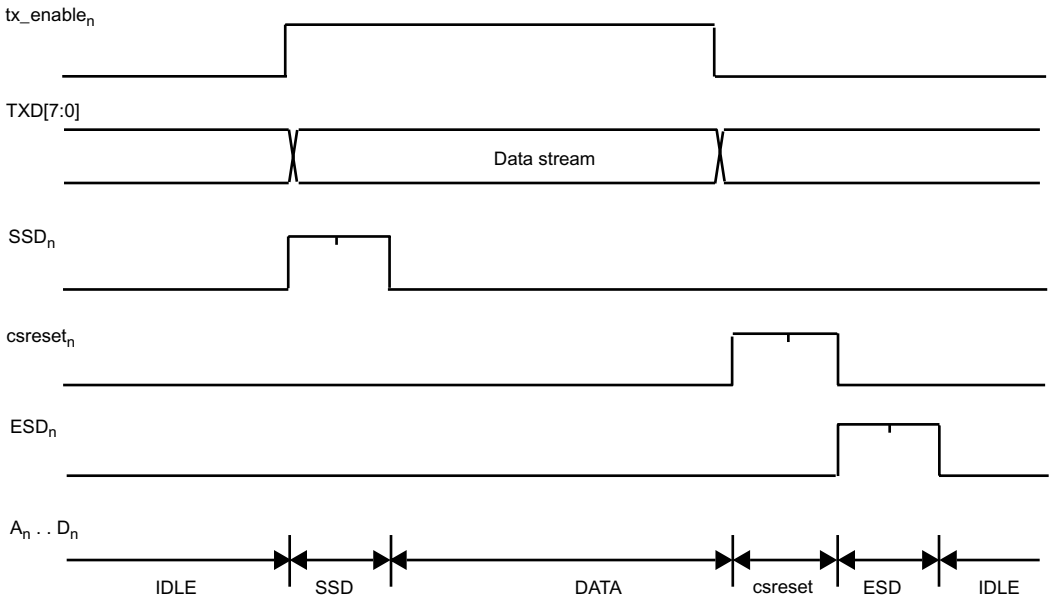


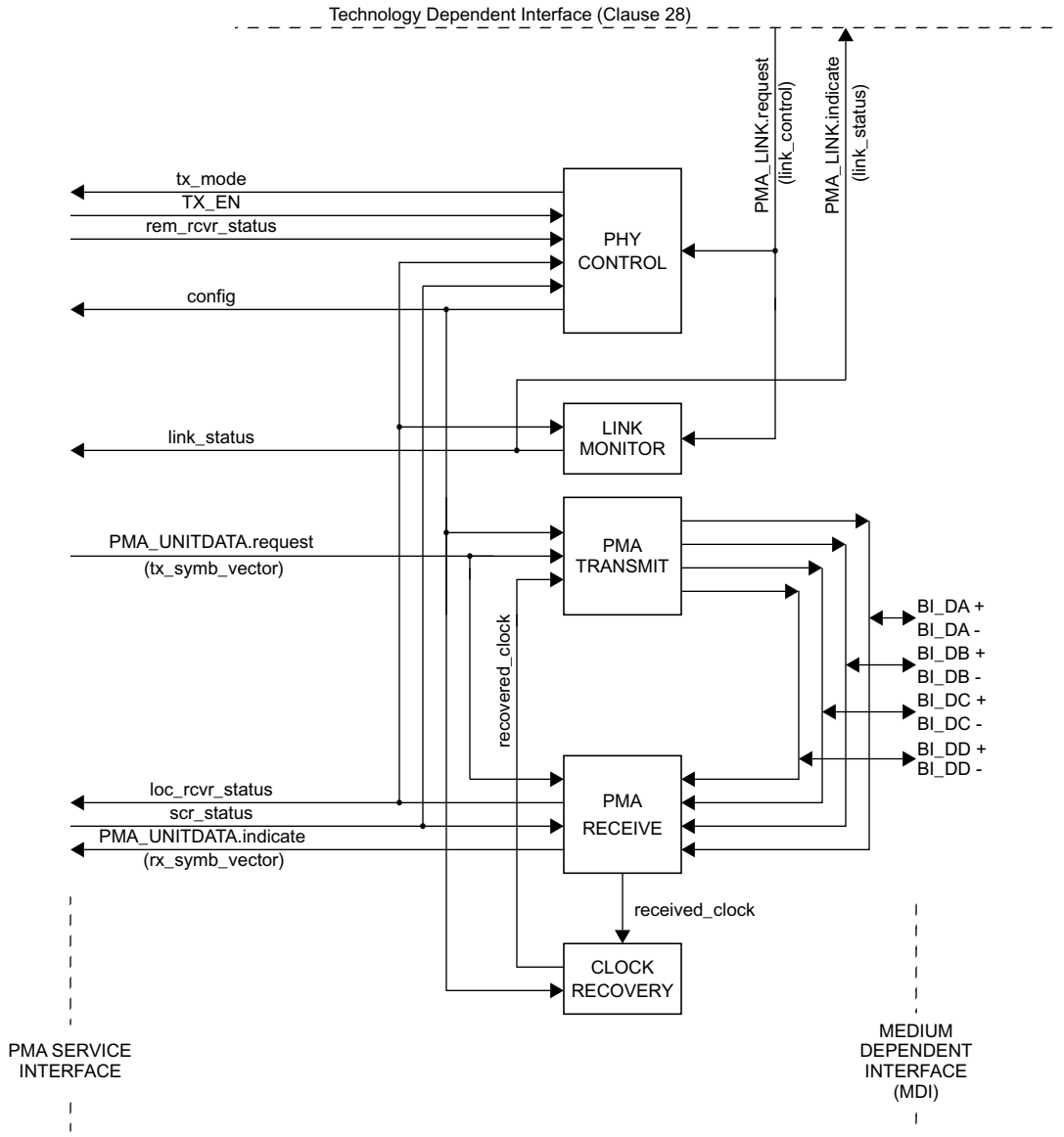
Figure 40–12—PCS sublayer to PMA timing

40.4 Physical Medium Attachment (PMA) sublayer

40.4.1 PMA functional specifications

The PMA couples messages from a PMA service interface specified in 40.2.2 to the 1000BASE-T baseband medium, specified in 40.7.

The interface between PMA and the baseband medium is the Medium Dependent Interface (MDI), which is specified in 40.8.



NOTE The recovered_clock arc is shown to indicate delivery of the recovered clock signal back to PMA TRANSMIT for loop timing.

Figure 40–13—PMA reference diagram

40.4.2 PMA functions

The PMA sublayer comprises one PMA Reset function and five simultaneous and asynchronous operating functions. The PMA operating functions are PHY Control, PMA Transmit, PMA Receive, Link Monitor, and Clock Recovery. All operating functions are started immediately after the successful completion of the PMA Reset function.

The PMA reference diagram, Figure 40–13, shows how the operating functions relate to the messages of the PMA Service interface and the signals of the MDI. Connections from the management interface, comprising the signals MDC and MDIO, to other layers are pervasive and are not shown in Figure 40–13. The management interface and its functions are specified in Clause 22.

40.4.2.1 PMA Reset function

The PMA Reset function shall be executed whenever one of the two following conditions occur:

a) Power on (see 36.2.5.1.3)
b) The receipt of a request for reset from the management entity

PMA Reset sets pcs_reset=ON while any of the above reset conditions hold true. All state diagrams take the open-ended pma_reset branch upon execution of PMA Reset. The reference diagrams do not explicitly show the PMA Reset function.

40.4.2.2 PMA Transmit function

The PMA Transmit function comprises four synchronous transmitters to generate four 5-level pulse-amplitude modulated signals on each of the four pairs BI_DA, BI_DB, BI_DC, and BI_DD. PMA Transmit shall continuously transmit onto the MDI pulses modulated by the quinary symbols given by tx_symb_vector[BI_DA], tx_symb_vector[BI_DB], tx_symb_vector[BI_DC] and tx_symb_vector[BI_DD], respectively. The four transmitters shall be driven by the same transmit clock, TX_TCLK. The signals generated by PMA Transmit shall follow the mathematical description given in 40.4.3.1, and shall comply with the electrical specifications given in 40.6.

When the PMA_CONFIG.indicate parameter config is MASTER, the PMA Transmit function shall source TX_TCLK from a local clock source while meeting the transmit jitter requirements of 40.6.1.2.5. When the PMA_CONFIG.indicate parameter config is SLAVE, the PMA Transmit function shall source TX_TCLK from the recovered clock of 40.4.2.6 while meeting the jitter requirements of 40.6.1.2.5.

40.4.2.3 PMA Receive function

The PMA Receive function comprises four independent receivers for quinary pulse-amplitude modulated signals on each of the four pairs BI_DA, BI_DB, BI_DC, and BI_DD. PMA Receive contains the circuits necessary to both detect quinary symbol sequences from the signals received at the MDI over receive pairs BI_DA, BI_DB, BI_DC, and BI_DD and to present these sequences to the PCS Receive function. The signals received at the MDI are described mathematically in 40.4.3.2. The PMA shall translate the signals received on pairs BI_DA, BI_DB, BI_DC, and BI_DB into the PMA_UNITDATA.indicate parameter rx_symb_vector with a symbol error rate of less than 10^{-10} over a channel meeting the requirements of 40.7.

To achieve the indicated performance, it is highly recommended that PMA Receive include the functions of signal equalization, echo and crosstalk cancellation, and sequence estimation. The sequence of code-groups assigned to tx_symb_vector is needed to perform echo and self near-end crosstalk cancellation.

The PMA Receive function uses the scr_status parameter and the state of the equalization, cancellation, and estimation functions to determine the quality of the receiver performance, and generates the loc_rcvr_status variable accordingly. The precise algorithm for generation of loc_rcvr_status is implementation dependent.

40.4.2.4 PHY Control function

PHY Control generates the control actions that are needed to bring the PHY into a mode of operation during which frames can be exchanged with the link partner. PHY Control shall comply with the state diagram description given in Figure 40–15.

During Auto-Negotiation PHY Control is in the DISABLE 1000BASE-T TRANSMITTER state and the transmitters are disabled. When the Auto-Negotiation process asserts link_control=ENABLE, PHY Control enters the SLAVE SILENT state. Upon entering this state, the maxwait timer is started and PHY Control forces transmission of zeros by setting tx_mode=SEND_Z. The transition out of the SLAVE SILENT state depends on whether the PHY is operating in MASTER or SLAVE mode. In MASTER mode, PHY Control transitions immediately to the TRAINING state. In SLAVE mode, PHY Control transitions to the TRAINING state only after the SLAVE PHY converges its decision feedback equilizer (DFE), acquires timing, and acquires its descrambler state, and sets scr_status=OK.

For the SLAVE PHY, the final convergence of the adaptive filter parameters is completed in the TRAINING state. The MASTER PHY performs all its receiver convergence functions in the TRAINING state. Upon entering the TRAINING state, the minwait_timer is started and PHY Control forces transmission into the idle mode by asserting tx_mode=SEND_I. After the PHY completes successful training and establishes proper receiver operations, PCS Transmit conveys this information to the link partner via transmission of the parameter loc_rcvr_status. (See $Sd_n[2]$ in 40.3.1.3.4.) The link partner's value for loc_rcvr_status is stored in the local device parameter rem_rcvr status. When the minwait_timer expires and the condition loc_rcvr_status=OK is satisfied, PHY Control transitions into either the SEND IDLE OR DATA state if rem_rcvr_status=OK or the SEND IDLE state if rem_rcvr_status=NOT_OK. On entry into either the SEND IDLE or SEND IDLE OR DATA states, the maxwait_timer is stopped and the minwait_timer is started.

The normal mode of operation corresponds to the SEND IDLE OR DATA state, where PHY Control asserts tx_mode=SEND_N and transmission of data over the link can take place. In this state, when no frames have to be sent, idle transmission takes place.

If unsatisfactory receiver operation is detected in the SEND IDLE OR DATA or SEND IDLE states (loc_rcvr_status=NOT_OK) and the minwait_timer has expired, transmission of the current frame is completed and PHY Control enters the SLAVE SILENT state. In the SEND IDLE OR DATA state, whenever a PHY that operates reliably detects unsatisfactory operation of the remote PHY (rem_rcvr_status=NOT_OK) and the minwait_timer has expired, it enters the SEND IDLE state where tx_mode=SEND_I is asserted and idle transmission takes place. In this state, encoding is performed with the parameter loc_rcvr_status=OK. As soon as the remote PHY signals satisfactory receiver operation (rem_rcvr_status=OK) and the minwait_timer has expired, the SEND IDLE OR DATA state is entered.

PHY Control may force the transmit scrambler state to be initialized to an arbitrary value by requesting the execution of the PCS Reset function defined in 40.3.1.1.

40.4.2.5 Link Monitor function

Link Monitor determines the status of the underlying receive channel and communicates it via the variable link_status. Failure of the underlying receive channel typically causes the PMA's clients to suspend normal operation.

The Link Monitor function shall comply with the state diagram of Figure 40–16.

Upon power on, reset, or release from power down, the Auto-Negotiation algorithm sets link_control=SCAN_FOR_CARRIER and, during this period, sends fast link pulses to signal its presence to a remote station. If the presence of a remote station is sensed through reception of fast link pulses, the Auto-Negotiation algorithm sets link_control=DISABLE and exchanges Auto-Negotiation information with the remote station. During this period, link_status=FAIL is asserted. If the presence of a remote 1000BASE-T station is established, the Auto-Negotiation algorithm permits full operation by setting link_control=ENABLE. As soon as reliable transmission is achieved, the variable link_status=OK is asserted, upon which further PHY operations can take place.

40.4.2.6 Clock Recovery function

The Clock Recovery function couples to all four receive pairs. It may provide independent clock phases for sampling the signals on each of the four pairs.

The Clock Recovery function shall provide clocks suitable for signal sampling on each line so that the symbol-error rate indicated in 40.4.2.3 is achieved. The received clock signal must be stable and ready for use when training has been completed (loc_rcvr_status=OK). The received clock signal is supplied to the PMA Transmit function by received_clock.

40.4.3 MDI

Communication through the MDI is summarized in 40.4.3.1 and 40.4.3.2.

40.4.3.1 MDI signals transmitted by the PHY

The quinary symbols to be transmitted by the PMA on the four pairs BI_DA, BI_DB, BI_DC, and BI_DD are denoted by tx_symb_vector[BI_DA], tx_symb_vector[BI_DB], tx_symb_vector[BI_DC], and tx_symb_vector[BI_DD], respectively. The modulation scheme used over each pair is 5-level Pulse Amplitude Modulation. PMA Transmit generates a pulse-amplitude modulated signal on each pair in the following form:

$$s(t) = \sum_k a_k h_1(t - kT)$$

In the above equation, a_k represents the quinary symbol from the set {2, 1, 0, –1, –2} to be transmitted at time kT, and $h_1(t)$ denotes the system symbol response at the MDI. This symbol response shall comply with the electrical specifications given in 40.6.

40.4.3.2 Signals received at the MDI

Signals received at the MDI can be expressed for each pair as pulse-amplitude modulated signals that are corrupted by noise as follows:

$$r(t) = \sum_k a_k h_2(t - kT) + w(t)$$

In this equation, $h_2(t)$ denotes the impulse response of the overall channel between the transmit symbol source and the receive MDI and $w(t)$ is a term that represents the contribution of various noise sources. The four signals received on pairs BI_DA, BI_DB, BI_DC, and BI_DD shall be processed within the PMA Receive function to yield the quinary received symbols rx_symb_vector[BI_DA], rx_symb_vector[BI_DB], rx_symb_vector[BI_DC], and rx_symb_vector[BI_DD].

40.4.4 Automatic MDI/MDI-X Configuration

Automatic MDI/MDI-X Configuration is intended to eliminate the need for crossover cables between similar devices. Implementation of an automatic MDI/MDI-X configuration is optional for 1000BASE-T devices. If an automatic configuration method is used, it shall comply with the following specifications. The assignment of pin-outs for a 1000BASE-T crossover function cable is shown in Table 40–12 in 40.8.

40.4.4.1 Description of Automatic MDI/MDI-X state machine

The Automatic MDI/MDI-X state machine facilitates switching the BI_DA(C)+ and BI_DA(C)– with the BI_DB(D)+ and BI_DB(D)– signals respectively prior to the auto-negotiation mode of operation so that FLPs can be transmitted and received in compliance with Clause 28 Auto-Negotiation specifications. The correct polarization of the crossover circuit is determined by an algorithm that controls the switching function. This algorithm uses an 11-bit Linear Feedback Shift Register (LFSR) to create a pseudo-random sequence that each end of the link uses to determine its proposed configuration. Upon making the selection to either MDI or MDI-X, the node waits for a specified amount of time while evaluating its receive channel to determine whether the other end of the link is sending link pulses or PHY-dependent data. If link pulses or PHY-dependent data are detected, it remains in that configuration. If link pulses or PHY-dependent data are not detected, it increments its LFSR and makes a decision to switch based on the value of the next bit. The state machine does not move from one state to another while link pulses are being transmitted.

40.4.4.2 Pseudo-random sequence generator

One possible implementation of the pseudo-random sequence generator using a linear-feedback shift register is shown in Figure 40–14. The bits stored in the shift register delay line at time n are denoted by S[10:0]. At each sample period, the shift register is advanced by one bit and one new bit represented by S[0] is generated. Switch control is determined by S[10].

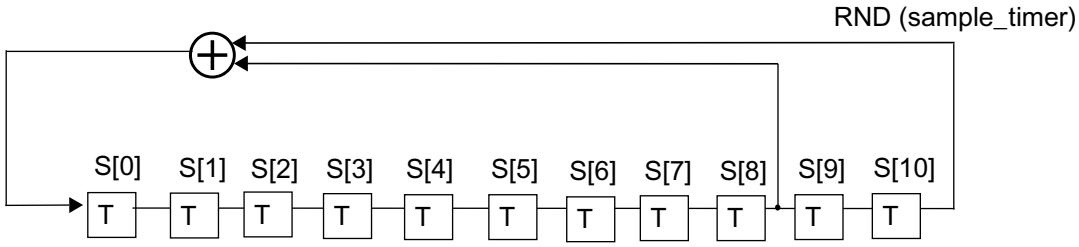


Figure 40–14—Automatic MDI/MDI-X linear-feedback shift register

40.4.5 State variables

40.4.5.1 State diagram variables

config

The PMA shall generate this variable continuously and pass it to the PCS via the PMA_CONFIG.indicate primitive.

Values: MASTER or SLAVE

link_control

This variable is defined in 28.2.6.2.

Link_Det

This variable indicates linkpulse = true or link_status = READY has occurred at the receiver since the last time sample_timer has been started.

Values: TRUE: linkpulse = true or link_status = READY has occurred since the last time sample_timer has been started.

FALSE: otherwise

linkpulse

This variable is defined in 28.2.6.3.

link_status

This variable is defined in 28.2.6.1.

loc_rcvr_status

Variable set by the PMA Receive function to indicate correct or incorrect operation of the receive link for the local PHY.

Values: OK: The receive link for the local PHY is operating reliably.
NOT_OK: Operation of the receive link for the local PHY is unreliable.

MDI_Status

This variable defines the condition of the Automatic MDI/MDI-X physical connection.

Values: MDI: The BI_DA, BI_DB, BI_DC, and BI_DD pairs follow the connections as described in the MDI column of Table 40–12.

MDI-X: The BI_DA, BI_DB, BI_DC, and BI_DD pairs follow the connections as described in the MDI-X column of Table 40–12.

pma_reset

Allows reset of all PMA functions.

Values: ON or OFF

Set by: PMA Reset

rem_rcvr_status

Variable set by the PCS Receive function to indicate whether correct operation of the receive link for the remote PHY is detected or not.

Values: OK: The receive link for the remote PHY is operating reliably.

NOT_OK: Reliable operation of the receive link for the remote PHY is not detected.

RND (sample_timer)

This variable is defined as bit S[10] of the LSFR described in 40.4.4.2

scr_status

The scr_status parameter as communicated by the PMA_SCRSTATUS.request primitive.

Values: OK: The descrambler has achieved synchronization.

NOT_OK: The descrambler is not synchronized.

T_Pulse

This variable indicates that a linkpulse is being transmitted to the MDI.

Values: TRUE: Pulse being transmitted to the MDI

FALSE: Otherwise

tx_enable
: The tx_enable parameter generated by PCS Transmit as specified in Figure 40–8.

  Values: TRUE or FALSE as per 40.3.3.1.

tx_mode
: PCS Transmit sends code-groups according to the value assumed by this variable.

  Values: SEND_N: This value is continuously asserted when transmission of sequences of code-groups representing a GMII data stream, control information, or idle mode is to take place.

  SEND_I: This value is continuously asserted when transmission of sequences of code-groups representing the idle mode is to take place.

  SEND_Z: This value is asserted when transmission of zero code-groups is to take place.

40.4.5.2 Timers

All timers operate in the manner described in 14.2.3.2 with the following addition. A timer is reset and stops counting upon entering a state where "stop timer" is asserted.

A_timer
: An asynchronous (to the Auto-Crossover State Machine) free-running timer that provides for a relatively arbitrary reset of the state machine to its initial state. This timer is used to reduce the probability of a lock-up condition where both nodes have the same identical seed initialization at the same point in time.

  Values: The condition A_timer_done becomes true upon timer expiration.

  Duration: This timer shall have a period of 1.3 s ± 25%.

  Initialization of A_timer is implementation specific.

maxwait_timer
: A timer used to limit the amount of time during which a receiver dwells in the SLAVE SILENT and TRAINING states. The timer shall expire 750 ± 10 ms if config = MASTER or 350 ± 5 ms if config = SLAVE. This timer is used jointly in the PHY Control and Link Monitor state diagrams. The maxwait_timer is tested by the Link Monitor to force link_status to be set to FAIL if the timer expires and loc_rcvr_status is NOT_OK. See Figure 40–15.

minwait_timer
: A timer used to determine the minimum amount of time the PHY Control stays in the TRAINING, SEND IDLE, or DATA states. The timer shall expire 1 ± 0.1μs after being started.

sample_timer
: This timer provides a long enough sampling window to ensure detection of Link Pulses or link_status, if they exist at the receiver.

  Values: The condition sample_timer_done becomes true upon timer expiration.

  Duration: This timer shall have a period of 62 ± 2 ms.

stabilize_timer
: A timer used to control the minimum time that loc_rcvr_status must be OK before a transition to Link Up can occur. The timer shall expire 1 ± 0.1 μs after being started.

40.4.6 State Diagrams

40.4.6.1 PHY Control state diagram

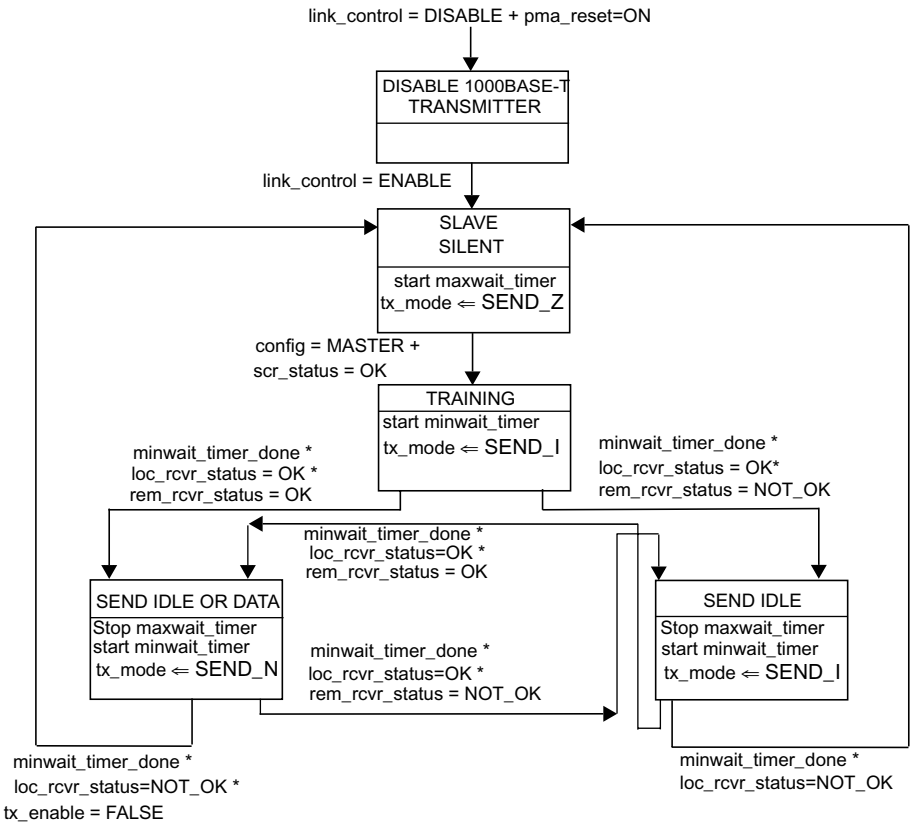
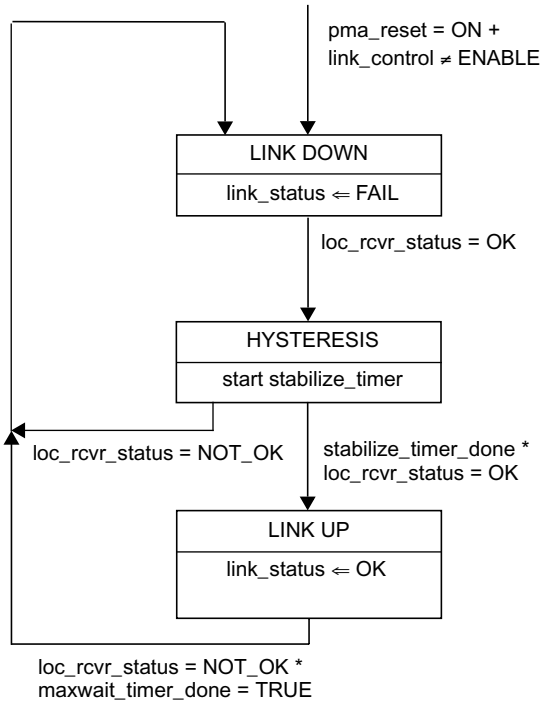


Figure 40–15—PHY Control state diagram

40.4.6.2 Link Monitor state diagram



NOTES
1 maxwait_timer is started in PHY Control state diagram (see Figure 40—15).≥
2 The variables link_control and link_status are designated as link_control_(1GigT) and link_status_(1GigT), respectively, by the Auto-Negotiation Arbitration state diagram (Figure 28—16)

Figure 40–16—Link Monitor state diagram

40.4.6.2.1 Auto Crossover state diagram

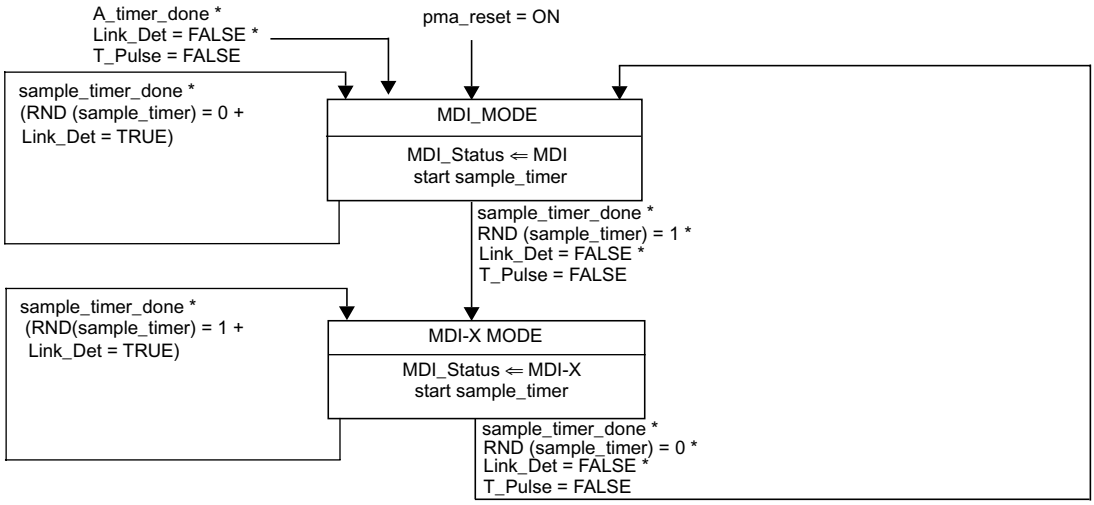


Figure 40–17—Auto Crossover state diagram

40.5 Management interface

1000BASE-T makes extensive use of the management functions provided by the MII Management Interface (see 22.2.4), and the communication and self-configuration functions provided by Auto-Negotiation (Clause 28.)

40.5.1 Support for Auto-Negotiation

All 1000BASE-T PHYs shall provide support for Auto-Negotiation (Clause 28) and shall be capable of operating as MASTER or SLAVE.

Auto-Negotiation is performed as part of the initial set-up of the link, and allows the PHYs at each end to advertise their capabilities (speed, PHY type, half or full duplex) and to automatically select the operating mode for communication on the link. Auto-negotiation signaling is used for the following two primary purposes for 1000BASE-T:

a) To negotiate that the PHY is capable of supporting 1000BASE-T half duplex or full duplex transmission.
b) To determine the MASTER-SLAVE relationship between the PHYs at each end of the link.

This relationship is necessary for establishing the timing control of each PHY. The 1000BASE-T MASTER PHY is clocked from a local source. The SLAVE PHY uses loop timing where the clock is recovered from the received data stream.

40.5.1.1 1000BASE-T use of registers during Auto-Negotiation

A 1000BASE-T PHY shall use the management register definitions and values specified in Table 40–3.

Table 40–3—1000BASE-T Registers

Register	Bit	Name	Description	Type [a]
0	0.15:0	MII control register	Defined in 28.2.4.1.1	RO
1	1.15:0	MII status register	Defined in 28.2.4.1.2	RO
4	4.15:0	Auto-Negotiation advertisement register	The Selector Field (4.4:0) is set to the appropriate code as specified in Annex 28A. The Technology Ability Field bits 4.12:5 are set to the appropriate code as specified in Annexes 28B and 28D. Bit 4.15 is set to logical one to indicate the desired exchange of Next Pages describing the gigabit extended capabilities.	R/W
5	5.15:0	Auto-Negotiation link partner ability register	Defined in 28.2.4.1.4. 1000BASE-T implementations do not use this register to store Auto-Negotiation Link Partner Next Page data.	RO
6	6.15:0	Auto-Negotiation expansion register	Defined in 28.2.4.1.5	RO
7	7.15:0	Auto-Negotiation Next Page transmit register	Defined in 28.2.4.1.6	R/W
8	8.15:0	Auto-Negotiation link partner Next Page register	Defined in 28.2.4.1.8	RO
9	9.15:13	Test mode bits	Transmitter test mode operations are defined by bits 9.15:13 as described in 40.6.1.1.2 and Table 40–7. The default values for bits 9.15:13 are all zero.	R/W
9	9.12	MASTER-SLAVE Manual Config Enable	1=Enable MASTER-SLAVE Manual configuration value 0=Disable MASTER-SLAVE Manual configuration value Default bit value is 0.	R/W
9	9.11	MASTER-SLAVE Config Value	1=Configure PHY as MASTER during MASTER-SLAVE negotiation, only when 9.12 is set to logical one. 0=Configure PHY as SLAVE during MASTER-SLAVE negotiation, only when 9.12 is set to logical one. Default bit value is 0.	R/W
9	9.10	Port type	Bit 9.10 is to be used to indicate the preference to operate as MASTER (multiport device) or as SLAVE (single-port device) if the MASTER-SLAVE Manual Configuration Enable bit, 9.12, is not set. Usage of this bit is described in 40.5.2. 1=Multiport device 0=single-port device	R/W
9	9.9	1000BASE-T Full Duplex	1 = Advertise PHY is 1000BASE-T full duplex capable. 0 = Advertise PHY is not 1000BASE-T full duplex capable.	R/W
9	9.8	1000BASE-T Half Duplex	1 = Advertise PHY is 1000BASE-T half duplex capable. 0 = Advertise PHY is not 1000BASE-T half duplex capable.	R/W

[a] R/W = Read/Write, RO = Read Only, SC = Self Clearing, LH = Latch High

Table 40–3—1000BASE-T Registers *(continued)*

Register	Bit	Name	Description	Type [a]
9	9.7:0	Reserved	Write as 0, ignore on read.	R/W
10	10.15	MASTER-SLAVE configuration fault	Configuration fault, as well as the criteria and method of fault detection, is PHY specific. The MASTER-SLAVE Configuration Fault bit will be cleared each time register 10 is read via the management interface and will be cleared by a 1000BASE-T PMA reset. This bit will self clear on Auto-Negotiation enable or Auto-Negotiation complete.This bit will be set if the number of failed MASTER-SLAVE resolutions reaches 7. For l000BASE-T, the fault condition will occur when both PHYs are forced to be MASTERs or SLAVEs at the same time using bits 9.12 and 9.11. Bit 10.15 should be set via the MASTER-SLAVE Configuration Resolution function described in 40.5.2. 1 = MASTER-SLAVE configuration fault detected 0 = No MASTER-SLAVE configuration fault detected	RO/LH/SC
10	10.14	MASTER-SLAVE configuration resolution	1 = Local PHY configuration resolved to MASTER 0 = Local PHY configuration resolved to SLAVE	RO
10	10.13	Local Receiver Status	1 = Local Receiver OK (loc_rcvr_status=OK) 0 = Local Receiver not OK (loc_rcvr_status=NOT_OK) Defined by the value of loc_rcvr_status as per 40.4.5.1.	RO
10	10.12	Remote Receiver Status	1 = Remote Receiver OK (rem_rcvr_status=OK) 0 = Remote Receiver not OK (rem_rcvr_status=NOT_OK) Defined by the value of rem_rcvr_status as per 40.4.5.1.	RO
10	10.11	LP 1000T FD	1 = Link Partner is capable of 1000BASE-T full duplex 0 = Link Partner is not capable of 1000BASE-T full duplex This bit is guaranteed to be valid only when the Page received bit (6.1) has been set to 1.	RO
10	10.10	LP 1000T HD	1 = Link Partner is capable of 1000BASE-T half duplex 0 = Link Partner is not capable of 1000BASE-T half duplex This bit is guaranteed to be valid only when the Page received bit (6.1) has been set to 1.	RO
10	10.9:8	Reserved	Reserved	RO
10	10.7:0	Idle Error Count	Bits 10.7:0 indicate the Idle Error count, where 10.7 is the most significant bit. These bits contain a cumulative count of the errors detected when the receiver is receiving idles and PMA_TXMODE.indicate is equal to SEND_N (indicating that both local and remote receiver status have been detected to be OK). The counter is incremented every symbol period that rxerror_status is equal to ERROR. These bits are reset to all zeros when the error count is read by the management function or upon execution of the PCS Reset function and are to be held at all ones in case of overflow (see 30.5.1.1.11).	RO/SC
15	15.15:12	Extended status register	See 22.2.4.4	RO

[a] R/W = Read/Write, RO = Read Only, SC = Self Clearing, LH = Latch High

40.5.1.2 1000BASE-T Auto-Negotiation page use

1000BASE-T PHYs shall exchange one Auto-Negotiation Base Page, a 1000BASE-T formatted Next Page, and two 1000BASE-T unformatted Next Pages in sequence, without interruption, as specified in Table 40–4. Additional Next Pages can be exchanged as described in Annex 40C.

Note that the Acknowledge 2 bit is not utilized and has no meaning when used for the 1000BASE-T message page exchange.

Table 40–4—1000BASE-T Base and Next Pages bit assignments

Bit	Bit definition	Register location
BASE PAGE		
D15	1 (to indicate that Next Pages follow)	
D14:D1	As specified in 28.2.1.2	Management register 4
PAGE 0 (Message Next Page)		
M10:M0	8	
PAGE 1 (Unformatted Next Page)		
U10:U5	Reserved transmit as 0	
U4	1000BASE-T half duplex (1 = half duplex and 0 = no half duplex)	GMII register 9.8 (MASTER-SLAVE Control register)
U3	1000BASE-T full duplex (1 = full duplex and 0 = no full duplex)	GMII register 9.9 (MASTER-SLAVE Control register)
U2	1000BASE-T port type bit (1 = multiport device and 0 = single-port device)	GMII register 9.10 (MASTER-SLAVE Control register)
U1	1000BASE-T MASTER-SLAVE Manual Configuration value (1 = MASTER and 0 = SLAVE.) This bit is ignored if 9.12 = 0.	GMII register 9.11 (MASTER-SLAVE Control register)
U0	1000BASE-T MASTER-SLAVE Manual Configuration Enable (1 = Manual Configuration Enable.) This bit is intended to be used for manual selection in a particular MASTER-SLAVE mode and is to be used in conjunction with bit 9.11.	GMII register 9.12 (MASTER-SLAVE Control register)
PAGE 2 (Unformatted Next Page)		
U10	1000BASE-T MASTER-SLAVE Seed Bit 10 (SB10) (MSB)	MASTER-SLAVE Seed Value (10:0)
U9	1000BASE-T MASTER-SLAVE Seed Bit 9 (SB9)	
U8	1000BASE-T MASTER-SLAVE Seed Bit 8 (SB8)	
U7	1000BASE-T MASTER-SLAVE Seed Bit 7 (SB7)	
U6	1000BASE-T MASTER-SLAVE Seed Bit 6 (SB6)	
U5	1000BASE-T MASTER-SLAVE Seed Bit 5 (SB5)	
U4	1000BASE-T MASTER-SLAVE Seed Bit 4 (SB4)	
U3	1000BASE-T MASTER-SLAVE Seed Bit 3 (SB3)	
U2	1000BASE-T MASTER-SLAVE Seed Bit 2 (SB2)	
U1	1000BASE-T MASTER-SLAVE Seed Bit 1 (SB1)	
U0	1000BASE-T MASTER-SLAVE Seed Bit 0 (SB0)	

40.5.1.3 Sending Next Pages

Implementors who do not wish to send additional Next Pages (i.e., Next Pages in addition to those required to perform PHY configuration as defined in this clause) can use Auto-Negotiation as defined in Clause 28 and the Next Pages defined in 40.5.1.2. Implementors who wish to send additional Next Pages are advised to consult Annex 40C.

40.5.2 MASTER-SLAVE configuration resolution

Since both PHYs that share a link segment are capable of being MASTER or SLAVE, a prioritization scheme exists to ensure that the correct mode is chosen. The MASTER-SLAVE relationship shall be determined during Auto-Negotiation using Table 40–5 with the 1000BASE-T Technology Ability Next Page bit values specified in Table 40–4 and information received from the link partner. This process is conducted at the entrance to the FLP LINK GOOD CHECK state shown in the Arbitration state diagram (Figure 28–13.)

The following four equations are used to determine these relationships:

manual_MASTER = U0 * U1
manual_SLAVE = U0 * !U1
single-port device = !U0 * !U2,
multiport device = !U0 * U2

where

U0 is bit 0 of unformatted page 1,
U1 is bit 1 of unformatted page 1, and
U2 is bit 2 of unformatted page 1 (see Table 40–4).

A 1000BASE-T PHY is capable of operating either as the MASTER or SLAVE. In the scenario of a link between a single-port device and a multiport device, the preferred relationship is for the multiport device to be the MASTER PHY and the single-port device to be the SLAVE. However, other topologies may result in contention. The resolution function of Table 40–5 is defined to handle any relationship conflicts.

Table 40–5—1000BASE-T MASTER-SLAVE configuration resolution table

Local device type	Remote device type	Local device resolution	Remote device resolution
single-port device	multiport device	SLAVE	MASTER
single-port device	manual_MASTER	SLAVE	MASTER
manual_SLAVE	manual_MASTER	SLAVE	MASTER
manual_SLAVE	multiport device	SLAVE	MASTER
multiport device	manual_MASTER	SLAVE	MASTER
manual_SLAVE	single-port device	SLAVE	MASTER
multiport device	single-port device	MASTER	SLAVE
multiport device	manual_SLAVE	MASTER	SLAVE

Table 40–5—1000BASE-T MASTER-SLAVE configuration resolution table *(continued)*

Local device type	Remote device type	Local device resolution	Remote device resolution
manual_MASTER	manual_SLAVE	MASTER	SLAVE
manual_MASTER	single-port device	MASTER	SLAVE
single-port device	manual_SLAVE	MASTER	SLAVE
manual_MASTER	multiport device	MASTER	SLAVE
multiport device	multiport device	The device with the higher SEED value is configured as MASTER, otherwise SLAVE.	The device with the higher SEED value is configured as MASTER, otherwise SLAVE.
single-port device	single-port device	The device with the higher SEED value is configured as MASTER, otherwise SLAVE	The device with the higher SEED value is configured as MASTER, otherwise SLAVE.
manual_SLAVE	manual_SLAVE	MASTER-SLAVE configuration fault	MASTER-SLAVE configuration fault
manual_MASTER	manual_MASTER	MASTER-SLAVE configuration fault	MASTER-SLAVE configuration fault

The rationale for the hierarchy illustrated in Table 40–5 is straightforward. A 1000BASE-T multiport device has higher priority than a single-port device to become the MASTER. In the case where both devices are of the same type, e.g., both devices are multiport devices, the device with the higher MASTER-SLAVE seed bits (SB0...SB10), where SB10 is the MSB, shall become the MASTER and the device with the lower seed value shall become the SLAVE. In case both devices have the same seed value, both should assert link_status_1GigT=FAIL (as defined in 28.3.1) to force a new cycle through Auto-Negotiation. Successful completion of the MASTER-SLAVE resolution shall be treated as MASTER-SLAVE configuration resolution complete.

The method of generating a random or pseudorandom seed is left to the implementor. The generated random seeds should belong to a sequence of independent, identically distributed integer numbers with a uniform distribution in the range of 0 to $2^{11}-2$. The algorithm used to generate the integer should be designed to minimize the correlation between the number generated by any two devices at any given time. A seed counter shall be provided to track the number of seed attempts. The seed counter shall be set to zero at start-up and shall be incremented each time a seed is generated. When MASTER-SLAVE resolution is complete, the seed counter shall be reset to 0 and bit 10.15 shall be set to logical zero. A MASTER-SLAVE resolution fault shall be declared if resolution is not reached after the generation of seven seeds.

The MASTER-SLAVE Manual Configuration Enable bit (control register bit 9.12) and the MASTER-SLAVE Config Value bit (control register bit 9.11) are used to manually set a device to become the MASTER or the SLAVE. In case both devices are manually set to become the MASTER or the SLAVE, this condition shall be flagged as a MASTER-SLAVE Configuration fault condition, thus the MASTER-SLAVE Configuration fault bit (status register bit 10.15) shall be set to logical one. The MASTER-SLAVE Configuration fault condition shall be treated as MASTER-SLAVE configuration resolution complete and link_status_1GigT shall be set to FAIL, because the MASTER-SLAVE relationship was not resolved. This will force a new cycle through Auto-Negotiation after the link_fail_inhibit_timer has expired. Determination of MASTER-SLAVE values occur on the entrance to the FLP LINK GOOD CHECK state (Figure 28–16) when the highest common denominator (HCD) technology is 1000BASE-T. The resulting MASTER-SLAVE value is used by the 1000BASE-T PHY control (40.4.2.4).

If MASTER-SLAVE Manual Configuration is disabled (bit 9.12 is set to 0) and the local device detects that both the local device and the remote device are of the same type (either multiport device or single-port device) and that both have generated the same random seed, it generates and transmits a new random seed for MASTER-SLAVE negotiation by setting link_status to FAIL and cycling through the Auto-Negotiation process again.

The MASTER-SLAVE configuration process returns one of the three following outcomes:

a) *Successful:* Bit 10.15 of the 1000BASE-T Status Register is set to logical zero and bit 10.14 is set to logical one for MASTER resolution or for logical zero for SLAVE resolution. 1000BASE-T returns control to Auto_Negotiation (at the entrance to the FLP LINK GOOD CHECK state in Figure 28–16) and passes the value MASTER or SLAVE to PMA_CONFIG.indicate (see 40.2.4.)
b) *Unsuccessful:* link_status_1GigT is set to FAIL and Auto-Negotiation restarts (see Figure 28–16.)
c) *Fault detected:* (This happens when both end stations are set for manual configuration and both are set to MASTER or both are set to SLAVE.) Bit 10.15 of the 1000BASE-T Status Register is set to logical one to indicate that a configuration fault has been detected. This bit also is set when seven attempts to configure the MASTER SLAVE relationship via the seed method have failed. When a fault is detected, link_status_1GigT is set to FAIL, causing Auto-Negotiation to cycle through again.

NOTE—MASTER-SLAVE arbitration only occurs if 1000BASE-T is selected as the highest common denominator; otherwise, it is assumed to have passed this condition.

40.6 PMA electrical specifications

This subclause defines the electrical characteristics of the PMA.

Common-mode tests use the common-mode return point as a reference.

40.6.1 PMA-to-MDI interface tests

40.6.1.1 Isolation requirement

The PHY shall provide electrical isolation between the port device circuits, including frame ground (if any) and all MDI leads. This electrical separation shall withstand at least one of the following electrical strength tests:

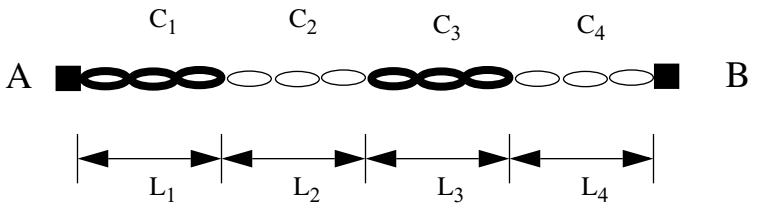
a) 1500 V rms at 50 Hz to 60 Hz for 60 s, applied as specified in Section 5.3.2 of IEC 60950: 1991.
b) 2250 Vdc for 60 s, applied as specified in Section 5.3.2 of IEC 60950: 1991.
c) A sequence of ten 2400 V impulses of alternating polarity, applied at intervals of not less than 1 s. The shape of the impulses shall be 1.2/50 μs (1.2 μs virtual front time, 50 μs virtual time or half value), as defined in IEC 60060.

There shall be no insulation breakdown, as defined in Section 5.3.2 of IEC 60950: 1991, during the test. The resistance after the test shall be at least 2 MΩ, measured at 500 Vdc.

40.6.1.1.1 Test channel

To perform the transmitter MASTER-SLAVE timing jitter tests described in this clause, a test channel is required to ensure that jitter is measured under conditions of poor signal to echo ratio. This test channel shall be constructed by combining 100 and 120 Ω cable segments that both meet or exceed ISO/IEC 11801 Category 5 specifications for each pair, as shown in Figure 40–18, with the lengths and additional restrictions on parameters described in Table 40–6. The ends of the test channel shall be terminated with connectors meeting or exceeding ANSI/TIA/EIA-568-A:1995 or ISO/IEC 11801:1995 Category 5 specifications. The

return loss of the resulting test channel shall meet the return loss requirements of 40.7.2.3 and the crosstalk requirements of 40.7.3.



Identical for each of the four pairs.

Figure 40–18—Test channel topology for each cable pair

Table 40–6—Test channel cable segment specifications

Cable segment	Length (meters)	Characteristic impedance (at frequencies > 1 MHz)	Attenuation (per 100 meters at 31.25 MHz)
1	L_1=1.20	120 ± 5Ω	7.8 to 8.8 dB
2	L_2=x	100 ± 5Ω	10.8 to 11.8 dB
3	L_3=1.48	120 ± 5Ω	7.8 to 8.8 dB
4	L_4=y	100 ± 5Ω	10.8 to 11.8 dB

NOTE—x is chosen so that the total delay of segments C1, C2, and C3, averaged across all pairs, is equal to 570 ns at 31.25 MHz; however, if this would cause the total attenuation of segments C1, C2, and C3, averaged across all pairs, to exceed the worst case insertion loss specified in 40.7.2.1 then x is chosen so that the total attenuation of segments C1, C2, and C3, averaged across all pairs, does not violate 40.7.2.1 at any frequencies. The value of y is chosen so that the total attenuation of segments C1, C2, C3, and C4, averaged across all pairs, does not violate 40.7.2.1 at any frequency (y may be 0).

40.6.1.1.2 Test modes

The test modes described below shall be provided to allow for testing of the transmitter waveform, transmitter distortion, and transmitted jitter.

For a PHY with a GMII interface, these modes shall be enabled by setting bits 9.13:15 (1000BASE-T Control Register) of the GMII Management register set as shown in Table 40–7. These test modes shall only change the data symbols provided to the transmitter circuitry and shall not alter the electrical and jitter characteristics of the transmitter and receiver from those of normal (non-test mode) operation. PHYs without a GMII shall provide a means to enable these modes for conformance testing.

Table 40–7—GMII management register settings for test modes

Bit 1 (9.15)	Bit 2 (9.14)	Bit 3 (9.13)	Mode
0	0	0	Normal operation
0	0	1	Test mode 1—Transmit waveform test
0	1	0	Test mode 2—Transmit jitter test in MASTER mode
0	1	1	Test mode 3—Transmit jitter test in SLAVE mode
1	0	0	Test mode 4—Transmitter distortion test
1	0	1	Reserved, operations not identified.
1	1	0	Reserved, operations not identified.
1	1	1	Reserved, operations not identified.

When test mode 1 is enabled, the PHY shall transmit the following sequence of data symbols A_n, B_n, C_n, D_n, of 40.3.1.3.6 continually from all four transmitters:

{{+2 followed by 127 0 symbols}, {–2 followed by 127 0 symbols},{+1 followed by 127 0 symbols}, {–1 followed by 127 0 symbols}, {128 +2 symbols, 128 –2 symbols, 128 +2 symbols, 128 –2 symbols}, {1024 0 symbols}}

This sequence is repeated continually without breaks between the repetitions when the test mode is enabled. A typical transmitter output is shown in Figure 40–19. The transmitter shall time the transmitted symbols from a 125.00 MHz ± 0.01% clock in the MASTER timing mode.

When test mode 2 is enabled, the PHY shall transmit the data symbol sequence {+2, –2} repeatedly on all channels. The transmitter shall time the transmitted symbols from a 125.00 MHz ± 0.01% clock in the MASTER timing mode.

When test mode 3 is enabled, the PHY shall transmit the data symbol sequence {+2, –2} repeatedly on all channels. The transmitter shall time the transmitted symbols from a 125.00 MHz ± 0.01% clock in the SLAVE timing mode. A typical transmitter output for transmitter test modes 2 and 3 is shown in Figure 40–20.

When test mode 4 is enabled, the PHY shall transmit the sequence of symbols generated by the following scrambler generator polynomial, bit generation, and level mappings:

$$g_{s1} = 1 + x^9 + x^{11}$$

The maximum-length shift register used to generate the sequences defined by this polynomial shall be updated once per symbol interval (8 ns). The bits stored in the shift register delay line at a particular time n are denoted by $Scr_n[10:0]$. At each symbol period the shift register is advanced by one bit and one new bit represented by $Scr_n[0]$ is generated. Bits $Scr_n[8]$ and $Scr_n[10]$ are exclusive OR'd together to generate the next $Scr_n[0]$ bit. The bit sequences, $x0_n$, $x1_n$, and $x2_n$, generated from combinations of the scrambler bits as shown in the following equations, shall be used to generate the quinary symbols, s_n, as shown in Table 40–8. The quinary symbol sequence shall be presented simultaneously to all transmitters. The transmitter shall time the transmitted symbols from a 125.00 MHz ± 0.01% clock in the MASTER timing mode. A typical transmitter output for transmitter test mode 4 is shown in Figure 40–21.

$$x0_n = Scr_n[0]$$

$$x1_n = Scr_n[1] \wedge Scr_n[4]$$

$$x2_n = Scr_n[2] \wedge Scr_n[4]$$

Table 40–8—Transmitter test mode 4 symbol mapping

x2n	x1n	x0n	quinary symbol, s_n
0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	–1
1	0	0	0
1	0	1	1
1	1	0	–2
1	1	1	–1

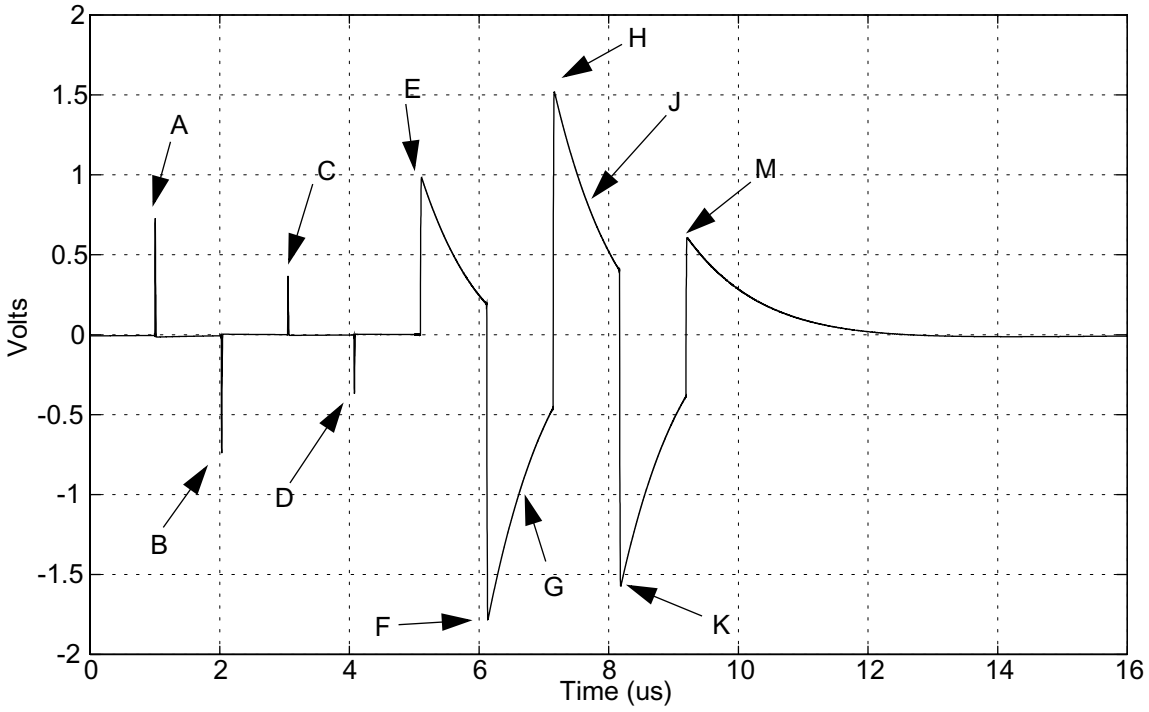


Figure 40–19—Example of transmitter test mode 1 waveform (1 cycle)

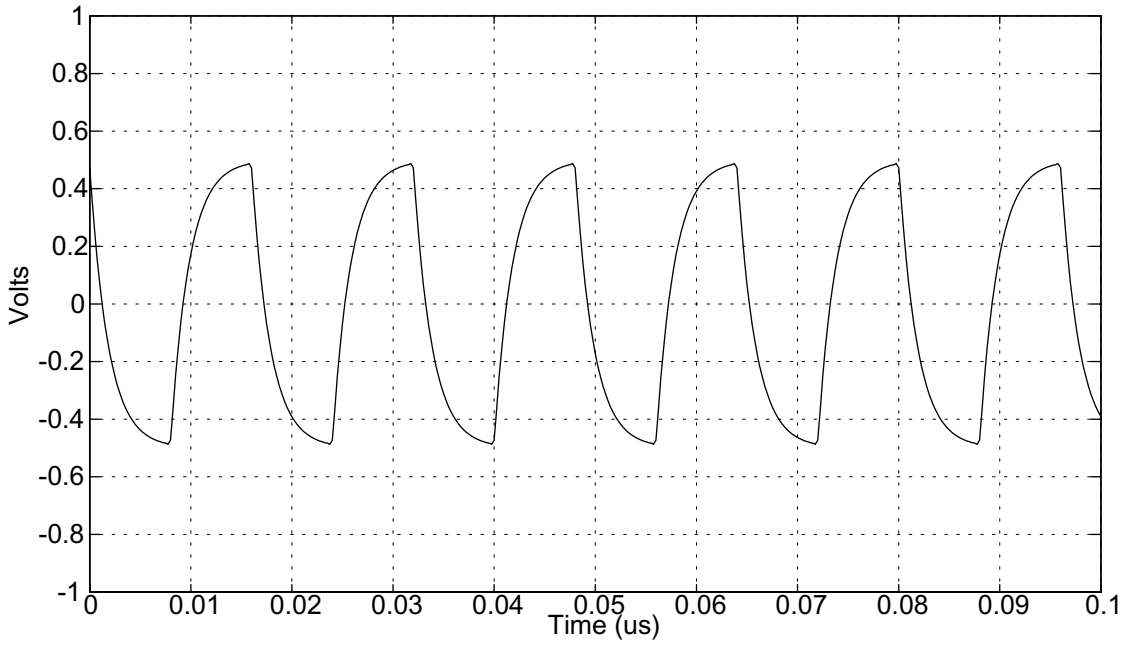


Figure 40–20—Example of transmitter test modes 2 and 3 waveform

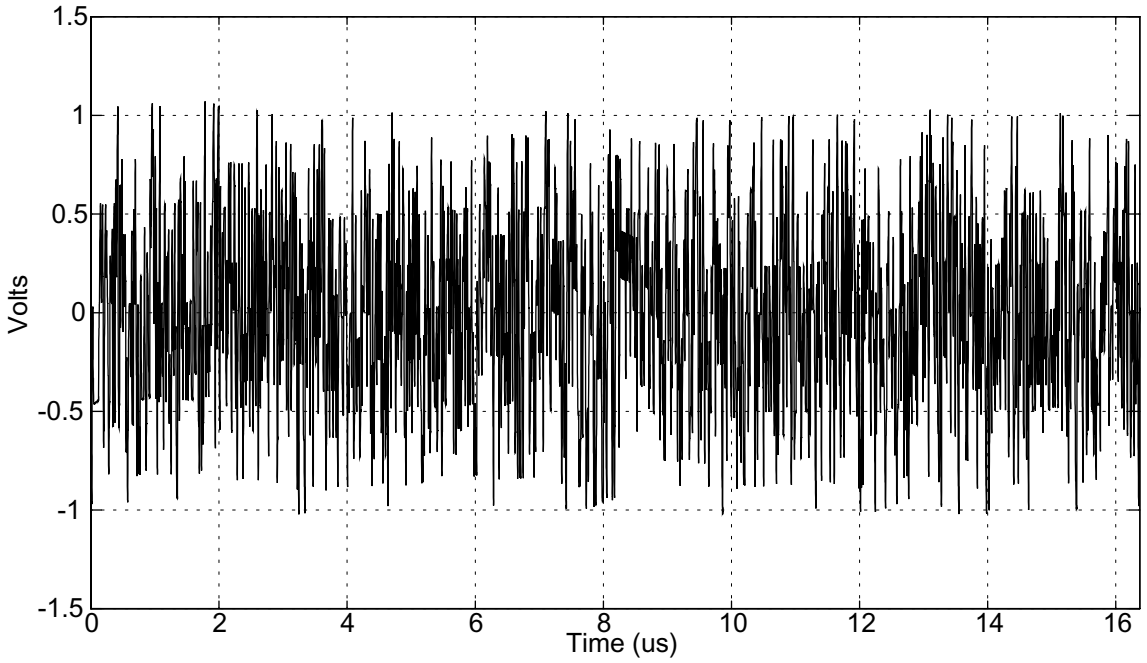


Figure 40–21—Example of Transmitter Test Mode 4 waveform (1 cycle)

40.6.1.1.3 Test Fixtures

The following fixtures (illustrated by Figure 40–22, Figure 40–23, Figure 40–24, and Figure 40–25), or their functional equivalents, shall be used for measuring the transmitter specifications described in 40.6.1.2.

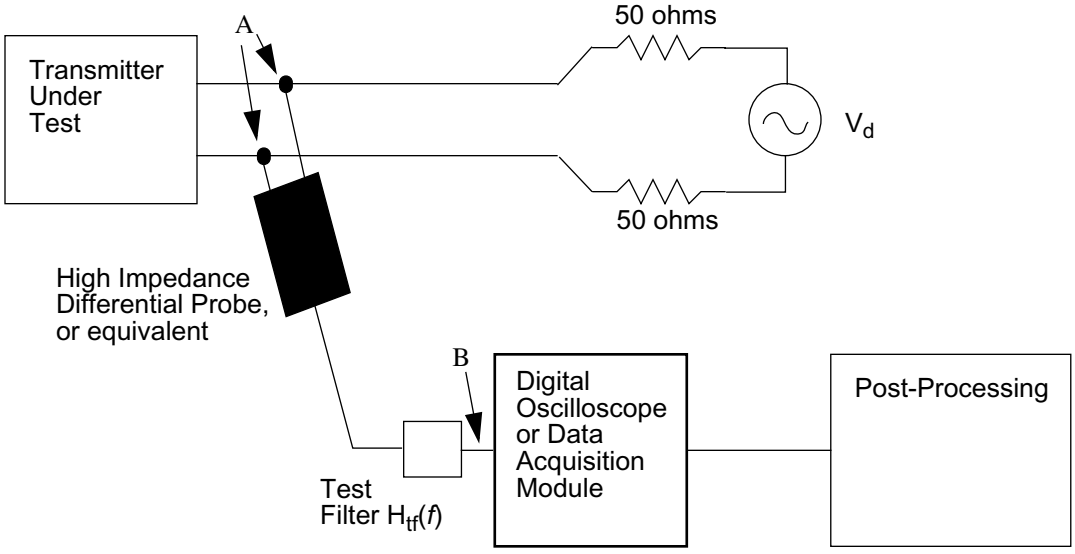


Figure 40–22—Transmitter test fixture 1 for template measurement

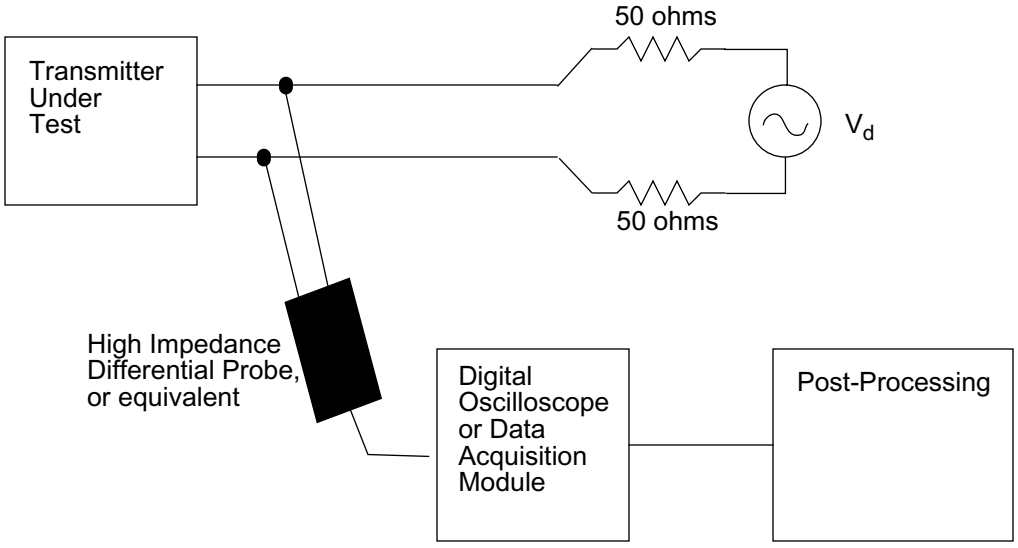


Figure 40–23—Transmitter test fixture 2 for droop measurement

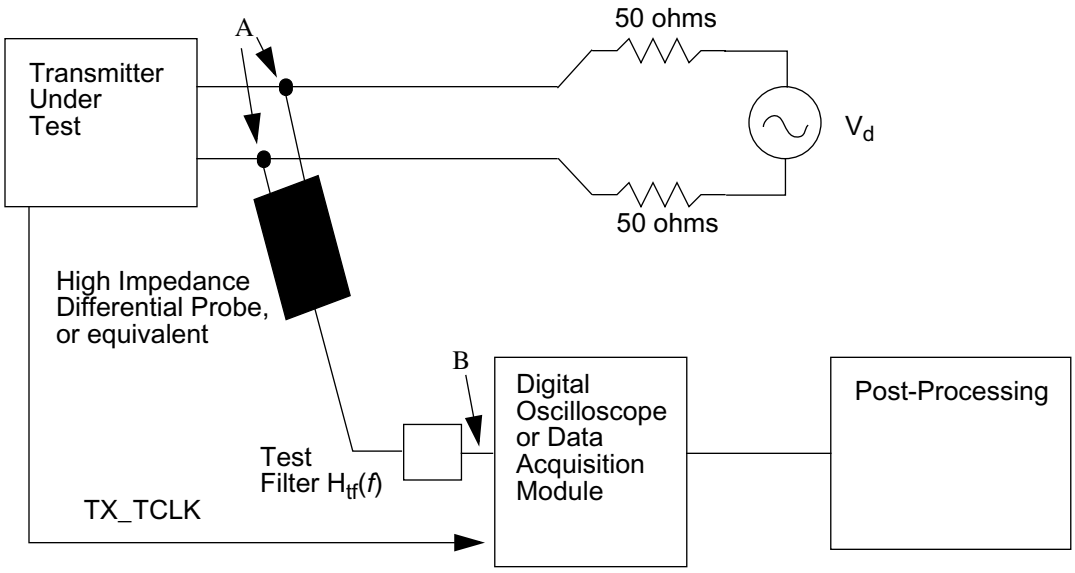


Figure 40–24—Transmitter test fixture 3 for distortion measurement

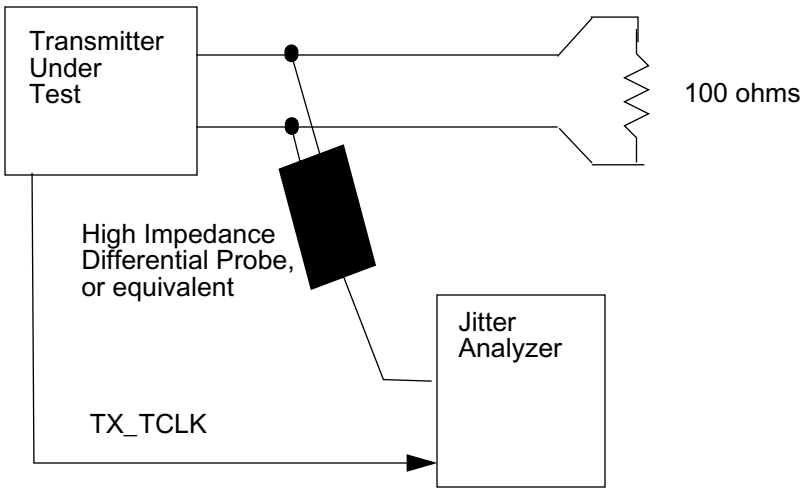


Figure 40–25—Transmitter test fixture 4 for transmitter jitter measurement

The test filter, $H_{tf}(f)$, used in transmitter test fixtures 1 and 3 may be located between the points A and B as long as the test filter does not significantly alter the impedance seen by the transmitter. The test filter may instead be implemented as a digital filter in the post processing block. The test filter shall have the following continuous time transfer function or its discrete time equivalent:

$$H_{tf}(f) = \frac{jf}{jf + 2\times10^{6}} \qquad f \text{ in Hz}$$

NOTE—j denotes the square root of –1.

The disturbing signal, V_d, shall have the characteristics listed in Table 40–9.

Table 40–9—V_d Characteristics

Characteristic	Transmit test fixture 1	Transmit test fixture 2	Transmit test fixture 3
Waveform	Sine wave		
Amplitude	2.8 volts peak-to-peak	2.8 volts peak-to-peak	5.4 volts peak-to-peak
Frequency	31.25 MHz	31.25 MHz	20.833 MHz (125/6 MHz)
Purity	All harmonics >40 dB below fundamental		

The post-processing block has two roles. The first is to remove the disturbing signal from the measurement. A method of removing the disturbing signal is to take a single shot acquisition of the transmitted signal plus test pattern, then remove the best fit of a sine wave at the fundamental frequency of the disturbing signal from the measurement. It will be necessary to allow the fitting algorithm to adjust the frequency, phase, and amplitude parameters of the sine wave to achieve the best fit.

The second role of the post-processing block is to compare the measured data with the templates, droop specification, or distortion specification.

Trigger averaging of the transmitter output to remove measurement noise and increase measurement resolution is acceptable provided it is done in a manner that does not average out possible distortions caused by the interaction of the transmitter and the disturbing voltage. For transmitter template and droop measurements, averaging can be done by ensuring the disturbing signal is exactly synchronous to the test pattern so that the phase of the disturbing signal at any particular point in the test pattern remains constant. Trigger averaging also requires a triggering event that is synchronous to the test pattern. A trigger pulse generated by the PHY would be ideal for this purpose; however, in practice, triggering off the waveform generated by one of the other transmitter outputs that does not have the disturbing signal present may be possible.

NOTE—The disturbing signal may be made synchronous to the test pattern by creating the disturbing signal using a source of the transmit clock for the PHY under test, dividing it down to the proper frequency for the disturbing signal, passing the result through a high Q bandpass filter to eliminate harmonics and then amplifying the result to the proper amplitude.

The generator of the disturbing signal must have sufficient linearity and range so it does not introduce any appreciable distortion when connected to the transmitter output (see Table 40–9). This may be verified by replacing the transmitter under test with another identical disturbing signal generator having a different frequency output and verifying that the resulting waveform's spectrum does not show significant distortion products.

Additionally, to allow for measurement of transmitted jitter in master and slave modes, the PHY shall provide access to the 125 MHz symbol clock, TX_TCLK, that times the transmitted symbols (see 40.4.2.2). The PHY shall provide a means to enable this clock output if it is not normally enabled.

40.6.1.2 Transmitter electrical specifications

The PMA shall provide the Transmit function specified in 40.4.2.2 in accordance with the electrical specifications of this clause.

Where a load is not specified, the transmitter shall meet the requirements of this clause with a 100 Ω resistive differential load connected to each transmitter output.

The tolerance on the poles of the test filters used in this subclause shall be ±1%.

Practical considerations prevent measurement of the local transmitter performance in the presence of the remotely driven signal in this standard; however, the design of the transmitter to tolerate the presence of the remotely driven signal with acceptable distortion or other changes in performance is a critical issue and must be addressed by the implementor. To this end, a disturbing sine wave is used to simulate the presence of a remote transmitter for a number of the transmitter tests described in the following subordinate subclauses.

40.6.1.2.1 Peak differential output voltage and level accuracy

The absolute value of the peak of the waveform at points A and B, as defined in Figure 40–19, shall fall within the range of 0.67 V to 0.82 V (0.75 V ± 0.83 dB). These measurements are to be made for each pair while operating in test mode 1 and observing the differential signal output at the MDI using transmitter test fixture 1 with no intervening cable.

The absolute value of the peak of the waveforms at points A and B shall differ by less than 1%.

The absolute value of the peak of the waveform at points C and D as defined in Figure 40–19 shall differ by less than 2% from 0.5 times the average of the absolute values of the peaks of the waveform at points A and B.

40.6.1.2.2 Maximum output droop

The magnitude of the negative peak value of the waveform at point G, as defined in Figure 40–19, shall be greater than 73.1% of the magnitude of the negative peak value of the waveform at point F. These measurements are to be made for each pair while in test mode 1 and observing the differential signal output at the MDI using transmit test fixture 2 with no intervening cable. Point G is defined as the point exactly 500 ns after point F. Point F is defined as the point where the waveform reaches its minimum value at the location indicated in Figure 40–19. Additionally, the magnitude of the peak value of the waveform at point J as defined in Figure 40–19 shall be greater than 73.1% of the magnitude of the peak value of the waveform at point H. Point J is defined as the point exactly 500 ns after point H. Point H is defined as the point where the waveform reaches its maximum value at the location indicated in Figure 40–19.

40.6.1.2.3 Differential output templates

The voltage waveforms around points A, B, C, D defined in Figure 40–19, after the normalization described herein, shall lie within the time domain template 1 defined in Figure 40–26 and the piecewise linear interpolation between the points in Table 40–10. These measurements are to be made for each pair while in test mode 1 and while observing the differential signal output at the MDI using transmitter test fixture 1 with no intervening cable. The waveforms may be shifted in time as appropriate to fit within the template.

The waveform around point A is normalized by dividing by the peak value of the waveform at A.

The waveform around point B is normalized by dividing by the negative of the peak value of the waveform at A.

The waveform around point C is normalized by dividing by 1/2 the peak value of the waveform at A.

The waveform around point D is normalized by dividing by the negative of 1/2 the peak value of the waveform at A.

The voltage waveforms around points F and H defined in Figure 40–19, after the normalization described herein, shall lie within the time domain template 2 defined in Figure 40–26 and the piecewise linear interpolation between the points in Table 40–11. These measurements are to be made for each pair while in test mode 1 and while observing the differential signal output at the MDI using transmitter test fixture 1 with no intervening cable. The waveforms may be shifted in time as appropriate to fit within the template.

The waveform around point F is normalized by dividing by the peak value of the waveform at F.

The waveform around point H is normalized by dividing by the peak value of the waveform at H.

NOTE—The templates were created with the following assumptions about the elements in the transmit path:

1) Digital Filter: $0.75 + 0.25\ z^{-1}$
2) Ideal DAC
3) Single pole continuous time low pass filter with pole varying from 70.8 MHz to 117 MHz or linear rise/fall time of 5 ns.
4) Single pole continuous time high-pass filter (transformer high pass) with pole varying from 1 Hz to 100 kHz.
5) Single pole continuous time high-pass filter (test filter) with pole varying from 1.8 MHz to 2.2 MHz.
6) Additionally, +0.025 was added to the upper template and –0.025 was added to the lower template to allow for noise and measurement error.

NOTE—The transmit templates are not intended to address electromagnetic radiation limits.

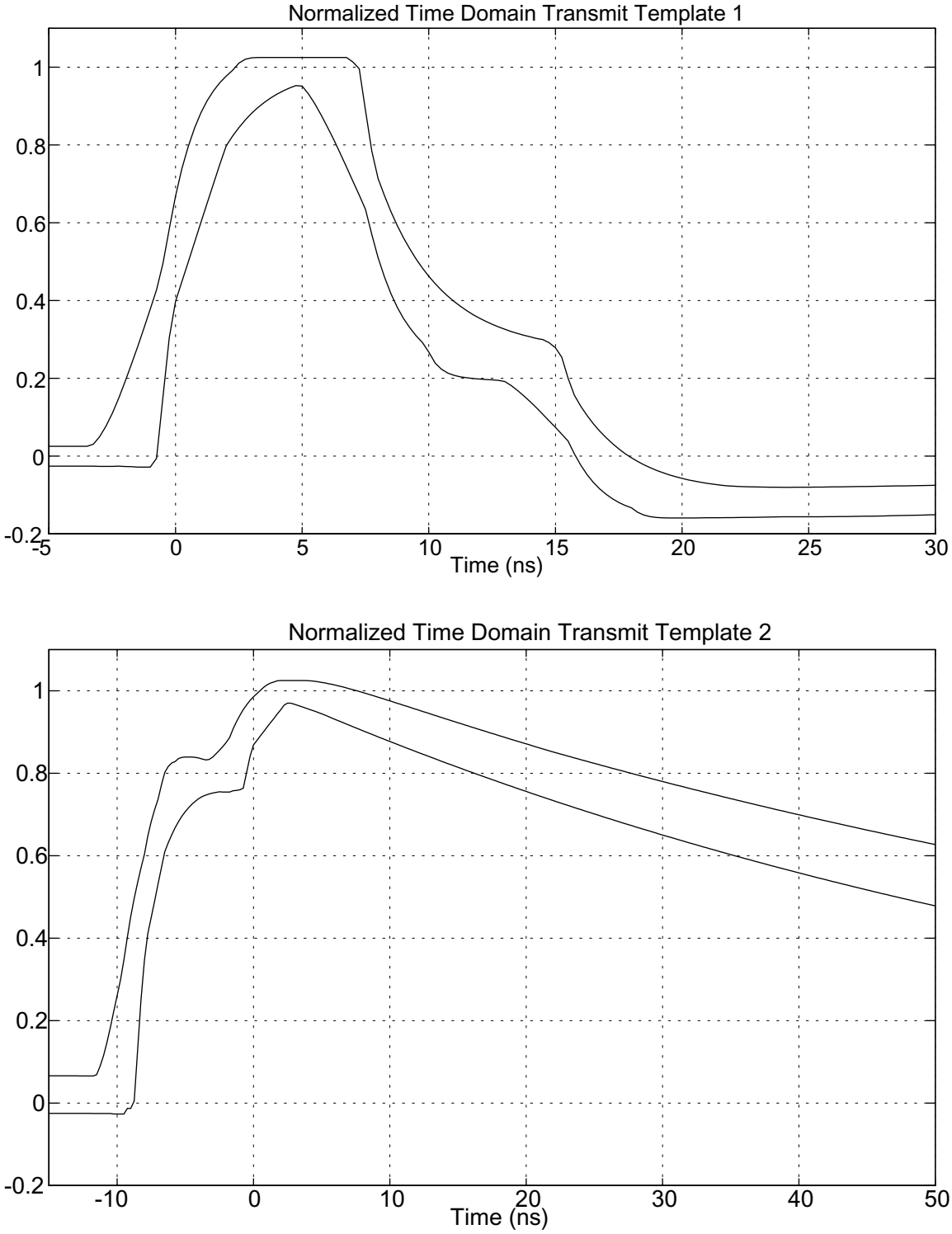


Figure 40–26—Normalized transmit templates as measured at MDI using transmit test fixture 1

NOTE—The ASCII for Tables 40–10 and 40–11 is available from http://www.ieee802.org/3/publication/index.html.[9]

Table 40–10—Normalized time domain voltage template 1

Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit	Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit
–5.00	0.025	–0.026	12.75	0.332	0.195
–4.75	0.025	–0.026	13.00	0.326	0.192
–4.50	0.025	–0.026	13.25	0.320	0.181
–4.25	0.025	–0.026	13.50	0.315	0.169
–4.00	0.025	–0.026	13.75	0.311	0.155
–3.75	0.025	–0.026	14.00	0.307	0.140
–3.50	0.025	–0.026	14.25	0.303	0.124
–3.25	0.031	–0.026	14.50	0.300	0.108
–3.00	0.050	–0.026	14.75	0.292	0.091
–2.75	0.077	–0.026	15.00	0.278	0.074
–2.50	0.110	–0.026	15.25	0.254	0.056
–2.25	0.148	–0.026	15.50	0.200	0.039
–2.00	0.190	–0.027	15.75	0.157	0.006
–1.75	0.235	–0.027	16.00	0.128	–0.023
–1.50	0.281	–0.028	16.25	0.104	–0.048
–1.25	0.329	–0.028	16.50	0.083	–0.068
–1.00	0.378	–0.028	16.75	0.064	–0.084
–0.75	0.427	–0.006	17.00	0.047	–0.098
–0.50	0.496	0.152	17.25	0.032	–0.110
–0.25	0.584	0.304	17.50	0.019	–0.119
0.00	0.669	0.398	17.75	0.007	–0.127
0.25	0.739	0.448	18.00	–0.004	–0.133
0.50	0.796	0.499	18.25	–0.014	–0.145
0.75	0.844	0.550	18.50	–0.022	–0.152
1.00	0.882	0.601	18.75	–0.030	–0.156
1.25	0.914	0.651	19.00	–0.037	–0.158
1.50	0.940	0.701	19.25	–0.043	–0.159

[9]*Copyright release for 802.3® template data:* Users of this standard may freely reproduce the template data in this subclause so it can be used for its intended purpose.

Table 40–10—Normalized time domain voltage template 1 *(continued)*

Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit	Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit
1.75	0.960	0.751	19.50	–0.048	–0.159
2.00	0.977	0.797	19.75	–0.053	–0.159
2.25	0.992	0.822	20.00	–0.057	–0.159
2.50	1.010	0.845	20.25	–0.061	–0.159
2.75	1.020	0.864	20.50	–0.064	–0.159
3.00	1.024	0.881	20.75	–0.067	–0.159
3.25	1.025	0.896	21.00	–0.070	–0.159
3.50	1.025	0.909	21.25	–0.072	–0.159
3.75	1.025	0.921	21.50	–0.074	–0.158
4.00	1.025	0.931	21.75	–0.076	–0.158
4.25	1.025	0.939	22.00	–0.077	–0.158
4.50	1.025	0.946	22.25	–0.078	–0.158
4.75	1.025	0.953	22.50	–0.079	–0.158
5.00	1.025	0.951	22.75	–0.079	–0.157
5.25	1.025	0.931	23.00	–0.079	–0.157
5.50	1.025	0.905	23.25	–0.080	–0.157
5.75	1.025	0.877	23.50	–0.080	–0.157
6.00	1.025	0.846	23.75	–0.080	–0.156
6.25	1.025	0.813	24.00	–0.080	–0.156
6.50	1.025	0.779	24.25	–0.080	–0.156
6.75	1.025	0.743	24.50	–0.080	–0.156
7.00	1.014	0.707	24.75	–0.080	–0.156
7.25	0.996	0.671	25.00	–0.080	–0.156
7.50	0.888	0.634	25.25	–0.080	–0.156
7.75	0.784	0.570	25.50	–0.080	–0.156
8.00	0.714	0.510	25.75	–0.079	–0.156
8.25	0.669	0.460	26.00	–0.079	–0.156
8.50	0.629	0.418	26.25	–0.079	–0.156
8.75	0.593	0.383	26.50	–0.079	–0.155

Table 40–10—Normalized time domain voltage template 1 *(continued)*

Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit	Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit
9.00	0.561	0.354	26.75	–0.079	–0.155
9.25	0.533	0.330	27.00	–0.078	–0.155
9.50	0.507	0.309	27.25	–0.078	–0.155
9.75	0.483	0.292	27.50	–0.078	–0.154
10.00	0.462	0.268	27.75	–0.078	–0.154
10.25	0.443	0.239	28.00	–0.077	–0.154
10.50	0.427	0.223	28.25	–0.077	–0.153
10.75	0.411	0.213	28.50	–0.077	–0.153
11.00	0.398	0.208	28.75	–0.076	–0.153
11.25	0.385	0.204	29.00	–0.076	–0.152
11.50	0.374	0.201	29.25	–0.076	–0.152
11.75	0.364	0.199	29.50	–0.076	–0.152
12.00	0.355	0.198	29.75	–0.075	–0.151
12.25	0.346	0.197	30.00	–0.075	–0.151
12.50	0.339	0.196			

Table 40–11—Normalized time domain voltage template 2

Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit	Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit
–15.00	0.066	–0.025	18.00	0.891	0.779
–14.50	0.066	–0.025	18.50	0.886	0.773
–14.00	0.066	–0.025	19.00	0.881	0.767
–13.50	0.066	–0.025	19.50	0.876	0.762
–13.00	0.066	–0.025	20.00	0.871	0.756
–12.50	0.066	–0.025	20.50	0.866	0.750
–12.00	0.066	–0.025	21.00	0.861	0.745
–11.50	0.069	–0.025	21.50	0.856	0.739
–11.00	0.116	–0.025	22.00	0.852	0.734
–10.50	0.183	–0.025	22.50	0.847	0.728

Table 40–11—Normalized time domain voltage template 2 *(continued)*

Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit	Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit
–10.00	0.261	–0.027	23.00	0.842	0.723
–9.50	0.348	–0.027	23.50	0.838	0.717
–9.00	0.452	–0.013	24.00	0.833	0.712
–8.50	0.535	0.130	24.50	0.828	0.707
–8.00	0.604	0.347	25.00	0.824	0.701
–7.50	0.683	0.451	25.50	0.819	0.696
–7.00	0.737	0.531	26.00	0.815	0.691
–6.50	0.802	0.610	26.50	0.811	0.686
–6.00	0.825	0.651	27.00	0.806	0.680
–5.50	0.836	0.683	27.50	0.802	0.675
–5.00	0.839	0.707	28.00	0.797	0.670
–4.50	0.839	0.725	28.50	0.793	0.665
–4.00	0.837	0.739	29.00	0.789	0.660
–3.50	0.832	0.747	29.50	0.784	0.655
–3.00	0.839	0.752	30.00	0.780	0.650
–2.50	0.856	0.755	30.50	0.776	0.645
–2.00	0.875	0.755	31.00	0.772	0.641
–1.50	0.907	0.758	31.50	0.767	0.636
–1.00	0.941	0.760	32.00	0.763	0.631
–0.50	0.966	0.803	32.50	0.759	0.626
0.00	0.986	0.869	33.00	0.755	0.621
0.50	1.001	0.890	33.50	0.751	0.617
1.00	1.014	0.912	34.00	0.747	0.612
1.50	1.022	0.933	34.50	0.743	0.607
2.00	1.025	0.954	35.00	0.739	0.603
2.50	1.025	0.970	35.50	0.734	0.598
3.00	1.025	0.967	36.00	0.730	0.594
3.50	1.025	0.962	36.50	0.727	0.589
4.00	1.025	0.956	37.00	0.723	0.585

Table 40–11—Normalized time domain voltage template 2 *(continued)*

Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit	Time, ns	Normalized transmit time domain template, upper limit	Normalized transmit time domain template, lower limit
4.50	1.023	0.950	37.50	0.719	0.580
5.00	1.020	0.944	38.00	0.715	0.576
5.50	1.017	0.937	38.50	0.711	0.571
6.00	1.014	0.931	39.00	0.707	0.567
6.50	1.010	0.924	39.50	0.703	0.563
7.00	1.005	0.917	40.00	0.699	0.558
7.50	1.001	0.910	40.50	0.695	0.554
8.00	0.996	0.903	41.00	0.692	0.550
8.50	0.991	0.897	41.50	0.688	0.546
9.00	0.986	0.890	42.00	0.684	0.541
9.50	0.981	0.884	42.50	0.680	0.537
10.00	0.976	0.877	43.00	0.677	0.533
10.50	0.970	0.871	43.50	0.673	0.529
11.00	0.965	0.864	44.00	0.669	0.525
11.50	0.960	0.858	44.50	0.666	0.521
12.00	0.954	0.852	45.00	0.662	0.517
12.50	0.949	0.845	45.50	0.659	0.513
13.00	0.944	0.839	46.00	0.655	0.509
13.50	0.938	0.833	46.50	0.651	0.505
14.00	0.933	0.827	47.00	0.648	0.501
14.50	0.928	0.820	47.50	0.644	0.497
15.00	0.923	0.814	48.00	0.641	0.493
15.50	0.917	0.808	48.50	0.637	0.490
16.00	0.912	0.802	49.00	0.634	0.486
16.50	0.907	0.796	49.50	0.631	0.482
17.00	0.902	0.791	50.00	0.627	0.478
17.50	0.897	0.785			

40.6.1.2.4 Transmitter distortion

When in test mode 4 and observing the differential signal output at the MDI using transmitter test fixture 3, for each pair, with no intervening cable, the peak distortion as defined below shall be less than 10 mV.

The peak distortion is determined by sampling the differential signal output with the symbol rate TX_TCLK at an arbitrary phase and processing a block of any 2047 consecutive samples with the MATLAB (see 1.3) code listed below or equivalent. Note that this code assumes that the differential signal has already been filtered by the test filter.

NOTE—The ASCII for the following MATLAB code is available from http://www.ieee802.org/3/publication/index.html.[10]

MATLAB code for Distortion Post Processing is as follows:

```
%
% Distortion Specification Post Processing
%

% Initialize Variables
clear
symbolRate=125e6;                                % symbol rate
dataFile=input('Data file name: ','s')

% Generate test pattern symbol sequence

scramblerSequence=ones(1,2047);
for i=12:2047
  scramblerSequence(i)=mod(scramblerSequence(i-11) + scramblerSequence(i-9),2);
end

for i=1:2047
  temp=scramblerSequence(mod(i-1,2047)+1) + ...
       2*mod(scramblerSequence(mod(i-2,2047)+1) + scramblerSequence(mod(i-5,2047)+1),2)
+ ...
      4*mod(scramblerSequence(mod(i-3,2047)+1) + scramblerSequence(mod(i-5,2047)+1),2);
  switch temp
    case 0,
      testPattern(i)=0;
    case 1,
      testPattern(i)=1;
    case 2,
      testPattern(i)=2;
    case 3,
      testPattern(i)=-1;
    case 4,
      testPattern(i)=0;
    case 5,
      testPattern(i)=1;
    case 6,
      testPattern(i)=-2;
    case 7,
      testPattern(i)=-1;
  end
end



% Input data file
fid=fopen(dataFile,'r');
sampledData=fscanf(fid,'%f');
```

[10]*Copyright release for MATLAB code*: Users of this standard may freely reproduce the MATLAB code in this subclause so it can be used for its intended purpose.


```
fclose(fid);
sampledData=sampledData.';


if (length(sampledData) < 2047)
  error('Must have 2047 consecutive samples for processing');
elseif (length(sampledData) > 2047)
  fprintf(1,'\n Warning - only using first 2047 samples in data file');
  sampledData=sampledData(1:2047);
end


% Fit a sine wave to the data and temporarily remove it to yield processed data

options=foptions;
options(1)=0;
options(2)=1e-8;
options(3)=1e-8;
options(14)=2000;
gradfun=zeros(0);
P=fmins('sinefit',[2.0 0 125/6.],options,gradfun,sampledData,symbolRate);

P

processedData=sampledData - ...
    P(1)*sin(2*pi*(P(3)*1e6*[0:2046]/symbolRate + P(2)*1e-9*symbolRate));



% LMS Canceller

numberCoeff=70;  % Number of coefficients in canceller
coefficients=zeros(1,numberCoeff);
delayLine=testPattern;

% Align data in delayLine to sampled data pattern
temp=xcorr(processedData,delayLine);
index=find(abs(temp)==max(abs(temp)));
index=mod(mod(length(processedData) - index(1),2047)+numberCoeff-10,2047);
delayLine=[delayLine((end-index):end) delayLine(1:(end-index-1))];


% Compute coefficients that minimize squared error in cyclic block

for i=1:2047
  X(i,:)=delayLine(mod([0:(numberCoeff-1)]+i-1,2047)+1);
end
coefficients=(inv(X.' * X)*(processedData*X).').';



% Canceller
for i=1:2047
   err(i)=processedData(i) - sum(delayLine(1+mod((i-1):(i+numberCoeff-2),2047)).*coef-
ficients);
end

% Add back temporarily removed sine wave

err=err+P(1)*sin(2*pi*(P(3)*1e6*[0:2046]./symbolRate + P(2)*1e-9*symbolRate));


% Re-fit sine wave and do a final removal
```

```
options=foptions;
options(1)=0;
options(2)=1e-12;
options(3)=1e-12;
options(14)=10000;
gradfun=zeros(0);
P=fmins('sinefit',[2.0 0 125/6.],options,gradfun,err,symbolRate);

P

processedData=sampledData - ...
    P(1)*sin(2*pi*(P(3)*1e6*[0:2046]/symbolRate + P(2)*1e-9*symbolRate));


% Compute coefficients that minimize squared error in cyclic block

coefficients=(inv(X.' * X)*(processedData*X).').';



% Canceller
for i=1:2047
   err(i)=processedData(i) - sum(delayLine(1+mod((i-1):(i+numberCoeff-2),2047)).*coef-
ficients);
end

% SNR Calculation
signal=0.5;
noise=mean(err.^2);

SNR=10*log10(signal./noise);

% Output Peak Distortion

peakDistortion=max(abs(err))


% Function for fitting sine wave
function err=sinefit(parameters,data,symbolRate)
err=sum((data- ...
    parameters(1)*sin(2*pi*(parameters(3)*1e6*[0:(length(data)-1)]/symbolRate + param-
eters(2)*1e-9*symbolRate))).^2);
```

40.6.1.2.5 Transmitter timing jitter

When in test mode 2 or test mode 3, the peak-to-peak jitter J_{txout} of the zero crossings of the differential signal output at the MDI relative to the corresponding edge of TX_TCLK is measured. The corresponding edge of TX_TCLK is the edge of the transmit test clock, in polarity and time, that generates the zero-crossing transition being measured.

When in the normal mode of operation as the MASTER, the peak-to-peak value of the MASTER TX_TCLK jitter relative to an unjittered reference shall be less than 1.4 ns. When the jitter waveform on TX_TCLK is filtered by a high-pass filter, $H_{jf1}(f)$, having the transfer function below, the peak-to-peak value of the resulting filtered timing jitter plus J_{txout} shall be less than 0.3 ns.

$$H_{jf1}(f) = \frac{\mathrm{j}f}{\mathrm{j}f + 5000} \qquad f \text{ in Hz}$$

When in the normal mode of operation as the SLAVE, receiving valid signals from a compliant PHY operating as the MASTER using the test channel defined in 40.6.1.1.1, with test channel port A connected to the SLAVE, the peak-to-peak value of the SLAVE TX_TCLK jitter relative to the MASTER TX_TCLK shall be less than 1.4 ns after the receiver is properly receiving the data and has set bit 10.13 of the GMII management register set to 1. When the jitter waveform on TX_TCLK is filtered by a high-pass filter, $H_{jf2}(f)$, having

the transfer function below, the peak-to-peak value of the resulting filtered timing jitter plus J_{txout} shall be no more than 0.4 ns greater than the simultaneously measured peak-to-peak value of the MASTER jitter filtered by $H_{jf1}(f)$.

$$H_{jf2}(f) = \frac{jf}{jf + 32000} \qquad f \text{ in Hz}$$

NOTE—j denotes the square root of –1.

For all high-pass filtered jitter measurements, the peak-to-peak value shall be measured over an unbiased sample of at least 10^5 clock edges. For all unfiltered jitter measurements, the peak-to-peak value shall be measured over an interval of not less than 100 ms and not more than 1 second.

40.6.1.2.6 Transmit clock frequency

The quinary symbol transmission rate on each pair of the master PHY shall be 125.00 MHz ± 0.01%.

40.6.1.3 Receiver electrical specifications

The PMA shall provide the Receive function specified in 40.4.2.3 in accordance with the electrical specifications of this clause. The patch cabling and interconnecting hardware used in test configurations shall be within the limits specified in 40.7.

40.6.1.3.1 Receiver differential input signals

Differential signals received at the MDI that were transmitted from a remote transmitter within the specifications of 40.6.1.2 and have passed through a link specified in 40.7 are translated into one of the PMA_UNITDATA.indicate messages with a 4-D symbol error rate less than 10^{-10} and sent to the PCS after link reset completion. Since the 4-D symbols are not accessible, this specification shall be satisfied by a frame error rate less than 10^{-7} for 125 octet frames.

40.6.1.3.2 Receiver frequency tolerance

The receive feature shall properly receive incoming data with a 5-level symbol rate within the range 125.00 MHz ± 0.01%.

40.6.1.3.3 Common-mode noise rejection

This specification is provided to limit the sensitivity of the PMA receiver to common-mode noise from the cabling system. Common-mode noise generally results when the cabling system is subjected to electromagnetic fields. Figure 40–27 shows the test configuration, which uses a capacitive cable clamp, that injects common-mode signals into a cabling system.

A 100-meter, 4-pair Category 5 cable that meets the specification of 40.7 is connected between two 1000BASE-T PHYs and inserted into the cable clamp. The cable should be terminated on each end with an MDI connector plug specified in 40.8.1. The clamp should be located a distance of ~20 cm from the receiver. It is recommended that the cable between the transmitter and the cable clamp be installed either in a linear run or wrapped randomly on a cable rack. The cable rack should be at least 3 m from the cable clamp. In addition, the cable clamp and 1000BASE-T receiver should be placed on a common copper ground plane and the ground of the receiver should be in contact with the ground plane. The chassis grounds of all test equipment used should be connected to the copper ground plane. No connection is required between the copper ground plane and an external reference. A description of the cable clamp, as well as the validation procedure, can be found in Annex 40B.

A signal generator with a 50 Ω impedance is connected to one end of the clamp and an oscilloscope with a 50 Ω input is connected to the other end of the clamp. The signal generator shall be capable of providing a sine wave signal of 1 MHz to 250 MHz. The output of the signal generator is adjusted for a voltage of 1.0 Vrms (1.414 Vpeak) on the oscilloscope.

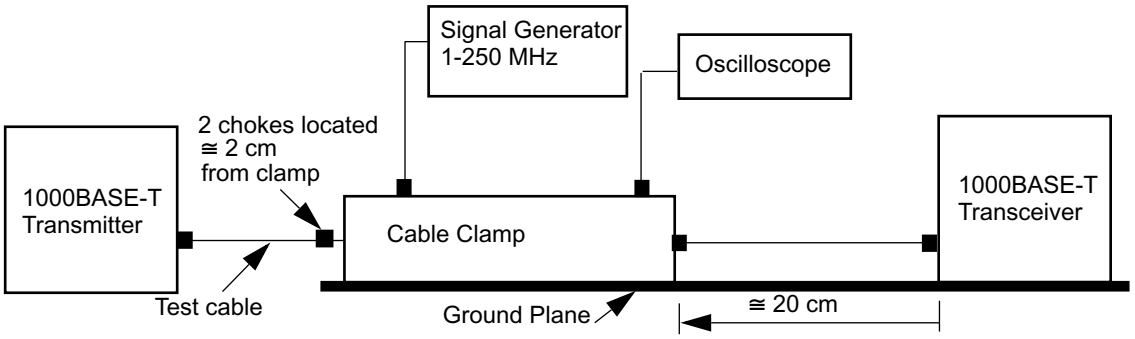


Figure 40–27—Receiver common-mode noise rejection test

While sending data from the transmitter, the receiver shall send the proper PMA_UNITDATA.indicate messages to the PCS as the signal generator frequency is varied from 1 MHz to 250 MHz.

NOTE—Although the signal specification is constrained within the 1–100 MHz band, this test is performed up to 250 MHz to ensure the receiver under test can tolerate out-of-band (100–250 MHz) noise.

40.6.1.3.4 Alien Crosstalk noise rejection

While receiving data from a transmitter specified in 40.6.1.2 through a link segment specified in 40.7 connected to all MDI duplex channels, a receiver shall send the proper PMA_UNITDATA.indicate message to the PCS when any one of the four pairs is connected to a noise source as described in Figure 40–28. Because symbol encoding is employed, this specification shall be satisfied by a frame error rate of less than 10^{-7} for 125 octet frames. The level of the noise signal at the MDI is nominally 25 mV peak-to-peak. (Measurements are to be made on each of the four pairs.) The noise source shall be connected to one of the MDI inputs using Category 5 balanced cable of a maximum length of 0.5 m.

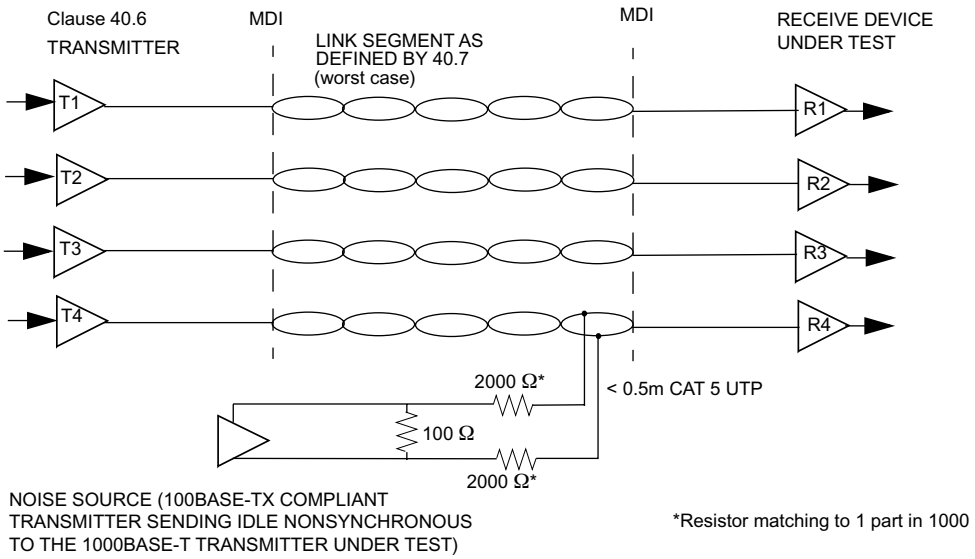


Figure 40–28—Differential mode noise rejection test

40.7 Link segment characteristics

1000BASE-T is designed to operate over a 4-pair Category 5 balanced cabling system. Each of the four pairs supports an effective data rate of 250 Mbps in each direction simultaneously. The term "link segment" used in this clause refers to four duplex channels. The term "duplex channel" will be used to refer to a single channel with full duplex capability. Specifications for a link segment apply equally to each of the four duplex channels. All implementations of the balanced cabling link shall be compatible at the MDI.

40.7.1 Cabling system characteristics

The cabling system used to support 1000BASE-T requires 4 pairs of Category 5 balanced cabling with a nominal impedance of 100 Ω. The cabling system components (cables, cords, and connectors) used to provide the link segment shall consist of Category 5 components as specified in ANSI/TIA/EIA-568-A:1995 and ISO/IEC 11801:1995. Additionally:

a) 1000BASE-T uses a star topology with Category 5 balanced cabling used to connect PHY entities.
b) 1000BASE-T is an ISO/IEC 11801 Class D application, with additional installation requirements and transmission parameters specified in Annex 40A.
c) The width of the PMD transmit signal spectrum is approximately 80 MHz.
d) The use of shielding is outside the scope of this standard.

40.7.2 Link transmission parameters

The transmission parameters contained in this subclause are specified to ensure that a Category 5 link segment of up to at least 100 m will provide a reliable medium. The transmission parameters of the link segment include insertion loss, delay parameters, characteristic impedance, NEXT loss, ELFEXT loss, and return loss.

Link segment testing shall be conducted using source and load impedances of 100 Ω. The tolerance on the poles of the test filter used in this subclause shall be no worse than 1%.

40.7.2.1 Insertion loss

The insertion loss of each duplex channel shall be less than

$$\text{Insertion\_Loss}(f) < 2.1\ f^{0.529} + 0.4/f \quad \text{(dB)}$$

at all frequencies from 1 MHz to 100 MHz. This includes the attenuation of the balanced cabling pairs, including work area and equipment cables plus connector losses within each duplex channel. The insertion loss specification shall be met when the duplex channel is terminated in 100 Ω.

NOTE—The above equation approximates the insertion loss specification at discrete frequencies for Category 5 100-meter links specified in ANSI/TIA/EIA-568-A Annex E and in TIA/EIA TSB 67.

40.7.2.2 Differential characteristic impedance

The nominal differential characteristic impedance of each link segment duplex channel, which includes cable cords and connecting hardware, is 100 Ω for all frequencies between 1 MHz and 100 MHz.

40.7.2.3 Return loss

Each link segment duplex channel shall meet or exceed the return loss specified in the following equation at all frequencies from 1 MHz to 100 MHz

$$\text{Return\_Loss}(f) \begin{cases} 15 & (1 - 20\ \text{MHz}) \\ 15 - 10\log_{10}(f/20) & (20 - 100\ \text{MHz}) \end{cases} \text{(dB)}$$

where f is the frequency in MHz. The reference impedance shall be 100 Ω.

40.7.3 Coupling parameters

In order to limit the noise coupled into a duplex channel from adjacent duplex channels, Near-End Crosstalk (NEXT) loss and Equal Level Far-End Crosstalk (ELFEXT) loss are specified for each link segment. Each duplex channel can be disturbed by more than one duplex channel. Requirements for Multiple Disturber Near-End Crosstalk (MDNEXT) are satisfied even when worst case conditions of differential pair-to-pair NEXT as specified under 40.7.3.1.1 occur. Therefore, there are no separate requirements for MDNEXT. Requirements for Multiple Disturber Equal-Level Far-End Crosstalk (MDELFEXT) loss are specified in 40.7.3.2.2.

40.7.3.1 Near-End Crosstalk (NEXT)

40.7.3.1.1 Differential Near-End Crosstalk

In order to limit the crosstalk at the near end of a link segment, the differential pair-to-pair Near-End Crosstalk (NEXT) loss between a duplex channel and the other three duplex channels is specified to meet the symbol error rate objective specified in 40.1. The NEXT loss between any two duplex channels of a link segment shall be at least

$$27.1 - 16.8\log_{10}(f/100)$$

where f is the frequency over the range of 1 MHz to 100 MHz.

NOTE—The above equation approximates the NEXT loss specification at discrete frequencies for Category 5 100-meter links specified in ANSI/TIA/EIA-568-A Annex E and in TSB-67.

40.7.3.2 Far-End Crosstalk (FEXT)

40.7.3.2.1 Equal Level Far-End Crosstalk (ELFEXT) loss

Equal Level Far-End Crosstalk (ELFEXT) loss is specified in order to limit the crosstalk at the far end of each link segment duplex channel and meet the BER objective specified in 40.6.1.3.1. Far-End Crosstalk (FEXT) is crosstalk that appears at the far end of a duplex channel (disturbed channel), which is coupled from another duplex channel (disturbing channel) with the noise source (transmitters) at the near end. FEXT loss is defined as

$$\text{FEXT\_Loss}(f) = 20\log_{10}[V\text{pds}(f)/V\text{pcn}(f)]$$

and ELFEXT_Loss is defined as

$$\text{ELFEXT\_Loss}(f) = 20\log_{10}[V\text{pds}(f)/V\text{pcn}(f)] - \text{SLS\_Loss}(f)$$

where

*V*pds	is the peak voltage of disturbing signal (near-end transmitter)
*V*pcn	is the peak crosstalk noise at far end of disturbed channel
SLS_Loss	is the insertion loss of disturbed channel in dB

The worst pair ELFEXT loss between any two duplex channels shall be greater than $17 - 20\log_{10}(f/100)$ dB where f is the frequency over the range of 1 MHz to 100 MHz.

40.7.3.2.2 Multiple Disturber Equal Level Far-End Crosstalk (MDELFEXT) loss

Since four duplex channels are used to transfer data between PMDs, the FEXT that is coupled into a data carrying channel will be from the three adjacent disturbing duplex channels. This specification is consistent with three channel-to-channel disturbers—one with a ELFEXT loss of at least $17 - 20\log_{10}(f/100)$ dB, one with a ELFEXT loss of at least $19.5 - 20\log_{10}(f/100)$ dB, and one with a ELFEXT loss of at least $23 - 20\log_{10}(f/100)$ dB. To ensure the total FEXT coupled into a duplex channel is limited, multiple disturber ELFEXT loss is specified as the power sum of the individual ELFEXT losses.

The Power Sum loss between a duplex channel and the three adjacent disturbers shall be

$$\text{PSELFEXT loss} > 14.4 - 20\log_{10}(f/100)\ \text{dB}$$

where f is the frequency over the range of 1 MHz to 100 MHz.

40.7.3.2.3 Multiple-Disturber Power Sum Equal Level Far-End Crosstalk (PSELFEXT) loss

PSELFEXT loss is determined by summing the magnitude of the three individual pair-to-pair differential ELFEXT loss values over the frequency range 1 to 100 MHz as follows:

$$\text{PSELFEXT\_Loss(f)} = -10\log_{10}\sum_{i=1}^{i=3} 10^{-(NL(f)i)/10}$$

where

NL(f)i is the magnitude of ELFEXT loss at frequency f of pair combination i
i is the 1, 2, or 3 (pair-to-pair combination)

40.7.4 Delay

In order to simultaneously send data over four duplex channels in parallel, the propagation delay of each duplex channel as well as the difference in delay between any two of the four channels are specified. This ensures the 1000 Mbps data that is divided across four channels can be properly reassembled at the far-end receiver. This also ensures the round-trip delay requirement for effective collision detection is met.

40.7.4.1 Maximum link delay

The propagation delay of a link segment shall not exceed 570 ns at all frequencies between 2 MHz and 100 MHz.

40.7.4.2 Link delay skew

The difference in propagation delay, or skew, between all duplex channel pair combinations of a link segment, under all conditions, shall not exceed 50 ns at all frequencies from 2 MHz to 100 MHz. It is a further functional requirement that, once installed, the skew between any two of the four duplex channels due to environmental conditions shall not vary more than 10 ns within the above requirement.

40.7.5 Noise environment

The 1000BASE-T noise environment consists of noise from many sources. The primary noise sources that impact the objective BER are NEXT and echo interference, which are reduced to a small residual noise using cancelers. The remaining noise sources, which are secondary sources, are discussed in the following list.

The 1000BASE-T noise environment consists of the following:

a) Echo from the local transmitter on the same duplex channel (cable pair). Echo is caused by the hybrid function used to achieve simultaneous bi-directional transmission of data and by impedance discontinuities in the link segment. It is impractical to achieve the objective BER without using echo cancellation. Since the symbols transmitted by the local disturbing transmitter are available to the cancellation processor, echo interference can be reduced to a small residual noise using echo cancellation methods.

b) Near-End Crosstalk (NEXT) interference from the local transmitters on the duplex channels (cable pairs) of the link segment. Each receiver will experience NEXT interference from three adjacent transmitters. NEXT cancelers are used to reduce the interference from each of the three disturbing transmitters to a small residual noise. NEXT cancellation is possible since the symbols transmitted by the three disturbing local transmitters are available to the cancellation processor. NEXT cancelers can reduce NEXT interference by at least 20 dB.

c) Far-End Crosstalk (FEXT) noise at a receiver is from three disturbing transmitters at the far end of the duplex channel (cable pairs) of the link segment. FEXT noise can be cancelled in the same way as echo and NEXT interference although the symbols from the remote transmitters are not immediately available. However, FEXT noise is much smaller than NEXT interference and can generally be tolerated.

d) Inter-Symbol Interference (ISI) noise. ISI is the extraneous energy from one signaling symbol that interferes with the reception of another symbol on the same channel.

e) Noise from non-idealities in the duplex channel, transmitters, and receivers; for example, DAC/ADC non-linearity, electrical noise (shot and thermal), and non-linear channel characteristics.

f) Noise from sources outside the cabling that couple into the link segment via electric and magnetic fields.

g) Noise from signals in adjacent cables. This noise is referred to as alien NEXT noise and is generally present when cables are bound tightly together. Since the transmitted symbols from the alien NEXT noise source are not available to the cancellation processor (they are in another cable), it is not possible to cancel the alien NEXT noise. To ensure robust operation the alien NEXT noise must meet the specification of 40.7.5.1.

40.7.6 External coupled noise

The noise coupled from external sources that is measured at the output of a filter connected to the output of the near end of a disturbed duplex channel should not exceed 40 mV peak-to-peak. The filter for this measurement is a fifth order Butterworth filter with a 3 dB cutoff at 100MHz.

40.8 MDI specification

This subclause defines the MDI. The link topology requires a crossover function in a DTE-to-DTE connection. See 40.4.4 for a description of the automatic MDI/MDI-X configuration.

40.8.1 MDI connectors

Eight-pin connectors meeting the requirements of subclause 3 and Figures 1 through 4 of IEC 60603-7: 1990 shall be used as the mechanical interface to the balanced cabling. The plug connector shall be used on the balanced cabling and the jack on the PHY. These connectors are depicted (for informational use only) in Figure 40–29 and Figure 40–30. The assignment of PMA signals to connector contacts for PHYs is shown in Table 40-12.

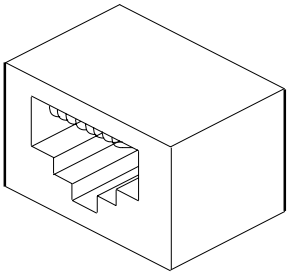


Figure 40–29—MDI connector

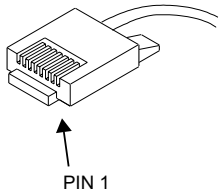


Figure 40–30—Balanced cabling connector

Table 40–12—Assignment of PMA signal to MDI and MDI-X pin-outs

Contact	MDI	MDI-X
1	BI_DA+	BI_DB+
2	BI_DA–	BI_DB–
3	BI_DB+	BI_DA+
4	BI_DC+	BI_DD+
5	BI_DC–	BI_DD–
6	BI_DB–	BI_DA–
7	BI_DD+	BI_DC+
8	BI_DD–	BI_DC–

40.8.2 Crossover function

Although the automatic MDI/MDI-X configuration (see 40.4.4) is not required for successful operation of 1000BASE-T, it is a functional requirement that a crossover function be implemented in every link segment to support the operation of Auto-Negotiation. The crossover function connects the transmitters of one PHY to the receivers of the PHY at the other end of the link segment. Crossover functions may be implemented internally to a PHY or else-where in the link segment. For a PHY that does not implement the crossover function, the MDI labels in the middle column of Table 40–12 refer to its own internal circuits. For PHYs that do implement the internal crossover, the MDI labels in the last column of Table 40–12 refer to the internal circuits of the remote PHY of the link segment. Additionally, the MDI connector for a PHY that implements the crossover function shall be marked with the graphical symbol X. The crossover function specified here is not compatible with the crossover function specified in 14.5.2 for pairs TD and RD.

When a link segment connects a single-port device to a multiport device, it is recommended that the crossover be implemented in the PHY local to the multiport device. If neither or both PHYs of a link segment contain internal crossover functions, an additional external crossover is necessary. It is recommended that the crossover be visible to an installer from one of the PHYs. When both PHYs contain internal crossovers,

it is further recommended that, in networks in which the topology identifies either a central backbone segment or a central device, the PHY furthest from the central element be assigned the external crossover to maintain consistency.

Implicit implementation of the crossover function within a twisted-pair cable or at a wiring panel, while not expressly forbidden, is beyond the scope of this standard.

40.8.3 MDI electrical specifications

The MDI connector (jack) when mated with a specified balanced cabling connector (plug) shall meet the electrical requirements for Category 5 connecting hardware for use with 100-ohm Category 5 cable as specified in ANSI/TIA/EIA-568-A:1995 and ISO/IEC 11801:1995.

The mated MDI/balanced cabling connector pair shall have a FEXT loss not less than $40 - 20\log_{10}(f/100)$ (where f is the frequency over the range 1 MHz to 100 MHz) between all contact pair combinations shown in Table 40–12.

No spurious signals shall be emitted onto the MDI when the PHY is held in power-down mode (as defined in 22.2.4.1.5) independent of the value of TX_EN, when released from power-down mode, or when external power is first applied to the PHY.

40.8.3.1 MDI return loss

The differential impedance at the MDI for each transmit/receive channel shall be such that any reflection due to differential signals incident upon the MDI from a balanced cabling having an impedance of 100 Ω ± 15% is attenuated, relative to the incident signal, at least 16 dB over the frequency range of 1.0 MHz to 40 MHz and at least $10 - 20\log_{10}(f/80)$ dB over the frequency range 40 MHz to 100 MHz (f in MHz). This return loss shall be maintained at all times when the PHY is transmitting data or control symbols.

40.8.3.2 MDI impedance balance

Impedance balance is a measurement of the impedance-to-ground difference between the two MDI contacts used by a duplex link channel and is referred to as common-mode-to-differential-mode impedance balance. Over the frequency range 1.0 MHz to 100.0 MHz, the common-mode-to-differential-mode impedance balance of each channel of the MDI shall exceed

$$34 - 19.2\log_{10}\left(\frac{f}{50}\right) \quad \text{dB}$$

where f is the frequency in MHz when the transmitter is transmitting random or pseudo random data. Test-mode 4 may be used to generate an appropriate transmitter output.

The balance is defined as

$$20\log_{10}\left(\frac{E_{cm}}{E_{dif}}\right)$$

where E_{cm} is an externally applied sine wave voltage as shown in Figure 40–31 and E_{dif} is the resulting waveform due only to the applied sine wave and not the transmitted data.

NOTES

1—Triggered averaging can be used to separate the component due to the applied common-mode sine wave from the transmitted data component.

2—The imbalance of the test equipment (such as the matching of the test resistors) must be insignificant relative to the balance requirements.

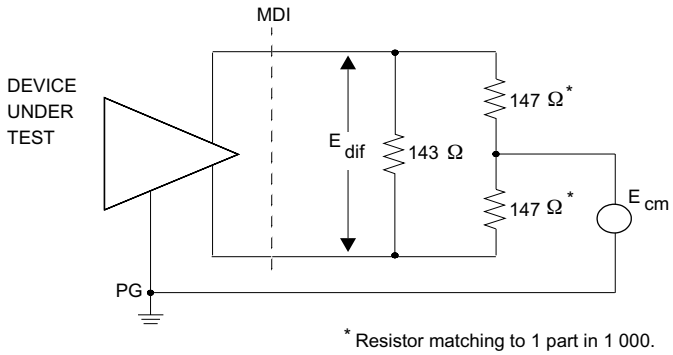


Figure 40–31—MDI impedance balance test circuit

40.8.3.3 MDI common-mode output voltage

The magnitude of the total common-mode output voltage, E_{cm_out}, on any transmit circuit, when measured as shown in Figure 40–32, shall be less than 50 mV peak-to-peak when transmitting data.

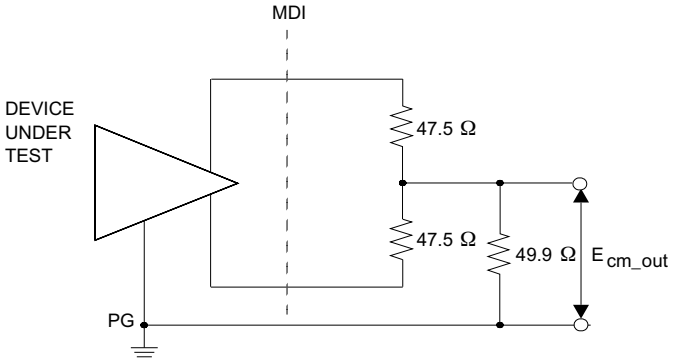


Figure 40–32—Common-mode output voltage test circuit

NOTE—The imbalance of the test equipment (such as the matching of the test resistors) must be insignificant relative to the balance requirements.

40.8.3.4 MDI fault tolerance

Each wire pair of the MDI shall, under all operating conditions, withstand without damage the application of short circuits of any wire to any other wire within the 4-pair cable for an indefinite period of time and shall resume normal operation after the short circuit(s) are removed. The magnitude of the current through such a short circuit shall not exceed 300 mA.

Each wire pair shall withstand without damage a 1000 V common-mode impulse applied at E_{cm} of either polarity (as indicated in Figure 40–33). The shape of the impulse shall be 0.3/50 μs (300 ns virtual front time, 50 μs virtual time of half value), as defined in IEC 60060.

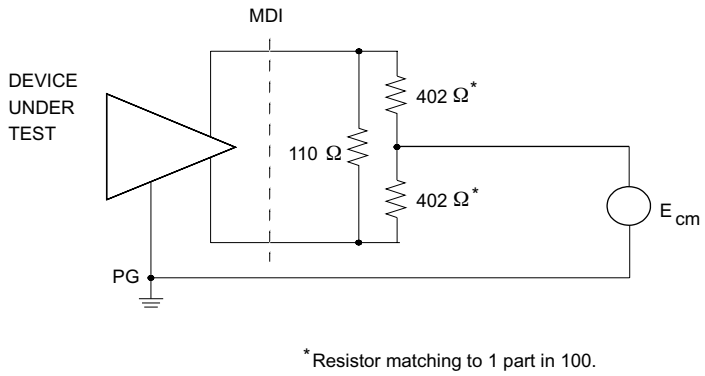


Figure 40–33—MDI fault tolerance test circuit

40.9 Environmental specifications

40.9.1 General safety

All equipment meeting this standard shall conform to IEC 60950: 1991.

40.9.2 Network safety

This subclause sets forth a number of recommendations and guidelines related to safety concerns; the list is neither complete nor does it address all possible safety issues. The designer is urged to consult the relevant local, national, and international safety regulations to ensure compliance with the appropriate requirements.

LAN cabling systems described in this subclause are subject to at least four direct electrical safety hazards during their installation and use. These hazards are as follows:

a) Direct contact between LAN components and power, lighting, or communications circuits.
b) Static charge buildup on LAN cabling and components.
c) High-energy transients coupled onto the LAN cabling system.
d) Voltage potential differences between safety grounds to which various LAN components are connected.

Such electrical safety hazards must be avoided or appropriately protected against for proper network installation and performance. In addition to provisions for proper handling of these conditions in an operational system, special measures must be taken to ensure that the intended safety features are not negated during installation of a new network or during modification or maintenance of an existing network.

40.9.2.1 Installation

It is a mandatory requirement that sound installation practice, as defined by applicable local codes and regulations, is followed in every instance in which such practice is applicable.

40.9.2.2 Installation and maintenance guidelines

It is a mandatory requirement that, during installation and maintenance of the cabling plant, care is taken to ensure that non-insulated network cabling conductors do not make electrical contact with unintended conductors or ground.

40.9.2.3 Telephony voltages

The use of building wiring brings with it the possibility of wiring errors that may connect telephony voltages to 1000BASE-T equipment. Other than voice signals (which are low voltage), the primary voltages that may be encountered are the "battery" and ringing voltages. Although there is no universal standard, the following maximums generally apply:

Battery voltage to a telephone line is generally 56 Vdc applied to the line through a balanced 400 Ω source impedance.

Ringing voltage is a composite signal consisting of an ac component and a dc component. The ac component is up to 175 V peak at 20 Hz to 60 Hz with a 100 Ω source resistance. The dc component is 56 Vdc with a 300 Ω to 600 Ω source resistance. Large reactive transients can occur at the start and end of each ring interval.

Although 1000BASE-T equipment is not required to survive such wiring hazards without damage, application of any of the above voltages shall not result in any safety hazard.

NOTE—Wiring errors may impose telephony voltages differentially across 1000BASE-T transmitters or receivers. Because the termination resistance likely to be present across a receiver's input is of substantially lower impedance than an off-hook telephone instrument, receivers will generally appear to the telephone system as off-hook telephones. Therefore, full-ring voltages will be applied for only short periods. Transmitters that are coupled using transformers will similarly appear like off-hook telephones (though perhaps a bit more slowly) due to the low resistance of the transformer coil.

40.9.3 Environment

40.9.3.1 Electromagnetic emission

A system integrating the 1000BASE-T PHY shall comply with applicable local and national codes for the limitation of electromagnetic interference.

40.9.3.2 Temperature and humidity

A system integrating the 1000BASE-T PHY is expected to operate over a reasonable range of environmental conditions related to temperature, humidity, and physical handling (such as shock and vibration). Specific requirements and values for these parameters are considered to be beyond the scope of this standard.

It is recommended that manufacturers indicate in the literature associated with the PHY the operating environmental conditions to facilitate selection, installation, and maintenance.

40.10 PHY labeling

It is recommended that each PHY (and supporting documentation) be labeled in a manner visible to the user with at least the following parameters:

a) Data rate capability in Mb/s
b) Power level in terms of maximum current drain (for external PHYs)
c) Port type (i.e., 1000BASE-T)
d) Any applicable safety warnings

40.11 Delay constraints

In half duplex mode, proper operation of a CSMA/CD LAN demands that there be an upper bound on the propagation delays through the network. This implies that MAC, PHY, and repeater implementors must conform to certain delay minima and maxima, and that network planners and administrators conform to constraints regarding the cabling topology and concatenation of devices. MAC constraints are specified in 35.2.4. Topological constraints are contained in Clause 42.

In full duplex mode, predictable operation of the MAC Control PAUSE operation (Clause 31, Annex 31B) also demands that there be an upper bound on the propagation delays through the network. This implies that MAC, MAC Control sublayer, and PHY implementors must conform to certain delay maxima, and that network planners and administrators conform to constraints regarding the cable topology and concatenation of devices.

The reference point for all MDI measurements is the peak point of the mid-cell transition corresponding to the reference code-bit, as measured at the MDI.

40.11.1 MDI to GMII delay constraints

Every 1000BASE-T PHY associated with a GMII shall comply with the bit delay constraints specified in Table 40–13 for half duplex operation and Table 40–14 for full duplex operation. These constraints apply for all 1000BASE-T PHYs. For any given implementation, the assertion and de-assertion delays on CRS shall be equal.

Table 40–13—MDI to GMII delay constraints (half duplex mode)

Sublayer measurement points	Event	Min (bit times)	Max (bit times)	Input timing reference	Output timing reference
GMII ⇔ MDI	TX_EN Sampled to MDI Output	—	84	GTX_CLK rising	1st symbol of SSD/CSReset/ CSExtend/ CSExtend_Err
	MDI input to CRS assert	—	244	1st symbol of SSD/CSReset	—
	MDI input to CRS de-assert	—	244	1st symbol of SSD/CSReset	—
	MDI input to COL assert	—	244	1st symbol of SSD/CSReset	—
	MDI input to COL de-assert	—	244	1st symbol of SSD/CSReset	—
	TX_EN sampled to CRS assert	—	16	GTX_CLK rising	—
	TX_EN sampled to CRS de-assert	—	16	GTX_CLK rising	—

40.11.2 DTE delay constraints (half duplex only)

Every DTE with a 1000BASE-T PHY shall comply with the bit delay constraints specified in Table 40–15 for half duplex operation.

Table 40–14—MDI to GMII delay constraints (full duplex mode)

Sublayer measurement points	Event	Min (bit times)	Max (bit times)	Input timing reference	Output timing reference
GMII ⇔ MDI	TX_EN Sampled to MDI Output	—	84	GTX_CLK rising	1st symbol of SSD/CSReset/ CSExtend/ CSExtend_Err
	MDI input to RX_DV de-assert	—	244	1st symbol of CSReset	—

Table 40–15— DTE delay constraints (half duplex mode)

Sublayer measurement points	Event	Min (bit times)	Max (bit times)	Input timing reference	Output timing reference
MAC ⇔ MDI	MAC transmit start to MDI output	—	132	—	1st symbol of SSD
	MDI input to collision detect	—	292	1st symbol of SSD	—
	MDI input to MDI output (nondeferred or Jam)	—	440	1st symbol of SSD	1st symbol of SSD
	MDI Input to MDI output (worse-case non-deferred transmit)	—	440	1st symbol of SSD	1st symbol of SSD

40.11.3 Carrier de-assertion/assertion constraint (half duplex mode)

To ensure fair access to the network, each DTE operating in half duplex mode shall, additionally, satisfy the following: (MAX MDI to MAC Carrier De-assert Detect) – (MIN MDI to MAC Carrier Assert Detect) < 16 Bit Times.

40.12 Protocol implementation conformance statement (PICS) proforma for Clause 40—Physical coding sublayer (PCS), physical medium attachment (PMA) sublayer and baseband medium, type 1000BASE-T[11]

The supplier of a protocol implementation that is claimed to conform to this clause shall complete the Protocol Implementation Conformance Statement (PICS) proforma listed in the following subclauses.

Instructions for interpreting and filling out the PICS proforma may be found in Clause 21.

[11]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so it can be used for its intended purpose and may further publish the completed PICS.

40.12.1 Identification

40.12.1.1 Implementation identification

Supplier	
Contact point for queries about the PICS	
Implementation Name(s) and Version(s)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Name(s)	
NOTES 1—Only the first three items are required for all implementations; other information may be completed as appropriate in meeting the requirements for the identification. 2—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

40.12.1.2 Protocol summary

Identification of protocol specification	IEEE Std 802.3-2002®, Clause 40, Physical coding sublayer (PCS), physical medium attachment (PMA) sublayer, and baseband medium, type 1000BASE-T
Identification of amendments and corrigenda to this PICS proforma which have been completed as part of this PICS	
Have any Exceptions items been required? No [ ] Yes [ ] (See Clause 21—The answer Yes means that the implementation does not conform to the standard)	
Date of Statement	

40.12.2 Major capabilities/options

Item	Feature	Subclause	Status	Support	Value/Comment
*GMII	PHY associated with GMII	40.1	O	Yes [ ] No [ ]	
*DTE	DTE with PHY not associated with GMII	40.1	O	Yes [ ] No [ ]	
AN	Support for Auto-Negotiation (Clause 28)	40.5.1	M	Yes [ ]	Required
OMS	Operation as MASTER or SLAVE	40.5.1	M	Yes [ ]	Required
*FDX	PHY supports full duplex mode	40.1	O	Yes [ ] No [ ]	
*HDX	PHY support half duplex mode	40.1	O	Yes [ ] No [ ]	
*INS	Installation / cabling	40.7	O	Yes [ ] No [ ]	Items marked with INS include installation practices and cabling specifications not applicable to a PHY manufacturer.
*AXO	Auto-Crossover	40.4.4	O	Yes [ ] No [ ]	PHY supports auto-crossover

40.12.3 Clause conventions

Item	Feature	Subclause	Status	Support	Value/Comment
CCO1	The values of all components in test circuits shall be	40.1.6	M	Yes [ ]	Accurate to within ±1% unless otherwise stated.

40.12.4 Physical Coding Sublayer (PCS)

Item	Feature	Subclause	Status	Support	Value/Comment
PCT1	The PCS shall	40.3.1.2	M	Yes [ ]	Implement the Data Transmission Enabling process as depicted in Figure 40–8 including compliance with the associated state variables specified in 40.3.3.
PCT2	PCS Transmit function shall	40.3.1.3	M	Yes [ ]	Conform to the PCS Transmit state diagram in Figure 40–9.
PCT3	PCS Transmit shall	40.4.5.1	M	Yes [ ]	Send code-groups according to the value assumed by the tx_mode variable.
PCT4	If the parameter config provided to the PCS by the PHY Control function via the PMA_CONFIG.indicate message assumes the value MASTER, PCS Transmit shall	40.3.1.3.1	M	Yes [ ]	Employ the transmitter side-stream scrambler generator polynomial specified for use with MASTER in 40.3.1.3.1.
PCT5	If the parameter config provided to the PCS by the PHY Control function via the PMA_CONFIG.indicate message assumes the value SLAVE, PCS Transmit shall	40.3.1.3.1	M	Yes [ ]	Employ the transmitter side-stream scrambler generator polynomial specified for use with SLAVE in 40.3.1.3.1.
PCT6	In no case shall	40.3.1.3.1	M	Yes [ ]	The scrambler state be initialized to all zeros.
PCT7	If tx_error_n=1 when the condition (tx_enable_n * tx_enable_{n-2}) = 1, error indication is signaled by means of symbol substitution, wherein the values of $Sd_n[5:0]$ are ignored during mapping and the symbols corresponding to the row denoted as “xmt_err” in Table 40–1 and Table 40–2 shall be used.	40.3.1.3.5	M	Yes [ ]	

Item	Feature	Subclause	Status	Support	Value/Comment
PCT8	If tx_error_n=0 when the variable $csreset_n$ = 1, the convolutional encoder reset condition is normal. This condition is indicated by means of symbol substitution, where the values of Sd_n[5:0] are ignored during mapping and the symbols corresponding to the row denoted as “CSReset” in Table 40–1 and Table 40–2 shall be used.	40.3.1.3.5	M	Yes [ ]	
PCT9	If tx_error_n=1 is asserted when the variable $csreset_n$ = 1, the convolutional encoder reset indicates carrier extension. In this condition, the values of Sd_n[5:0] are ignored during mapping and the symbols corresponding to the row denoted as “CSExtend” in Table 40–1 and Table 40–2 shall be used when TXD_n = 0x’0F, and the row denoted as “CSExtend_Err” in Table 40–1 and Table 40–2 shall be used when $TXD_n \neq$ 0x’0F.	40.3.1.3.5	M	Yes [ ]	
PCT10	In case carrier extension with error is indicated during the first octet of CSReset, the error condition shall be encoded during the second octet of CSReset, and during the subsequent two octets of the End-of-Stream delimiter.	40.3.1.3.5	M	Yes [ ]	
PCT11	The symbols corresponding to the SSD1 row in Table 40–1 shall be used when the condition (tx_enable_n) * (!tx_enable_{n-1}) = 1.	40.3.1.3.5	M	Yes [ ]	
PCT12	The symbols corresponding to the SSD2 row in Table 40–1 shall be used when the condition (tx_enable_{n-1})* (!tx_enable_{n-2}) = 1.	40.3.1.3.5	M	Yes [ ]	
PCT13	If carrier extend error is indicated during ESD, the symbols corresponding to the ESD_Ext_Err row in Table 40–1 shall be used.	40.3.1.3.5	M	Yes [ ]	
PCT14	The symbols corresponding to the ESD1 row in Table 40–1 shall be used when the condition (!tx_enable_{n-2}) * (tx_enable_{n-3}) = 1, in the absence of carrier extend error indication at time n.	40.3.1.3.5	M	Yes [ ]	

Item	Feature	Subclause	Status	Support	Value/Comment
PCT15	The symbols corresponding to the ESD2_Ext_0 row in shall be used when the condition (!tx_enable_{n-3}) * (tx_enable_{n-4}) * (!tx_error_{n}) * (!tx_error_{n-1}) = 1.	40.3.1.3.5	M	Yes [ ]	
PCT16	The symbols corresponding to the ESD2_Ext_1 row in Table 40–1 shall be used when the condition (!tx_enable_{n-3}) * (tx_enable_{n-4}) * (!tx_error_{n}) * (tx_error_{n-1}) * (tx_error_{n-2}) * (tx_error_{n-3})= 1.	40.3.1.3.5	M	Yes [ ]	
PCT17	The symbols corresponding to the ESD2_Ext_2 row in Table 40–1 shall be used when the condition (!tx_enable_{n-3}) * (tx_enable_{n-4}) * (tx_error_{n}) * (tx_error_{n-1}) * (tx_error_{n-2}) * (tx_error_{n-3}) * (TXD_{n}=0x0F)= 1, in the absence of carrier extend error indication.	40.3.1.3.5	M	Yes [ ]	

40.12.4.1 PCS receive functions

Item	Feature	Subclause	Status	Support	Value/Comment
PCR1	PCS Receive function shall	40.3.1.4	M	Yes [ ]	Conform to the PCS Receive state diagram shown in Figure 40–10a including compliance with the associated state variables as specified in 40.3.3.
PCR2	The PHY shall	40.3.1.4.2	M	Yes [ ]	Descramble the data stream and return the proper sequence of data bits RXD<7:0> to the GMII.
PCR3	For side-stream descrambling, the MASTER PHY shall employ	40.3.1.4.2	M	Yes [ ]	The receiver scrambler generator polynomial specified for MASTER operation in 40.3.1.4.2.
PCR4	For side-stream descrambling, the SLAVE PHY shall employ	40.3.1.4.2	M	Yes [ ]	The receiver scrambler generator polynomial specified for SLAVE operation in 40.3.1.4.2.

40.12.4.2 Other PCS functions

Item	Feature	Subclause	Status	Support	Value/Comment
PCO1	The PCS Reset function shall	40.3.1.1	M	Yes [ ]	Be executed any time “power on” or receipt of a request for reset from the management entity occurs, including compliance with the associated state variables as specified in 40.3.3.
PCO2	The PCS shall	40.3.1.5	M	Yes [ ]	Implement the Carrier Sense process as depicted in Figure 40–11, including compliance with the associated state variables as specified in 40.3.3.
PCO3	Symb-timer shall be generated	40.3.3.3	M	Yes [ ]	Synchronously with TX_TCLK.

40.12.5 Physical Medium Attachment (PMA)

Item	Feature	Subclause	Status	Support	Value/Comment
PMF1	PMA Reset function shall be executed	40.4.2.1	M	Yes [ ]	At power on and upon receipt of a reset request from the management entity or from PHY Control.
PMF2	PMA Transmit shall	40.4.2.2	M	Yes [ ]	Continuously transmit onto the MDI pulses modulated by the quinary symbols given by tx_symb_vector[BI_DA], tx_symb_vector[BI_DB], tx_symb_vector[BI_DC], and tx_symb_vector[BI_DD], respectively.
PMF3	The four transmitters shall be driven by the same transmit clock, TX_TCLK	40.4.2.2	M	Yes [ ]	
PMF4	PMA Transmit shall	40.4.2.2	M	Yes [ ]	Follow the mathematical description given in 40.4.3.1.
PMF5	PMA Transmit shall comply with	40.4.2.2	M	Yes [ ]	The electrical specifications given in 40.6.
PMF6	When the PMA_CONFIG.indicate parameter config is MASTER, the PMA Transmit function shall	40.4.2.2	M	Yes [ ]	Source the transmit clock TX_TCLK from a local clock source while meeting the transmit jitter requirements of 40.6.1.2.5.

Item	Feature	Subclause	Status	Support	Value/Comment
PMF7	When the PMA_CONFIG.indicate parameter config is SLAVE, the PMA Transmit function shall	40.4.2.2	M	Yes [ ]	Source the transmit clock TX_TCLK from the recovered clock of 40.4.2.5 while meeting the jitter requirements of 40.6.1.2.5.
PMF8	PMA Receive function shall	40.4.2.3	M	Yes [ ]	Translate the signals received on pairs BI_DA BI_DB, BI_DC and BI_DD into the PMA_UNITDATA.indicate parameter rx_symb_vector with a symbol error rate of less than one part in 10^{10}.
PMF9	PHY Control function shall	40.4.2.4	M	Yes [ ]	Comply with the state diagram descriptions given in Figure 40–15.
PMF10	The Link Monitor function shall	40.4.2.5	M	Yes [ ]	Comply with the state diagram shown in Figure 40–16.
PMF11	Clock Recovery function shall provide	40.4.2.6	M	Yes [ ]	Provide clocks suitable for signal sampling on each line so that the symbol-error rate indicated in 40.4.2.3 is achieved.
PMF12	The symbol response shall comply with	40.4.3.1	M	Yes [ ]	The electrical specifications given in 40.6.
PMF13	The four signals received on pairs BI_DA, BI_DB, BI_DC, and BI_DD shall be processed within the PMA Receive function to yield	40.4.3.2	M	Yes [ ]	The quinary received symbols rx_symb_vector[BI_DA], rx_symb_vector[BI_DB], rx_symb_vector[BI_DC], and rx_symb_vector[BI_DD].
PMF14	If an automatic configuration method is used, it shall	40.4.4	M	Yes [ ]	Comply with the specifications of 40.4.4.
PMF15	The PMA shall	40.4.5.1	M	Yes [ ]	Generate the config variable continuously and pass it to the PCS via the PMA_CONFIG.indicate primitive.
PMF16	The variable link_det shall take the value	40.4.5.1	AXO:M	N/A [ ] Yes [ ]	TRUE or FALSE as per 40.4.4.1.
PMF17	The variable MDI_status shall take the value	40.4.5.1	AXO:M	N/A [ ] Yes [ ]	MDI or MDI-X as per Table 40–12.
PMF18	PCS Transmit shall	40.4.5.1	M	Yes [ ]	Send code-groups according to the value assumed by tx_mode.
PMF19	The A_timer shall have a period of	40.4.5.2	AXO:M	N/A [ ] Yes [ ]	1.3s ± 25%.
PMF20	The maxwait_timer timer shall expire	40.4.5.2	M	Yes [ ]	750 ± 10 ms if config = MASTER or 350 ± 5ms if config = SLAVE

Item	Feature	Subclause	Status	Support	Value/Comment
PMF21	The minwait_timer timer shall expire	40.4.5.2	M	Yes [ ]	1 ±0.1μs after being started.
PMF22	The sample_timer shall have a period of	40.4.5.2	AXO:M	N/A [ ] Yes [ ]	62 ± 2ms.
PMF23	The stabilize_timer shall expire	40.4.5.2	M	Yes [ ]	1 ± 0.1 μs after being started.

40.12.6 Management interface

Item	Feature	Subclause	Status	Support	Value/Comment
MF1	All 1000BASE-T PHYs shall provide support for Auto-Negotiation (Clause 28) and shall be capable of operating as MASTER or SLAVE.	40.5.1	M	Yes [ ]	
MF2	A 100BASE-T PHY shall	40.5.1.1	M	Yes [ ]	Use the management register definitions and values specified in Table 40–3.

40.12.6.1 1000BASE-T Specific Auto-Negotiation Requirements

Item	Feature	Subclause	Status	Support	Value/Comment
AN1	1000BASE-T PHYs shall	40.5.1.2	M	Yes [ ]	Exchange one Auto-Negotiation Base Page, a 1000BASE-T formatted Next Page, and two 1000BASE-T unformatted Next Pages in sequence, without interruption, as specified in Table 40–4.
AN2	The MASTER-SLAVE relationship shall be determined during Auto-Negotiation	40.5.2	M	Yes [ ]	Using Table 40–5 with the 1000BASE-T Technology Ability Next Page bit values specified in Table 40–4 and information received from the link partner.
AN3	Successful completion of the MASTER-SLAVE resolution shall	40.5.2	M	Yes [ ]	Be treated as MASTER-SLAVE configuration resolution complete.
AN4	A seed counter shall be provided to	40.5.2	M	Yes [ ]	Track the number of seed attempts.
AN5	At start-up, the seed counter shall be set to	40.5.2	M	Yes [ ]	Zero.
AN6	The seed counter shall be incremented	40.5.2	M	Yes [ ]	Every time a new random seed is sent.

Item	Feature	Subclause	Status	Support	Value/Comment
AN7	When MASTER-SLAVE resolution is complete, the seed counter shall be reset to 0 and bit 10.15 shall be set to logical zero.	40.5.2	M	Yes [ ]	
AN8	Maximum seed attempts before declaring a MASTER_SLAVE configuration Resolution Fault	40.5.2	M	Yes [ ]	Seven.
AN9	During MASTER_SLAVE configuration, the device with the higher seed value shall	40.5.2	M	Yes [ ]	Become the MASTER.
AN10	During MASTER_SLAVE configuration, the device with the lower seed value shall	40.5.2	M	Yes [ ]	Become the SLAVE.
AN11	Both PHYs set in manual mode to be either MASTER or SLAVE shall be treated as	40.5.2	M	Yes [ ]	MASTER-SLAVE resolution fault (failure) condition.
AN12	MASTER-SLAVE resolution fault (failure) condition shall result in	40.5.2	M	Yes [ ]	MASTER-SLAVE Configuration Resolution Fault bit (10.15) to be set to logical one.
AN13	MASTER-SLAVE Configuration resolution fault condition shall be treated as	40.5.2	M	Yes [ ]	MASTER-SLAVE Configuration Resolution complete.
AN14	MASTER-SLAVE Configuration resolution fault condition shall	40.5.2	M	Yes [ ]	Cause link_status_1000BASE-T to be set to FAIL.

40.12.7 PMA Electrical Specifications

Item	Feature	Subclause	Status	Support	Value/Comment
PME15	The PHY shall provide electrical isolation between	40.6.1.1	M	Yes [ ]	The port device circuits including frame ground, and all MDI leads.
PME16	PHY-provided electrical separation shall withstand at least one of three electrical strength tests	40.6.1.1	M	Yes [ ]	a) 1500 V rms at 50Hz to 60Hz for 60 s, applied as specified in Section 5.3.2 of IEC 60950:1991. b) 2250 Vdc for 60 s, applied as specified in Section 5.3.2 of IEC 60950: 1991. c) A sequence of ten 2400 V impulses of alternating polarity, applied at intervals of not less than 1 s. The shape of the impulses shall be 1.2/50 μs. (1.2 μs virtual front time, 50 μs virtual time or half value), as defined in IEC 60060.
PME17	There shall be no insulation breakdown as defined in Section 5.3.2 of IEC 60950, during the test.	40.6.1.1	M	Yes [ ]	
PME18	The resistance after the test shall be at least	40.6.1.1	M	Yes [ ]	2 MΩ, measured at 500 Vdc.
PME19	The transmitter MASTER-SLAVE timing jitter test channel shall	40.6.1.1.1	M	Yes [ ]	Be constructed by combining 100 Ω and 120 Ω cable segments that meet or exceed ISO.IEC 11801 Category 5 specifications for each pair as shown in Figure 40–18 with the lengths and additional restrictions on parameters described in Table 40–6.
PME20	The ends of the MASTER-SLAVE timing jitter test channel shall	40.6.1.1.1	M	Yes [ ]	Be connectorized with connectors meeting or exceeding ANSI/TIA/EIA-568-A:1995 or ISO/IEC 11801:1995 Category 5 specifications.
PME21	The return loss of the MASTER-SLAVE timing jitter test channel shall	40.6.1.1.1	M	Yes [ ]	Meet the return loss requirements of 40.7.2.3.
PME22	The return loss of the MASTER-SLAVE timing jitter test channel shall	40.6.1.1.1	M	Yes [ ]	Meet the crosstalk requirements of 40.7.3 on each pair.
PME23	The test modes described in 40.6.1.1.2 shall be provided for testing of the transmitted waveform, transmitter distortion and transmitted jitter.	40.6.1.1.2	M	Yes [ ]	

Item	Feature	Subclause	Status	Support	Value/Comment
PME24	For a PHY with a GMII interface the test modes shall be enabled by	40.6.1.1.2	M	Yes [ ]	Setting bits 9:13-15 (1000BASE-T Control Register) of the GMII Management register set as shown in Table 40–7.
PME25	The test modes shall only change the data symbols provided to the transmitter circuitry and shall not alter the electrical and jitter characteristics of the transmitter and receiver from those of normal operation.	40.6.1.1.2	M	Yes [ ]	
PME26	A PHY without a GMII shall provide a means to enable the test modes for conformance testing.	40.6.1.1.2	M	Yes [ ]	
PME27	When transmit test mode 1 is enabled, the PHY shall transmit	40.6.1.1.2	M	Yes [ ]	The sequence of data symbols specified in 40.6.1.1.2 continuously from all four transmitters.
PME28	When in test mode 1, the transmitter shall time the transmitted symbols	40.6.1.1.2	M	Yes [ ]	From a 125.00 MHz ± 0.01% clock in the MASTER timing mode.
PME29	When test mode 2 is enabled, the PHY shall transmit	40.6.1.1.2	M	Yes [ ]	The data symbol sequence {+2,–2} repeatedly on all four channels.
PME30	When in test mode 2, the transmitter shall time the transmitted symbols	40.6.1.1.2	M	Yes [ ]	From a 125.00 MHz ± 0.01% clock in the MASTER timing mode.
PME31	When transmit test mode 3 is enabled, the PHY shall transmit	40.6.1.1.2	M	Yes [ ]	The data symbol sequence {+2,–2} repeatedly on all four channels.
PME32	When in test mode 3, the transmitter shall time the transmitted symbols	40.6.1.1.2	M	Yes [ ]	From a 125 MHz ± 1% clock in the SLAVE timing mode.
PME33	When test mode 4 is enabled, the PHY shall transmit	40.6.1.1.2	M	Yes [ ]	The data symbols generated by the scrambler polynomial specified in 40.6.1.1.2.
PME34	When test mode 4 is enabled, the PHY shall	40.6.1.1.2	M	Yes [ ]	Use the bit sequences generated by the scrambler bits shown in 40.6.1.1.2 to generate the quinary symbols, s_n, as shown in Table 40–8.
PME35	When test mode 4 is enabled, the maximum-length shift register used to generate the sequences defined by this polynomial shall be	40.6.1.1.2	M	Yes [ ]	Updated once per symbol interval (8 ns).

Item	Feature	Subclause	Status	Support	Value/Comment
PME36	When test mode 4 is enabled, the bit sequences, $x0_n$, $x1_n$, and $x2_n$, generated from combinations of the scrambler bits shown in 40.6.1.1.2 shall be	40.6.1.1.2	M	Yes [ ]	Used to generate the quinary symbols, s_n, as shown in Table 40–8.
PME37	When test mode 4 is enabled, the quinary symbol sequence shall be	40.6.1.1.2	M	Yes [ ]	Presented simultaneously to all transmitters.
PME38	When in test mode 4, the transmitter shall time the transmitted symbols	40.6.1.1.2	M	Yes [ ]	From a 125.00 MHz ± 0.01% clock in the MASTER timing mode.
PME39	The test fixtures defined in Figure 40–22, Figure 40–23, Figure 40–24, and Figure 40–25 or their functional equivalents shall be used for measuring transmitter specifications.	40.6.1.1.3	M	Yes [ ]	
PME40	The test filter used in transmitter test fixtures 1 and 3 shall	40.6.1.1.3	M	Yes [ ]	Have the continuous time transfer function specified in 40.6.1.1.3 or its discrete time equivalent.
PME41	The disturbing signal V_d shall	40.6.1.1.3	M	Yes [ ]	Have the characteristics listed in Table 40–9.
PME42	To allow for measurement of transmitted jitter in MASTER and SLAVE modes the PHY shall provide access to the 125 MHz symbol clock, TX_TCLK that times the transmitted symbols.	40.6.1.1.3	M	Yes [ ]	
PME43	To allow for measurement of transmitted jitter in MASTER and SLAVE modes the PHY shall provide a means to enable the TX_TCLK output if it is not normally enabled.	40.6.1.1.3	M	Yes [ ]	
PME44	The PMA shall	40.6.1.2	M	Yes [ ]	Provide the Transmit function specified in 40.4.2.2 in accordance with the electrical specifications of this clause.
PME45	Where a load is not specified, the transmitter shall	40.6.1.2	M	Yes [ ]	Meet the requirements of this clause with a 100 Ω resistive differential load connected to each transmitter output.
PME46	The tolerance on the poles of the test filters used in 40.6 shall be ± 1%.	40.6.1.2	M	Yes [ ]	

Item	Feature	Subclause	Status	Support	Value/Comment
PME47	When in transmit test mode 1 and observing the differential signal output at the MDI using test fixture 1, for each pair, with no intervening cable, the absolute value of the peak of the waveform at points A and B as defined in Figure 40–19 shall fall within	40.6.1.2.1	M	Yes [ ]	The range of 0.67 V to 0.82 V (0.75 V ± 0.83 dB).
PME48	The absolute value of the peak of the waveforms at points A and B shall	40.6.1.2.1	M	Yes [ ]	Differ by less than 1%.
PME49	The absolute value of the peak of the waveform at points C and D as defined in Figure 40–19 shall differ	40.6.1.2.1	M	Yes [ ]	From 0.5 times the average of the absolute values of the peaks of the waveform at points A and B by less than 2%.
PME50	When in transmit test mode 1 and observing the differential transmitted output at the MDI, for either pair, with no intervening cabling, the peak value of the waveform at point F as defined in Figure 40–19 shall be	40.6.1.2.2	M	Yes [ ]	Greater than 73.1% of the magnitude of the negative peak value of the waveform at point F. Point G is defined as the point exactly 500 ns after point F. Point F is defined as the point where the waveform reaches it's minimum value at the location indicated in Figure 40–19.
PME51	When in transmit test mode 1 and observing the differential transmitted output at the MDI, for either pair, with no intervening cabling, the peak value of the waveform at point J as defined in Figure 40–19 shall be	40.6.1.2.2	M	Yes [ ]	Greater than 73.1% of the magnitude of the peak value of the waveform at point H. Point J is defined as the point exactly 500 ns after point H. Point H is defined as the point where the waveform reaches it's maximum value at the location indicated in Figure 40–19.
PME52	When in test mode 1 and observing the differential signal output at the MDI using transmitter test fixture 1, for each pair, with no intervening cable, the voltage waveforms at points A, B, C, D defined in Figure 40–19, after the normalization described within the referenced subclause, shall	40.6.1.2.3	M	Yes [ ]	Lie within the time domain template 1 defined in Figure 40–26 and the piecewise linear interpolation between the points in Table 40–10. The waveforms may be shifted in time as appropriate to fit within the template.
PME53	When in test mode 1 and observing the differential signal output at the MDI using transmitter test fixture 1, for each pair, with no intervening cable, the voltage waveforms at points F and H defined in Figure 40–19, after the normalization described within the referenced subclause, shall	40.6.1.2.3	M	Yes [ ]	Lie within the time domain template 2 defined in Figure 40–26 and the piecewise linear interpolation between the points in Table 40–11. The waveforms may be shifted in time as appropriate to fit within the template.

Item	Feature	Subclause	Status	Support	Value/Comment
PME54	When in test mode 4 and observing the differential signal output at the MDI using transmitter test fixture 3, for each pair, with no intervening cable, the peak distortion as defined below shall be	40.6.1.2.4	M	Yes [ ]	Less than 10 mV.
PME55	When in the normal mode of operation as the MASTER, the peak-to-peak value of the MASTER TX_TCLK jitter relative to an unjittered reference shall be	40.6.1.2.5	M	Yes [ ]	Less than 1.4 ns.
PME56	When the jitter waveform on TX_TCLK is filtered by a high-pass filter, $H_{jf1}(f)$ having the transfer function specified in 40.6.1.2.5, the peak-to-peak value of the resulting filtered timing jitter plus J_{txout}, shall be	40.6.1.2.5	M	Yes [ ]	Less than 0.3 ns.
PME57	When in the normal mode of operation as the SLAVE, receiving valid signals from a compliant PHY operating as the MASTER using the test channel defined in 40.6.1.1.1, with test channel port A connected to the SLAVE, the peak-to-peak value of the SLAVE TX_TCLK jitter relative to the MASTER TX_TCLK shall be	40.6.1.2.5	M	Yes [ ]	Less than 1.4 ns after the receiver is properly receiving the data and has set bit 10.13 of the GMII management register set to 1.
PME58	When the jitter waveform on TX_TCLK is filtered by a high-pass filter, $H_{jf2}(f)$, having the transfer function specified in 40.6.1.2.5, the peak-to-peak value of the resulting filtered timing jitter plus J_{txout} shall be	40.6.1.2.5	M	Yes [ ]	No more than 0.4 ns greater than the simultaneously measured peak-to-peak value of the MASTER jitter filtered by $H_{jf1}(f)$
PME59	For all jitter measurements the peak-to-peak value shall be	40.6.1.2.5	M	Yes [ ]	Measured over an unbiased sample of at least 10^5 clock edges.
PME60	For all unfiltered jitter measurements the peak-to-peak value shall be	40.6.1.2.5	M	Yes [ ]	Measured over an interval of not less than 100 ms and not more than 1 second.
PME61	The quinary symbol transmission rate on each pair of the MASTER PHY shall be	40.6.1.2.6	M	Yes [ ]	125.00 MHz ± 0.01%
PME62	The PMA shall provide the Receive function specified in 40.3.1.4 in accordance with the electrical specifications of this clause.	40.6.1.3	M	Yes [ ]	

Item	Feature	Subclause	Status	Support	Value/Comment
PME63	The patch cabling and interconnecting hardware used in test configurations shall be	40.6.1.3	M	Yes [ ]	Within the limits specified in 40.7.
PME64	Differential signals received on the receive inputs that were transmitted within the specifications given in 40.6.1.2 and have then passed through a link compatible with 40.7, shall be translated into	40.6.1.3.1	M	Yes [ ]	One of the PMA_UNITDATA.indicate messages with a 4-D symbol rate error less than 10^{-10} and sent to the PCS after link bring-up. Since the 4-D symbols are not accessible, this specification shall be satisfied by a frame error rate less than 10^{-7} for 125 octet frames.
PME65	The receive feature shall	40.6.1.3.2	M	Yes [ ]	Properly receive incoming data with a 5-level symbol rate within the range 125.00 MHz ± 0.01%.
PME66	The signal generator for the common-mode test shall be	40.6.1.3.3	M	Yes [ ]	Capable of providing a sine wave signal of 1 MHz to 250 MHz.
PME67	While sending data from the transmitter the receiver shall	40.6.1.3.3	M	Yes [ ]	Send the proper PMA_UNITDATA.indicate messages to the PCS as the signal generator frequency is varied from 1 MHz to 250 MHz.
PME68	While receiving data from a transmitter specified in 40.6.1.2 through a link segment specified in 40.7 connected to all MDI duplex channels, a receiver shall	40.6.1.3.4	M	Yes [ ]	Send the proper PMA_UNITDATA.indicate message to the PCS when any one of the four pairs is connected to a noise source as described in Figure 40–28.
PME69	The alien crosstalk test specified in 40.6.1.3.4 shall be satisfied by	40.6.1.3.4	M	Yes [ ]	A frame error rate of less than 10^{-7} for 125 octet frames
PME70	The noise source shall be	40.6.1.3.4	M	Yes [ ]	Connected to one of the MDI inputs using Category 5 balanced cable of a maximum length of 0.5 m.

40.12.8 Characteristics of the link segment

Item	Feature	Subclause	Status	Support	Value/Comment
LKS1	All implementations of the balanced cabling link shall	40.7.1	M	Yes [ ]	Be compatible at the MDI.
LKS2	1000BASE-T links shall	40.7.1	M	Yes [ ]	Consist of Category 5 components as specified in ANSI/TIA/EIA-568-A:1995 and ISO/IEC 11801:1995.
LKS3	Link segment testing shall be conducted using	40.7.2	M	Yes [ ]	Source and load impedances of 100 Ω.
LKS4	The tolerance on the poles of the test filter used in this section shall be	40.7.2		Yes [ ]	± 1%.
LKS5	The insertion loss of each duplex channel shall be	40.7.2.1	M	Yes [ ]	Less than $2.1 f^{0.529} + 0.4/f$ (dB) at all frequencies from 1 MHz to 100 MHz. This includes the attenuation of the balanced cabling pairs, connector losses, and patch cord losses of the duplex channel.
LKS6	The insertion loss specification shall be met when	40.7.2.1	M	Yes [ ]	The duplex channel is terminated in 100 Ω.
LKS7	The return loss of each duplex channel shall be	40.7.2.3	M	Yes [ ]	As specified in 40.7.2.3 at all frequencies from 1 MHz to 100 MHz.
LKS8	The reference impedance for return loss measurement shall be	40.7.2.3	M	Yes [ ]	100 Ω.
LKS9	The NEXT loss between duplex channel pairs of a link segment shall be	40.7.3.1.1	M	Yes [ ]	At least $27.1 - 16.8\log_{10}(f/100)$ (where f is the frequency in MHz over the frequency range 1 MHz to 100 MHz.)
LKS10	The worst case ELFEXT loss between duplex channel pairs of a link segment shall be	40.7.3.2	M	Yes [ ]	Greater than $17 - 20\log_{10}(f/100)$ dB (where f is the frequency in MHz) over the frequency range 1 MHz to 100 MHz.
LKS11	The Power Sum loss between a duplex channel and the three adjacent disturbers shall be	40.7.3.2.2	M	Yes [ ]	Greater than $14.4 - 20\log_{10}(f/100)$ dB where f is the frequency in MHz over the frequency range of 1 MHz to 100 MHz.

Item	Feature	Subclause	Status	Support	Value/Comment
LKS12	The propagation delay of a link segment shall	40.7.4.1	M	Yes [ ]	Not exceed 570 ns at all frequencies from 2 MHz to 100 MHz.
LKS13	The difference in propagation delay, or skew, between all duplex channel pair combinations of a link segment under all conditions shall not exceed	40.7.4.2	M	Yes [ ]	50 ns at all frequencies between 2 MHz and 100 MHz.
LKS14	Once installed, the skew between pairs due to environmental conditions shall not vary	40.7.4.2	M	Yes [ ]	More than ± 10 ns.

40.12.9 MDI requirements

Item	Feature	Subclause	Status	Support	Value/Comment
MDI1	MDI connector	40.8.1	M	Yes [ ]	8-Way connector as per IEC 60603-7: 1990.
MDI2	Connector used on cabling	40.8.1	M	Yes [ ]	Plug.
MDI3	Connector used on PHY	40.8.1	M	Yes [ ]	Jack (as opposed to plug).
MDI4	MDI connector	40.8.2	M	Yes [ ]	A PHY that implements the crossover function shall be marked with the graphical symbol X.
MDI5	The MDI connector (jack) when mated with a balanced cabling connector (plug) shall	40.8.3	M	Yes [ ]	Meet the electrical requirements for Category 5 connecting hardware for use with 100 Ω Category 5 cable as specified in ANSI/TIA/EIA-568-A:1995 and ISO/IEC 11801:1995.
MDI6	The mated MDI connector and balanced cabling connector shall	40.8.3	M	Yes [ ]	Not have a FEXT loss greater than $40 - 20\log_{10}(f/100)$ over the frequency range 1 MHz to 100 MHz between all contact pair combinations shown in Table 40–12.
MDI7	No spurious signals shall be emitted onto the MDI when the PHY is held in power down mode as defined in 22.2.4.1.5, independent of the value of TX_EN, when released from power down mode, or when external power is first applied to the PHY.	40.8.3	M	Yes [ ]	

Item	Feature	Subclause	Status	Support	Value/Comment
MDI8	The differential impedance as measured at the MDI for each transmit/receive channel shall be such that	40.8.3.1	M	Yes [ ]	Any reflection due to differential signals incident upon the MDI from a balanced cabling having an impedance of 100 Ω ± 15% is at least 16 dB over the frequency range of 2.0 MHz to 40 MHz and at least $10 - 20\log_{10}(f/80)$ dB over the frequency range 40 MHz to 100 MHz (f in MHz).
MDI9	This return loss shall be maintained	40.8.3.1	M	Yes [ ]	At all times when the PHY is transmitting data or control symbols.
MDI10	The common-mode to differential-mode impedance balance of each transmit output shall exceed	40.8.3.2	M	Yes [ ]	The value specified by the equations specified in 40.8.3.2. Test mode 4 may be used to generate an appropriate transmitter output.
MDI11	The magnitude of the total common-mode output voltage, E_{cm_out}, on any transmit circuit, when measured as shown in Figure 40–32, shall be	40.8.3.3	M	Yes [ ]	Less than 50 mv peak-to-peak when transmitting data.
MDI12	Each wire pair of the MDI shall	40.8.3.4	M	Yes [ ]	Withstand without damage the application of short circuits across the MDI port for an indefinite period of time without damage.
MDI13	Each wire pair of the MDI shall resume	40.8.3.4	M	Yes [ ]	Normal operation after such faults are removed.
MDI14	The magnitude of the current through the short circuit specified in PME64 shall not exceed	40.8.3.4	M	Yes [ ]	300 mA.
MDI15	Each wire pair shall withstand without damage	40.8.3.4	M	Yes [ ]	A 1000 V common-mode impulse of either polarity (E_{cm} as indicated in Figure 40–33).
MDI16	The shape of the impulse shall be	40.8.3.4	M	Yes [ ]	0.3/50 μs (300 ns virtual front time, 50 μs virtual time of half value), as defined in IEC 60060.

40.12.10 General safety and environmental requirements

Item	Feature	Subclause	Status	Support	Value/Comment
ENV1	Conformance to safety specifications	40.9.1	M	Yes [ ]	IEC 60950.
ENV2	Installation practice	40.9.2.1	INS:M	N/A [ ] Yes [ ]	Sound practice, as defined by applicable local codes.
ENV3	Care taken during installation to ensure that non-insulated network cabling conductors do not make electrical contact with unintended conductors or ground.	40.9.2.2	INS:M	N/A [ ] Yes [ ]	
ENV4	1000BASE-T equipment shall be capable of withstanding a telephone battery supply from the outlet as described in 40.9.2.3.	40.9.2.3	M	Yes [ ]	
ENV5	A system integrating the 1000BASE-T PHY shall comply with applicable local and national codes for the limitation of electromagnetic interference.	40.9.3.1	INS:M	N/A [ ] Yes [ ]	

40.12.11 Timing requirements

Item	Feature	Subclause	Status	Support	Value/Comment
TR1	Every 1000BASE-T PHY associated with a GMII shall	40.11.1	M	Yes [ ]	Comply with the bit delay constraints specified in Table 40–13 for half duplex operation and Table 40–14 for full duplex operation. These constraints apply for all 1000BASE-T PHYs.
TR2	For any given implementation, the assertion delays on CRS shall	40.11.1	M	Yes [ ]	Be equal.
TR3	Every DTE with a 1000BASE-T PHY shall	40.11.2	M	Yes [ ]	Comply with the bit delay constraints specified in Table 40–15.
TR4	To ensure fair access to the network, each DTE operating in half duplex mode shall, additionally, satisfy the following:	40.11.3	M	Yes [ ]	(MAX MDI to MAC Carrier De-assert Detect) – (MIN MDI to MAC Carrier Assert Detect) < 16 Bit Times.

41. Repeater for 1000 Mb/s baseband networks

41.1 Overview

41.1.1 Scope

Clause 41 defines the functional and electrical characteristics of a repeater for use with ISO/IEC 8802-3 1000 Mb/s baseband networks. A repeater for any other ISO/IEC 8802-3 network type is beyond the scope of this clause. The relationship of this standard to the entire ISO/IEC 8802-3 CSMA/CD LAN standard is shown in Figure 41–1. The purpose of the repeater is to provide a simple, inexpensive, and flexible means of coupling two or more segments.

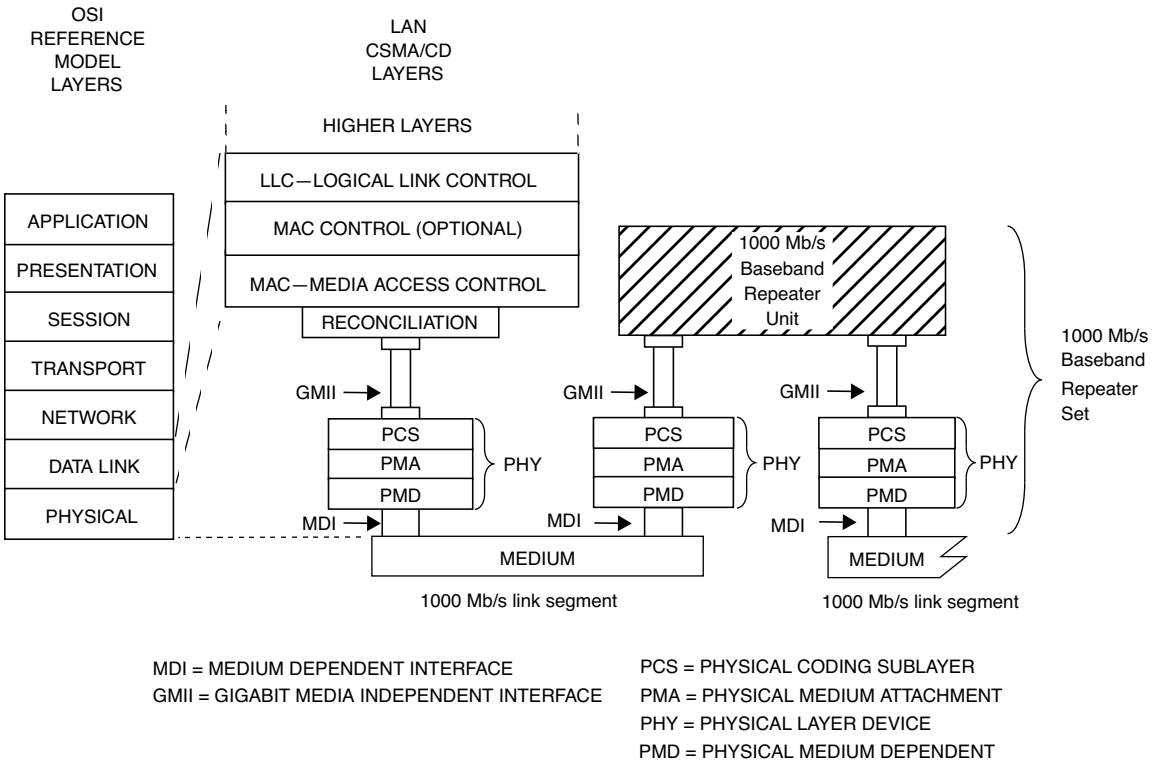


Figure 41–1—1000 Mb/s repeater set relationship to the ISO/IEC OSI reference model

41.1.1.1 Repeater set

Repeater sets are an integral part of all 1000 Mb/s baseband networks with more than two DTEs and are used to extend the physical system topology by providing a means of coupling two or more segments. A single repeater set is permitted within a single collision domain to provide the maximum connection path length. Allowable topologies contain only one operative signal path between any two points on the network. A repeater set is not a station and does not count toward the overall limit of 1024 stations on a network.

A repeater set can receive and decode data from any segment under worst-case noise, timing, and signal amplitude conditions. It retransmits the data to all other segments attached to it with timing, amplitude, and coding restored. The retransmission of data occurs simultaneously with reception. If a collision occurs, the repeater set propagates the collision event throughout the network by transmitting a Jam signal. A repeater set also provides a degree of protection to a network by isolating a faulty segment's carrier activity from propagating through the network.

41.1.1.2 Repeater unit

A repeater unit is a subset of a repeater set containing all the repeater-specific components and functions, exclusive of PHY components and functions. A repeater unit connects to the PHYs using the Gigabit Media Independent Interface (GMII) defined in Clause 35.

41.1.2 Application perspective

This subclause states the broad objectives and assumptions underlying the specification defined through Clause 41.

41.1.2.1 Objectives

a) Provide physical means for coupling two or more LAN segments at the Physical Layer.
b) Support interoperability of independently developed physical, electrical, and optical interfaces.
c) Provide a communication channel with a mean bit error rate, at the physical service interface equivalent to that for the attached PHY.
d) Provide for ease of installation and service.
e) Ensure that fairness of DTE access is not compromised.
f) Provide for low-cost networks, as related to both equipment and cabling.

41.1.2.2 Compatibility considerations

All implementations of the repeater set shall be compatible at the MDI. The repeater set is defined to provide compatibility among devices designed by different manufacturers. Designers are free to implement circuitry within the repeater set in an application-dependent manner provided the appropriate PHY specifications are met.

41.1.2.2.1 Internal segment compatibility

Implementations of the repeater set that contain a MAC layer for network management or other purposes, irrespective of whether they are connected through an exposed repeater port or are internally ported, shall conform to the requirements of Clause 30 on that port if repeater management is implemented.

41.1.3 Relationship to PHY

A close relationship exists between Clause 41 and the GMII clause (Clause 35) and the PHY clauses (Clauses 36–39 for 1000BASE-X PHYs and Clause 40 for 1000BASE-T PHYs). The PHY's PMA, PCS, and MDI specification provide the actual medium attachment, including drivers, receivers, and Medium Interface Connectors for the various supported media. The repeater clause does not define a new PHY; it utilizes the existing PHYs complete and without modification. The repeater_mode variable in each PHY is set, so that the CRS signal of the GMII is asserted only in response to receive activity (see 36.2.5.1.3).

41.2 Repeater functional specifications

A repeater set provides the means whereby data from any segment can be received under worst-case noise, timing, and amplitude conditions and then retransmitted with timing and amplitude restored to all other attached segments. Retransmission of data occurs simultaneously with reception. If a collision occurs, the repeater set propagates the collision event throughout the network by transmitting a Jam signal. If an error is received by the repeater set, no attempt is made to correct it and it is propagated throughout the network by transmitting an explicit error code.

The repeater set provides the following functional capability to handle data flow between ports:

a) *Signal restoration*. Provides the ability to restore the timing and amplitude of the received signal prior to retransmission.
b) *Transmit function*. Provides the ability to output signals on the appropriate port and encoded appropriately for that port. Details of signal processing are described in the specifications for the PHYs.
c) *Receive function*. Provides the ability to receive input signals presented to the ports. Details of signal processing are described in the specifications for the PHYs.
d) *Data-Handling function*. Provides the ability to transfer code-elements between ports in the absence of a collision.
e) *Received Event-Handling requirement*. Provides the ability to derive a carrier signal from the input signals presented to the ports.
f) *Collision-Handling function*. Provides the ability to detect the simultaneous reception of frames at two or more ports and then to propagate a Jam message to all connected ports.
g) *Error-Handling function*. Provides the ability to prevent substandard links from generating streams of false carrier and interfering with other links.
h) *Partition function*. Provides the ability to prevent a malfunctioning port from generating an excessive number of consecutive collisions and indefinitely disrupting data transmission on the network.
i) *Receive Jabber function*. Provides the ability to interrupt the reception of abnormally long streams of input data.

41.2.1 Repeater functions

The repeater set shall provide the Signal Restoration, Transmit, Receive, Data Handling, Received Event Handling, Collision Handling, Error Handling, Partition, and Receive Jabber functions. The repeater is transparent to all network acquisition activity and to all DTEs. The repeater will not alter the basic fairness criterion for all DTEs to access the network or weigh it toward any DTE or group of DTEs regardless of network location.

The Transmit and Receive functional requirements are specified by the PHY clauses, Clause 40 for 1000BASE-T and Clauses 36 to 39 for 1000BASE-X.

41.2.1.1 Signal restoration functional requirements

41.2.1.1.1 Signal amplification

The repeater set (including its integral PHYs) shall ensure that the amplitude characteristics of the signals at the MDI outputs of the repeater set are within the tolerances of the specification for the appropriate PHY type. Therefore, any loss of signal-to-noise ratio due to cable loss and noise pickup is regained at the output of the repeater set as long as the incoming data is within system specification.

41.2.1.1.2 Signal wave-shape restoration

The repeater set (including its integral PHYs) shall ensure that the wave-shape characteristics of the signals at the MDI outputs of a repeater set are within the specified tolerance for the appropriate PHY type. Therefore, any loss of wave-shape due to PHYs and media distortion is restored at the output of the repeater set.

41.2.1.1.3 Signal retiming

The repeater set (including its integral PHYs) shall ensure that the timing of the encoded data output at the MDI outputs of a repeater set are within the specified tolerance for the appropriate PHY type. Therefore, any receive jitter from the media is removed at the output of the repeater set.

41.2.1.2 Data-handling functional requirements

41.2.1.2.1 Data frame forwarding

The repeater set shall ensure that the data frame received on a single input port is distributed to all other output ports in a manner appropriate for the PHY type of that port. The data frame is that portion of the packet after the SFD and before the end-of-frame delimiter. The only exceptions to this rule are when contention exists among any of the ports, when the receive port is partitioned as defined in 41.2.1.6, when the receive port is in the Jabber state as defined in 41.2.1.7, or when the receive port is in the Link Unstable state as defined in 41.2.1.5.1. Between unpartitioned ports, the rules for collision handling (see 41.2.1.4) take precedence.

41.2.1.2.2 Received code violations

The repeater set shall ensure that any code violations received while forwarding a packet are propagated to all outgoing segments. These code violations shall be replaced by a code-group that provide an explicit indication that an error was received, as appropriate for the outgoing PHY type. Once a received code violation has been replaced by a code-group indicating a receive error, this substitution shall continue for the remainder of the received event regardless of its content. The only exception to this rule is when contention exists among any of the ports, where the rules for collision handling (see 41.2.1.4) then take precedence.

41.2.1.3 Received event-handling functional requirements

41.2.1.3.1 Received event handling

For all its ports, the repeater set shall detect received events by monitoring the port for any assertion of the GMII CRS signal that is the result of receive activity. The repeater_mode variable in the PHY shall be set to ensure that the CRS signal is not asserted in response to transmit activity. Received events include both the data frame and any encapsulation of the data frame such as Preamble, SFD, start and end of packet delimiters, carrier extension symbols, and error propagation symbols. A received event is exclusive of the IDLE pattern. Upon detection of a received event from one port, the repeater set shall repeat all received signals in the data frame from that port to the other ports as described in Figure 41–2.

41.2.1.3.2 Preamble regeneration

The repeater set shall output preamble as appropriate for the outgoing PHY type followed by the SFD. The duration of the output preamble shall not vary more than 8 bit times from the duration of the received preamble.

41.2.1.3.3 Start-of-packet propagation delay

The start-of-packet propagation delay for a repeater set is the time delay between the start of a received event on a repeated-from (input) port to the start of transmit on the repeated-to (output) port (or ports). This parameter is referred to as the SOP delay, and is measured at the MDI of the repeater ports. The maximum value of this delay is constrained such that the sum of the SOP delay and SOJ delay shall not exceed the value specified in 41.2.1.4.3.

41.2.1.3.4 Start-of-packet variability

The start-of-packet variability for a repeater set is defined as the total worst-case difference between start-of-packet propagation delays for successive received events separated by 112 bit times or less at the same input port. The variability shall be less than or equal to 16 bit times.

41.2.1.4 Collision-handling functional requirements

41.2.1.4.1 Collision detection

The repeater performs collision detection by monitoring all its enabled input ports for received events. When the repeater detects received events on more than one input port, it shall enter a collision state and transmit the Jam message to all of its output ports.

41.2.1.4.2 Jam generation

While a collision is occurring between any of its ports, the repeater unit shall transmit the Jam message to all of the ports. The Jam message shall be transmitted in accordance with the repeater state diagram in Figure 41–2. The Jam message is signalled across the GMII using the Transmit Error Propagation encoding if the collision is detected during Normal Data Transmission, or using the Carrier Extend Error encoding if the collision is detected during Carrier Extension.

41.2.1.4.3 Start-of-collision-jam propagation delay

The start-of-collision Jam propagation delay for a repeater set is the time delay between the start of the second received event (that results in a collision) to arrive at its port and the start of Jam out on all ports. This parameter is referred to as the SOJ delay, and is measured at the MDI of the repeater ports. The sum of the SOP delay and SOJ delay shall not exceed 976 bit times (BT).

41.2.1.4.4 Cessation-of-collision Jam propagation delay

The cessation-of-collision Jam propagation delay for a repeater set is the time delay between the end of the received event that creates a state such that Jam should end at a port and the end of Jam at that port. The states of the input signals that should cause Jam to end are covered in detail in the repeater state diagram in Figure 41–2. This parameter is referred to as the EOJ delay. This delay shall not exceed the SOP delay.

41.2.1.5 Error-handling functional requirements

41.2.1.5.1 Carrier integrity functional requirements

It is desirable that the repeater set protect the network from some transient fault conditions that would disrupt network communications. Potential likely causes of such conditions are DTE and repeater power-up and power-down transients, cable disconnects, and faulty wiring.

The repeater unit shall provide a self-interrupt capability at each port, as described in Figure 41–5, to prevent a segment's spurious carrier activity from propagating through the network.

At each port the repeater shall count consecutive false carrier events signalled across the GMII. The count shall be incremented on each false carrier event and shall be reset on reception of a valid carrier event. In addition, each port shall have a false carrier timer, which is enabled at the beginning of a false carrier event and reset at the conclusion of such an event. A repeater unit shall transmit the Jam signals to all ports for the duration of the false carrier event or until the duration of the event exceeds the time specified by the false_carrier_timer (see 41.2.2.1.4), whichever is shorter. The Jam message shall be transmitted in accordance with the repeater state diagram in Figure 41–2. The LINK UNSTABLE condition shall be detected when the False Carrier Event Count equals the value FCELimit (see 41.2.2.1.1) or the duration of a false carrier event exceeds the time specified by the false_carrier_timer. In addition, the LINK UNSTABLE condition shall be detected upon power-up reset.

Upon detection of LINK UNSTABLE at a port, the repeater unit shall perform the following:

a) Inhibit sending further messages from that port to the repeater unit.
b) Inhibit sending further output messages to that port from the repeater unit.
c) Continue to monitor activity on that port.

The repeater unit shall exit the LINK UNSTABLE condition at the port when one of the following is met:

a) The repeater has detected no activity (Idle) for more than the time specified by ipg_timer plus idle_timer (see 41.2.2.1.4) on port X.
b) A valid carrier event with a duration greater than the time specified by valid_carrier_timer (see 41.2.2.1.4) has been received, preceded by no activity (Idle) for more than the time specified by ipg_timer (see 41.2.2.1.4) on port X.

The false_carrier_timer duration is longer than the maximum round-trip latency from a repeater to a DTE, but less than a slot time. This allows a properly functioning DTE to respond to the Jam message by detecting collision and terminating the transmission prior to the expiration of the timer. The upper limit on the false_carrier_timer prevents the Jam message from exceeding the maximum fragment size.

The combination of the ipg_timer, idle_timer, and valid_carrier_timer filter transient activity that can occur on a link during power cycles or mechanical connection. The duration of the ipg_timer is greater than two-thirds of the minimum IPG, and less than the minimum IPG less some shrinkage. The idle_timer is specified as approximately 320 μs based upon empirical data on such transients. The valid_carrier_timer duration is less than the duration of a minimum valid carrier event, but long enough to filter most spurious carrier events (note that there can be no valid collision fragments on an isolated link in a single repeater topology). The range of the valid_carrier_timer is specified to be the same as the false_carrier_timer range for the convenience of implementations.

41.2.1.5.2 Speed handling

If the PHY has the capability of detecting speeds other than 1000 Mb/s, then the repeater set shall have the capability of blocking the flow of non-1000 Mb/s signals. The incorporation of 1000 Mb/s and 100 Mb/s or 10 Mb/s repeater functionality within a single repeater set is beyond the scope of this standard.

41.2.1.6 Partition functional requirements

It is desirable that the repeater set protect the network from some fault conditions that would disrupt network communications. A potentially likely cause of this condition could be due to a cable fault.

The repeater unit shall provide a self-interrupt capability at each port, as described in Figure 41–4, to prevent a faulty segment's carrier activity from propagating through the network. The repeater unit shall count consecutive collision events at each port. The count shall be incremented on each transmission that suffers a collision and shall be reset on a successful transmission or reception. If this count equals or exceeds the value CELimit (see 41.2.2.1.1), the Partition condition shall be detected. In addition, the partition condition shall be detected due to a carrier event of duration in excess of jabber_timer in which a collision has occurred.

Upon detection of Partition at a port, the repeater unit shall perform the following:

a) Inhibit sending further input messages from that port to the repeater unit.
b) Continue to output messages to that port from the repeater unit.
c) Continue to monitor activity on that port.

The repeater unit shall reset the Partition function at the port when one of the following conditions is met:

— On power-up reset.

— The repeater has detected activity on the port for more than the number of bits specified for no_collision_timer (see 41.2.2.1.4) without incurring a collision.
— The repeater has transmitted on the port for more than the number of bits specified for no_collision_timer (see 41.2.2.1.4) without incurring a collision.

The no_collision_timer duration is longer than the maximum round-trip latency from a repeater to a DTE (maximum time required for a repeater to detect a collision), and less than the minimum valid carrier event duration (slot time plus header_size minus preamble shrinkage).

41.2.1.7 Receive jabber functional requirements

The repeater unit shall provide a self-interrupt capability at each port, as described in Figure 41–3, to prevent an illegally long reception of data from propagating through the network. The repeater unit shall provide a window of duration jabber_timer bit times (see 41.2.2.1.4) during which the input messages from a port may be passed on to other repeater unit functions. If a reception exceeds this duration, the jabber condition shall be detected.

Upon detection of the jabber condition at a port, the repeater unit shall perform the following:

a) Inhibit sending further input messages from that port to the repeater unit.
b) Inhibit sending further output messages to that port from the repeater unit.

The repeater shall reset the Jabber function at the port, and re-enable data transmission and reception, when either one of the following conditions is met:

— On power-up reset.
— When carrier is no longer detected at that port.

The lower bound of the jabber_timer is longer than the carrier event of a maximum length burst. The upper bound is large enough to permit a wide variety of implementations.

41.2.2 Detailed repeater functions and state diagrams

A precise algorithmic definition is given in this subclause, providing a complete procedural model for the operation of a repeater, in the form of state diagrams. Note that whenever there is any apparent ambiguity concerning the definition of repeater operation, the state diagrams should be consulted for the definitive statement.

The model presented in this subclause is intended as a primary specification of the functions to be provided by any repeater unit. It is important to distinguish, however, between the model and a real implementation. The model is optimized for simplicity and clarity of presentation, while any realistic implementation should place heavier emphasis on such constraints as efficiency and suitability to a particular implementation technology.

It is the functional behavior of any repeater set implementation that is expected to match the standard, not the internal structure. The internal details of the procedural model are useful only to the extent that they help specify the external behavior clearly and precisely. For example, the model uses a separate Receive Port Jabber state diagram for each port. However, in actual implementation, the hardware may be shared.

The notation used in the state diagram follows the conventions of 1.2.1. Note that transitions shown without source states are evaluated at the completion of every state and take precedence over other transition conditions.

41.2.2.1 State diagram variables

41.2.2.1.1 Constants

CELimit

The number of consecutive Collision Events that must occur before a segment is partitioned.

Values: Positive integer greater than 60.

FCELimit

The number of consecutive False Carrier Events that must occur before a segment is isolated.

Value: 2.

41.2.2.1.2 Variables

begin

The Interprocess flag controlling state diagram initialization values.

Values: true
false

CRS(X), RXD(X), RX_DV(X), RX_ER(X), TXD(X), TX_EN(X), TX_ER(X)

GMII signals received from or sent to the PHY at port X (see Clause 35). The repeater_mode variable in the PHY is set to ensure that the CRS(X) signal is asserted in response to receive activity only.

RXERROR(X)

A combination of the GMII signal encodings indicating that the PHY has detected a Data Error, Carrier Extend Error, or False Carrier Error.

Value: RXERROR(X) ⇐ ((RX_ER(X) = true) *
((RX_DV(X) = true) + (RXD(X) = FalseCarrier) + (RXD(X) = CarrierExtendError)))

TX(X)

A combination of the GMII signal encodings indicating that port X is transmitting a frame.

Value: TX(X) ⇐ ((TX_EN(X) = true) + (TX_ER(X) = true))

isolate(X)

Flag from Carrier Integrity state diagram for port X, which determines whether a port should be enabled or disabled.

Values: true; the Carrier Integrity Monitor has determined the port should be disabled.
false; the Carrier Integrity Monitor has determined the port should be enabled.

force_jam(X)

Flag from Carrier Integrity state diagram for port X, which causes the Repeater Unit to enter the Jam state.

Values: true; the port is in the False Carrier state.
false; the port is not in the False Carrier state.

jabber(X)

Flag from Receive Timer state diagram for port X which indicates that the port has received excessive length activity.

Values: true; port has exceeded the continuous activity limit.
false; port has not exceeded the continuous activity limit.

link_status(X)

Indication from the Auto-Negotiation process (Clauses 28 and 37) that Auto-Negotiation has completed and the priority resolution function has determined that the link will be operated in half duplex mode.

Values: OK; the link is operational in half duplex mode.
FAIL; the link is not operational in half duplex mode.

partition(X)

Flag from Partition state diagram for port X, which determines whether a port receive path should be enabled or disabled.

Values: true; port has exceeded the consecutive collision limit.
false; port has not exceeded the consecutive collision limit.

41.2.2.1.3 Functions

port(Test)

A function that returns the designation of a port passing the test condition. For example, port(CRS = true) returns the designation: X for a port for which CRS is asserted. If multiple ports meet the test condition, the Port function will be assigned one and only one of the acceptable values.

41.2.2.1.4 Timers

All timers operate in the same fashion. A timer is reset and starts timing upon entering a state where "start x_timer" is asserted. At time "x" after the timer has been started, "x_timer_done" is asserted and remains asserted until the timer is reset. At all other times, "x_timer_not_done" is asserted.

When entering a state where "start x_timer" is asserted, the timer is reset and restarted even if the entered state is the same as the exited state.

The timers used in the repeater state diagrams are defined as follows:

false_carrier_timer

Timer for length of false carrier (41.2.1.5.1) that must be present to set isolate(X) to true. The timer is done when it reaches 3600–4000 BT.

idle_timer

Timer for length of time without carrier activity that must be present to set isolate(X) to false. The timer is done when it reaches 240 000–400 000 BT.

ipg_timer

Timer for length of time without carrier activity that must be present before carrier integrity tests (41.2.1.5.1) are re-enabled. The timer is done when it reaches 64–86 BT.

jabber_timer

Timer for length of carrier which must be present before the Jabber state is entered (41.2.1.7). The timer is done when it reaches 80 000–150 000 BT.

no_collision_timer

Timer for length of packet without collision before partition(X) is set to false (41.2.1.6). The timer is done when it reaches 3600–4144 BT.

valid_carrier_timer

Timer for length of valid carrier that must be present to cause isolate(X) to be set to false at the end of the carrier event. The timer is done when it reaches 3600–4000 BT.

41.2.2.1.5 Counters

CE(X)

Consecutive port Collision Event count for port X. Partitioning occurs on a terminal count of CELimit being reached.

Values: Non-negative integers up to a terminal count of CELimit.

FCE(X)

False Carrier Event count for port X. Isolation occurs on a terminal count of FCELimit being reached.

Values: Non-negative integers up to a terminal count of FCELimit.

41.2.2.1.6 Port designation

Ports are referred to by number. Port information is obtained by replacing the X in the desired function with the number of the port of interest. Ports are referred to in general as follows:

X

Generic port designator. When X is used in a state diagram, its value is local to that diagram and not global to the set of state diagrams.

N

Identifies the port that caused the exit from the IDLE or JAM states of Figure 41–2. The value is assigned in the term assignment statement on the transition out of these states (see 1.2.1 for State Diagram Conventions).

ALL

Indicates all repeater ports are to be considered. The test passes when all ports meet the test conditions.

ALLXJIPN

The test passes when all ports, excluding those indicated by J, I, P, or N, meet the test conditions. One or more of the J, I, P, or N indications are used to exclude from the test ports with Jabber = true, Isolate = true, Partition = true, or port N, respectively.

ANY

Indicates all ports are to be considered. The test passes when one or more ports meet the test conditions.

ANYXJIPN

The test passes when one or more ports, excluding those indicated by J, I, P, or N, meet the test conditions. One or more of the J, I, P, or N indications are used to exclude from the test ports with Jabber = true, Isolate = true, Partition = true, or port N, respectively.

ONLY1

Indicates all ports except those with Jabber = true, Isolate = true, or Partition = true are to be considered. The test passes when one and only one port meet the test conditions.

41.2.2.2 State diagrams

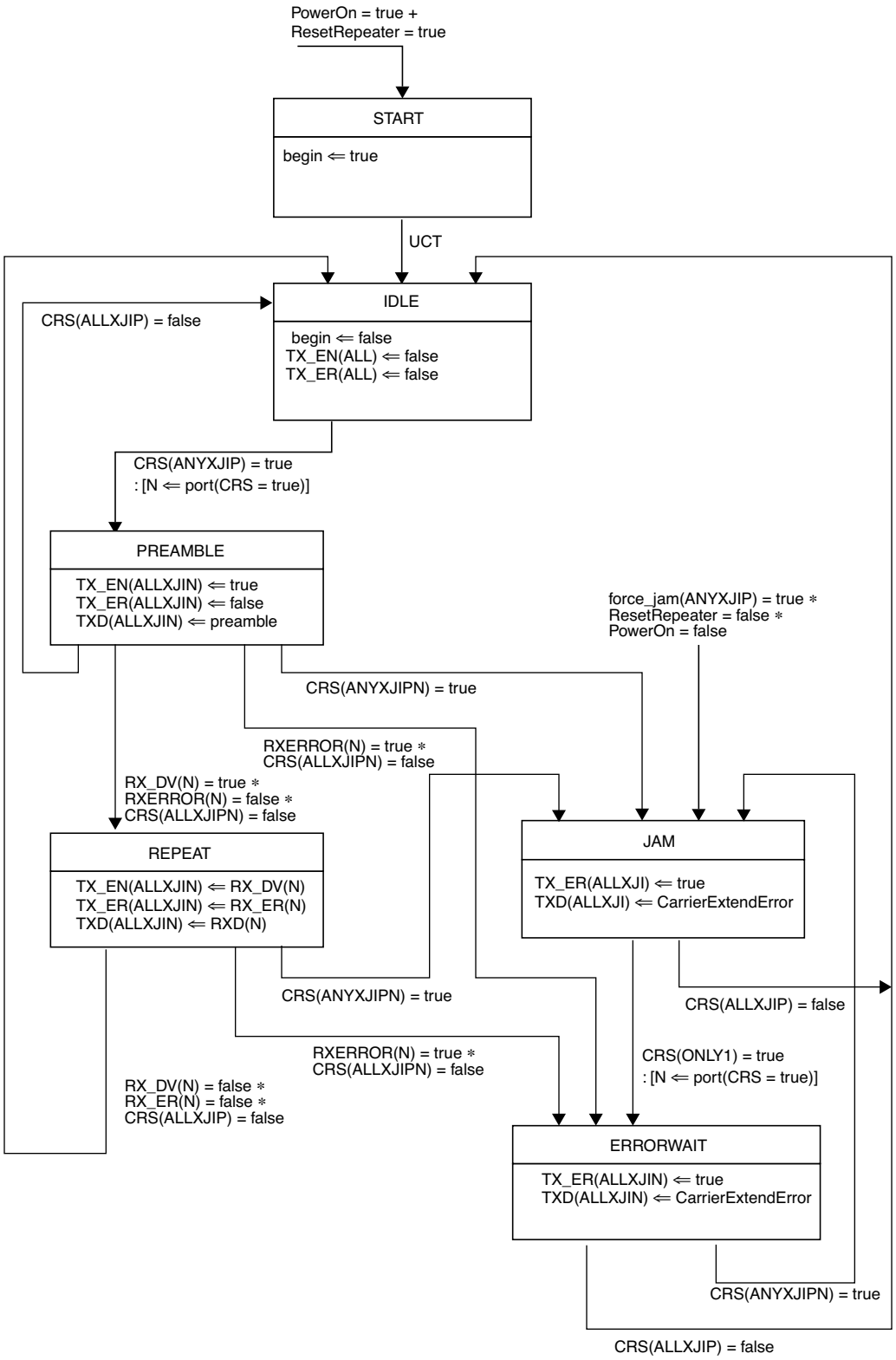


Figure 41–2—Repeater unit state diagram

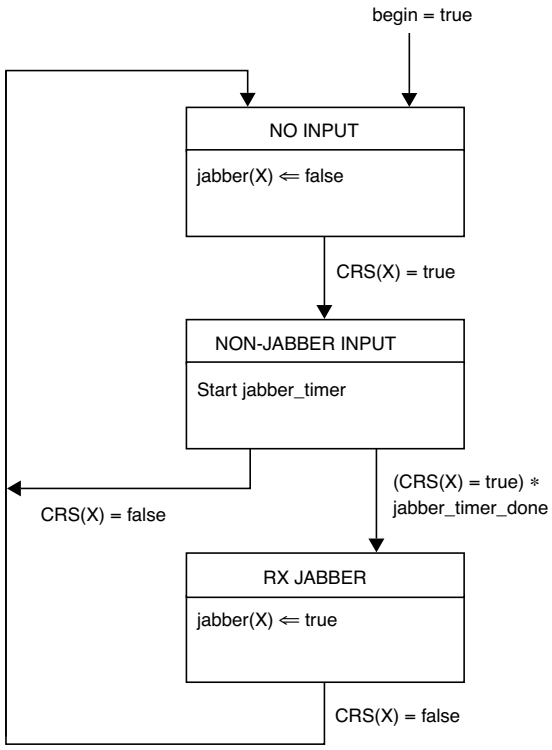


Figure 41–3—Receive timer state diagram for port X

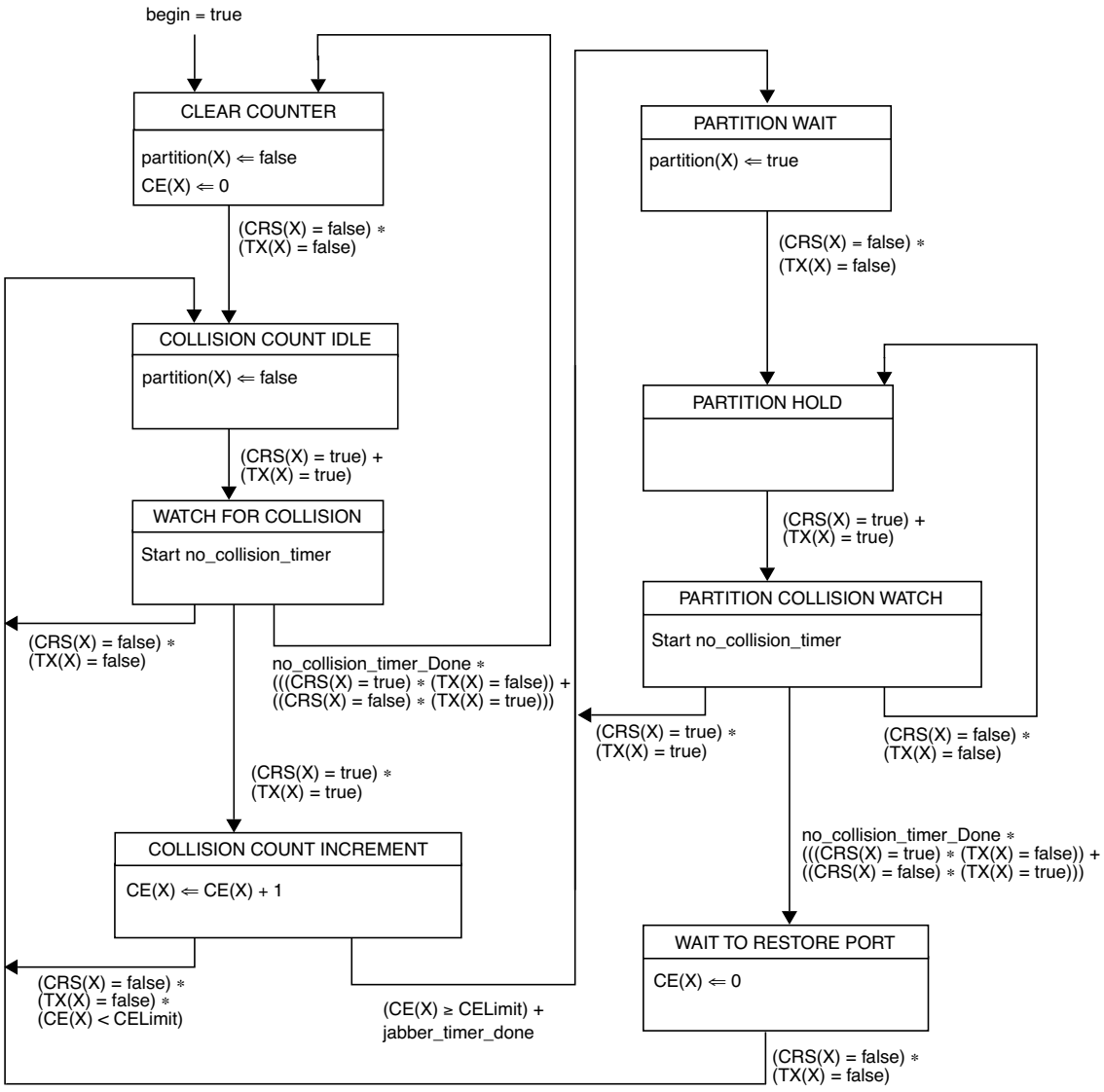


Figure 41–4—Partition state diagram for port X

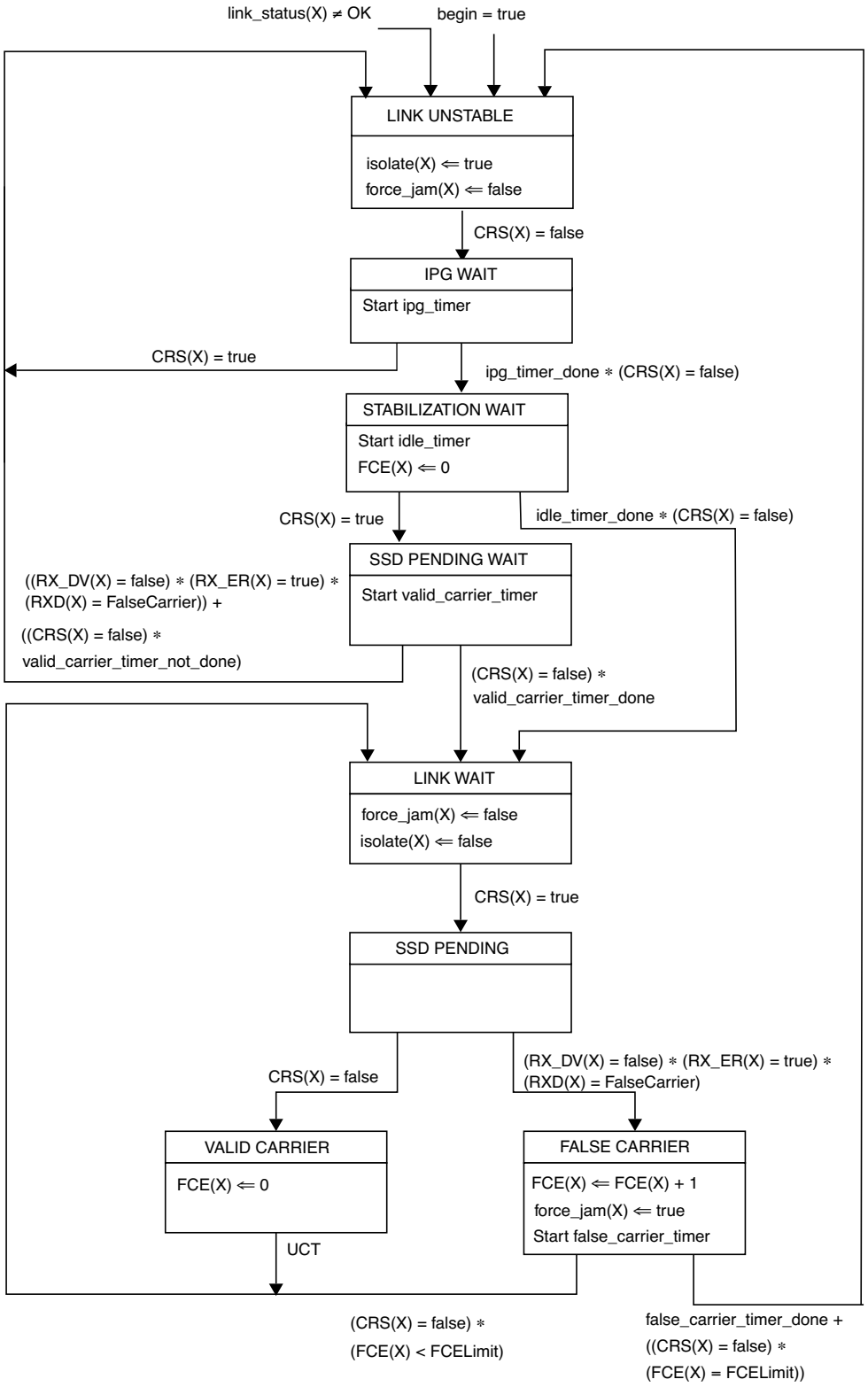


Figure 41–5—Carrier integrity monitor state diagram for port X

41.3 Repeater electrical specifications

41.3.1 Electrical isolation

Network segments that have different isolation and grounding requirements shall have those requirements provided by the port-to-port isolation of the repeater set.

41.4 Environmental specifications

41.4.1 General safety

All equipment meeting this standard shall conform to IEC 60950: 1991.

41.4.2 Network safety

This subclause sets forth a number of recommendations and guidelines related to safety concerns; the list is neither complete nor does it address all possible safety issues. The designer is urged to consult the relevant local, national, and international safety regulations to ensure compliance with the appropriate requirements.

LAN cable systems described in this subclause are subject to at least four direct electrical safety hazards during their installation and use. These hazards are as follows:

a) Direct contact between LAN components and power, lighting, or communications circuits.
b) Static charge buildup on LAN cables and components.
c) High-energy transients coupled onto the LAN cable system.
d) Voltage potential differences between safety grounds to which the various LAN components are connected.

Such electrical safety hazards must be avoided or appropriately protected against for proper network installation and performance. In addition to provisions for proper handling of these conditions in an operational system, special measures must be taken to ensure that the intended safety features are not negated during installation of a new network or during modification or maintenance of an existing network. Isolation requirements are defined in 41.4.3.

41.4.2.1 Installation

Sound installation practice, as defined by applicable local codes and regulations, shall be followed in every instance in which such practice is applicable.

41.4.2.2 Grounding

The safety ground, or chassis ground for the repeater set, shall be provided through the main ac power cord via the third wire ground as defined by applicable local codes and regulations.

If the MDI connector should provide a shield connection, the shield may be connected to the repeater safety ground. A network segment connected to the repeater set through the MDI may use a shield. If both ends of the network segment have a shielded MDI connector available, then the shield may be grounded at both ends according to local regulations and ISO/IEC 11801: 1995, and as long as the ground potential difference between both ends of the network segment is less than 1 V rms.

WARNING

It is assumed that the equipment to which the repeater is attached is properly grounded and not left floating nor serviced by a "doubly insulated ac power distribution system." The use of floating or insulated equipment, and the consequent implications for safety, are beyond the scope of this standard.

41.4.2.3 Installation and maintenance guidelines

During installation and maintenance of the cable plant, care should be taken to ensure that uninsulated network cable connectors do not make electrical contact with unintended conductors or ground.

41.4.3 Electrical isolation

There are two electrical power distribution environments to be considered that require different electrical isolation properties:

a) *Environment A*. When a LAN or LAN segment, with all its associated interconnected equipment, is entirely contained within a single low-voltage power distribution system and within a single building.
b) *Environment B*. When a LAN crosses the boundary between separate power distribution systems or the boundary of a single building.

41.4.3.1 Environment A requirements

Attachment of network segments via repeater sets requires electrical isolation of 500 V rms, one-minute withstand, between the segment and the protective ground of the repeater unit.

41.4.3.2 Environment B requirements

The attachment of network segments that cross environment B boundaries requires electrical isolation of 1500 Vrms, one-minute withstand, between each segment and all other attached segments and also the protective ground of the repeater unit.

The requirements for interconnected electrically conducting LAN segments that are partially or fully external to a single building environment may require additional protection against lightning strike hazards. Such requirements are beyond the scope of this standard. It is recommended that the above situation be handled by the use of nonelectrically conducting segments (e.g., fiber optic).

It is assumed that any nonelectrically conducting segments will provide sufficient isolation within that media to satisfy the isolation requirements of environment B.

41.4.4 Reliability

A two-port repeater set shall be designed to provide a mean time between failure (MTBF) of at least 50 000 hours of continuous operation without causing a communications failure among stations attached to the network medium. Repeater sets with more than two ports shall add no more than 3.46×10^{-6} failures per hour for each additional port.

The repeater set electronics should be designed to minimize the probability of component failures within the repeater electronics that prevent communications among other PHYs on the individual segments. Connectors and other passive components comprising the means of connecting the repeater to the cable should be designed to minimize the probability of total network failure.

41.4.5 Environment

41.4.5.1 Electromagnetic emission

The repeater shall comply with applicable local and national codes for the limitation of electromagnetic interference.

41.4.5.2 Temperature and humidity

The repeater is expected to operate over a reasonable range of environmental conditions related to temperature, humidity, and physical handling (such as shock and vibration). Specific requirements and values for these parameters are considered to be beyond the scope of this standard.

It is recommended that manufacturers indicate in the literature associated with the repeater the operating environmental conditions to facilitate selection, installation, and maintenance.

41.5 Repeater labeling

It is required that each repeater (and supporting documentation) shall be labeled in a manner visible to the user with these parameters:

a) Crossover ports appropriate to the respective PHY shall be marked with an X.

Additionally it is recommended that each repeater (and supporting documentation) also be labeled in a manner visible to the user with at least these parameters:

b) Data rate capability in Mb/s
c) Any applicable safety warnings
d) Port type, i.e., 1000BASE-CX, 1000BASE-SX, 1000BASE-LX, and 1000BASE-T
e) Worst-case bit time delays between any two ports appropriate for
    1) Start-of-packet propagation delay
    2) Start-of-collision Jam propagation delay
    3) Cessation-of-collision Jam propagation delay

41.6 Protocol Implementation Conformance Statement (PICS) proforma for Clause 41, Repeater for 1000 Mb/s baseband networks[12]

41.6.1 Introduction

The supplier of a protocol implementation that is claimed to conform to Clause 41, Repeater for 1000 Mb/s baseband networks, shall complete the following Protocol Implementation Conformance Statement (PICS) proforma.

A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

[12]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this subclause so that it can be used for its intended purpose and may further publish the completed PICS.

41.6.2 Identification

41.6.2.1 Implementation identification

Supplier	
Contact point for enquiries about the PICS	
Implementation Name(s) and Version(s)	
Other information necessary for full identification—e.g., name(s) and version(s) for machines and/or operating systems; System Names(s)	
NOTE 1—Only the first three items are required for all implementations; other information may be completed as appropriate in meeting the requirements for the identification. NOTE 2—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

41.6.2.2 Protocol summary

Identification of protocol standard	IEEE Std 802.3-2002®, Clause 41, Repeater for 1000 Mb/s baseband networks
Identification of amendments and corrigenda to this PICS proforma that have been completed as part of this PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21; the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®.)	

Date of Statement	

41.6.3 Major capabilities/options

Item	Feature	Subclause	Value/Comment	Status	Support
*SXP	Repeater supports 1000BASE-SX connections	41.1.2.2		O	Yes [ ] No [ ]
*LXP	Repeater supports 1000BASE-LX connections	41.1.2.2		O	Yes [ ] No [ ]
*CXP	Repeater supports 1000BASE-CX connections	41.1.2.2		O	Yes [ ] No [ ]
*TP	Repeater supports 1000BASE-T connections	41.1.2.2		O	Yes [ ] No [ ]
*PHYS	PHYs capable of detecting non 1000 Mb/s signals	41.2.1.5.2		O	Yes [ ] No [ ]

In addition, the following predicate name is defined for use when different implementations from the set above have common parameters:

*XP:SXP or LXP or CXP

41.6.4 PICS proforma tables for the Repeater for 1000 Mb/s baseband networks

41.6.4.1 Compatibility considerations

Item	Feature	Subclause	Value/Comment	Status	Support
CC1	1000BASE-SX port compatible at the MDI	41.1.2.2		SXP:M	Yes [ ] N/A [ ]
CC2	1000BASE-LX port compatible at the MDI	41.1.2.2		LXP:M	Yes [ ] N/A [ ]
CC3	1000BASE-CX port compatible at the MDI	41.1.2.2		CXP:M	Yes [ ] N/A [ ]
CC4	1000BASE-T port compatible at the MDI	41.1.2.2		TP:M	Yes [ ] N/A [ ]
CC5	Internal segment compatibility	41.1.2.2.1	Internal port meets Clause 30 when repeater management implemented	M	Yes [ ]

41.6.4.2 Repeater functions

Item	Feature	Subclause	Value/Comment	Status	Support
RF1	Signal Restoration	41.2.1		M	Yes [ ]
RF2	Data Handling	41.2.1		M	Yes [ ]
RF3	Received Event Handling	41.2.1		M	Yes [ ]
RF4	Collision Handling	41.2.1		M	Yes [ ]
RF5	Error Handling	41.2.1		M	Yes [ ]
RF6	Partition	41.2.1		M	Yes [ ]
RF7	Received Jabber	41.2.1		M	Yes [ ]
RF8	Transmit	41.2.1		M	Yes [ ]
RF9	Receive	41.2.1		M	Yes [ ]

41.6.4.3 Signal restoration function

Item	Feature	Subclause	Value/Comment	Status	Support
SR1	Output amplitude as required by 1000BASE-SX	41.2.1.1.1		SXP:M	Yes [ ] N/A [ ]
SR2	Output amplitude as required by 1000BASE-LX	41.2.1.1.1		LXP:M	Yes [ ] N/A [ ]
SR3	Output amplitude as required by 1000BASE-CX	41.2.1.1.1		CXP:M	Yes [ ] N/A [ ]
SR4	Output amplitude as required by 1000BASE-T	41.2.1.1.1		TP:M	Yes [ ] N/A [ ]
SR5	Output signal wave-shape as required by 1000BASE-SX	41.2.1.1.2		SXP:M	Yes [ ] N/A [ ]
SR6	Output signal wave-shape as required by 1000BASE-LX	41.2.1.1.2		LXP:M	Yes [ ] N/A [ ]
SR7	Output signal wave-shape as required by 1000BASE-CX	41.2.1.1.2		CXP:M	Yes [ ] N/A [ ]
SR8	Output signal wave-shape as required by 1000BASE-T	41.2.1.1.2		TP:M	Yes [ ] N/A [ ]
SR9	Output data timing as required by 1000BASE-SX	41.2.1.1.3		SXP:M	Yes [ ] N/A [ ]
SR10	Output data timing as required by 1000BASE-LX	41.2.1.1.3		LXP:M	Yes [ ] N/A [ ]
SR11	Output data timing as required by 1000BASE-CX	41.2.1.1.3		CXP:M	Yes [ ] N/A [ ]
SR12	Output data timing as required by 1000BASE-T	41.2.1.1.3		TP:M	Yes [ ] N/A [ ]

41.6.4.4 Data-Handling function

Item	Feature	Subclause	Value/Comment	Status	Support
DH1	Data frames forwarded to all ports except receiving port	41.2.1.2.1		M	Yes [ ]
DH2	Code Violations forwarded to all transmitting ports	41.2.1.2.2		M	Yes [ ]
DH3	Received Code Violation forwarded as code-group explicitly indicating received error	41.2.1.2.2		M	Yes [ ]
DH4	Code element substitution for remainder of packet after received Code Violation	41.2.1.2.2		M	Yes [ ]

41.6.4.5 Receive Event-Handling function

Item	Feature	Subclause	Value/Comment	Status	Support
RE1	Detect all received events	41.2.1.3.1		M	Yes [ ]
RE2	Repeat all received signals	41.2.1.3.1		M	Yes [ ]
RE3	Preamble repeated as required	41.2.1.3.2		M	Yes [ ]
RE4	Start-of-packet propagation delay	41.2.1.3.3	SOP + SOJ ≤ 976 BT	M	Yes [ ]
RE5	Start-of-packet variability	41.2.1.3.4	SOP variation ≤ 16 BT	M	Yes [ ]
RE6	PHY repeater_mode variable	41.2.1.3.1	Shall be set to ensure CRS signal not asserted in response to transmit activity	M	Yes [ ]
RE7	Output preamble variation	41.2.1.3.2	Variation between received and transmitted preamble ≤ 8 BT	M	Yes [ ]

41.6.4.6 Collision-Handling function

Item	Feature	Subclause	Value/Comment	Status	Support
CO1	Collision Detection	41.2.1.4.1	Receive event on more than one port	M	Yes [ ]
CO2	Jam Generation	41.2.1.4.2	Transmit Jam message while collision is detected	M	Yes [ ]
CO3	Collision-Jam Propagation delay	41.2.1.4.3	SOP + SOJ ≤ 976 BT	M	Yes [ ]
CO4	Cessation of Collision Propagation delay	41.2.1.4.4	EOJ ≤ SOP	M	Yes [ ]

41.6.4.7 Error-Handling function

Item	Feature	Subclause	Value/Comment	Status	Support
EH1	Carrier Integrity function implementation	41.2.1.5.1	Self-interrupt of data reception	M	Yes [ ]
EH2	False Carrier Event count for Link Unstable detection	41.2.1.5.1	False Carrier Event count equals FCELimit	M	Yes [ ]
EH3	False carrier count reset	41.2.1.5.1	Count reset on valid carrier	M	Yes [ ]
EH4	False carrier timer for Link Unstable detection	41.2.1.5.1	False carrier of length in excess of false_carrier_timer	M	Yes [ ]
EH5	Jam message duration	41.2.1.5.1	Equals duration of false carrier event, but not greater than duration of false_carrier_timer	M	Yes [ ]
EH6	Link Unstable detection	41.2.1.5.1	False Carrier Event count equals FCELimit or False carrier exceeds the false_carrier_timer or power-up reset	M	Yes [ ]
EH7	Messages sent to repeater unit in Link Unstable state	41.2.1.5.1	Inhibited sending messages to repeater unit	M	Yes [ ]
EH8	Messages sent from repeater unit in Link Unstable state	41.2.1.5.1	Inhibited sending output messages	M	Yes [ ]
EH9	Monitoring activity on a port in Link Unstable state	41.2.1.5.1	Continue monitoring activity at that port	M	Yes [ ]
EH10	Reset of Link Unstable state	41.2.1.5.1	No activity for more than ipg_timer plus idle_timer or Valid carrier event of duration greater than valid_carrier_timer preceded by Idle of duration greater than ipg_timer	M	Yes [ ]
EH11	Block flow of non-1000 Mb/s signals	41.2.1.5.2		M	Yes [ ]

41.6.4.8 Partition function

Item	Feature	Subclause	Value/Comment	Status	Support
PA1	Partition function implementation	41.2.1.6	Self-interrupt of data reception	M	Yes [ ]
PA2	Consecutive Collision Event count for entry into partition state	41.2.1.6	Consecutive Collision Event count equals or exceeds CELimit	M	Yes [ ]
PA3	Excessive receive duration with collision for entry into partition state.	41.2.1.6	Reception duration in excess of jabber_timer with collision	M	Yes [ ]
PA4	Consecutive Collision Event counter incrementing	41.2.1.6	Count incremented on each transmission that suffers a collision	M	Yes [ ]
PA5	Consecutive Collision Event counter reset	41.2.1.6	Count reset on successful transmission or reception	M	Yes [ ]
PA6	Messages sent to repeater unit in Partition state	41.2.1.6	Inhibited sending messages to repeater unit	M	Yes [ ]
PA7	Messages sent from repeater unit in Partition state	41.2.1.6	Continue sending output messages	M	Yes [ ]
PA8	Monitoring activity on a port in Partition state	41.2.1.6	Continue monitoring activity at that port	M	Yes [ ]
PA9	Reset of Partition state	41.2.1.6	Power-up reset or transmitting or detecting activity for greater than duration no_collision_timer without a collision	M	Yes [ ]

41.6.4.9 Receive Jabber function

Item	Feature	Subclause	Value/Comment	Status	Support
RJ1	Receive Jabber function implementation	41.2.1.7	Self-interrupt of data reception	M	Yes [ ]
RJ2	Excessive receive duration timer for Receive Jabber detection	41.2.1.7	Reception duration in excess of jabber_timer	M	Yes [ ]
RJ3	Messages sent to repeater unit in Receive Jabber state	41.2.1.7	Inhibit sending input messages to repeater unit	M	Yes [ ]
RJ4	Messages sent from repeater unit in Receive Jabber state	41.2.1.7	Inhibit sending output messages	M	Yes [ ]
RJ5	Reset of Receive Jabber state	41.2.1.7	Power-up reset or Carrier no longer detected	M	Yes [ ]

41.6.4.10 Repeater state diagrams

Item	Feature	Subclause	Value/Comment	Status	Support
SD1	Repeater unit state diagram	41.2.2.2	Meets the requirements of Figure 41–2	M	Yes [ ]
SD2	Receive timer for port X state diagram	41.2.2.2	Meets the requirements of Figure 41–3	M	Yes [ ]
SD3	Repeater partition state diagram for port X	41.2.2.2	Meets the requirements of Figure 41–4	M	Yes [ ]
SD4	Carrier integrity monitor for port X state diagram	41.2.2.2	Meets the requirements of Figure 41–5	M	Yes [ ]

41.6.4.11 Repeater electrical

Item	Feature	Subclause	Value/Comment	Status	Support
EL1	Port-to-port isolation	41.3.1	Satisfies isolation and grounding requirements for attached network segments	M	Yes [ ]
EL2	Safety	41.4.1	IEC 60950: 1991	M	Yes [ ]
EL3	Installation practices	41.4.2.1	Sound, as defined by local code and regulations	M	Yes [ ]
EL4	Grounding	41.4.2.2	Chassis ground provided through ac mains cord	M	Yes [ ]
EL5	Two-port repeater set MTBF	41.4.4	At least 50 000 hours	M	Yes [ ]
EL6	Additional port effect on MTBF	41.4.4	No more than 3.46×10^{-6} increase in failures per hour	M	Yes [ ]
EL7	Electromagnetic interference	41.4.5.1	Comply with local or national codes	M	Yes [ ]

41.6.4.12 Repeater labeling

Item	Feature	Subclause	Value/Comment	Status	Support
LB1	Crossover ports	41.5	Marked with an X	M	Yes [ ]
LB2	Data Rate	41.5	1000 Mb/s	O	Yes [ ] No [ ]
LB3	Safety warnings	41.5	Any applicable	O	Yes [ ] No [ ]
LB4	Port Types	41.5	1000BASE-SX, 1000BASE-LX, 1000BASE-CX, or 1000BASE-T	O	Yes [ ] No [ ]
LB5	Worse-case start-of-packet propagation delay	41.5	Value in bit times (BT)	O	Yes [ ] No [ ]
LB6	Worse-case start-of-collision-Jam propagation delay	41.5	Value in BT	O	Yes [ ] No [ ]
LB7	Worse-case Cessation-of-Collision Jam propagation delay	41.5	Value in BT	O	Yes [ ] No [ ]

42. System considerations for multisegment 1000 Mb/s networks

42.1 Overview

This clause provides information on building 1000 Mb/s networks. The 1000 Mb/s technology is designed to be deployed in both homogenous 1000 Mb/s networks and 10/100/1000 Mb/s mixed networks using bridges and/or routers. Network topologies can be developed within a single 1000 Mb/s collision domain, but maximum flexibility is achieved by designing multiple collision domain networks that are joined by bridges and/or routers configured to provide a range of service levels to DTEs. For example, a combined 1000BASE-T/100BASE-T/10BASE-T system built with repeaters and bridges can deliver dedicated or shared service to DTEs at 1000 Mb/s, 100 Mb/s, or 10 Mb/s.

Linking multiple collision domains with bridges maximizes flexibility. Bridged topology designs can provide single data rate (Figure 42–1) or multiple data rate (Figure 42–2) services.

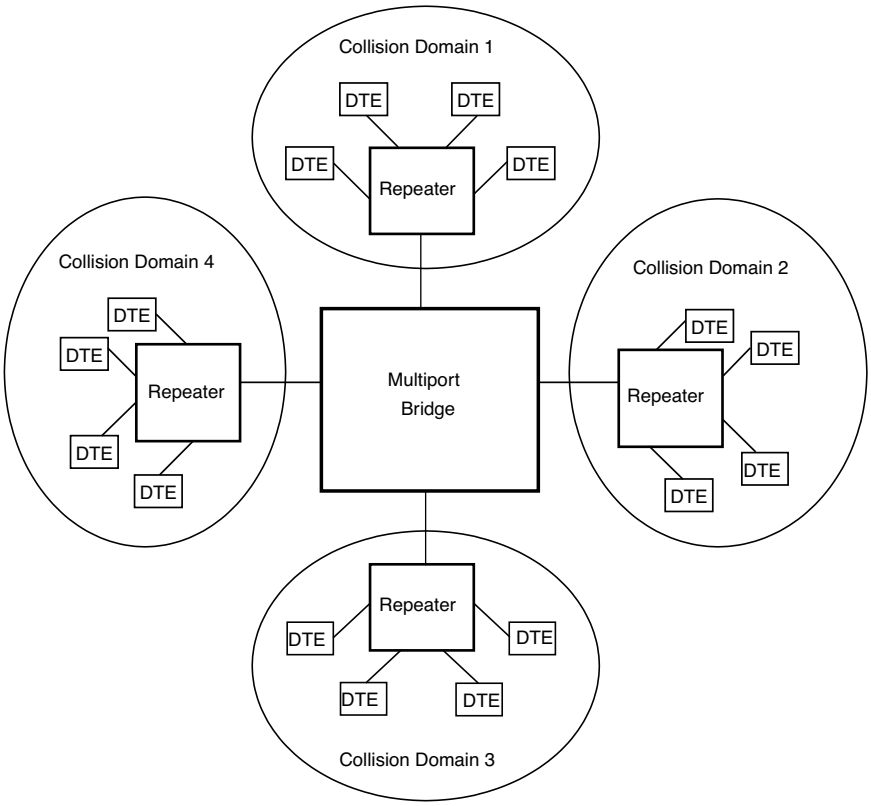


Figure 42–1—1000 Mb/s multiple collision domain topology using multiport bridge

Individual collision domains can be linked by single devices (as shown in Figure 42–1 and Figure 42–2) or by multiple devices from any of several transmission systems. The design of multiple-collision-domain networks is governed by the rules defining each of the transmission systems incorporated into the design.

The design of shared bandwidth 10 Mb/s collision domains is defined in Clause 13; the design of shared bandwidth 100 Mb/s CSMA/CD collision domains is defined in Clause 29; the design of shared bandwidth 1000 Mb/s CSMA/CD collision domains is defined in the following subclauses.

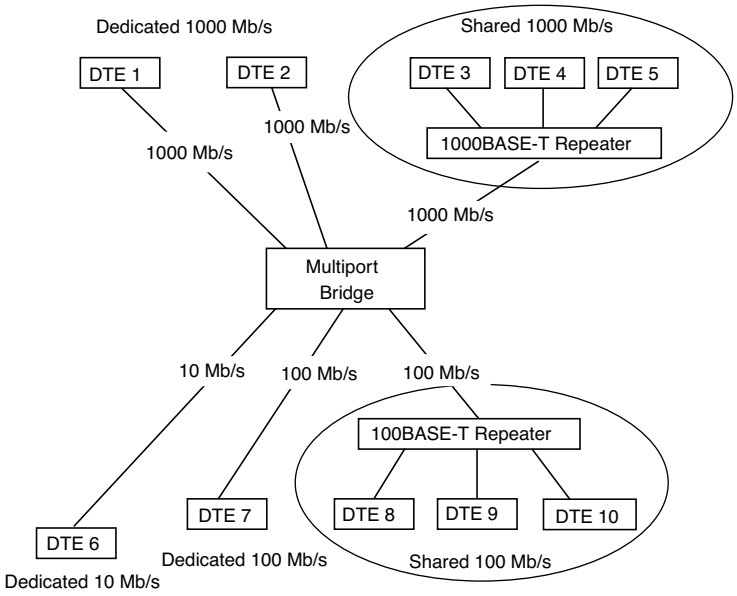


Figure 42–2—Multiple data rate, multiple collision domain topology using multiport bridge

42.1.1 Single collision domain multisegment networks

This clause provides information on building 1000 Mb/s CSMA/CD multisegment networks within a single collision domain. The proper operation of a CSMA/CD network requires the physical size of the collision domain to be limited in order to meet the round-trip propagation delay requirements of 4.2.3.2.3 and 4.4.2.1, and requires the number of repeaters to be limited to one so as not to exceed the InterFrameGap shrinkage noted in 4.4.2.4.

This clause provides two network models. Transmission System Model 1 is a set of configurations that have been validated under conservative rules and have been qualified as meeting the requirements set forth above. Transmission System Model 2 is a set of calculation aids that allow those configuring a network to test a proposed configuration against a simple set of criteria that allows it to be qualified. Transmission System Model 2 validates an additional broad set of topologies that are fully functional and do not fit within the simpler, but more restrictive rules of Model 1.

The physical size of a CSMA/CD network is limited by the characteristics of individual network components. These characteristics include the following:

a) Media lengths and their associated propagation time delay.
b) Delay of repeater units (start-up, steady-state, and end of event).
c) Delay of MAUs and PHYs (start-up, steady-state, and end of event).
d) Interpacket gap shrinkage due to repeater units.
e) Delays within the DTE associated with the CSMA/CD access method.
f) Collision detect and deassertion times associated with the MAUs and PHYs.

Table 42–1 summarizes the delays, measured in Bit Times (BTs), for 1000 Mb/s media segments.

Table 42–1—Delays for network media segments Model 1

Media type	Maximum number of PHYs per segment	Maximum segment length (m)	Maximum medium round-trip delay per segment (BT)
Category 5 UTP Link Segment (1000BASE-T)	2	100	1112
Shielded Jumper Cable Link Segment (1000BASE-CX)	2	25	253
Optical Fiber Link Segment (1000BASE-SX, 1000BASE-LX)	2	316[a]	3192

[a]May be limited by the maximum transmission distance of the link.

42.1.2 Repeater usage

Repeaters are the means used to connect segments of a network medium together into a single collision domain. Different physical signaling systems (e.g., 1000BASE-CX, 1000BASE-SX, 1000BASE-LX, 1000BASE-T) can be joined into a common collision domain using a repeater. Bridges can also be used to connect different signaling systems; however, if a bridge is so used, each LAN connected to the bridge will comprise a separate collision domain.

42.2 Transmission System Model 1

The following network topology constraints apply to networks using Transmission System Model 1.

a) Single repeater topology maximum.
b) Link distances not to exceed the lesser of 316 m or the segment lengths as shown in Table 42–1.

42.3 Transmission System Model 2

Transmission System Model 2 is a single repeater topology with the physical size limited primarily by round-trip collision delay. A network configuration must be validated against collision delay using a network model for a 1000 Mb/s collision domain. The modeling process is quite straightforward and can easily be done either manually or with a spreadsheet.

The model proposed here is derived from the one presented in 13.4. Modifications have been made to accommodate adjustments for DTE, repeater, and cable speeds.

For a network consisting of two 1000BASE-T DTEs as shown in Figure 42–3, a crossover connection may be required if the auto-crossover function is not implemented. See 40.4 and 40.8.

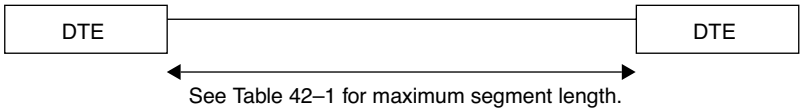


Figure 42–3—Model 1: Two DTEs, no repeater

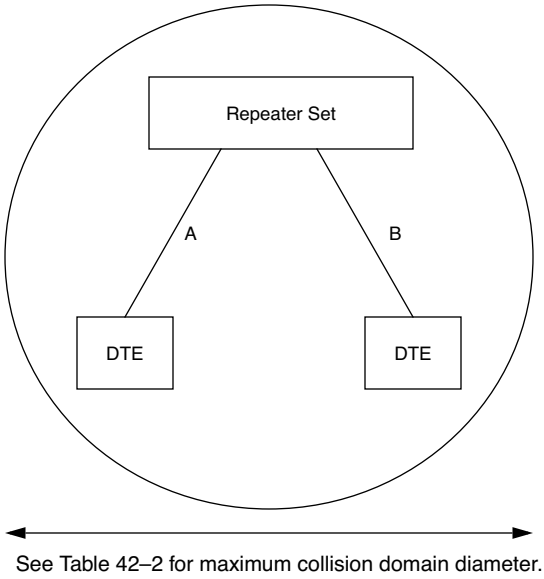


Figure 42–4—Model 1: Single repeater

Table 42–2—Maximum Model 1 collision domain diameter[a]

Configuration	Category 5 UTP (T)	Shielded Jumper Cable (CX)	Optical Fiber (SX/LX)	Mixed Category 5 and Fiber (T and SX/LX)	Mixed Shielded Jumper and Fiber (CX and SX/LX)
DTE-DTE (see Figure 42–3)	100	25	316[b]	NA	NA
One repeater (see Figure 42–4)	200	50	220	210[c]	220[d]

[a]In meters.
[b]May be limited by the maximum transmission distance of the link.
[c]Assumes 100 m of Category 5 UTP and one Optical Fiber link of 110 m.
[d]Assumes 25 m of Shielded Jumper Cable and one Optical Fiber link of 195 m.

42.3.1 Round-trip collision delay

For a network configuration to be valid, it must be possible for any two DTEs on the network to properly arbitrate for the network. When two or more stations attempt to transmit within a slot time interval, each station must be notified of the contention by the returned "collision" signal within the "collision window" (see 4.1.2.2). Additionally, the maximum length fragment created on a 1000 Mb/s network must contain less than 512 bytes after the Start Frame Delimiter (SFD). These requirements limit the physical diameter (maximum distance between DTEs) of a network. The maximum round-trip delay must be qualified between all pairs of DTEs in the network. In practice this means that the qualification must be done between those that, by inspection of the topology, are candidates for the longest delay. The following network modeling methodology is provided to assist that calculation.

42.3.1.1 Worst-case path delay value (PDV) selection

The worst-case path through a network to be validated is identified by examination of aggregate DTE delays, cable delays, and repeater delay. The worst case consists of the path between the two DTEs at opposite ends of the network that have the longest round-trip time. Figure 42–5 shows a schematic representation of a one-repeater path.

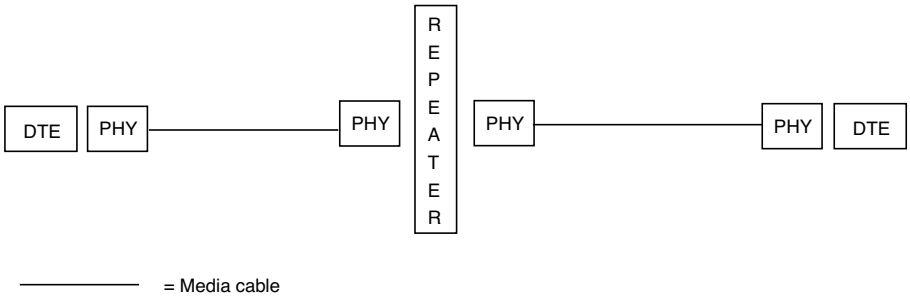


Figure 42–5——System Model 2: Single repeater

42.3.1.2 Worst-case PDV calculation

Once a set of paths is chosen for calculation, each is checked for validity against the following formula:

PDV = ∑link delays (LSDV) + repeater delay + DTE delays + safety margin

Values for the formula variables are determined by the following method:

a) Determine the delay for each link segment (Link Segment Delay Value, or LSDV), using the formula

LSDV=2 (for round-trip delay) × segment length × cable delay for this segment

NOTE 1—Length is the sum of the cable lengths between the PHY interfaces at the repeater and PHY interfaces at the farthest DTE. All measurements are in meters.

NOTE 2—Cable delay is the delay specified by the manufacturer or the maximum value for the type of cable used as shown in Table 42–3. For this calculation, cable delay must be specified in bit times per meter (BT/m). Table 42–4 can be used to convert values specified relative to the speed of light (%c) or nanoseconds per meter (ns/m).

NOTE 3—When actual cable lengths or propagation delays are not known, use the Max delay in bit times as specified in Table 42–3 for copper cables. Delays for fiber should be calculated, as the value found in Table 42–3 will be too large for most applications.

NOTE 4—The value found in Table 42–3 for Shielded Jumper Cable is the maximum delay for cable with solid dielectric. Cables with foam dielectric may have a significantly smaller delay.

b) Sum together the LSDVs for all segments in the path.
c) Determine the delay for the repeater. If model-specific data is not available from the manufacturer, enter the appropriate default value from Table 42–3.
d) Use the DTE delay value shown in Table 42–3 unless your equipment manufacturer defines a different value. If the manufacturer's supplied values are used, the DTE delays of both ends of the worst-case path should be summed together.
e) Decide on appropriate safety margin—0 to 40 bit times—for the PDV calculation. Safety margin is used to provide additional margin to accommodate unanticipated delay elements, such as extra-long connecting cable runs between wall jacks and DTEs. (A safety margin of 32 BT is recommended.)
f) Insert the values obtained through the calculations above into the following formula to calculate the PDV. (Some configurations may not use all the elements of the formula.)

PDV = ∑link delays (LSDV) + repeater delay + DTE delay + safety margin

g) If the PDV is less than 4096, the path is qualified in terms of worst-case delay.
h) Late collisions and/or CRC errors may be indications that path delays exceed 4096 BT.

Table 42–3—Network component delays, Transmission System Model 2

Component	Round-trip delay in bit times per meter (BT/m)	Maximum round-trip delay in bit times (BT)
Two DTEs		864
Category 5 UTP Cable segment	11.12	1112 (100 m)
Shielded Jumper Cable segment	10.10	253 (25 m)
Optical Fiber Cable segment	10.10	1111 (110 m)
Repeater		976

Table 42–4—Conversion table for cable delays

Speed relative to c	ns/m	BT/m
0.4	8.34	8.34
0.5	6.67	6.67
0.51	6.54	6.54
0.52	6.41	6.41
0.53	6.29	6.29
0.54	6.18	6.18
0.55	6.06	6.06
0.56	5.96	5.96
0.57	5.85	5.85
0.58	5.75	5.75
0.5852	5.70	5.70
0.59	5.65	5.65
0.6	5.56	5.56
0.61	5.47	5.47
0.62	5.38	5.38
0.63	5.29	5.29

Table 42–4—Conversion table for cable delays *(continued)*

Speed relative to c	ns/m	BT/m
0.64	5.21	5.21
0.65	5.13	5.13
0.654	5.10	5.10
0.66	5.05	5.05
0.666	5.01	5.01
0.67	4.98	4.98
0.68	4.91	4.91
0.69	4.83	4.83
0.7	4.77	4.77
0.8	4.17	4.17
0.9	3.71	3.71

42.4 Full duplex 1000 Mb/s topology limitations

Unlike half duplex CSMA/CD networks, the physical size of full duplex 1000 Mb/s networks is not limited by the round-trip collision propagation delay. Instead, the maximum link length between DTEs is limited only by the signal transmission characteristics of the specific link.

43. Link Aggregation

43.1 Overview

This clause defines an optional Link Aggregation sublayer for use with CSMA/CD MACs. Link Aggregation allows one or more links to be aggregated together to form a Link Aggregation Group, such that a MAC Client can treat the Link Aggregation Group as if it were a single link. To this end, it specifies the establishment of DTE to DTE logical links, consisting of N parallel instances of full duplex point-to-point links operating at the same data rate.

43.1.1 Terminology

In this clause, unless otherwise noted, the term *link* refers to an *Aggregation Link* and the term *port* refers to *Aggregation Port,* as defined in 1.4. This allows for better readability of this clause while avoiding conflicting use of these terms in other clauses of this standard. Similarly, the term *Key* when used in this clause is synonymous with *Aggregation Key,* and the term *System* is synonymous with *Aggregation System*.

43.1.2 Goals and objectives

Link Aggregation, as specified in this clause, provides the following:

a) **Increased bandwidth**—The capacity of multiple links is combined into one logical link.
b) **Linearly incremental bandwidth**—Bandwidth can be increased in unit multiples as opposed to the order-of-magnitude increase available through Physical Layer technology options (10 Mb/s, 100 Mb/s, 1000 Mb/s, etc.).
c) **Increased availability**—The failure or replacement of a single link within a Link Aggregation Group need not cause failure from the perspective of a MAC Client.
d) **Load sharing**—MAC Client traffic may be distributed across multiple links.
e) **Automatic configuration**—In the absence of manual overrides, an appropriate set of Link Aggregation Groups is automatically configured, and individual links are allocated to those groups. If a set of links can aggregate, they will aggregate.
f) **Rapid configuration and reconfiguration**—In the event of changes in physical connectivity, Link Aggregation will quickly converge to a new configuration, typically on the order of 1 second or less.
g) **Deterministic behavior**—Depending on the selection algorithm chosen, the configuration can be made to resolve deterministically; i.e., the resulting aggregation can be made independent of the order in which events occur, and be completely determined by the capabilities of the individual links and their physical connectivity.
h) **Low risk of duplication or mis-ordering**—During both steady-state operation and link (re)configuration, there is a high probability that frames are neither duplicated nor mis-ordered.
i) **Support of existing IEEE 802.3® MAC Clients**—No change is required to existing higher-layer protocols or applications to use Link Aggregation.
j) **Backwards compatibility with aggregation-unaware devices**—Links that cannot take part in Link Aggregation—either because of their inherent capabilities, management configuration, or the capabilities of the devices to which they attach—operate as normal, individual IEEE 802.3® links.
k) **Accommodation of differing capabilities and constraints**—Devices with differing hardware and software constraints on Link Aggregation are, to the extent possible, accommodated.
l) **No change to the IEEE 802.3® frame format**—Link aggregation neither adds to, nor changes the contents of frames exchanged between MAC Clients.
m) **Network management support**—The standard specifies appropriate management objects for configuration, monitoring, and control of Link Aggregation.

Link Aggregation, as specified in this clause, does not support the following:

n) **Multipoint Aggregations**—The mechanisms specified in this clause do not support aggregations among more than two Systems.
o) **Dissimilar MACs**—Link Aggregation is supported only on links using the IEEE 802.3® MAC.
p) **Half-duplex operation**—Link Aggregation is supported only on point-to-point links with MACs operating in full duplex mode.
q) **Operation across multiple data rates**—All links in a Link Aggregation Group operate at the same data rate (e.g., 10 Mb/s, 100 Mb/s, or 1000 Mb/s).

43.1.3 Positioning of Link Aggregation within the IEEE 802.3® architecture

Link Aggregation comprises an optional sublayer between a MAC Client and the MAC (or optional MAC Control sublayer). Figure 43–1 depicts the positioning of the Link Aggregation sublayer in the CSMA/CD layer architecture, and the relationship of that architecture to the Data Link and Physical Layers of the OSI Reference Model. The figure also shows the ability of the Link Aggregation sublayer to combine a number of individual links in order to present a single MAC interface to the MAC Client.

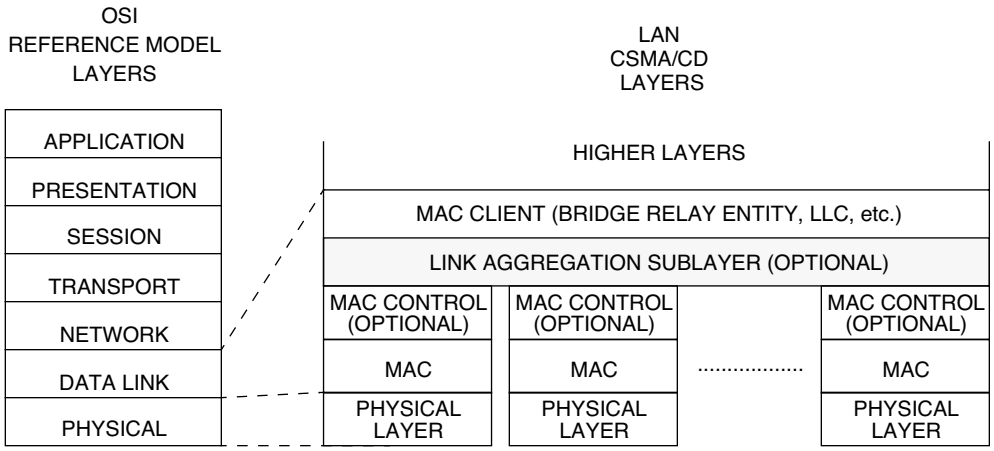


Figure 43–1—Architectural positioning of Link Aggregation sublayer

Figure 43–2 depicts the major blocks that form the Link Aggregation sublayer, and their interrelationships.

It is possible to implement the optional Link Aggregation sublayer for some ports within a System while not implementing it for other ports; i.e., it is not necessary for all ports in a System to be subject to Link Aggregation. A conformant implementation is not required to be able to apply the Link Aggregation sublayer to every port.

43.1.4 State diagram conventions

Many of the functions specified in this clause are presented in state diagram notation. All state diagrams contained in this clause use the notation and conventions defined in 21.5. In the event of a discrepancy between the text description and the state diagram formalization of a function, the state diagrams take precedence.

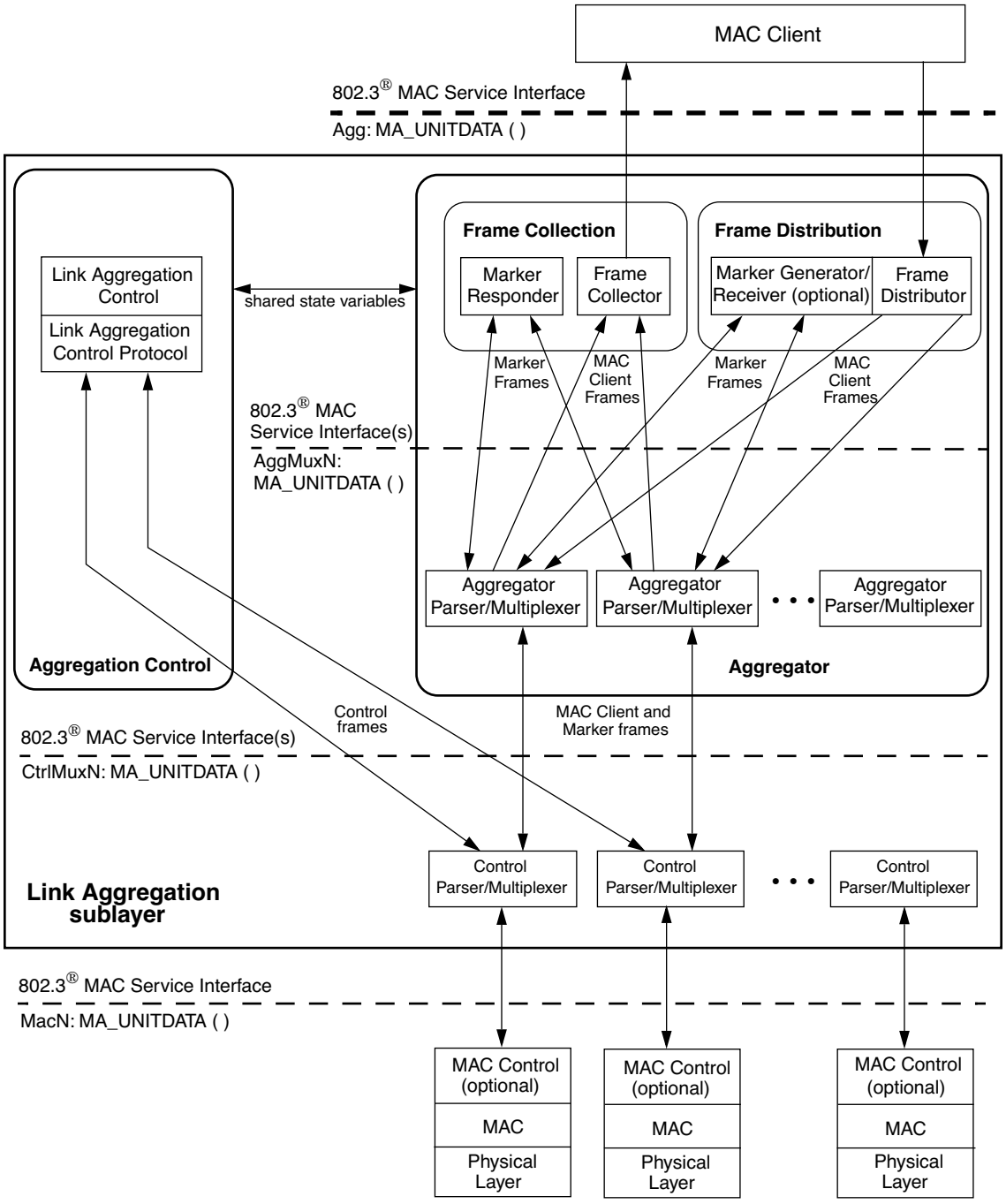


Figure 43–2—Link Aggregation sublayer block diagram

43.2 Link Aggregation operation

As depicted in Figure 43–2, the Link Aggregation sublayer comprises the following functions:

a) *Frame Distribution*. This block is responsible for taking frames submitted by the MAC Client and submitting them for transmission on the appropriate port, based on a frame distribution algorithm employed by the Frame Distributor. Frame Distribution also includes an optional *Marker Generator/Receiver* used for the Marker protocol. (See 43.2.4, 43.2.5, and 43.5.)

b) *Frame Collection*. This block is responsible for passing frames received from the various ports to the MAC Client. Frame Collection also includes a *Marker Responder,* used for the Marker protocol. (See 43.2.3 and 43.5.)

c) *Aggregator Parser/Multiplexers*. On transmit, these blocks simply pass frame transmission requests from the Distributor, Marker Generator, and/or Marker Responder to the appropriate port. On receive, these blocks distinguish among Marker Request, Marker Response, and MAC Client PDUs, and pass each to the appropriate entity (Marker Responder, Marker Receiver, and Collector, respectively).

d) *Aggregator.* The combination of Frame Distribution and Collection, along with the Aggregator Parser/Multiplexers, is referred to as the Aggregator.

e) *Aggregation Control*. This block is responsible for the configuration and control of Link Aggregation. It incorporates a *Link Aggregation Control Protocol* (LACP) that can be used for automatic communication of aggregation capabilities between Systems and automatic configuration of Link Aggregation.

f) *Control Parser/Multiplexers*. On transmit, these blocks simply pass frame transmission requests from the Aggregator and Control entities to the appropriate port. On receive, these blocks distinguish Link Aggregation Control PDUs from other frames, passing the LACPDUs to the appropriate sublayer entity, and all other frames to the Aggregator.

43.2.1 Principles of Link Aggregation

Link Aggregation allows a MAC Client to treat a set of one or more ports as if it were a single port. In doing so, it employs the following principles and concepts:

a) A MAC Client communicates with a set of ports through an Aggregator, which presents a standard IEEE 802.3® service interface to the MAC Client. The Aggregator binds to one or more ports within a System.

b) It is the responsibility of the Aggregator to distribute frame transmissions from the MAC Client to the various ports, and to collect received frames from the ports and pass them to the MAC Client transparently.

c) A System may contain multiple Aggregators, serving multiple MAC Clients. A given port will bind to (at most) a single Aggregator at any time. A MAC Client is served by a single Aggregator at a time.

d) The binding of ports to Aggregators within a System is managed by the Link Aggregation Control function for that System, which is responsible for determining which links may be aggregated, aggregating them, binding the ports within the System to an appropriate Aggregator, and monitoring conditions to determine when a change in aggregation is needed.

e) Such determination and binding may be under manual control through direct manipulation of the state variables of Link Aggregation (e.g., Keys) by a network manager. In addition, automatic determination, configuration, binding, and monitoring may occur through the use of a Link Aggregation Control Protocol (LACP). The LACP uses peer exchanges across the links to determine, on an ongoing basis, the aggregation capability of the various links, and continuously provides the maximum level of aggregation capability achievable between a given pair of Systems.

f) Frame ordering must be maintained for certain sequences of frame exchanges between MAC Clients (known as conversations, see 1.4). The Distributor ensures that all frames of a given conversation are passed to a single port. For any given port, the Collector is required to pass frames to the MAC Client in the order that they are received from that port. The Collector is otherwise free to select frames received from the aggregated ports in any order. Since there are no means for frames to be misordered on a single link, this guarantees that frame ordering is maintained for any conversation.

g) Conversations may be moved among ports within an aggregation, both for load balancing and to maintain availability in the event of link failures.

h) This standard does not impose any particular distribution algorithm on the Distributor. Whatever algorithm is used should be appropriate for the MAC Client being supported.

i) Each port is assigned a unique, globally administered MAC address. This MAC address is used as the source address for frame exchanges that are initiated by entities within the Link Aggregation sublayer itself (i.e., LACP and Marker protocol exchanges).

NOTE—The LACP and Marker protocols use a multicast destination address for all exchanges, and do not impose any requirement for a port to recognize more than one unicast address on received frames.

j) Each Aggregator is assigned a unique, globally administered MAC address; this address is used as the MAC address of the aggregation from the perspective of the MAC Client, both as a source address for transmitted frames and as the destination address for received frames. The MAC address of the Aggregator may be one of the MAC addresses of a port in the associated Link Aggregation Group (see 43.2.10).

43.2.2 Service interfaces

The MAC Client communicates with the Aggregator using the standard service interface specified in Clause 2. Similarly, Link Aggregation communicates internally (between Frame Collection/Distribution, the Aggregator Parser/Multiplexers, the Control Parser/Multiplexers, and Link Aggregation Control) and with its bound ports using the same, standard service interface. No new interlayer service interfaces are defined for Link Aggregation.

Since Link Aggregation uses four instances of the MAC Service Interface, it is necessary to introduce a notation convention so that the reader can be clear as to which interface is being referred to at any given time. A prefix is therefore assigned to each service primitive, indicating which of the four interfaces is being invoked, as depicted in Figure 43–2. The prefixes are as follows:

a) *Agg:,* for primitives issued on the interface between the MAC Client and the Link Aggregation sublayer.

b) *AggMuxN:,* for primitives issued on the interface between Aggregator Parser/Multiplexer N and its internal clients (where N is the port number associated with the Aggregator Parser/Multiplexer).

c) *CtrlMuxN:,* for primitives issued on the interface between Control Parser/Multiplexer N and its internal clients (where N is the port number associated with the Control Parser/Multiplexer).

d) *MacN:,* for primitives issued on the interface between underlying MAC N and its Control Parser/Multiplexer (where N is the port number associated with the underlying MAC).

MAC Clients may generate Agg:MA_DATA.request primitives for transmission on an aggregated link. These are passed by the Frame Distributor to a port selected by the distribution algorithm. MacN:MA_DATA.indication primitives signifying received frames are passed unchanged from a port to the MAC Client by the Frame Collector.

MAC Clients that generate MA_CONTROL.request primitives (and which expect MA_CONTROL.indication primitives in response) cannot communicate through a Link Aggregation sublayer. They must communicate directly with the MAC Control entity through which these control primitives are to be sent and received.

The multiplexing of such MAC Control clients with a Link Aggregation sublayer for simultaneous use of a single port is outside the scope of this standard.

The multiplexing of MAC Clients with a Link Aggregation sublayer for simultaneous use of an individual MAC that is also part of a Link Aggregation Group is outside the scope of this standard.

43.2.3 Frame Collector

A Frame Collector is responsible for receiving incoming frames (i.e., AggMuxN:MA_DATA.indications) from the set of individual links that form the Link Aggregation Group (through each link's associated Aggregator Parser/Multiplexer) and delivering them to the MAC Client. Frames received from a given port are delivered to the MAC Client in the order that they are received by the Frame Collector. Since the Frame Distributor is responsible for maintaining any frame ordering constraints, there is no requirement for the Frame Collector to perform any reordering of frames received from multiple links.

The Frame Collector shall implement the function specified in the state diagram shown in Figure 43–3 and the associated definitions contained in 43.2.3.1.

43.2.3.1 Frame Collector state diagram

43.2.3.1.1 Constants

CollectorMaxDelay
: In tens of microseconds, the maximum time that the Frame Collector may delay the delivery of a frame received from an Aggregator Parser to its MAC Client. Value is assigned by management or administration policy.
Value: Integer

43.2.3.1.2 Variables

DA
SA
m_sdu
status
: The parameters of the MA_DATA.indication primitive, as defined in Clause 2.

BEGIN
: A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.
Value: Boolean

43.2.3.1.3 Messages

Agg:MA_DATA.indication
AggMuxN:MA_DATA.indication
: The service primitives used to pass a received frame to a client with the specified parameters.

43.2.3.1.4 State diagram

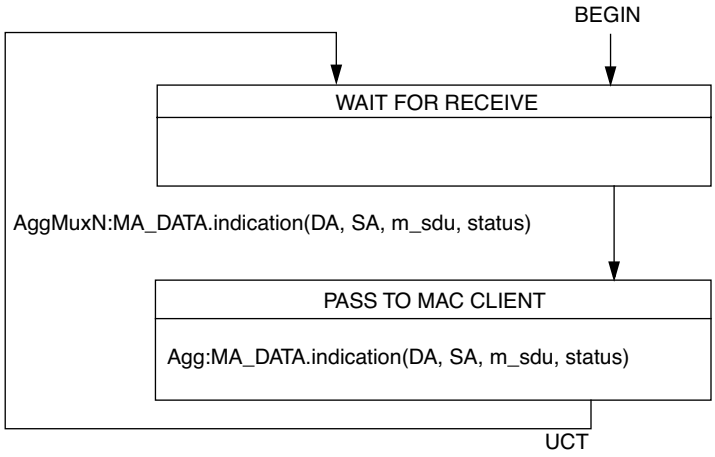


Figure 43–3—Frame Collector state diagram

The architectural models of the IEEE 802.3® MAC and Link Aggregation do not make any provision for queueing of frames between the link and the MAC Client. Furthermore, the state machine conventions used in this standard assume that actions within a state occur instantaneously (see 21.5). However, practical implementations of Link Aggregation will typically incur both queueing and delay in the Frame Collector. In order to ensure that frame delivery is not delayed indefinitely (which could cause a frame ordering problem when moving conversations from one link to another), the Frame Collector shall, upon receiving a frame from an Aggregator Parser, either deliver the frame to its MAC Client, or discard the frame within a CollectorMaxDelay time. The Frame Distributor (within the Partner System at the other end of the link) can assume that all frames transmitted on a given link have been either received by its Partner's MAC Client or discarded after a CollectorMaxDelay plus the propagation delay of the link. The use of CollectorMaxDelay is further discussed in 43A.3.

43.2.4 Frame Distributor

The Frame Distributor is responsible for taking outgoing frames from the MAC Client and transmitting them through the set of links that form the Link Aggregation Group. The Frame Distributor implements a distribution function (algorithm) responsible for choosing the link to be used for the transmission of any given frame or set of frames.

This standard does not mandate any particular distribution algorithm(s); however, any distribution algorithm shall ensure that, when frames are received by a Frame Collector as specified in 43.2.3, the algorithm shall not cause

a) Mis-ordering of frames that are part of any given conversation, or
b) Duplication of frames.

The above requirement to maintain frame ordering is met by ensuring that all frames that compose a given conversation are transmitted on a single link in the order that they are generated by the MAC Client; hence, this requirement does not involve the addition (or modification) of any information to the MAC frame, nor any buffering or processing on the part of the corresponding Frame Collector in order to re-order frames. This approach to the operation of the distribution function permits a wide variety of distribution and load balancing algorithms to be used, while also ensuring interoperability between devices that adopt differing algorithms.

NOTE—The subject of distribution algorithms and maintenance of frame ordering is discussed in Annex 43A.

The Frame Distributor shall implement the function specified in the state diagram shown in Figure 43–4 and the associated definitions contained in 43.2.4.1.

43.2.4.1 Frame Distributor state diagram

43.2.4.1.1 Variables

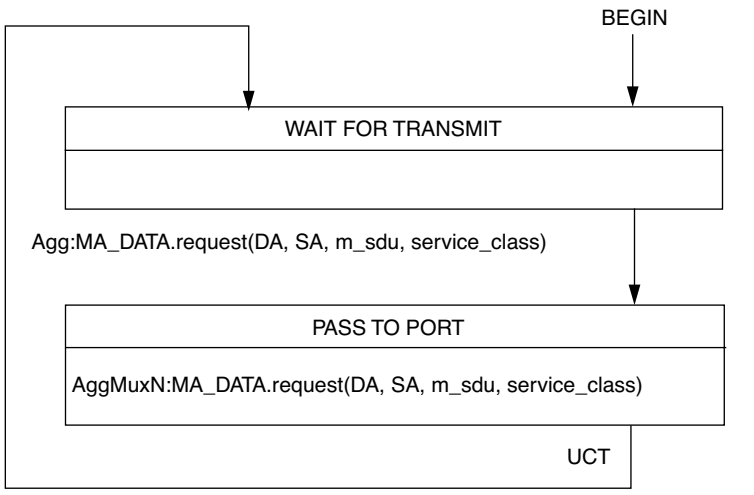
DA
SA
m_sdu
service_class
: The parameters of the MA_DATA.request primitive, as defined in Clause 2.

BEGIN
: A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.
Value: Boolean

43.2.4.1.2 Messages

Agg:MA_DATA.request
AggMuxN:MA_DATA.request
: The service primitives used to transmit a frame with the specified parameters.

43.2.4.1.3 State diagram



If a client issues an Agg:MA_DATA.request primitive that contains no SA parameter, the AggMuxN:MA_DATA.request primitive generated shall use the Aggregator's MAC address for the SA.

NOTE—The algorithm that the Frame Distributor uses to select the value of N in AggMuxN:MA_DATA.request for a given frame is unspecified.

Figure 43–4—Frame Distributor state diagram

43.2.5 Marker Generator/Receiver (optional)

The optional Marker Generator is used by the Marker protocol, as specified in 43.5. When implemented and so requested by the Distribution algorithm, the Marker Generator shall issue an AggMuxN:MA_DATA.request primitive, with an m_sdu containing a Marker PDU as defined in 43.5.3, to the port associated with the conversation being marked, subject to the timing restrictions for Slow Protocols specified in Annex 43B.

The optional Marker Receiver is used by the Marker protocol, as specified in 43.5. It receives Marker Response PDUs from the Aggregator Parser.

43.2.6 Marker Responder

The Marker Responder is used by the Marker protocol, as specified in 43.5. The Marker Responder receives Marker PDUs (generated by a Partner System's Marker Generator), and transmits a Marker Response PDU through the same port from which the Marker PDU was received. While implementation of the Marker Generator/Receiver is optional, the ability to respond to a Marker PDU (the Marker Responder) is mandatory. An implementation conformant to this clause shall implement the Marker Responder as specified in 43.5.4.2, thus ensuring that implementations that need to make use of the protocol can do so.

43.2.7 Aggregator Parser/Multiplexer

On transmission, the Aggregator Multiplexer shall provide transparent pass-through of frames submitted by the Marker Responder and optional Marker Generator to the port specified in the transmission request. The Aggregator Multiplexer shall provide transparent pass-through of frames submitted by the Frame Distributor to the port specified in the transmission request only when the port state is Distributing (see 43.4.15); otherwise, such frames shall be discarded.

On receipt, the Aggregator Parser decodes frames received from the Control Parser, passes those frames destined for the Marker Responder or Marker Receiver to the selected entity, and discards frames with invalid Slow Protocol subtype values (see Table 43B–2). The Aggregator Parser shall pass all other frames to the Frame Collector for passage to the MAC Client only when the port state is Collecting (see 43.4.15); otherwise, such frames shall be discarded. The Aggregator Parser shall implement the function specified in the state diagram shown in Figure 43–5 and the associated definitions contained in 43.2.7.1.

43.2.7.1 Aggregator Parser state diagram

43.2.7.1.1 Constants

Slow_Protocols_Multicast
: The value of the Slow Protocols Multicast address. (See Table 43B–1.)

Slow_Protocols_Type
: The value of the Slow Protocols Length/Type field. (See Annex 43B.)

Marker_subtype
: The value of the Subtype field for the Marker protocol. (See 43.5.3.)
Value: Integer
2

Marker_Information
: The encoding of the Marker Information TLV_type field. (See 43.5.3.)
Value: Integer
1

Marker_Response_Information
The encoding of the Marker Response Information TLV_type field. (See 43.5.3.)
Value: Integer
2

43.2.7.1.2 Variables

DA
SA
m_sdu
status
The parameters of the MA_DATA.indication primitive as defined in Clause 2.

Length/Type
The value of the Length/Type field in a received frame.
Value: Integer

Subtype
The value of the octet following the Length/Type field in a Slow Protocol frame. (See Annex 43B.)
Value: Integer

TLV_type
The value contained in the octet following the Version Number in a received Marker or Marker Response frame. This identifies the "type" for the Type/Length/Value (TLV) tuple. (See 43.5.3.)
Value: Integer

BEGIN
A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.
Value: Boolean

43.2.7.1.3 Messages

CtrlMuxN:MA_DATA.indication
AggMuxN:MA_DATA.indication
The service primitives used to pass a received frame to a client with the specified parameters.

43.2.8 Aggregator

An *Aggregator* comprises an instance of a Frame Collection function, an instance of a Frame Distribution function and one or more instances of the Aggregator Parser/Multiplexer function for a Link Aggregation Group. A single Aggregator is associated with each Link Aggregation Group. An Aggregator offers a standard IEEE 802.3® MAC service interface to its associated MAC Client; access to the MAC service by a MAC Client is always achieved via an Aggregator. An Aggregator can therefore be considered to be a *logical MAC*, bound to one or more ports, through which the MAC client is provided access to the MAC service.

A single, individual MAC address is associated with each Aggregator (see 43.2.10).

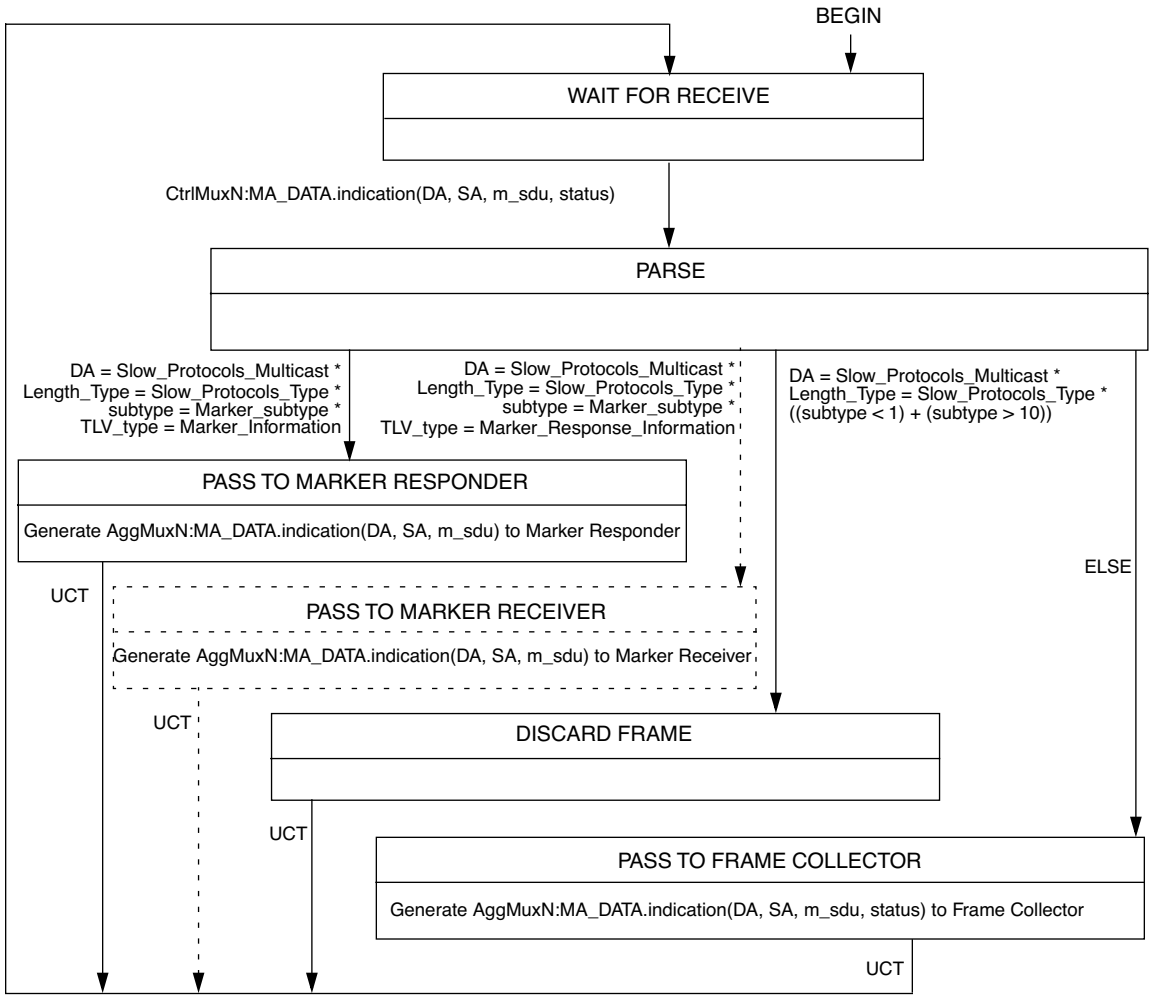
An Aggregator is available for use by the MAC Client if the following are all true:

a) It has one or more attached ports.
b) The Aggregator has not been set to a disabled state by administrative action (see 30.7.1.1.13).

c) The collection and/or distribution function associated with one or more of the attached ports is enabled (see 30.7.1.1.14).

NOTE—To simplify the modeling and description of the operation of Link Aggregation, it is assumed that there are as many Aggregators as there are ports in a given System; however, this is not a requirement of this standard. Aggregation of two or more ports consists of changing the bindings between ports and Aggregators such that more than one port is bound to a single Aggregator. The creation of any aggregations of two or more links will therefore result in one or more Aggregators that are bound to more than one port, and one or more Aggregators that are not bound to any port. An Aggregator that is not bound to any port appears to a MAC Client as a MAC interface to an inactive port. During times when the bindings between ports and Aggregators are changing, or as a consequence of particular configuration choices, there may be occasions when one or more ports are not bound to any Aggregator.

43.2.8.1 State diagram



If the optional Marker Receiver is not implemented, Marker Responses shall be passed to the Frame Collector. If the port state is not Collecting, all frames that would have been passed to the MAC Client through the Collector will be discarded.

Figure 43–5—Aggregator Parser state diagram

43.2.9 Control Parser/Multiplexer

On transmission, the Control Multiplexer shall provide transparent pass-through of frames submitted by the Aggregator and Link Aggregation Control Protocol to the port specified in the transmission request.

On receipt, the Control Parser decodes frames received from the various ports in the Link Aggregation Group, passes those frames destined for the Link Aggregation Control Protocol to the appropriate entity, and passes all other frames to the Aggregator Parser. The Control Parser shall implement the function specified by the state diagram shown in Figure 43–6 and the associated definitions contained in 43.2.9.1.

43.2.9.1 Control Parser state diagram

43.2.9.1.1 Constants

Slow_Protocols_Multicast
: The value of the Slow Protocols Multicast address. (See Table 43B–1.)

Slow_Protocols_Type
: The value of the Slow Protocols Length/Type field. (See Table 43B–2.)

LACP_subtype
: The value of the Subtype field for the Link Aggregation Control Protocol. (See Table 43B–3.)
Value: Integer
1

43.2.9.1.2 Variables

DA
SA
m_sdu
status
: The parameters of the MA_DATA.indication primitive, as defined in Clause 2.

Length/Type
: The value of the Length/Type field in a received frame.
Value: Integer

Subtype
: The value of the octet following the Length/Type field in a Slow Protocol frame.
(See Annex 43B.)
Value: Integer

BEGIN
: A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.
Value: Boolean

43.2.9.1.3 Messages

MacN:MA_DATA.indication
CtrlMuxN:MA_DATA.indication
: The service primitives used to pass a received frame to a client with the specified parameters.

43.2.9.1.4 State diagram

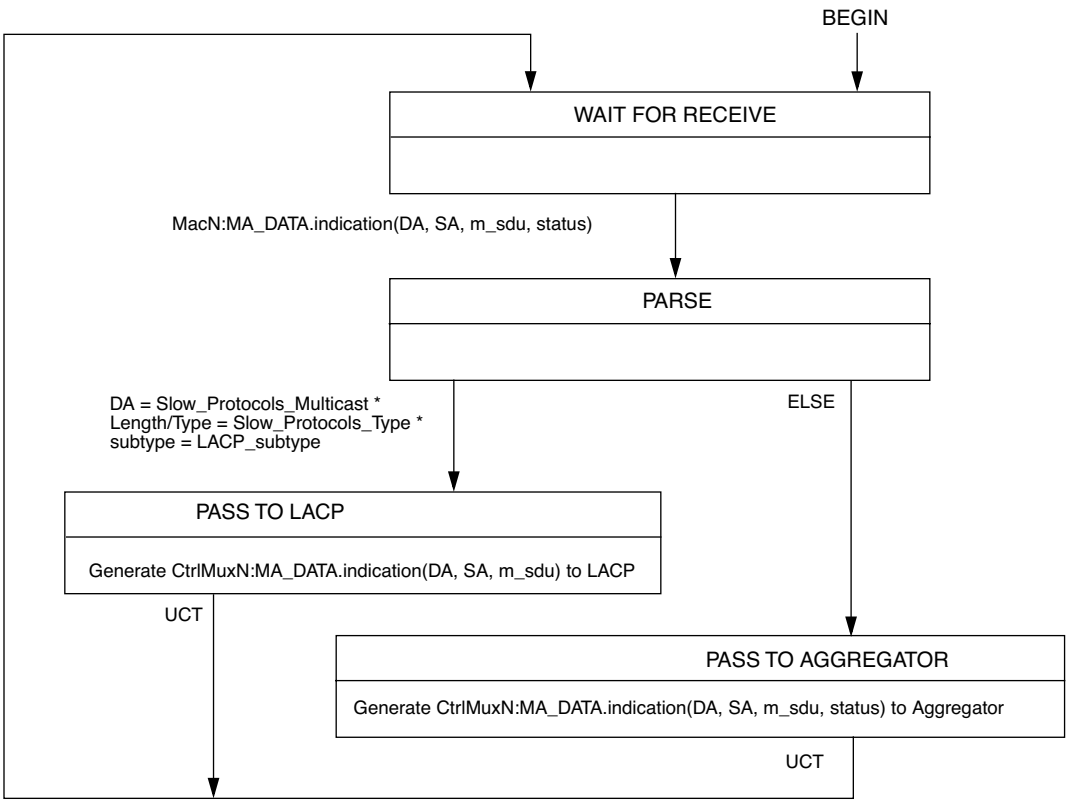


Figure 43–6—Control Parser state diagram

43.2.10 Addressing

Each IEEE 802.3® MAC has an associated globally-unique individual MAC address, whether that MAC is used for Link Aggregation or not (see Clause 4).

Each Aggregator to which one or more ports are attached has an associated globally-unique individual MAC address (see 43.3.3). The MAC address of the Aggregator may be the globally-unique individual MAC addresses of one of the MACs in the associated Link Aggregation Group, or it may be a distinct MAC address. The manner in which such addresses are chosen is not otherwise constrained by this standard.

Protocol entities sourcing frames from within the Link Aggregation sublayer (e.g., LACP and the Marker protocol) use the MAC address of the MAC within an underlying port as the source address in frames transmitted through that port. The MAC Client sees only the Aggregator and not the underlying MACs, and therefore uses the Aggregator's MAC address as the source address in transmitted frames. If a MAC Client submits a frame to the Aggregator for transmission without specifying a source address, the Aggregator inserts its own MAC address as the source address for transmitted frames.

NOTE—This behavior causes the Aggregator to behave the same way as a standard MAC with regard to frames submitted by its client.

43.3 Link Aggregation Control

Link Aggregation Control configures and controls the Link Aggregation sublayer using static information local to the control function and dynamic information exchanged by means of the Link Aggregation Control Protocol.

For each aggregatable port in the System, Link Aggregation Control

a) Maintains configuration information (reflecting the inherent properties of the individual links as well as those established by management) to control aggregation.
b) Exchanges configuration information with other Systems to allocate the link to a Link Aggregation Group.

NOTE—A given link is allocated to, at most, one Link Aggregation Group at a time. The allocation mechanism attempts to maximize aggregation, subject to management controls.

c) Attaches the port to the Aggregator used by the Link Aggregation Group, and detaches the port from the Aggregator when it is no longer used by the Group.
d) Uses information from the Partner System's Link Aggregation Control entity to enable or disable the Aggregator's Collector and Distributor.

The operation of Link Aggregation Control involves the following activities, which are described in detail in the subclauses that follow:

e) Checking that candidate links can actually be aggregated.
f) Controlling the addition of a link to a Link Aggregation Group, and the creation of the group if necessary.
g) Monitoring the status of aggregated links to ensure that the aggregation is still valid.
h) Removing a link from a Link Aggregation Group if its membership is no longer valid, and removing the group if it no longer has any member links.

In order to allow Link Aggregation Control to determine whether a set of links connect to the same System, and to determine whether those links are compatible from the point of view of aggregation, it is necessary to be able to establish

i) A globally unique identifier for each System that participates in Link Aggregation (see 43.3.2).
j) A means of identifying the set of capabilities associated with each port and with each Aggregator, as understood by a given System.
k) A means of identifying a Link Aggregation Group and its associated Aggregator.

System identification allows the detection of links that are connected in a loopback configuration (i.e., both ends of the same link are connected to the same System).

43.3.1 Characteristics of Link Aggregation Control

Link Aggregation Control provides a configuration capability that is

a) **Automatic.** In the absence of manual override controls, an appropriate set of Link Aggregation Groups is automatically configured, and individual links are allocated to those groups. If a set of links can aggregate, they do aggregate.
b) **Continuous.** Manual intervention or initialization events are not a requirement for correct operation. The configuration mechanism continuously monitors for changes in state that require reconfiguration. The configuration functions detect and correct misconfigurations by performing reconfiguration and/or by taking misconfigured links out of service.

c) **Deterministic.** The configuration can resolve deterministically; i.e., the configuration achieved can be made independent of the order in which events occur, and be completely determined by the combination of the capabilities of the individual links and their physical connectivity.
d) **Controllable.** The configuration capabilities accommodate devices with differing hardware and software constraints on Link Aggregation.
e) **Compatible.** Links that cannot take part in Link Aggregation, either because of their inherent capabilities or of the capabilities of the devices to which they attach, operate as normal IEEE 802.3® links. The introduction of Link Aggregation capability at one or both ends of a link should not result in a degradation of the perceived performance of the link.
f) **Rapid.** The configuration resolves rapidly to a stable configuration. Convergence can be achieved by the exchange of three LACPDUs, without dependence on timer values.

with

g) **Low risk of misdelivery.** The operation of the (re-)configuration functions minimizes the risk of frames being delivered to the wrong Aggregator.
h) **Low risk of duplication or mis-ordering.** The operation of the (re-)configuration functions minimizes the risk of frame duplication and frame mis-ordering.
i) **Low protocol overhead.** The overhead involved in external communication of configuration information between devices is small.

43.3.2 System identification

The globally unique identifier used to identify a System shall be the concatenation of a globally administered individual MAC address and the System Priority. The MAC address chosen may be the individual MAC address associated with one of the ports of the System.

Where it is necessary to perform numerical comparisons between System Identifiers, each System Identifier is considered to be an eight octet unsigned binary number, constructed as follows:

a) The two most significant octets of the System Identifier comprise the System Priority. The System Priority value is taken to be an unsigned binary number; the most significant octet of the System Priority forms the most significant octet of the System Identifier.
b) The third most significant octet of the System Identifier is derived from the initial octet of the MAC address; the least significant bit of the octet is assigned the value of the first bit of the MAC address, the next most significant bit of the octet is assigned the value of the next bit of the MAC address, and so on. The fourth through eighth octets are similarly assigned the second through sixth octets of the MAC address.

43.3.3 Aggregator identification

Each Aggregator to which one or more ports are attached shall be assigned a unique, globally administered individual MAC address. The MAC address assigned to the Aggregator may be the same as the MAC address assigned to one of its bound ports. No Aggregator shall be assigned a MAC address that is the same as that of a port bound to a different Aggregator within the System. When receiving frames, a port is never required to recognize more than one unicast address, i.e., the Aggregator's MAC address.

NOTE—This ensures that Aggregators can be uniquely addressed, and allows (but does not require) the unique address to be allocated from the same set of addresses as are assigned to the ports. It also acknowledges the fact that locally administered addresses may be used in particular implementations or environments. The stated restriction on the allocation of MAC addresses to Aggregators may have implications with regard to the choice of selection algorithm.

An aggregator also shall be assigned an integer identifier that is used by Link Aggregation Control to uniquely identify the aggregator within the System. This value will typically be the same as the interface identifier (ifIndex) used for management purposes.

43.3.4 Port identification

Link Aggregation Control uses a Port Identifier, comprising the concatenation of a Port Priority and a Port Number, to identify the port. Port Numbers (and hence, Port Identifiers) shall be uniquely assigned within a System. Port Number 0 shall not be assigned to any port.

When it is necessary to perform numerical comparisons between Port Identifiers, each Port Identifier is considered to be a four octet unsigned binary number constructed as follows:

a) The most significant and second most significant octets are the first and second most significant octets of the Port Priority, respectively.
b) The third and fourth most significant octets are the first and second most significant octets of the Port Number, respectively.

43.3.5 Capability identification

The ability of one port to aggregate with another is summarized by a simple integer parameter, known as a Key. This facilitates communication and comparison of aggregation capabilities, which may be determined by a number of factors, including

a) The port's physical characteristics, such as data rate, duplexity, point-to-point or shared medium.
b) Configuration constraints established by the network administrator.
c) Use of the port by higher layer protocols (e.g. assignment of Network Layer addresses).
d) Characteristics or limitations of the port implementation itself.

Two Keys shall be associated with each port: an operational Key and an administrative Key. The operational Key is the Key that is currently in active use for the purposes of forming aggregations. The administrative Key allows manipulation of Key values by management. The administrative and operational Keys assigned to a port may differ

e) If the operation of the implementation is such that an administrative change to a Key value cannot be immediately reflected in the operational state of the port.
f) If the System supports the dynamic manipulation of Keys, as discussed in 43.6.2, either to accurately reflect changes in operational capabilities of the port (for example, as a result of Auto-Negotiation), or to provide a means of handling constraints on aggregation capability.

A given Key value is meaningful only in the context of the System that allocates it; there is no global significance to Key values. Similarly, the relationship between administrative and operational Key values is meaningful only in the context of the System that allocates it. When a System assigns an administrative Key value to a set of ports, it signifies that the set of ports have the potential to aggregate together, subject to the considerations discussed in 43.6.2. When a System assigns an operational Key value to a set of ports, it signifies that, in the absence of other constraints, the current operational state of the set of ports allows any subset of that set of ports (including the entire set) to be aggregated together from the perspective of the System making the assignment. The set of such ports that will actually be aggregated will be those that terminate at a common Partner System, and for which that Partner System has assigned a common operational Key value, local to that Partner. The set of ports in a given System that share the same operational Key value are said to be members of the same Key Group.

A System may determine that a given link is not able to be aggregated with other links. Such links are referred to as Individual links (as opposed to Aggregatable links). A System may declare a link to be Individual if the inherent properties of the link allow its use as part of an aggregation, but the system is aware of no other links that are capable of aggregating with this link (e.g., the System has allocated a unique operational Key value to the link).

The capability information communicated between Systems, therefore, includes this local knowledge of the aggregation capability of the link in addition to the operational Key value; i.e., whether the System considers the link to be Aggregatable or Individual.

An administrative Key value and an operational Key value shall also be associated with each Aggregator. The operational Key is the Key that is currently in active use for the purposes of forming aggregations. The administrative Key allows manipulation of Key values by management. The values of administrative and operational Key for an Aggregator may differ in the same manner as that of port Keys, per item e) and item f), in this subclause. Ports that are members of a given Key Group can only be bound to Aggregators that share the same operational Key value.

All Keys are 16-bit identifiers. All values except the null value (all zeroes) are available for local use.

NOTE—This model allows for two convenient initial configurations. The first is achieved by assigning each port an initial administrative and operational Key value identical to its port number, and assigning the same port numbers as Keys to the corresponding Aggregators for each port. A device with this initial configuration will bring up all links as individual, non-aggregated links. The second is achieved by assigning the same administrative and operational Key values to all ports with a common set of capabilities, and also to all Aggregators. A device with this initial configuration will attempt to aggregate together any set of links that have the same Partner System ID and operational Key, and for which both Systems are prepared to allow aggregation.

43.3.6 Link Aggregation Group identification

A Link Aggregation Group consists of either

a) One or more Aggregatable links that terminate in the same pair of Systems and whose ports belong to the same Key Group in each System, or
b) An Individual link.

43.3.6.1 Construction of the Link Aggregation Group Identifier

A unique Link Aggregation Group Identifier (LAG ID) is constructed from the following parameters for each of the communicating Systems:

a) The System Identifier
b) The operational Key assigned to the ports in the LAG
c) The Port Identifier, if the link is identified as an Individual link

The local System's values for these parameters shall be non-zero. In cases where the local System is unable to determine the remote System's values for these parameters by exchange of protocol information, administrative values are used in the construction of the LAG ID. The value of these administrative parameters for the remote System may be configured as zero, provided that the port(s) concerned are also configured to be Individual.

A compound identifier formed from the System Identifiers and Key values alone is sufficient to identify a LAG comprising Aggregatable links. However, such an identifier is not sufficient for a LAG comprising a single Individual link where the Partner System Identifier and operational Key may be zero. Even if these are non-zero there may be multiple Individual Links with the same System Identifier and operational Key combinations, and it becomes necessary to include Port Identifiers to provide unique LAG IDs.

Given

d) S and T are System Identifiers,
e) K and L are the operational Keys assigned to a LAG by S and T respectively, and
f) P and Q are the Port Identifiers of the ports being attached if the LAG comprises a single Individual Link and zero if the LAG comprises one or more Aggregatable links,

then the general form of the unique LAG ID is [(SKP), (TLQ)].

To simplify comparison of LAG IDs it is conventional to order these such that S is the numerically smaller of S and T.

43.3.6.2 Representation of the Link Aggregation Group Identifier

In order to allow for convenient transcription and interpretation by human network personnel, this standard provides a convention for representing compound LAG IDs. Using this format

a) All fields are written as hexadecimal numbers, 2 digits per octet, in canonical format.
b) Octets are presented in order, from left to right. Within fields carrying numerical significance (e.g., priority values), the most significant octet is presented first, and the least significant octet last.
c) Within fields that carry MAC addresses, successive octets are separated by dashes (-), in accordance with the hexadecimal representation for MAC addresses defined in IEEE Std 802-1990™ .
d) Parameters of the LAG ID are separated by commas.

For example, consider the parameters for the two Partners in a Link Aggregation Group shown in Table 43–1.

Table 43–1—Example Partner Parameters

	Partner SKP	Partner TLQ
System Parameters (S, T)	System Priority = 0x8000 (see 43.4.2.2) System Identifier = AC-DE-48-03-67-80	System Priority = 0x8000 (see 43.4.2.2) System Identifier = AC-DE-48-03-FF-FF
Key Parameter (K, L)	Key = 0x0001	Key = 0x00AA
Port Parameters (P, Q)	Port Priority = 0x80 (see 43.4.2.2) Port Number = 0x0002	Port Priority = 0x80 (see 43.4.2.2) Port Number = 0x0002

The complete LAG ID derived from this information is represented as follows, for an Individual link:

[(SKP), (TLQ)] = [(8000,AC-DE-48-03-67-80,0001,80,0002), (8000,AC-DE-48-03-FF-FF,00AA,80,0002)]

The corresponding LAG ID for a set of Aggregatable links is represented as follows:

[(SKP), (TLQ)] = [(8000,AC-DE-48-03-67-80,0001,00,0000), (8000,AC-DE-48-03-FF-FF,00AA,00,0000)]

NOTE—The difference between the two representations is that, for an Aggregatable link, the port identifier components are zero.

It is recommended that this format be used whenever displaying LAG ID information for use by network personnel.

43.3.7 Selecting a Link Aggregation Group

Each port is selected for membership in the Link Aggregation Group uniquely identified by the LAG ID (composed of operational information, both derived from local administrative parameters and received through the Link Aggregation Control Protocol). Initial determination of the LAG ID is delayed to allow receipt of such information from a peer Link Aggregation Control entity; in the event such information is not received, locally configured administrative defaults are assumed for the remote port's operational parameters.

Where a particular link is known to be Individual, the complete LAG ID is not required to select the Link Aggregation Group since the link will not be aggregated with any other.

43.3.8 Agreeing on a Link Aggregation Group

Before frames are distributed and collected from a link, both the local Link Aggregation Control entity and its remote peer (if present) need to agree on the Link Aggregation Group. The Link Aggregation Control Protocol allows each of the communicating entities to check their peer's current understanding of the LAG ID, and facilitates rapid exchange of operational parameters while that understanding differs from their own. The protocol entities monitor their operation and, if agreement is not reached (perhaps due to an implementation failure), management is alerted.

The ability of LACP to signal that a particular link is Individual can accelerate the use of the link since, if both Link Aggregation Control entities know that the link is Individual, full agreement on the LAG ID is not necessary.

43.3.9 Attaching a link to an Aggregator

Once a link has selected a Link Aggregation Group, Link Aggregation Control can attach that link to a compatible Aggregator. An Aggregator is compatible if

a) The Aggregator's operational Key matches the port's operational Key, and
b) All other links currently attached to the Aggregator have selected the same Link Aggregation Group.

If several compatible Aggregators exist, Link Aggregation Control may employ a locally determined algorithm, either to ensure deterministic behavior (i.e., independence from the order in which Aggregators become available) or to maximize availability of the aggregation to a MAC Client. If no compatible Aggregator exists, then it is not possible to enable the link until such a time as a compatible Aggregator becomes available.

NOTE—In a properly configured System, there should always be a suitable Aggregator available with the proper Key assigned to serve a newly created Link Aggregation Group, so the unavailability of a compatible Aggregator is normally a temporary state encountered while links are moved between Aggregators. However, given the flexibility of the Key scheme, and given that in some implementations there may not be enough Aggregators to service a given configuration of links, it is possible to create configurations in which there is no Aggregator available to serve a newly identified LAG, in which case the links that are members of that LAG cannot become active until such a time as the configuration is changed to free up an appropriate Aggregator.

Links that are not successful candidates for aggregation (e.g., links that are attached to other devices that cannot perform aggregation or links that have been manually configured to be non-aggregatable) are enabled to operate as individual IEEE 802.3® links. For consistency of modeling, such a link is regarded as being attached to a compatible Aggregator that can only be associated with a single link. That is, from the perspective of Link Aggregation, non-aggregated links are not a special case; they compose an aggregation with a maximum membership of one link.

More than one link can select the same Link Aggregation Group within a short period of time and, as these links detach from their prior Aggregators, additional compatible Aggregators can become available. In order to avoid such events causing repeated configuration changes, Link Aggregation Control applies hysteresis to the attachment process and allows multiple links to be attached to an Aggregator at the same time.

43.3.10 Signaling readiness to transfer user data

Once a link has been attached to an Aggregator (43.3.9) compatible with the agreed-upon Link Aggregation Group (43.3.8), each Link Aggregation Control entity signals to its peer its readiness to transfer user data to and from the Aggregator's MAC Client. In addition to allowing time for the organization of local Aggregator resources, including the possibility that a compatible Aggregator may not exist, explicit signaling of readiness to transfer user data can be delayed to ensure preservation of frame ordering and prevention of frame duplication. Link Aggregation Control will not signal readiness until it is certain that there are no frames in transit on the link that were transmitted while the link was a member of a previous Link Aggregation Group. This may involve the use of an explicit Marker protocol that ensures that no frames remain to be received at either end of the link before reconfiguration takes place. The operation of the Marker protocol is described in 43.5. The decision as to when, or if, the Marker protocol is used is entirely dependent upon the nature of the distribution algorithm that is employed.

43.3.11 Enabling Collection and Distribution

Every Aggregator can enable or disable Collection and Distribution of frames for each port that is attached to the Aggregator. Initially, both Collection and Distribution are disabled. Once the Link Aggregation Control entity is ready to transfer user data using the link and its peer entity has also signaled readiness, the process of enabling the link can proceed. The Collector is enabled (thus preparing it to receive frames sent over the link by the remote Aggregator's Distributor) and that fact is communicated to the Partner. Once the received information indicates that the remote Aggregator's Collector is enabled, the Distributor is also enabled.

NOTE—This description assumes that the implementation is capable of controlling the state of the transmit and receive functions of the MAC independently. In an implementation where this is not possible, the transmit and receive functions are enabled or disabled together. The manner in which this is achieved is detailed in the description of the Mux machine (see 43.4.15).

If at least one port's Mux in the Link Aggregation Group is Collecting, then the Receive state of the corresponding Aggregator will be Enabled. If at least one port's Mux in the Link Aggregation Group is Distributing, then the Transmit state of the corresponding Aggregator will be Enabled.

43.3.12 Monitoring the membership of a Link Aggregation Group

Each link is monitored in order to confirm that the Link Aggregation Control functions at each end of the link still agree on the configuration information for that link. If the monitoring process detects a change in configuration that materially affects the link's membership in its current LAG, then it may be necessary to remove the link from its current LAG and to move it to a new LAG.

43.3.13 Detaching a link from an Aggregator

A port may be detached from the Aggregator used by its Link Aggregation Group as a result of protocol (e.g., Key) changes, or because of System constraints (e.g., exceeding a maximum allowable number of aggregated links, or device failures) at either end of the link. Both classes of events will cause the LAG ID information for the link to change, and it will be necessary for Link Aggregation Control to detach the link from its current Aggregator and move it to a new LAG (if possible). At the point where the change is detected, the Collecting and Distributing states for the port are set to FALSE. The Frame Distribution function is informed that the link is no longer part of the group, the changed configuration information is

communicated to the corresponding Link Aggregation Partner, then the Frame Collection function is informed that the link is no longer part of the group.

Once a link has been removed from its Aggregator, the link can select its new Link Aggregation Group and then attach to a compatible Aggregator, as described in 43.3.7 and 43.3.9.

Any conversation that is reallocated to a different link as a result of detaching a link from an Aggregator shall have its frame ordering preserved. This may involve the use of the Marker protocol to ensure that no frames that form part of that conversation remain to be received at either end of the old link before the conversation can proceed on the new link.

43.3.14 Configuration and administrative control of Link Aggregation

Administrative configuration facilities allow a degree of control to be exerted over the way that links may be aggregated. In particular, administrative configuration allows

a) The Key values associated with a port to be identified or modified.
b) The Key values associated with an Aggregator to be identified or modified.
c) Links to be identified as being incapable of aggregation.
d) Link Aggregation Control Protocol parameters to be identified or modified.

43.3.15 Link Aggregation Control state information

The Link Aggregation Control function maintains the following information with respect to each link:

a) The identifier of the Link Aggregation Group to which it currently belongs.
b) The identifier of the Aggregator associated with that Link Aggregation Group.
c) The status of interaction between the Frame Collection function of the Aggregator and the link (Collecting TRUE or Collecting FALSE). Collecting TRUE indicates that the receive function of this link is enabled with respect to its participation in an aggregation; i.e., received frames will be passed up to the Aggregator for collection.
d) The status of interaction between the Frame Distribution function of the Aggregator and the link (Distributing TRUE or Distributing FALSE). Distributing TRUE indicates that the transmit function of this link is enabled with respect to its participation in an aggregation; i.e., frames may be passed down from the Aggregator's distribution function for transmission.

This state information is communicated directly between Link Aggregation Control and the Aggregator through shared state variables without the use of a formal service interface.

The Link Aggregation Control function maintains the following information with respect to each Aggregator:

e) The status of the Frame Collection function (Receive Enabled or Receive Disabled).
f) The status of the Frame Distribution function (Transmit Enabled or Transmit Disabled).

These status values are exactly the logical OR of the Collection and Distribution status of the individual links associated with that Aggregator; i.e., if one or more links in the Link Aggregation Group are Collecting, then the Aggregator is Receive Enabled, and if one or more links are Distributing, then the Aggregator is Transmit Enabled.

The Transmit and Receive status of the Aggregator effectively govern the point at which the Aggregator becomes available for use by the MAC Client, or conversely, the point at which it ceases to be available.

43.4 Link Aggregation Control Protocol (LACP)

The Link Aggregation Control Protocol (LACP) provides a standardized means for exchanging information between Partner Systems on a link to allow their Link Aggregation Control instances to reach agreement on the identity of the Link Aggregation Group to which the link belongs, move the link to that Link Aggregation Group, and enable its transmission and reception functions in an orderly manner.

43.4.1 LACP design elements

The following considerations were taken into account during the development of the protocol described in this subclause:

a) The protocol depends upon the transmission of information and state, rather than the transmission of commands. LACPDUs sent by the first party (the Actor) convey to the second party (the Actor's protocol Partner) what the Actor knows, both about its own state and that of the Partner.
b) The information conveyed in the protocol is sufficient to allow the Partner to determine what action to take next.
c) Active or passive participation in LACP is controlled by LACP_Activity, an administrative control associated with each port, that can take the value Active LACP or Passive LACP. Passive LACP indicates the port's preference for not transmitting LACPDUs unless its Partner's control value is Active LACP (i.e., a preference not to speak unless spoken to). Active LACP indicates the port's preference to participate in the protocol regardless of the Partner's control value (i.e., a preference to speak regardless).
d) Periodic transmission of LACPDUs occurs if the LACP_Activity control of either the Actor or the Partner is Active LACP. These periodic transmissions will occur at either a slow or fast transmission rate depending upon the expressed LACP_Timeout preference (Long Timeout or Short Timeout) of the Partner System.
e) In addition to periodic LACPDU transmissions, the protocol transmits LACPDUs when there is a Need To Transmit (NTT) something to the Partner; i.e., when the Actor's state changes or when it is apparent from the Partner's LACPDUs that the Partner does not know the Actor's current state.
f) The protocol assumes that the rate of LACPDU loss is very low.

There is no explicit frame loss detection/retry mechanism employed by the LACP; however, if information is received from the Partner indicating that it does not have up-to-date information on the Actor's state, or if the next periodic transmission is due, then the Actor will transmit a LACPDU that will correctly update the Partner.

43.4.2 LACPDU structure and encoding

43.4.2.1 Transmission and representation of octets

All LACPDUs comprise an integral number of octets. The bits in each octet are numbered from 0 to 7, where 0 is the low-order bit. When consecutive octets are used to represent a numerical value, the most significant octet is transmitted first, followed by successively less significant octets.

When the encoding of (an element of) a LACPDU is depicted in a diagram

a) Octets are transmitted from top to bottom.
b) Within an octet, bits are shown with bit 0 to the left and bit 7 to the right, and are transmitted from left to right.
c) When consecutive octets are used to represent a binary number, the octet transmitted first has the more significant value.

d) When consecutive octets are used to represent a MAC address, the least significant bit of the first octet is assigned the value of the first bit of the MAC address, the next most significant bit the value of the second bit of the MAC address, and so on through the eighth bit. Similarly the least significant through most significant bits of the second octet are assigned the value of the ninth through seventeenth bits of the MAC address, and so on for all the octets of the MAC address.

43.4.2.2 LACPDU structure

LACPDUs are basic IEEE 802.3® frames; they shall not be tagged (See Clause 3). The LACPDU structure shall be as shown in Figure 43–7 and as further described in the following field definitions:

a) *Destination Address (DA)*. The DA in LACPDUs is the Slow_Protocols_Multicast address. Its use and encoding are specified in Annex 43B.

b) *Source Address (SA)*. The SA in LACPDUs carries the individual MAC address associated with the port through which the LACPDU is transmitted.

c) *Length/Type*. LACPDUs are always Type encoded, and carry the Slow_Protocols_Type field value. The use and encoding of this type is specified in Annex 43B.

d) *Subtype*. The Subtype field identifies the specific Slow Protocol being encapsulated. LACPDUs carry the Subtype value 0x01.

e) *Version Number.* This identifies the LACP version; implementations conformant to this version of the standard carry the value 0x01.

f) *TLV_type = Actor Information*. This field indicates the nature of the information carried in this TLV-tuple. Actor information is identified by the value 0x01.

g) *Actor_Information_Length*. This field indicates the length (in octets) of this TLV-tuple, Actor information uses a length value of 20 (0x14).

h) *Actor_System_Priority*. The priority assigned to this System (by management or administration policy), encoded as an unsigned integer.

i) *Actor_System*. The Actor's System ID, encoded as a MAC address.

j) *Actor_Key*. The operational Key value assigned to the port by the Actor, encoded as an unsigned integer.

k) *Actor_Port_Priority*. The priority assigned to this port by the Actor (the System sending the PDU; assigned by management or administration policy), encoded as an unsigned integer.

l) *Actor_Port*. The port number assigned to the port by the Actor (the System sending the PDU), encoded as an unsigned integer.

m) *Actor_State*. The Actor's state variables for the port, encoded as individual bits within a single octet, as follows and as illustrated in Figure 43–8:
   1) *LACP_Activity* is encoded in bit 0. This flag indicates the Activity control value with regard to this link. Active LACP is encoded as a 1; Passive LACP is encoded as a 0.
   2) *LACP_Timeout* is encoded in bit 1. This flag indicates the Timeout control value with regard to this link. Short Timeout is encoded as a 1; Long Timeout is encoded as a 0.
   3) *Aggregation* is encoded in bit 2. If TRUE (encoded as a 1), this flag indicates that the System considers this link to be *Aggregatable*; i.e., a potential candidate for aggregation. If FALSE (encoded as a 0), the link is considered to be *Individual*; i.e., this link can be operated only as an individual link.
   4) *Synchronization* is encoded in bit 3. If TRUE (encoded as a 1), the System considers this link to be *IN_SYNC*; i.e., it has been allocated to the correct Link Aggregation Group, the group has been associated with a compatible Aggregator, and the identity of the Link Aggregation Group is consistent with the System ID and operational Key information transmitted. If FALSE (encoded as a 0), then this link is currently *OUT_OF_SYNC;* i.e., it is not in the right Aggregation.

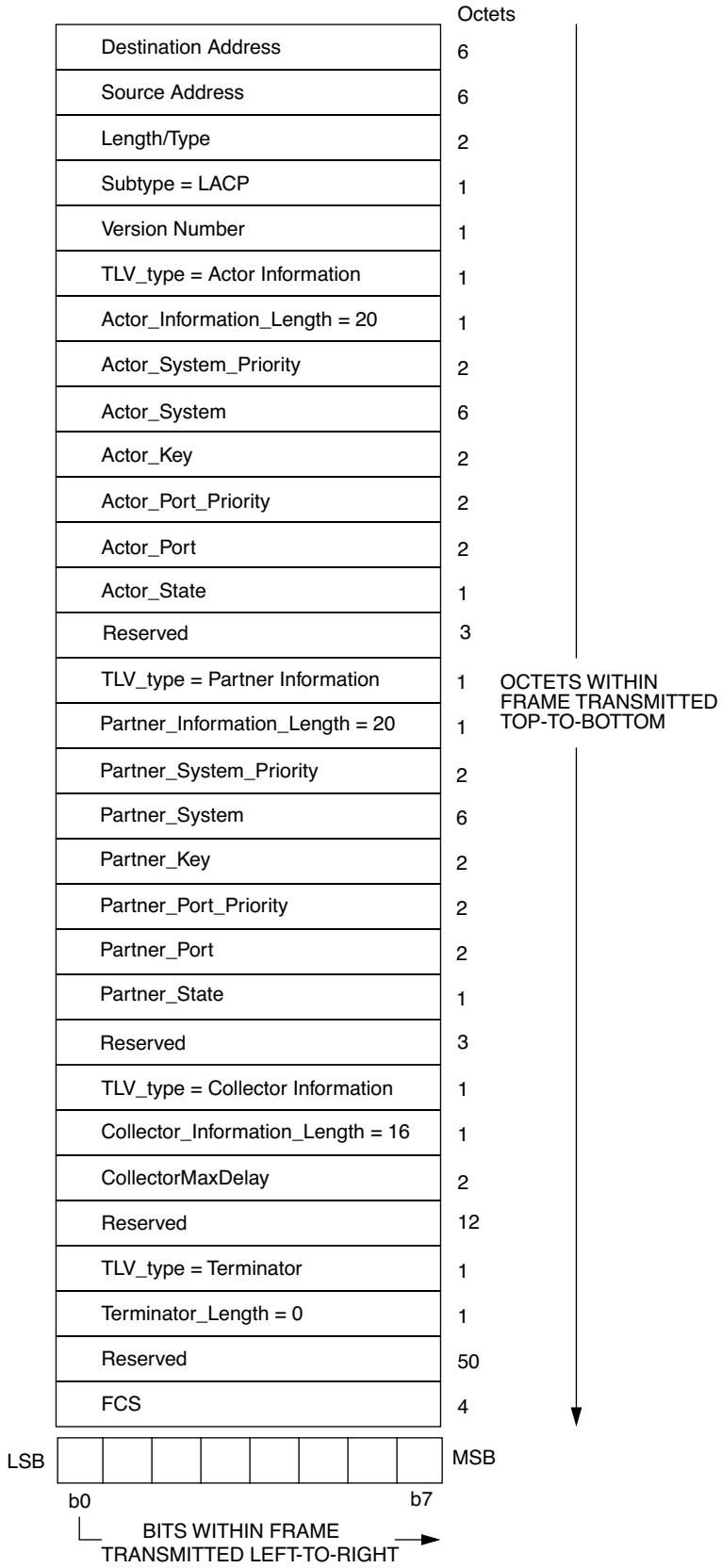
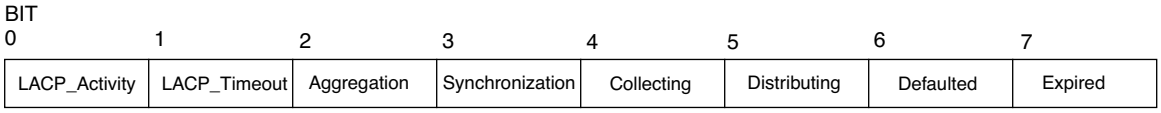


Figure 43–7—LACPDU structure

5) *Collecting* is encoded in bit 4. TRUE (encoded as a 1) means collection of incoming frames on this link is definitely enabled; i.e., collection is currently enabled and is not expected to be disabled in the absence of administrative changes or changes in received protocol information. Its value is otherwise FALSE (encoded as a 0);
6) *Distributing* is encoded in bit 5. FALSE (encoded as a 0) means distribution of outgoing frames on this link is definitely disabled; i.e., distribution is currently disabled and is not expected to be enabled in the absence of administrative changes or changes in received protocol information. Its value is otherwise TRUE (encoded as a 1);
7) *Defaulted* is encoded in bit 6. If TRUE (encoded as a 1), this flag indicates that the Actor's Receive machine is using Defaulted operational Partner information, administratively configured for the Partner. If FALSE (encoded as a 0), the operational Partner information in use has been received in a LACPDU;
8) *Expired* is encoded in bit 7. If TRUE (encoded as a 1), this flag indicates that the Actor's Receive machine is in the EXPIRED state; if FALSE (encoded as a 0), this flag indicates that the Actor's Receive machine is not in the EXPIRED state.

NOTE—The received values of Defaulted and Expired state are not used by LACP; however, knowing their values can be useful when diagnosing protocol problems.



NOTE—Bit ordering within this field is as specified in 43.4.2.1.

Figure 43–8—Bit encoding of the Actor_State and Partner_State fields

n) *Reserved*. These 3 octets are reserved for use in future extensions to the protocol. They shall be ignored on receipt and shall be transmitted as zeroes to claim compliance with Version 1 of this protocol.
o) *TLV_type = Partner Information*. This field indicates the nature of the information carried in this TLV-tuple. Partner information is identified by the integer value 0x02.
p) *Partner_Information_Length*. This field indicates the length (in octets) of this TLV-tuple, Partner information uses a length value of 20 (0x14).
q) *Partner_System_Priority*. The priority assigned to the Partner System (by management or administration policy), encoded as an unsigned integer.
r) *Partner_System*. The Partner's System ID, encoded as a MAC address.
s) *Partner_Key*. The operational Key value assigned to the port associated with this link by the Partner, encoded as an unsigned integer.
t) *Partner_Port_Priority*. The priority assigned to this port by the Partner (by management or administration policy), encoded as an unsigned integer.
u) *Partner_Port*. The port number associated with this link assigned to the port by the Partner, encoded as an unsigned integer.
v) *Partner_State*. The Actor's view of the Partner's state variables, depicted in Figure 43–8 and encoded as individual bits within a single octet, as defined for Actor_State.
w) *Reserved*. These 3 octets are reserved for use in future extensions to the protocol. They shall be ignored on receipt and shall be transmitted as zeroes to claim compliance with Version 1 of this protocol.
x) *TLV_type = Collector Information*. This field indicates the nature of the information carried in this TLV-tuple. Collector information is identified by the integer value 0x03.

y) *Collector_Information_Length*. This field indicates the length (in octets) of this TLV-tuple. Collector information uses a length value of 16 (0x10).

z) *CollectorMaxDelay*. This field contains the value of CollectorMaxDelay (43.2.3.1.1) of the station transmitting the LACPDU, encoded as an unsigned integer number of tens of microseconds. The range of values for this parameter is 0 to 65 535 tens of microseconds (0.65535 seconds).

aa) *Reserved*. These 12 octets are reserved for use in future extensions to the protocol. They shall be ignored on receipt and shall be transmitted as zeroes to claim compliance with Version 1 of this protocol.

ab) *TLV_type = Terminator*. This field indicates the nature of the information carried in this TLV-tuple. Terminator (end of message) information is identified by the integer value 0x00.

ac) *Terminator_Length*. This field indicates the length (in octets) of this TLV-tuple. Terminator information uses a length value of 0 (0x00).

NOTE—The use of a Terminator_Length of 0 is intentional. In TLV encoding schemes it is common practice for the terminator encoding to be 0 both for the type and the length.

ad) *Reserved*. These 50 octets are reserved for use in future extensions to the protocol. They are ignored on receipt and are transmitted as zeroes to claim compliance with Version 1 of this protocol.

NOTE—The Reserved octets are included in all valid LACPDUs in order to force the TLV lengths to multiples of 4 octets, and to force a fixed PDU size of 128 octets, regardless of the version of the protocol. Hence, a Version 1 implementation is guaranteed to be able to receive version N PDUs successfully, although version N PDUs may contain additional information that cannot be interpreted (and will be ignored) by the Version 1 implementation. A crucial factor in ensuring backwards compatibility is that any future version of the protocol is required not to re-define the structure or semantics of information defined for the previous version; it may only add new information elements to the previous set. Hence, in a version N PDU, a Version 1 implementation can expect to find the Version 1 information in exactly the same places as in a Version 1 PDU, and can expect to interpret that information as defined for Version 1. Future versions of this protocol expect to have the Reserved octets available for their use.

ae) *FCS*. This field is the Frame Check Sequence, typically generated by the underlying MAC.

43.4.3 LACP state machine overview

The operation of the protocol is controlled by a number of state machines, each of which performs a distinct function. These state machines are for the most part described on a per-port basis; any deviations from per-port description are highlighted in the text. Events (such as expiration of a timer or received LACPDUs) may cause state transitions and also cause actions to be taken; those actions may include the need for transmission of a LACPDU containing repeated or new information. Periodic and event-driven transmissions are controlled by the state of a Need-To-Transmit (NTT) variable (see 43.4.7), generated by the state machines as necessary.

The state machines are as follows:

a) *Receive machine (RX—43.4.12)*. This state machine receives LACPDUs from the Partner, records the information contained, and times it out using either Short Timeouts or Long Timeouts, according to the setting of LACP_Timeout. It evaluates the incoming information from the Partner to determine whether the Actor and Partner have both agreed upon the protocol information exchanged to the extent that the port can now be safely used, either in an aggregation with other ports or as an individual port; if not, it asserts NTT in order to transmit fresh protocol information to the Partner. If the protocol information from the Partner times out, the Receive machine installs default parameter values for use by the other state machines.

b) *Periodic Transmission machine (43.4.13)*. This state machine determines whether the Actor and its Partner will exchange LACPDUs periodically in order to maintain an aggregation (periodic LACPDU exchanges occur if either or both are configured for Active LACP).

c) *Selection Logic (43.4.14)*. The Selection Logic is responsible for selecting the Aggregator to be associated with this port.

d) *Mux machine (MUX—43.4.15)*. This state machine is responsible for attaching the port to a selected Aggregator, detaching the port from a de-selected Aggregator, and for turning collecting and distributing at the port on or off as required by the current protocol information.

e) *Transmit machine (TX—43.4.16)*. This state machine handles the transmission of LACPDUs, both on demand from the other state machines, and on a periodic basis.

Figure 43–9 illustrates the relationships among these state machines and the flow of information between them. The set of arrows labelled Partner State Information represents new Partner information, contained in an incoming LACPDU or supplied by administrative default values, being fed to each state machine by the Receive machine. The set of arrows labelled Actor State Information represents the flow of updated Actor state information between the state machines. Transmission of LACPDUs occurs either as a result of the Periodic machine determining the need to transmit a periodic LACPDU, or as a result of changes to the Actor's state information that need to be communicated to the Partner. The need to transmit a LACPDU is signalled to the Transmit machine by asserting NTT. The remaining arrows represent shared variables in the state machine description that allow a state machine to cause events to occur in another state machine.

NOTE—The arrows marked Ready_N show that information derived from the operation of another port or ports can affect the operation of a port's state machine. See the definition of the Ready and Ready_N variables in 43.4.8.

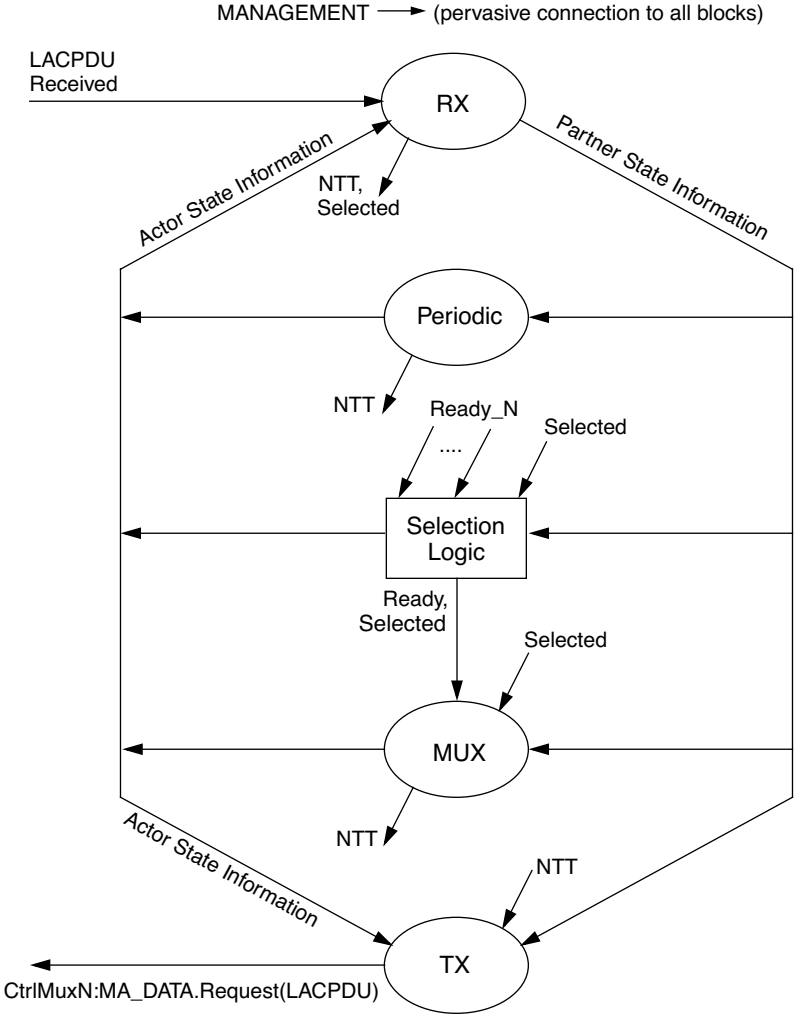


Figure 43–9—Interrelationships among state machines

Two further state machines are defined for diagnostic purposes. They are as follows:

f) *Actor and Partner Churn Detection machines (43.4.17)*. These state machines make use of the IN_SYNC and OUT_OF_SYNC states generated by the Actor and Partner Mux machines in order to detect the situation where the state machines are unable to resolve the state of a given link; e.g., because the Partner System repeatedly sends conflicting information in its LACPDUs. As this situation can occur in a normally functioning link, particularly where either or both participating Systems have constrained aggregation capability (see 43.6), these state machines simply detect the presence of such a condition and signal its existence to management.

43.4.4 Constants

All timers specified in this subclause have an implementation tolerance of ± 250 ms.

Fast_Periodic_Time
The number of seconds between periodic transmissions using Short Timeouts.
Value: Integer
1

Slow_Periodic_Time
The number of seconds between periodic transmissions using Long Timeouts.
Value: Integer
30

Short_Timeout_Time
The number of seconds before invalidating received LACPDU information when using Short Timeouts (3 x Fast_Periodic_Time).
Value: Integer
3

Long_Timeout_Time
The number of seconds before invalidating received LACPDU information when using Long Timeouts (3 x Slow_Periodic_Time).
Value: Integer
90

Churn_Detection_Time
The number of seconds that the Actor and Partner Churn state machines wait for the Actor or Partner Sync state to stabilize.
Value: Integer
60

Aggregate_Wait_Time
The number of seconds to delay aggregation, to allow multiple links to aggregate simultaneously.
Value: Integer
2

43.4.5 Variables associated with the System

Actor_System
The MAC address component of the System Identifier of the System.
Value: 48 bits
Assigned by administrator or System policy.

Actor_System_Priority
The System Priority of the System.
Value: Integer
Assigned by administrator or System policy.

43.4.6 Variables associated with each Aggregator

Aggregator_MAC_address
The MAC address assigned to the Aggregator.
Value: 48 bits
Assigned by administrator or System policy.

Aggregator_Identifier
Used to uniquely identify an Aggregator within a System.
Value: Integer
Assigned by administrator or System policy.

Individual_Aggregator
The aggregation capability of the Aggregator.
Value: Boolean
TRUE if the port attached to this Aggregator is not capable of aggregation with any other port.
FALSE if the port(s) attached to this Aggregator are capable of aggregation with other ports.

Actor_Admin_Aggregator_Key
The administrative Key value associated with the Aggregator.
Value: Integer
Assigned by administrator or System policy.

Actor_Oper_Aggregator_Key
The operational Key value associated with the Aggregator.
Value: Integer
Assigned by the Actor.

Partner_System
The MAC address component of the System Identifier of the remote System to which the Aggregator is connected. If the Aggregator has no attached ports, this variable is set to 0x00-00-00-00-00-00.
Value: 48 bits

Partner_System_Priority
The System Priority of the remote System to which the Aggregator is connected. If the Aggregator has no attached ports, this variable is set to zero.
Value: Integer

Partner_Oper_Aggregator_Key
The operational Key assigned to an aggregation by the remote System to which this Aggregator is connected. If the Aggregator has no attached ports, this variable is set to zero.
Value: Integer

Receive_State
The Receive_State of the Aggregator will be Enabled if one or more ports attached to the Aggregator are Collecting (i.e., Actor_Oper_Port_State.Collecting is TRUE for any port). Otherwise, Receive_State is Disabled.
Values: Enabled or Disabled

Transmit_State
The Transmit_State of the Aggregator will be Enabled if one or more ports attached to the Aggregator are Distributing (i.e., Actor_Oper_Port_State.Distributing is TRUE for any port). Otherwise, Transmit_State is Disabled.
Values: Enabled or Disabled

LAG_Ports
The set of ports that belong to the Link Aggregation Group.
Value: Integer Array

43.4.7 Variables associated with each port

Actor_Port_Number
The port number assigned to the port.
Value: Integer
Assigned by administrator or System policy.

Actor_Port_Priority
The priority value assigned to the port, used to converge dynamic Key changes.
Value: Integer
Assigned by administrator or System policy.

Actor_Port_Aggregator_Identifier
The identifier of the Aggregator that this port is attached to.
Value: Integer

NTT
Need To Transmit flag.
Value: Boolean
TRUE indicates that there is new protocol information that should be transmitted on the link, or that the Partner needs to be reminded of the old information.
FALSE otherwise.

Actor_Admin_Port_Key
The administrative value of Key assigned to this port by administrator or System policy.
Value: Integer

Actor_Oper_Port_Key
The operational value of Key assigned to this port by the Actor.
Value: Integer

Actor_Admin_Port_State
The administrative values of the Actor's state parameters. This consists of the following set of variables, as described in 43.4.2.2:
LACP_Activity
LACP_Timeout
Aggregation
Synchronization

Collecting
Distributing
Defaulted
Expired
Value: 8 bits

Actor_Oper_Port_State
The operational values of the Actor's state parameters. This consists of the following set of variables, as described in 43.4.2.2:
LACP_Activity
LACP_Timeout
Aggregation
Synchronization
Collecting
Distributing
Defaulted
Expired
Value: 8 bits

Partner_Admin_System
Default value for the MAC address component of the System Identifier of the Partner, assigned by administrator or System policy for use when the Partner's information is unknown or expired.
Value: 48 bits

Partner_Oper_System
The operational value of the MAC address component of the System Identifier of the Partner. The Actor sets this variable either to the value received from the Partner in an LACPDU, or to the value of Partner_Admin_System.
Value: 48 bits

Partner_Admin_System_Priority
Default value for the System Priority component of the System Identifier of the Partner, assigned by administrator or System policy for use when the Partner's information is unknown or expired.
Value: Integer

Partner_Oper_System_Priority
The operational value of the System Priority of the Partner. The Actor sets this variable either to the value received from the Partner in an LACPDU, or to the value of Partner_Admin_System_Priority.
Value: Integer

Partner_Admin_Key
Default value for the Partner's Key, assigned by administrator or System policy for use when the Partner's information is unknown or expired.
Value: Integer

Partner_Oper_Key
The operational value of the Key value assigned to this link by the Partner. The Actor sets this variable either to the value received from the Partner in an LACPDU, or to the value of Partner_Admin_Key.
Value: Integer

Partner_Admin_Port_Number

Default value for the Port Number component of the Partner's Port Identifier, assigned by administrator or System policy for use when the Partner's information is unknown or expired.
Value: Integer

Partner_Oper_Port_Number

The operational value of the port number assigned to this link by the Partner. The Actor sets this variable either to the value received from the Partner in an LACPDU, or to the value of Partner_Admin_Port_Number.
Value: Integer

Partner_Admin_Port_Priority

Default value for the Port Priority component of the Partner's Port Identifier, assigned by administrator or System policy for use when the Partner's information is unknown or expired.
Value: Integer

Partner_Oper_Port_Priority

The operational value of the priority value assigned to this link by the Partner, used to converge dynamic Key changes. The Actor sets this variable either to the value received from the Partner in an LACPDU, or to the value of Partner_Admin_Port_Priority.
Value: Integer

Partner_Admin_Port_State

Default value for the Partner's state parameters, assigned by administrator or System policy for use when the Partner's information is unknown or expired. The value consists of the following set of variables, as described in 43.4.2.2:

LACP_Activity
LACP_Timeout
Aggregation
Synchronization
Collecting
Distributing
Defaulted
Expired

The value of Collecting shall be set the same as the value of Synchronization.
Value: 8 bits

Partner_Oper_Port_State

The operational value of the Actor's view of the current values of the Partner's state parameters. The Actor sets this variable either to the value received from the Partner in an LACPDU, or to the value of Partner_Admin_Port_State. The value consists of the following set of variables, as described in 43.4.2.2:

LACP_Activity
LACP_Timeout
Aggregation
Synchronization
Collecting
Distributing
Defaulted
Expired

Value: 8 bits

port_enabled
A variable indicating that the physical layer has indicated that the link has been established and the port is operable.
Value: Boolean
TRUE if the physical layer has indicated that the port is operable.
FALSE otherwise.

NOTE—The means by which the value of the port_enabled variable is generated by the underlying MAC and/or PHY is implementation-dependent.

43.4.8 Variables used for managing the operation of the state machines

BEGIN
This variable indicates the initialization (or reinitialization) of the LACP protocol entity. It is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.
Value: Boolean

LACP_Enabled
This variable indicates that the port is operating the LACP. If the port is operating in half duplex, the value of LACP_Enabled shall be FALSE. Otherwise, the value of LACP_Enabled shall be TRUE.
Value: Boolean

actor_churn
This variable indicates that the Actor Churn Detection machine has detected that a local port configuration has failed to converge within a specified time, and that management intervention is required.
Value: Boolean

partner_churn
This variable indicates that the Partner Churn Detection machine has detected that a remote port configuration has failed to converge within a specified time, and that management intervention is required.
Value: Boolean

Ready_N
The port asserts Ready_N TRUE to indicate to the Selection Logic that the wait_while_timer has expired and it is waiting (i.e., the port is in the WAITING state) to attach to an Aggregator. Otherwise, its value is FALSE. There is one Ready_N value for each port.
Value: Boolean

Ready
The Selection Logic asserts Ready TRUE when the values of Ready_N for all ports that are waiting to attach to a given Aggregator are TRUE. If any of the values of Ready_N for the ports that are waiting to attach to that Aggregator are FALSE, or if there are no ports waiting to attach to that Aggregator, then the value of Ready is FALSE.
Value: Boolean

Selected
A value of SELECTED indicates that the Selection Logic has selected an appropriate Aggregator. A value of UNSELECTED indicates that no aggregator is currently selected. A value of STANDBY indicates that although the Selection Logic has selected an appropriate Aggregator, aggregation restrictions currently prevent the port from being enabled as part of the aggrega-

tion, and so the port is being held in a standby condition. This variable can only be set to SELECTED or STANDBY by the operation of the port's Selection Logic. It can be set to UNSELECTED by the operation of the port's Receive machine, or by the operation of the Selection Logic associated with another port.

NOTE—Setting Selected UNSELECTED in the Selection Logic associated with another port occurs if the Selection Logic determines that the other port has a stronger claim to attach to this port's current Aggregator.

Value: SELECTED, UNSELECTED, or STANDBY

port_moved
This variable is set to TRUE if the Receive machine for a port is in the PORT_DISABLED state, and the combination of Partner_Oper_System and Partner_Oper_Port_Number in use by that port has been received in an incoming LACPDU on a different port. This variable is set to FALSE once the INITIALIZE state of the Receive machine has set the Partner information for the port to administrative default values.
Value: Boolean

43.4.9 Functions

recordPDU
This function records the parameter values for the Actor carried in a received LACPDU (Actor_Port, Actor_Port_Priority, Actor_System, Actor_System_Priority, Actor_Key, and Actor_State variables) as the current Partner operational parameter values (Partner_Oper_Port_Number, Partner_Oper_Port_Priority, Partner_Oper_System, Partner_Oper_System_Priority, Partner_Oper_Key, and Partner_Oper_Port_State variables with the exception of Synchronization) and sets Actor_Oper_Port_State.Defaulted to FALSE.

This function also updates the value of the Partner_Oper_Port_State.Synchronization using the parameter values carried in received LACPDUs. Parameter values for the Partner carried in the received PDU (Partner_Port, Partner_Port_Priority, Partner_System, Partner_System_Priority, Partner_Key, and Partner_State.Aggregation) are compared to the corresponding operational parameter values for the Actor (Actor_Port_Number, Actor_Port_Priority, Actor_System, Actor_System_Priority, Actor_Oper_Port_Key, and Actor_Oper_Port_State.Aggregation). Partner_Oper_Port_State.Synchronization is set to TRUE if all of these parameters match, Actor_State.Synchronization in the received PDU is set to TRUE, and LACP will actively maintain the link in the aggregation.

Partner_Oper_Port_State.Synchronization is also set to TRUE if the value of Actor_State.Aggregation in the received PDU is set to FALSE (i.e., indicates an Individual link), Actor_State.Synchronization in the received PDU is set to TRUE, and LACP will actively maintain the link.

Otherwise, Partner_Oper_Port_State.Synchronization is set to FALSE.

LACP is considered to be actively maintaining the link if either the PDU's Actor_State.LACP_Activity variable is TRUE or both the Actor's Actor_Oper_Port_State.LACP_Activity and the PDU's Partner_State.LACP_Activity variables are TRUE.

recordDefault
This function records the default parameter values for the Partner carried in the Partner Admin parameters (Partner_Admin_Port_Number, Partner_Admin_Port_Priority, Partner_Admin_System, Partner_Admin_System_Priority, Partner_Admin_Key, and

Partner_Admin_Port_State) as the current Partner operational parameter values (Partner_Oper_Port_Number, Partner_Oper_Port_Priority, Partner_Oper_System, Partner_Oper_System_Priority, Partner_Oper_Key, and Partner_Oper_Port_State) and sets Actor_Oper_Port_State.Defaulted to TRUE.

update_Selected

This function updates the value of the Selected variable, using parameter values from a newly received LACPDU. The parameter values for the Actor carried in the received PDU (Actor_Port, Actor_Port_Priority, Actor_System, Actor_System_Priority, Actor_Key, and Actor_State.Aggregation) are compared with the corresponding operational parameter values for the port's Partner (Partner_Oper_Port_Number, Partner_Oper_Port_Priority, Partner_Oper_System, Partner_Oper_System_Priority, Partner_Oper_Key, and Partner_Oper_Port_State.Aggregation). If one or more of the comparisons show that the value(s) received in the PDU differ from the current operational values, then Selected is set to UNSELECTED. Otherwise, Selected remains unchanged.

update_Default_Selected

This function updates the value of the Selected variable, using the Partner administrative parameter values. The administrative values (Partner_Admin_Port_Number, Partner_Admin_Port_Priority, Partner_Admin_System, Partner_Admin_System_Priority, Partner_Admin_Key, and Partner_Admin_Port_State.Aggregation) are compared with the corresponding operational parameter values for the Partner (Partner_Oper_Port_Number, Partner_Oper_Port_Priority, Partner_Oper_System, Partner_Oper_System_Priority, Partner_Oper_Key, and Partner_Oper_Port_State.Aggregation). If one or more of the comparisons shows that the administrative value(s) differ from the current operational values, then Selected is set to UNSELECTED. Otherwise, Selected remains unchanged.

update_NTT

This function updates the value of the NTT variable, using parameter values from a newly received LACPDU. The parameter values for the Partner carried in the received PDU (Partner_Port, Partner_Port_Priority, Partner_System, Partner_System_Priority, Partner_Key, Partner_State.LACP_Activity, Partner_State.LACP_Timeout, Partner_State.Synchronization, and Partner_State.Aggregation) are compared with the corresponding operational parameter values for the Actor (Actor_Port_Number, Actor_Port_Priority, Actor_System, Actor_System_Priority, Actor_Oper_Port_Key, Actor_Oper_Port_State.LACP_Activity, Actor_Oper_Port_State.LACP_Timeout, Actor_Oper_Port_State.Synchronization, and Actor_Oper_Port_State.Aggregation). If one or more of the comparisons show that the value(s) received in the PDU differ from the current operational values, then NTT is set to TRUE. Otherwise, NTT remains unchanged.

Attach_Mux_To_Aggregator

This function causes the port's Control Parser/Multiplexer to be attached to the Aggregator Parser/Multiplexer of the selected Aggregator, in preparation for collecting and distributing frames.

Detach_Mux_From_Aggregator

This function causes the port's Control Parser/Multiplexer to be detached from the Aggregator Parser/Multiplexer of the Aggregator to which the port is currently attached.

Enable_Collecting

This function causes the Aggregator Parser of the Aggregator to which the port is attached to start collecting frames from the port.

Disable_Collecting
: This function causes the Aggregator Parser of the Aggregator to which the port is attached to stop collecting frames from the port.

Enable_Distributing
: This function causes the Aggregator Multiplexer of the Aggregator to which the port is attached to start distributing frames to the port.

Disable_Distributing
: This function causes the Aggregator Multiplexer of the Aggregator to which the port is attached to stop distributing frames to the port.

Enable_Collecting_Distributing
: This function causes the Aggregator Parser of the Aggregator to which the port is attached to start collecting frames from the port, and the Aggregator Multiplexer to start distributing frames to the port.

Disable_Collecting_Distributing
: This function causes the Aggregator Parser of the Aggregator to which the port is attached to stop collecting frames from the port, and the Aggregator Multiplexer to stop distributing frames to the port.

43.4.10 Timers

current_while_timer
: This timer is used to detect whether received protocol information has expired. If Actor_Oper_State.LACP_Timeout is set to Short Timeout, the timer is started with the value Short_Timeout_Time. Otherwise, it is started with the value Long_Timeout_Time (see 43.4.4).

actor_churn_timer
: This timer is used to detect Actor churn states. It is started using the value Churn_Detection_Time (see 43.4.4).

periodic_timer (time_value)
: This timer is used to generate periodic transmissions. It is started using the value Slow_Periodic_Time or Fast_Periodic_Time (see 43.4.4), as specified in the Periodic Transmission state machine.

partner_churn_timer
: This timer is used to detect Partner churn states. It is started using the value Churn_Detection_Time (see 43.4.4).

wait_while_timer
: This timer provides hysteresis before performing an aggregation change, to allow all links that will join this Aggregation to do so. It is started using the value Aggregate_Wait_Time (see 43.4.4).

43.4.11 Messages

CtrlMuxN:MA_DATA.indicate(LACPDU)
: This message is generated by the Control Parser as a result of the reception of a LACPDU, formatted as defined in 43.4.2.

43.4.12 Receive machine

The Receive machine shall implement the function specified in Figure 43–10 with its associated parameters (43.4.4 through 43.4.11).

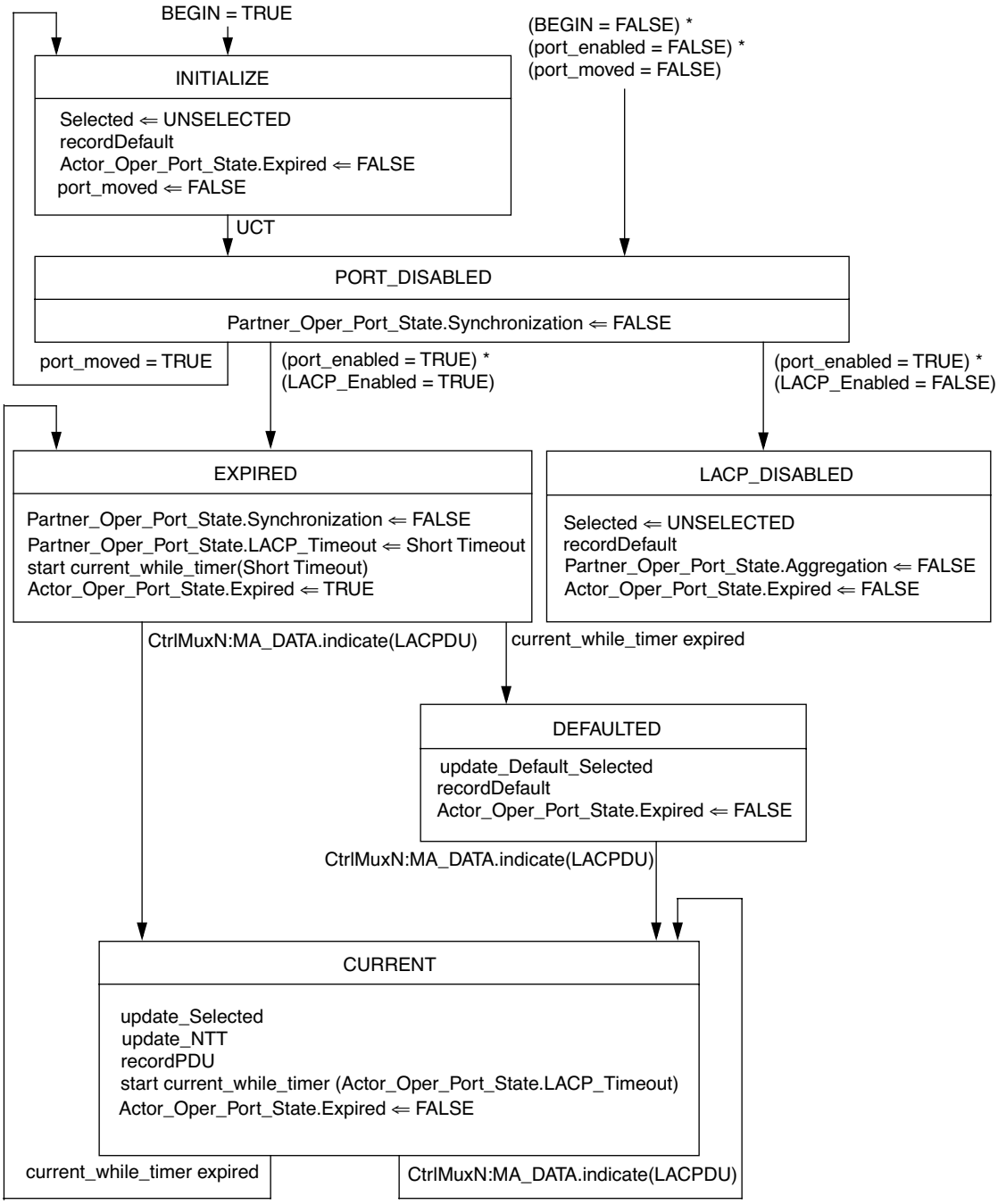


Figure 43–10—Receive machine state diagram

On receipt of a LACPDU, the state machine enters the CURRENT state. The update_Selected function sets the Selected variable UNSELECTED if the Link Aggregation Group identified by the combination of the protocol Partner's own information and the Actor's own information has changed. The Selected variable is used by the Mux machine (43.4.15).

NOTE—The Receive machine may set the Selected variable to UNSELECTED; however, setting this variable to SELECTED or STANDBY is the responsibility of the Selection Logic.

The update_NTT function is used to determine whether further protocol transmissions are required; NTT is set to TRUE if the Partner's view of the Actor's operational parameters is not up to date. The recordPDU function records the information contained in the LACPDU in the Partner operational variables, and the current_while timer is started. The value used to start the timer is either Short_Timeout_Time or Long_Timeout_Time, depending upon the Actor's operational value of LACP_Timeout.

In the process of executing the recordPDU function, a Receive machine compliant to this standard shall not validate the Version Number, TLV_type, or Reserved fields in received LACPDUs. The same actions are taken regardless of the values received in these fields. A Receive machine may validate the Actor_Information_Length, Partner_Information_Length, Collector_Information_Length, or Terminator_Length fields. These behaviors, together with the constraint on future protocol enhancements, are discussed in 43.4.2.2.

NOTE—The rules expressed above allow Version 1 devices to be compatible with future revisions of the protocol.

If no LACPDU is received before the current_while timer expires, the state machine transits to the EXPIRED state. The Partner_Oper_Port_State.Synchronization variable is set to FALSE, the current operational value of the Partner's LACP_Timeout variable is set to Short Timeout, and the current_while timer is started with a value of Short_Timeout_Time. This is a transient state; the LACP_Timeout settings allow the Actor to transmit LACPDUs rapidly in an attempt to re-establish communication with the Partner.

If no LACPDU is received before the current_while timer expires again, the state machine transits to the DEFAULTED state. The recordDefault function overwrites the current operational parameters for the Partner with administratively configured values. This allows configuration of aggregations and individual links when no protocol Partner is present, while still permitting an active Partner to override default settings. The update_Default_Selected function sets the Selected variable UNSELECTED if the Link Aggregation Group has changed. Since all operational parameters are now set to locally administered values, there can be no disagreement as to the Link Aggregation Group, so the Partner_Oper_Port_State.Synchronization variable is set to TRUE.

If the port becomes inoperable and a BEGIN event has not occurred, the state machine enters the PORT_DISABLED state. Partner_Oper_Port_State.Synchronization is set to FALSE. This state allows the current Selection state to remain undisturbed, so that, in the event that the port is still connected to the same Partner and Partner port when it becomes operable again, there will be no disturbance caused to higher layers by unneccessary re-configuration. If the same Actor System ID and Port are seen in a LACPDU received on a different Port (port_moved is set to TRUE), this indicates that the physical connectivity has changed, and causes the state machine to enter the INITIALIZE state. This state is also entered if a BEGIN event occurs.

The INITIALIZE state causes the administrative values of the Partner parameters to be used as the current operational values, and sets Selected to UNSELECTED. These actions force the Mux machine to detach the port from its current Aggregator. The variable port_moved is set to FALSE; if the entry to INITIALIZE occurred as a result of port_moved being set to TRUE, then the state machine will immediately transition back to the PORT_DISABLED state.

If the port is operating in half duplex, the operation of LACP is disabled on the port (LACP_Enabled is FALSE) and the LACP_DISABLED state is entered. This state is entered following a BEGIN or Port Enabled event. This state is similar to the DEFAULTED state, except that the port is forced to operate as an Individual port, as the value of Partner_Oper_Port_State.Aggregation is forced to Individual. Exit from this state occurs on a BEGIN or Port Disabled event.

43.4.13 Periodic Transmission machine

The Periodic Transmission machine shall implement the function specified in Figure 43–11 with its associated parameters (43.4.4 through 43.4.11).

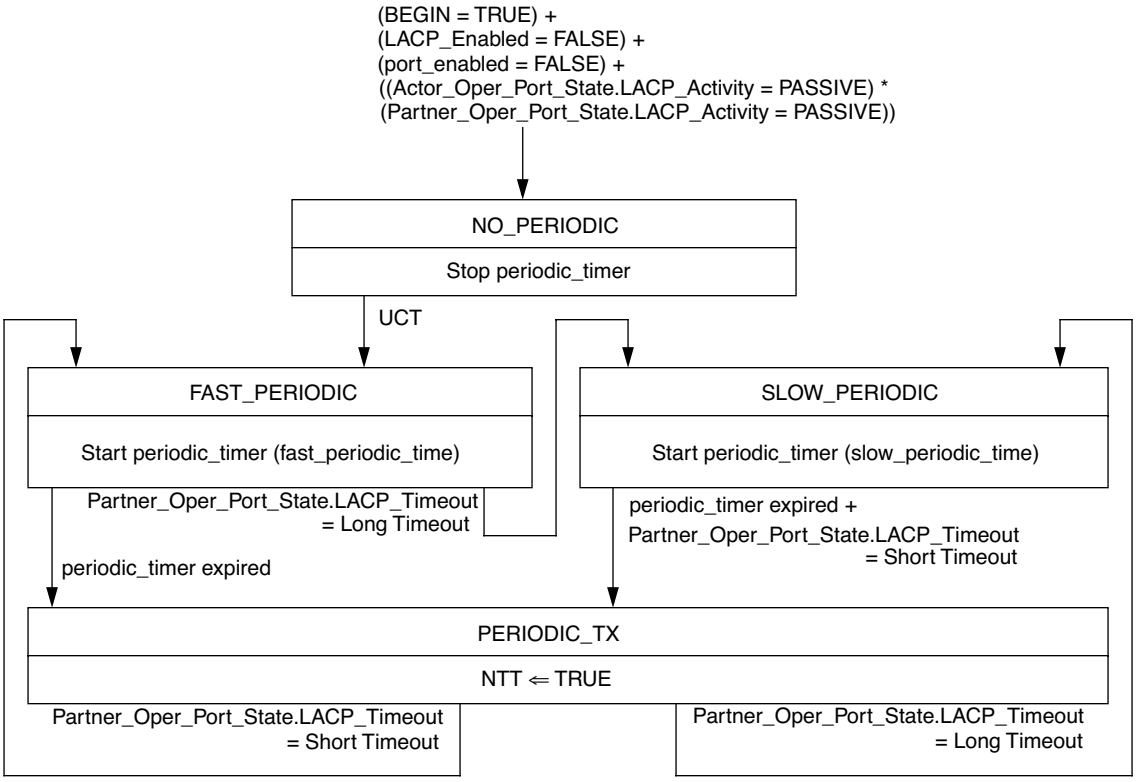


Figure 43–11—Periodic Transmission machine state diagram

The Periodic Transmission machine establishes the desire of the Actor and Partner to exchange periodic LACPDUs on the link in order to maintain an aggregation, and establishes how often those periodic transmissions should occur. Periodic transmissions will take place if either participant so wishes. Transmissions occur at a rate determined by the Partner; this rate is linked to the speed at which the Partner will time out received information.

The state machine has four states. They are as follows:

a) *NO_PERIODIC*. While in this state, periodic transmissions are disabled.
b) *FAST_PERIODIC*. While in this state, periodic transmissions are enabled at a fast transmission rate.
c) *SLOW_PERIODIC*. While in this state, periodic transmissions are enabled at a slow transmission rate.
d) *PERIODIC_TX*. This is a transitory state entered on periodic_timer expiry, that asserts NTT and then exits to FAST_PERIODIC or SLOW_PERIODIC depending upon the Partner's LACP_Timeout setting.

The values of Partner_Oper_Port_State.LACP_Activity and Actor_Oper_Port_State.LACP_Activity determine whether periodic transmissions take place. If either or both parameters are set to Active LACP, then periodic transmissions occur; if both are set to Passive LACP, then periodic transmissions do not occur. Similarly, if either of the LACP_Enabled or port_enabled variables is set to FALSE, indicating that LACP has been disabled on the port or that the port is non-operational, then no periodic transmissions take place.

If periodic transmissions are enabled, the rate at which they take place is determined by the value of the Partner_Oper_Port_State.LACP_Timeout variable. If this variable is set to Short Timeout, then the value fast_periodic_time is used to determine the time interval between periodic transmissions. Otherwise, slow_periodic_time is used to determine the time interval.

43.4.14 Selection Logic

The Selection Logic selects a compatible Aggregator for a port, using the port's LAG ID. The Selection Logic may determine that the link should be operated as a standby link if there are constraints on the simultaneous attachment of ports that have selected the same Aggregator.

NOTE—There will never be more than one Aggregator with the same LAG ID, but there may be none. Normally, the latter will be a temporary state, caused by the fact that it takes a finite time for ports to be moved to the correct aggregators during reconfiguration.

The Mux machine controls the process of attaching the port to a selected Aggregator, after first detaching the port from any prior Aggregator if the port's LAG ID has changed.

NOTE—A port is always detached from its prior Aggregator when the LAG ID changes, even if the same Aggregator is selected later; to do otherwise would be to risk misdelivery of frames. Selection of a new Aggregator cannot take place until the port is detached from any prior Aggregator; other Aggregators may become free while the port is detaching, and other ports may attach to some of the available Aggregators during this time interval.

The operation of the Selection Logic is separated into the following two subclauses:

a) The requirements for the correct operation of the Selection Logic are defined in 43.4.14.1.
b) The recommended default operation of the Selection Logic is described in 43.4.14.2.

This separation reflects the fact that a wide choice of selection rules is possible within the proper operation of the protocol. An implementation that claims conformance to this standard may support selection rules other than the recommended default; however, any such rules shall meet the requirements stated in 43.4.14.1.

43.4.14.1 Selection Logic—Requirements

Aggregation is represented by a port selecting an appropriate Aggregator, and then attaching to that Aggregator. The following are required for correct operation of the selection and attachment logic:

a) The implementation shall support at least one Aggregator per System.
b) Each port shall be assigned an operational Key (43.3.5). Ports that can aggregate together are assigned the same operational Key as the other ports with which they can aggregate; ports that cannot aggregate with any other port are allocated unique operational Keys.
c) Each Aggregator shall be assigned an operational Key.
d) Each Aggregator shall be assigned an identifier that distinguishes it among the set of Aggregators in the System.
e) A port shall only select an Aggregator that has the same operational Key assignment as its own operational Key.

f) Subject to the exception stated in item g), ports that are members of the same Link Aggregation Group (i.e., two or more ports that have the same Actor System ID, Actor Key, Partner System ID, and Partner Key, and that are not required to be Individual) shall select the same Aggregator.

g) Any pair of ports that are members of the same Link Aggregation Group, but are connected together by the same link, shall not select the same Aggregator (i.e., if a loopback condition exists between two ports, they shall not be aggregated together. For both ports, the Actor System ID is the same as the Partner System ID; also, for port A, the Partner's port identifier is port B, and for port B, the Partner's port identifier is port A).

NOTE—This exception condition prevents the formation of an aggregated link, comprising two ends of the same physical link aggregated together, in which all frames transmitted through an Aggregator are immediately received through the same Aggregator. However, it permits the aggregation of multiple links that are in loopback; for example, if port A is looped back to port C and port B is looped back to port D, then it is permissible for A and B (or A and D) to aggregate together, and for C and D (or B and C) to aggregate together.

h) Any port that is required to be Individual (i.e., the operational state for the Actor or the Partner indicates that the port is Individual) shall not select the same Aggregator as any other port.

i) Any port that is Aggregatable shall not select an Aggregator to which an Individual port is already attached.

j) If the above conditions result in a given port being unable to select an Aggregator, then that port shall not be attached to any Aggregator.

k) If there are further constraints on the attachment of ports that have selected an Aggregator, those ports may be selected as standby in accordance with the rules specified in 43.6.1. Selection or deselection of that Aggregator can cause the Selection Logic to re-evaluate the ports to be selected as standby.

l) The Selection Logic operates upon the operational information recorded by the Receive state machine, along with knowledge of the Actor's own operational configuration and state. The Selection Logic uses the LAG ID for the port, determined from these operational parameters, to locate the correct Aggregator to attach the port to.

m) The Selection Logic is invoked whenever a port is not attached to and has not selected an Aggregator, and executes continuously until it has determined the correct Aggregator for the port.

NOTE—The Selection Logic may take a significant time to complete its determination of the correct Aggregator, as a suitable Aggregator may not be immediately available, due to configuration restrictions or the time taken to re-allocate ports to other Aggregators.

n) Once the correct Aggregator has been determined, the variable Selected shall be set to SELECTED or to STANDBY (43.4.8, 43.6.1).

NOTE—If Selected is SELECTED, the Mux machine will start the process of attaching the port to the selected Aggregator. If Selected is STANDBY, the Mux machine holds the port in the WAITING state, ready to be attached to its Aggregator once its Selected state changes to SELECTED.

o) The Selection Logic is responsible for computing the value of the Ready variable from the values of the Ready_N variable(s) associated with the set of ports that are waiting to attach to the same Aggregator (see 43.4.8).

p) Where the selection of a new Aggregator by a port, as a result of changes to the selection parameters, results in other ports in the System being required to re-select their Aggregators in turn, this is achieved by setting Selected to UNSELECTED for those other ports that are required to re-select their Aggregators.

NOTE—The value of Selected is set to UNSELECTED by the Receive machine for the port when a change of LAG ID is detected.

q) A port shall not be enabled for use by the MAC Client until it has both selected and attached to an Aggregator.

43.4.14.2 Selection Logic—Recommended default operation

The recommended default behavior provides an element of determinism (i.e., history independence) in the assignment of ports to Aggregators. It also has the characteristic that no additional MAC addresses are needed, over and above those already assigned to the set of underlying MACs.

NOTE—This standard does not specify any alternative selection rules beyond the recommended set. A wide variety of selection rules are possible within the scope of the requirements stated in 43.4.14.1. In particular, it is possible within these requirements to support implementations that provide fewer Aggregators than ports, as well as implementations designed to minimize configuration changes at the expense of less deterministic behavior.

Each port has an Aggregator associated with it, (i.e., the number of Aggregators in the System equals the number of ports supported). Each port/Aggregator pair is assigned the same operational Key and port number. When there are multiple ports in an aggregation, the Aggregator that the set of ports selects is the Aggregator with the same port number as the lowest-numbered port in the aggregation. Note that this lowest numbered port may not be in a state that allows data transfer across the link; however, it has selected the Aggregator in question. This is illustrated in Figure 43–12.

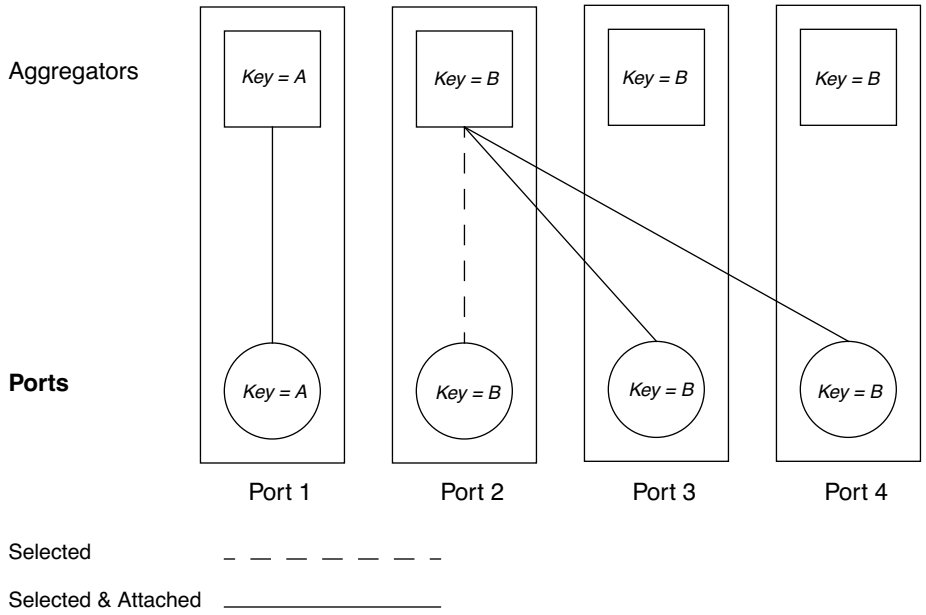


Figure 43–12—Selection of Aggregators

If the port is Individual, then the Aggregator selected is always the port's own Aggregator. Otherwise, an Aggregator is selected from the set of Aggregators corresponding to the set of ports that will form the aggregation. The Aggregator selected is the lowest numbered Aggregator with the same selection parameters as those of the port. These selection parameters are

a) The Actor's System ID
b) The Actor's operational Key
c) The Partner's System ID
d) The Partner's operational Key
e) The Individual_Aggregator state (which must be FALSE)

43.4.15 Mux machine

The Mux machine shall implement the function specified in either of the Mux machine state diagrams, Figure 43–13 and Figure 43–14, with their associated parameters (43.4.4 through 43.4.11).

The state machine conventions in 21.5 assert that all in-state actions are instantaneous, are performed in sequence, and are performed prior to evaluating any exit conditions. While the Mux machine will operate correctly if all actions can be performed instantaneously, this will not be realistic in many implementations. Correct operation is maintained even if actions are not completed instantaneously, as long as each action completes prior to initiating the next sequential action, and all actions complete prior to evaluating any exit conditions.

The independent control state diagram (Figure 43–13) is suitable for use implementations in which it is possible to control enabling and disabling of frame collection from a port, and frame distribution to a port, independently. The coupled control state diagram (Figure 43–14) is suitable for use implementations where collecting and distributing cannot be controlled independently with respect to a port. It is recommended that the independent control state diagram be implemented in preference to the coupled control state diagram.

The value of the Selected variable may be changed by the following:

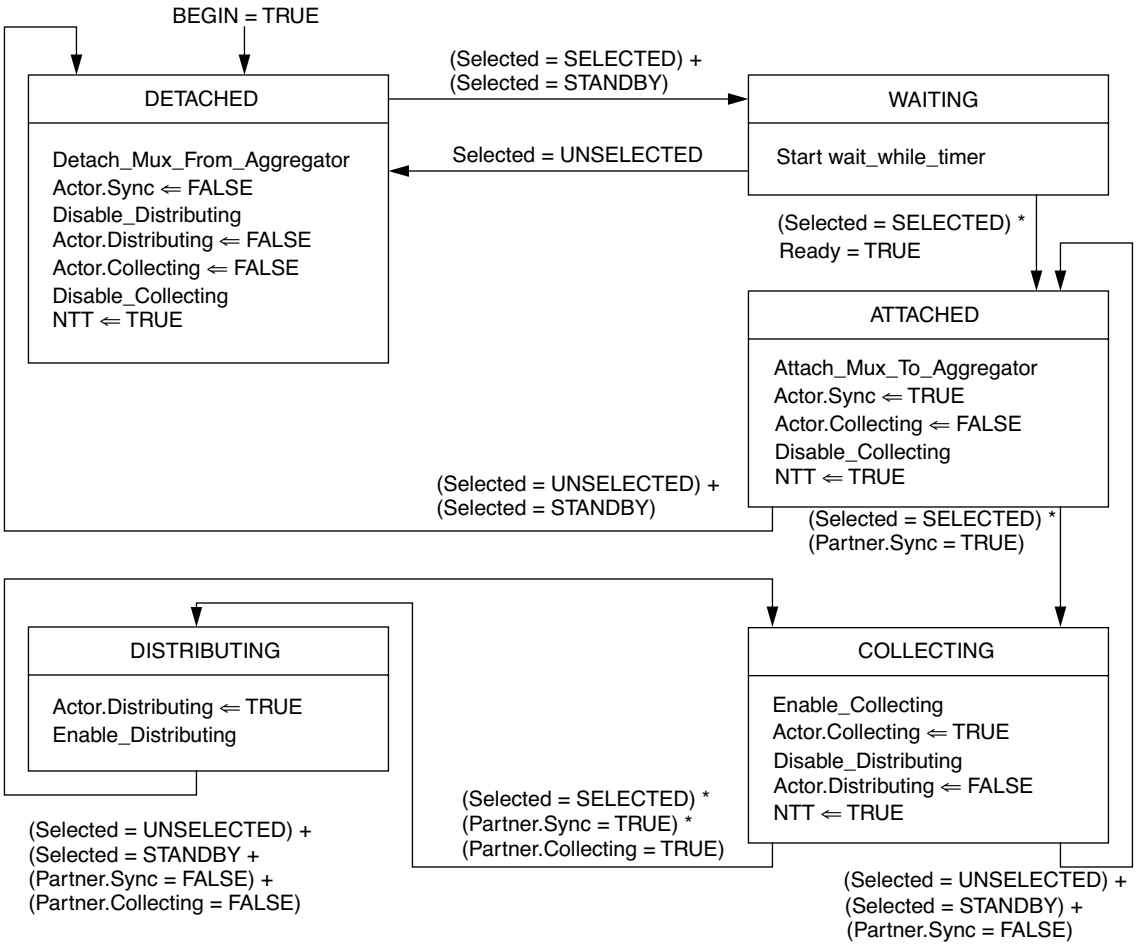
a) *The Receive machine*. The Receive machine can set Selected to UNSELECTED at any time if any of the following change: The Partner System ID, the Partner Priority, the Partner Key, the Partner_State.Aggregation, the Actor System ID, the Actor Priority, the Actor Key, or the Actor_State.Aggregation.
b) *The Selection Logic, in the process of selecting an Aggregator.* The Selection Logic will select an Aggregator when the Mux machine is in the DETACHED state and the value of the Selected variable is UNSELECTED.
c) *The Selection Logic, in the process of selecting or de-selecting standby links*. If the value of the Selected variable is SELECTED or STANDBY, the Selection Logic can change the value to STANDBY or SELECTED.

The Mux machine enters the WAITING state from the DETACHED state if the Selection Logic determines that Selected is either SELECTED or STANDBY. The WAITING state provides a holding state for the following two purposes:

d) If Selected is SELECTED, the wait_while_timer forces a delay to allow for the possibility that other ports may be reconfiguring at the same time. Once the wait_while_timer expires, and once the wait_while_timers of all other ports that are ready to attach to the same Aggregator have expired, the process of attaching the port to the Aggregator can proceed, and the state machine enters the ATTACHED state. During the waiting time, changes in selection parameters can occur that will result in a re-evaluation of Selected. If Selected becomes UNSELECTED, then the state machine re-enters the DETACHED state. If Selected becomes STANDBY, the operation is as described in item e).

   NOTE—This waiting period reduces the disturbance that will be visible to higher layers; for example, on start-up events. However, the selection need not wait for the entire waiting period in cases where it is known that no other ports will attach; for example, where all other ports with the same operational Key are already attached to the Aggregator.

e) If Selected is STANDBY, the port is held in the WAITING state until such a time as the selection parameters change, resulting in a re-evaluation of the Selected variable. If Selected becomes UNSELECTED, the state machine re-enters the DETACHED state. If SELECTED becomes SELECTED, then the operation is as described in item d). The latter case allows a port to be brought into operation from STANDBY with minimum delay once Selected becomes SELECTED.



The following abbreviations are used in this diagram:
Actor.Sync: Actor_Oper_Port_State.Synchronization
Actor.Collecting: Actor_Oper_Port_State.Collecting
Actor.Distributing: Actor_Oper_Port_State.Distributing
Partner.Sync: Partner_Oper_Port_State.Synchronization
Partner.Collecting: Partner_Oper_Port_State.Collecting

Figure 43–13—Mux machine state diagram (independent control)

On entry to the ATTACHED state, the Mux machine initiates the process of attaching the port to the selected Aggregator. Once the attachment process has completed, the value of Actor_Oper_Port_State.Synchronization is set to TRUE indicating that the Actor considers the port to be IN_SYNC, and Actor_Oper_Port_State.Collecting is set to FALSE. Collection of frames from the port is disabled. In the coupled control state machine, Distribution of frames to the port is also disabled, and Actor_Oper_Port_State.Distributing is set to FALSE.

A change in the Selected variable to UNSELECTED or to STANDBY causes the state machine to enter the DETACHED state. The process of detaching the port from the Aggregator is started. Once the detachment process is completed, Actor_Oper_Port_State.Synchronization is set to FALSE indicating that the Actor considers the port to be OUT_OF_SYNC, distribution of frames to the port is disabled, Actor_Oper_Port_State.Distributing and Actor_Oper_Port_State.Collecting are set to FALSE, and collection of frames from the port is disabled. The state machine remains in the DISABLED state until such time as the Selection logic is able to select an appropriate Aggregator.

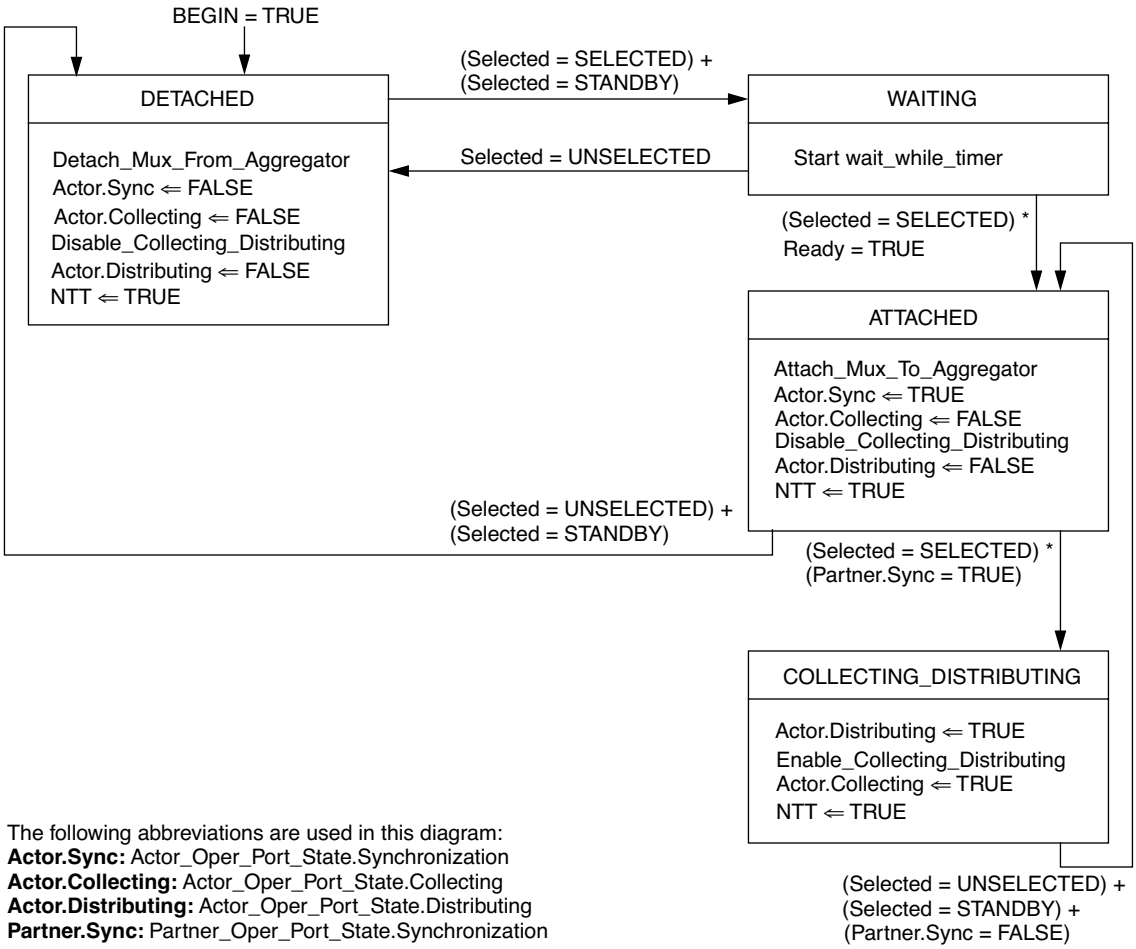


Figure 43–14—Mux machine state diagram (coupled control)

While in the ATTACHED state, a TRUE value for Partner_Oper_Port_State.Synchronization causes the state machine to transition to the COLLECTING state (independent control) or the COLLECTING_DISTRIBUTING state (coupled control).

In the COLLECTING state, collection of frames from the port is enabled, then Actor_Oper_Port_State.Collecting is set to TRUE, then distribution of frames to the port is disabled and Actor_Port_Oper_State.Distributing is set to FALSE. The state machine will return to the ATTACHED state if Selected changes to UNSELECTED or STANDBY, or if the Partner's synchronization state becomes FALSE. Once the Partner has signalled that collecting has been enabled (Partner_Oper_Port_State.Collecting is TRUE), the state machine transitions to the DISTRIBUTING state. The value of Actor_Oper_Port_State.Distributing is set to TRUE, and then distribution of frames to the port is enabled. From DISTRIBUTING, a return to the COLLECTING state occurs if the value of Selected becomes UNSELECTED or STANDBY, if the Partner's synchronization state becomes FALSE, or if the Partner signals that collection has been disabled (Partner_Oper_Port_State.Collecting is FALSE).

In the COLLECTING_DISTRIBUTING state, the value of Actor_Oper_Port_State.Distributing is set to TRUE, distribution of frames to the port and collection of frames from the port are both enabled, and then Actor_Oper_Port_State.Collecting is set to TRUE. The state machine will return to the ATTACHED state if Selected changes to UNSELECTED or STANDBY, or if the Partner's synchronization state becomes FALSE.

The sequence of operations and transitions defined for the COLLECTING and DISTRIBUTING states in the independent control version of this state machine ensures that frames are not distributed to a port until the Partner has enabled collection, and that distribution is stopped as soon as the Partner's state indicates that collection has been disabled. This sequence minimizes the possibility that frames will be misdelivered during the process of bringing the port into operation or taking the port out of operation. In the coupled control version of the state machine, the COLLECTING and DISTRIBUTING states merge together to form the combined state, COLLECTING_DISTRIBUTING. As independent control is not possible, the coupled control state machine does not wait for the Partner to signal that collection has started before enabling both collection and distribution.

The NTT variable is set to TRUE in the DETACHED, ATTACHED, COLLECTING, and COLLECTING_DISTRIBUTING states in order to ensure that the Partner is made aware of the changes in the Actor's state variables that are caused by the operations performed in those states.

43.4.16 Transmit machine

When the Transmit machine creates a LACPDU for transmission, it shall fill in the following fields with the corresponding operational values for this port:

a) Actor_Port and Actor_Port_Priority
b) Actor_System and Actor_System_Priority
c) Actor_Key
d) Actor_State
e) Partner_Port and Partner_Port_Priority
f) Partner_System and Partner_System_Priority
g) Partner_Key
h) Partner_State
i) CollectorMaxDelay

When the Periodic machine is in the NO_PERIODIC state, the Transmit machine shall

— Not transmit any LACPDUs, and
— Set the value of NTT to FALSE.

When the LACP_Enabled variable is TRUE and the NTT (43.4.7) variable is TRUE, the Transmit machine shall ensure that a properly formatted LACPDU (43.4.2) is transmitted (i.e., issue a Ctrl-MuxN:MA_DATA.Request(LACPDU) service primitive), subject to the restriction that no more than three LACPDUs may be transmitted in any Fast_Periodic_Time interval. If NTT is set to TRUE when this limit is in force, the transmission shall be delayed until such a time as the restriction is no longer in force. The NTT variable shall be set to FALSE when the Transmit machine has transmitted a LACPDU.

If the transmission of a LACPDU is delayed due to the above restriction, the information sent in the LACPDU corresponds to the operational values for the port at the time of transmission, not at the time when NTT was first set to TRUE. In other words, the LACPDU transmission model is based upon the transmission of state information that is current at the time an opportunity to transmit occurs, as opposed to queuing messages for transmission.

When the LACP_Enabled variable is FALSE, the Transmit machine shall not transmit any LACPDUs and shall set the value of NTT to FALSE.

43.4.17 Churn Detection machines

If implemented, the Churn Detection machines shall implement the functions specified in Figure 43–15 and Figure 43–16 with their associated parameters (43.4.4 through 43.4.11). Implementation of the Churn Detection machines is mandatory if the associated management functionality (the Aggregation Port Debug Information package) is implemented; otherwise, implementation of the Churn Detection machines is optional.

The Churn Detection machines detect the situation where a port is operable, but the Actor and Partner have not attached the link to an Aggregator and brought the link into operation within a bounded time period. Under normal operation of the LACP, agreement between Actor and Partner should be reached very rapidly. Continued failure to reach agreement can be symptomatic of device failure, of the presence of non-standard devices, or of misconfiguration; it can also be the result of normal operation in cases where either or both Systems are restricted in their ability to aggregate. Detection of this condition is signalled by the Churn Detection machines to management in order to prompt administrative action to further diagnose and correct the fault.

NOTE—One of the classes of problems that will be detected by this machine is the one where the implementation has been designed to support a limited number of Aggregators (fewer than the number of ports—see 43.6.4.2) and the physical topology is such that one or more ports end up with no Aggregator to attach to. This may be the result of a wiring error or an error in the allocation of operational Key values to the ports and Aggregators. Alternatively, failure of a link to aggregate may be the result of a link being placed in standby mode by a System that has hardware limitations placed on its aggregation ability, leading it to make use of the techniques described in 43.6 to find the ideal configuration. Given that the time taken by an aggregation-constrained System to stabilize its configuration may be relatively large, the churn detection timers allow 60 seconds to elapse before a Churn condition is signalled.

The symptoms that the Actor Churn Detection state machine detects is that the Actor's Mux has determined that it is OUT_OF_SYNC, and that condition has not resolved itself within a period of time equal to Short_Timeout_Time (43.4.4). Under normal conditions, this is ample time for convergence to take place. Similarly, the Partner Churn Detection state machine detects a failure of the Partner's Mux to synchronize.

The Actor Churn Detection state machine is depicted in Figure 43–15. The Partner Churn Detection state machine is depicted in Figure 43–16.

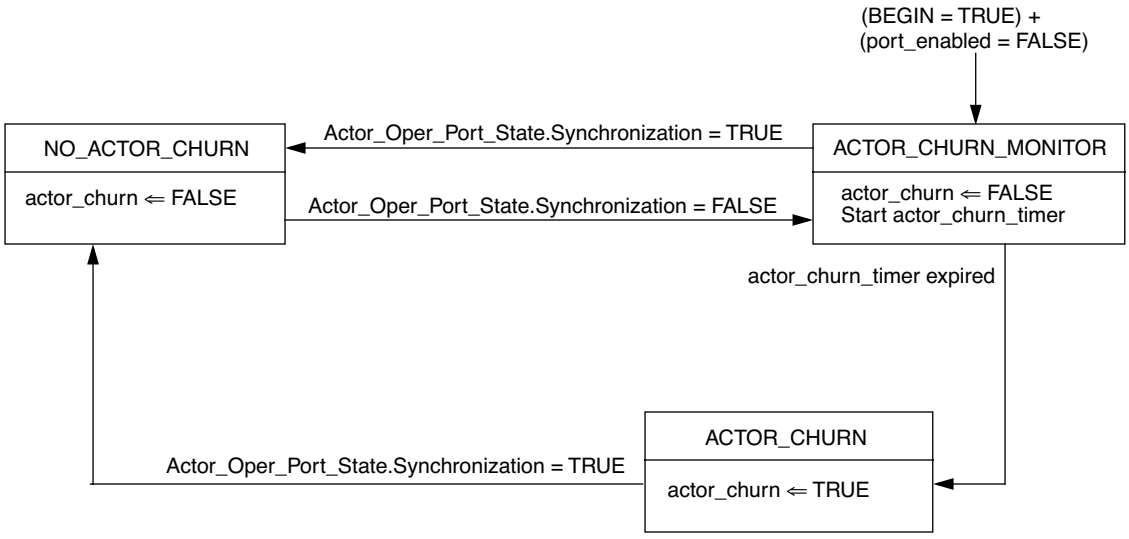


Figure 43–15—Actor Churn Detection machine state diagram

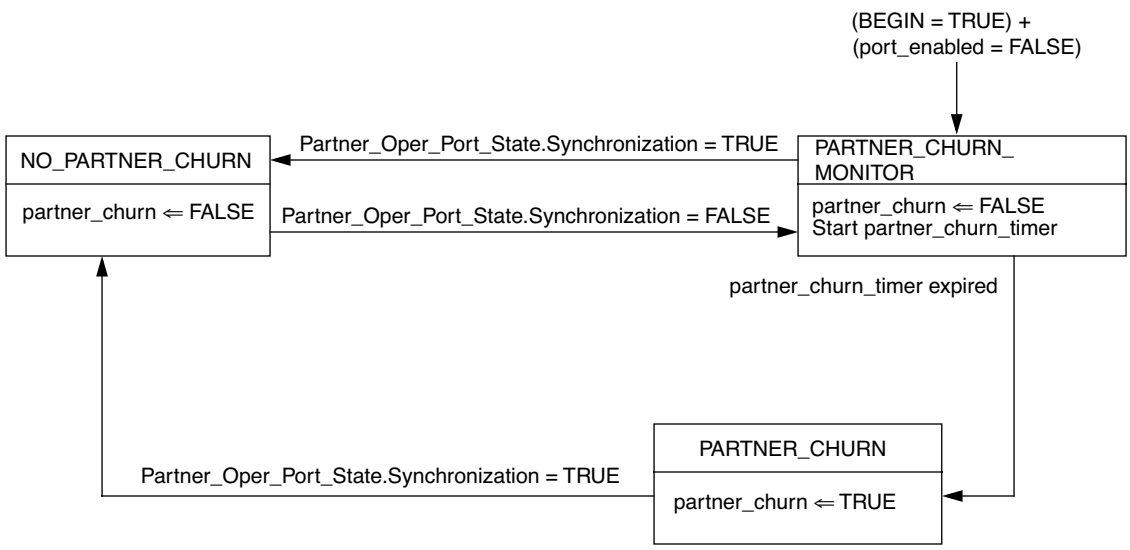


Figure 43–16—Partner Churn Detection machine state diagram

43.5 Marker protocol

43.5.1 Introduction

The Marker protocol allows the distribution function of an Actor's Link Aggregation sublayer to request the transmission of a Marker PDU on a given link. The Marker PDU is received by the Partner's collection function and a Marker Response PDU is returned on the same link to the initiating Actor's distribution function. Marker and Marker Response PDUs are treated by the underlying MACs at each end of the link as normal MAC Client PDUs; i.e., there is no prioritization or special treatment of Marker or Marker Response PDUs relative to other frames. Marker/Marker Response PDUs are subject to the operation of flow control, where supported on the link. Hence, if the distribution function requests transmission of a Marker PDU on a given link and does not transmit any further MAC Client PDUs that relate to a given set of conversations until the corresponding Marker Response PDU is received from that link, then it can be certain that there are no MAC Client PDUs related to those conversations still to be received by the Partner's collection function. The use of the Marker protocol can therefore allow the Distribution function a means of determining the point at which a given set of conversations can safely be reallocated from one link to another without the danger of causing frames in those conversations to be mis-ordered at the collector.

NOTE—The use of the Marker protocol is further discussed in Annex 43A.

The operation of the Marker protocol is unaffected by any changes in the Collecting and Distributing states associated with the port. Therefore, Marker and Marker Response PDUs can be sent on a port whose distribution function is disabled; similarly, such PDUs can be received and passed to the relevant Aggregator's collection or distribution function on a port whose collection function is disabled.

The use of the Marker protocol is optional; however, the ability to respond to Marker PDUs, as defined for the operation of the Marker Responder (see 43.5.4.1 and 43.5.4.2), is mandatory. Some distribution algorithms may not require the use of a marker; other mechanisms (such as timeouts) may be used as an alternative. As the specification of distribution algorithms is outside the scope of this standard, no attempt is made to specify how, when, or if the Marker protocol is used. (See Annex 43A for an informative discussion of distribution algorithms.).

The Marker protocol does not provide a guarantee of a response from the Partner; no provision is made for the consequences of frame loss or for the failure of the Partner System to respond correctly. Implementations that make use of this protocol must therefore make their own provision for handling such cases.

43.5.2 Sequence of operations

Figure 43–17 illustrates the sequence of marker operations between an initiating and responding System. Time is assumed to flow from the top of the diagram to the bottom.

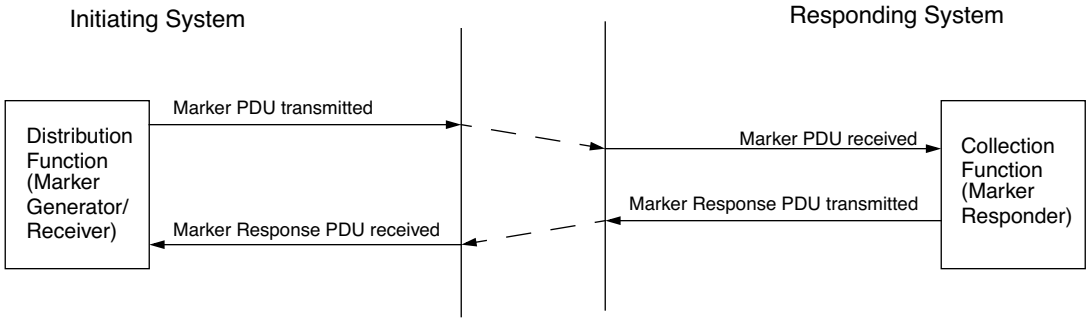


Figure 43–17—Marker protocol time sequence diagram

43.5.3 Marker and Marker Response PDU structure and encoding

43.5.3.1 Transmission and representation of octets

All Marker and Marker Response PDUs comprise an integral number of octets. The bits in each octet are numbered from 0 to 7, where 0 is the low-order bit. When consecutive octets are used to represent a numerical value, the most significant octet is transmitted first, followed by successively less significant octets.

When the encoding of (an element of) a Marker or Marker Response PDU is depicted in a diagram, then

a) Octets are transmitted from top to bottom.
b) Within an octet, bits are shown with bit 0 to the left and bit 7 to the right, and are transmitted from left to right.
c) Numerical values are encoded as binary numbers.

43.5.3.2 Marker and Marker Response PDU structure

Marker PDUs and Marker Response PDUs are basic IEEE 802.3® frames; they shall not be tagged (see Clause 3). The Marker PDU and Marker Response PDU structure shall be as shown in Figure 43–18 and as further described in the following field definitions:

a) *Destination Address*. The DA in all Marker and Marker Response PDUs is the Slow_Protocols_Multicast address. Its use and encoding are specified in Annex 43B.
b) *Source Address*. The SA in Marker and Marker Response PDUs is the individual MAC address associated with the port from which the PDU is transmitted.
c) *Length/Type*. Marker and Marker Response PDUs are always Type encoded, and carry the Slow Protocol type field. The use and encoding of this type field is specified in Annex 43B.
d) *Subtype*. The Subtype field identifies the specific Slow Protocol being encapsulated. Both Marker and Marker Response PDUs carry the Marker_subtype value 0x02.

e) *Version number*. This identifies the Marker protocol version; implementations conformant to this version of the standard carry the value 0x01.

f) *TLV_type = Marker Information/Marker Response Information*. This indicates the nature of the information carried in this TLV-tuple. Marker Information is encoded as the integer value 0x01; Marker Response Information is encoded as the integer value 0x02.

g) *Marker_Information_Length/Marker_Response_Information_Length*. This field indicates the length (in octets) of this TLV-tuple. Both Marker and Marker Response information use a length value of 16 (0x10).

h) *Requester_Port*. The port number assigned to the port by the Requester (the System sending the initial Marker PDU), encoded as an unsigned integer.

Marker PDU	Octets	Marker Response PDU
Destination Address	6	Destination Address
Source Address	6	Source Address
Length/Type	2	Length/Type
Subtype = Marker	1	Subtype = Marker
Version Number	1	Version Number
TLV_type = Marker Information	1	TLV_type = Marker Response Information
Marker_Information_Length= 16	1	Marker_Response_Information_Length = 16
Requester_Port	2	Requester_Port
Requester_System	6	Requester_System
Requester_Transaction_ID	4	Requester_Transaction_ID
Pad = 0	2	Pad = 0
TLV_type = Terminator	1	TLV_type = Terminator
Terminator_Length = 0	1	Terminator_Length = 0
Reserved	90	Reserved
FCS	4	FCS

Figure 43–18—Marker PDU and Marker Response PDU structure

i) *Requester_System*. The Requester's System ID, encoded as a MAC address.

j) *Requester_Transaction_ID*. The transaction ID allocated to this request by the requester, encoded as an integer.

k) *Pad*. This field is used to align TLV-tuples on 16-byte memory boundaries. It is transmitted as zeroes in Marker PDUs; in Marker Response PDUs, this field may be transmitted as zeroes, or with the contents of this field from the Marker PDU triggering the response. The field is ignored on receipt in all cases.

NOTE—The difference in handling of the Pad field in Marker Response PDUs allows an implementation to reflect the contents of the received Marker PDU in its response, without enforcing the requirement to transmit the field as zeroes.

l) *TLV_type = Terminator*. This field indicates the nature of the information carried in this TLV-tuple. Terminator (end of message) information carries the integer value 0x00.

m) *Terminator_Length*. This field indicates the length (in octets) of this TLV-tuple. Terminator information uses a length value of 0 (0x00).

n) *Reserved*. These 90 octets are reserved for use in future extensions to the protocol. It is transmitted as zeroes in Marker PDUs; in Marker Response PDUs, this field may be either transmitted as zeroes, or with the contents of this field from the Marker PDU triggering the response. The field is ignored on receipt in all cases.

NOTE—These trailing reserved octets are included in all Marker and Marker Response PDUs in order to force a fixed PDU size, regardless of the version of the protocol. Hence, a Version 1 implementation is guaranteed to be able to receive version N PDUs successfully, although version N PDUs may contain additional information that cannot be interpreted (and will be ignored) by the Version 1 implementation. A crucial factor in ensuring backwards compatibility is that any future version of the protocol is required not to re-define the structure or semantics of information defined for the previous version; it may only add new information elements to the previous set. Hence, in a version N PDU, a Version 1 implementation can expect to find the Version 1 information in exactly the same places as in a Version 1 PDU, and can expect to interpret that information as defined for Version 1.

o) *FCS*. This field is the Frame Check Sequence, typically generated by the underlying MAC.

43.5.4 Protocol definition

43.5.4.1 Operation of the marker protocol

Marker PDUs may be generated by the Frame Distribution function to provide a sequence marker in the stream of frames constituting a conversation or set of conversations. Received Marker PDUs are delivered to the Marker Responder within the Frame Collection function of the Partner System.

On receipt of a valid Marker PDU, the Frame Collection function issues a Marker Response PDU, in the format specified in Figure 43–18, to the same port from which the Marker PDU was received. The **Requester_Port, Requester_System,** and **Requester_Transaction_ID** parameter in the Marker Response PDU carry the same values as those received in the corresponding Marker PDU.

Received Marker Response PDUs are passed to the Marker Receiver within the Frame Distribution function. Implementation of the Marker Generator and Receiver is optional.

The Marker Generator, if implemented, shall comply with the frame rate limitation constraint for Slow Protocols, as specified in Annex 43B.3. A Marker Responder may (but is not required to) control its Marker Response transmissions to conform to this Slow Protocols timing constraint when faced with Marker messages not in compliance with this constraint (i.e., to send fewer Marker Response PDUs than Marker PDUs received). If the Marker Responder is controlling its responses in this manner, Marker Response PDUs corresponding to Marker PDUs received in excess of the Slow Protocols timing constraint shall not be sent.

NOTE—It is important that Marker Response PDUs not be queued indefinitely, and sent long after the corresponding Marker PDU that triggered the response.

Frames generated by the Marker Responder do not count towards the rate limitation constraint for Slow Protocols, as specified in Annex 43B.3.

43.5.4.2 Marker Responder state diagram

The Marker Responder shall implement the function specified in Figure 43–19, with its associated parameters (43.5.4.2.1 through 43.5.4.2.3).

43.5.4.2.1 Constants

Slow_Protocols_Multicast
: The value of the Slow Protocols reserved multicast address. (See Table 43B–1.)

43.5.4.2.2 Variables

DA
SA
m_sdu
service_class
status
: The parameters of the MA_DATA.request and MA_DATA.indication primitives, as defined in Clause 2.

43.5.4.2.3 Messages

AggMuxN:MA_DATA.request
: The service primitive used to transmit a frame with the specified parameters.

AggMuxN:MA_DATA.indication
: The service primitive used to pass a received frame to a client with the specified parameters.

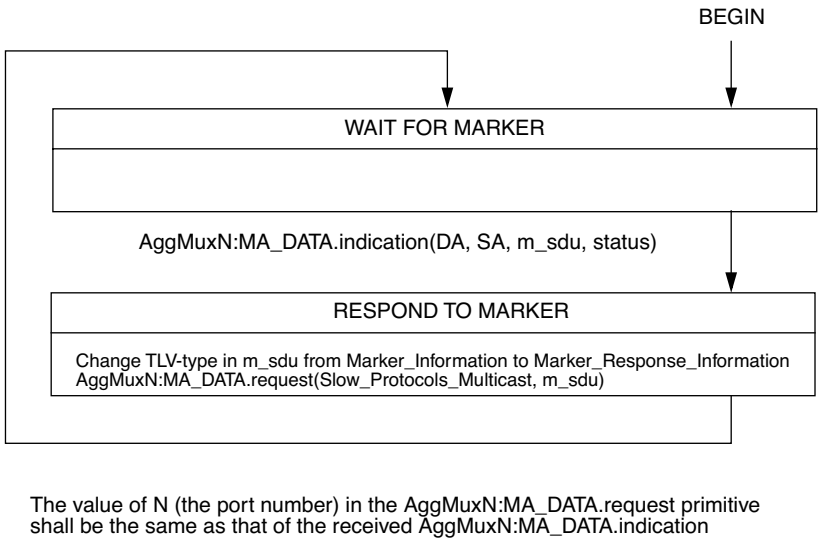


Figure 43–19—Marker Responder state diagram

Upon receipt of an AggMuxN:MA_DATA.indication primitive, the Marker Responder shall not validate the Version Number, Pad, or Reserved fields in the contained Marker Request PDU. The same actions are taken regardless of the values received in these fields. A Marker Responder may validate the Marker_Information_Length field. These behaviors, together with the constraint on future protocol enhancements discussed in the Note in 43.5.3.2, allow Version 1 devices to be compatible with future revisions of the protocol.

43.6 Configuration capabilities and restrictions

43.6.1 Use of system and port priorities

The formulation chosen for the Link Aggregation Group identifier (43.3.6) has the consequence that it is not possible to represent two or more Link Aggregation Groups, comprising aggregatable links, that share the same combination of {SK, TL}. Hence, placing configuration restrictions on the size of an aggregation (e.g.,

for a Key Group containing 6 members, restricting the size of any aggregation to 4 members or fewer) is only possible if it is also acceptable that only one Link Aggregation Group can be constructed from that Key Group for a given {SK, TL}. In practice, this restriction can be somewhat alleviated by subdividing Key Groups and allocating different operational Keys to each subdivision; however, this is, in general, only useful if the form of the size restriction is closely bound to physical subdivisions in the implementation (e.g., it might be possible to aggregate only those links that are on the same interface card).

In Systems that have limited aggregation capability of this form, the following algorithm shall be used to determine the subset of ports that will be aggregated together:

a) The System Aggregation Priority of each System is an eight octet binary number, formed by using the Actor_System_Priority as the two most significant octets and the Actor's System ID as the least significant six octets. For a given Actor and Partner, the System with the numerically lower value of System Aggregation Priority has the higher priority.
b) The Port Aggregation Priority of each port is a four octet binary number, formed by using the Actor_Port_Priority as the two most significant octets and the port number as the two least significant octets. For any given set of ports, the port with the numerically lower value of Port Aggregation Priority has the higher priority.
c) Ports shall be selected for aggregation by each System based upon the Port Aggregation Priority assigned by the System with the higher System Aggregation Priority, starting with the highest priority port of the System with the higher priority, and working downward through the ordered list of Port Aggregation Priority values for the N ports, applying the particular constraints imposed on the System concerned.
d) For each link that a given System cannot include in the aggregation, the Selection Logic identifies the Selection state of the corresponding port as STANDBY, preventing the link from becoming active. The synchronization state signalled in transmitted LACPDUs for such links will be OUT_OF_SYNC.
e) The selection algorithm is reapplied upon changes in the membership of the Link Aggregation Group (for example, if a link fails, or if a new link joins the group) and any consequent changes to the set of active links are made accordingly.

A port that is selected as standby as a result of limitations on aggregation capability can be viewed as providing a "hot standby" facility, as it will be able to take part in the aggregation upon failure of one of the active links in the aggregation. The ability to hold links in a standby mode in this way provides the possibility of using LACP even where the System is incapable of supporting distribution and collection with more than one port. Parallel links could be automatically configured as standby links, and deployed to mask link failures without any disruption to higher layer protocols.

43.6.2 Dynamic allocation of operational Keys

In some circumstances, the use of System and port priorities may prove to be insufficient to generate the optimum aggregation among the set of links connecting a pair of Systems. A System may have a limited aggregation capability that cannot be simply expressed as a limit on the total number of links in the aggregation. The full description of its restrictions may be that it can only aggregate together particular subsets of links, and the sizes of the subsets need not all be the same.

NOTE—An example would be an implementation organized such that, for a set of four links A through D, it would be possible to operate with {A+B+C+D} as a single aggregation, or operate with {A+B} and {C+D} as two separate aggregations, or operate as four individual links; however, all other aggregation possibilities (such as {A+C} and {B+D}) would not be achievable by the implementation.

In such circumstances, it is permissible for the System with the higher System Aggregation Priority (i.e., the numerically lower value) to dynamically modify the operational Key value associated with one or more of

the ports; the System with the lower priority shall not attempt to modify operational Key values for this purpose. Operational Key changes made by the higher priority System should be consistent with maintaining its highest priority port in the aggregate as an active link (i.e., in the IN_SYNC state). Successive operational Key changes, if they occur, should progressively reduce the number of ports in the aggregation. The original operational Key value should be maintained for the highest priority port thought to be aggregatable.

NOTE—Restricting operational Key changes in the manner described prevents the case where both Partner Systems involved have limited capability and both attempt to make operational Key changes; this could be a non-converging process, as a change by one participant can cause the other participant to make a change, which in turn causes the first participant to make a change—and so on, ad infinitum.

This approach effectively gives the higher priority System permission to search the set of possible configurations, in order to find the best combination of links given its own and its Partner's configuration constraints. The reaction of the Partner System to these changes can be determined by observing the changes in the synchronization state of each link. A System performing operational Key changes should allow at least 4 seconds for the Partner System to change an OUT_OF_SYNC state to an IN_SYNC state.

In the course of normal operation a port can dynamically change its operating characteristics (e.g., data rate, full or half duplex operation). It is permissible (and appropriate) for the operational Key value associated with such a port to change with the corresponding changes in the operating characteristics of the link, so that the operational Key value always correctly reflects the aggregation capability of the link. Operational Key changes that reflect such dynamic changes in the operating characteristics of a link may be made by either System without restriction.

43.6.3 Link Aggregation on shared-medium links

The Link Aggregation Control Protocol cannot detect the presence of multiple Aggregation-aware devices on the same link. Hence, shared-medium links shall be treated as Individual, with transmission/reception of LACPDUs disabled on such ports.

43.6.4 Selection Logic variants

Two variants of the Selection Logic rules are described as follows:

a) The first accommodates implementations that may wish to operate in a manner that minimizes disturbance of existing aggregates, at the expense of the deterministic characteristics of the logic described in 43.4.14.2.
b) The second accommodates implementations that may wish to limit the number of Aggregators that are available for use to fewer than the number of ports supported.

43.6.4.1 Reduced reconfiguration

By removing the constraint that the Aggregator chosen is always the lowest numbered Aggregator associated with the set of ports in an aggregation, an implementation can minimize the degree to which changes in the membership of a given aggregation result in changes of connectivity at higher layers. As there would still be the same number of Aggregators and ports with a given operational Key value, any port will still always be able to find an appropriate Aggregator to attach to, however the configuration achieved over time (i.e., after a series of link disconnections, reconnections, or reconfigurations) with this relaxed set of rules would not necessarily be the same as the configuration achieved if all Systems involved were reset, given the rules stated in 43.4.15.

43.6.4.2 Limited Aggregator availability

By removing the constraint that there are always as many Aggregators as ports, an implementation can limit the number of MAC Client interfaces available to higher layers while maintaining the ability for each Aggregator to serve multiple ports. This has the same effect as removing the assumption that Aggregators and their associated ports have the same operational Key value; Aggregators can be effectively disabled (and therefore ignored) by configuring their Keys to be different from any operational Key value allocated to any of the ports.

In this scenario, any port(s) that cannot find a suitable Aggregator to attach to will simply wait in the DETACHED state until an Aggregator becomes available, with a synchronization state of OUT_OF_SYNC.

43.7 Protocol Implementation Conformance Statement (PICS) proforma for Clause 43, Aggregation of Multiple Link Segments[13]

43.7.1 Introduction

The supplier of an implementation that is claimed to conform to Clause 43, Link Aggregation, shall complete the following Protocol Implementation Conformance Statement (PICS) proforma.

A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

43.7.2 Identification

43.7.2.1 Implementation identification

Supplier (Note 1)	
Contact point for queries about the PICS (Note 1)	
Implementation Name(s) and Version(s) (Notes 1 and 3)	
Other information necessary for full identification—e.g., name(s) and version(s) of machines and/or operating system names (Note 2)	
NOTES 1—Required for all implementations. 2—May be completed as appropriate in meeting the requirements for the identification. 3—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

[13] *Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this annex so that it can be used for its intended purpose and may further publish the completed PICS.

43.7.2.2 Protocol summary

Identification of protocol specification	IEEE Std 802.3-2002®, Clause 43, Link Aggregation.
Identification of amendments and corrigenda to the PICS proforma which have been completed as part of the PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21: the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®, Clause 43, Link Aggregation.)	

Date of Statement	

43.7.3 Major capabilities/options

Item	Feature	Subclause	Value/Comment	Status	Support
*MG	Marker Generator/Receiver	43.2.5		O	Yes [ ] No [ ]
*AM	Aggregation Port Debug Information package support	30.7		O	Yes [ ] No [ ]
*CM	Churn Detection machine	43.4.17	Required if Aggregation Port Debug Information package supported	AM: M !AM: O	N/A [ ] M:Yes [ ] Yes [ ] No [ ]

43.7.4 Frame Collector

Item	Feature	Subclause	Value/Comment	Status	Support
FC1	Frame Collector function	43.2.3	As specified in the state diagram shown in Figure 43–3 and associated definitions in 43.2.3.1	M	Yes [ ]
FC2	Frame Collector function—CollectorMaxDelay	43.2.3.1.4	Deliver or discard frames within CollectorMaxDelay	M	Yes [ ]

43.7.5 Frame Distributor

Item	Feature	Subclause	Value/Comment	Status	Support
	Distribution algorithm ensures the following, when frames received by Frame Collector:	43.2.4			
FD1	Frame mis-ordering		None	M	Yes [ ]
FD2	Frame duplication		None	M	Yes [ ]
FD3	Frame Distributor function	43.2.4	Function as specified in the state diagram shown in Figure 43–4 and associated definitions in 43.2.4.1	M	Yes [ ]

43.7.6 Marker protocol

Item	Feature	Subclause	Value/Comment	Status	Support
MGR1	Marker Generator/Receiver	43.2.5		MG:M	N/A [ ] M:Yes [ ]
MGR2	Marker Responder	43.2.6	Function specified in 43.5.4.2	M	Yes [ ]

43.7.7 Aggregator Parser/Multiplexer

Item	Feature	Subclause	Value/Comment	Status	Support
APM1	Aggregator Multiplexer	43.2.7	Transparent pass-through of frames	M	Yes [ ]
APM2	Aggregator Multiplexer	43.2.7	Discard of TX frames when port not Distributing	M	Yes [ ]
APM3	Aggregator Parser	43.2.7	Function specified by state diagram shown in Figure 43–5 and associated definitions in 43.2.7.1.	M	Yes [ ]
APM4	Aggregator Parser	43.2.7	Discard of RX frames when port not Collecting	M	Yes [ ]

43.7.8 Control Parser/Multiplexer

Item	Feature	Subclause	Value/Comment	Status	Support
CPM1	Control Multiplexer	43.2.9	Transparent pass-through of frames	M	Yes [ ]
CPM2	Control Parser	43.2.9	Function specified by state diagram shown in Figure 43–6 and associated definitions in 43.2.9.1.	M	Yes [ ]

43.7.9 System identification

Item	Feature	Subclause	Value/Comment	Status	Support
SID1	Globally unique identifier	43.3.2	Globally administered individual MAC address plus System Priority	M	Yes [ ]
SID2	MAC address chosen	43.3.2	MAC address associated with one of the ports	O	Yes [ ] No [ ]

43.7.10 Aggregator identification

Item	Feature	Subclause	Value/Comment	Status	Support
AID1	Globally unique identifier	43.3.3	Globally administered individual MAC address	M	Yes [ ]
AID2	Integer identifier	43.3.3	Uniquely identifies the Aggregator within the System	M	Yes [ ]
*AID3	Unique identifier allocated	43.3.3	Unique identifier assigned to one of its bound ports	O	Yes [ ] No [ ]
AID4			Unique identifier not assigned to any other Aggregator	!AID3 :M	N/A [ ] M:Yes [ ]

43.7.11 Port identification

Item	Feature	Subclause	Value/Comment	Status	Support
PID1	Port identifiers	43.3.4	Unique within a System; port number 0 not used for any port	M	Yes [ ]

43.7.12 Capability identification

Item	Feature	Subclause	Value/Comment	Status	Support
CID1	Administrative and operational Key values associated with each port	43.3.5		M	Yes [ ]
CID2	Administrative and operational Key values associated with each Aggregator	43.3.5		M	Yes [ ]

43.7.13 Link Aggregation Group identification

Item	Feature	Subclause	Value/Comment	Status	Support
LAG1	LAG ID component values	43.3.6.1	Actor's values non-zero. Partner's admin values only zero for Individual ports	M	Yes [ ]

43.7.14 Detaching a link from an Aggregator

Item	Feature	Subclause	Value/Comment	Status	Support
DLA1	Effect on conversation reallocated to a different link	43.3.13	Frame ordering preserved	M	Yes [ ]

43.7.15 LACPDU structure

Item	Feature	Subclause	Value/Comment	Status	Support
LPS1	LACPDU Frame format	43.4.2.2	Not Tagged	M	Yes [ ]
LPS2	LACPDU structure	43.4.2.2	As shown in Figure 43–7 and as described	M	Yes [ ]
LPS3	LACPDU structure	43.4.2	All Reserved octets ignored on receipt and transmitted as zero	M	Yes [ ]

43.7.16 State machine variables

Item	Feature	Subclause	Value/Comment	Status	Support
SMV1	Partner_Admin_Port_State	43.4.7	Collecting set to the same value as Synchronization	M	Yes [ ]
SMV2	LACP_Enabled	43.4.8	FALSE for half duplex ports, otherwise TRUE	M	Yes [ ]

43.7.17 Receive machine

Item	Feature	Subclause	Value/Comment	Status	Support
RM1	Receive machine	43.4.12	As defined in Figure 43–10 and associated parameters	M	Yes [ ]
RM2	Validation of LACPDUs		No validation of Version Number, TLV_type, or Reserved fields	M	Yes [ ]

43.7.18 Periodic Transmission machine

Item	Feature	Subclause	Value/Comment	Status	Support
PM1	Periodic Transmission machine	43.4.13	As defined in Figure 43–11 and associated parameters	M	Yes [ ]

43.7.19 Selection Logic

Item	Feature	Subclause	Value/Comment	Status	Support
	Selection logic requirements	43.4.14.1			
SLM1	Aggregator support		At least one Aggregator per System	M	Yes [ ]
SLM2	Port Keys		Each port assigned an operational Key	M	Yes [ ]
SLM3	Aggregator Keys		Each Aggregator assigned an operational Key	M	Yes [ ]
SLM4	Aggregator Identifiers		Each Aggregator assigned an identifier	M	Yes [ ]
SLM5	Aggregator selection		If same Key assignment as port	M	Yes [ ]
SLM6	Ports that are members of the same Link Aggregation Group		Ports select same Aggregator	M	Yes [ ]
SLM7	Pair of ports connected in loopback		Not select same Aggregator as each other	M	Yes [ ]
SLM8	Port required to be Individual		Not select same Aggregator as any other port	M	Yes [ ]
SLM9	Port is Aggregatable		Not select same Aggregator as any Individual port	M	Yes [ ]
SLM10	Port unable to select an Aggregator		Port not attached to any Aggregator	M	Yes [ ]
SLM11	Further aggregation constraints		Ports may be selected as standby	O	Yes [ ] No [ ]
SLM12	Selected variable		Set to SELECTED or STANDBY once Aggregator is determined	M	Yes [ ]
SLM13	Port enabled		Only when selected and attached to an Aggregator	M	Yes [ ]
SLM14	Recommended default operation of Selection Logic	43.4.14.2	Meets requirements of 43.4.14.2	O	Yes [ ] No [ ]

43.7.20 Mux machine

Item	Feature	Subclause	Value/Comment	Status	Support
XM1	Mux machine	43.4.15	As defined in Figure 43–13 or Figure 43–14, and associated parameters	M	Yes [ ]

43.7.21 Transmit machine

Item	Feature	Subclause	Value/Comment	Status	Support
	Transmitted in outgoing LACPDUs	43.4.16			
TM1	Actor_Port and Actor_Port_Priority			M	Yes [ ]
TM2	Actor_System and Actor_System_Priority			M	Yes [ ]
TM3	Actor_Key			M	Yes [ ]
TM4	Actor_State			M	Yes [ ]
TM5	Partner_Port and Partner_Port_Priority			M	Yes [ ]
TM6	Partner_System and Partner_System_Priority			M	Yes [ ]
TM7	Partner_Key			M	Yes [ ]
TM8	Partner_State			M	Yes [ ]
TM9	CollectorMaxDelay			M	Yes [ ]
TM10	Action when Periodic machine is in the NO_PERIODIC state	43.4.16	Set NTT to FALSE, do not transmit	M	Yes [ ]
TM11	Action when LACP_Enabled is TRUE, NTT is TRUE, and not rate limited	43.4.16	Properly formatted LACPDU transmitted	M	Yes [ ]
TM12	Action when LACP_Enabled is TRUE and NTT is TRUE, when rate limit is in force	43.4.16	Transmission delayed until limit is no longer in force	M	Yes [ ]
TM13	Action when LACPDU has been transmitted	43.4.16	Set NTT to FALSE	M	Yes [ ]
TM14	Action when LACP_Enabled is FALSE	43.4.16	Set NTT to FALSE, do not transmit	M	Yes [ ]

43.7.22 Churn Detection machines

Item	Feature	Subclause	Value/Comment	Status	Support
CM1	Churn Detection machines	43.4.17	As defined in Figure 43–15 and Figure 43–16	CM:M	N/A [ ] Yes [ ]

43.7.23 Marker protocol

Item	Feature	Subclause	Value/Comment	Status	Support
FP1	Respond to all received Marker PDUs	43.5.1	As specified by 43.5.4	M	Yes [ ]
FP2	Use of the Marker protocol	43.5.1	As specified by 43.5.4	O	Yes [ ] No [ ]
FP3	MARKER.request service primitives request rate	43.5.4.1	Maximum of five during any one-second period	MG:M	N/A [ ] Yes [ ]
FP4	Marker PDU Frame format	43.5.3.2	Not Tagged	MG:M	N/A [ ] Yes [ ]
FP5	Marker Response PDU Frame format	43.5.3.2	Not Tagged	M	Yes [ ]
FP6	Marker PDU structure	43.5.3.2	As shown in Figure 43–18 and as described	MG:M	N/A [ ] Yes [ ]
FP7	Marker Response PDU structure	43.5.3.2	As shown in Figure 43–18 and as described	M	Yes [ ]
FP8	Marker Responder State Diagram	43.5.4.2	As specified in Figure 43–19 and 43.5.4.2.1 through 43.5.4.2.3	M	Yes [ ]
FP9	Validation of Marker Request PDUs	43.5.4.2.3	Marker Responder shall not validate the Version Number, Pad, or Reserved fields	M	Yes [ ]

43.7.24 Configuration capabilities and restrictions

Item	Feature	Subclause	Value/Comment	Status	Support
CCR1	Algorithm used to determine subset of ports that will be aggregated in Systems that have limited aggregation capability	43.6.1	As specified in items a) to e) of 43.6.1	M	Yes[ ]
CCR2	Key value modification to generate optimum aggregation	43.6.2		O	Yes [ ] No [ ]
CCR3	Key value modification when System has higher System Aggregation Priority			CCR2:M	N/A [ ] M:Yes [ ]
CCR4	Key value modification when System has lower System Aggregation Priority			CCR2:X	N/A [ ] X:Yes [ ]

43.7.25 Link Aggregation on shared-medium links

Item	Feature	Subclause	Value/Comment	Status	Support
LSM1	Shared-medium links—Configuration	43.6.3	Configured as Individual links	M	Yes [ ]
LSM2	Shared-medium links—Operation of LACP	43.6.3	LACP is disabled	M	Yes [ ]

Annex 36A

(informative)

Jitter test patterns

This annex defines test patterns that allow 1000BASE-X PMDs to be tested for compliance while in a system environment. The patterns may be implemented at a bit, code-group, or frame level and may be used for transmitter testing. The receiver may not have the capability to accept these diagnostic sequences; however, system debug can be improved if a receiver is able to test for one or more of these patterns and report bit errors (e.g. 8B/10B decoder errors) back to the user.

36A.1 High-frequency test pattern

The intent of this test pattern is to test random jitter (RJ) at a BER of 10^{-12}, and also to test the asymmetry of transition times. This high-frequency test pattern generates a one, or light on, for a duration of 1 bit time, followed by a zero, or light off, for a duration of 1 bit time. This pattern repeats continuously for the duration of the test. For example: 1010101010101010101010101010101010101010...

NOTE—This pattern can be generated by the repeated transmission of the D21.5 code-group. Disparity rules are followed.

36A.2 Low-frequency test pattern

The intent of this test pattern is to test low-frequency RJ and also to test PLL tracking error. This low frequency test pattern generates a one, or light on, for a duration of 5 bit times, followed by a zero, or light off, for a duration of 5 bit times. This pattern repeats continuously for the duration of the test. For example: 1111100000111110000011111000001111100000...

NOTE—This pattern can be generated by the repeated transmission of the K28.7 code-group. Disparity rules are followed.

36A.3 Mixed frequency test pattern

The intent of this test pattern is to test the combination of RJ and deterministic jitter (DJ). This mixed frequency test pattern generates a one, or light on, for a duration of 5 bit times, followed by a zero, or light off, for a duration of 1 bit times, followed by a one for 1 bit time followed by a zero for 1 bit time followed by a one for 2 bit times followed by a zero for 5 bit times followed by a one for 1 bit time followed by a zero for 1 bit time followed by a one for 1 bit time followed by a zero for 2 bit times. This pattern repeats continuously for the duration of the test. For example: 1111101011000001010011111010110000010100...

NOTE—This pattern can be generated by the repeated transmission of the K28.5 code-group. Disparity rules are followed.

36A.4 Long continuous random test pattern

The long continuous random test pattern is a random test pattern intended to provide broad spectral content and minimal peaking that can be used for the measurement of jitter at either a component or system level.

NOTE—The derivation of this pattern may be found in *Fibre Channel Jitter Working Group Technical Report* [B36]. This technical report modified the original RPAT as defined by Fibre Channel so that it would maintain its intended qualities but fit into a Fibre Channel frame. This annex uses similar modifications to fit the same RPAT into an 802.3® frame.

The long continuous random test pattern consists of a continuous stream of identical packets, separated by a minimum IPG. Each packet is encapsulated within SPD and EPD delimiters as specified in Clause 36 in the ordinary way. The contents of each packet is composed of the following octet sequences, as observed at the GMII, before 8B/10B coding.

Each packet in the long continuous random test pattern consists of 8 octets of PREAMBLE/SFD, followed by 1512 data octets (126 repetitions of the 12-octet modified RPAT sequence), plus 4 CRC octets, followed by a minimum IPG of 12 octets of IDLE.

PREAMBLE/SFD:

55 55 55 55 55 55 55 D5

MODIFIED RPAT SEQUENCE (LOOP 126 TIMES)

BE D7 23 47 6B 8F B3 14 5E FB 35 59

CRC

94 D2 54 AC

IPG (TX_EN and TX_ER low)

00 00 00 00 00 00 00 00 00 00 00 00

END

36A.5 Short continuous random test pattern

The short continuous random test pattern is a random test pattern intended to provide broad spectral content and minimal peaking that can be used for the measurement of jitter at either a component or system level.

NOTE—The derivation of this pattern may be found in *Fibre Channel Jitter Working Group Technical Report* [B36]. This technical report modified the original RPAT as defined by Fibre Channel so that it would maintain its intended qualities but fit into a Fibre Channel frame. This annex uses similar modifications to fit the same RPAT into an 802.3® frame.

The short continuous random test pattern consists of a continuous stream of identical packets, separated by a minimum IPG. Each packet is encapsulated within SPD and EPD delimiters as specified in Clause 36 in the ordinary way. The contents of each packet is composed of the following octet sequences, as observed at the GMII, before 8B/10B coding.

Each packet in the short continuous random test pattern consists of 8 octets of PREAMBLE/SFD, followed by 348 data octets (29 repetitions of the 12-octet modified RPAT sequence), plus 4 CRC octets, followed by a minimum IPG of 12 octets of IDLE.

The format of this packet is such that the PCS will generate the following ordered sets for the IPG: /T/ /R/ /I1/ /I2/ /I2/ /I2/ /I2/

PREAMBLE/SFD:

55 55 55 55 55 55 55 D5

MODIFIED RPAT SEQUENCE (LOOP 29 TIMES)

BE D7 23 47 6B 8F B3 14 5E FB 35 59

CRC

2F E0 AA EF

IPG (TX_EN and TX_ER low)

00 00 00 00 00 00 00 00 00 00 00 00

END

Annex 36B

(informative)

8B/10B transmission code running disparity calculation examples

Detection of a invalid code-group in the 8B/10B transmission code does not necessarily indicate that the code-group in which the error was detected was the one in which the error occurred. Invalid code-groups may result from a prior error that altered the running disparity of the bit stream but that did not result in a detectable error at the code-group in which the error occurred. The examples shown in Tables 36B–1 and 36B–2 exhibit this behavior. The example shown in Table 36B–3 exhibits the case where a bit error in a received code-group is detected in that code-group, affects the next code group, and error propagation is halted upon detection of the running disparity error.

Table 36B–1—RD error detected two code-groups following error

Stream	Code-group			Code-group			Code-group		
	RD		RD		RD			RD	
Transmitted code-group	–	D21.1	–		D10.2	–		D23.5	+
Transmitted bit stream	–	101010 - 1001	–		010101 – 0101	–		111010 + 1010	+
Received bit stream	–	101010 - 10**1**1[a]	+		010101 + 0101	+		111010 **+**[b] 1010	+
Received code-group	–	D21.0	+		D10.2	+		invalid code-group[b]	+[c]

[a]Bit error introduced (1001 ⇒ 10**1**1)
[b]Nonzero disparity blocks must alternate in polarity (+ ⇒ –). In this case, RD does not alternate (+ ⇒ **+**), the received code group is not found in the Current RD+ column in either Table 36–1a or Table 36–2, and an invalid code-group is recognized.
[c]Running disparity is calculated on the received code-group regardless of the validity of the received code-group. Nonzero disparity blocks prevent the propagation of errors and normalize running disparity to the transmitted bit stream (i.e., equivalent to the received bit stream had an error not occurred).

Table 36B–2—RD error detected in next code-group following error

Stream	Code-group			Code-group			Code-group		
	RD		RD		RD			RD	
Transmitted code-group	–	D21.1	–		D23.4	–		D23.5	+
Transmitted bit stream	–	101010 – 1001	–		111010 + 0010	–		111010 + 1010	+
Received bit stream	–	101010 – 10**1**1[a]	+		111010 **+**[b] 0010	–		111010 + 1010	+
Received code-group	–	D21.0	+		invalid code-group[b]	–		D23.5	+

[a]Bit error introduced (1001 ⇒ 10**1**1)
[b]Nonzero disparity blocks must alternate in polarity (+ ⇒ –).

Table 36B–3—A single bit error affects two received code-groups

Stream	Code-group			Code-group			Code-group		
	RD		RD		RD			RD	
Transmitted code-group	–	D3.6 (FCS3)	–	K29.7 (/T/)		–	K23.7 (/R/)		–
Transmitted bit stream	–	110001 - 0110	–	101110 + 1000		–	111010 + 1000		–
Received bit stream	–	110001 - 011**1**[a]	+[b]	101110	+[c] 1000	–	111010 + 1000		–
Received code-group	–	invalid code-group[d]	+	invalid code-group[e]		–	K23.7 (/R/)		–

[a]Bit error introduced (0110 ⇒ 011**1**).
[b]Nonzero disparity blocks must alternate in polarity (– ⇒ +). Received RD differs from transmitted RD.
[c]Nonzero disparity blocks must alternate in polarity (+ ⇒ –). Invalid code-group due to RD error since RD remains at +.
[d]Received code-group is not found in either Table 36–1a or Table 36–2.
[e]Nonzero disparity blocks prevent the propagation of errors and normalize running disparity to the transmitted bit stream (i.e. equivalent to the received bit stream had an error not occurred). All code-groups contained in PCS End_of_Packet delimiters (/T/R/R or /T/R/K28.5/) include nonzero disparity blocks.

Annex 38A

(informative)

Fiber launch conditions

38A.1 Overfilled Launch

Overfilled Launch (OFL), as described in IEC 60793-1-4 [B24], is the standard launch used to define optical fiber bandwidth. This launch uniformly overfills the fiber both angularly and spatially. It excites both radial and azimuthal modes of the fiber equally, thus providing a reproducible bandwidth which is insensitive to small misalignments of the input fiber. It is also relatively insensitive to microbending and macrobending when they are not sufficient to affect power distribution carried by the fiber. A restricted launch gives a less reproducible bandwidth number and is dependent on an exact definition of the launch. Overfilled launch is commonly used to measure the bandwidth of LED-based links.

38A.2 Radial Overfilled Launch (ROFL)

A Radial Overfilled Launch is created when a multimode optical fiber is illuminated by the coherent optical output of a source operating in its lowest order transverse mode in a manner that excites predominantly the radial modes of the multimode fiber. This contrasts with the OFL, which is intended to excite both radial and azimuthal modes of the fiber equally. In practice an ROFL is obtained when

a) A spot of laser light is projected onto the core of the multimode fiber,
b) The laser spot is approximately symmetrical about the optical center of the multimode fiber,
c) The optical axis of both the fiber and the laser beam are approximately aligned,
d) The angle of divergence of the laser beam is less than the numerical aperture of the multimode fiber,
e) The laser spot is larger than the core of the multimode fiber.

An ROFL cannot be created using a multi-transverse mode laser or by simply projecting a speckle pattern through an aperture.

Annex 40A

(informative)

Additional cabling design guidelines

This annex provides additional cabling guidelines when installing a new Category 5 balanced cabling system or using an existing Category 5 balanced cabling system. These guidelines are intended to supplement those in Clause 40. 1000BASE-T is designed to operate over 4-pair unshielded twisted-pair cabling systems that meet both the Category 5 requirements described in ANSI/TIA/EIA-568A (1995) and ISO/IEC 11801:1995, and the additional transmission parameters of return loss, ELFEXT loss, and MDELFEXT loss specified in 40.7. There are additional steps that may be taken by network designers that provide additional operating margins and ensure the objective BER of 10^{-10} is achieved. Cabling systems that meet or exceed the specifications in 40.7 for a worst case 4-connector topology are recommended for new installations. Whether installing a new Category 5 balanced cabling system or reusing one that is already installed, it is *highly recommended* that the cabling system be measured/certified before connecting 1000BASE-T equipment following the guidelines in (proposed) ANSI/TIA/EIA TSB95.

40A.1 Alien crosstalk

40A.1.1 Multipair cabling (i.e., greater than 4-pair)

Multiple Gigabit Ethernet links [(n*4-Pair) with n greater than 1] should not share a common sheath as in a 25-pair binder group in a multipair cable. When the multipair cable is terminated into compliant connecting hardware (TIA does not specify 25 position connecting hardware), the NEXT loss contributions between the adjacent 4-pair gigabit ethernet link, from connecting hardware and the cable combined, cannot be completely cancelled.

40A.1.2 Bundled or hybrid cable configurations

Another source of alien crosstalk can occur in a bundled or hybrid cable configuration where two or more 4-pair cables are assembled together.

In order to limit the noise coupled between adjacent 1000BASE-T link segments in a bundled or hybrid cable configuration, the PSNEXT loss between a 1000BASE-T duplex channel in a link segment and all duplex channels in adjacent 1000BASE-T link segments should be greater than $35 - 15*\log(f/100)$ (dB) at all frequencies from 1 MHz to 100 MHz.

40A.2 Cabling configurations

The primary application for the Clause 40 specification is expected to be between a workstation and the local telecommunications closet. In commercial buildings this application is generally referred to as the horizontal cabling subsystem. As specified in ANSI/TIA/EIA-568-A and ISO/IEC 11801: 1995 the maximum length of a horizontal subsystem building wiring channel is 100 m. The channel consists of cords, cables, and connecting hardware. The maximum configuration for this channel is shown in Figure 40A–1.

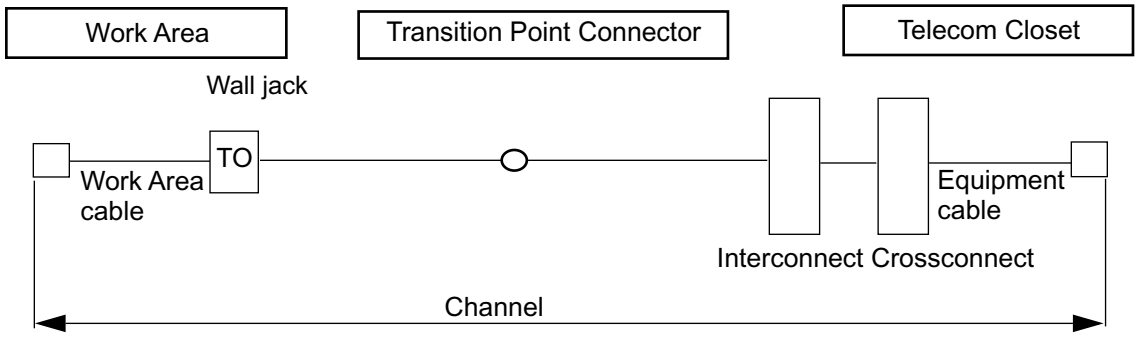


Figure 40A–1—Maximum horizontal subsystem configuration

On the other hand, a minimum configuration can be achieved by removing the patch cord and transition point, which is shown in Figure 40A–2.

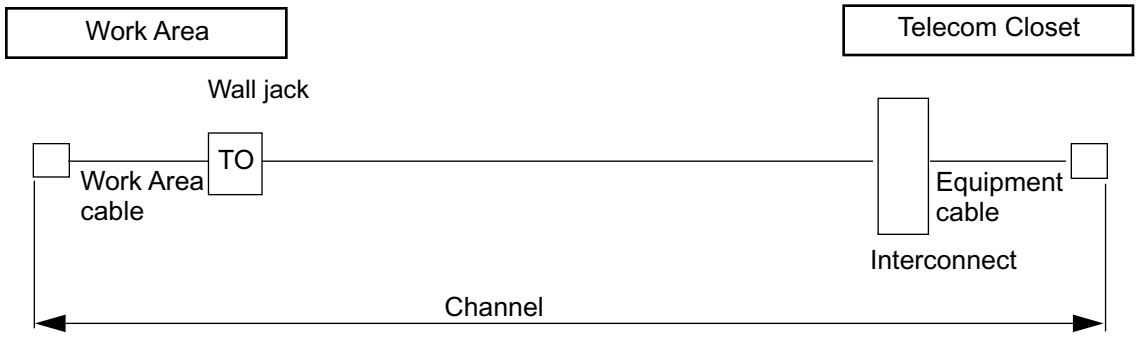


Figure 40A–2—Minimum horizontal subsystem configuration

1000BASE-T is designed to operate over a channel that meets the specifications of 40.7 and the channel configuration shown in Figure 40A–1. However, if the channel specifications of 40.8 cannot be met when using the channel configuration shown in Figure 40A–1, the configuration shown in Figure 40A–2 is recommended. This optimized channel for a 1000BASE-T link segment deletes the transition point and runs an equipment patch cord directly between the LAN equipment and the connector termination of the permanent link. This reduces the number of connectors and their associated crosstalk in the link. The minimum link configuration:

a) Minimizes crosstalk, both near-end and far-end, which maximizes the BER margin; and
b) Minimizes link insertion loss.

Annex 40B

(informative)

Description of cable clamp

This annex describes the cable clamp used in the common-mode noise rejection test of 40.6.1.3.3, which is used to determine the sensitivity of the 1000BASE-T receiver to common-mode noise from the link segment. As shown in Figure 40B–1, the clamp is 300 mm long, 58 mm wide, 54 mm high with a center opening of 6.35 mm (0.25 in). The clamp consists of two halves that permit the insertion of a cable into the clamp.

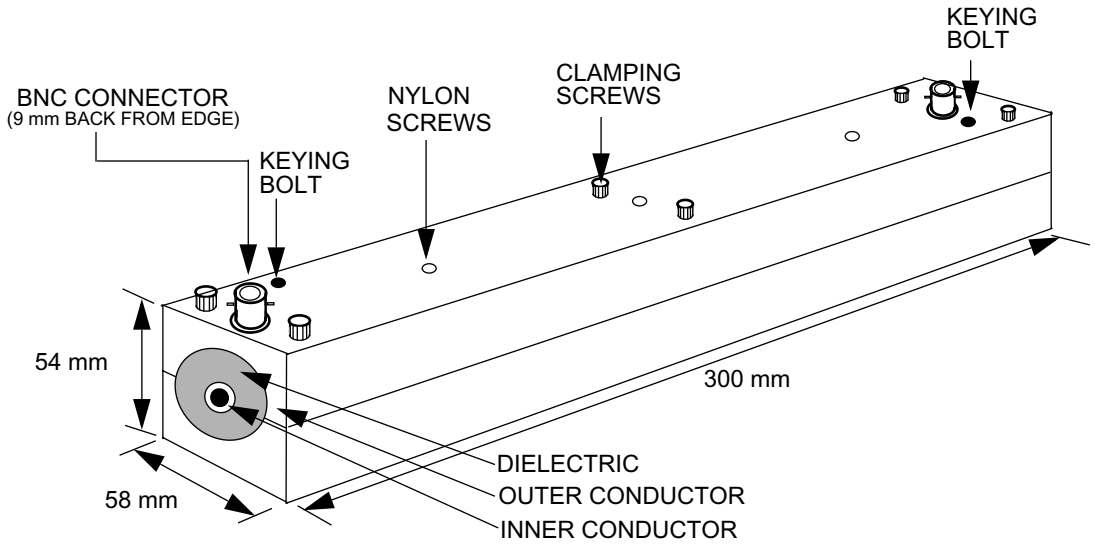


Figure 40B–1—Cable clamp

The clamp has a copper center conductor and an aluminum outer conductor with a high density polyethylene dielectric. The following is a review of the construction and materials of the clamp:

a) *Inner conductor*—Copper tubing with an inner diameter of 6.35 mm (0.25 in) and an outer diameter of 9.4 mm (0.37 in).
b) *Outer conductor*—Aluminum bar that is 300 mm long and approximately 54 mm by 58 mm. The bar is milled to accept the outer diameter of the dielectric material.
c) *Dielectric*—High Density Polyethylene (Residual, TypeF) with dielectric constant of 2.32. An outside diameter of 33.5 mm and an inner diameter that accepts the outside diameter of the copper inner conductor.
d) *Connectors*—BNC connectors are located 9 mm (0.39 in) from each end of the clamp and are recessed into the outer conductor. The center conductor of the connector is connected to the inter conductor as shown in Figure 40B–2.
e) *Clamping screws*—Six screws are used to connect the two halves of the clamp together after the cable has been inserted. Although clamping screws are shown in Figure 40B–1, any clamping method may be used that ensures the two halves are connected electrically and permits quick assembly and disassembly.
f) *Nylon screws*—Used to align and secure the inner conductor and dielectric to the outer conductor. The use and location of the screws is left to the manufacturer.
g) *Keying bolts*—Two studs used to align the two halves of the clamp.

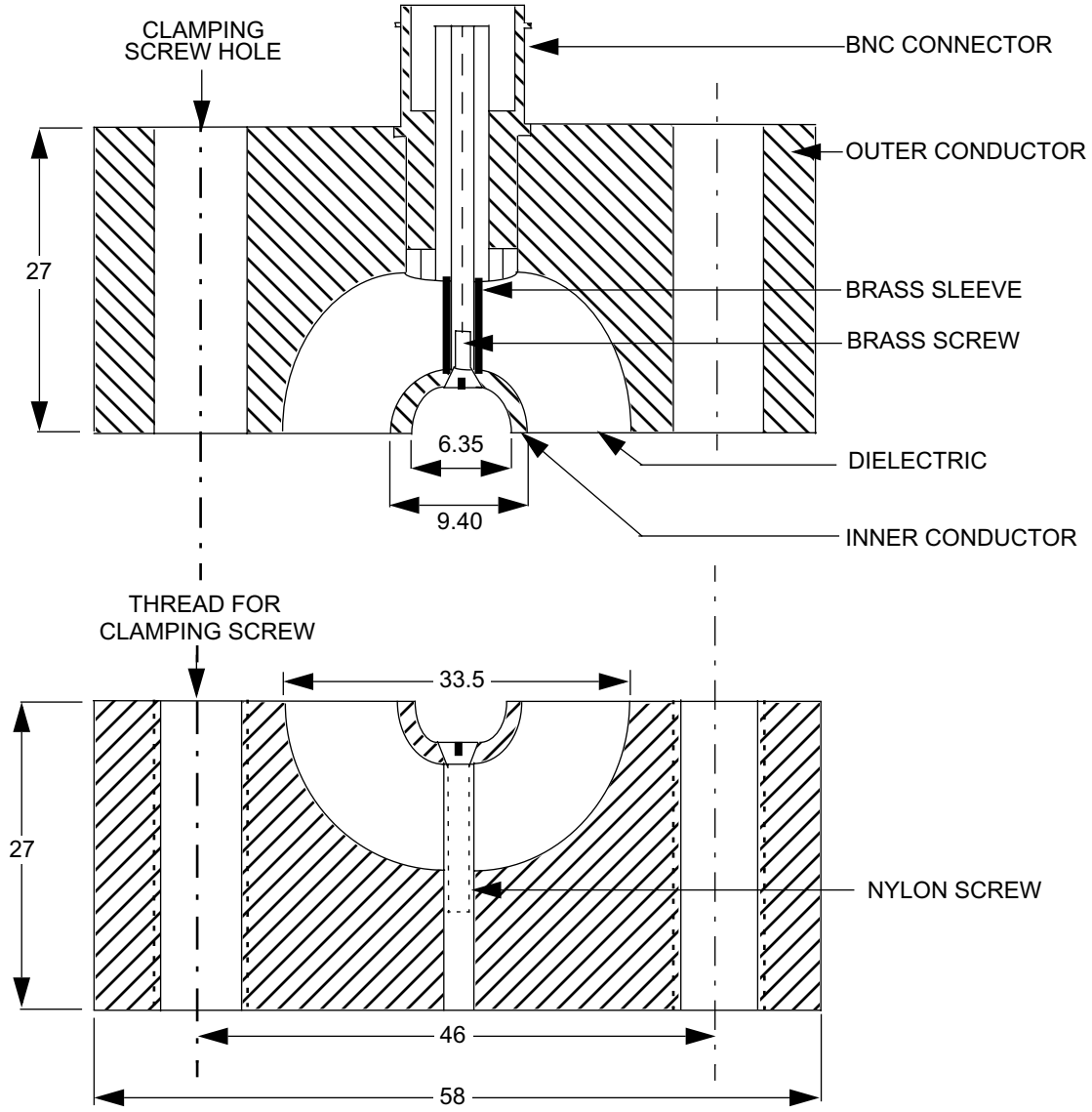


Figure 40B–2—Cross-section of cable clamp

As shown in Figure 40B–2 the inner conductor on the bottom half of the clamp extends slightly (~ 0.1mm) above the dielectric to ensure there is good electrical connection with the inner conductor of the top half of the clamp along the full length of the conductor when the two halves are clamped together.

The electrical parameters of the clamp between 1MHz and 250 MHz are as follows:

a) Insertion loss: < 0.2 dB
b) Return loss: > 20.0 dB

40B.1 Cable clamp validation

In order to ensure the cable clamp described above is operating correctly, the following test procedure is provided. Prior to conducting the following test shown in Figure 40B–3, the clamp should be tested to ensure the insertion loss and return loss are as specified above. The cable clamp validation test procedure uses a well-balanced 4-pair Category 5 unshielded test cable or better that meets the specifications of 40.7. The test hardware consists of the following:

a) *Resistor network*—Network consists of three 50 ±0.1% Ω resistors; two resistors are connected in series as a differential termination for cable pairs and the other resistor is connected between the two and the ground plane as a common-mode termination.

b) *Balun*—3 ports, laboratory quality with a 100 Ω differential input, 50 Ω differential output, and a 50 Ω common-mode output:

Insertion Loss (100 Ω balanced <->50 Ω unbalanced): <1.2 dB (1-350MHz)

Return Loss: >20 dB (1-350 MHz)

Longitudinal Balance: >50 dB (1-350 MHz)

c) *Test cable*—4-pair 100 Ω UTP category 5 balanced cable at least 30 m long.

d) *Chokes (2)*—Wideband Ferrite Material:

Inter-diameter: 6.35 to 6.86 mm

Impedance: 250 Ω @ 100 MHz

e) *Ground plane*—Copper sheet or equivalent.

f) *Signal generator*

g) *Oscilloscope*

h) *Receiver*

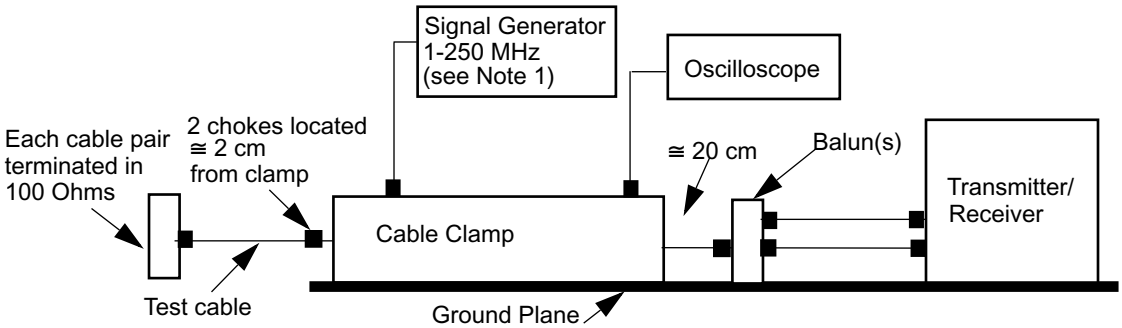


Figure 40B–3—Cable clamp validation test configuration

With the test cable inserted in the cable clamp, a signal generator with a 50 Ω output impedance is connected to one end of the cable clamp and an oscilloscope with a 50 Ω input impedance is connected to the other end. The signal generator shall be capable of providing a sine wave signal of 1 MHz to 250 MHz. The output of the signal generator is adjusted for a voltage of 1.0 Vrms (2.83 Vpp) at 20 MHz on the oscilloscope. The remainder of the test is conducted without changing the signal generator voltage. The cable pairs not connected to the balun are terminated in a resistor network, although when possible it is recommended that each cable pair be terminated in a balun. It very important that the cable clamp, balun, receiver, and resistor networks have good contact with the ground plane. The two chokes, which are located next to each other, are located approximately 2.0 cm from the clamp. The cable between the clamp and the balun should be straight and not in contact with the ground plane.

The differential-mode and common-mode voltage outputs of the balun should meet the limits shown in Table 40B–1 over the frequency range 1 MHz to 250 MHz for each cable pair. The differential mode voltage at the output of the balun must be increased by 3 dB to take into account the 100-to-50 impedance matching loss of the balun.

Table 40B–1—Common- and differential-mode output voltages

Frequency (*f*)	Common-mode voltage	Differential-mode voltage
1-30 MHz	<0.1+0.97(f/30) Vpp	<2.4 + 19.68 (f/30) mVpp
30-80 MHz	<1.07 Vpp	<22 mVpp
80-250 MHz	<1.07 – 0.6 (f-80)/170 Vpp	<22 mVpp

NOTE—Prior to conducting the validation test the cable clamp should be tested without the cable inserted to determine the variation of the signal generator voltage with frequency at the output of the clamp. The signal generator voltage should be adjusted to 1 Vrms (2.83 Vpp) at 20 MHz on the oscilloscope. When the frequency is varied from 20 MHz to 250 MHz, the voltage on the oscilloscope should not vary more than ±7.5%. If the voltage varies more than ±7.5%, then a correction factor must be applied at each measurement frequency.

Annex 40C

(informative)

Add-on interface for additional Next Pages

This annex describes a technique for implementing Auto-Negotiation for 1000BASE-T when the implementor wishes to send additional Next Pages (other than those required to configure for 1000BASE-T operation). To accomplish this mode of Auto-Negotiation, the implementor must ensure that the three Next Pages required for 1000BASE-T configuration are sent first.

The add-on interface described in this annex shows one technique for supporting the transmission of additional Next Pages. This mechanism utilizes the existing Clause 28 Auto-Negotiation mechanism and variables defined in Clause 28. Its purpose is merely to provide optional NEXT PAGE WAIT responses to the Auto-Negotiation Arbitration state diagram (see Figure 28–16).

The add-on interface for Auto-Negotiation is intended to interface directly between the defined registers and the Auto-Negotiation mechanism defined in Clause 28. The mechanism described includes five main blocks (see Figure 40C–1).

The first three blocks are used by the MASTER-SLAVE resolution function. They are used to generate and store random seeds and to resolve the status of the MASTER-SLAVE relationship. Their operation is described later in this annex. The final two blocks, the transmit state machine for the 1000BASE-T Next Page exchange and the receive state machine for the 1000BASE-T Next Page exchange, are described in this annex.

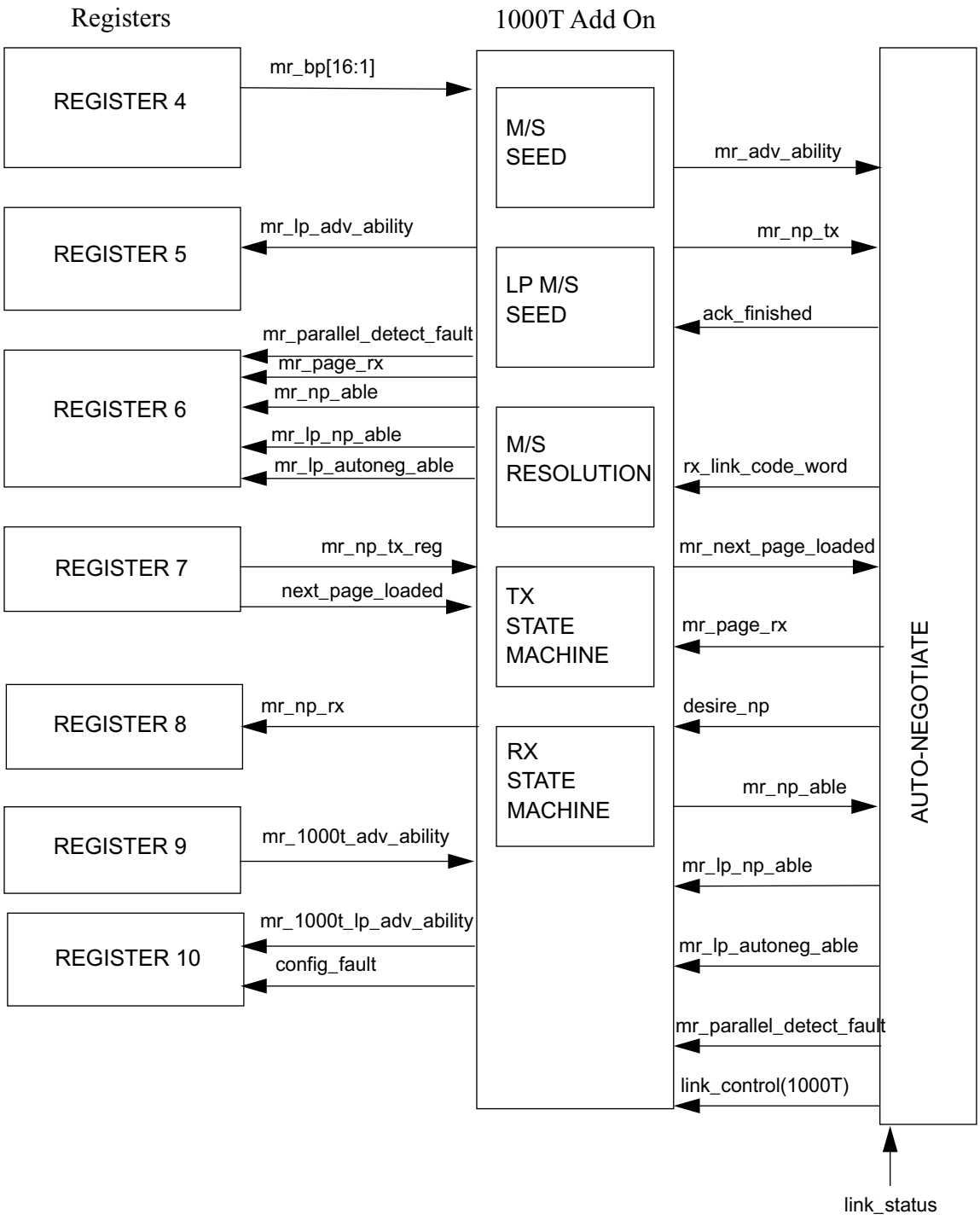


Figure 40C–1—Auto-negotiate add-on diagram for 1000BASE-T

NOTE—When the exchange of Next Pages is complete, the MASTER-SLAVE relationship can be determined using Table 40–5 with the 1000BASE-T Technology Ability Next Page bit values specified in Table 40–4 and information received from the link partner. This process is conducted at the entrance to the FLP LINK GOOD CHECK state shown in the Auto-Negotiation Arbitration State Diagram (Figure 28–16).

40C.1 State variables

mr_bp
: This variable is used as an intermediate signal from register 4. Normally register 4 would directly source the mr_adv_ability information. This information is now sourced from the transmit state machine.

mr_1000t_adv_ability
: A 16-bit array used to store and indicate the contents of register 9.

mr_1000t_lp_adv_ability
: A 16-bit array used to write information to register 10.

mr_np_tx_reg
: This variable is an intermediate signal from register 7. Normally register 7 would directly source the information to the Auto-Negotiation function via mr_np_tx. This information is now sourced from the transmit state machine.

mr_np_rx
: A 16-bit array used to write information to register 8.
Values: Zeros; data bit is logical zero.
One; data bit is logical one.

config_fault
: This variable indicates the result of the MASTER-SLAVE resolution function.

next_page_loaded
: This variable is an intermediate signal from register 7. Normally register 7 would directly source the information to the Auto-Negotiation function via mr_next_page_loaded. This information is now sourced from the transmit state machine.

reg_random
: An 11-bit array used to store the received random seed from the link partner. It is used by the MASTER-SLAVE resolution function.

1000T_capable
: This variable is used merely to show the local device is 1000Base-T capable. It is shown to illustrate the path that a non-1000Base-T device would take within the auto negotiation mechanism.

ATMP_CNT
: This variable is used to count the number of failed MASTER-SLAVE resolutions. It has a maximum value of 7.

All other signals are defined in Clause 28.

40C.2 State diagrams

40C.2.1 Auto-Negotiation Transmit state machine add-on for 1000BASE-T

The Auto-Negotiation transmit state machine add-on (see Figure 40C–2) is responsible for sending the Base Page, 1000BASE-T Next Pages, as well as additional Next Pages as specified by the management interface. 1000BASE-T Next Pages will automatically be sent by the PHY whenever there are no additional Next Pages to be sent. If the user desires to send additional Next Pages, then the user must first send three pages of any format. Management will automatically replace these three pages with the appropriate 1000BASE-T Message Page and the two following unformatted pages and then will send the additional Next Pages as specified by the user. All other steps are performed by the management interface. The management interface is now required to complete the Next Page exchange by sourcing its own NULL page. This is shown in Figure 40C–2 for illustration only.

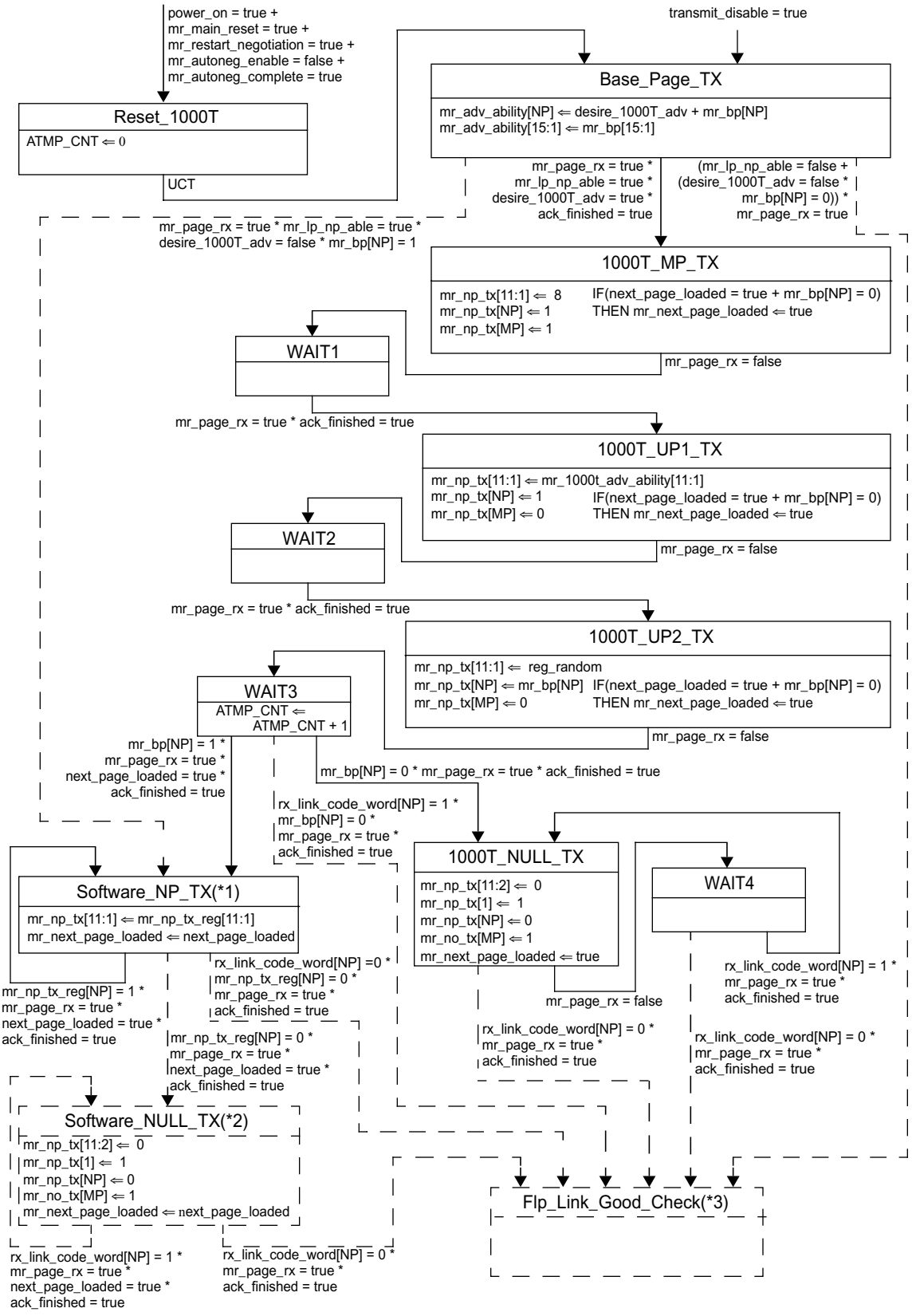


Figure 40C–2—Auto-Negotiation Transmit state diagram add-on for 1000BASE-T

NOTES for Figure 40C–2

1—(Software_NP_TX) If the user desires to send additional Next Pages, then the contents of the first three Next Pages will be overwritten by the three 1000BASE-T Next Pages. In this case, the user is responsible for stepping through the Next Page sequence (by creating the initial three Next Pages to be overwritten by the three 1000BASE-T Next Pages); otherwise the process is automatic. (next_page_loaded signals clear operation as per Clause 28.)

2—(Software_NULL_TX) This is shown for illustration only. This is done in software and is required.

3—(Flp_Link_Good_Check) This is shown for illustration only. This state is from the Auto-Negotiation arbitration state diagram and indicates the conclusion of pages being sent. (The transition 1000T_capable = false is to show sequence for non 1000BASE-T implementations.)

40C.2.2 Auto-Negotiation receive state diagram add-on for 1000BASE-T

The Auto-Negotiation receive state machine add-on for 1000BASE-T Next Pages (see Figure 40C–3) is responsible for receiving the Base Page, 1000BASE-T Next Pages, and any additional Next Pages received. 1000BASE-T Next Pages will automatically be received whenever the user does not wish to participate in Next Page exchanges. In this case, the appropriate 1000BASE-T message page and its two unformatted pages will automatically be received and stored in their appropriate registers. Any additional Next Pages received will still be placed in register 8, but will be overwritten automatically when a new page is received. The net result is that the management interface does not need to poll registers 6 and 8. The information in register 8 will be meaningless in this case. If the user desires to participate in additional Next Page exchanges via setting the appropriate bit in register 4, the user now becomes responsible (as was previously the case) for defining how this will be accomplished. In this situation, the first three Next Pages received may be 1000BASE-T and should be discarded. This information will automatically be stored internally in the appropriate register 10 and reg_random. The management interface/user can ignore the information received for the 1000BASE-T Next Pages.

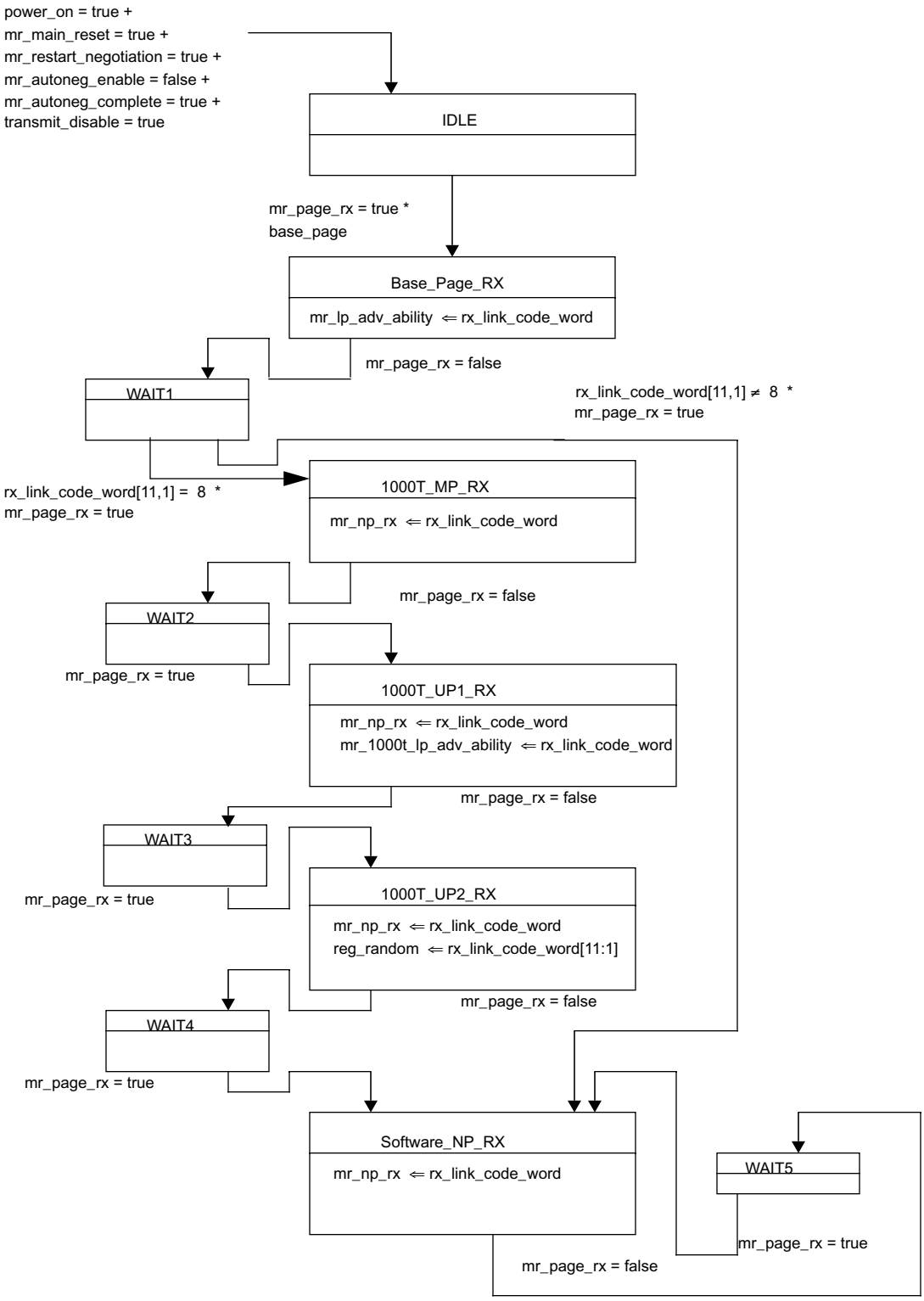


Figure 40C–3—Auto-Negotiation Receive state diagram add-on for 1000BASE-T

Annex 43A

(informative)

Collection and Distribution functions

43A.1 Introduction

The specification of the Collection and Distribution functions was defined with the following considerations in mind:

a) Frame duplication is not permitted.
b) Frame ordering must be preserved in aggregated links. Strictly, the MAC service specification (ISO/IEC 15802-1) states that order must be preserved for frames with a given SA, DA, and priority; however, this is a tighter constraint than is absolutely necessary. There may be multiple, logically independent conversations in progress between a given SA-DA pair at a given priority; the real requirement is to maintain ordering within a conversation, though not necessarily between conversations.
c) A single algorithm can be defined for the collection function that is independent of the distribution function(s) employed by the Partner System.
d) In the interests of simplicity and scalability, the collection function should not perform re-assembly functions, re-order received frames, or modify received frames. Distribution functions, therefore, do not make use of segmentation techniques, do not label or otherwise modify transmitted frames in any way, and must operate in a manner that will inherently ensure proper ordering of received frames with the specified collector.
e) The distribution and collection functions need to be capable of handling dynamic changes in aggregation membership.
f) There are expected to be many different topologies and many different types of devices in which Link Aggregation will be employed. It is therefore unlikely that a single distribution function will be applicable in all cases.

A simple collection function has been specified. The Collector preserves the order of frames received on a given link, but does not preserve frame ordering amongst links. The distribution function maintains frame ordering by

— Transmitting frames of a given conversation on a single link at any time.
— Before changing the link on which frames of a given conversation are transmitted, ensuring that all previously transmitted frames of that conversation have been received to a point such that any subsequently transmitted frames received on a different links will be delivered to the MAC Client at a later time.

Given the wide variety of potential distribution algorithms, the normative text in Clause 43 specifies only the requirements that such algorithms must meet, and not the details of the algorithms themselves. To clarify the intent, this informative annex gives examples of distribution algorithms, when they might be used, and the role of the Marker protocol (43.5) in their operation. The examples are not intended to be either exhaustive or prescriptive; implementors may make use of any distribution algorithms as long as the requirements of Clause 43 are met.

43A.2 Port selection

A distribution algorithm selects the port used to transmit a given frame, such that the same port will be chosen for subsequent frames that form part of the same conversation. The algorithm may make use of information carried in the frame in order to make its decision, in combination with other information associated with the frame, such as its reception port in the case of a MAC Bridge.

The algorithm may assign one or more conversations to the same port, however, it must not allocate some of the frames of a given conversation to one port and the remainder to different ports. The information used to assign conversations to ports could include the following:

a) Source MAC address
b) Destination MAC address
c) The reception port
d) The type of destination address (individual or group MAC address)
e) Ethernet Length/Type value (i.e., protocol identification)
f) Higher layer protocol information (e.g., addressing and protocol identification information from the LLC sublayer or above)
g) Combinations of the above

One simple approach applies a hash function to the selected information to generate a port number. This produces a deterministic (i.e., history independent) port selection across a given number of ports in an aggregation. However, as it is difficult to select a hash function that will generate a uniform distribution of load across the set of ports for all traffic models, it might be appropriate to weight the port selection in favor of ports that are carrying lower traffic levels. In more sophisticated approaches, load balancing is dynamic; i.e., the port selected for a given set of conversations changes over time, independent of any changes that take place in the membership of the aggregation.

43A.3 Dynamic reallocation of conversations to different ports

It may be necessary for a given conversation or set of conversations to be moved from one port to one or more others, as a result of

a) An existing port being removed from the aggregation,
b) A new port being added to the aggregation, or
c) A decision on the part of the Distributor to re-distribute the traffic across the set of ports.

Before moving conversation(s) to a new port, it is necessary to ensure that all frames already transmitted that are part of those conversations have been successfully received. The following procedure shows how the Marker protocol (43.5) can be used to ensure that no mis-ordering of frames occurs:

1) Stop transmitting frames for the set of conversations affected. If the MAC Client requests transmission of further frames that are part of this set of conversations, these frames are discarded.
2) Start a timer, choosing the timeout period such that, if the timer expires, the destination System can be assumed either to have received or discarded all frames transmitted prior to starting the timer.
3) Use the Marker protocol to send a Marker PDU on the port previously used for this set of conversations.
4) Wait until either the corresponding Marker Response PDU is received or the timer expires.
5) Restart frame transmission for the set of conversations on the newly selected port.

The appropriate timeout value depends on the connected devices. For example, the recommended maximum Bridge Transit Delay is 1 second; if the receiving device is a MAC Bridge, it may be expected to have

forwarded or discarded all frames received more than 1 second ago. The appropriate timeout value for other circumstances could be smaller or larger than this by several orders of magnitude. For example, if the two Systems concerned are high-performance end stations connected via Gigabit Ethernet links, then timeout periods measured in milliseconds might be more appropriate. In order to allow an appropriate timeout value to be determined, the Frame Collector parameter CollectorMaxDelay (see 43.2.3) defines the maximum delay that the collector can introduce between receiving a frame from a port and either delivering it to the MAC Client or discarding it. This value will be dependent upon the particular implementation choices that have been made in a System. As far as the operation of the Collector state machine is concerned, CollectorMaxDelay is a constant; however, a management attribute, aAggCollectorMaxDelay (30.7.1.1.32), is provided that allows interrogation and administative control of its value. Hence, if a System knows the value of CollectorMaxDelay that is in use by a Partner System, it can set the value of timeout used when flushing a link to be equal to that value of CollectorMaxDelay, plus sufficient additional time to allow for the propagation delay experienced by frames between the two Systems. A value of zero for the CollectorMaxDelay parameter indicates that the delay imposed by the Collector is less than the resolution of the parameter (10 microseconds). In this case, the delay that must be considered is the physical propagation delay of the channel. Allowing management manipulation of CollectorMaxDelay permits fine-tuning of the value used in those cases where it may be difficult for the equipment to pre-configure a piece of equipment with a realistic value for the physical propagation delay of the channel.

The Marker protocol provides an optimization that can result in faster reallocation of conversations than would otherwise be possible—without the use of markers, the full timeout period would always have to be used in order to be sure that no frames remained in transit between the local Distributor and the remote Collector. The timeout described recovers from loss of Marker or Marker Response PDUs that can occur.

43A.4 Topology considerations in the choice of distribution algorithm

Figure 43A–1 gives some examples of different aggregated link scenarios. In some cases, it is possible to use distribution algorithms that use MAC frame information to allocate conversations to links; in others, it is necessary to make use of higher-layer information.

In example A, there is a many-to-many relationship between end stations communicating over the aggregated link. It would be possible for each switch to allocate conversations to links simply on the basis of source or destination MAC addresses.

In examples B and C, a number of end stations communicate with a single server via the aggregated link. In these cases, the distribution algorithm employed in the server or in Switch 2 can allocate traffic from the server on the basis of destination MAC address; however, as one end of all conversations constitutes a single server with a single MAC address, traffic from the end stations to the server would have to be allocated on the basis of source MAC address. These examples illustrate the fact that different distribution algorithms can be used in different devices, as appropriate to the circumstances. The collection function is independent of the distribution function(s) that are employed.

In examples D and E, assuming that the servers are using a single MAC address for all of their traffic, the only appropriate option is for the distribution algorithm used in the servers and switches to make use of higher-layer information (e.g., Transport Layer socket identifiers) in order to allocate conversations to links. Alternatively, in example E, if the servers were able to make use of multiple MAC addresses and allocate conversations to them, then the switches could revert to MAC Address-based allocation.

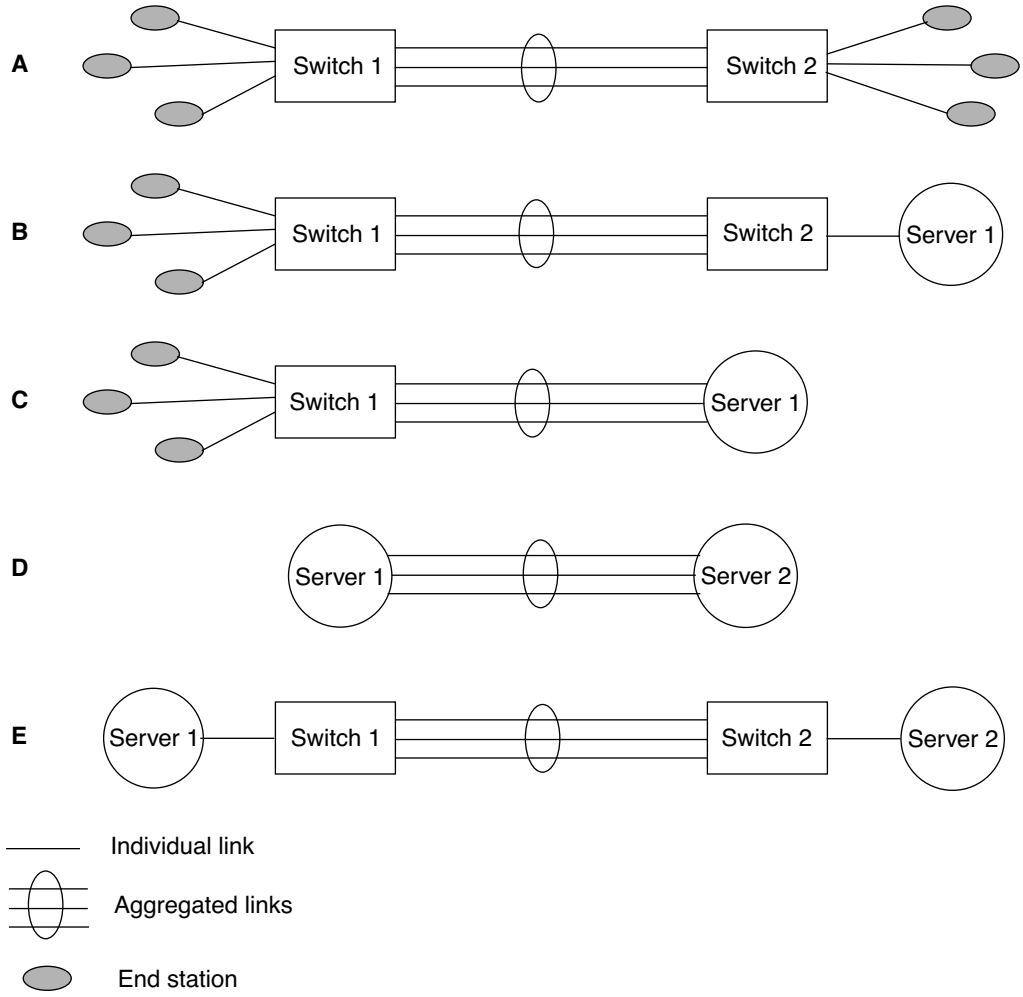


Figure 43A–1—Link aggregation topology examples

Annex 43B

(normative)

Requirements for support of Slow Protocols

43B.1 Introduction and rationale

There are two distinct classes of protocols used to control various aspects of the operation of IEEE 802.3® devices. They are as follows:

a) Protocols such as the MAC Control PAUSE operation (Annex 31B) that need to process and respond to PDUs rapidly in order to avoid performance degradation. These are likely to be implemented as embedded hardware functions, making it relatively unlikely that existing equipment could be easily upgraded to support additional such protocols.

   NOTE—This consideration was one of the contributing factors in the decision to use a separate group MAC address to support LACP and the Marker protocol, rather than re-using the group MAC address currently used for PAUSE frames.

b) Protocols such as LACP, with less stringent frequency and latency requirements. These may be implemented in software, with a reasonable expectation that existing equipment be upgradeable to support additional such protocols, depending upon the approach taken in the original implementation.

In order to place some realistic bounds upon the demands that might be placed upon such a protocol implementation, this annex defines the characteristics of this class of protocols and identifies some of the behaviors that an extensible implementation needs to exhibit.

43B.2 Slow Protocol transmission characteristics

Protocols that make use of the addressing and protocol identification mechanisms identified in this annex are subject to the following constraints:

a) No more than 5 frames shall be transmitted in any one-second period.
b) The maximum number of Slow Protocols is 10.

   NOTE—This is the maximum number of Slow Protocols that use the specified protocol type defined here. That is, there may be more than 10 slow protocols in the universe, but no more than 10 may map to the same Ethernet Length/Type field.

c) The MAC Client data generated by any of these protocols shall be in the normal length range for an IEEE 802.3® MAC frame, as specified in 4.4.2. It is recommended that the maximum length for a Slow Protocol frame be limited to 128 octets.

   NOTE—The Slow Protocols specified in Clause 43 (i.e., LACP and Marker) conform to this recommended maximum.

d) PDUs generated by these protocols shall use the Basic and not the Tagged frame format (see Clause 3).

The effect of these restrictions is to restrict the bandwidth consumed and performance demanded by this set of protocols; the absolute maximum traffic loading that would result is 50 maximum length frames per second per link.

43B.3 Addressing

The Slow_Protocols_Multicast address has been allocated exclusively for use by Slow Protocols PDUs; its value is identified in Table 43B–1.

Table 43B–1—Slow_Protocols_Multicast address

Name	Value
Slow_Protocols_Multicast address	01-80-C2-00-00-02

NOTES

1—This address is within the range reserved by ISO/IEC 15802-3 (MAC Bridges) for link-constrained protocols. As such, frames sent to this address will not be forwarded by conformant MAC Bridges; its use is restricted to a single link.

2—Although the two currently existing Slow Protocols (i.e., LACP and the Marker protocol) always use this MAC address as the destination address in transmitted PDUs, this may not be true for all Slow Protocols. In some yet-to-be-defined protocol, unicast addressing may be appropriate and necessary. Rather, the requirement is that this address not be used by any protocols that are not Slow Protocols.

43B.4 Protocol identification

All Slow Protocols use Type-field encoding of the Length/Type field, and use the Slow_Protocols_Type value as the primary means of protocol identification; its value is shown in Table 43B–2.

Table 43B–2—Slow_Protocols_Type field

Name	Value
Slow_Protocols_Type field	88-09

The first octet of the MAC Client data following the Length/Type field is a protocol subtype identifier that distinguishes between different Slow Protocols. Table 43B–3 identifies the semantics of this subtype.

NOTE—Although this mechanism is defined as part of an IEEE 802.3® standard, it is the intent that the reserved code points in this table will be made available to protocols defined by other working groups within IEEE 802®, should this mechanism be appropriate for their use.

Table 43B–3—Slow Protocols subtypes

Protocol Subtype value	Protocol name
0	Unused—Illegal value
1	Link Aggregation Control Protocol (LACP)
2	Link Aggregation—Marker Protocol
3	Reserved for future use
4	Reserved for future use
5	Reserved for future use
6	Reserved for future use
7	Reserved for future use
8	Reserved for future use
9	Reserved for future use
10	Reserved for future use
11–255	Unused—Illegal values

43B.5 Handling of Slow Protocol frames

Any received MAC frame that carries the Slow_Protocols_Type field value is assumed to be a Slow Protocol frame. An implementation that claims conformance to this standard shall handle all Slow Protocol frames as follows:

a) Discard any Slow Protocol frame that carries an illegal value of Protocol Subtype (see Table 43B–3). Such frames shall not be passed to the MAC Client.

b) Pass any Slow Protocol frames that carry Protocol Subtype values that identify supported Slow Protocols to the protocol entity for the identified Slow Protocol.

c) Pass any Slow Protocol frames that carry Protocol Subtype values that identify unsupported Slow Protocols to the MAC Client.

NOTE—The intent of these rules is twofold. First, they rigidly enforce the maximum number of Slow Protocols, ensuring that early implementations of this mechanism do not become invalidated as a result of "scope creep." Second, they make it clear that the appropriate thing to do in any embedded frame parsing mechanism is to pass frames destined for unsupported protocols up to the MAC Client rather than discarding them, thus allowing for the possibility that, in soft configurable systems, the MAC Client might be enhanced in the future in order to support protocols that were not implemented in the hardware.

43B.6 Protocol Implementation Conformance Statement (PICS) proforma for Annex 43B, Requirements for support of Slow Protocols[14]

43B.6.1 Introduction

The supplier of an implementation that is claimed to conform to Annex 43B shall complete the following Protocol Implementation Conformance Statement (PICS) proforma.

A detailed description of the symbols used in the PICS proforma, along with instructions for completing the PICS proforma, can be found in Clause 21.

43B.6.2 Identification

43B.6.2.1 Implementation identification

Supplier (Note 1)	
Contact point for queries about the PICS (Note 1)	
Implementation Name(s) and Version(s) (Notes 1 and 3)	
Other information necessary for full identification—e.g., name(s) and version(s) of machines and/or operating system names (Note 2)	
NOTES 1—Required for all implementations. 2—May be completed as appropriate in meeting the requirements for the identification. 3—The terms Name and Version should be interpreted appropriately to correspond with a supplier's terminology (e.g., Type, Series, Model).	

43B.6.2.2 Protocol summary

Identification of protocol specification	IEEE Std 802.3-2002®, Annex 43B, Requirements for support of Slow Protocols.
Identification of amendments and corrigenda to the PICS proforma which have been completed as part of the PICS	
Have any Exception items been required? No [ ] Yes [ ] (See Clause 21: the answer Yes means that the implementation does not conform to IEEE Std 802.3-2002®, Annex 43B, Requirements for support of Slow Protocols.)	

Date of Statement	

[14]*Copyright release for PICS proformas:* Users of this standard may freely reproduce the PICS proforma in this annex so that it can be used for its intended purpose and may further publish the completed PICS.

43B.6.2.3 Transmission characteristics

Item	Feature	Subclause	Value/Comment	Status	Support
SP1	Transmission rate	43B.2	Max 5 frames in any one-second period	M	Yes [ ]
SP2	Frame size	43B.2	Normal IEEE 802.3® frame size range (see 4.4.2)	M	Yes [ ]
SP3	Frame format	43B.2	Basic (not Tagged) frame format	M	Yes [ ]

43B.6.2.4 Frame handling

Item	Feature	Subclause	Value/Comment	Status	Support
FH1	Handling of Slow Protocol frames	43B.5	As specified in 43B.5	M	Yes [ ]

Annex 43C

(informative)

LACP standby link selection and dynamic Key management

43C.1 Introduction

While any two ports on a given system that have been assigned the same administrative Key may be capable of aggregation, it is not necessarily the case that an arbitrary selection of such ports can be aggregated. (Keys may have been deliberately assigned to allow one link to be operated specifically as a hot standby for another). A system may reasonably limit the number of ports attached to a single Aggregator, or the particular way more than two ports can be combined.

In cases where both communicating systems have constraints on aggregation, it is necessary for them both to agree to some extent on the links to be selected for aggregation and on which not to use. Otherwise it might be possible for the two systems to make different selections, possibly resulting in no communication at all.

When one or more links have to be selected as standby, it is possible that they could be used as part of a different Link Aggregation Group. For this to happen, one or another of the communicating systems has to change the operational Key values used for the ports attached to those links.

If the operational Key values were to be changed independently by each system, the resulting set of aggregations could be unpredictable. It is possible that numerous aggregations, each containing a single link, may result. Worse, with no constraint on changes, the process of both systems independently searching for the best combination of operational Key values may never end.

This annex describes protocol rules for standby link selection and dynamic Key management. It provides examples of a dynamic Key management algorithm applied to connections between systems with various aggregation constraints.

43C.2 Goals

The protocol rules presented

a) Enable coordinated, predictable, and reproducible standby link selections.
b) Permit predictable and reproducible partitioning of links into aggregations by dynamic Key management.

They do not require

c) A LACP system to understand all the constraints on aggregations of multiple ports that might be imposed by other systems.
d) Correct configuration of parameters, i.e., they retain the plug and play attributes of LACP.

43C.3 Standby link selection

Every link between systems operating LACP is assigned a unique priority. This priority comprises (in priority order) the System Priority, System ID, Port Priority, and Port Number of the higher-priority system. In priority comparisons, numerically lower values have higher priority.

Ports are considered for active use in an aggregation in link priority order, starting with the port attached to the highest priority link. Each port is selected for active use if preceding higher priority selections can also be maintained, otherwise the port is selected as standby.

43C.4 Dynamic Key management

Dynamic Key management changes the Key values used for links that either system has selected as a standby to allow use of more links. Whether this is desirable depends on their use. For example, if a single spanning tree is being used throughout the network, separating standby links into a separate aggregation serves little purpose. In contrast, if equal cost load sharing is being provided by routing, making additional bandwidth available in a separate Link Aggregation Group may be preferable to holding links in standby to provide link resilience.

The communicating system with the higher priority (as determined by System Priority and unique System ID) controls dynamic Key changes. Dynamic Key changes may only be made by this controlling system.

NOTE—The controlling system can observe the port priorities assigned by the Partner system, if it wishes to take these into account.

This rule prevents the confusion that could arise if both systems change Keys simultaneously. In principle the controlling system might search all possible Key combinations for the best way to partition the links into groups. In practice the number of times that Keys may have to be changed to yield acceptable results is small.

After each Key change, the controlling system assesses which links are being held in standby by its Partner. Although there is no direct indication of this decision, standby links will be held OUT_OF_SYNC. After matched information is received from the protocol Partner, and before acting on this information, a “settling time” allows for the Partner’s aggregate wait delay, and for the selected links to be aggregated. Twice the Aggregate Wait Time (the expiry period for the wait_while_timer), i.e., 4 seconds, should be ample. If matched Partner information indicates that all the links that the Actor can make active have been brought IN_SYNC, it can proceed to change Keys on other links without further delay.

43C.5 A dynamic Key management algorithm

The following algorithm is simple but effective.

After the “settling time” (see 43C.4) has elapsed, the controlling system scans its ports in the Link Aggregation Group (i.e., all those ports with a specific operational Key value that have the same Partner System Priority, System ID, and Key) in descending priority order.

For each port, it may wish to know

a) Is the port (i.e., the Actor) *capable* of being aggregated with the ports already selected for aggregation with the current Key? Alternatively is the Actor *not capable* of this aggregation?
b) Is the port’s Partner IN_SYNC or is the Partner OUT_OF_SYNC?

And as it inspects each port it may

c) *Select* the port to be part of the aggregation with the current Key.
d) *Retain* the current Key for a further iteration of the algorithm, without selecting the port to be part of the current aggregation.
e) *Change* the operational Key to a new value. Once a new value is chosen, all the ports in the current Link Aggregation Group that have their Keys changed will be changed to this new value.

As the ports are scanned for the first time

1) The highest priority port is always selected.

   If it is capable and IN_SYNC, move to step 2).

   Otherwise, **change** the operational Key of all other ports (if any) in this Link Aggregation Group, and apply this dynamic Key algorithm to those ports, beginning with step 1), after the settling time.

2) Move to the next port.

   IF there is a next port, continue at step 3).

   Otherwise, dynamic Key changes for ports with this operational Key are complete.

   Note that ports that were once in the same aggregation may have had their operational Keys changed to (further) new values. If so, apply the dynamic Key management algorithms to those ports, beginning with step 1), after the settling time.

3) If this port is capable and IN_SYNC:

      select it, and repeat from step 2).

   If this port is OUT_OF_SYNC:

      change the operational Key, and repeat from step 2).

   If this port is not capable but IN_SYNC:

      change the operational Key, move to step 4).

4) Move to the next port.

   If there is a next port, continue at step 5).

   Otherwise If there are still ports in the current Link Aggregation Group (which will have the current operational Key), wait for the settling time and appy the dynamic Key management algorithm, beginning with the first such port, at step 3).

   Otherwise, dynamic Key changes for ports with this operational Key are complete.

5) **If** this port is capable:

      retain the current Key and repeat from step 2).

   Otherwise, **change** the operational Key and repeat from step 2).

This procedure is repeated until no OUT_OF_SYNC links remain, or a limit on the number of steps has been reached.

If the Partner's System ID changes on any link at any time, the Actor's operational Key for that link should revert to the administrative Key value, and the dynamic Key procedure should be rerun. This may involve changing the operational Key values for all the links that were assigned Key values subsequent to the change in Key for the link with the new Partner.

43C.6 Example 1

Two systems, A and B, are connected by four parallel links. Each system can support a maximum of two links in an aggregation. They are connected as shown in Figure 43C–1. System A is the higher priority system.

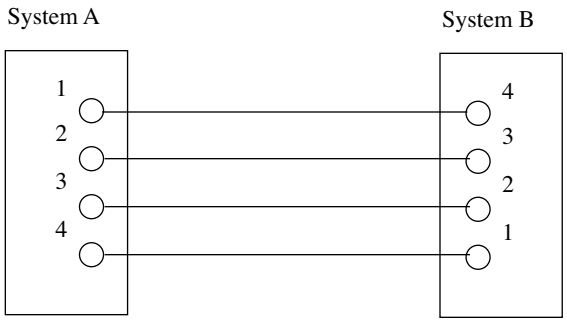


Figure 43C–1—Example 1

The administrative Key for all of System A and System B's ports is 1. Neither system knows before the configuration is chosen that all its ports would attach to links of the same Partner system. Equally, if the links were attached to two different systems, it is not known which pair of links (e.g., 1 and 2, or 1 and 4) would be attached to the same Partner. So choosing the administrative Keys values to be identical for four ports, even though only two could be actively aggregated, is very reasonable.

If there was no rule for selecting standby links, System A and System B might have both selected their own ports 1 and 2 as the active links, and there would be no communication. With the rule, the links A1-B4 and A2-B3 will become active, while A3-B2 and A4-B1 will be standby.

Since System A is the higher-priority system, System B's operational Key values will remain 1 while System A may dynamically change Keys, though it may choose to retain the standby links. Following the Key management algorithm suggested, System A would be able to change the Keys for A3 and A4 in a little over 2 seconds (depending on how fast System B completes the process of attaching its ports to the selected Aggregator) after the connections were first made, and both aggregations could be operating within 5 seconds.

If System A's aggregations were to be constrained to a maximum of three links, rather than two, while System B's are still constrained to two, the suggested algorithm would delay for 4 seconds before changing Keys. Both aggregations could be operating within 7 seconds.

43C.7 Example 2

A system has the odd design constraint that each of its four ports may be aggregated with one other as follows:

a) Port 1 with port 2, or port 4.
b) Port 2 with port 3, or port 1.
c) Port 3 with port 4, or port 2.
d) Port 4 with port 1, or port 3.

This is equivalent to each port being able to aggregate with either neighbor, understanding the ports to be arranged in a circle.

Two such systems are connected with four parallel links as shown in Figure 43C–2.

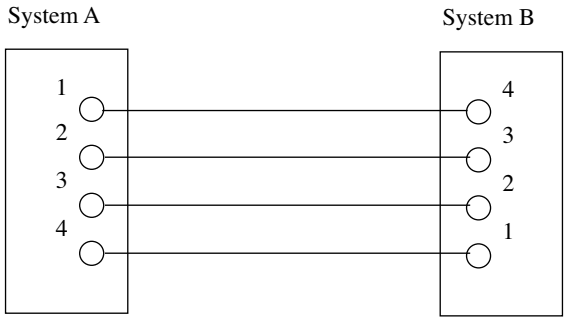


Figure 43C–2—Example 2a

Just as for Example 1, links A1-B4 and A2-B3 become active without changing the operational Key from its original administrative value. The Key for A3 and A4 is changed as soon as they become active, and a few seconds later A3-B2 and A4-B1 become active in a separate aggregation.

If the two systems had been connected as shown in Figure 43C–3:

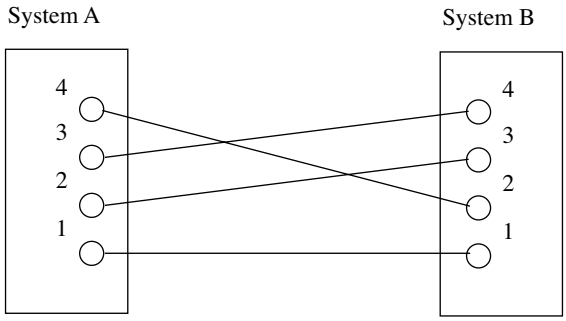


Figure 43C–3—Example 2b

the following would occur, assuming that System A operates the simple algorithm already described.

Initially System A advertises an operational Key equal to the administrative Key value of 1 on all ports. System B first selects B1 as active; since the link connects to A1 it has the highest priority. The next highest priority link is B3-A2, but System B cannot aggregate B3 with B1, so System B makes this port standby. System B can aggregate B4-A3, so the port is made active. Finally if B4 is aggregated with B1, B2 cannot be aggregated, so B2 is made standby.

System A, observing the resulting synchronization status from System B, assigns a Key value of 2 to ports 2 and 3, retaining the initial Key of 1 for ports 1 and 4. System B will remove B4 from the aggregation with B1, and substitute B2. B3 and B4 will be aggregated. In the final configuration A1-B1 and A4-B2 are aggregated, as are A2-B3 and A3-B4.