Ethernet evolutions

Fulvio Risso

Politecnico di Torino

An historic perspective (1)

- Ethernet was very successful
  - On the wired side, basically no competitors
- New problems soon appeared
  - Necessity of more speed
    - FDDI was used to create backbones, but expensive and not appropriate for desktops
  - Necessity to interconnect more networks
    - FDDI was not able to interconnect Token Ring networks
      - MTUs: FDDI 4352, 4Mb Token Ring 4464, 16Mb Token Ring 17,914

An historic perspective (2)

- Features not negotiable in new LAN standards
  - Keep compatibility with frame format
  - Preserve investments in human workforce
    - Hard to convert people to new technologies
- Two new specs appeared
  - 100VG-Anylan (Hewlett Packard)
    - Demand Priority Protocol, a sort of polling-based mechanism
  - Fast Ethernet (3Com)
    - Classical Ethernet, but ported to 100Mbps

An historic perspective (3)

- 100VG-Anylan disappeared soon
- FastEthernet: slow adoption at the beginning
  - FDDI better for backbone (robustness and network diameter)
  - No such need of speed at the desktop
  - Mainstream starting from ‘95-’96
  - No interconnection for Token Ring networks (MTU issues)
    - Killed Token Ring Technology
- Later, Gigabit Ethernet
  - Mainly for backbones, now also for desktop
- Now, 10GB Ethernet
  - Intended to cover MANs as well
- Even faster Ethernet are in the pipeline

Fast Ethernet: IEEE 802.3u

- Characteristics
  - Same frames, same CSMA/CD algorithm
- Basically, Ethernet moved at 100 Mbps

- Everything remains the same except
  - Collision domain (see later)
  - Physical layers
    - Often derived from other existing technologies (e.g., FDDI)

Fast Ethernet: physical layers

- 100BASE-T4 (twisted pair cable, 4 pairs)
  - 8B/6T: 37.5MHz
- 100BASE-TX (twisted pair cable, 2 pairs)
  - 4B/5B + MLT-3: 31.25MHz
- 100BASE-FX (fiber)
  - 4B/5B
- TX, FX: derived from TP-PMD/PMD of FDDI (ISO 9314-3) with minor modifications
  - Other standards have been defined, but never implemented
    - The existing ones worked extremely well

Fast Ethernet and Collision Domain

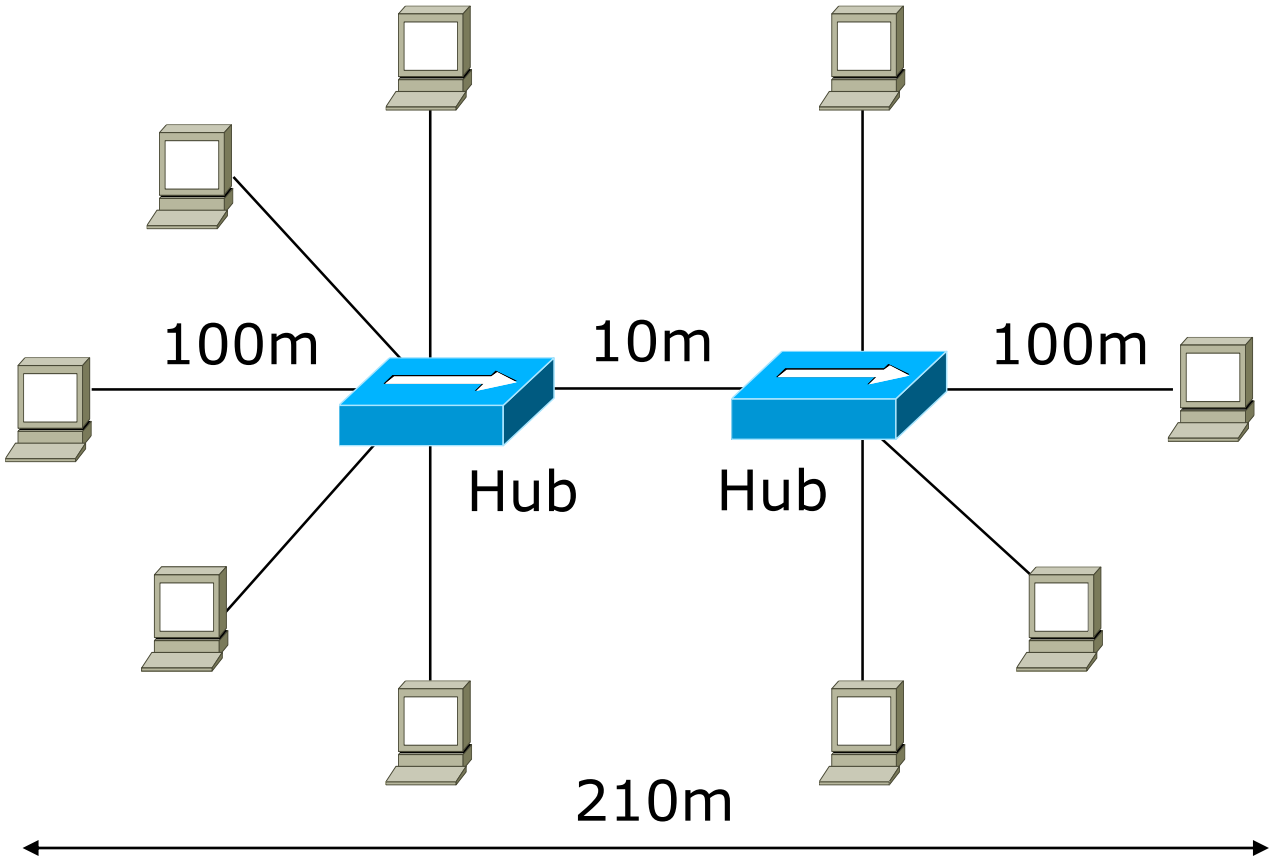
- Four parameters are strictly correlated in the CSMA/CD algorithm
  - Signal propagation speed
  - Bit time (i.e., bandwidth)
  - Minimum frame
  - Maximum allowed distance
- Since
  - It’s hard to change the signal propagation speed
  - The minimum frame must be kept as is for backward compatibility
- We have…
  - That any increase in the bandwidth will trigger a decrease in the maximum network diameter, and vice versa

Fast Ethernet vs Ethernet (10Mbps)

- Differences
  - 10x increase in speed
    - Data Rate 100Mb/s
    - Bit time 10ns
    - Inter-frame gap 0.96μs
    - Slot time 5.12μs (512 bits / 64 bytes)
  - /10 in distance (200m + 20m)
    - Reduced collision domain
    - Basically, Host – hub – host
    - Rather limiting
- Introduces "Full Duplex" mode
  - No CSMA/CD on that link

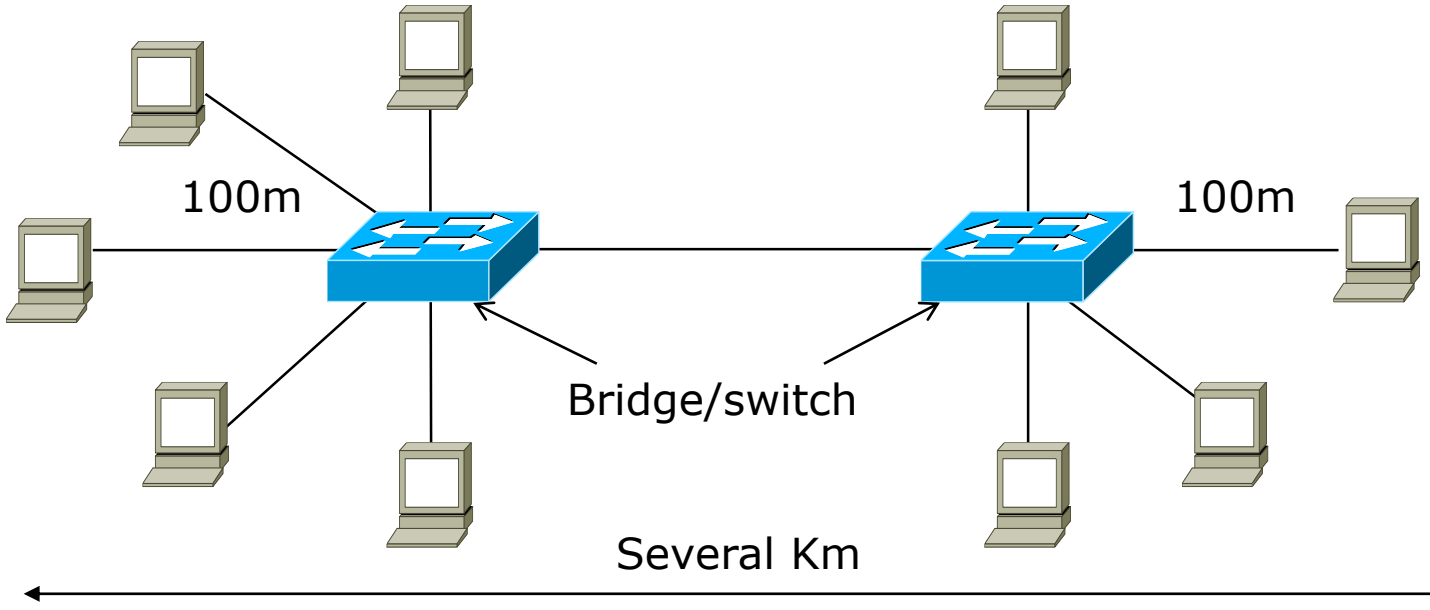
Fast Ethernet topology (1)

- Limited network size, but still usable
  - Compatible with structured cabling limits (100m)



Fast Ethernet topology (2)

- Bridges/switches becomes common in those times
  - No CSMA/CD → less limitations to the network diameter
    - The limit becomes the attenuation on cables
  - Finally FastEthernet was ready for backbones as well
- Very complex topologies (several bridges in cascade)



Gigabit Ethernet: IEEE 802.3z (1)

- 1Gbps
- Characteristics
  - Same frame
    - Required to maintain interoperability with other Ethernet standards
    - Same format, 64/1518 bytes frame size
  - Same CSMA/CD algorithm
- At a first sight, FastEthernet at 10x speed and 10/ distance
- Note: compatibility at frame level is more important than compatibility at CSMA/CD level

GbE and collision domain (1)

- A simple extension of FastEthernet originates a toy network
  - Usual rule: speed 10x, network diameter 10/
  - Collision domain was ridiculous (~20m), if other parameters kept unchanged
- Needed to define some tricks in order to increase the network diameter
  - Modification in the slot-time (presented later)
- But… was CSMA/CD still needed in GbE?

GbE and collision domain (2)

- A lesson learned from Fast Ethernet
  - Very slow adoption at the beginning
    - The limited network diameter was really a show stopper
    - Is it reasonable to have a backbone max 200m wide?
      - Most people still used FDDI, which did not have such limitations
  - FastEthernet become successful only when switches become mainstream
  - The idea of even faster Ethernet standards (e.g. GbE) with even a smaller network diameter was known to be very stupid from the beginning
- So, why did Gigabit Ethernet use CSMA/CD?
  - Not really needed from the technical point of view
    - A pure switched network does not use CSMA/CD and it works
  - Economic reasons?

GbE and collision domain (3)

- So, CSMA/CD specified anyway
  - With some additional tricks to have a reasonable network diameters
- And so there were repeaters, etc.

- Nobody never implemented those specs
  - All GbE product did not use the CSMA/CD
  - Worked only in a pure switched (i.e., full-duplex) environment

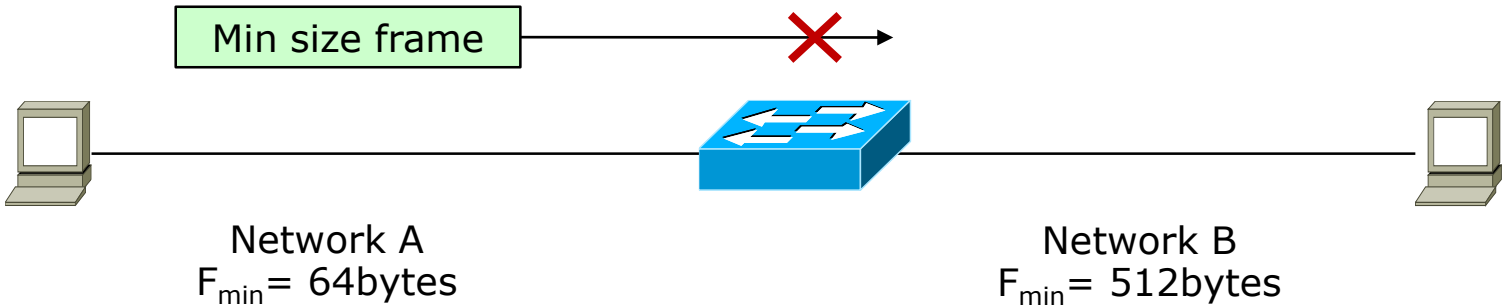
Gigabit Ethernet vs previous Ethernets

- In fact, GbE brings some new idea compared to FastEthernet, which is simply a 10x Ethernet
  - Increased slot time and added Carrier Extension
  - Added Frame Bursting
  - Full-duplex becomes, in fact, the standard operating mode (CSMA/CD no longer used in practice)

- Why Gigabit Ethernet?
  - Well, hardware is cheap
  - Market demand (and vendor offer)
  - May be useful in the server domain and for backbone links

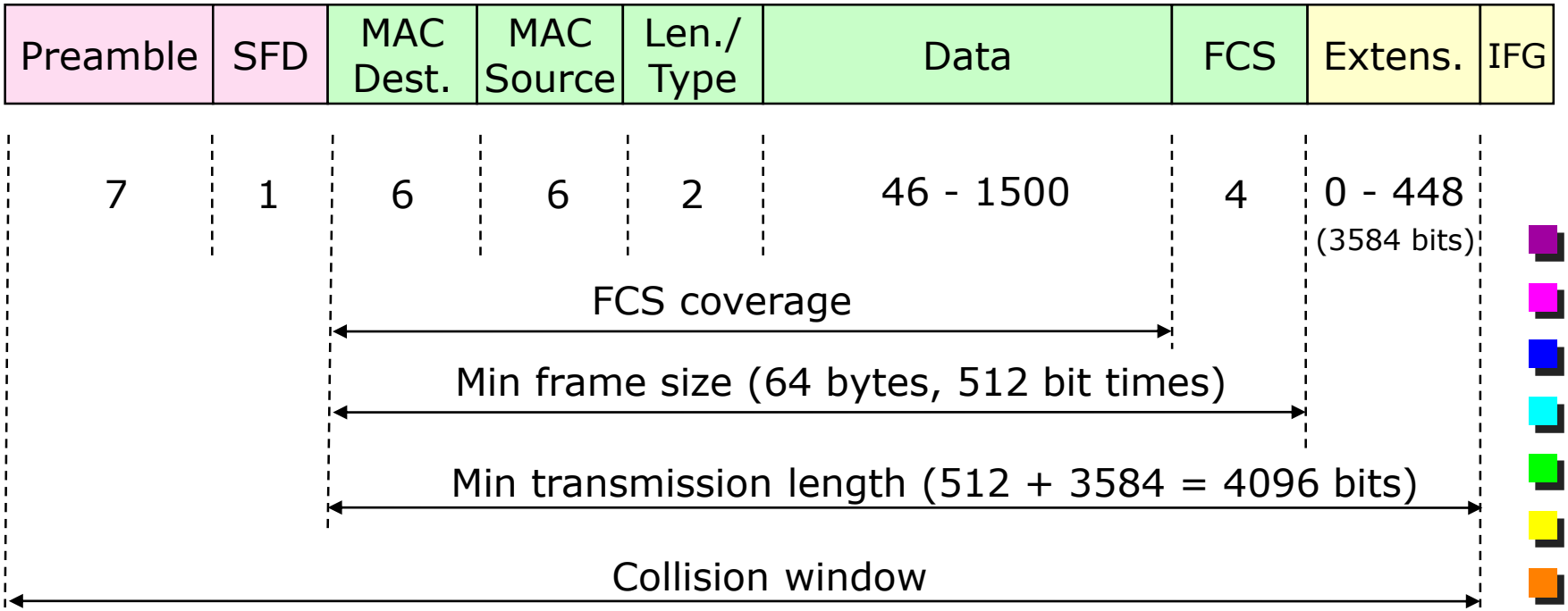
Gigabit Ethernet: Carrier Extension (1)

- The minimum frame size of 64 bytes limits the network diameter
  - Need to increase the minimum duration of the transmission
    - Please note that "Min duration of the transmission" is different from "minimum frame size"
  - But… cannot increase minimum frame size (for compatibility)
    - If so, how can we transport a 64B Ethernet frame into GE?
  - We increase the *slot time*



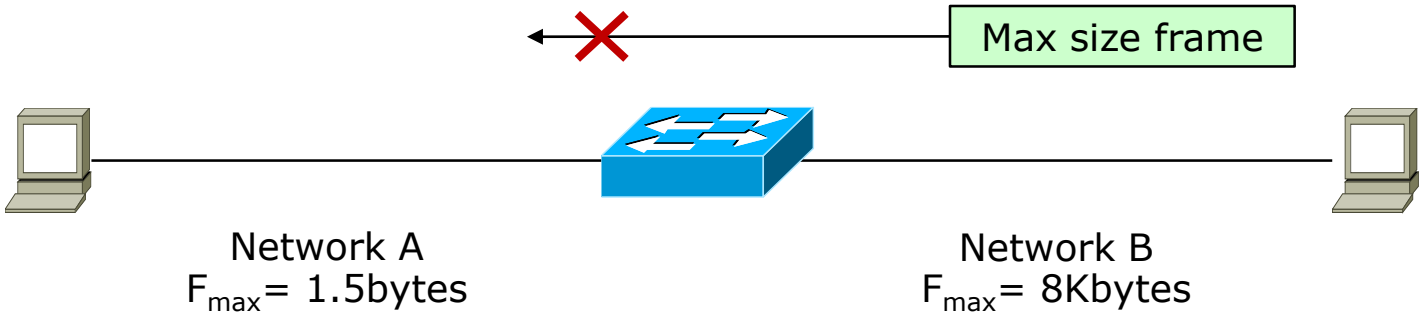
Gigabit Ethernet: Carrier Extension (2)

- Pad short frames to reach at least 4096 bit times (slot time)
  - Predefined sequence of symbols
    - Valid at physical layer, but not used when transporting Eth data
- Frames bigger than 4096 bits (512 *bytes*) are not extended



Gigabit Ethernet: Frame Bursting (1)

- The maximum frame of 1518 bytes is obsolete
  - In Ethernet, 1518 bytes → 1.2ms channel occupancy
    - Note: we have to count also PRE, SFD and IFG for channel occupancy
  - Reasonable to guarantee statistical demultiplexing
  - Cannot increase the maximum frame size
    - If so, how can we transport a large GE frame into Ethernet?
  - We concatenate several frames one after the other: *frame bursting*

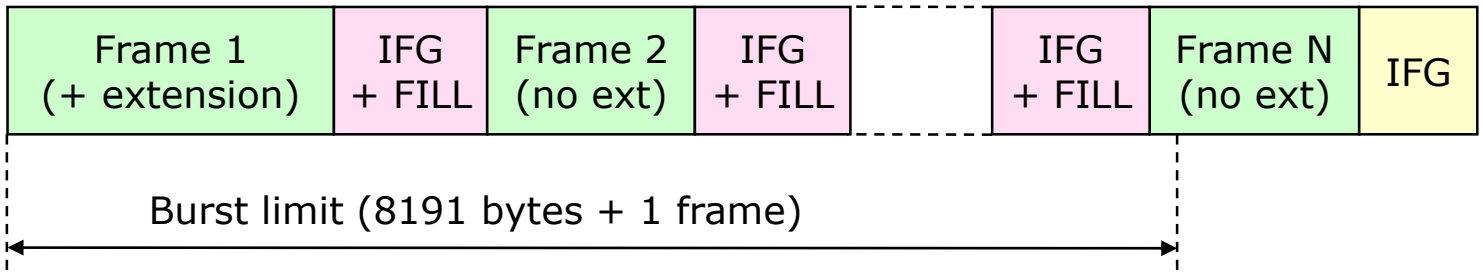


Gigabit Ethernet: Frame Bursting (2)

- Gigabit Ethernet allows an host to transmit several consecutive frames without releasing the channel
  - Burst-limit equal to 65528 bits (8192 Bytes – 1) + 1 frame
  - The frame currently under transmission when the 8192 limit is passed is still allowed to go on
- Advantages
  - Carrier extension (optionally) present only after the first frame
  - No "lost" time in contention after each frame (only after the burst)
  - Throughput increases especially in case of short frames
  - **The maximum frame size is still 1518 bytes**
    - Compatibility with older Ethernet networks
  - Very simple implementation (a counter when exceeding 8192 bytes)

Gigabit Ethernet: Frame Bursting (3)

- Mechanism
  - First frame must last at least the collision window (i.e. must be extended if shorter than slot time)
  - Inter-Frame Gap is still present, but the physical coding is different in order to distinguish this case from the "standard" IFG
    - Called "Filling Extension" (indicated as "IFG+FILL" in the picture)
    - Required in order to delimit frames
    - Always 96 bit times
  - Other hosts must wait till the frame ends (with IFG)
  - **All frames include SFD, Preamble and the actual frame**



Gigabit Ethernet: Frame Bursting (4)

- Was that needed?
  - More efficiency
    - Well, not much bytes saved in the headers
      - All the headers remain the same, except for the carrier extension in the first frame
        - By the way… do you expect small or big packets when you need frame bursting?
      - Even Preamble and SFD are there
    - And then… who cares when your efficiency (with 1500B packets) is already 98%?
- Or… was Frame Bursting added mostly for other reasons?

GbE: recap of most important parameters

- Slot time increased to 4096 bits (512 *bytes*)
  - ~200m diameter (star-based topology: 100m + hub + 100m)
- IFG kept at 96 bit times
- Bit time reduced to 1/10, hence speed increased 10x

	Ethernet	Fast Ethernet	Gigabit Ethernet
Transmission speed	10 Mbps	100 Mbps	1 Gbps
Bit time	100 ns	10 ns	1 ns
Inter-frame gap	9.6 us	0.96 us	96 ns
Slot time	51.2 us	5.12 us	4.096 us

Working modes

- Shared mode (i.e., CSMA/CD) to be used with repeaters
  - Not used
  - Not implemented in any commercial product
- GbE usually deployed in Full Duplex mode
  - No carrier extension
    - Collisions does not exist
  - No burst mode
    - Contention does not exist

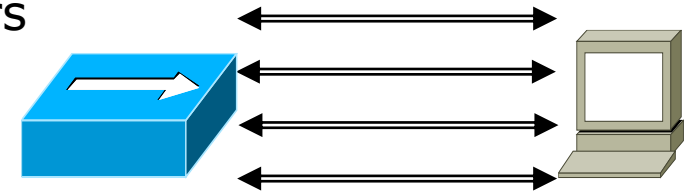
Gigabit Ethernet: Physical layer

Standard	Cabling	Use	Max length	Coding
1000BASE-SX	MMF 50/125 um (400 MHz * Km, 850nm) MMF 50/125 um (500 MHz * Km, 850nm) MMF 62.5/125 um (160 MHz * Km, 850nm) MMF 62.5/125 um (200 MHz * Km, 850nm)	2 fibers	550m 550m 220m 275m	8B10B
1000BASE-LX	MMF 50/125 um (4/500 MHz * Km, 1300nm) MMF 62.5/125 um (500 MHz * Km, 1300nm) SMF 10/125 um	2 fibers	550m 550m 5Km	8B10B
1000BASE-CX	STP (jumper cable), 150Ohm	2 pairs	25m	8B10B
1000BASE-T	UTP (balanced 100 Ohm, Cat. 5E)	4 pairs	100m	PAM

MMF = Multi Mode Fiber
SMF = Single Mode Fiber

1000BASE-T (IEEE 802.3ab)

- Full-duplex transmission over 4 pairs
  - 250 Mb/s per pair
  - Hybrid transformers



- PAM5 Line coding (5-level Pulse Amplitude Modulation)
  - A signal of 5 different levels is transmitted over the 4 pairs
    - 5^4=625 possible symbols, of which only 256 are valid
    - Each pair transports 2 bits → 125 Mbaud (250Mbps) per pair
    - Redundancy used for control codes
- Cat 5 UTP has to pass additional tests compared to the ones defined by TIA/EIA ISB95

1000BASE-X

- Sub Standard
  - 1000BASE-CX (copper short range)
  - 1000BASE-SX (short wavelength)
  - 1000BASE-LX (long wavelength)
- Based on Fiber Channel (FC) Physical Layer
  - Code 8B10B
  - Redundancy code: control symbol and transitions

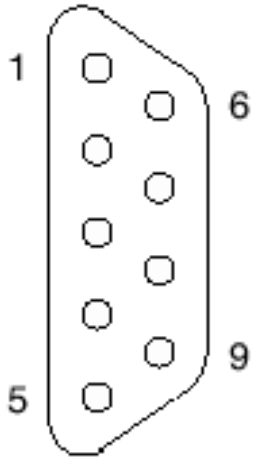
1000BASE-CX connectors

Type 1 connector

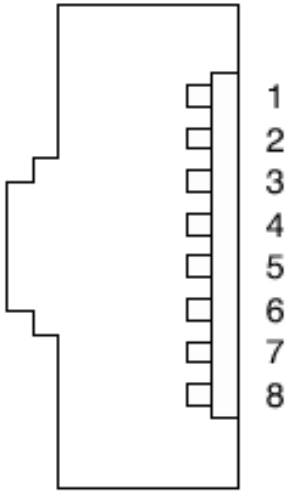
1: Transmission +
6: Transmission -

Shell: shield

5: Reception -
9: Reception +

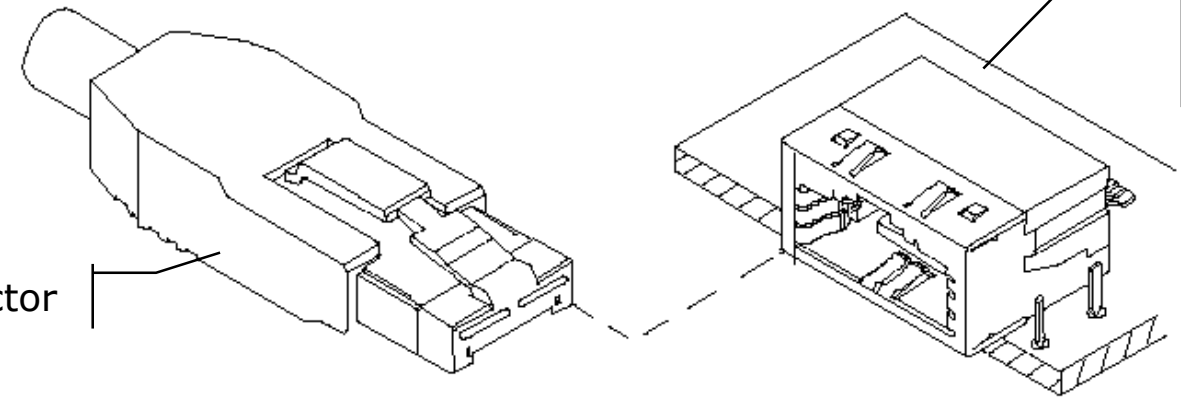


Type 2 connector

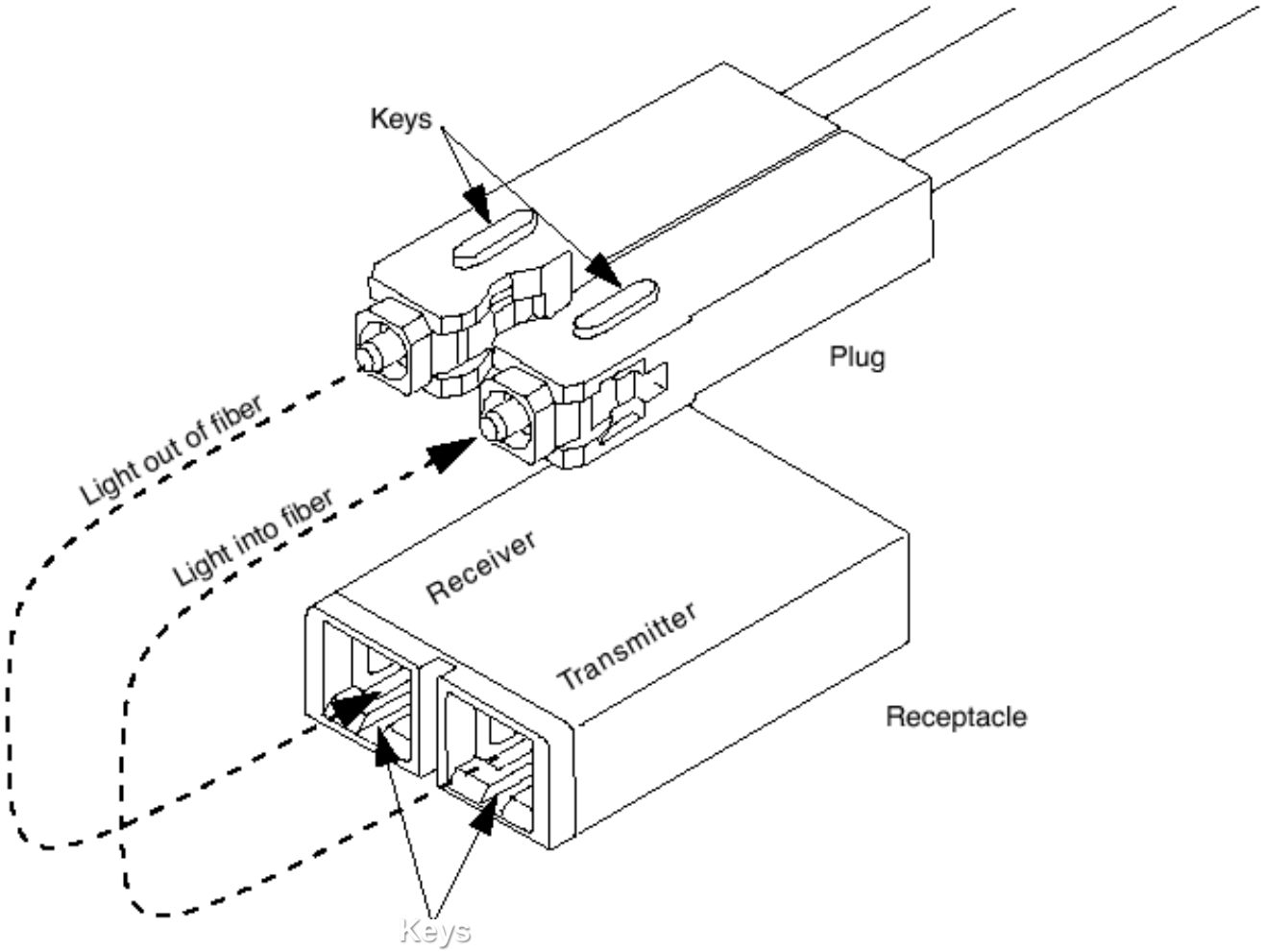


1: Transmission +
3: Transmission -

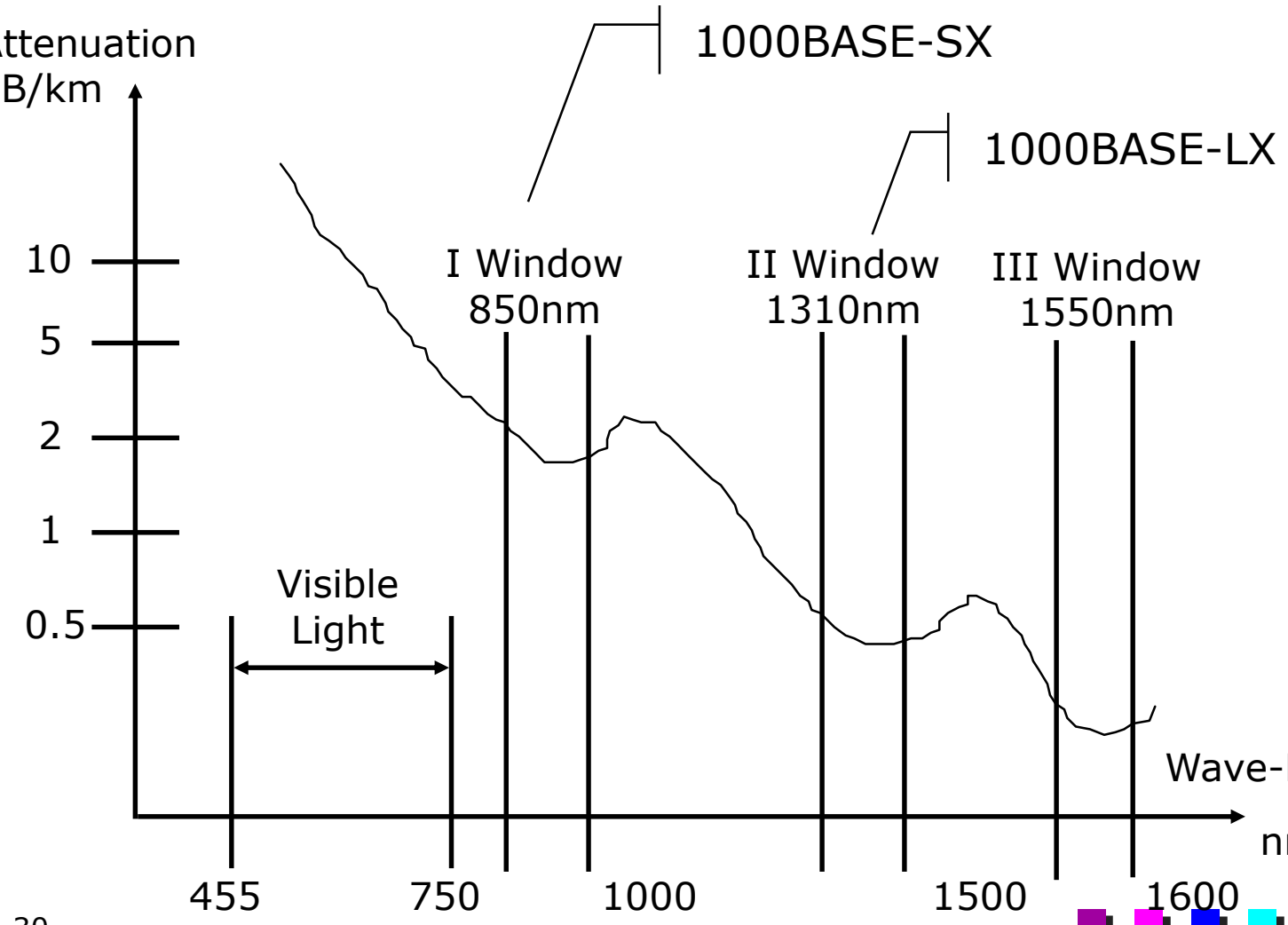
6: Reception -
7: Reception +



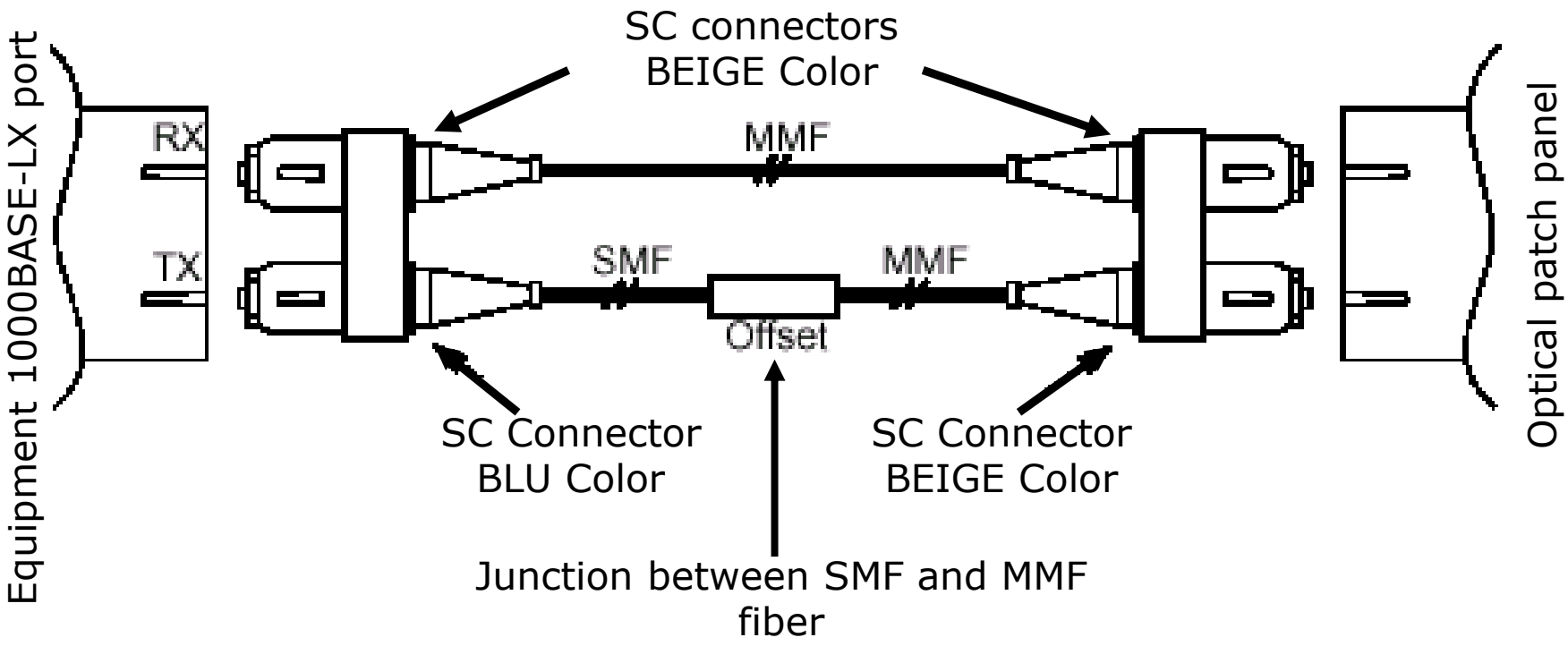
1000BASE-SX and 1000BASE-LX connectors



Wave-Length and standard



1000BASE-LX & multimode fiber: Mode Conditioning Patch Cord



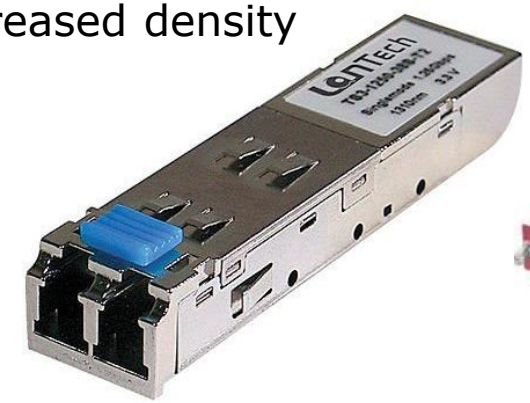
MMF = Multi Mode Fiber
SMF = Single Mode Fiber

Non standard products

- 1310 nm single-mode fiber: 10 Km
  - Example Cisco GBIC 1000BASE-LX/LH
- 1550 nm single-mode fiber dispersion shift: 100 Km
  - Example Cisco GBIC 1000BASE-LZ
- Interoperability between products of different vendors is not guaranteed

GbE and GBIC

- The most part of the Ethernet evolution is now in the physical levels
  - Domain of electronic/telecommunication engineering
  - Computer engineers do not have much to say
    - Ethernet principles (i.e., framing) are always the same
- GBIC are a common solution for having the possibility to update the physical layer without having to update the rest of the equipment
  - Mini-GBIC for increased density



10 Gigabit Ethernet - IEEE 802.3ae

- IEEE 802.3 frame
- Full-duplex mode
  - No repeater
  - No CSMA/CD
  - No carrier extension
- Keep Ethernet's good reputation
  - 10 times more efficient
  - 3 times more expensive
- Break into metropolitan network (MAN) and wide area network (WAN) markets
  - Price/Bandwidth ratio is better than traditional solutions (e.g. SONET/SDH)

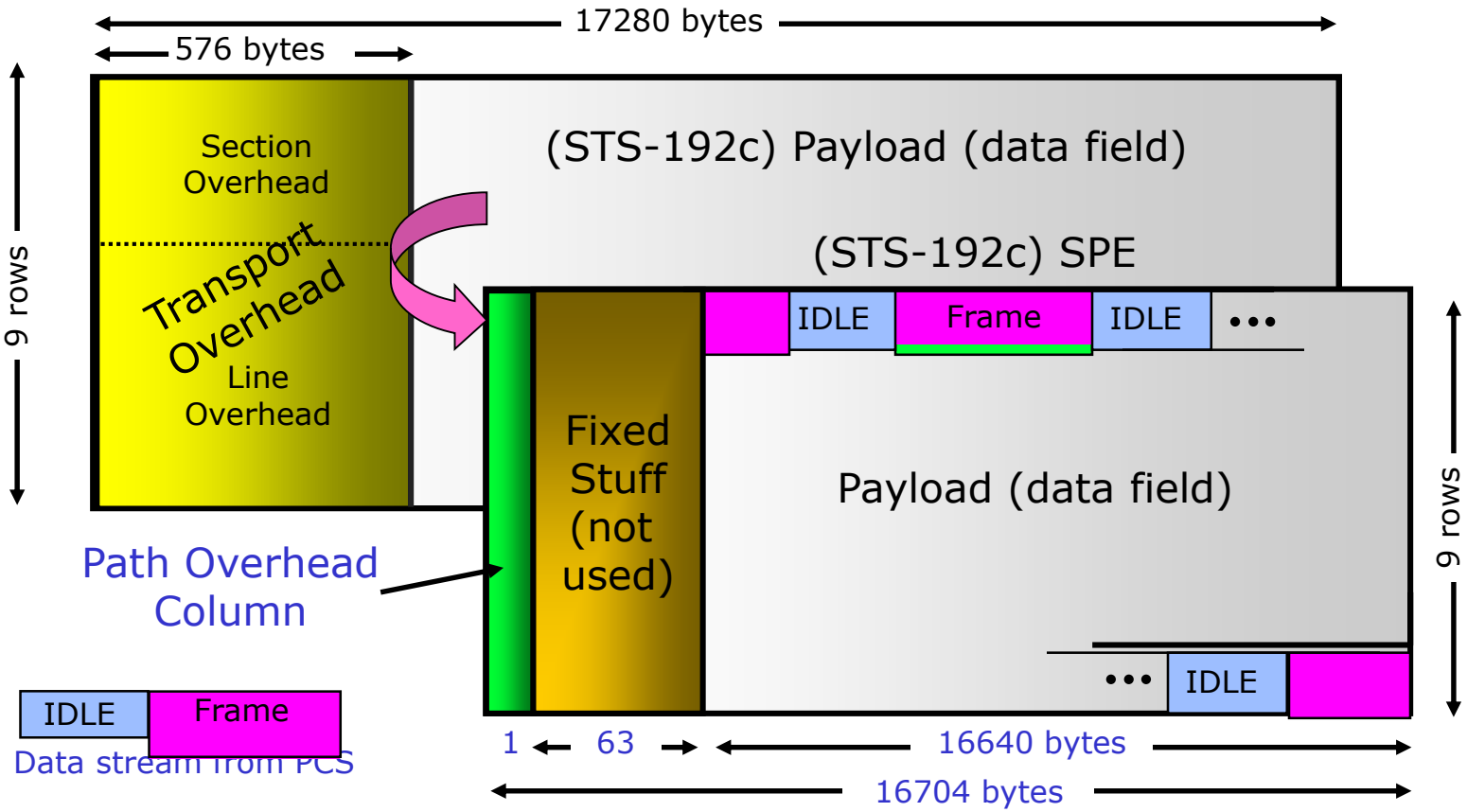
Why 10GbE?

- Not for the desktop
  - No need for such this speed
  - Not compatible with structured cabling (twisted pair)
- Targets
  - Datacenters
  - Backbones
    - If hosts are at 1Gbps, backbone needs to be faster
  - MAN and WAN
    - Ethernet increasingly present there

WAN PHY

- Enables transport over existent MAN and WAN infrastructure
  - DWDM (Dense Wavelength Division Multiplexing)
- Enables existent MAN and WAN component reuse
  - SONET/SDH transceivers and circuitry
- Different transmission speed (9.6 Gb/s) respect to LAN PHY's speed
- WAN PHY and LAN PHY common properties→ market is waiting for components with both functionalities
  - 10GBASE-R and 10GBASE-W in particular
- WIS (WAN Interface Sublayer) tunes PCS' signal
  - Bit scrambling
  - SONET/SDH headers

10GE frame over SONET/SDH



STS-192c = Synchronous Transport Signal – of level 192, c = concatenated

SPE = Synchronous Payload Envelope

10GE and SONET/SDH

- Simplified version of SONET/SDH
  - Avoid imposed complexities required by SONET/SDH
  - Limit component cost
  - Keeps resiliency (SONET or DWDM rings)
- Only some header's fields are used
- High precision synchronization has been removed
  - No Stratum-1 clock (10^{-12} precision)
- Frames are generated and forwarded by 10GE devices in asynchronous mode using
  - SONET/SDH framing
  - Limited SONET/SDH management functionalities

Physical layer

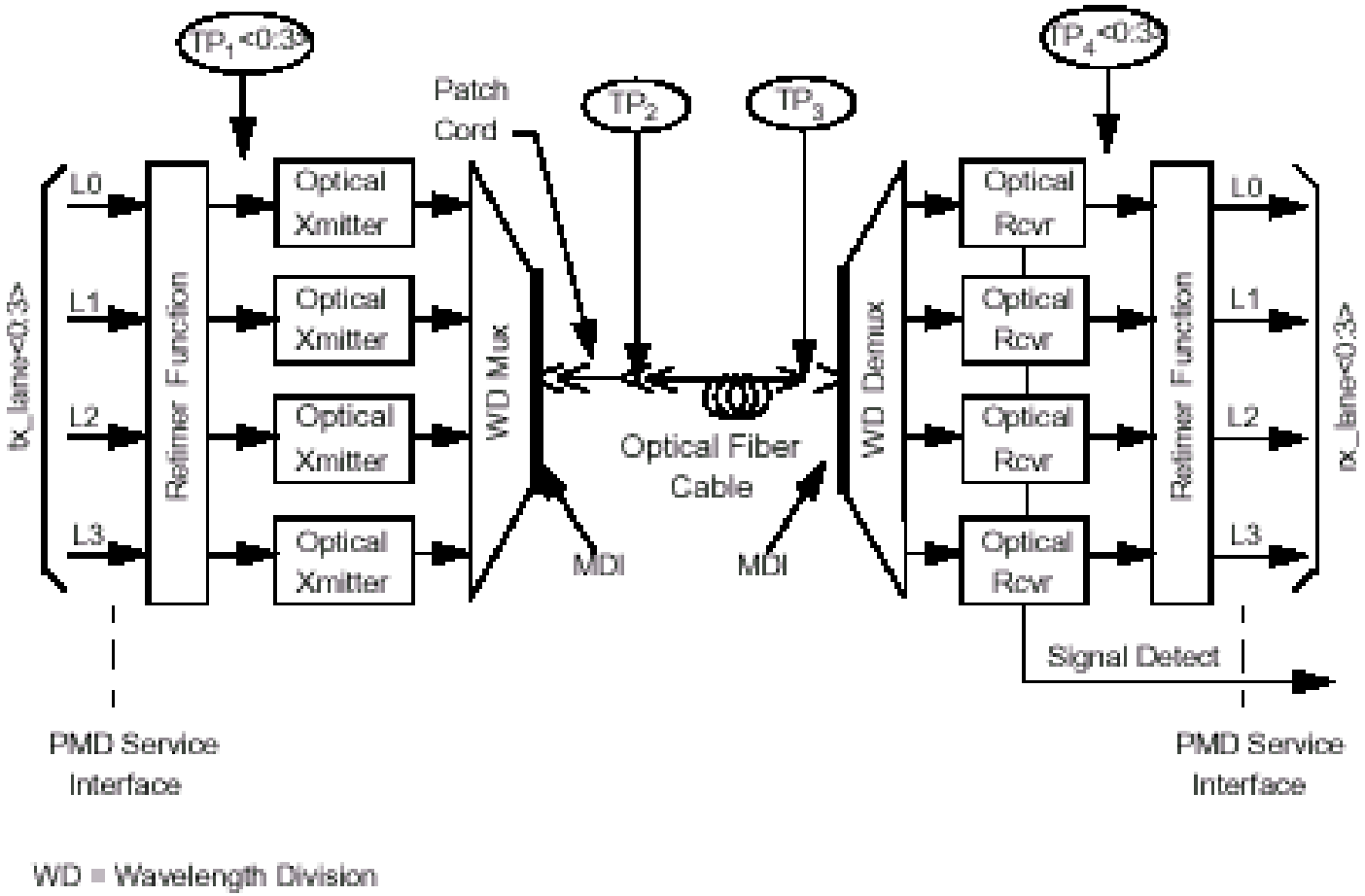
Standard	Fiber	Max length	Window	Usage	Coding
10GBASE-SR	Multimode 62.5 μm	26 – 33 m	850 nm	Building (horizontal wiring)	64B/66B
	Multimode 50 μm	66 – 300 m			
10GBASE-LR	Monomode (10 μm)	10 Km	1310 nm	Area	64B/66B
10GBASE-ER	Monomode (10 μm)	40 Km	1550 nm	Metropolitan	64B/66B
10GBASE-LX4	Multimode 62.5 μm	300 m	1310 nm	Building (orizontal wiring)	FC 10G: 8B10B
	Multimode 50 μm	240 – 300 m			
	Monomode (10 μm)	10 Km		Area	
10GBASE-SW	Multimode 62.5 μm	26 – 33 m	850 nm	Building (orizontal wiring)	64B/66B SONET/SDH framing
	Multimode 50 μm	66 – 300 m			
10GBASE-LW	Monomode (10 μm)	10 Km	1310 nm	Area	64B/66B SONET/SDH framing
10GBASE-EW	Monomode (10 μm)	40 Km	1550 nm	Metropolitan	64B/66B SONET/SDH framing

10GBASE-X

- Copper
- Coding derived from 10G FC (Fiber Channel at 10 Gb/s)
- 32 bit blocks are encoded in 4 blocks of 10 bit each
- Sent over 4 lanes
  - 3.125 Gbaud per lane
- Redundancy used for control codes
  - For example idle signal act as inter-frame gap

10GBASE-LX4

- Fiber



Further evolutions

- Higher speeds (40Gbps, 100Gbps) in the pipeline
  - Datacenter, MAN/WAN
- For the first time, not a 10x
  - Market demand for speed > 10Gbps
  - Specs at 100Gbps not ready yet (and extremely expensive)
  - Decided to go over an intermediate speed
    - 40Gbps was already been developed for DWDM
- Parallelism will be probably a must
  - The same as it happens for CPU, where multicore is the way to increase speed
  - We probably need to use “multi link/path” to be able to drive to upper speed

Readings

- Interesting reading from Wired (July 2011)
  - It covers the most important aspect of the Ethernet history, with enough technical details a lot of background information
  - Suggested for who would like to know also *why* a technology looks like as it is, and not just *how* it looks like
  - Speed Matters: How Ethernet Went From 3Mbps to 100Gbps… and Beyond, from Wired:
    - http://www.wired.com/epicenter/2011/07/speed-matters/all/1

Conclusions

- Ethernet is the facto standard for cabled LANs
  - Widely diffused on MAN as well
  - Progressively used on WAN
- Probably will replace also other standards
  - E.g. Fibre Channel, Infiniband
  - New features (“lossless Ethernet”) being added
- CSMA/CD no longer present
  - Its influence (e.g. min frame) still present nowadays
  - Everything “switched”, Full Duplex
- Framing is basically what remains from Ethernet DIX
  - New Ethernet standards mostly in the domain of electronic/telecommunication engineering
  - Ironically, the only thing that was changed by IEEE in 802.3
- Higher speeds (40Gbps, 100Gbps) in the pipeline