1 A brief history of "Ethernet" (from a car manufacturer's perspective)

1.1 From the beginning

In 1969 employees at AT&T/Bell Labs developed the first version of Unix. The original intention was to aid the company's internal development of software on and for multiple platforms, but over time Unix evolved to be a very widespread and powerful operating system, which facilitated distributed computing. An important reason for the successful proliferation of Unix was that for antitrust reasons AT&T was neither allowed to sell Unix nor to keep the intellectual property to itself [1]. In consequence Unix – in source code – was shared with everybody interested.

It was especially, but not only, embraced by universities and the community that evolved provided the basis for the computing environment we are used to today and in which also Ethernet has its place. At a time when computing was dominated by large, proprietary, and very expensive mainframe computers few people could use, Unix created a demand for Local Area Networking (LAN) while at the same time providing an affordable, common platform for developing it [2]. As one example, a group at the University of California, Berkeley created a Unix derivative. The Berkeley Software Distribution (BSD) was first released in 1978 and its evolutions became as established as the "BSD-style license" attached to it [3]. Another example is the TCP protocol. The first version of this, published in 1974, was implemented for Unix by the University of Stanford by 1979 [4]. Later in 1989, the then up-to-date TCP/IP code for Unix from AT&T was placed in the public domain and thus significantly helped to distribute the TCP/IP Internet Protocol Suite [5].

The advent of Unix represents an important milestone in the early days of computing. It coincides with the point in time in which a significant number of public as well as proprietary research projects were initiated to investigate methods to interchange data locally and at higher speeds than could be provided for by the telephone system [6]. One of the most momentous projects was the one at Xerox PARC. Xerox needed a solution for data transmission between its first personal computer workstations (called "Xerox Alto"), its laser printers, and the early Internet. Thus, Ethernet was invented (1973), patented (1975), and published (1976) [7].

The general opinion (see e.g. [8]) is that the foundation of Ethernet's later success was laid almost as early in time as this, because of the following two choices:

1. **Opening the technology to others.** At the time, it was common for computer companies to try to bind customers to their products by using proprietary technologies or at least restricting competition with the licensing policy of their patents. Xerox held the patents on Ethernet, but there seems to have been an early understanding that they would profit more from the network effects of a widely deployed Ethernet than from selling the technology itself.[1] Seven years after the invention, on 30 September 1980, Xerox published the "DIX Standard" on Ethernet [9] jointly with the Digital Equipment Corporation (DEC) and Intel. They also offered the technology for adoption to the IEEE 802 group, very shortly after the IEEE 802 group had been founded.[2] With several competing technologies being proposed and followed up, it was by no means evident that Ethernet would prevail. But it did, and one of the reasons attributed to this is that Xerox followed a relaxed licensing policy while not trying to dominate the standardization effort [6]. In the authors' view this is an attitude as little self-evident then as it is now.
2. **Limiting the technical solution to the task at hand.** Ethernet addressed, and still does address, the communication mechanisms needed on the lower 1½ layers of the ISO/OSI layering model only (see also Figure 1.4 in Section 1.2.1), at a time when the ISO/OSI layering model had yet to be completed. It provided a container that gets a packet through a network with multiple participants, but is as independent from the application layer as possible [10]. Even today there is still a tendency to define all layers of a communication system. What allegedly provides the advantages of complete control over the whole communication stack generally makes the system less flexible and less adaptable to future, and hence unknown, requirements. Indeed, Ethernet's adaptability has proven itself to the extent that it is now being introduced in a completely different physical and application environment: in automotive.

In the years that followed, the IEEE became the host for the development of Ethernet. In 1983, IEEE 802.3 published the first of many Ethernet Standards, 10BASE-5 for 10 Mbps over thick coax cable [11]. In the same year already at least 21 companies were mentioned in the trade press to be developing and/or manufacturing Ethernet products [6]. When on 1 January 1984 the AT&T monopoly ended, the existing installed telephone wiring became useable for competing services and applications [12] and a whole new range of possibilities opened to the networking world. Thus in 1987 SynOptics, a Xerox spinout, was the first company to prove the feasibility of transmitting Ethernet at 10 Mbps over telephone wires with a proprietary Ethernet product [6]. The IEEE ratified the respective 10BASE-T standard in September 1990. This did require some effort, owing to the many proprietary versions that had evolved, but when successful, it sealed the victory over other networking technologies in the market [13]. The first optical version of Ethernet was published in 1993 as 10BASE-F.

Meanwhile, the world around Ethernet did not stand still, but continued to provide means for and create demands for networking. Various evolutions of TCP and IP were developed and in October 1989 the IETF published the complete set of protocols in the TCP/IP Internet protocol suite [14, 15]. As mentioned, the success of TCP/IP was fueled by AT&T's public domain implementation of TCP/IP on Unix [5]. In 1991, the

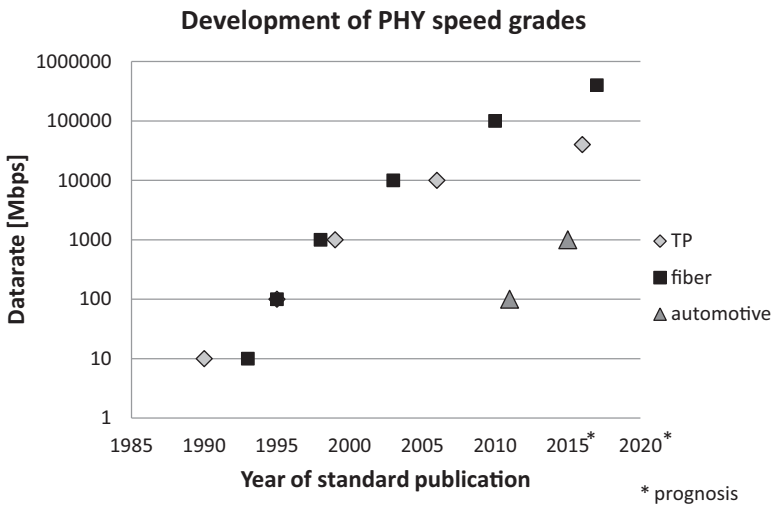

Figure 1.1 Timeline of major PHY speed grades.

TIA published a standard for inexpensive UTP wiring: TIA/EIA-568. Even today, it is impossible to imagine an Ethernet network without the 8P8C/RJ-45 connector described in that standard. The World Wide Web was launched in 1994 [13] and the IETF released a specification for IPv4 routers in June 1995 [16]; the Windows 95 Service Pack-1 released on 14 February 1996 automatically included the Microsoft Internet Explorer 2.0 (i.e. built-in TCP/IP networking), bringing the Internet to the masses [17]. Internet Explorer had been available before, but needed to be purchased extra.

Subsequently, the IEEE amended and enhanced Ethernet, proving Ethernet's adaptability. First, IEEE 802.3 added, and continues to add, new speed grades. Figure 1.1 gives an overview of these for copper and fiber channels. Under discussion today are 40 Gbps for transmission over twisted pair cables and 400 Gbps for optical communication. In 1997, IEEE 802.3 enabled full duplex communication and flow control to replace the shared media approach prevailing until then. New functionalities that have been added are auto-negotiation in 1995, Power-over-Ethernet (PoE) in 2003, and Energy Efficient Ethernet (EEE) in 2010. New use cases that have been considered by the IEEE include Ethernet in the First Mile (EFM, 2004, see also Section 1.2.4), Ethernet over copper backplane (2007), and finally in 2013, a good 15 years after the respective activity had been started for data centers, IEEE 802.3 set up a task force to develop a Reduced Twisted Pair Gigabit Ethernet (RTPGE) suitable for automotive. Figure 1.2 gives an overview of Physical Layer (PHY) variants developed or under development.

In addition to the directly PHY related activities, the IEEE has worked, and is still working on, Quality of Service (QoS) schemes for Ethernet and other management functions. In Ethernet basically the only quality control provided is a CRC check at the receiver, which has no other consequences than offering the possibility to discard the packets with detected errors. A pure IEEE 802.3 measure was taken in 1998, when IEEE 802.3 agreed on a packet extension, in order to incorporate an IEEE 802.1Q header

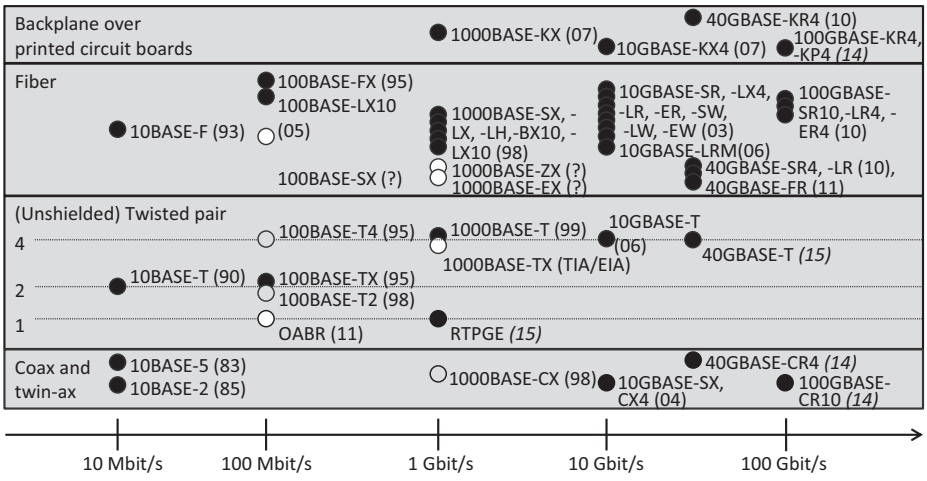

Figure 1.2 Overview of Ethernet PHY variants: black = IEEE versions; grey = versions without market relevance; white = non-IEEE versions; (expected) year of release in brackets when known [18].

consisting of 802.1 Virtual LAN (VLAN) and priority information. Another important concept was established in 2011, when the IEEE (mainly in 802.1) finalized the first set of standards summarized under Audio Video Bridging (AVB). AVB aims at improving the quality of audio and video transmissions over an Ethernet network (for more details see Section 5.1). At the time of writing in 2013 further enhancements on the AVB/QoS functionalities were being standardized under the name of Time Sensitive Networking (TSN).

1.2 The meaning of "Ethernet"

The term "Ethernet" was first used in 1973, the name referring to the belief of nineteenth century physicists that there is a passive medium between Sun and Earth which allows electromagnetic waves to propagate everywhere, which they called the "lumeniferous Ethernet." The coax used for the inventors' communication system was equally passive and they also intended their data packets to go everywhere [13]. Nevertheless, the IEEE never officially adopted the name (although, unofficially, it did). As an open standards body the IEEE did not want to give the impression of favoring any company in particular. Despite the fact that Xerox had relinquished their trademark on the name, IEEE 802.3 was called "Carrier Sense Multiple Access with Collision Detection (CSMA/CD)" instead [10].

As a consequence, in the various application fields and industries, the name "Ethernet" is used with different meanings, some of which have very little in common with what is specified in IEEE 802.3. The following sections will therefore address how Ethernet is used in the IEEE, in some other industries, and in the "Automotive Ethernet" discussed in this book.

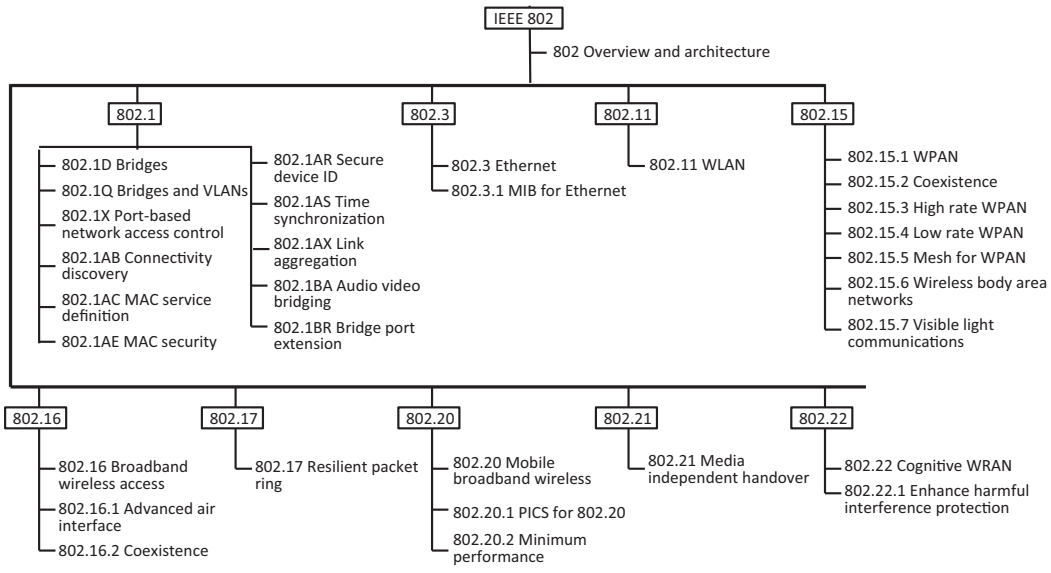

Figure 1.3 The ongoing IEEE 802 standardization activities in 2012.
Source: IEEE 802.3-2012, IEEE Standard for Ethernet, reprinted with permission of IEEE, 3 Park Avenue, New York, NY 10016-5997, USA, Copyright 2012, by IEEE.

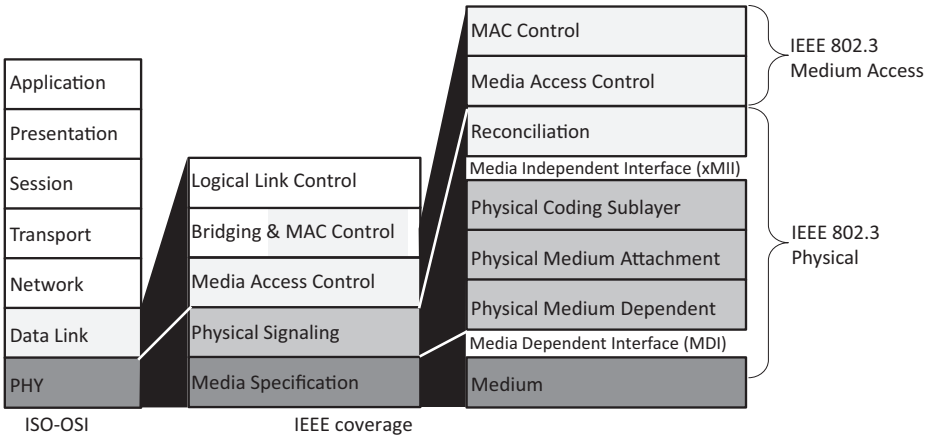

Figure 1.4 Ethernet in the IEEE (e.g. [19]).

1.2.1 Ethernet in the IEEE

Ethernet is standardized in IEEE 802.3 (see also Figure 1.3). This comprises the complete Physical Layer (PHY) and those parts of the Data Link Layer (DLL) that are technology specific, like the packet format and the medium access method chosen (see also Figure 1.4). Various other aspects also in the IEEE standards, e.g. in IEEE 802.1, affect the implementation of an Ethernet-based communication system. While being relevant, these standards are applicable to all technologies addressed in 802 and therefore not "IEEE Ethernet" specific. This is the same for the Logical Link Control (LLC), whose

Preamble	SFD	Destination MAC address	Source MAC address	Optional 802.1Q Header	Length or Ethertype	Payload/LLC	CRC/ FCS	Inter Frame Gap
Bytes: 7	1	6	6	4	2	42/46-1500	4	min.12

Figure 1.5 Elements of an Ethernet frame/packet.

standardization has been concluded in IEEE 802.2 and whose task is to harmonize various methods of medium access towards the network layer [10, 19].

One of the main inventions of the original Ethernet was sharing the media with the help of a Carrier Sense Multiple Access with Collision Detection (CSMA/CD) mechanism. CSMA/CD was based on the ALOHA method, which had been developed at the University of Hawaii a few years earlier as a multi-user access method and which more or less simply proposed retransmissions in case collisions were detected [13]. In the case of CSMA/CD this was enhanced by additionally establishing whether the channel is occupied before the start of a transmission. If the channel is sensed available, the transmitter is allowed to send its packet. Nevertheless, even in this case collisions can occur, such as when another unit had also sensed the channel available and started transmitting simultaneously. Both transmitters would detect this and, in consequence, go into a random back-off period that would increase its potential length with the number of collisions having occurred for one packet [10].

Today, it is hard to find any Ethernet installation that still uses the CSMA/CD method.[3] The vast majority of Ethernet networks are installed as switched networks with Point-to-Point (P2P) links. In these networks, only the PHYs of two units are connected directly and switches in the receiving unit forward the packets according to their addressing between the other unit's internal PHYs.[4] The so-called "full duplex"[5] operation provides significant advantages in terms of timing and supported link segment lengths [10], so that today the CSMA/CD mode has become obsolete. Also in "full duplex" the MAC is responsible for receiving and transmitting packets. With full duplex, a new sublayer was added: the MAC Control (Medium Access Control Control!). The general purpose of the MAC Control layer was to allow for the interception of Ethernet packets in the case of specific requirements. In the case of full duplex it enables flow control. In order to allow for limited resources in terms of the buffering and switching bandwidth, the MAC Control provides the mechanisms to decide when packets are being sent [19].

The most pronounced and stable element of Ethernet is the Ethernet frame/Ethernet packet (see Figure 1.5). The packet starts with a preamble and the Start Frame Delimiter (SFD), which together help synchronize incoming data in the case of CSMA/CD operation. With CSMA/CD no longer deployed, they have become obsolete but are kept for backward compatibility reasons. Starting with 100BASE-TX more complex signal encoding has been used, which allows the deployment of special symbols to detect the beginning and end of a packet.

Each Ethernet interface[6] is assigned a unique serial number consisting of 48 bits, often referred to as the "MAC address" or the "hardware address." Following the preamble, every packet contains information on where the packet is to be sent and the device that sent it, using the respective MAC addresses. End node MACs initially only read up

to the destination address to evaluate whether a packet is intended for this end node (as direct-, multi-, or broadcast). If the address matches, the packet is read completely; if the address does not match the packet is ignored. Switch nodes evaluate both the destination address, deciding which port to send the packet to, and the source address, remembering for future incoming packets on which port to find the addressee with that address. This means that there normally is a learning period after start-up in a switched Ethernet network.

The next four bytes represent an optional IEEE 802.1Q header. The first two bytes identify that this indeed is an 802.1Q header. The remaining two provide the Tag Control Information (TCI) and are divided into three bits for the priority information according to the 802.1p standard, one bit representing the Drop Eligible Identifier (DEI), and 12 bits for the Virtual LAN identifier, which specifies to which Virtual LAN (VLAN) the packet belongs [20]. VLANs represent an important concept for partitioning a physical LAN into various logical domains on layer 2 (see also Section 5.2).

The next field indicates either the length of the packet or the (Ether-)type. The Ethertype states what type of data to expect in the payload in respect to the higher layers. It covers content like IP (v4 or v6) or certain AVB packets, but also various proprietary types that have accumulated over time. Ethernet had been designed as a container for whatever data that needs to be transmitted; for example, several of the Industrial Ethernet variants – e.g. Profinet, EtherCat, Sercos, Powerlink, High-Speed Ethernet (HSE) – have their own Ethertype (see also Section 1.2.2). The IEEE 802.1Q identifier mentioned above has the Ethertype 0x8100. A list of Ethertypes is maintained by the IEEE [21]. When the field represents the length, the content is a number equal to or less than 1500 (see next paragraph). In this case, the IEEE 802.3 LLC protocol can be used to identify the type of data that is being transmitted.

The payload has a minimum size of 42 bytes when the 802.1Q header is present and 46 bytes when it is not.[7] Should the data needing to be sent be shorter than the minimum, then the remaining bytes are filled with padding. The maximum payload length is 1500 bytes. Note that the payload represents user data only from a layer 2 perspective. Various headers from other layers, like the IP or UDP headers, will further reduce the bytes available for the actual application.

Finally, the packet is terminated with a Cyclic Redundancy Check (CRC) called the Frame Check Sequence (FCS). The FCS checks the integrity of the various bits of the packet (other than preamble and SFD). Following the packet there must be an interframe gap of a minimum of 12 bytes. With a fully loaded payload this means that the header/payload efficiency is larger than 97%.

Table 1.1 provides an overview of the main components of Ethernet and how they have changed over time. As has been visualized in Figure 1.2, Ethernet has been developed for various media and almost all but the original one are being addressed today. As a consequence of higher data rates and advancements in signal processing, the physical signaling has changed with the media and has also been standardized in various forms. The original media access mechanism vanished. Nevertheless, the principle that Ethernet performs no quality control in form of acknowledgements or retransmits, as well as its "container"-function, has been kept. If needed, retransmits have to be initiated on higher

Table 1.1 Comparison of the four main Ethernet components as defined in 10BASE-5 and the "IEEE Ethernet" today

	Ethernet in 10BASE-5	IEEE Ethernet today
Frame	26 bytes overhead, 46–1500 bytes payload	Optional 4 bytes for 802.1Q header added
MAC	CSMA/CD, best-effort traffic without acknowledgements	Full duplex and flow control, best-effort traffic without acknowledgements
Signaling	Manchester encoding	Various, e.g. PAM-2/3/4/5, DSQ128, NRZ
Media	Coax	UTP, fiber, backplane, Twinax

layers. Likewise the Ethernet packet has remained almost unchanged, with only the addition of the optional 802.1Q header.

1.2.2 Ethernet in industrial automation

Communication in industrial automation is generally structured hierarchically (see also Figure 1.6). The lowest level of communication happens between sensor or actuator and the low-level controller [22, 23]. The amount of data transmitted with every cycle can consist of a few bits only. Nevertheless, the communication needs to be cost efficient and the response time short. Cycle times for tasks like motion control can be much less than 1 ms with a synchronization accuracy of not more than 1 μs [24]. At a machine level, more intelligent field devices like I/O stations, operator panels, and Programmable Logic Controllers (PLCs) exchange data. For most tooling machines or remote I/O a response time of below 10 ms is required. At the floor (or "field") level, automation and operator stations communicate with PCs. A response time of 100 ms is sufficient for activities like process monitoring and thus most processes in process automation and building control [24]. Often the floor level is subdivided into smaller "cells" and larger "areas." This allows the separation of critical from not-so-critical cells and, in the case

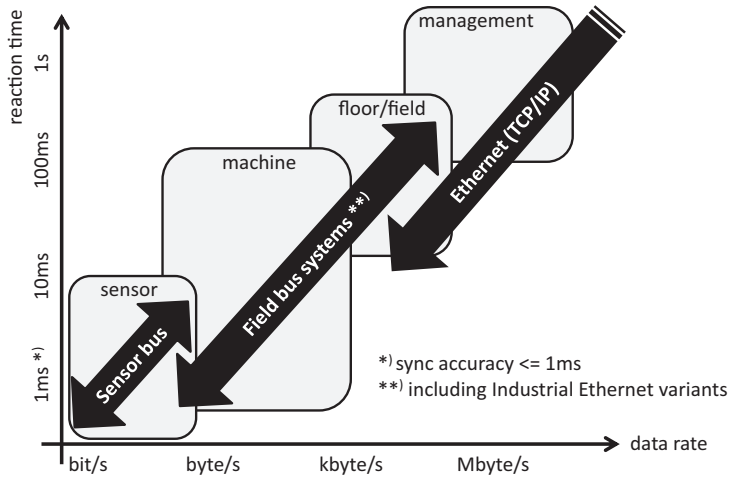

Figure 1.6 Hierarchical approach to factory communication [24, 32].

of issues, enables them to be isolated as well as repaired without affecting the whole production. At the highest, management-level orders, reports, quality statistics, etc. are handled. The requirements for the reaction time are less critical, while the packet size and amount of data increase.

The process of industrialization is the foundation of wealth in occidental society. Hence, right from the beginning of the industrial revolution, efforts have been made to improve and optimize production processes. Naturally, the possibilities of computerization were explored from the early days and the foundations for hierarchical communication were laid in the early 1970s. After the 1960s had brought a number of inventions impacting industrial manufacturing – mini computers, robots, computer-controlled "NC" machines, and especially Programmable Logic Controllers (PLCs) – there was a need for efficient communication between the units as well as the possibility for decentralizing their control [25]. It was found that decentralization improved the quality and availability of process observation and control as well as unburdened the central computer. At the same time it removed the need to use a star architecture, and thus reduced the amount of cabling [26]. The first commercially available distributed computer control systems were introduced by Honeywell and Yokogawa in 1975 [25].

The rest of industry followed and in the 1980s every company in the automation business seemed to develop their own "fieldbus"[8] system in order to support the respective communication in manufacturing plants. The large number of fieldbus variants (>50 [27]) nevertheless did not appeal to the customers. In the case of technical problems manufacturing plant owners need access to replacements fast – potentially from a different vendor – to minimize the risk and impact of downtimes. In consequence, suppliers published their specifications [28], which helped to establish fieldbus systems in industrial automation. Up till today fieldbus connected nodes represent the majority of new as well as existing nodes in industrial plants [29]. At the same time, efforts towards standardization were made, although the outcome of those is a somewhat double-edged sword: when the IEC finally adopted its IEC 61158 standard on 31 December 2000, it contained no less than 18 variants [24]. The possibility to have interoperable solutions in general and the possibility to have perfectly fitting solutions for different use cases, was obviously more important and more advantageous than to have a single solution that covers all [30]. The respective standardization efforts in IEEE (802.4) were finally disbanded in 2004 [31].[9]

Fieldbusses can fulfill very small reaction time requirements (see also Figure 1.6). Investigations into the use of fieldbus technologies showed the advantages of using one technology only [33]. Nevertheless, many publications mention the additional use of a separate sensor bus for cost reasons (e.g. [32]). On top, the standard Ethernet TCP/IP is used to integrate the management level, which makes it three technologies at minimum. The desire for seamless communication over all hierarchy levels and parts of the production process for complexity and cost reasons is easy to understand, and this made "Ethernet," being part of the system anyway, an obvious choice. Standard Ethernet TCP/IP is nevertheless non-deterministic and reaction times can be above 100 ms, although there are simple means to reduce this, like using UDP instead of TCP or restricting the possible traffic in local sections of the network. With the resulting

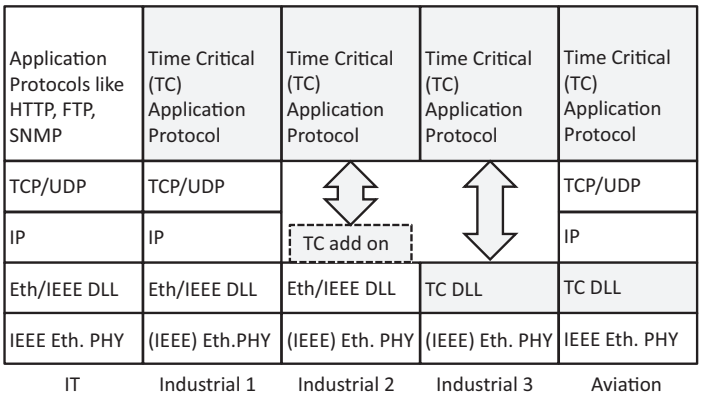

Figure 1.7 Conceptual real-time variants in Industrial Ethernet [23, 24, 34].

reaction time of 10 ms [23] a significant number of applications in industrial automation can be covered.

To make Ethernet (even) more suitable for real-time applications and to fulfill various additional requirements on robustness, functional safety, high availability, and security combined with low latency, "Industrial Ethernet"[10] solutions were developed. Figure 1.7 shows the different concepts behind them. In the simplest case a protocol specifically catering for time-critical use cases is used on the application layers ("Industrial 1"). The next option ("Industrial 2") is to have the time-critical traffic bypass the IP and TCP/UDP layers and to directly communicate with the data link layer. The reaction time can thus potentially be shortened down to 1 ms [23]. This bypass concept is also used in the IEEE 802.1 Audio Video Bridging (AVB) and will be discussed in more detail in Section 5.1. Nevertheless, the addition of a non-standard add-on to the data link layer, as might be the case in an Industrial Ethernet version, is not under discussion for automotive. In the last variant depicted ("Industrial 3") the data link layer is redefined in order to accommodate the real-time requirements directly in the MAC. This implies the most significant changes that might even affect the implementation in hardware down to the PHY. Even though the aviation industry does not reuse any of the Industrial Ethernet variants for the communication between avionics systems in an aircraft (see also Section 1.2.3), the "Aviation" structure depicted in Figure 1.7 is in the end just another version of "Industrial 3." One of the basic principles behind almost all Industrial Ethernet versions is that the "IT" part of the communication is used for best-effort traffic and that Standard Ethernet hardware is used for the PHY. Note that special variants of cabling and connectors generally are used for robustness in the physically harsh environment of industrial manufacturing [24].

Industry thus does not only use a large number of fieldbus variants but also various incompatible types of "Industrial Ethernet." Twenty-nine versions of real-time Ethernet are listed in [35], and seven are mentioned with respect to their market share in [36]. From an automotive perspective, this is surprising, because many incompatible networking technologies result in additional costs and overhead. Even if the costs for the networking