

Chapter 2

Ethernet

2.1 Introduction

Local area networks (LANs) are networks that cover a small area as in a department, in a company, or university. In the early 1980s, the three major local area networks were Ethernet (IEEE standard 802.3), Token Ring (802.5 and used extensively by IBM), and Token Bus (802.4, intended for manufacturing plants). However, over the years, Ethernet [149] has become the most popular wired local area network standard. While maintaining a low cost, it has gone through six versions, most ten times faster than the previous version (10 Mbps, 100 Mbps, 1 Gbps, 10 Gbps, 40 Gbps, 100 Gbps, and in the works 400 Gbps).

Ethernet was invented at the Xerox Palo Alto Research Center (PARC) by Metcalfe and Boggs [92]. It is similar in spirit to the earlier Aloha radio protocol, though the scale is smaller. IEEE's 802.3 committee produced the first Ethernet standard. Xerox never produced Ethernet commercially but other companies did.

In going from one Ethernet version to the next, the IEEE 802.3 committee sought to make each version similar to the previous ones and to use existing technology. In the following, we now discuss the various versions of Ethernet.

2.2 10 Mbps Ethernet

Back in the 1980s, Ethernet was originally wired using coaxial cable. As in Fig. 2.1a, a coaxial cable was snaked through the floor or ceiling and computers attached to it along its length. The coaxial cable acted as a private radio channel that each computer would monitor. If a station had a packet to send, it would send it

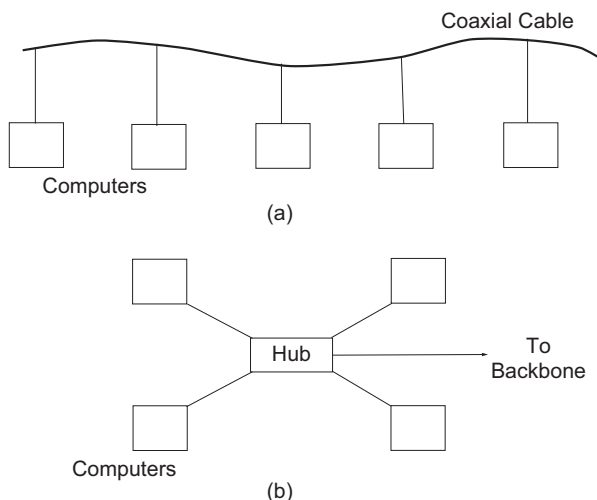


Fig. 2.1 Ethernet wiring using (a) coaxial cable and (b) hub topology

immediately if the channel was idle. If the station sensed the channel to be busy, it would wait until the channel was free. In all of this, only one transmission can be on the channel at one time.

A problem occurs if two or more stations sense the channel to be idle at about the same time and attempt to transmit simultaneously. The packets overlap in the cable and are garbled. This is a collision. The stations involved, using analog electronics, can detect the collision, stop transmitting, and reschedule their transmissions.

Thus, the price one pays for this completely decentralized access protocol is the presence of utilization lowering collisions. The protocol used goes by the name 1-persistent CSMA/CD (Carrier Sense Multiple Access with Collision Detection). The name is pretty much self-explanatory except that 1-persistent refers to the fact that a station with a packet to send attempts this on an idle channel with a probability of 1.0. In a CSMA/CD protocol, if the bit rate is 10 Mbps, the actual useful information transport can be significantly less because of collisions (or occasional idleness).

In the case of a collision, the rescheduling algorithm used is called Binary Exponential Backoff. Under this protocol, two or more stations experiencing a collision randomly reschedule over a time window with a default of $51 \mu\text{s}$ for a 500m network. If a station becomes involved in a second collision, it doubles its time window size and attempts again to randomly reschedule its transmission. Windows may be doubled in size up to ten times. Once a packet is successfully transmitted, the time window size drops back to the default (smallest) value for that packet's station. Thus, this protocol at a station has no long term memory regarding past transmissions.

Table 2.1 Ethernet frame format

Field	Length
Preamble	7 bytes
Frame delimiter	1 byte
Destination address	2 or 6 bytes
Source address	2 or 6 bytes
Data length	2 bytes
Data	up to 1500 bytes
Pad	variable
CRC Checksum	4 bytes

Fig. 2.2 Manchester encoding

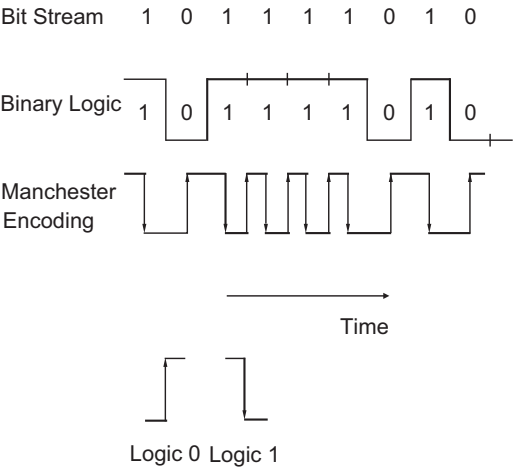


Table 2.1 above shows the fields in the 10 Mbp Ethernet frame. A frame is the name for a packet at the data link layer. The preamble is for communication receiver synchronization purposes. Addresses are either local (2 bytes) or global (6 bytes). Note that Ethernet addresses are different from IP addresses. Different amounts of data can be accommodated up to 1500 bytes. Transmissions longer than 1500 bytes of data must be segmented into multiple packets. The pad field is used to guarantee that the frame is at least 64 bytes in length (minimum frame size) if the frame would be less than 64 bytes in length. Finally the checksum is based on CRC error detecting coding.

A problem with digital receivers is that they require many 0 to 1 and 1 to 0 transitions to properly lock onto a signal. But long runs of 1's or 0's are not uncommon in data. To provide many transitions between logic levels, even if the data has a long run of one logic level, 10 Mbps Ethernet uses Manchester encoding.

Referring to Fig. 2.2, under Manchester encoding, if a logic 0 needs to be sent, a transition is made for 0 to 1 (low to high voltage) and if a logic 1 needs to be sent, the opposite transition is made for 1 to 0 (high to low voltage). The voltage level makes a return to its original level at the end of a bit as necessary. Note that the “signaling rate” is variable. That is, the number of transitions per second is twice

Table 2.2 Original ethernet wiring

Cable	Type	Maximum size
10Base5	Thick coax	500 m
10Base2	Thin coax	200 m
10Base-T	Twisted pair	100 m
10Base-F	Fiber optics	2 km

the data rate for long runs of a logic level and is equal to the data rate if the logic level alternates. For this reason, Manchester encoding is said to have an efficiency of 50%. More modern signaling codes, such as 4B5B, achieve a higher efficiency (see Fast Ethernet below).

During the 1980s, Ethernets were wired with linear coaxial cables. Today hubs are commonly used (Fig. 2.2b). These are boxes (some smaller than a cigar box) that computers tie into, in a star type wiring pattern, with the hub at the center of the star.

A hub may internally have multiple cards, each of which has multiple external Ethernet connections. A high speed (in the gigabits) proprietary bus interconnects the cards. Cards may mimic a CSMA/CD Ethernet with collisions (shared hub) or use buffers at each input (switched hub). In a switched hub, multiple packets may be received simultaneously without collisions, raising throughput at the expense of some delay.

The next table (Table 2.2) illustrates Ethernet wiring. In “10 Base5”, the 10 stands for 10 Mbps and the 5 for the 500 m maximum size. Used in the early 1980s, 10 Base5 used vampire taps which would puncture the cable. Also at the time, 10 Base2 used T junctions and BNC connectors as wiring hardware. Today, 10 Base-T is the most common wiring solution for 10 Mbps Ethernet. Fiber optics, 10 Base-F, was only intended for runs between buildings, but a higher data rate protocol would probably be used today for this purpose.

2.3 Fast Ethernet

As the original 10 Mbps Ethernet became popular and the years passed, traffic on Ethernet networks continued to grow. To maintain performance, network administrators were forced to segment Ethernet networks into smaller networks (each handling a smaller number of computers) connected by a spaghetti-like arrangement of inter-connecting repeaters, bridges, and routers. In 1992, IEEE assigned the 802.3 committee the task of developing a faster local area network protocol.

The committee agreed on a 100 Mbps protocol that would incorporate as much of the existing Ethernet protocol/technology as possible to gain acceptance and so that they could move quickly. The resulting protocol, IEEE 802.3u, was called Fast Ethernet.

Fast Ethernet is only implemented with hubs, in a star topology (Fig. 2.1b). There are three major wiring options (Table 2.3).

Table 2.3 Fast ethernet wiring

Cable	Type	Maximum Size
100Base-T4	Twisted pair	100 m
100Base-TX	Twisted pair	100 m
100Base-FX	Fiber optics	2 km

The original Ethernet has a data rate of 10 Mbps and a maximum signaling rate of 20 MHz (recall that the Manchester encoding used was 50% efficient). Fast Ethernet 100 Base-T4 with its data rate of 100 Mbps has a signaling speed of 25 MHz, not 200 MHz. How is this accomplished?

Fast Ethernet 100 Base-T4 actually uses four twisted pairs per cable. Three twisted pairs carry signals from its hub to a PC. Each of the three twisted pairs uses ternary (not binary) signaling using 3 logic levels. Thus, one of $3 \times 3 \times 3 = 27$ symbols can be sent at once. Only 16 symbols are used though, which is equivalent to sending 4 bits at once. With 25 MHz clocking $25\text{ MHz} \times 4\text{ bits}$ yields a data rate of 100 Mbps. The channel from the PC to hub operates at 33 MHz. For most PC applications, an asymmetrical connection with more capacity from hub to PC for downloads is acceptable. Category 3 or 5 unshielded twisted pair wiring is used for 100 Base-T4.

An alternative to 100 Base-T4 is 100 Base-TX. This uses two twisted pairs, with 100 Mbps in each direction. However, 100 Base-T4 has a signaling rate of only 125 MHz. It accomplishes this using 4B5B (Four Bit Five Bit) encoding rather than Manchester encoding. Under 4B5B, every four bits is mapped into five bits in such a way that there are many transitions for digital receivers to lock onto, irrespective of the actual data stream. Since four bits are mapped into five bits, 4B5B is 80% efficient. Thus, 125 MHz times 0.8 yields 100 Mbps.

Finally, 100 Base-FX uses two strands of the lower performing multi-mode fiber. It has 100 Mbps in both directions and is for runs (say between buildings) of up to 2 km.

It should be noted that Fast Ethernet uses the signaling method for twisted pair (for 100 Base-TX) and fiber (100 Base-FX) borrowed from FDDI. The FDDI protocol was a 100 Mbps token ring protocol used as a backbone in the 1980s.

To maintain channel efficiency (utilization) at 100 Mbps for CSMA/CD implementations, versus the original 10 Mbps, the maximum network size of Fast Ethernet is about ten times smaller than that of the original Ethernet. The trade-off can be seen in the Ethernet design equation [122].

2.4 Gigabit Ethernet

The ever growing amount of network traffic brought on by the growth of applications and more powerful computers motivated a revised, faster version of Ethernet. Approved in 1998, the next version of Ethernet operates at 1000 Mbps or 1 Gbps

and is known as Gigabit Ethernet, or 802.3z. As much as possible, the Ethernet committee sought to utilize existing Ethernet features.

Gigabit Ethernet wiring is either between two computers directly or, as is more common, in a star topology with a hub or switch in the center of the star. In this connection, it is appropriate to say something about the distinction between a hub and switch. A shared medium hub uses the established CSMA/CD protocol so collisions can occur. At most, one attached station can successfully transmit through the hub at a time, as one would expect with CSMA/CD. The half duplex Gigabit Ethernet mode uses shared medium hubs.

A “switch,” on the other hand, does not use CSMA/CD. Rather, the use of buffers means multiple attached stations may send and receive distinct communications to/from the switch at the same time. The use of multiple simultaneous transmissions means that switch throughput is substantially greater than that of a single input line. Level 2 switches are usually implemented in software, level 3 switches implement routing functions in hardware [144]. Full duplex Gigabit Ethernet most often uses switches.

In terms of wiring, Gigabit Ethernet has two fiber optic options (1000 Base-SX and 1000 Base-LX), a copper option (1000 Base-CX) and a twisted pair option (1000-Base T).

The Gigabit Ethernet fiber option deserves some comment. It makes use of 8B10B encoding, which is similar in its operation to Fast Ethernet’s 4B5B. Under 8B10B, eight bits (1 byte) are mapped into 10 bits. The extra redundancy this involves allows each 10 bits not to have an excessive number of bits of the same type in a row or too many bits of one type in each of 10 bits. Thus, there are sufficient transitions from 1 to 0 and 0 to 1 or the data stream even if the data has a long run of 1’s and 0’s.

Gigabit Ethernet using twisted pair uses five logic levels on each wire. Four of the logic levels convey data and the fifth is for control signaling. With four data logic levels, two bits are communicated at once or eight bits over all four wires at a time. Thus the signaling rate is 1 Gbps/8 or 125 MHz.

In terms of utilization under CSMA/CD operation, if the maximum segment size had been reduced by a factor of 10 as was done in going from the original Ethernet to Fast Ethernet, only very small gigabit networks could have been supported. To compensate for the ten times increase in data rate relative to Fast Ethernet, the minimum frame size for Gigabit Ethernet was increased (by a factor of eight) to 512 bytes from Fast Ethernet’s 512 bits (see [122] for a discussion of the Ethernet equation that governs this).

Another technique that helps Gigabit Ethernet’s efficiency is frame bursting. Under frame bursting, a series of frames are sent in a single burst.

Gigabit Ethernet’s range is at least 500 m for most of the fiber options and about 200 m for twisted pair [144, 149].

2.5 10 Gigabit Ethernet

Considering the improvement in Ethernet data rate over the years, it is not too surprising that a 10 Gbps Ethernet was developed [143, 154]. Continuing the increases in data rate by a factor of ten that have characterized the Ethernet standards, 10 Gbps (or 10,000Mbps) Ethernet is ten times faster than Gigabit Ethernet. Applications include backbones, campus size networks, and metropolitan and wide area networks. This latter application is aided by the fact that the 10 Gbps data rate is comparable with a basic SONET fiber optic transmission standard rate. See a later chapter for more information on SONET.

There are eight implementations of 10 Gbps Ethernet. It can use four transceiver types (one four wavelength parallel system and three serial systems with a number of multi-mode and single mode fiber options). Like earlier versions of Ethernet, it uses CRC error coding. It operates in full duplex non-CSMA/CD mode. It can go more than 40 km via single mode fiber.

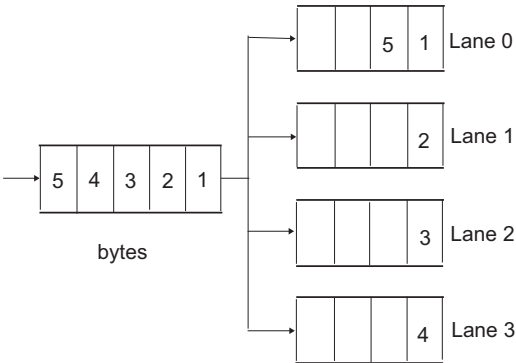
To lower the speed at which the MAC (Media Access Control) layer processes the data stream, the MAC operates in parallel on four 2.5 Gbps streams (lanes). As illustrated in Fig. 2.3, bytes in an arriving 10 Gbps serial transmission are placed in parallel in the four lanes.

There is a 12 byte inter-packet gap (IPG) which is the minimum gap between packets. Normally, it would not be easy to predict the ending byte lane of the previous packet, so it would be difficult to determine the starting lane of the next transmission. The solution is to have a starting byte in a packet always occupy lane 0. The IPG is found using a pad (add in extra 1 to 3 bytes), a shrink (subtract 1 to 3 bytes), or through combination averaging (average of 12 bytes achieved through a combination of pads and shrinks). Note that padding introduces extra overhead in some implementations.

In terms of the protocol stack, this can be visualized as in Fig. 2.4. The PCS, PMA, and PMD sublayers use parallel lanes for processing. In terms of the sublayers, they are:

Reconciliation: Command translator that maps terminology and commands in MAC into electrical format appropriate for physical layer.

Fig. 2.3 Four parallel lanes for 10 Gigabit Ethernet



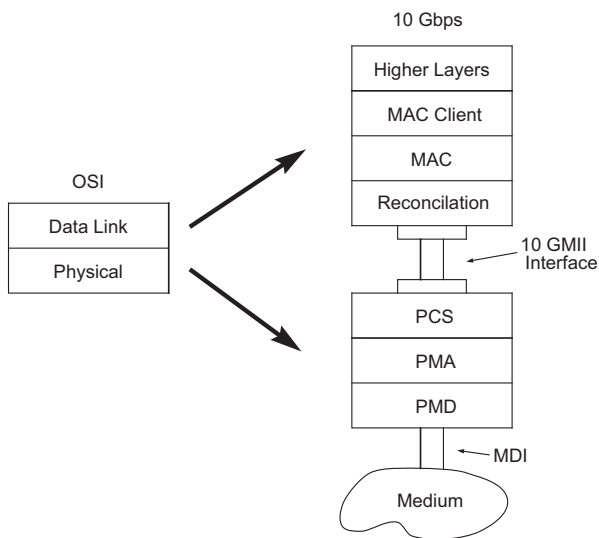


Fig. 2.4 Protocol stack for 10 Gbps Ethernet

PCS: Physical Coding Sublayer.

PMA: Physical Medium Attachment (at transmitter serialize code groups into bit stream, at receiver synchronization for data decoding).

PMD: Physical Medium Dependent (includes amplification, modulation, wave shaping).

MDI: Medium Dependent Interface (i.e., connector).

2.6 40/100 Gigabit Ethernet

Over the years Ethernet has been attractive to users because of its relatively low cost, robustness, and its ability to provide an interoperable network service. Users have also liked the wide vendor availability of Ethernet related products. However, even with the release of gigabit and 10 Gbps Ethernet demand for bandwidth continued to grow. Network equipment shipments can grow at a 17% a year rate. Internet traffic grows at 75–125% a year. Computer performance doubles every 24 months. A 2008 projection was that within 4 years 40 Gbps would be needed.

One of the applications driving this growth is the increasing use of data centers (see the latter chapter on this topic). These facilities house server farms for hosting web services and cloud computing services. Projections indicate a need for 100 Gbps of data transfer capacity from switch to switch. Also 100 Gbps will have applications between buildings, within campuses, and for metropolitan area networks (MAN) and wide area networks (WAN).

In July 2006 a committee was convened to explore increasing the data rate of Ethernet beyond 10 Gbps. In 2010 standards for 40 Gbps and 100 Gbps Ethernet were approved. This discussion is based on [35, 101].

2.6.1 40/100 Gigabit Technology

In implementing 40 and 100 gigabit Ethernet some of the objectives are:

- MAC (medium access control) data rates of 40 and 100 gigabit per second.
- Full duplex is only supported (i.e., two way communication).
- Maintain the existing minimum and maximum frame length.
- Use the current frame format and MAC layer.
- Optical transport network (OTN) support.

A variety of transmission media can carry 40 and 100 gigabit Ethernet as Table 2.4 illustrates.

Figure 2.5 illustrates the protocol stack for 40 and 100 gigabit Ethernet. In the figure one has the physical coding sublayer (PCS), the forward error correction

Table 2.4 40/100 Gbps Ethernet

40 Gbps	100 Gbps
≥ 10 km single mode fiber	≥ 40 km single mode fiber
≥ 100 m multi-mode fiber	≥ 10 km single mode fiber
≥ 10 m copper cable	≥ 100 m multi-mode fiber
≥ 1 m backplane	≥ 10 m copper cable

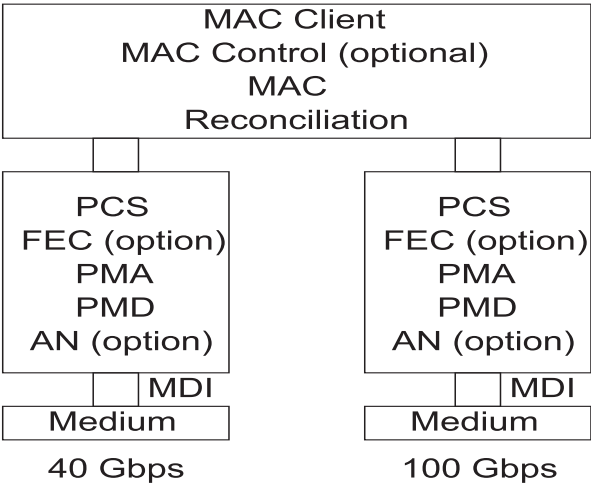


Fig. 2.5 40 and 100 Gbps Ethernet protocol functions

sublayer (FEC), physical medium attachment sublayer (PMA), physical medium dependent sublayer (PMD), and the auto-negotiation sublayer (AN). Here also MDI is the medium dependent interface or the connector.

In the physical coding sublayer 64B/66B coding is used, mapping 64 bits into 66 bits to provide enough transitions between 0 and 1 for digital receivers. As in 10 gigabit Ethernet, the concept of parallel lanes is used in 40 and 100 gigabit Ethernet. A 66 bit block is distributed in round robin fashion on the PCS lanes. Specifically for 40 gigabit Ethernet there are 4 PCS lanes that support 1, 2, or 4 channels or wavelengths. For 100 gigabit Ethernet there are 20 PCS lanes that support 1, 2, 4, 5, 10, or 20 channels or wavelengths. As an example, a 100 gigabit Ethernet may use 5 parallel wavelengths over a fiber, each carrying 20 Gbps.

Created PCS lanes can be multiplexed into any interface width that is supported. There is a unique lane marker for each PCS lane which is inserted periodically. Bandwidth for the lane marker comes from periodically deleting the inter-packet gap (IPG) in a lane. All bits in the same lane follow the same physical path no matter how multiplexing is done.

The receiver reassembles the PCS lanes by demultiplexing bits and also realigns the PCS lanes taking into account the skewness of the lanes. Advantages of this include the fact that all encoding, deskew, and scrambling functions are implemented on a CMOS device located on the host and there is minimal bit processing except for using an optical module for multiplexing.

Finally, clocking takes place at 1/64th of the data rate (625 MHz for 40 gigabits and 1.5625 GHz for 100 gigabits). More information on 40 and 100 gigabit Ethernet can be found on www.ethernetalliance.org.

2.7 Higher Ethernet Speeds

2.7.1 Introduction

In 2012 the IEEE 802.3 Ethernet committee did a data rate growth assessment and came to the following conclusions [36]:

- The need for increased data rates was increasing more rapidly for “network aggregation nodes,” than for end-station applications. Network aggregation nodes combine traffic from very many end users. Perhaps the most important example of this is data centers.
- The compound annual growth rate (CAGR) that should be supported is 58%. The biggest growth rates were in data intensive science and in the financial industry.
- In 2012 the 802.3 committee projected a need in data rate capacities of 1 Tbps (i.e., 10^{12} bps) by 2015 and 10 Tbps by 2020.

As John D'Ambrosia and Paul Mooney put it in a 2013 white paper [36]:

“Bandwidth growth is unrelenting everywhere across Ethernet networking. Every day, *more* users are *more* quickly accessing the Internet in *more* ways, to utilize *more* applications and consume *more* content that demands *more* bandwidth every day. More, more, more...”

In fact there is a cyclic action where enabling higher data rates enables applications which creates a need for increased data rates leading to enabling higher data rates and on and on... [36].

2.7.2 The Road to Higher Speeds

The initial thought of the Ethernet community was that it was time to create a 400 Gbps version of Ethernet. The IEEE P802.3bs 400 Gigabit Ethernet (400 GbE) committee (known as the 400 Gbps Ethernet study group) first met in May 2014. One might ask why not go for a ten fold increase of speed from 100 Gbps Ethernet to 1 Tbps Ethernet? The feeling was that a 400 Gbps implementation was doable with existing technology whereas a 1 Tbps implementation would involve new technology and a need for more research and development.

As of 2016 there were a number of parallel efforts going on under the aegis of IEEE standards bodies. These include a 200 Gbps version of Ethernet as well as a 400 Gbps version. Moreover higher data rates are built out of parallel lanes or channels. In this light, to be started is work on a 50 Gbps version of Ethernet which can be used as a building block for higher rates.

Both multi-mode and single mode fiber versions of higher data rate Ethernets are being considered.

In terms of goals for higher rates, one has [37] (<http://www.wikipedia.org>):

- Full duplex operation only (as has been true since 10 Gbps Ethernet).
- MAC data rates of 200 and 400 Gbps.
- Use existing Ethernet frame format using the Ethernet MAC.
- Keep the minimum and maximum frame size of the existing standard.
- The bit error rate should be 10^{-13} or the frame loss equivalent. Note that for 10, 40 and 100 Gbps Ethernet versions the bit error rate was 10^{-12} .
- OTN (optical transport network for Ethernet) should be supported. Energy-Efficient Ethernet (EEE) is optionally supported.
- Optional 400 Gbps attachment unit interfaces for chip-to-chip and chip-to-module applications.
- The physical layer supports link distances in Table 2.5.

While the implementation details need to be filled in, the trend for Ethernet is a march to higher and higher data rates.

Table 2.5 200/400 Gbps Ethernet

200 Gbps	400 Gbps
≥ 500 m single mode fiber	≥ 100 m multi-mode fiber
≥ 2 km single mode fiber	≥ 500 m single mode fiber
≥ 10 km single mode fiber	≥ 2 km single mode fiber
	≥ 10 km single mode fiber

2.8 Conclusion

For 35 years Ethernet has continually transformed itself by way of higher data rates to meet increasing demand for networking services. It will be interesting to see what the future holds.



<http://www.springer.com/978-3-319-53102-1>

Introduction to Computer Networking

Robertazzi, Th.G.

2017, XIII, 154 p. 42 illus., 8 illus. in color., Hardcover

ISBN: 978-3-319-53102-1