Ethernet Protocol

Quick Explanation

ELECTGON
www.electgon.com
ma_ext@gmx.net

06.06.2019

Contents

Abstract

Digital Data transmission is standardized and practically used in networks communications according to OSI layers model. This means that data shall be processed according to specifications of that standard in order to comply with this efficient OSI model. A basic protocol that is used almost in all network transmissions is the Ethernet protocol which will discussed in this document along with its main features.

1 Ethernet Data Link Layer

Data Link layer specifications can be implemented using many protocols like Token Ring, Token Bus, Ethernet, TDMA, CDMA, etc. Ethernet was produced as a protocol that can be used to control transmission of packets within Local Area Networks (LAN). It organizes operation of Data Link Layer (layer 2) and also Physical Layer (Layer 1) as well. This protocol is now covered by IEEE 802.3 to describe different families or technologies used to implement that protocol to support additional network media and higher data rate capabilities. Figure1 shows structure of OSI model with more details about Data Link layer. It can be shown that this layer is divided into Logic Link Control (LLC) and Media Access Control (MAC).

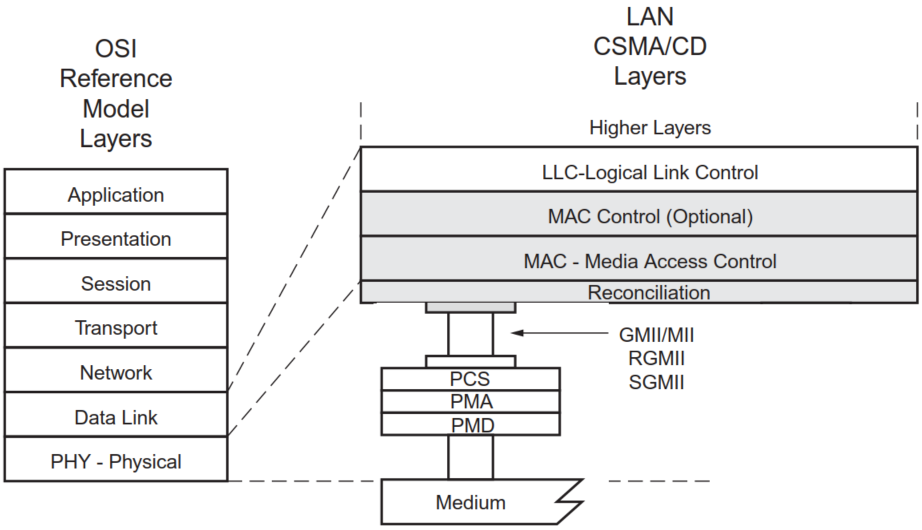

Figure 1: Structure of OSI Model and Layer 2

LLC sublayer is responsible for creating the frames to be sent over the network. In that way it handles communication between upper and lower layers. The MAC sublayer communicates directly with the physical layer. It is responsible for:

- Data encapsulation, including frame assembly before transmission, and frame parsing/error detection during and after reception.
- Media access control, including initiation of frame transmission and recovery from transmission failure.

Ethernet standard is defined to cover MAC sublayer and Layer 1. It fulfills specifications of the MAC sublayer by defining a basic frame format. Figure 2 shows the basic format of the Ethernet frame.

DIX Ethernet Frame

Preamble	Destination Address	Source Address	Type	Data	Cyclical Redundancy Check
8 bytes	6 bytes	6 bytes	2 bytes	Up to 1500 bytes	4 bytes

IEEE 802.3 Ethernet Frame

Preamble	Start Frame Delimiter	Destination Address	Source Address	Length	LLC	Data	Pad	Frame Check Sequence
7 bytes	1 byte	6 bytes	6 bytes	2 bytes		Up to 1500 bytes		4 bytes

Figure 2: Ethernet Frame

Preamble is a 7 byte sequence used to synchronize frame sending/receiving between sender and receiver. These 7 bytes are 0x55, 0x55, 0x55, 0x55, 0x55, 0x55 which are in binary sequence of 0 and 1.

SFD (Start of Frame Delimiter) is one byte used to indicate frame data is starting after this byte. For Ethernet it is 0xD5.

Destination Address is the address of the other terminal to which the frame is sent. The address here is the MAC address of the device.

Source Address is the MAC address of the sender.

Length/Type Ethernet II considered these bytes to represent EtherType while the original IEEE 802.3 framing considered these octets to represent the size of the payload in bytes. Thus, values of 1500 and below for this field indicate that the field is used as the size of the payload of the Ethernet frame while values of 1536 and above indicate that the field is used to represent EtherType. The interpretation of values 1501–1535, inclusive, is undefined.
Below are some values defined for indicating type of the Ethernet.

EtherType	Protocol
0x0800	Internet Protocol version 4 (IPv4)
0x0806	Address Resolution Protocol (ARP)
0x0842	Wake-on-LAN
0x22F3	IETF TRILL Protocol
0x6003	DECnet Phase IV
0x8035	Reverse Address Resolution Protocol
0x809B	AppleTalk (Ethertalk)
0x80F3	AppleTalk Address Resolution Protocol (AARP)
0x8100	VLAN-tagged frame (IEEE 802.1Q) and Shortest Path Bridging IEEE 802.1aq
0x8137	IPX
0x86DD	Internet Protocol Version 6 (IPv6)
0x8808	Ethernet flow control
0x8892	PROFINET Protocol
0x88A2	ATA over Ethernet
0x88A4	EtherCAT Protocol
0x88E5	MAC security (IEEE 802.1AE)
0x88E7	Provider Backbone Bridges (PBB) (IEEE 802.1ah)
0x88F7	Precision Time Protocol (PTP) over Ethernet (IEEE 1588)
0x9100	VLAN-tagged (IEEE 802.1Q) frame with double tagging

Table 1: Some EtherTypes Codes

What is meant by Ethernet type here is the different protocols that are standardized based upon Ethernet technology. These extended standards are using Ethernet frame structure for sending its information and consequently they are using Ethernet physical medium specifications for communication. This extended Ethernet classification is different than Ethernet classification according to its speed which shall be covered in physical layer implementation of the Ethernet.

Data: is the Payload obtained from higher layers to be sent over the Ethernet.

FCS: This 4 bytes part is a checksum appended to end of frame to check accuracy of received data of the frame. For Ethernet, it is using CRC-32. It worth mentioning here that for the CRC part, the most significant bit is sent first.

The other part of Ethernet standard is covering Layer 1 specifications. This layer depends on physical implementation of the medium. Because of that, Ethernet standard has been evolved over many years. First Ethernet implementation started with 3 Mbps in 1973. When it started as industry standard and started with 10 Mbps. Figure 3 shows evolution of Ethernet Data Rate over time.

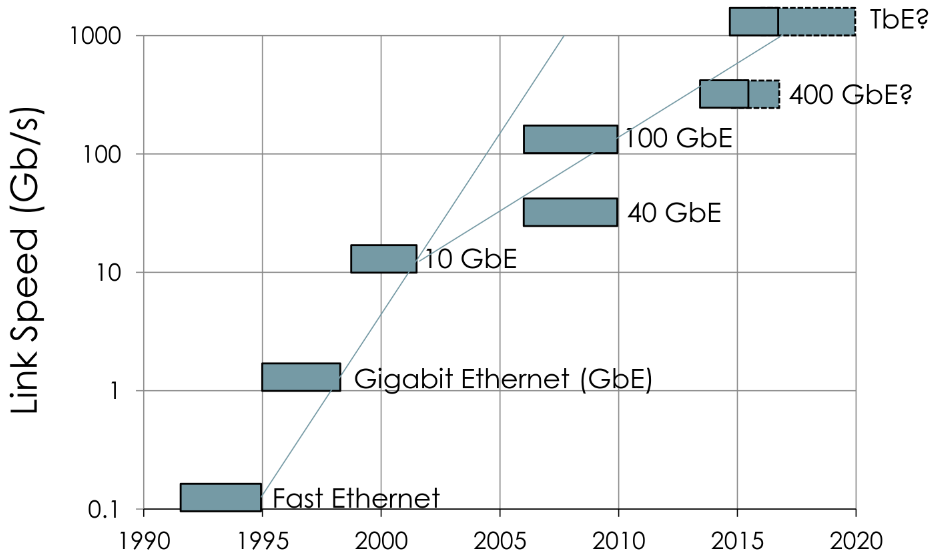

Figure 3: Ethernet Data Rate Evolution [1]

Because of these different implementations of the Ethernet, Layer 1 is organizing different implementation techniques by defining sublayer for the implementation methodology. This will be discussed shortly in the following lines.

2 Reconciliation Sublayer

This sublayer is medium independent and responsible for reconciling the MII with the 802.3u MAC sublayer, so that the latter sees the same physical layer service as 802.3. The reason of having this sublayer is the Ethernet physical interface is implemented with different data rate. Some data rates are using 4 bits to be transmitted, others are using 8 bits as will be shown next.

3 Media Independent Interface

Interfaces between MAC component and Physical component are proposed with different technologies that are listed in table 2

	Supported Data Rate	Total No. of Pins	No. of Data Pins	Clock Frequency
MII	10 M 100 M	16 Pins	TX: 4 pins RX: 4 pins	2.5 MHz for 10 M 25 MHz for 100 M
RMII	10 M 100 M	8 Pins	TX: 2 pins RX: 2 pins	50 MHz for 10/100 M
GMII	10M 100M 1G	24 Pins*	TX: 8 pins RX: 8 pins	2.5 MHz for 10 M 25 MHz for 100 M 125 MHz for 1 G
RGMII	10M 100M/ 1G	12 Pins	TX: 4 pins RX: 4 pins	2.5 MHz for 10 M 25 MHz for 100 M 125 MHz for 1 G
SGMII	10M 100M 1G	8 Pins	TX: 2 pins RX: 2 pins (each is Differential Pair)	625 MHz for all
XGMII	10 G **	74 Pins	TX: 32 pins RX: 32 pins	156.25 MHz

Notes	* it is mentioned in all online resources that total pins are 24 pins, Maybe because transmitter signals have 2 different clock signals for 1 G transmission and another for 10/100 Transmission. I assume then they count for either one of those clocks. ** Not sure if it is limited to 10 G only or it can cover lower data rates also. *** In IEEE standard for Ethernet there are also XLGMII, CGMII.

Table 2: Ethernet Physical Interfaces

3.1 RGMII In Scope

IEEE has defined standard for the gigabit Ethernet in 1998 (IEEE Std 802.3z). This standard introduces GMII as the interface between MAC layer and physical layer. Although this interface provides efficient description for this accepted data rate, however it suffers from increased number of pins connections between MAC and PHY layers. Therefore another specification has been produced by Broadcom, HP and Marvell companies to provide RGMII as an adoption for GMII. This is very useful and efficient solution in cases when MAC and PHY are implemented in different hardware.

Originally, the GMII interface depends on the following timing diagram when sending or receiving Ethernet frames.

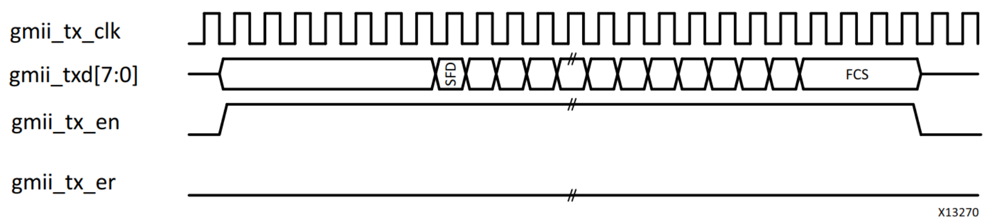

Figure 4: GMII TX Timing Diagram [5]

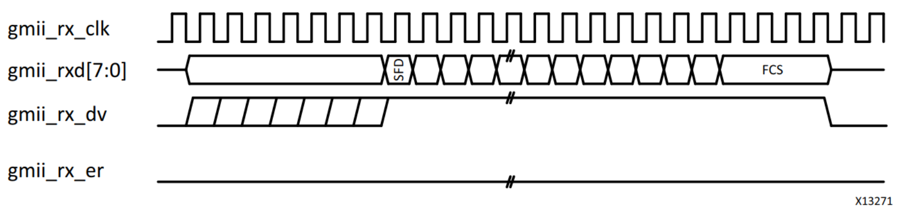

Figure 5: GMII RX Timing Diagram [5]

RGMII interface defined in Specifications V2.0 contains signals definition as shown in table 3

Signal Name	RTBI	RGMII	Description
TXC	**MAC**	**MAC**	The transmit reference clock will be 125Mhz, 25Mhz, or 2.5Mhz +- 50ppm depending on speed.
TD[3:0]	**PCS**	**MAC**	In RTBI mode, contains bits 3:0 on ↑ of TXC and bits 8:5 on ↓ of TXC. In RGMII mode, bits 3:0 on ↑ of TXC, bits7:4 on ↓ of TXC
TX_CTL	**PCS**	**MAC**	In RTBI mode, contains the fifth bit on ↑ of TXC and tenth bit on ↓ of TXC. In RGMII mode, TXEN on ↑ of TXC, and a logical derivative of TXEN and TXERR on ↓ of TXC as described in section 3.4
RXC	**PHY**	**PHY**	The continuous receive reference clock will be 125Mhz, 25Mhz, or 2.5Mhz +- 50ppm. and shall be derived from the received data stream
RD[3:0]	**PHY**	**PHY**	In RTBI mode, contains bits 3:0 on ↑ of RXC and bits 8:5 on ↓ of RXC. In RGMII mode, bits 3:0 on ↑ of RXC, bits7:4 on ↓ of RXC
RX_CTL	**PHY**	**PHY**	In RTBI mode, contains the fifth bit on ↑ of RXC and tenth bit on ↓ of RXC. In RGMII mode, RXDV on ↑ of RXC, and a derivative of RXDV and RXERR on ↓ of RXC as described in section 3.4

Table 3: RGMII Signals Description

Which means that Transmission path (from MAC to PHY) will have 4 bits for data transmission (TXD), 1 bit for clock (TXC) and 1 bit for control (TX_CTL). Receiver path has the same structure. RGMII supports 10/100/1000 Mbps transmission. For the case of 10 or 100 Mbps, transmitted clock on TXC will 2.5 or 25 MHz respectively and the data will be transmitted on the rising edge of the clock. In case of 1 Gbps, transmitted clock is 125 MHz and data is transmitted on both rising and falling edges.

Since data size is reduced (from 8 bits in GMII to 4 bits in RGMII), RGMII specifies transmitting lowest 4 bits first then highest 4 bits.

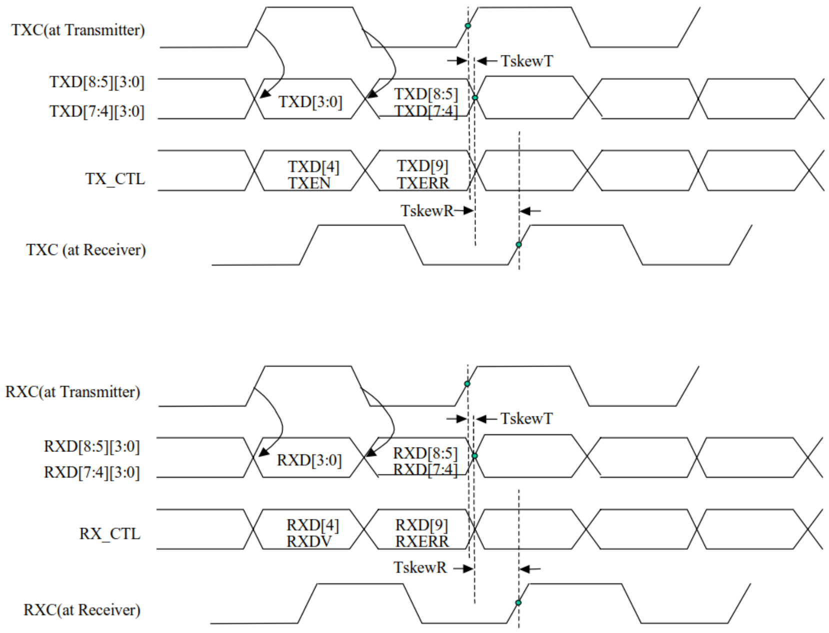

Figure 6: 1G Transmission in RGMII

That is how it is reducing the 8 bits into 4 bits. GMII standard has 2 control signals gmii_tx_en and gmii_tx_er. RGMII reduces these control signals by sending gmii_tx_en on the rising edge. And sending a logical derivative of gmii_tx_en and gmii_tx_er in the falling edges. Accodrding to Specifications 2.0 this logical derivative is Xoring those signals. Figure 7 shows how a valid and errored frame shall be sent.

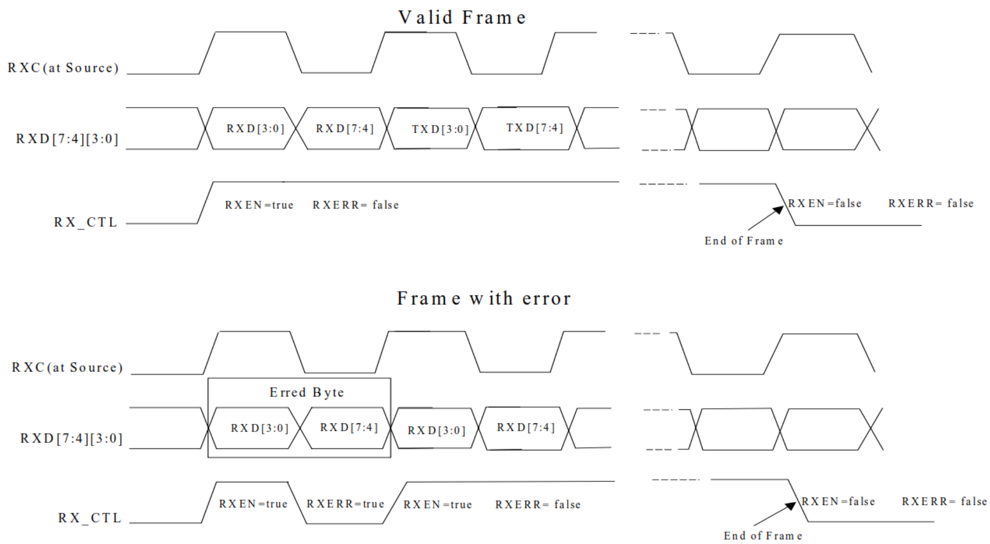

Figure 7: RGMII Frames

RGMII specification describes also timing specs of the data, control and clock signals in terms of setup, hold and skew times between those signals. What is important to point out here is there should be 2 ns setup time and 2 ns hold times between data and transmitted clock (the same applies also for setup /hold times between TX_CTL and TXC). Note that it is meant here that the transmitted data should arrive at the receiver with that setup and hold times specifications.

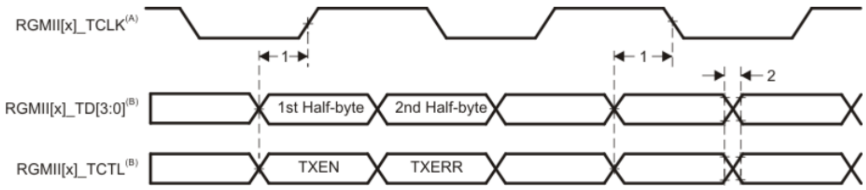

Figure 8: RGMII Setup Time [6]

4 Physical Coding Sublayer (PCS)

The media-dependent physical coding sublayer (PCS) provides encoding/decoding of the bit stream.

Depending on which link speed (Data Rate) is used, the encoding of the serial bits coming from MII sublayer is carried out in PCS sublayer. For 10 Mbps, Manchester coding is used. For 100 Mbps, 4B/5B coding is used. For 1 G, 8B/10B coding is used in case of fiber medium and Trellis modulation with PAM-5 is used in case of copper medium. For 10 G, 64B/66B coding is used.

5 Physical Medium Attachment (PMA)

This sublayer contains the signal transmitters and receivers (transceivers) so it is responsible for serializing/de-serializing encoded bits.

PMA sublayer has a mode of operation; either half duplex or full duplex. This shall be taken into consideration while building the physical layer. Most of Ethernet standard specifying full duplex implementation. For our scope here 1 Gigabit Ethernet was used in full duplex operation. Figure 9 shows connection of this standard.

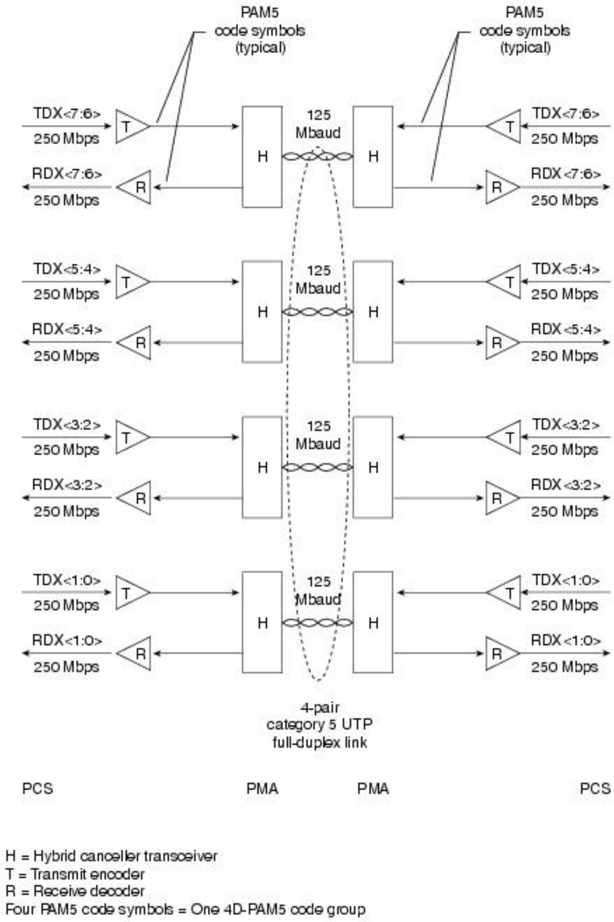

Figure 9: 1G Full Duplex Connection [7]

6 Communication Medium

Physical medium used currently for Ethernet are Copper or Fiber. Together with available data rates this makes a lot of Ethernet technologies. To organize these different technologies, figure 10 shows Ethernet nomenclature used to identify each technology

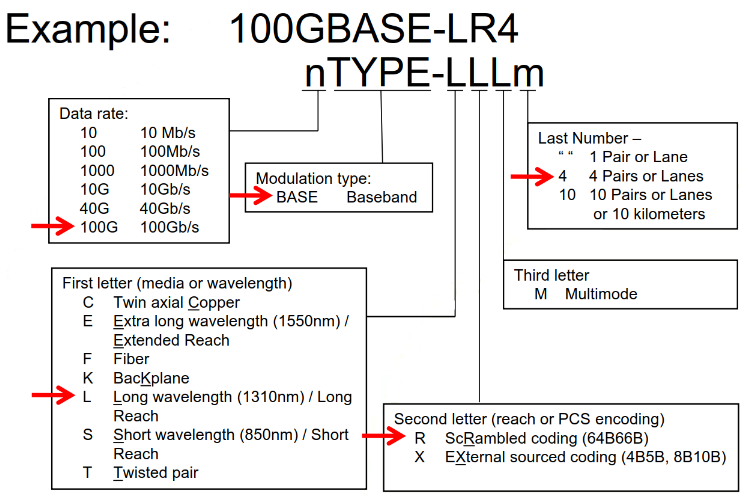

Figure 10: Ethernet Nomenclature

6.1 Twisted Pair Cabling

Twisted Pair Cabling is the most commonly deployed form of Ethernet and includes:

- 10BASE-T – 10Mb/s to 100 meters
- 100BASE-T – 100Mb/s to 100 meters
- 1000BASE-T – 1Gb/s to 100 meters
- 10GBASE-T – 10Gb/s to 100 meters

It has several categories as shown in figure 11.

TIA/EIA Category	ISO/IEC Class	Application	Bandwidth (MHz)	# of Wire Pairs
CAT 1	A	Obsolete. Used for telephones and door bells.	0.1	1 or 2
CAT 2	B	Obsolete. Used in ARCnet and 4 Mb/s Token Ring.	1	2
CAT 3	C	10BASE-T	16	2
CAT 4	N/A	Not used in Ethernet. Copper cabling designed for Token Ring	N/A	4
CAT 5	N/A	Replaced by CAT 5e	N/A	4
CAT 5e	D	Enhanced CAT 5 screened for high bandwidth	100	4
CAT 6	E	1000BASE-T	250	4
CAT 6A	EA	10GBASE-T	500	4
CAT 7	F	10GBASE-T	600	4
CAT 7A	FA	10GBASE-T	1000	4

Figure 11: Twisted Pair Categories

6.2 Fiber Cabling

There are two main types of fiber optic cables

- Single Mode Fiber (SMF)
- Multi-Mode Fiber (MMF)

The difference is basically in the size of the core.

Optical fibers used in Ethernet come in multiple categories. ISO/IEC 11801 specifies Categories OM1, OM2, OM3, OS1 and OS2. OM1, 2, 3 are categories for Multimode fiber while OS1, OS2 are categories for single mode fiber. Figure 12 shows applications of these categories in Ethernet.

Type	Application	Bandwidth Length Product (MHz*km or GHz*m)	Core / Cladding Diameter (um)
OM1	Obsolete. Used for FDDI.	160-200	62.5/125
OM2	Used for 100BASE-FX to 1000BASE-SX.	400-500	50/125
OM3	Used for 10GBASE-SR and higher speeds.	2000	50/125
OM4	Used for 10GBASE-SR and higher speeds.	4700	50/125
OS1	Standard single-mode fiber.	Nearly infinite	9/125
OS2	Reduced loss fiber not typically used in Ethernet	Nearly infinite	9/125

Figure 12: Fiber Cables Categories

6.2.1 Single-Mode Fiber Variants

Single-mode fiber is used for long distance links within large data centers and for links in campus or metro areas such as:

- 100BASE-LX – 100Mb/s to at least 5 kilometers
- 1000BASE-LX – 1Gb/s to at least 5 kilometers
- 10GBASE-LR – 10Gb/s to at least 10 kilometers
- 10GBASE-ER – 10Gb/s to at least 40 kilometers
- 40GBASE-FR – 40Gb/s to at least 2 kilometers
- 40GBASE-LR – 40Gb/s to at least 10 kilometers
- 100GBASE-LR – 100Gb/s to at least 10 kilometers
- 100GBASE-ER – 100Gb/s to at least 40 kilometers

6.2.2 Multimode Fiber Variants

Multimode fiber has enabled longer distances at higher speeds within the data center such as:

- 100BASE-FX – 100Mb/s up to 2 kilometers
- 1000BASE-SX – 1Gb/s up to 550 meters
- 10GBASE-SR – 10Gb/s up to 300 meters
- 40GBASE-SR4 – 40Gb/s up to 100 meters of OM3

- 100GBASE-SR10 – 100Gb/s up to 100 meters of OM3
- 40GBASE-SR4 – 40Gb/s up to 150 meters of OM4
- 100GBASE-SR10 – 100Gb/s up to 150 meters of OM4

6.3 Twinax Copper Cable

Twinax is a shielded copper cable that has twin conductors with good electrical properties that enables these short reach applications at high speed:

- 1000BASE-CX – 1Gb/s up to 25 meters
- 10GBASE-CX4 – 10Gb/s up to 15 meters
- SFP+ Direct Attach Cable – 10Gb/s to 7 meters
- 40GBASE-CR4 – 40Gb/s up to 7 meters
- 100GBASE-CR10 – 100Gb/s up to 7 meters

7 Medium Dependent Interface (MDI)

The medium-dependent interface (MDI) is the cable connector between the signal transceivers and the link.

For twisted pairs cabling RJ45 connector is used.

Pin	Label
1	TP0+
2	TP0-
3	TP1+
4	TP2+
5	TP2-
6	TP1-
7	TP3+
8	TP3-

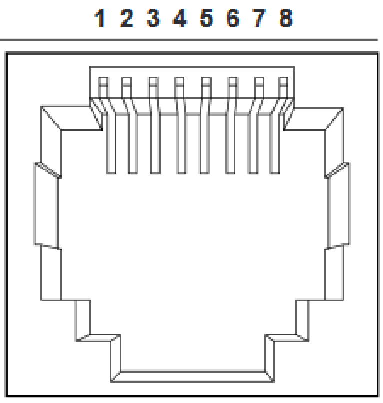

Figure 13: RJ45 Connector

For Fiber cabling there are various connectors that change over time as the technology progresses. The most common connectors are SC, ST, and LC, although many other connectors exist. These connectors vary in shape and size, but all are designed to allow easy interconnection between cables and other devices.

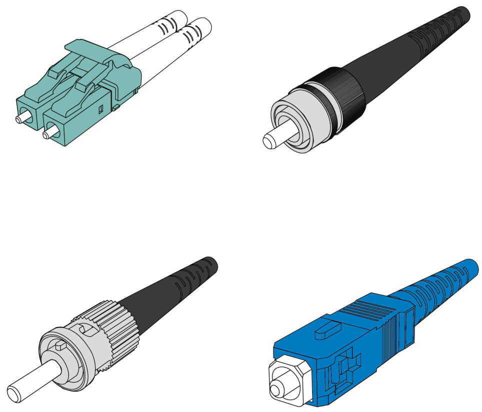

Figure 14: Fibre Cables Connectors [8]

Each connector needs adopter then to be plugged in.

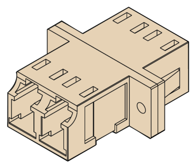

Figure 15: SFP Connector Adopter [8]

These connectors then can be connected to the physical optical port which is called Optical Form Factor

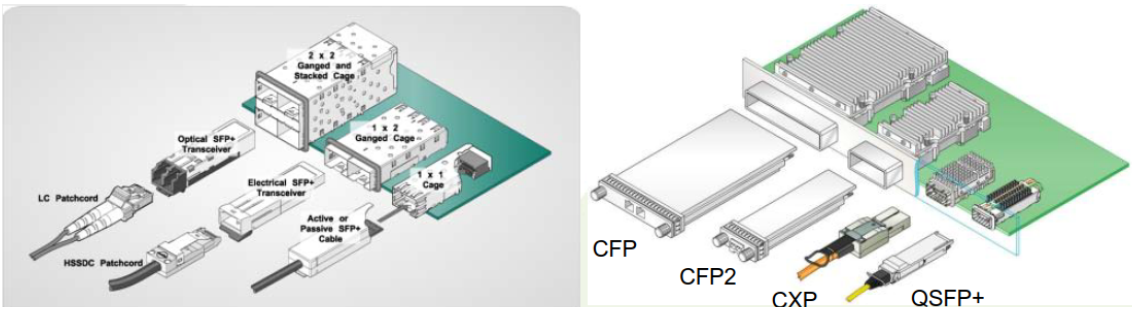

Figure 16: Optical Ports Connectors [1]

8 Autonegotiation

As result of having multiple Ethernet speeds, the connected devices have to know at which speed they are going to talk to each other. This point initially wasn't needed as the Ethernet communication was carried out at 10 Mbps. Now we have many Ethernet speed, the connected device shall setup or prepare the communication. So they have to agree on which speed data will be exchanged also if it is full or half duplex. This process is called then auto-negotiation (also known as NWay algorithm).

As an optional sublayer in 10 Mbps speed, connected devices detect an active link through the transmission and reception of Link Integrity Pulses LIP (also known as Normal Link Pulses NLP). The Autonegotiation sublayer became mandatory for other Ethernet speeds. By producing 100 Mbps speed, the NLP has been replaced by other sequence of pulses called Fast Link Pulses FLP. The devices are exchanging bursts of FLP to initiate the communication. One FLP burst consists of series of 33 pulses and it is 2sms long. Total transmission interval is about 16 ms. Figure 17 depicts these pulses.

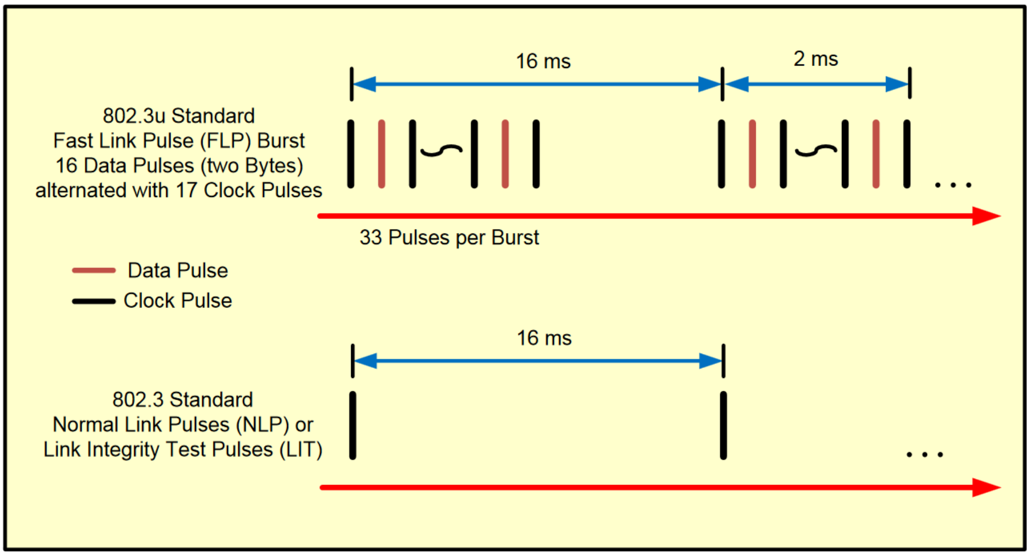

Figure 17: FLP Pulses [9]

The individual pulses alternate between clock pulses and data pulses with the first and all successive odd numbered pulses being clock pulses.

Bibliography

[1] http://www.ethernetalliance.org

[2] https://en.wikipedia.org/wiki/Media-independent_interface

[3] http://appmarkers.tistory.com/entry/MII-RMII-GMII-RGMII

[4] http://humblesblog.blogspot.de/2013/01/rgmiisgmiixaui.html

[5] https://www.xilinx.com/support/documentation/application_notes/xapp209.pdf

[6] http://www.ti.com/product/AM3359/datasheet/peripheral_information_and_timings

[7] http://docwiki.cisco.com/wiki/Ethernet_Technologies

[8] http://www.fibrefab.com/product/lc-connectors

[9] http://supportsystem.livehelpnow.net/resources/14629/ethernet%20autonegotation%20wp[1].pdf