The Ethernet Evolution

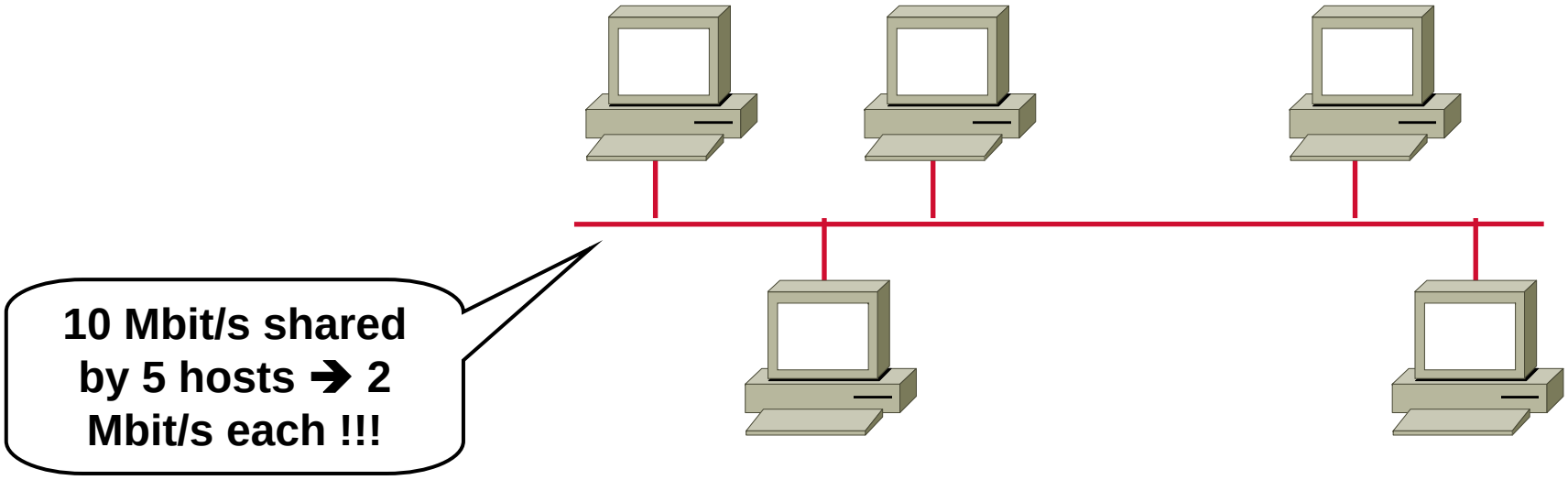
The 180 Degree Turn

“Use common sense in routing cable. Avoid wrapping coax around sources of strong electric or magnetic fields. Do not wrap the cable around flourescent light ballasts or cyclotrons, for example.”

Ethernet Headstart Product Information and Installation Guide, Bell Technologies, pg. 11

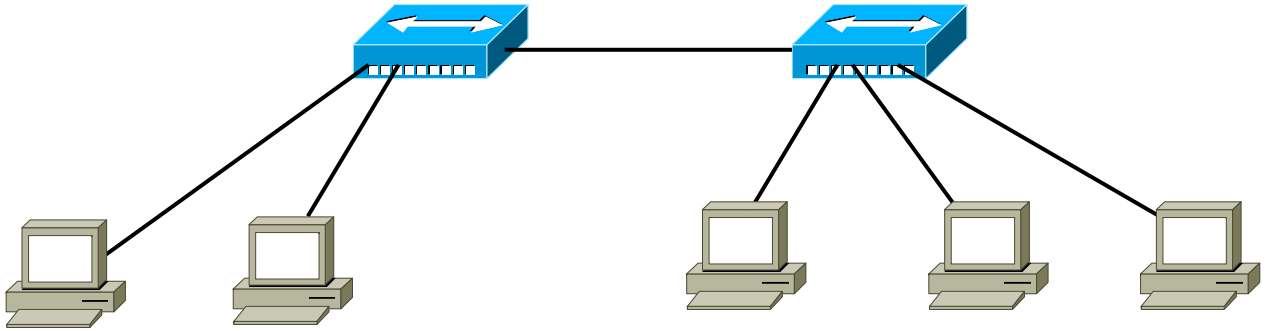
History: Initial Idea

- Shared media ➔ CSMA/CD as access algorithm
- COAX Cables
- Half duplex communication
- Low latency ➔ No networking nodes (except repeaters)
- One collision domain and also one broadcast domain



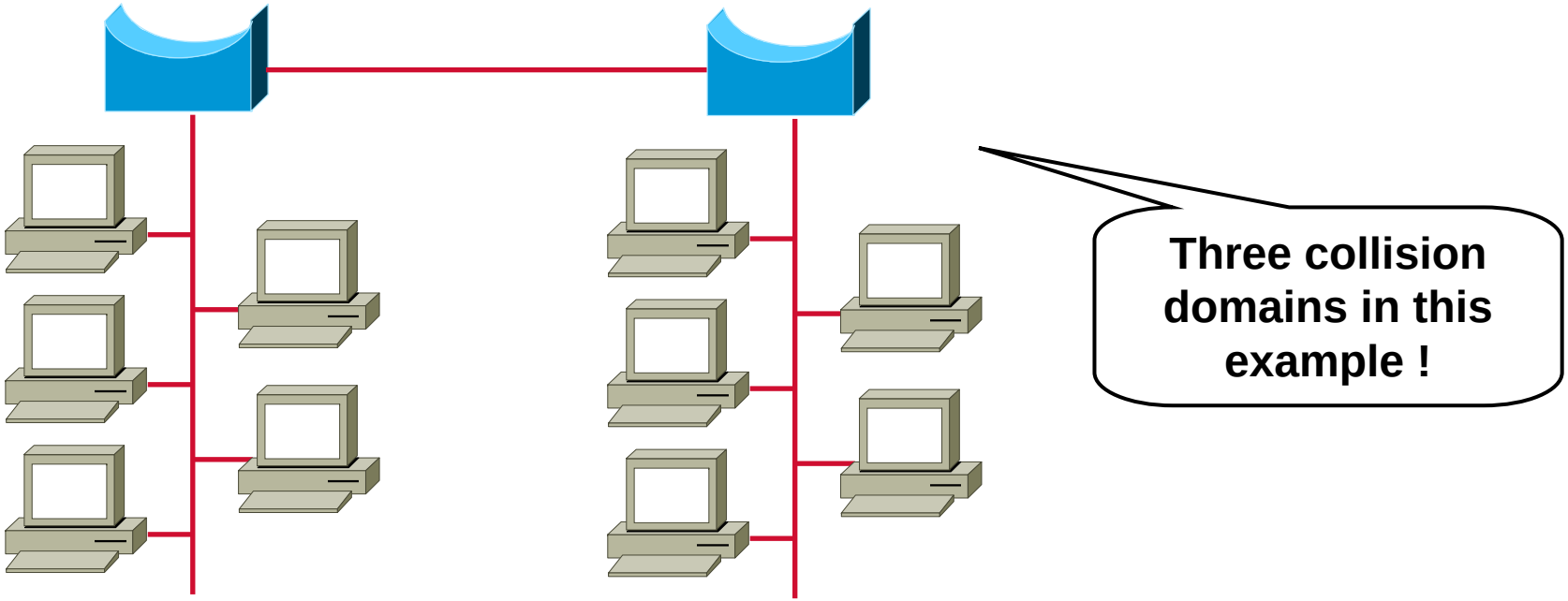
History: Multiport Repeaters

- Demand for structured cabling (voice-grade twisted-pair)
  - 10BaseT (Cat3, Cat4, ...)
- Multiport repeater ("Hub") created
- Still one collision domain ("CSMA/CD in a box")



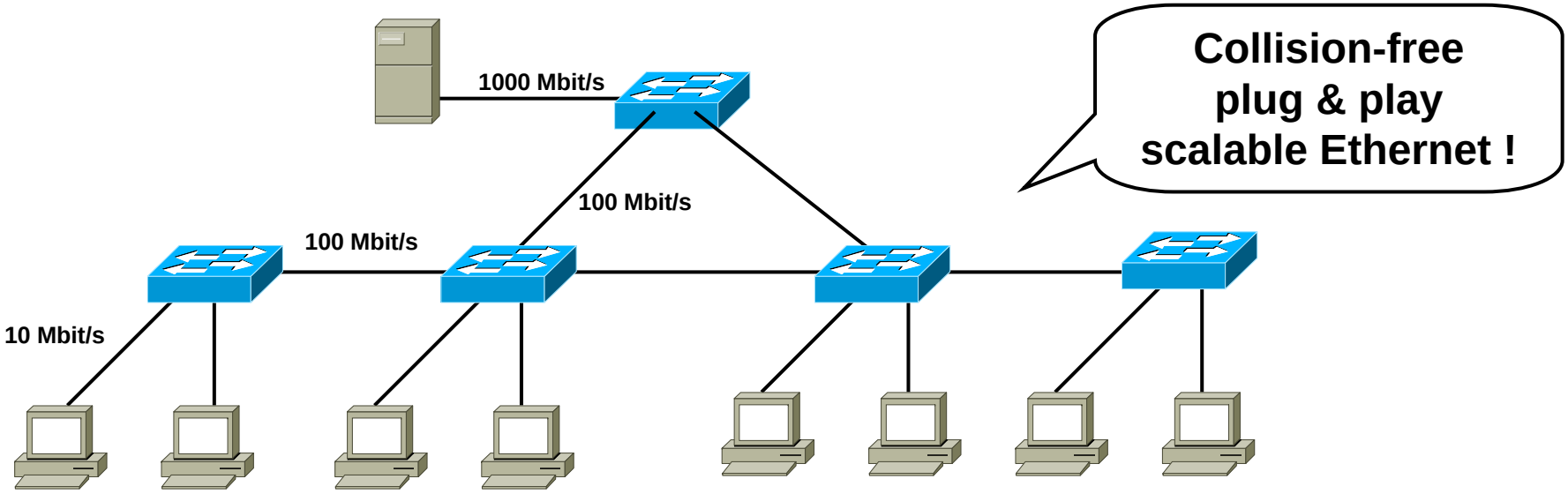
History: Bridges

- Store and forwarding according destination MAC address
- Separated collision domains
- Improved network performance
- Still one broadcast domain

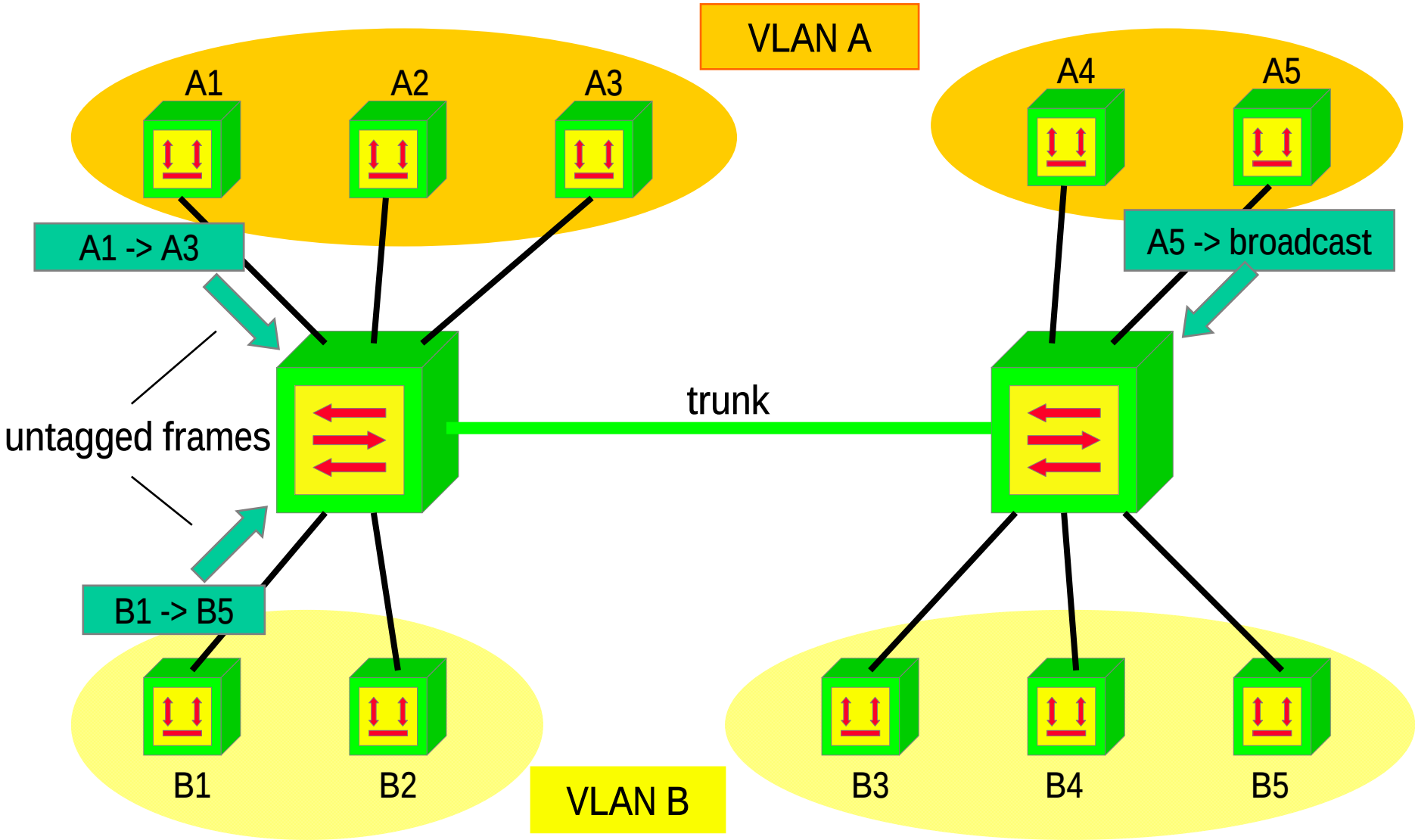


History: Switches

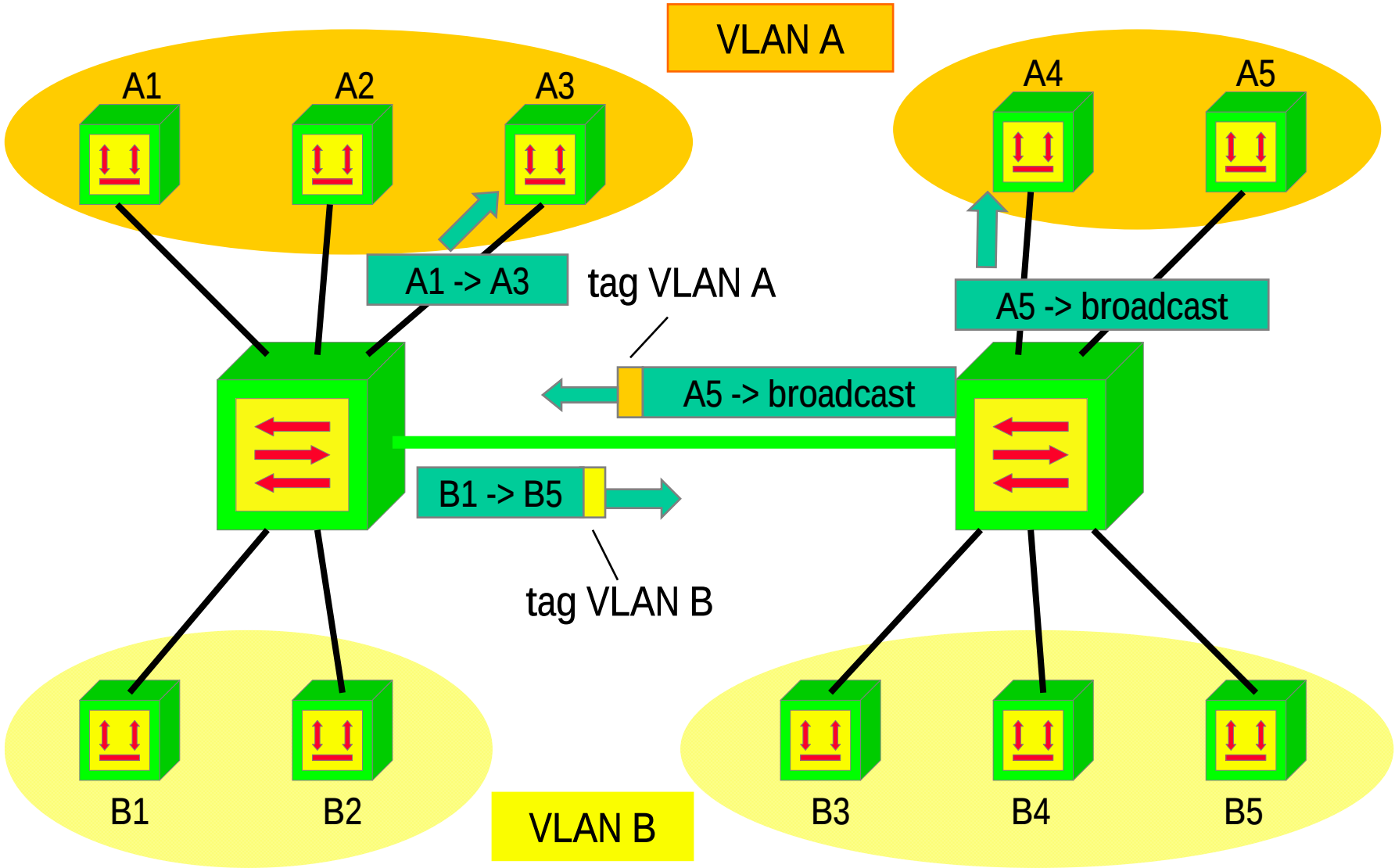
- Switch = Multiport Bridges with HW acceleration
- Full duplex ➔ Collision-free Ethernet ➔ No CSMA/CD necessary anymore
- Different data rates at the same time supported
  - Autonegotiation
- VLAN splits LAN into several broadcast domains



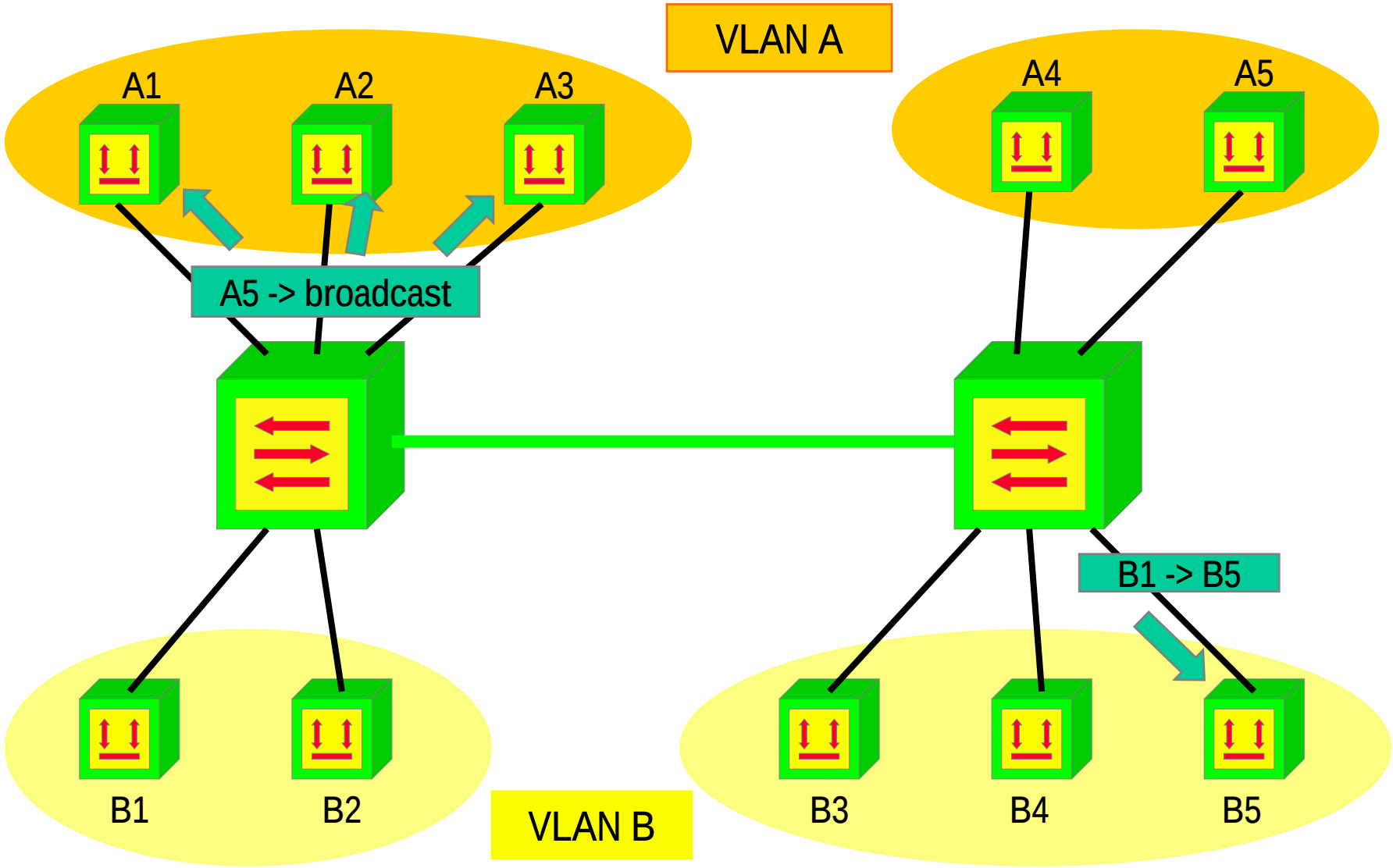
VLAN Operation (1)



VLAN Operation (2)



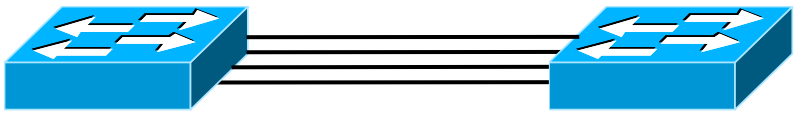
VLAN Operation (3)



Today

- No collisions ➔ no distance limitations !
- Gigabit Ethernet becomes WAN technology !
  - Over 100 km link span already
- Combine several links to "Etherchannels"
  - Link Aggregation Control Protocol (LACP, IEEE 802.3ad)
  - Cisco proprietary: Port Aggregation Protocol (PAgP)
  - HP: Mesh (like L2-routing over 5-8 hops)

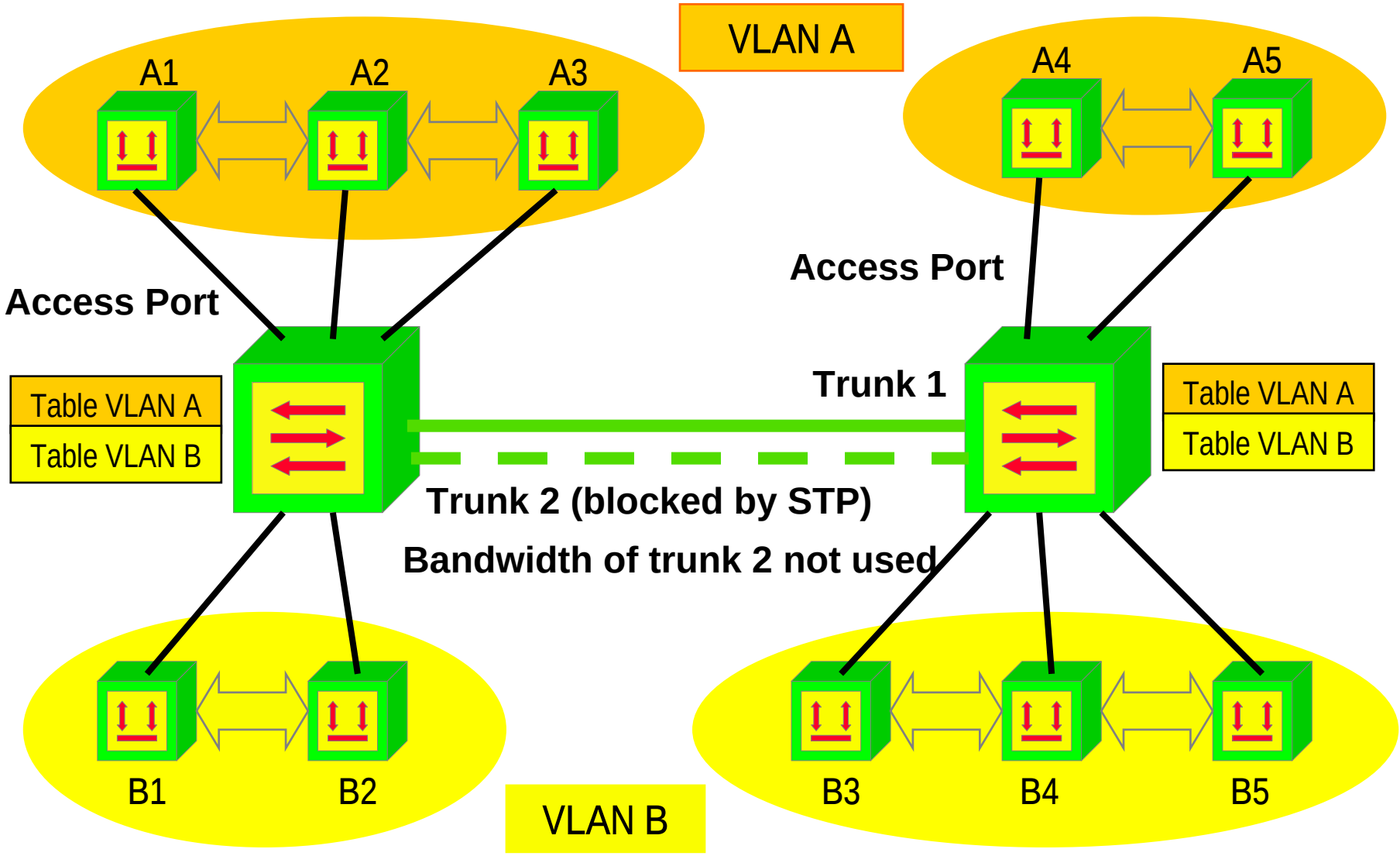
1 Gbit/s or even 10 Gbit/s long reach connection !!!



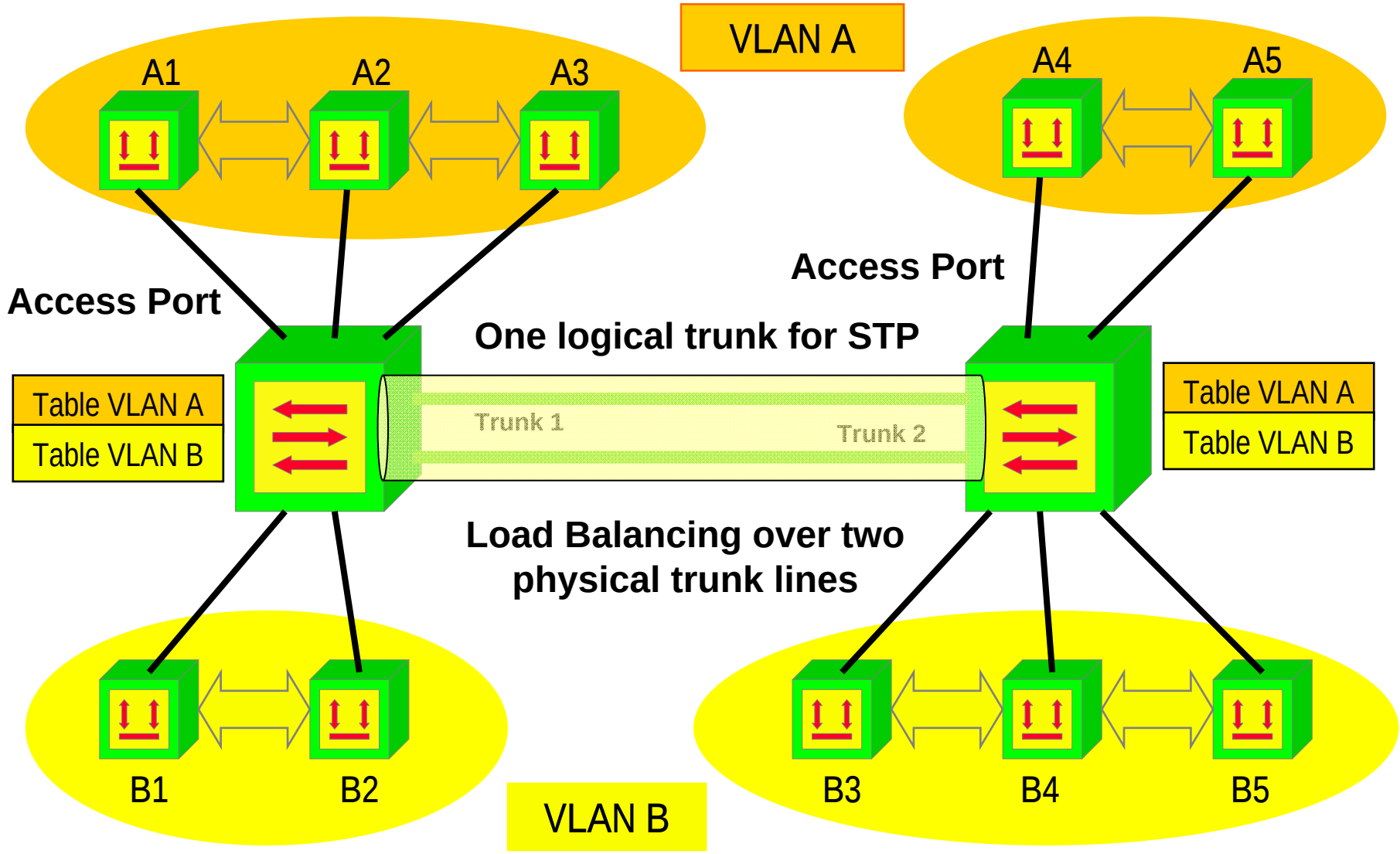
Ether Channel

Note: Spanning Tree regards this as one logical link!
=> Load balancing!

Trunking without LACP / FEC / GEC



Trunking with LACP / FEC / GEC



What About Gigabit Hubs?

- Would limit network diameter to 20-25 meters (Gigabit Ethernet)
- Solutions
  - Frame Bursting
  - Carrier Extension
- No GE-Hubs available on the market today ➔ forget it!
- No CSMA/CD defined for 10GE (!)

CSMA/CD Restrictions (Half Duplex Mode)

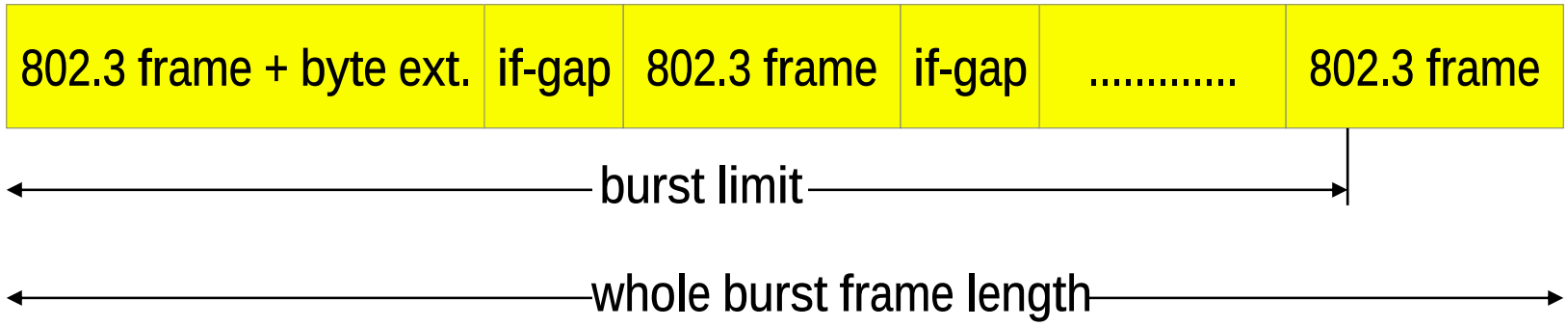
- **Solutions to increase the maximal net expansion:**
  - Carrier Extension:
    - extension bytes appended to (and removed from) the Ethernet frame by the physical layer
    - frame exists a longer period of time on the medium
  - Frame Bursting:
    - to minimize the extension bytes overhead, station may chain several frames together and transmit them at once ("burst").

Frame Bursting 1

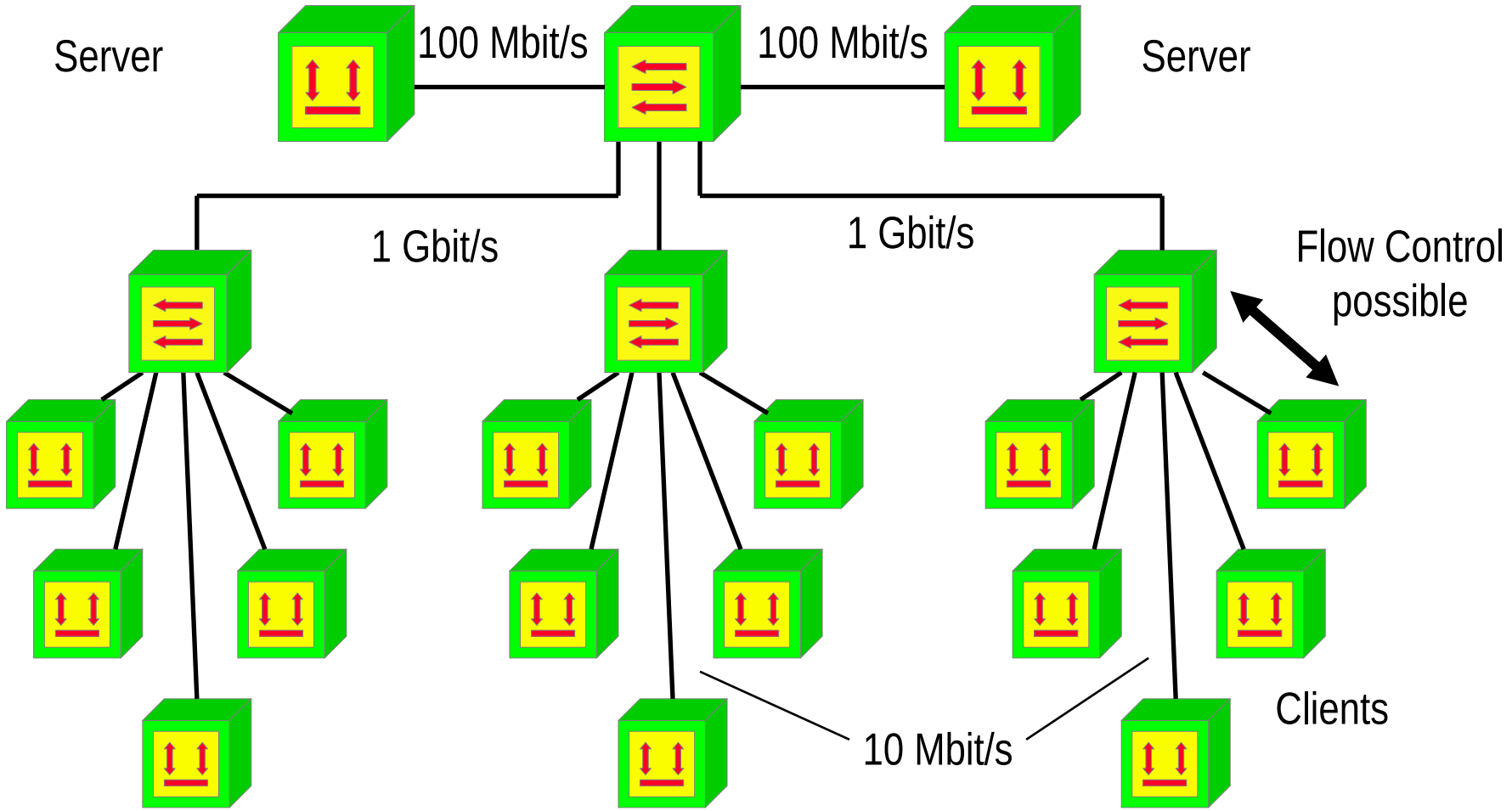
- **With both methods the minimal frame length is increased from 512 to 4096 bits**
  - = 512 bytes
  - The corresponding time is called slottime
- **If a station decides to chain several frames to a burst frame, the first frame inside the burst frame must have a length of at least 512 bytes**
  - By using extension bytes if necessary
- **The next frames (inside the burst frame) can have normal length (i.e. at least 64 bytes)**

Frame Bursting 2

- **Station may chain frames up to 8192 bytes (=burst limit)**
  - Also may finish the transmission of the last frame even beyond the burst limit
- **So the whole burst frame length must not exceed 8192+1518 bytes**
  - Incl. interframe gap of 0.096 µs = 12 bytes

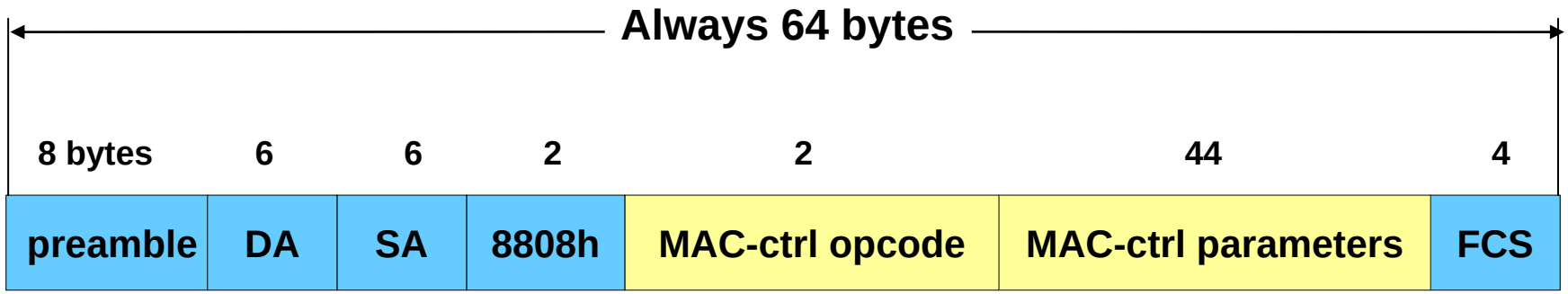


Ethernet Switching <-> Flow Control



MAC Control Frames

- **Additional functionality easily integrated**
- **Currently only Pause-Frame supported**



MAC-ctrl opcode ..... Defines function of control frame

MAC-ctrl parameters .... control parameter data (always filled up to 44 bytes)

The Pause Command 1

- **on receiving the pause command**
  - station stops sending normal frames for a given time which is specified in the MAC-control parameter field
- **this pause time is a multiple of the slot time**
  - 4096 bit-times when using Gigabit Ethernet or 512 bit-times with conventional 802.3
- **paused station waits**
  - until pause time expires or an additional MAC-control frame arrives with pause time = 0
  - note: paused stations are still allowed to send MAC-control-frames (to avoid blocking of LAN)

The Pause Command 2

- **destination address is either**
  - address of destination station or
  - broadcast address or
  - special multicast address 01-80-C2-00-00-01
- **this special multicast address prevents bridges to transfer associated pause-frames to not concerned network segments**
- **hence flow-control (with pause commands) affects only the own segment**

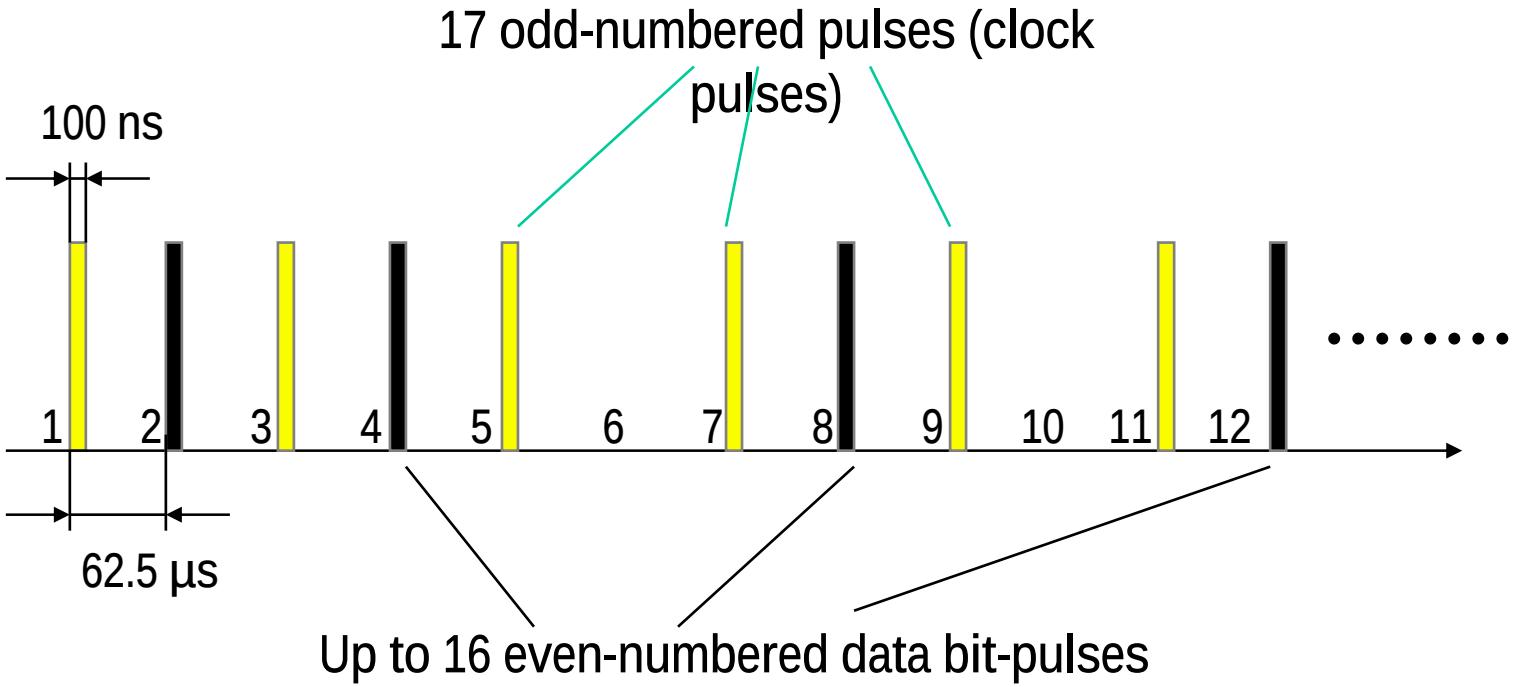
Auto Negotiation

- Enables each two Ethernet devices to exchange information about their capabilities
  - Signal rate, CSMA/CD, half- or full-duplex
- Using Link-Integrity-Test-Pulse-Sequence
  - Normal-Link-Pulse (NLP) technique is used in 10BaseT to check the link state (green LED)
  - 10 Mbit/s LAN devices send every 16.8 ms a 100ns lasting NLP, no signal on the wire means disconnected

Fast Link Pulses

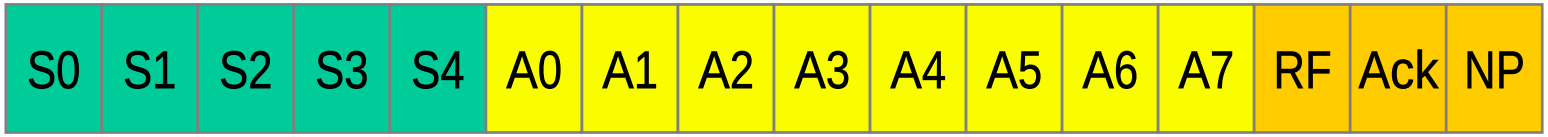
- Modern Ethernet NICs send bursts of Fast-Link-Pulses (FLP) consisting of 17-33 NLPs for Autonegotiation signalling
- Each representing a 16 bit word
  - GE sends several "pages"

FLP Burst Coding



= 1 1 0 1 0 1 ....

Base Page



Selector field

Technology ability field

provides selection of up to 32 different message types; currently only 2 selector codes available:

10000....IEEE 802.3

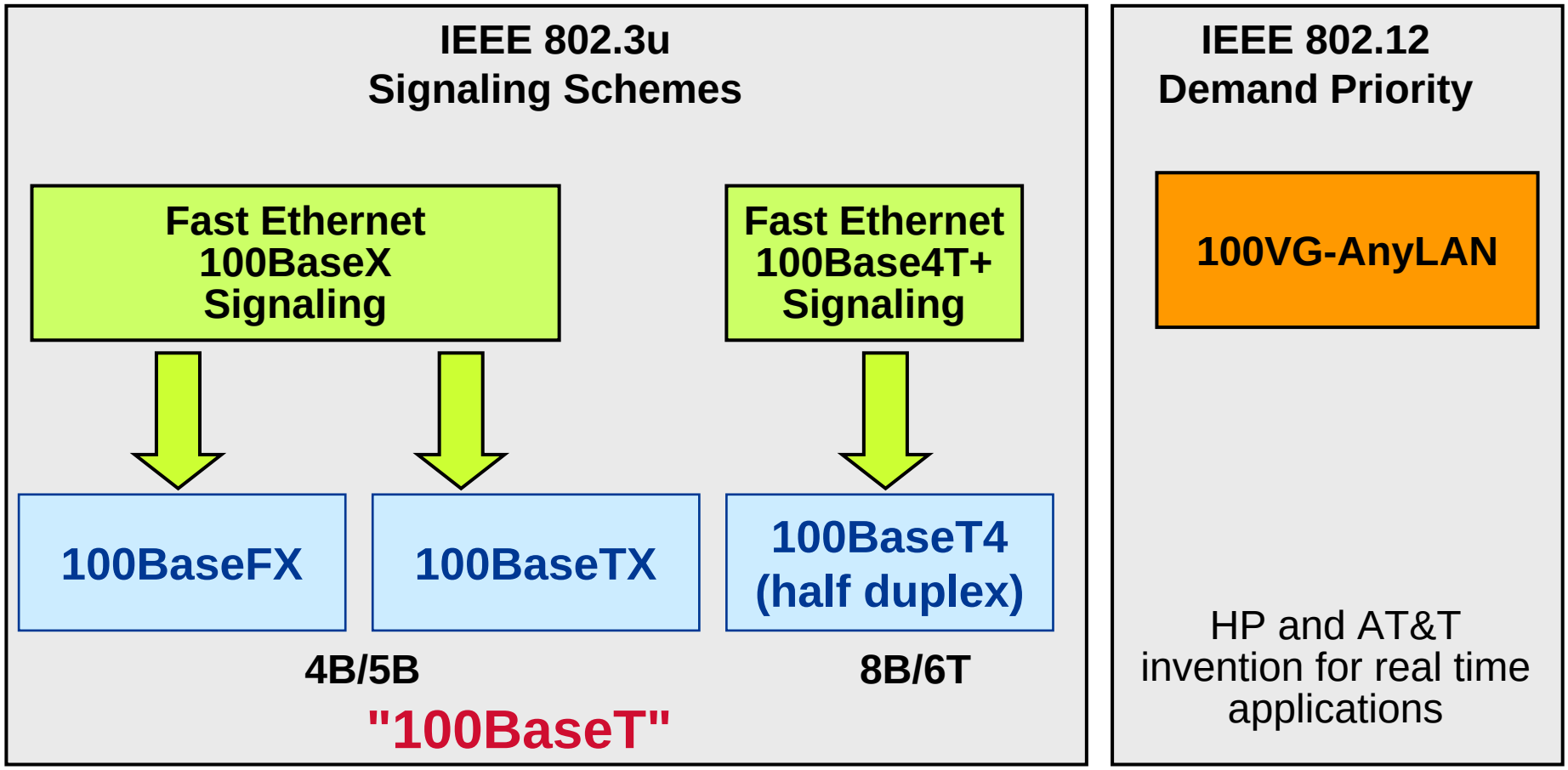
01000....IEEE 802.9

(ISLAN-16T)

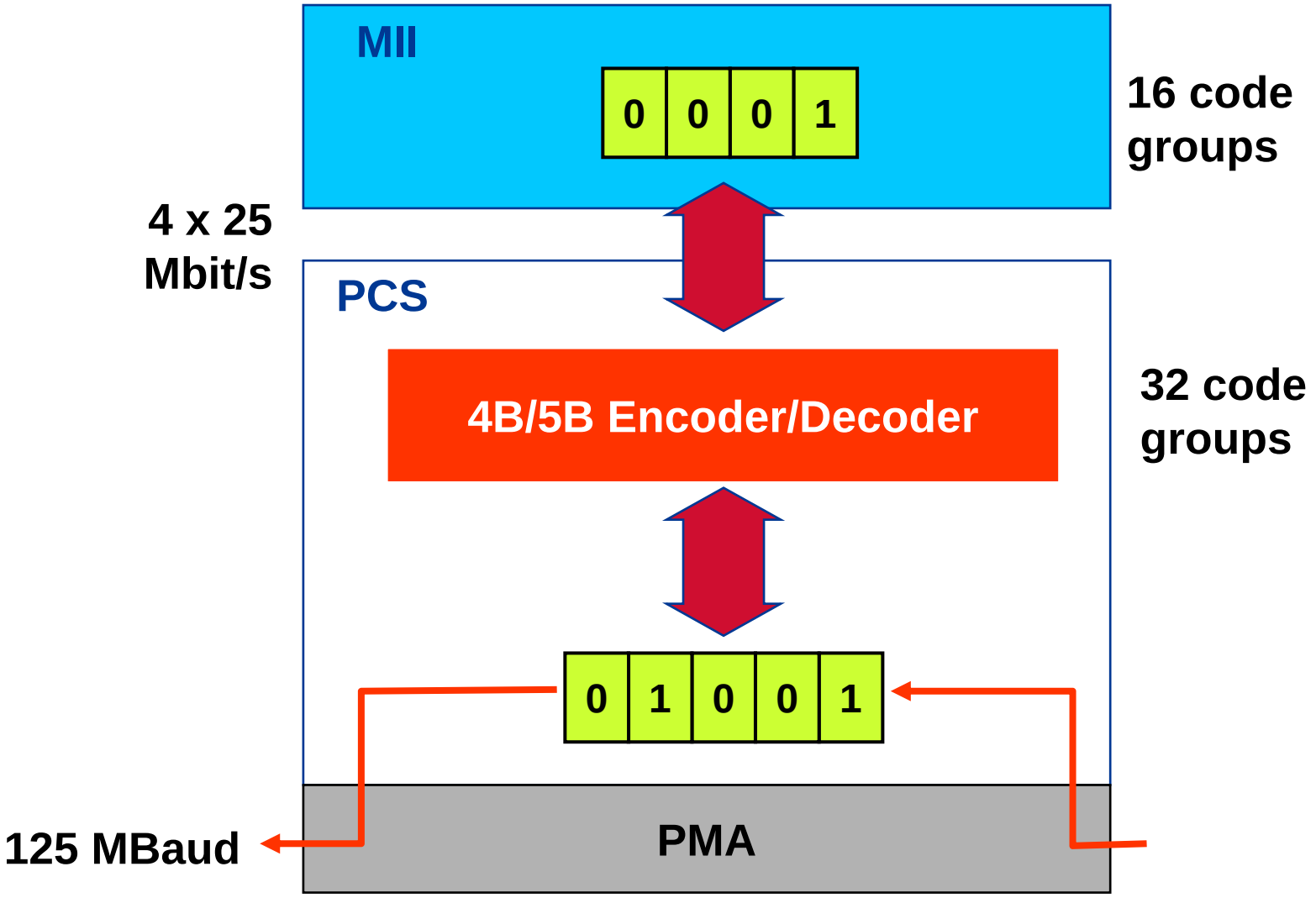
(ISO-Ethernet)

Bit	Technology
A0	10BaseT
A1	10BaseT-full duplex
A2	100BaseTx
A3	100BaseTx-full duplex
A4	100BaseT4
A5	Pause operation for full duplex links
A6	reserved
A7	reserved

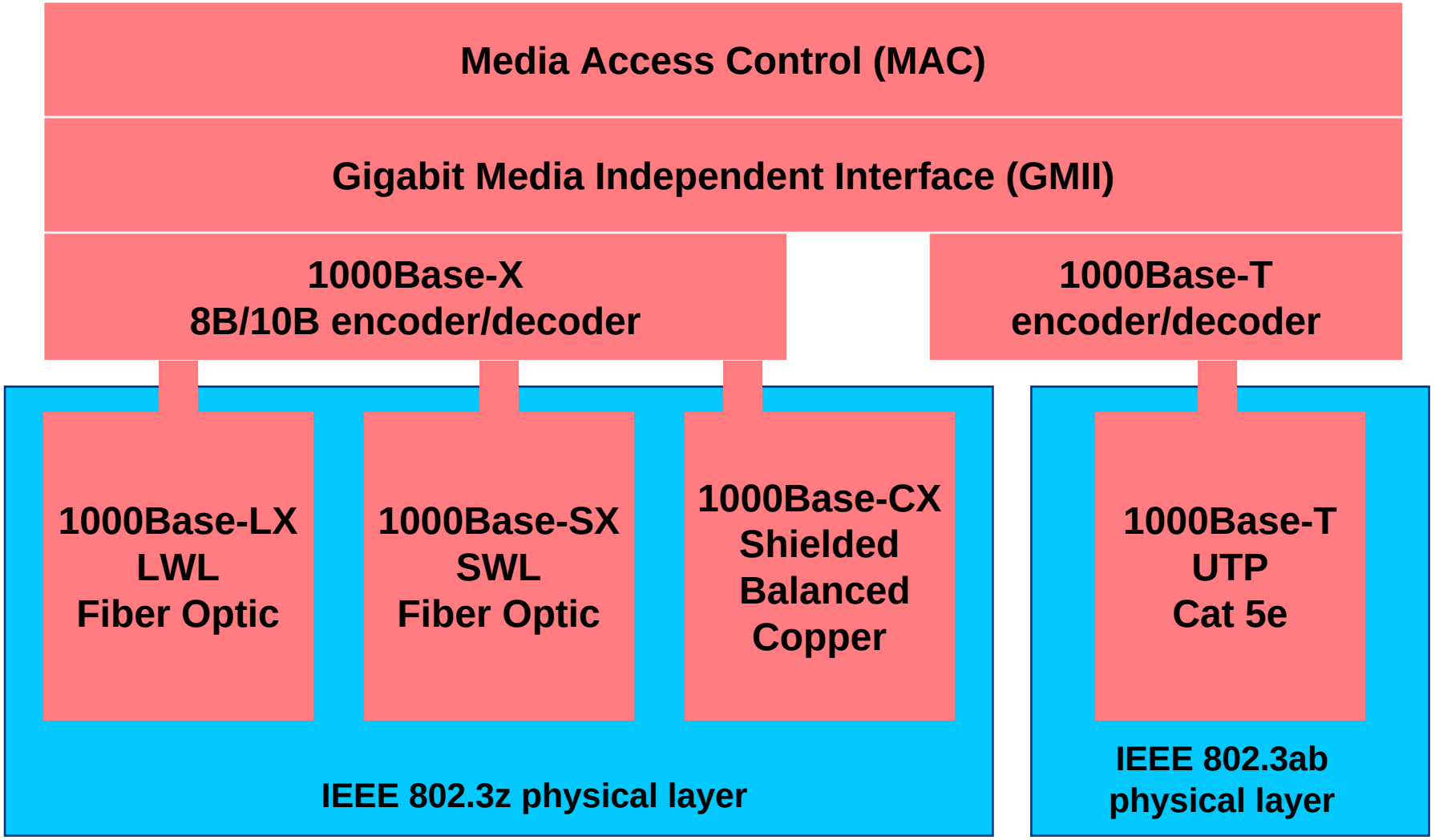
100 Mbit Ethernet Overview



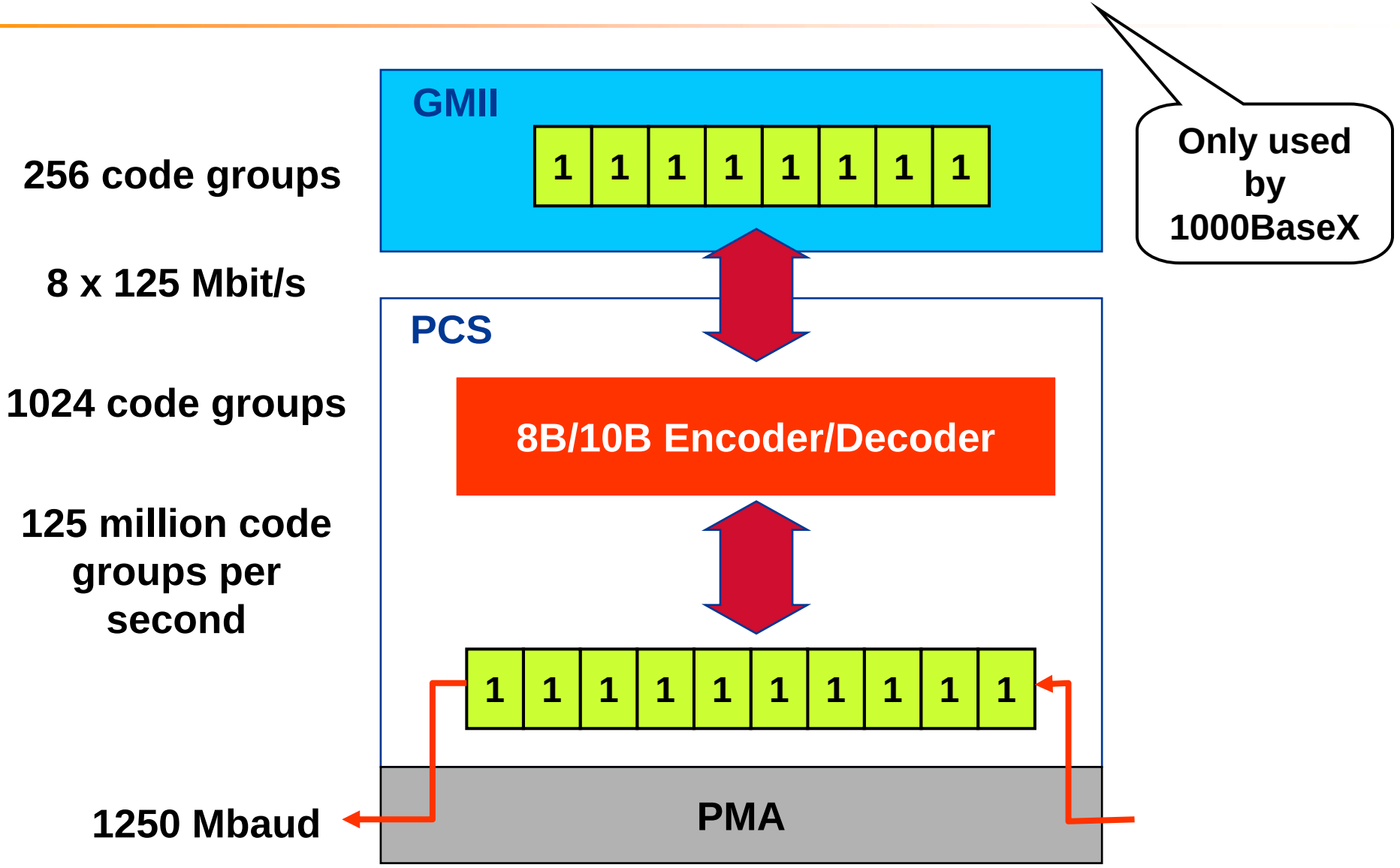
4B/5B Coding


MII
0 0 0 1
16 code groups
4 x 25 Mbit/s
PCS
4B/5B Encoder/Decoder
32 code groups
0 1 0 0 1
125 MBaud
PMA

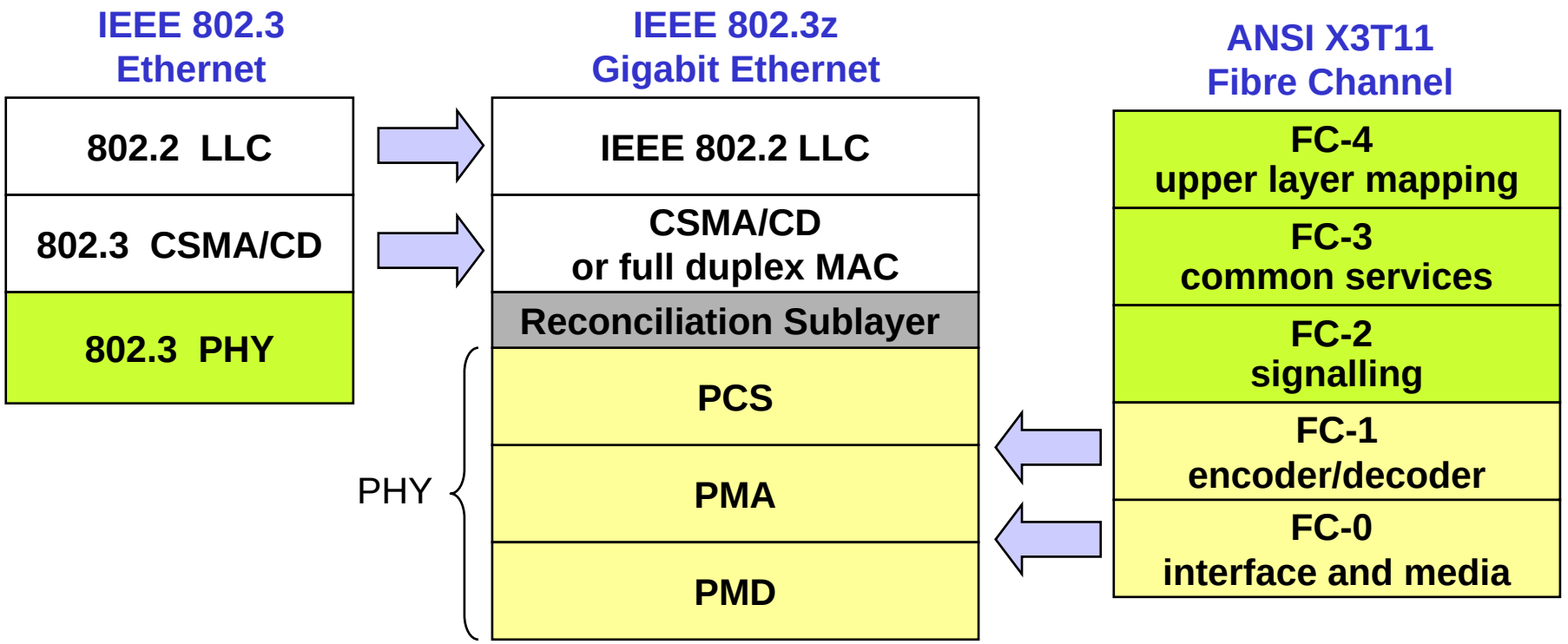
Gigabit Ethernet



GE 8B/10B Coding



GE Signaling



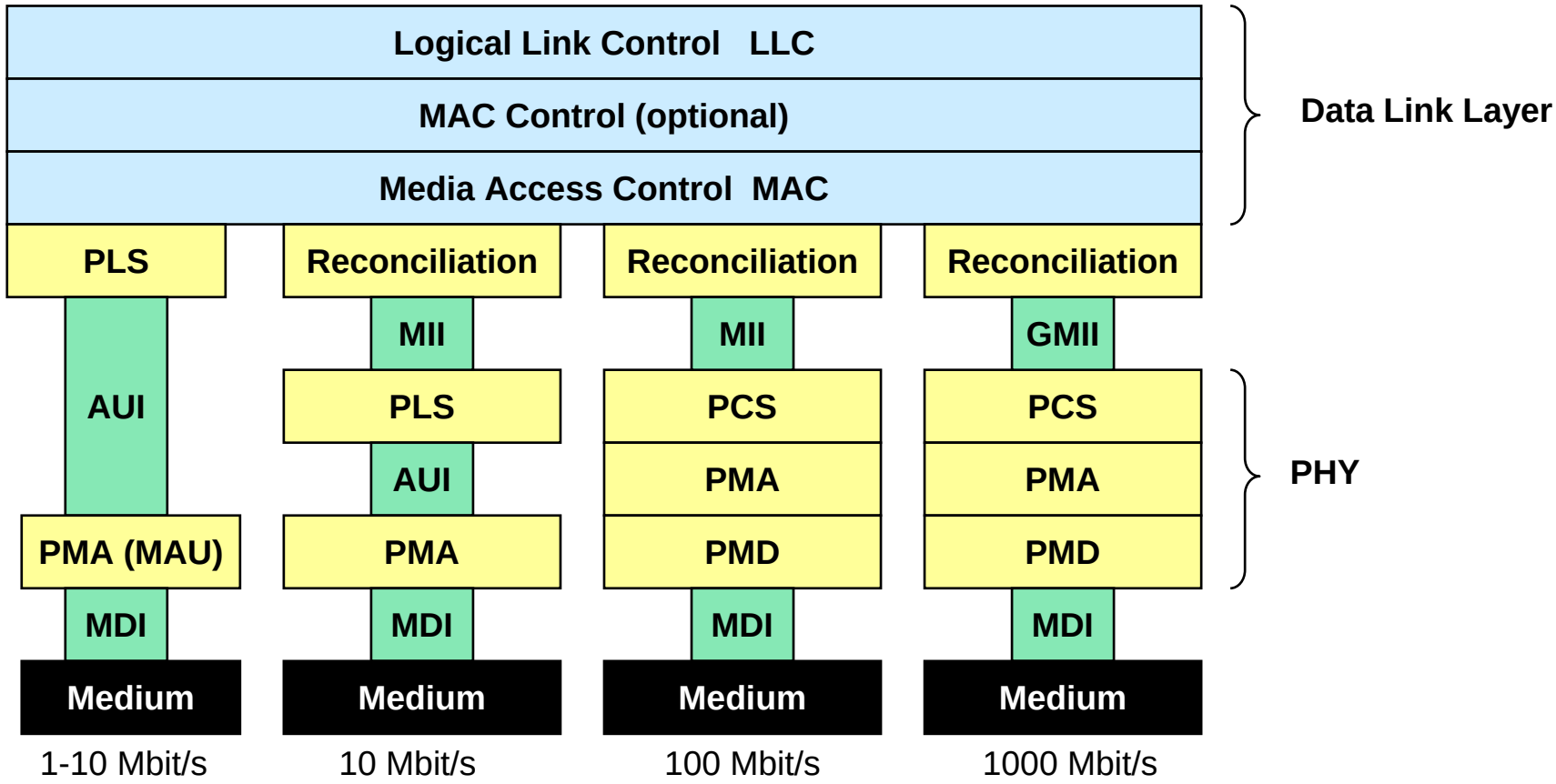
1000BaseX

- Two different wavelengths supported
- Full duplex only
  - 1000Base-SX: short wave, 850 nm MMF, up to 550m max. distance
  - 1000Base-LX: long wave, 1300 nm MMF or SMF, up to 5km max. distance
- 1000Base-CX:
  - Twinax Cable (high quality 150 Ohm balanced shielded copper cable)
  - About 25 m distance limit, DB-9 or the newer HSSDC connector

1000BaseT

- Defined by 802.3ab task force
- UTP
  - Uses all 4 line pairs simultaneously for duplex transmission! (echo cancellation)
  - 5 level PAM coding
    - 4 levels encode 2 bits + extra level used for Forward Error Correction (FEC)
  - Signal rate: 4 x 125 Mbaud = 4 x 250Mbit/s data rate
    - Cat. 5 links, max 100 m; all 4pairs, cable must conform to the requirements of ANSI/TIA/EIA-568-A
  - Only 1 CSMA/CD repeater allowed in a collision domain
  - up to 100m max. distance

Several Physical Media Supported



AUI Attachment Unit Interface, **PLS** Physical Layer Signaling, **MDI** Medium Dependent Interface **PCS** Physical Coding Sublayer, **MII** Media Independent Interface, **GMII** Gigabit Media Independent Interface, **PMA** Physical Medium Attachment, **MAU** Medium Attachment Unit, **PMD** Physical Medium Dependent

10 Gigabit Ethernet / IEEE 802.3ae

- Only optical support
  - 850nm (MM) / 1310nm /1550 nm (SM only)
  - No copper PHY anymore !
- Different implementations at the moment – standardization not finished!
- 8B/10B (IBM), SONET/SDH support, …
- XAUI ("Zowie") instead of GMII

10 Gigabit Ethernet (IEEE 802.3ae)

- **Preserves Ethernet framing**
- **Maintains the minimum and maximum frame size of the 802.3 standard**
- **Supports only full-duplex operation**
  - CSMA/CD protocol was dropped
- **Focus on defining the physical layer**
  - Four new optical interfaces (PMD)
    - To operate at various distances on both single-mode and multi-mode fibers
  - Two families of physical layer specifications (PHY) for LAN and WAN support
  - Properties of the PHY defined in corresponding PCS
    - Encoding and decoding functions

PMDs

- **10GBASE-L**
  - SM-fiber, 1300nm band, maximum distance 10km
- **10GBASE-E**
  - SM-fiber, 1550nm band, maximum distance 40km
- **10GBASE-S**
  - MM-fiber, 850nm band, maximum distance 26 – 82m
  - With laser-optimized MM up to 300m
- **10GBASE-LX4**
  - For SM- and MM-fiber, 1300nm
  - Array of four lasers each transmitting 3,125 Gbit/s and four receivers arranged in WDM (Wavelength-Division Multiplexing) fashion
  - Maximum distance 300m for legacy FDDI-grade MM-fiber
  - Maximum distance 10km for SM-fiber

WAN PHY / LAN PHY and their PCS

- **LAN-PHY**
  - 10GBASE-X
  - 10GBASE-R
    - 64B/66B coding running at 10,3125 Gbit/s
- **WAN-PHY**
  - 10GBASE-W
    - 64B/66B encoded payload into SONET concatenated STS192c frame running at 9,953 Gbit/s
    - Adaptation of 10Gbit/s to run over traditional SDH links

IEEE 802.3ae PMDs, PHYs, PCSs

		PCS		
PMD	10GBASE-E	10GBASE-ER		10GBASE-EW
	10GBASE-L	10GBASE-LR		10GBASE-LW
	10GBASE-S	10GBASE-SR		10GBASE-SW
	10GBASE-L4		10GBASE-LX4	
		LAN PHY		WAN PHY

10 Gigabit Ethernet over Copper

- **IEEE 802.3ak defined in 2004**
  - 10GBASE-CX4
  - Four pairs of twin-axial copper wiring with IBX4 connector
  - Maximum distance of 15m
- **IEEE 802.3an working group**
  - 10GBASE-T
  - CAT6 UTP cabling with maximum distance of 55m to 100m
  - CAT7 cabling with maximum distance of 100m
  - Standard ratification expected in July 2006

Note

- GE and 10GE use synchronous physical sublayer !!!
- Recommendation: Don't use GE over copper wires
  - Radiation/EMI
  - Grounding problems
  - High BER
  - Thick cable bundles (especially Cat-7)

Summary

- Ethernet evolved in the opposite direction:
  - Collision free
  - WAN qualified
  - Switched
- Several coding styles ➔ Complex PHY architecture
- Plug & play through autonegotiation
- Much simpler than ATM but no BISDN solution – might change!

Quiz

- Why tends high-speed Ethernet to synchronous PHY?
- Can I attach a 100 Mbit/s port to a 1000 Mbit/s port via fiber?
- What is the idea of Etherchannels? (Maximum bit rate, difference to multiple parallel links)