



第16章 高速局域网

南京大学计算机系 黄皓教授

2007年9月11日 星期二

|

2007年9月14日 星期五



参考文献

1. Rich Seifert, 千兆以太网-技术与应用, 机械工业出版社, 2000年。
2. 敖志刚, 万兆位以太网及其实用技术, 电子工业出版社, 2007年7月。



Contents

- Introduction
- High Speed LANs
- Ethernet
- Token Ring
- Fibre Channel



16. 0 Introduction

- Range of technologies
 - Fast and Gigabit Ethernet
 - Fibre Channel
 - High Speed Wireless LANs



16.1 High Speed LANs — Why High Speed LANs?

- Office LANs used to provide basic connectivity
 - Connecting PCs and terminals to mainframes and midrange systems that ran corporate applications
 - Providing workgroup connectivity at departmental level
 - Traffic patterns light
 - Emphasis on file transfer and electronic mail
- Speed and power of PCs has risen
 - Graphics-intensive applications and GUIs
- MIS organizations recognize LANs as essential
 - Began with client/server computing
 - Now dominant architecture in business environment
 - Intranetworks

Frequent transfer of large volumes of data



16.1 High Speed LANs — Applications Requiring High Speed LANs

- Centralized server farms
 - User needs to draw huge amounts of data from multiple centralized servers
 - E.g. Color publishing
 - Servers contain tens of gigabytes of image data
 - Downloaded to imaging workstations
- Power workgroups
 - Small number of cooperating users
 - Draw massive data files among workstations
 - E.g. Software development group testing new software version or computer-aided design (CAD) running simulations
- High-speed local backbone
 - Processing demand grows
 - LANs proliferate at site
 - **High-speed interconnection** is necessary



16.2 Ethernet

- ALOHA
- Slotted-ALOHA
- CSMA
 - ☐ Nonpersistent
 - ☐ 1-persistent
 - ☐ p-persistent
- CSMA/CD

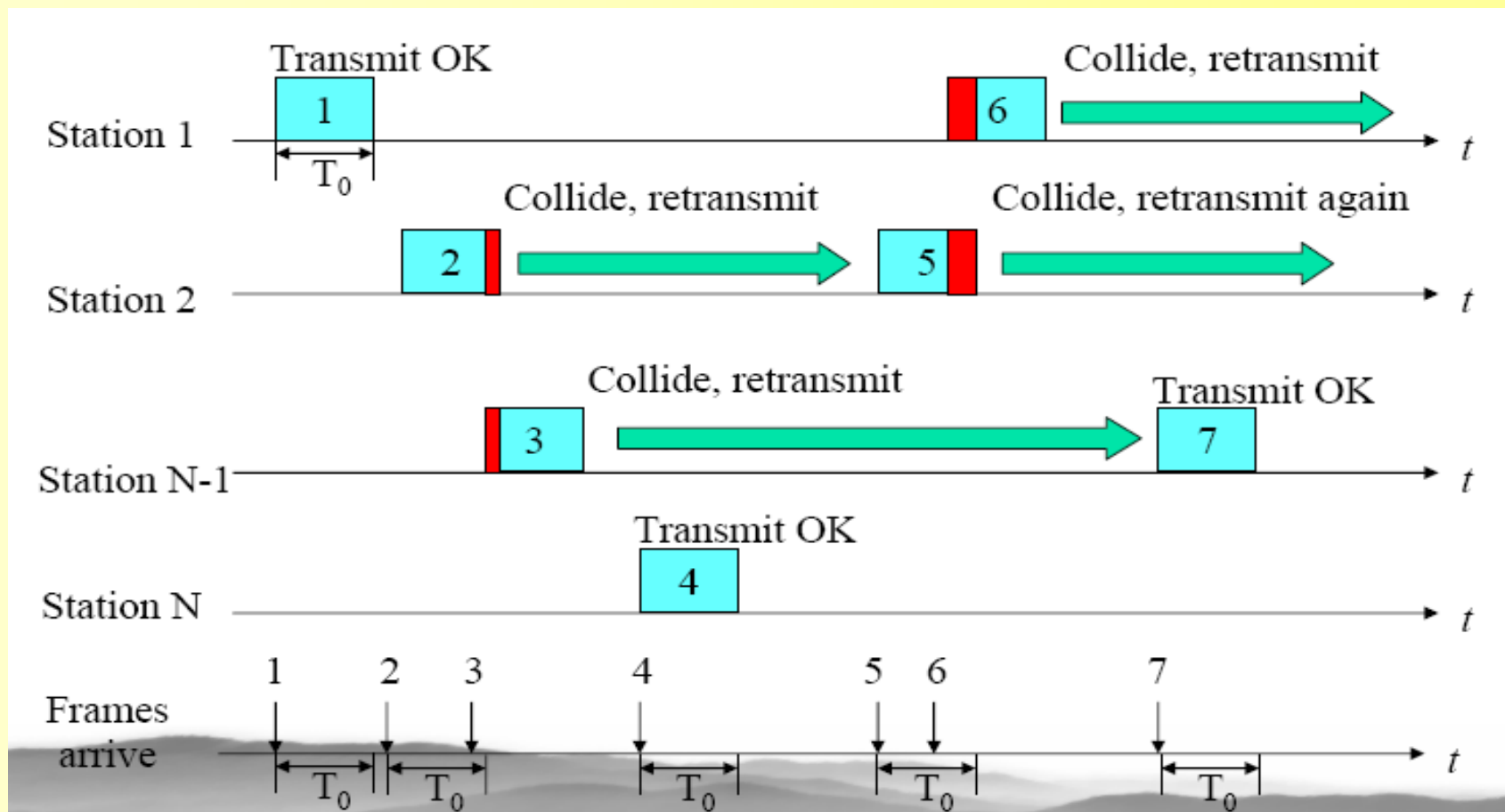


CSMA/CD — ALOHA

- Sender
 - ☐ When station has frame, it sends
 - ☐ Station listens (for max round trip time) plus small increment
 - ☐ If ACK, fine. If not, retransmit
 - ☐ If no ACK after repeated transmissions, give up
- Receiver
 - ☐ Use frame check sequence (as in HDLC)
 - ☐ If frame OK and address matches receiver, send ACK
 - ☐ Otherwise, the receiver just ignore the frame.
- Frame may be damaged by noise or collision
 - ☐ Another station transmitting at the same time
- Max utilization 18%



Illustration of ALOHA



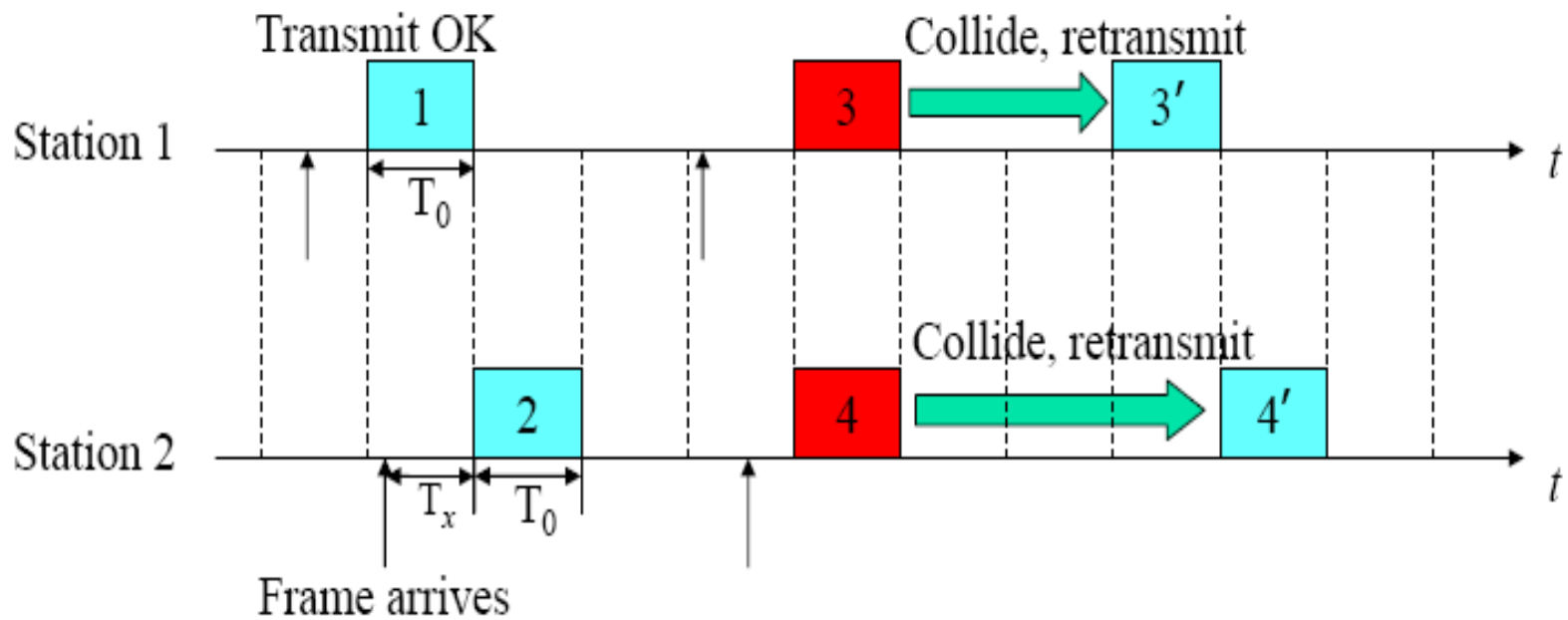


Slotted ALOHA

- Time in uniform slots equal to frame transmission time
- Need central clock (or other sync mechanism)
- Transmission begins at slot boundary
- Frames either miss or overlap totally
- Max utilization 37%



16.2 Ethernet





- Suppose
 - Propagation time is much less than transmission time
 - All stations know that a transmission has started almost immediately
- Method
 - First listen for clear medium (carrier sense)
 - If medium idle, transmit
 - Wait reasonable time (round trip plus ACK contention)
 - No ACK then retransmit
- If two stations start at the same instant, collision
- Max utilization depends on propagation time (medium length) vs. frame length
 - Longer frame and shorter propagation gives better utilization

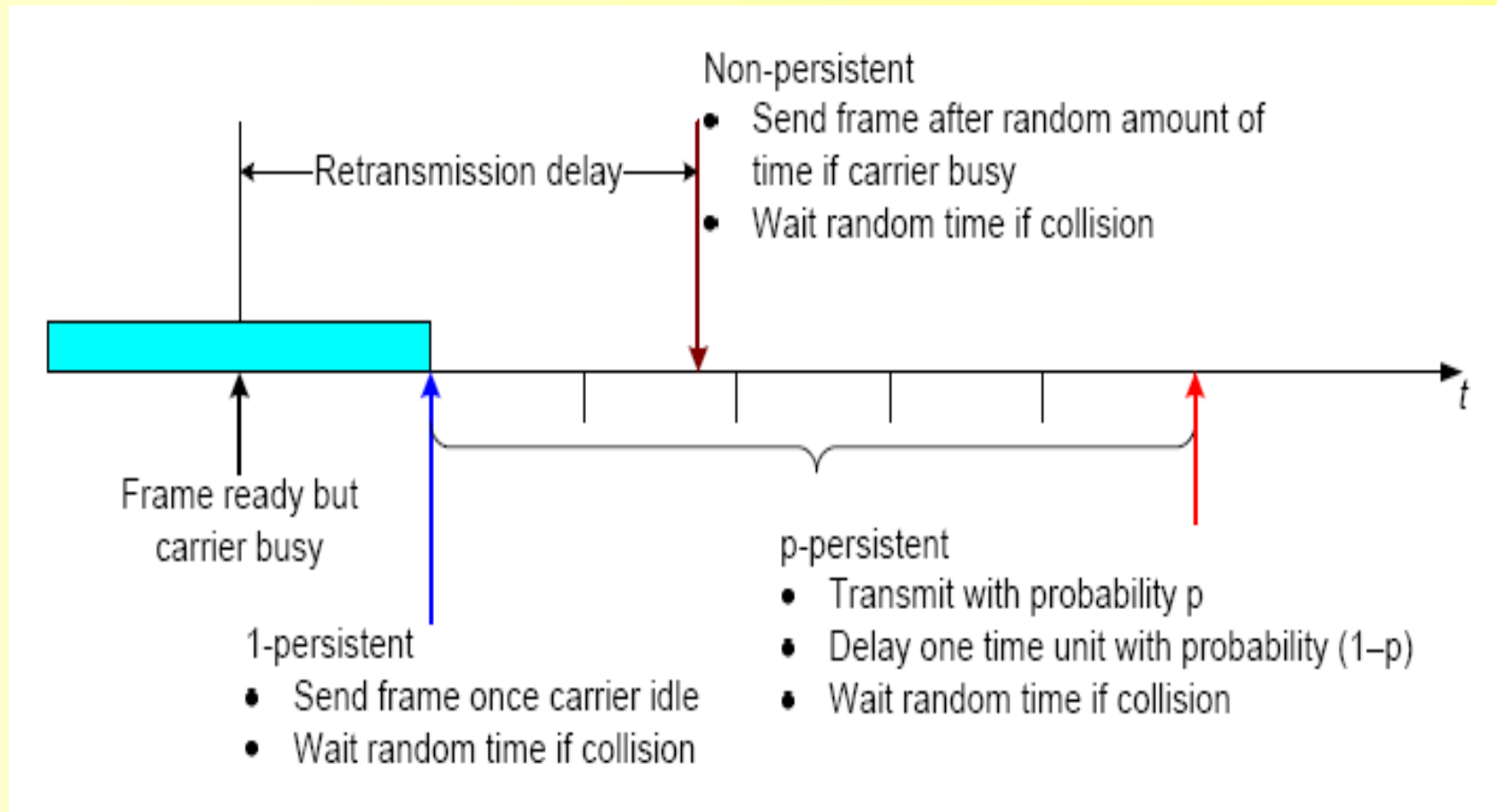


Nonpersistent CSMA

- Station wishing to transmit listens
 - 1. If medium is idle, transmit; otherwise, go to 2
 - 2. If medium is busy, wait amount of time drawn from probability distribution (retransmission delay) and repeat 1
- Random delays reduces probability of collisions
 - Consider two stations become ready to transmit at same time while another transmission is in progress
 - If both stations delay same time before retrying, both will attempt to transmit at same time
- Capacity is wasted because medium will remain idle following end of transmission
 - Even if one or more stations waiting
- Nonpersistent stations are deferential



Nonpersistent CSMA





1-persistent CSMA

To avoid idle channel time, 1-persistent protocol used

- A station wishing to transmit listens and obeys following:
 1. If medium idle, transmit; otherwise, go to step 2
 2. If medium busy, listen until idle; then transmit immediately
- 1-persistent stations selfish
- If two or more stations waiting, collision guaranteed
 - Gets sorted out after collision



P-persistent CSMA

- Compromise that attempts to reduce collisions Like nonpersistent
- And reduce idle time Like 1-persistent
- Rules:
 1. If medium idle, transmit with probability p , and delay one time unit with probability $(1 - p)$
 - Time unit typically maximum propagation delay
 2. If medium busy, listen until idle and repeat step 1
 3. If transmission is delayed one time unit, repeat step 1
- What is an effective value of p ?



Value of p ? (1)

- Avoid instability under heavy load
- n stations waiting to send
- End of transmission, expected number of stations attempting to transmit is number of stations ready times probability of transmitting
 - np
- If $np > 1$, on average, there will be a collision
- Repeated attempts to transmit almost guaranteeing more collisions
- Retries compete with new transmissions
- Eventually, all stations trying to send
 - Continuous collisions; zero throughput



Value of p ? (2)

- So $np < 1$ for expected peaks of n
- If heavy load expected, p small
- However, as p made smaller, stations wait longer
- At low loads, this gives very long delays

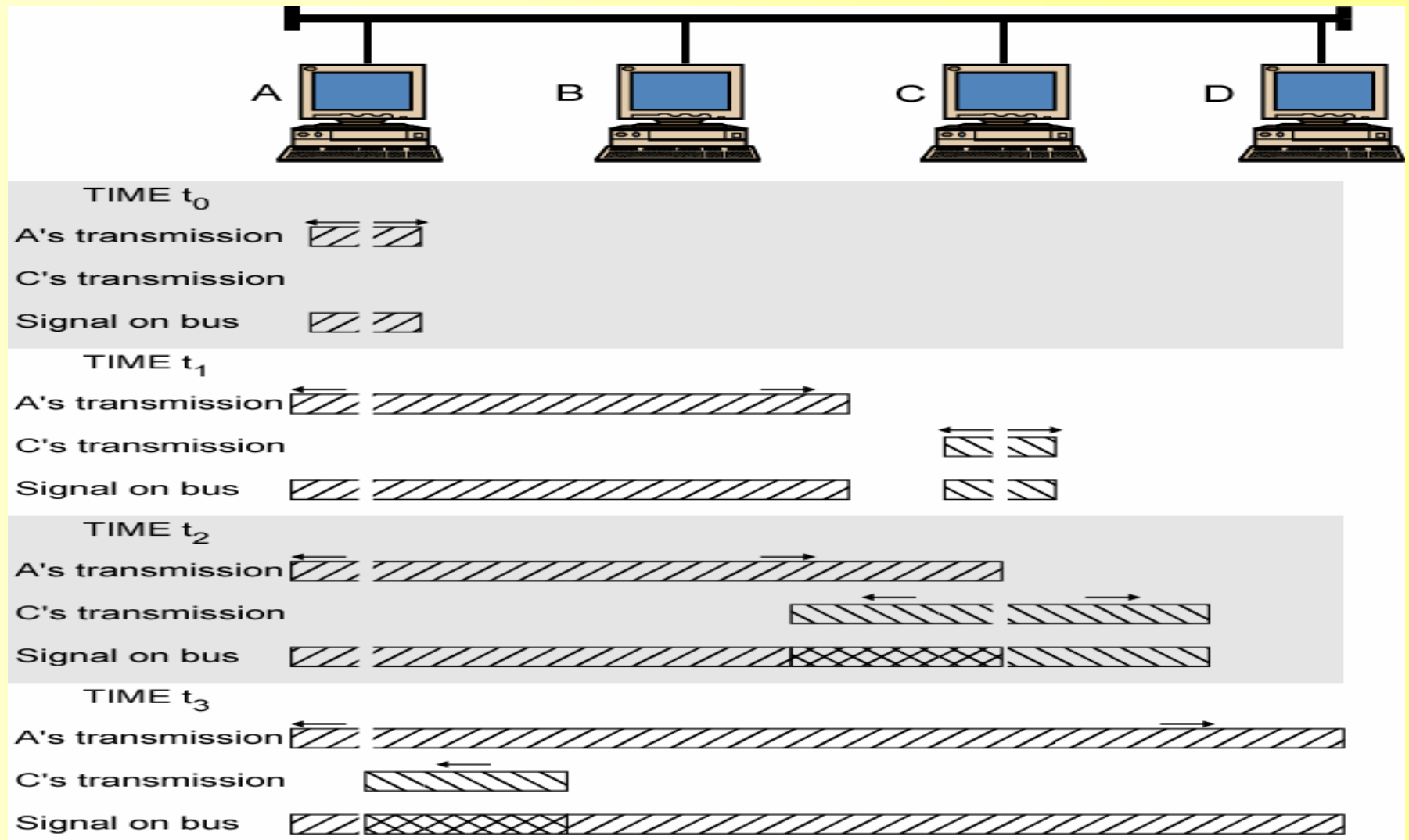


CSMA/CD

- With CSMA, collision occupies medium for duration of transmission
 - Stations listen whilst transmitting
1. If medium idle, transmit, otherwise, step 2;
 2. If busy, listen for idle, then transmit;
 3. If collision detected, jam then cease transmission;
 4. After jam, wait random time (backoff) then start from step 1;



CSMA/CD operation





Which Persistence Algorithm?

- IEEE 802.3 uses 1-persistent
- Both nonpersistent and p-persistent have performance problems
- 1-persistent ($p = 1$) seems more unstable than p-persistent
 - Greed of the stations
 - But wasted time due to collisions is short (if frames long relative to propagation delay)
 - With random backoff, unlikely to collide on next tries
 - To ensure backoff maintains stability, IEEE 802.3 and Ethernet use binary exponential backoff



Binary Exponential Backoff

- Attempt to transmit repeatedly if repeated collisions
- First 10 attempts, mean value of random delay doubled
- Value then remains same for 6 further attempts
- After 16 unsuccessful attempts, station gives up and reports error
- As congestion increases, stations back off by larger amounts to reduce the probability of collision.
- 1-persistent algorithm with binary exponential backoff efficient over wide range of loads
 - Low loads, 1-persistence guarantees station can seize channel once idle
 - High loads, at least as stable as other techniques
- Backoff algorithm gives last-in, first-out effect
- Stations with few collisions transmit first

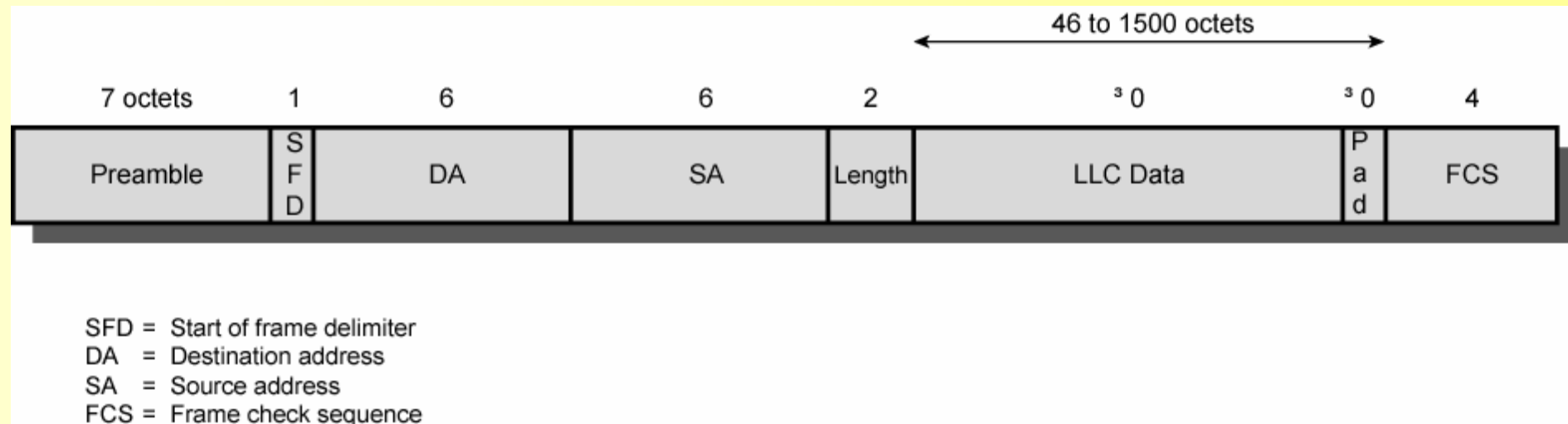


Collision Detection

- On baseband bus, collision produces much higher signal voltage than signal
- Collision detected if cable signal greater than single station signal
- Signal attenuated over distance
- Limit distance to 500m (10Base5) or 200m (10Base2)
- For twisted pair (star-topology) activity on more than one port is collision
- A special collision presence signal is generated



IEEE 802.3 Frame Format





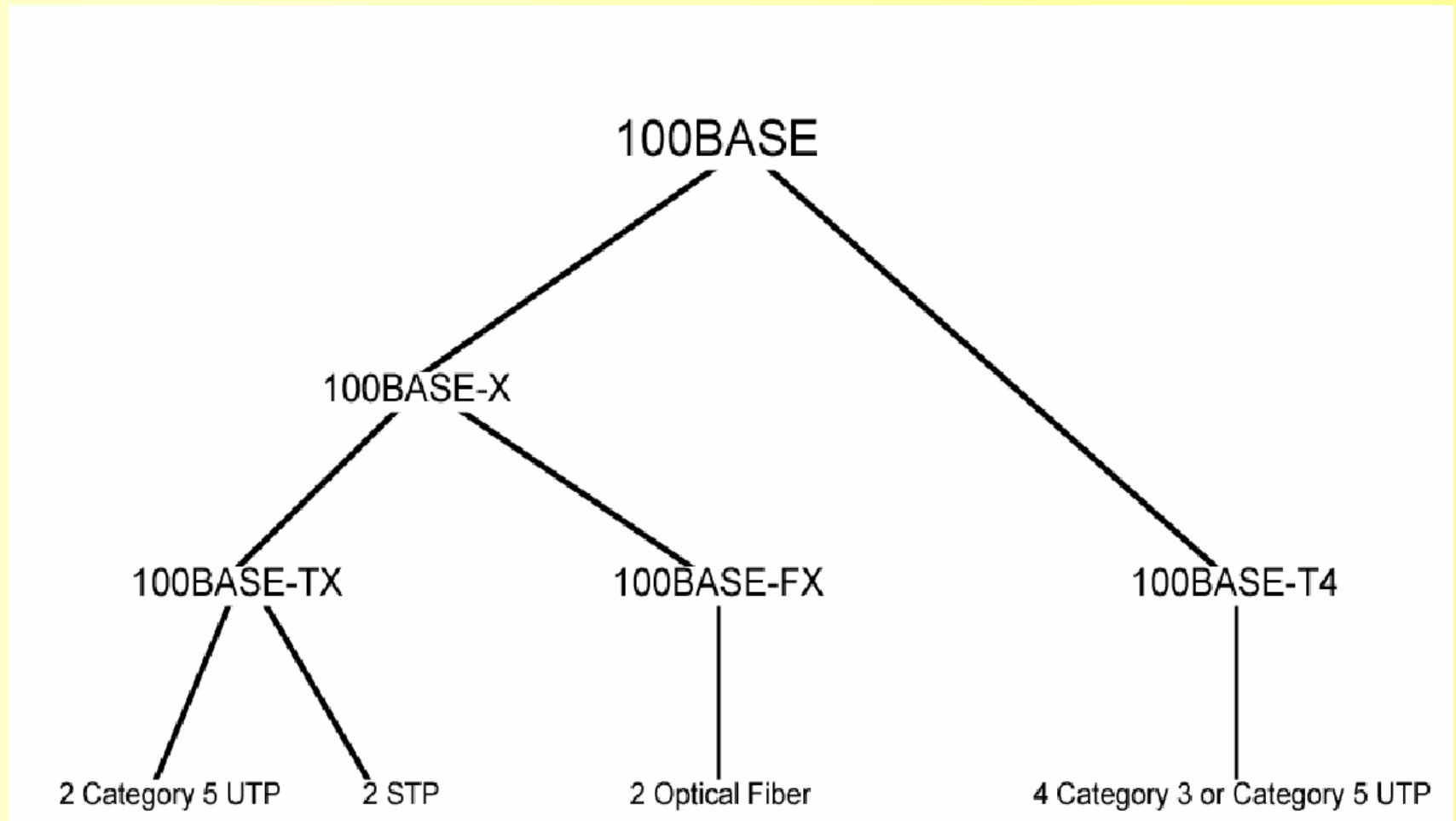
10Mbps Specification (Ethernet)

<data rate><Signaling method><Medium>

	10Base5	10Base2	10Base-T	10Base-F
Medium	Coaxial	Coaxial	UTP	850nm fiber
Signaling	Baseband	Baseband	Baseband	Manchester
	Manchester	Manchester	On/Off	
Topology	Bus	Bus	Star	Star
Nodes	100	30	-	33



Fast Ethernet 100BASE-T Options





- IEEE专门成立了快速以太网研究组评估以太网传输速率提升到100 Mbps的可行性。
- 100Base-T和100VG—AnyLAN(适用于令牌环网)。



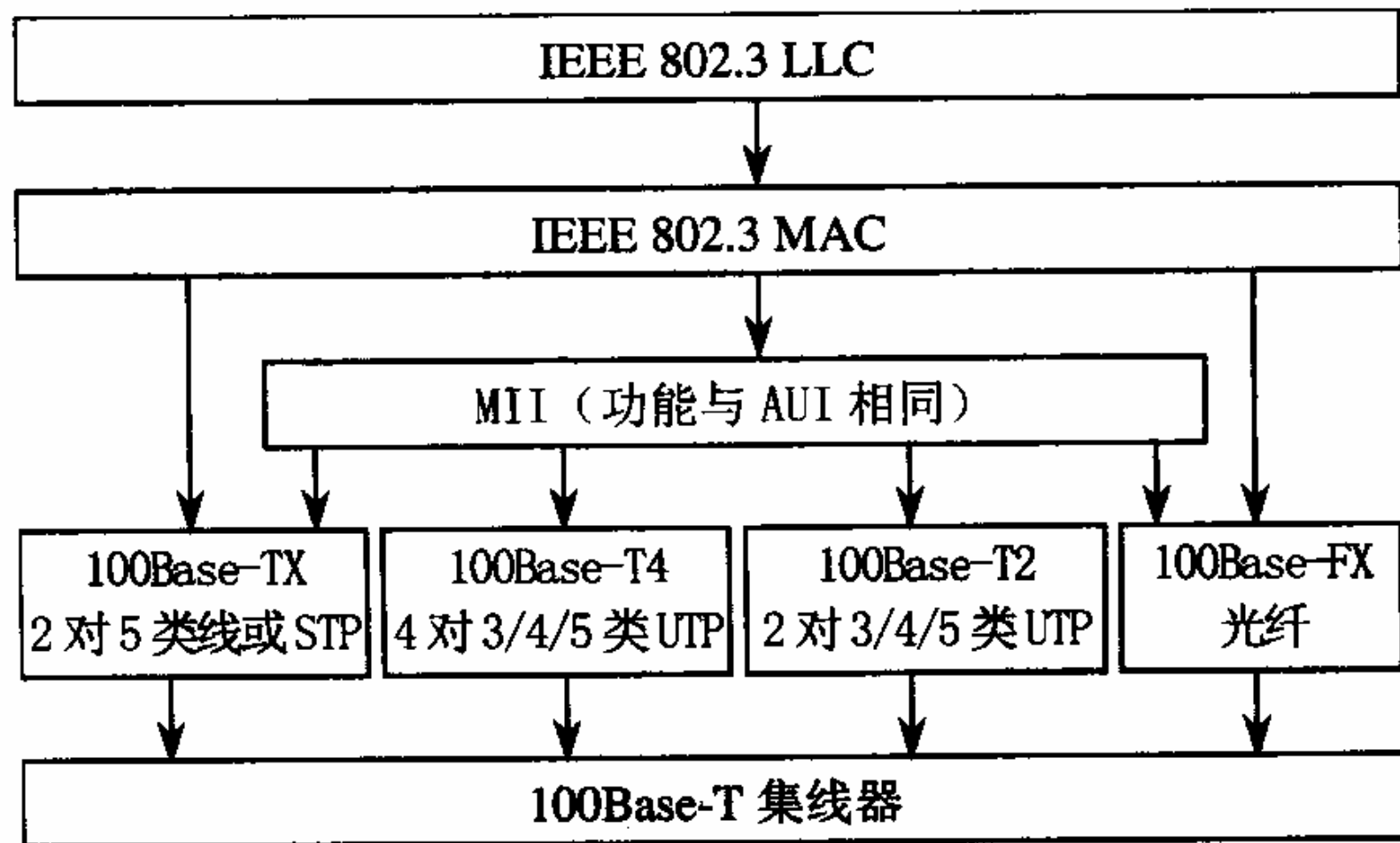
快速以太网特点

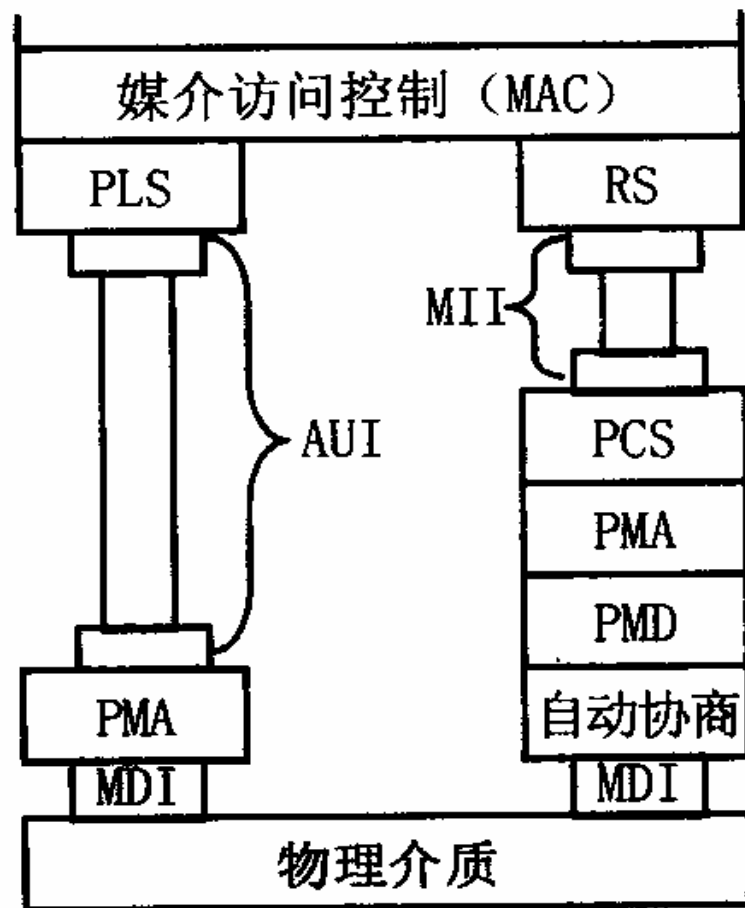
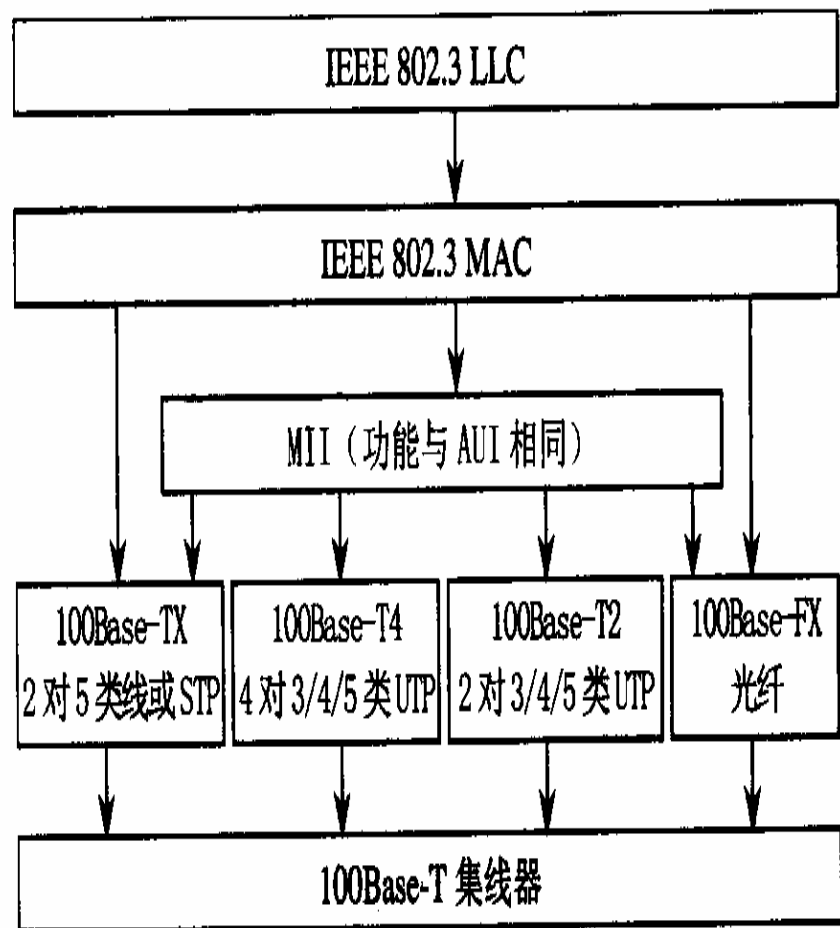
- 采用所有一般以太网做介质，从而保护了现有网络投资。
- 采用现在流行的SNMP的网管软件和以太网管理信息库(MIB, Management Information Base)，所以完全兼容于现有的网管产品。
- 由于采用CSMA/CD协议，可与10Base-T并行工作，避免了协议转换造成的系统开销，因此效率更高。
- 快速以太网还提供全双工通信，总带宽达到200 Mbps。
- 快速以太网有自动协商的功能，能够自动适应电缆两端最高可用的通信速率，能方便地与10 Mbps以太网连接通信。



快速以太网保留了传统以太网的所有特性

- **相同的数据帧格式。**快速以太网的IEEE 802.3u标准在LLC子层使用IEEE 802.2标准
- **相同的介质访问控制方式。**在MAC子层使用CSMA/CD方法，
- **相同的组网方法。**
- 将每个比特的发送时间由100 ns降低到10 ns。
- 在物理层做了一些必要的调整，定义了新的物理层标准(100Base-T)。
- 定义Medium Independent Interface(MII)。将MAC子层和物理层分开。







增加了自动协商功能(Auto Negotiate)

- 快速以太网被设计成能够同时支持10 Mbps和100 Mbps两种工作速度。
- 自动协商的内容
 - 辨识双方是10Mbps、100Mbps还是双速设备
 - 确认双方模式是半双工、全双工还是支持两种模式
 - 确认双方的物理层规范类型



100BASE-X Media

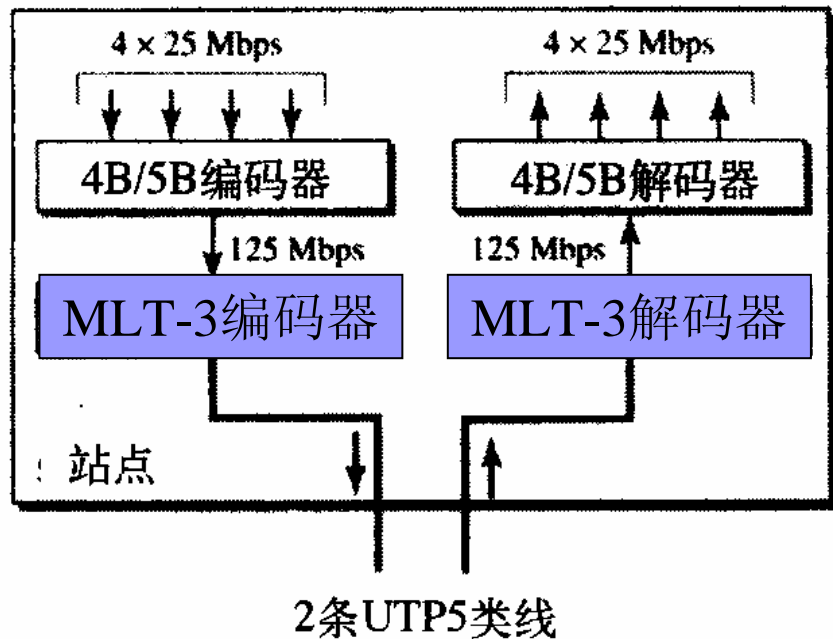
- Two physical medium specifications
- 100BASE-TX
 - Two pairs of twisted-pair cable
 - One pair for transmission and one for reception
 - STP and Category 5 UTP allowed
 - The MTL-3 signaling scheme is used
- 100BASE-FX
 - Two optical fiber cables
 - One for transmission and one for reception
 - Intensity modulation used to convert 4B/5B-NRZI code group stream into optical signals
 - 1 represented by pulse of light
 - 0 by either absence of pulse or very low intensity pulse



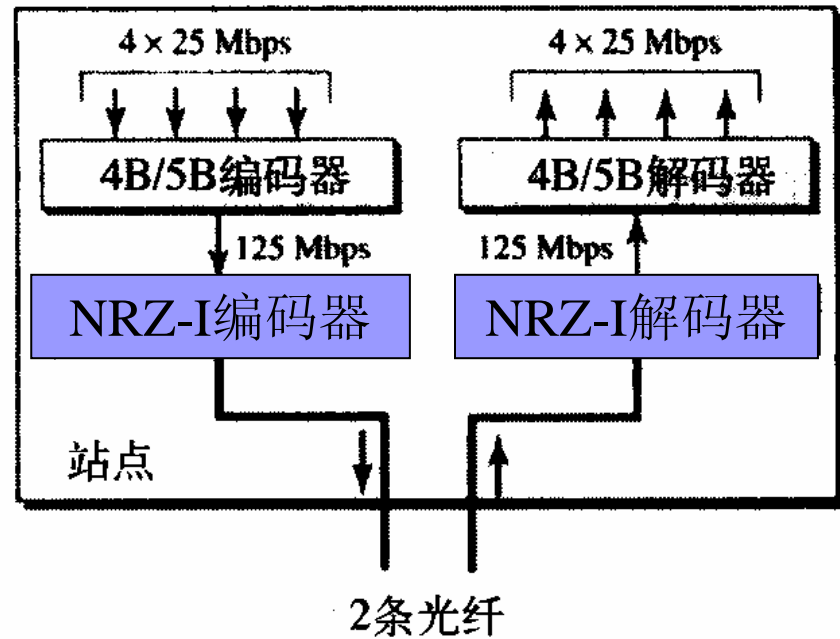
100BASE-T4

- 100-Mbps over lower-quality Cat 3 UTP
 - Taking advantage of large installed base
 - Cat 5 optional
 - Does not transmit continuous signal between packets
 - Useful in battery-powered applications
- Can not get 100 Mbps on single twisted pair
 - Data stream split into three separate streams
 - Each with an effective data rate of 33.33 Mbps
 - Four twisted pairs used
 - Data transmitted and received using three pairs
 - Two pairs configured for bidirectional transmission
- NRZ encoding not used
 - Would require signaling rate of 33 Mbps on each pair
 - Does not provide synchronization
 - Ternary signaling scheme (8B6T)

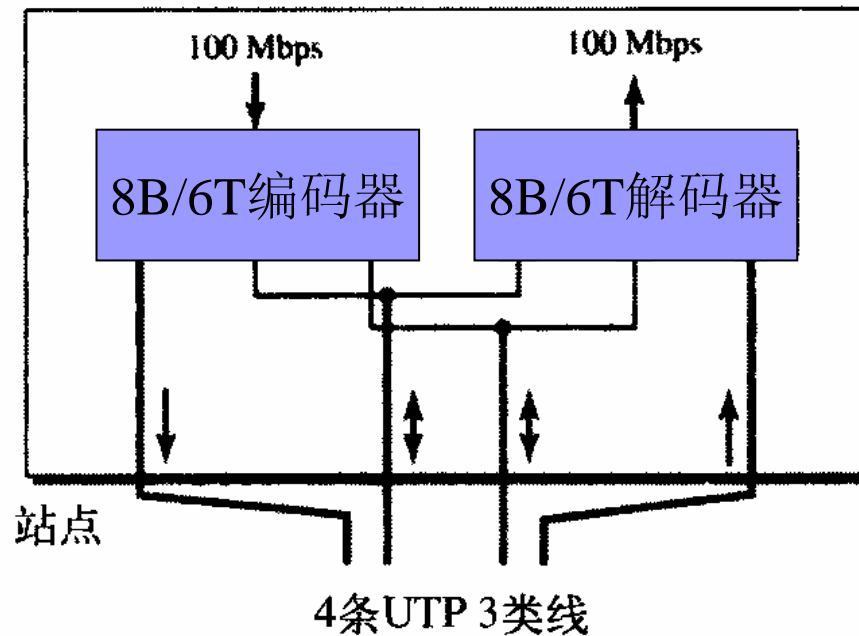
100Base-TX



100Base-FX



100Base-T4





Full Duplex Operation

- Traditional Ethernet half duplex
 - Either transmit or receive but not both simultaneously
- With full-duplex, station can transmit and receive simultaneously
- 100-Mbps Ethernet in full-duplex mode, theoretical transfer rate 200 Mbps
- Attached stations must have full-duplex adapter cards
- Must use switching hub
 - Each station constitutes separate collision domain
 - In fact, no collisions
 - CSMA/CD algorithm no longer needed
 - 802.3 MAC frame format used
 - Attached stations can continue CSMA/CD

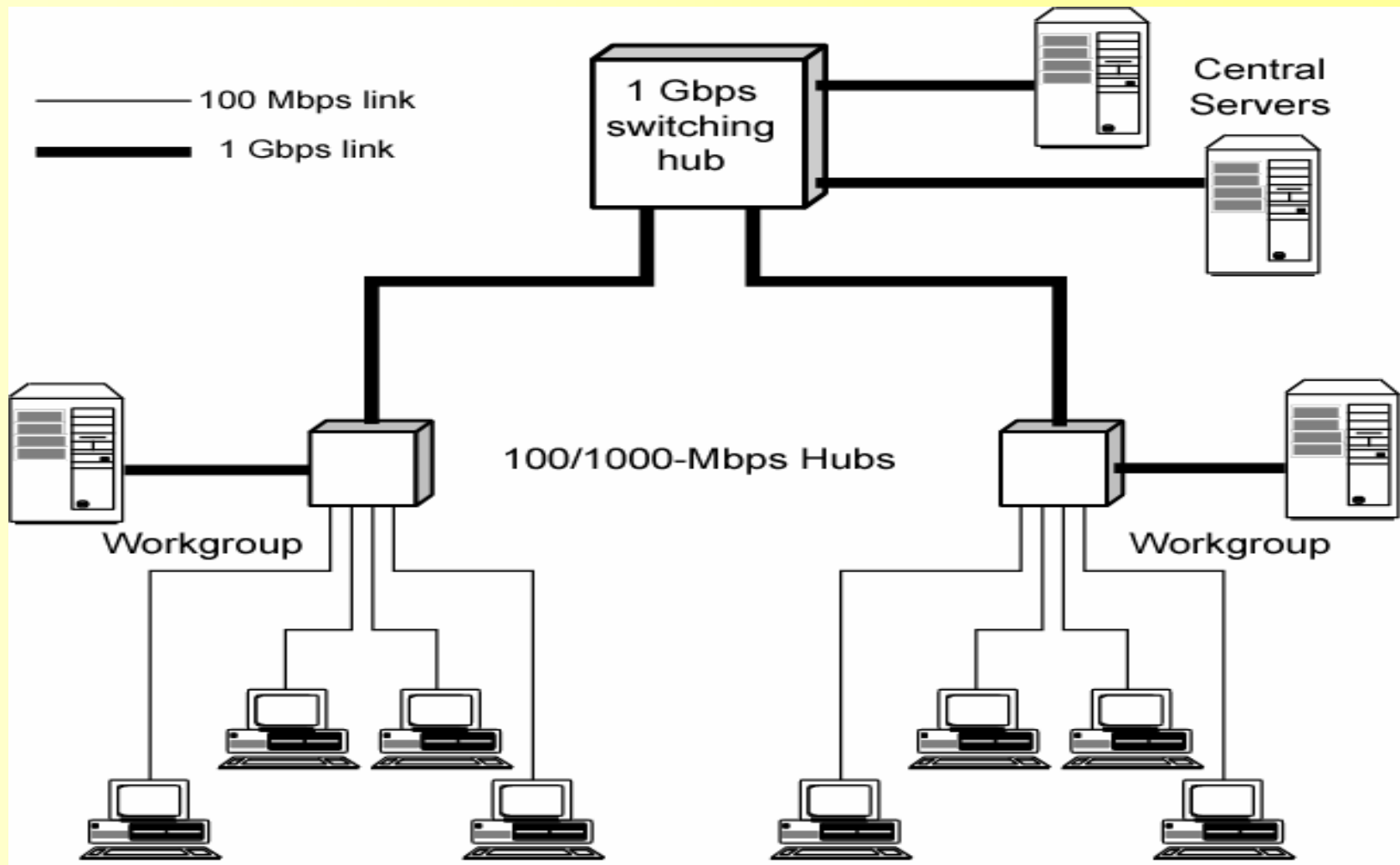


Mixed Configurations

- Fast Ethernet supports mixture of existing 10-Mbps LANs and newer 100-Mbps LANs
- E.g. 100-Mbps backbone LAN to support 10-Mbps hubs
 - Stations attach to 10-Mbps hubs using 10BASE-T
 - Hubs connected to switching hubs using 100BASE-T
 - Support 10-Mbps and 100-Mbps
 - High-capacity workstations and servers attach directly to 10/100 switches
 - Switches connected to 100-Mbps hubs using 100-Mbps links
 - 100-Mbps hubs provide building backbone
 - Connected to router providing connection to WAN



Gigabit Ethernet Configuration





千兆以太网

- 对传输速度更高的需求使得千兆以太网(1000Mbps)应运而生。IEEE委员会称之为标准802.3z。
- 千兆以太网设计的目标:
 1. 将数据速率升级到1千兆。
 2. 使其与标准以太网或快速以太网相兼容。
 3. 使用相同的48位地址。
 4. 使用相同的帧格式。
 5. 保留帧长度的最大值和最小值。
 6. 支持快速以太网中定义的自动协商。



全双工模式

- 在全双工模式中，有一个中心交换机将所有的电脑或其他交换机连接起来。
- 每个交换机的每个进入端口都有缓存区，使数据在传输前得以存储。
- 在这种模式中不存在冲突。也就是说CSMA/CD是不必要的。电缆长度的最大值取决于电缆中信号的衰减程度，而不是冲突检测过程。



半双工模式

- 传统方法
- 载波扩展方法
- 帧突发方法



Gigabit Ethernet - Differences

- Carrier extension
- At least 4096 bit-times long (512 bit-times for 10Mbps/100Mbps)
 - The frame length of a transmission is longer than the propagation time at 1 Gbps
- Frame bursting
 - Frame bursting avoids the overhead of carrier extension
 - A single station has a number of small frames ready to send
 - 1500 bytes



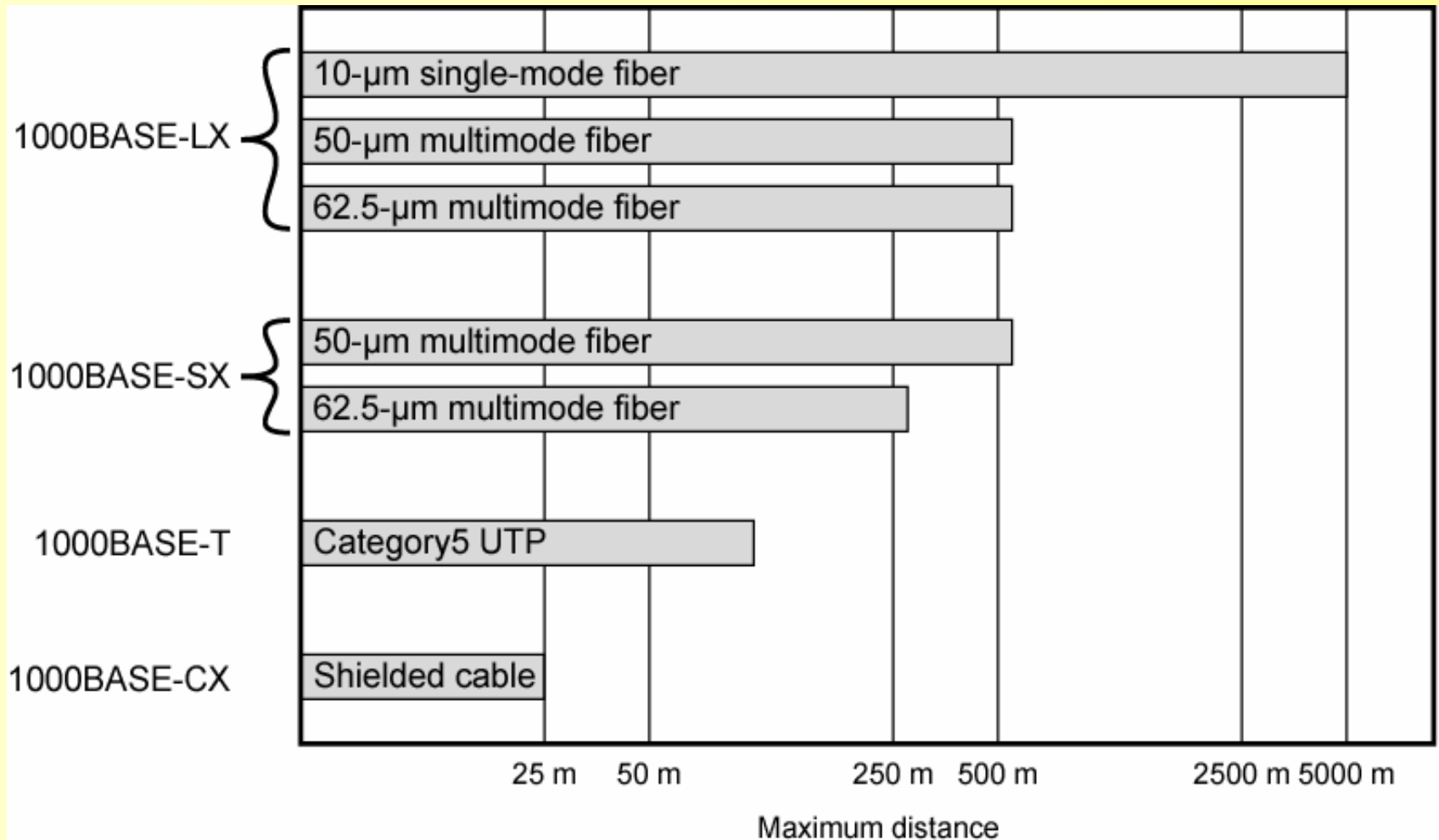
Gigabit Ethernet – Physical

- 1000Base-SX
 - Short wavelength, multimode fiber
- 1000Base-LX
 - Long wavelength, Multi or single mode fiber
- 1000Base-CX
 - Copper jumpers <25m, shielded twisted pair
- 1000Base-T
 - 4 pairs, cat 5 UTP

- Signaling - 8B/10B



Gbit Ethernet Medium Options (log scale)





10Gbps Ethernet - Uses

- High-speed, local backbone interconnection between large-capacity switches
- Server farm
- Campus wide connectivity
- Enables Internet service providers (ISPs) and network service providers (NSPs) to create very high-speed links at very low cost
- Allows construction of (MANs) and WANs
 - Connect geographically dispersed LANs between campuses or points of presence (PoPs)
- Ethernet competes with ATM and other WAN technologies
- 10-Gbps Ethernet provides substantial value over ATM



10Gbps Ethernet - Advantages

- No expensive, bandwidth-consuming conversion between Ethernet packets and ATM cells
- Network is Ethernet, end to end
- IP and Ethernet together offers QoS and traffic policing approach ATM
- Advanced traffic engineering technologies available to users and providers
- Variety of standard optical interfaces (wavelengths and link distances) specified for 10 Gb Ethernet
- Optimizing operation and cost for LAN, MAN, or WAN

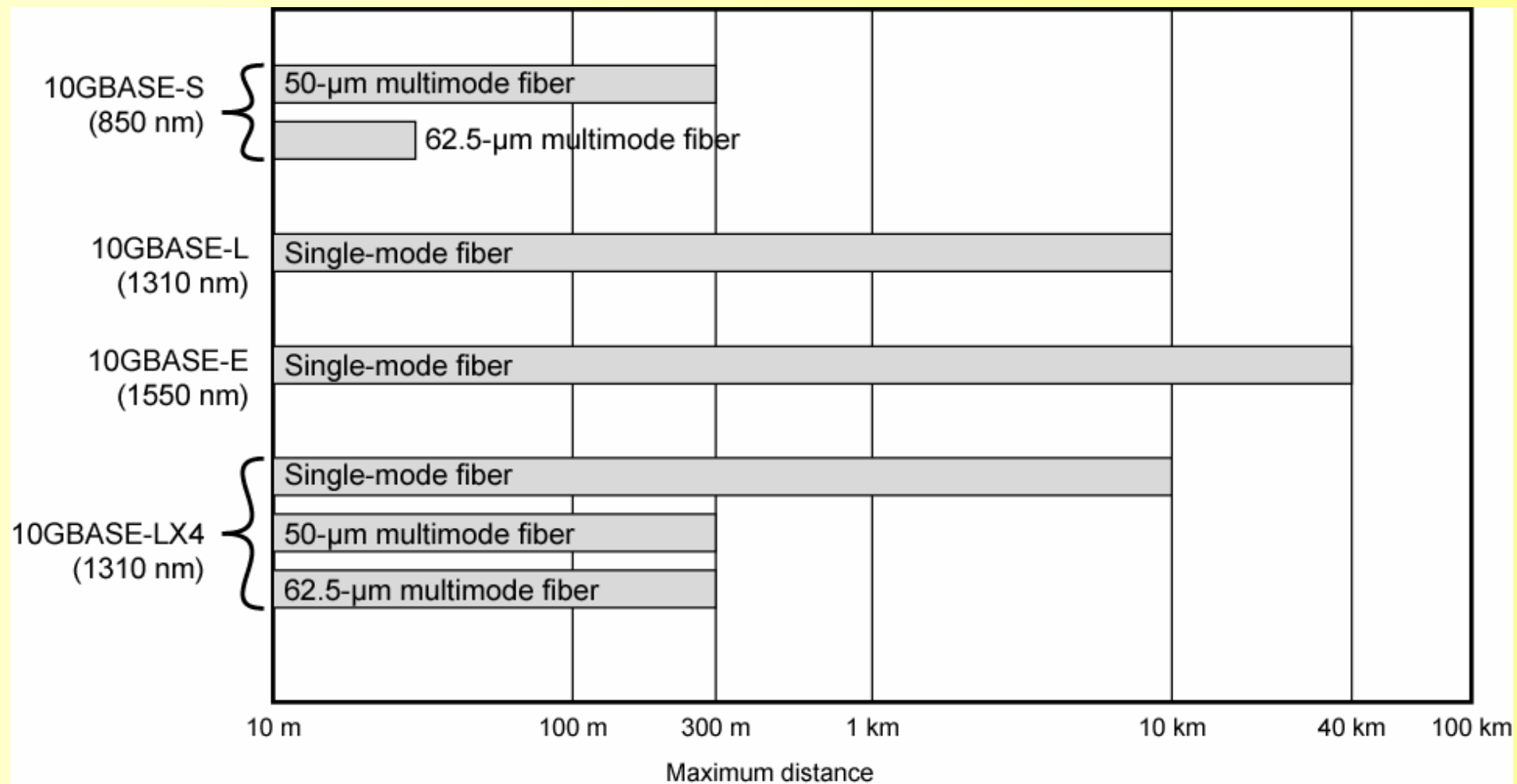


10Gbps Ethernet - Advantages

- Maximum link distances cover 300 m to 40 km
- Full-duplex mode only
- 10GBASE-S (short):
 - 850 nm on multimode fiber
 - Up to 300 m
- 10GBASE-L (long)
 - 1310 nm on single-mode fiber
 - Up to 10 km
- 10GBASE-E (extended)
 - 1550 nm on single-mode fiber
 - Up to 40 km
- 10GBASE-LX4:
 - 1310 nm on single-mode or multimode fiber
 - Up to 10 km
 - Wavelength-division multiplexing (WDM) bit stream across four light waves



10Gbps Ethernet Distance Options (log scale)





16.3 Token Ring (802.5)

- Developed from IBM's commercial token ring
- Because of IBM's presence, token ring has gained broad acceptance
- Currently, large installed base of token ring products
- Never achieved popularity of Ethernet
- Market share likely to decline

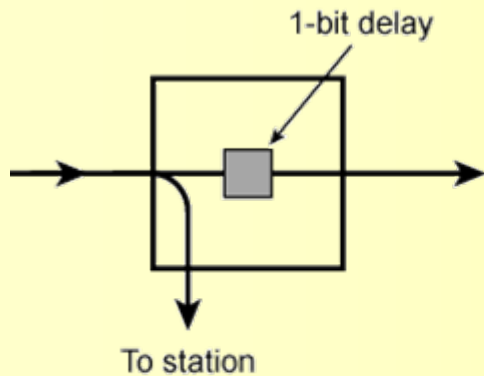


Ring Operation

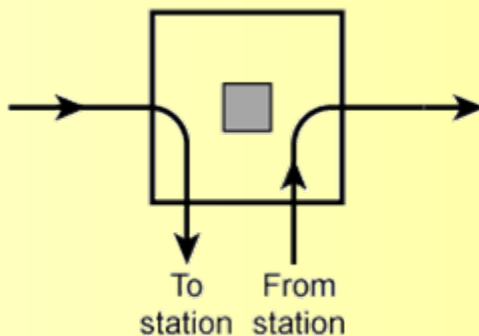
- Each **repeater** connects to two others via **unidirectional transmission links**
 - Form single closed path
- Repeater acts as attachment point
- Data transferred bit by bit from one repeater to the next
 - Repeater regenerates and retransmits each bit
 - Repeater performs data insertion, data reception, data removal
- Frame **removed by transmitter** after one trip round ring



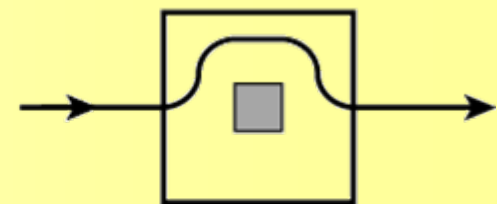
Ring Repeater States



(a) Listen state



(b) Transmit state



(c) Bypass state



Listen State Functions

- Scan passing bit stream for **patterns**
 - Address of attached station vs. destination address
 - Token permission to transmit
- Copy incoming bit and send to attached station
 - If destination address matched
 - Whilst forwarding each bit
- **Modify bit** as it passes
 - e.g. to indicate a packet has been copied (ACK)



Transmit State Functions

- When
 - Station has data
 - Repeater has permission
- May receive incoming bits
 - If ring bit length shorter than frame
 - Pass back to station for checking (ACK)
 - May be more than one frame on ring
 - Buffer other's frame for retransmission later



Bypass State

- Signals propagate past repeater with no delay (other than propagation delay)
- Partial solution to reliability problem
- Improved performance

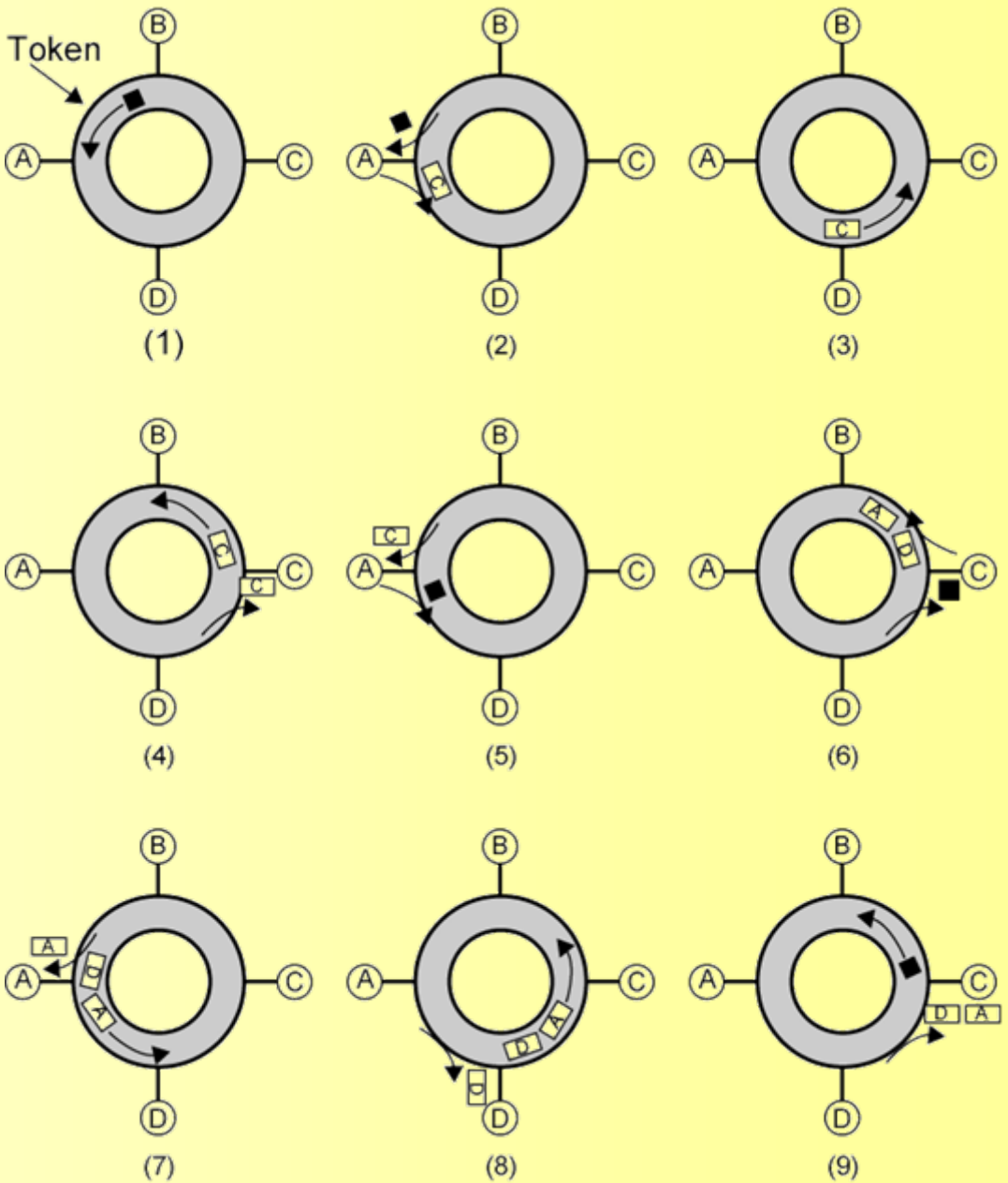


802.5 MAC Protocol

- Small frame (**token**) circulates when idle
- Station waits for token
- Changes one bit in token to make it **SOF for data frame**
- Append rest of data frame
- Frame makes round trip and is absorbed by transmitting station
- Station then **inserts new token** when transmission has finished (leading edge of returning frame arrives)
- Under light loads, some inefficiency
- Under heavy loads, **round robin**



Token Ring Operation



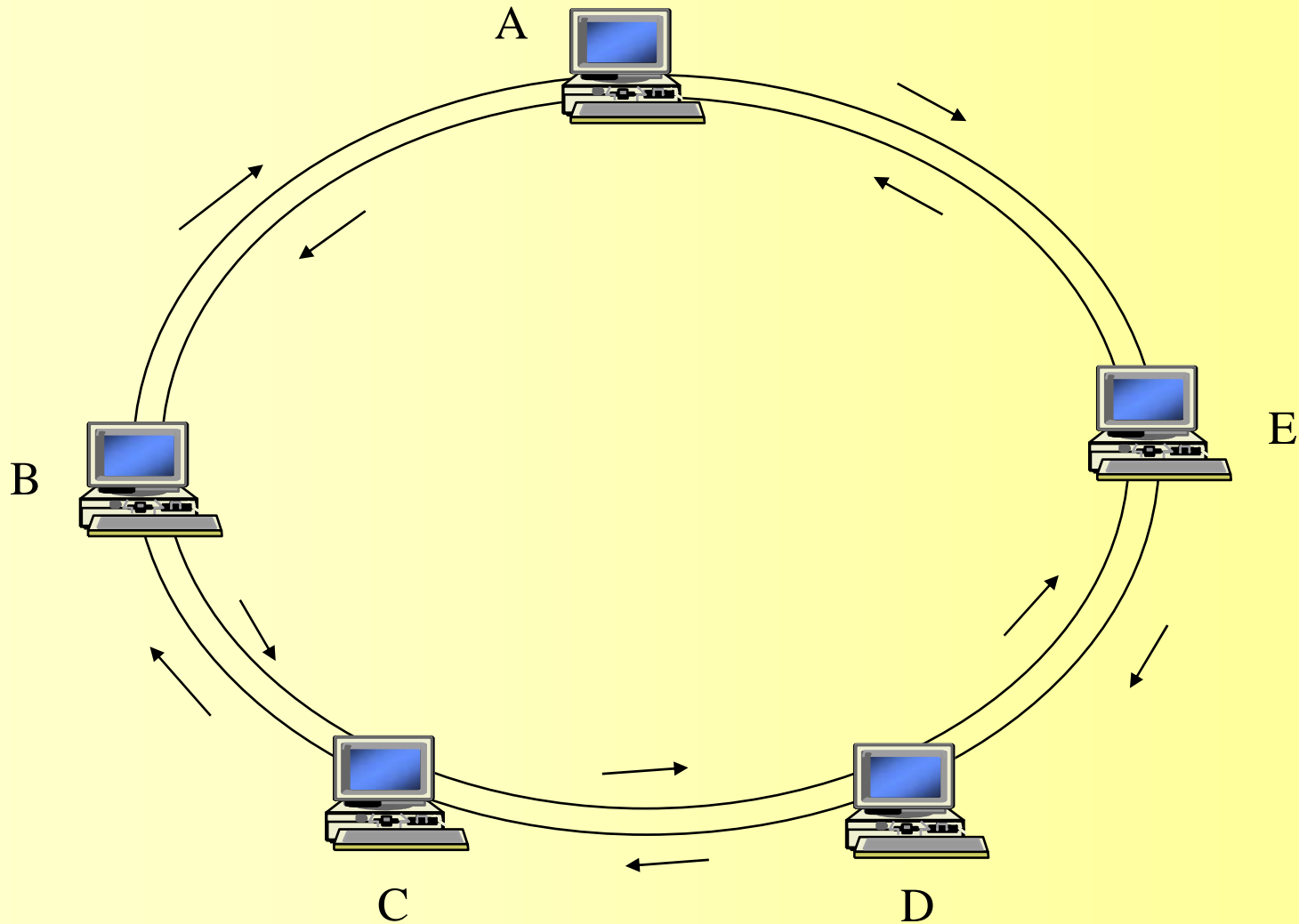


FDDI

- 100 Mbps Token Ring
- Use multi-mode or single-mode optical fiber transmission links
- Span up to 200 kms and permits up to 500 stations
- LAN and MAN applications



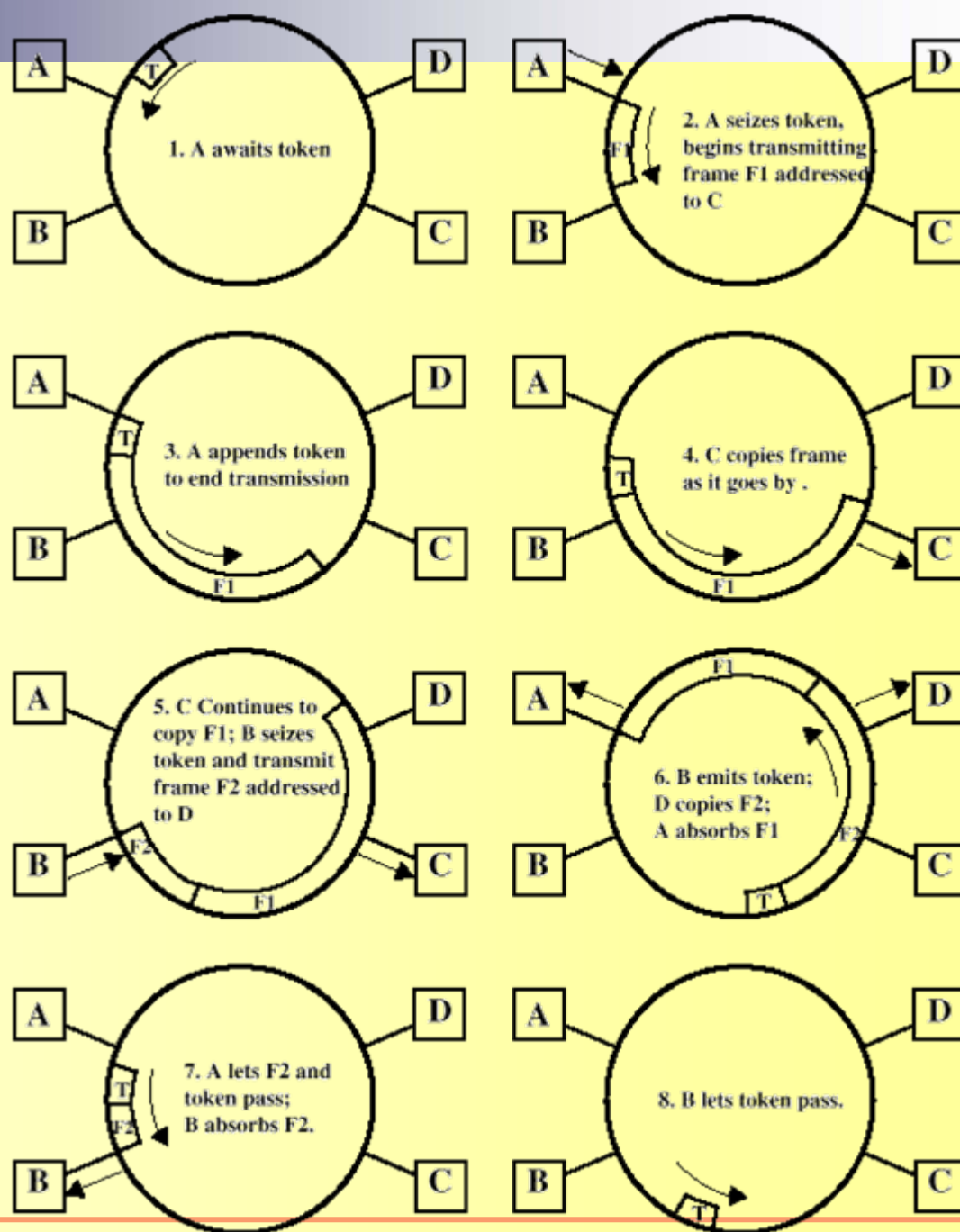
FDDI Token Ring





FDDI Operation

- 在一圈的过程中，每个准备好要发送数据的站都可以发送数据。
- 环路上存在一个以上的数据帧。与802.5的主要区别。

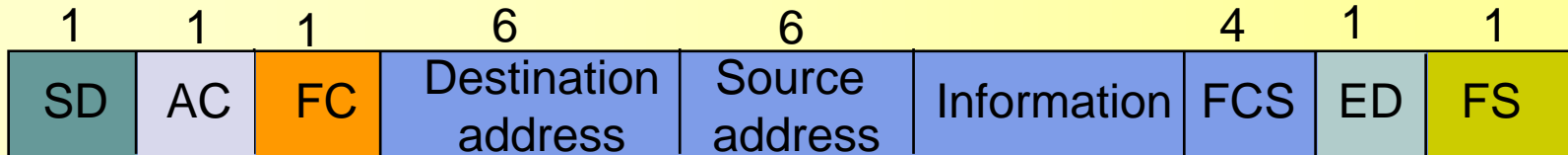




16.3 Token Ring

Token Ring Frame Format (1)

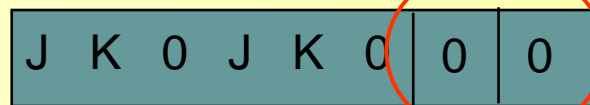
Data frame format



Token frame format

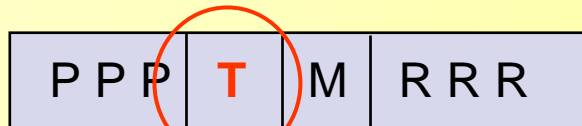


Starting
Delimiter



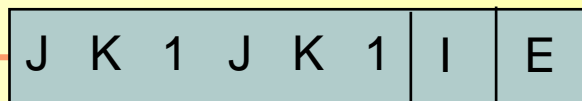
J, K non-data symbols (line code)
J begins as "0" but no transition
K begins as "1" but no transition

Access
Control



PPP=priority; **T=token bit**
M=monitor bit; RRR=reservation
T=0 token; T=1 data

Ending
Delimiter

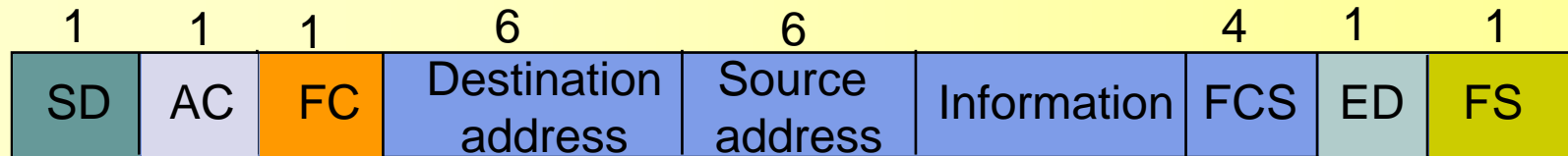


I = intermediate-frame bit
E = error-detection bit

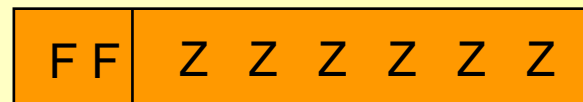


Token Ring Frame Format (2)

Data frame format



Frame control



FF = frame type; FF=01 data frame
FF=00 MAC control frame
ZZZZZZ type of MAC control

Addressing

48 bit format as in 802.3

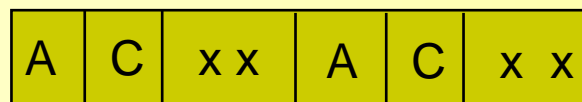
Information

Length limited by allowable token holding time

FCS

CCITT-32 CRC

Frame Status

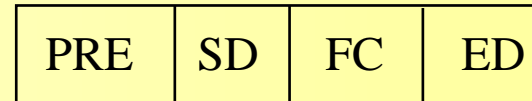


A = address-recognized bit
xx = undefined
C = frame-copied bit

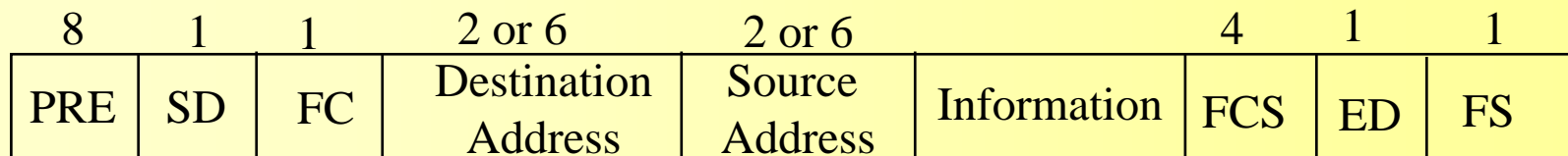


FDDI Frame Format

Token Frame Format



Data Frame Format



Preamble

Frame Control

CLFFZZZZ

C = Sync / Async

L = Address length (16 or 48 bits)

FF = LLC/MAC control/reserved frame type

- No AC field
- FC changed
- Add PRE field
- DA and SA changed



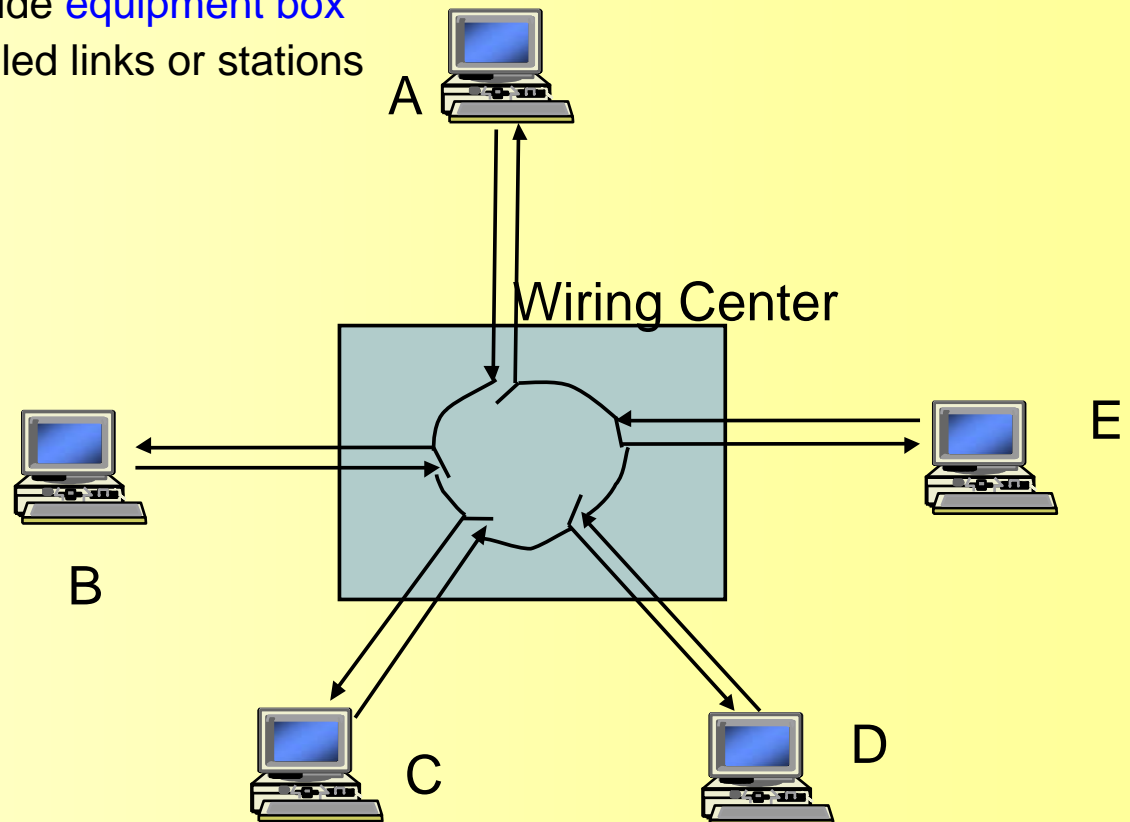
Potential Ring Problems

- Break in any link **disables** network
- Repeater failure disables network
- **Installation** of new repeater to attach new station requires identification of two topologically adjacent repeaters
- Timing jitter
- Method of recovering **lost token** required
- Method of removing circulating frames required
- Mostly solved with **star-ring architecture**



Star Topology Ring LAN

- Stations connected in star fashion to wiring closet
 - Can use existing telephone wiring
- Ring implemented inside **equipment box**
- Relays can bypass failed links or stations





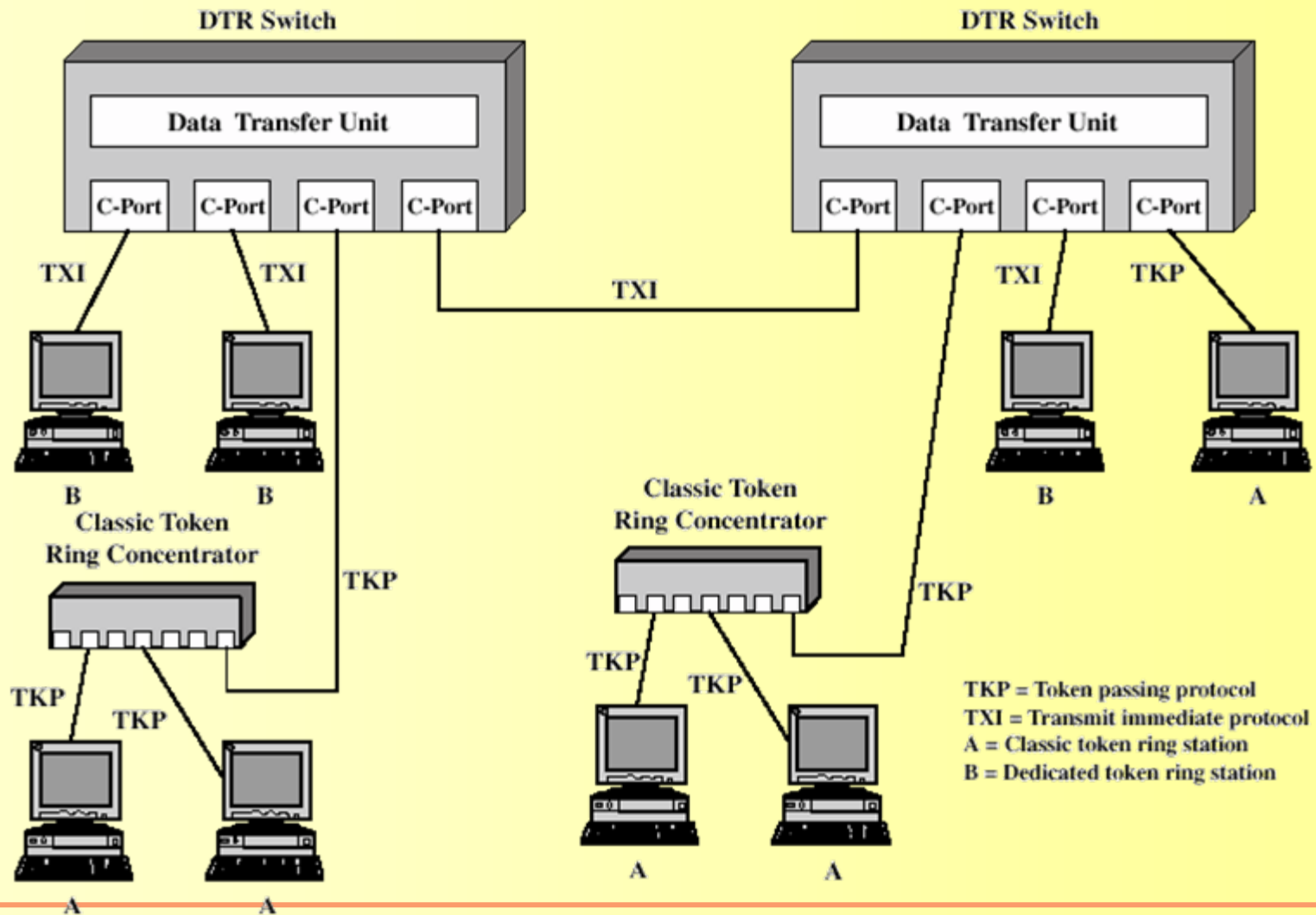
Dedicated Token Ring

- Central hub
 - Also acts as switch
- Full duplex point to point link
- Concentrator acts as **frame level repeater**
- No token passing



16.3 Token Ring

A Dedicated Token Ring Configuration





802.5 Physical Layer

Data Rate (Mbps)	4	16	100	100	1000
Medium	UTP, STP, Fiber	UTP, STP, Fiber	UTP, STP	Fiber	Fiber
Signaling	Differential Manchester	Differential Manchester	MLT-3	4B5B NRZI	8B/10B
Max Frame Len	4,550	18,200	18,200	18,200	18,200
Access Control	TR or DTR	TR or DTR	DTR	DTR	DTR

- Note: 1 Gbit specified in 2001
 - Uses 802.3 physical layer specification



16.4 Fibre Channel – Background

- I/O channel
 - ☐ Direct point to point or multipoint communications link
 - ☐ Hardware based
 - ☐ High Speed
 - ☐ Very short distance
 - ☐ User data moved from source buffer to destination buffer
- Network connection
 - ☐ Interconnected access points
 - ☐ Software based protocol
 - ☐ Flow control, error detection, and recovery
 - ☐ End systems connections



Fibre Channel

- Best of both technologies
- Channel oriented
 - Data type qualifiers for routing frame payload
 - Link level constructs associated with I/O operations
 - Protocol interface specifications to support existing I/O architectures
 - e.g. SCSI (Small Computer System Interface)
- Network oriented
 - Full multiplexing between multiple destinations
 - Peer to peer connectivity
 - Internetworking to other connection technologies



Fibre Channel Requirements

- Full duplex links with two fibers per link
- 100 Mbps to 800 Mbps on single line
 - Full duplex 200 Mbps to 1600 Mbps per link
- Up to 10 km
- Small connectors
- High-capacity utilization, distance insensitivity
- Greater connectivity than existing multidrop channels
- Broad availability, i.e. standard components
- Multiple cost/performance levels, from small systems to supercomputers
- Carry multiple existing interface command sets for existing channel and network protocols



Fibre Channel Facilities

- Uses **generic transport mechanism** based on point-to-point links and a switching network
- Supports simple encoding and framing scheme
- In turn supports a variety of channel and network protocols

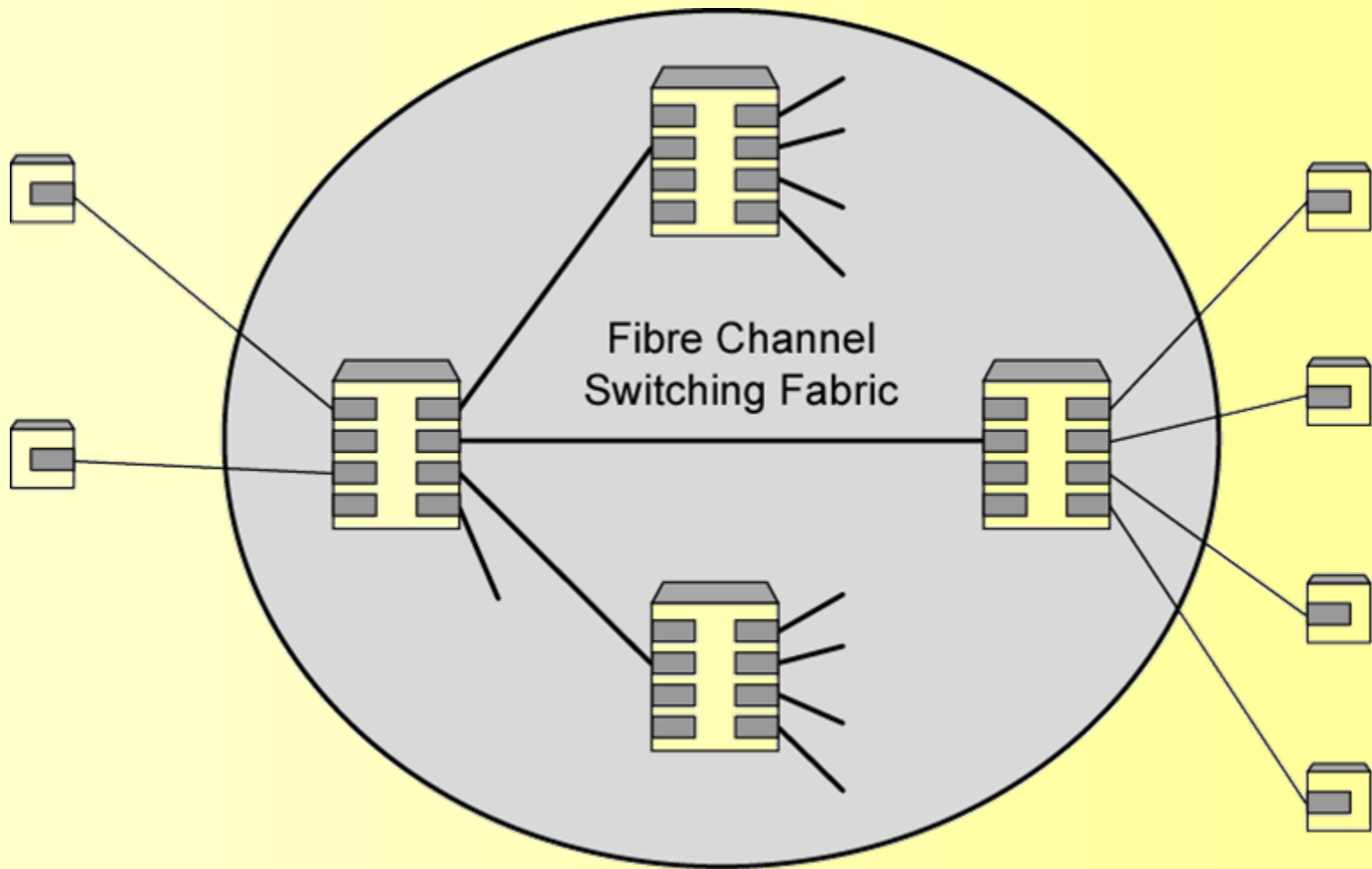


Fibre Channel Elements

- End systems – Nodes
 - Include one or more N_ports for interconnection
- Switched elements – the network or fabric
 - Collection of switched elements form the network or fabric
 - Each switched element has multiple F_ports for interconnection
- Communication across point-to-point links
 - Bidirectional links between ports

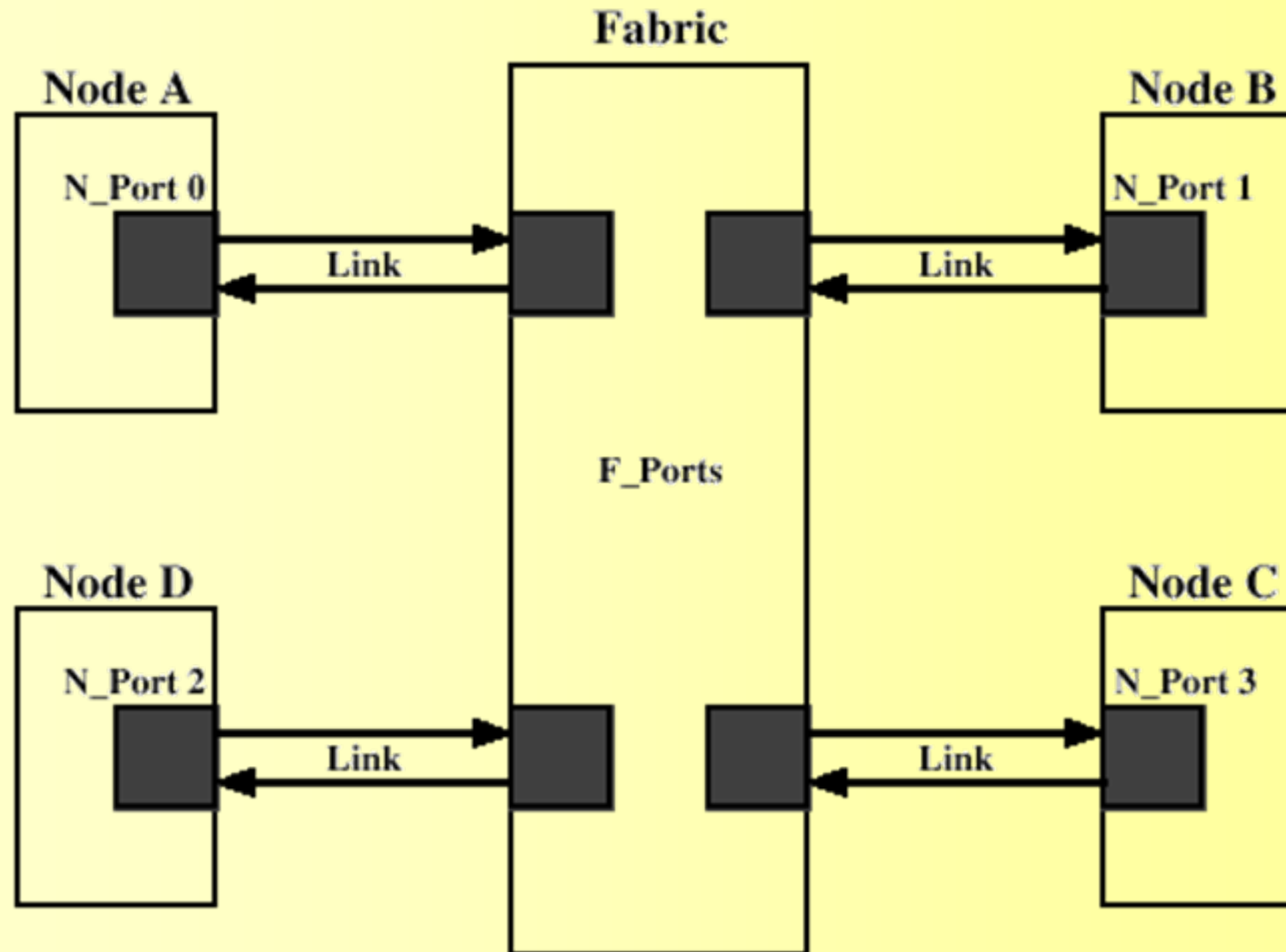


Fibre Channel Network





Fibre Channel Port Types





Fibre Channel Protocol Architecture (1)

■ FC-0 Physical Media

- ☐ Optical fiber for long distance
- ☐ coaxial cable for high speed short distance
- ☐ STP for lower speed short distance

■ FC-1 Transmission Protocol

- ☐ 8B/10B signal encoding

■ FC-2 Framing Protocol

- ☐ Topologies
- ☐ Framing formats
- ☐ Flow and error control
- ☐ Sequences and exchanges (logical grouping of frames)



Fibre Channel Protocol Architecture (2)

■ FC-3 Common Services

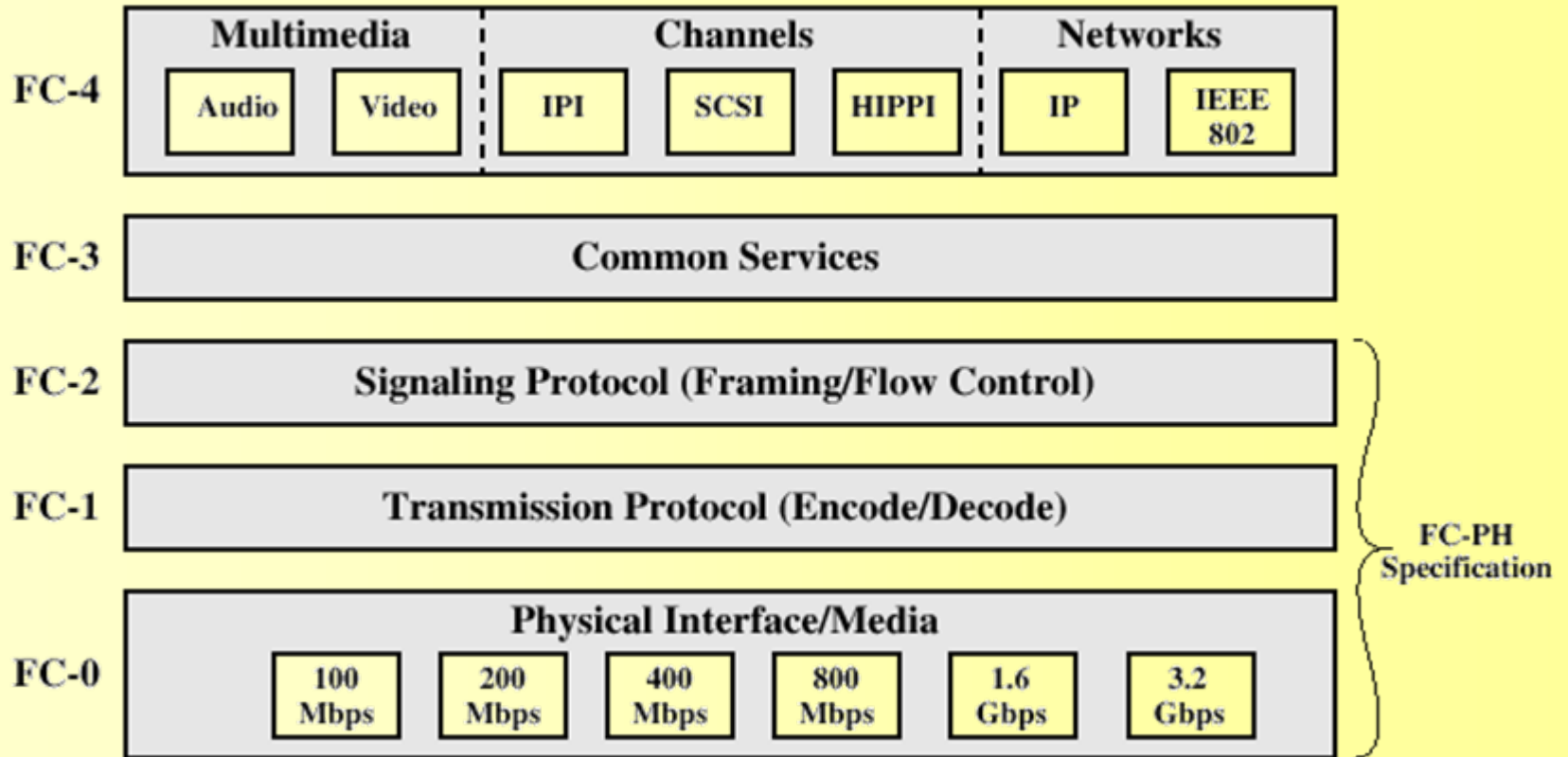
- Including multicasting

■ FC-4 Mapping

- Mapping of channel and network services onto fibre channel
- e.g. IEEE 802, ATM, IP, SCSI



Fiber Channel Levels





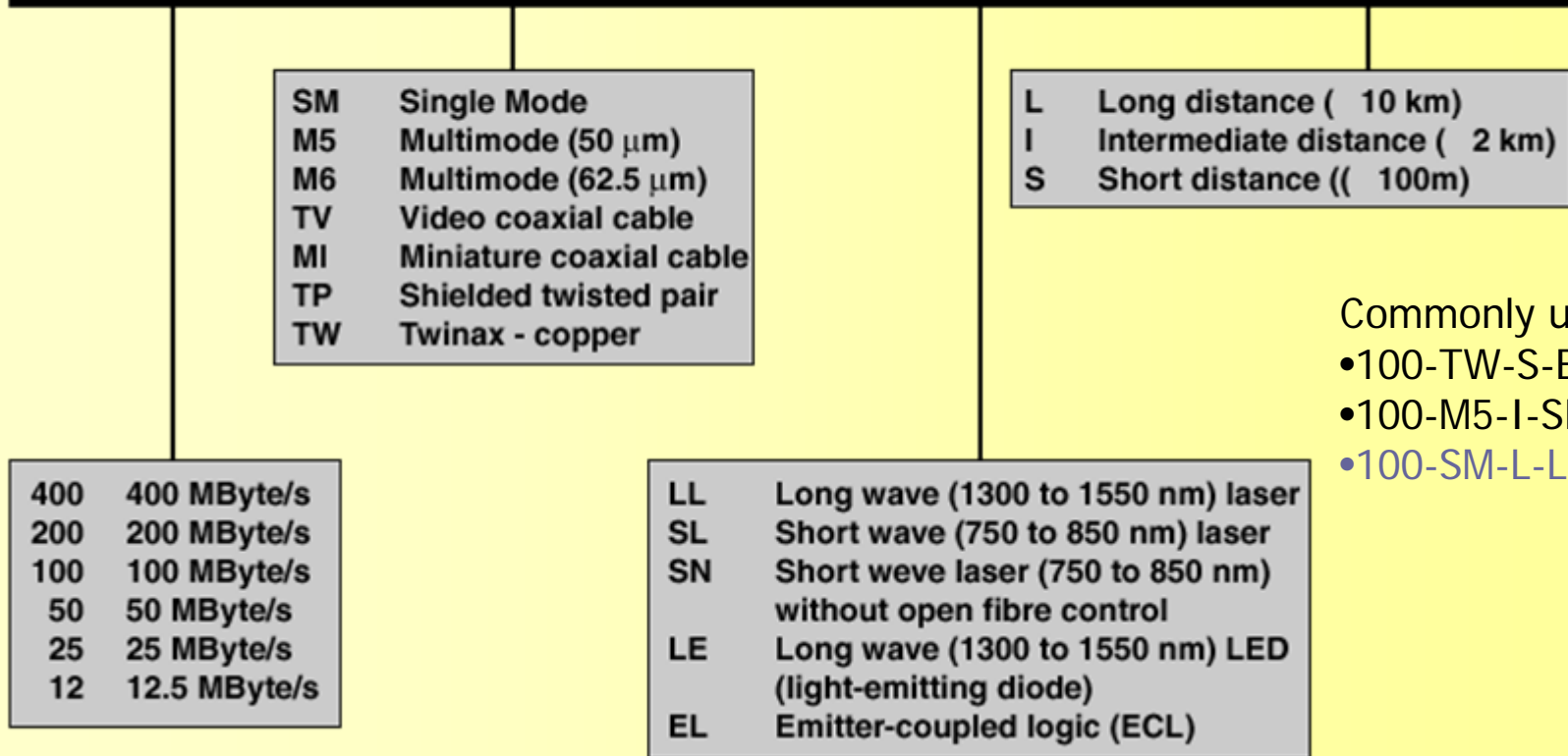
Fibre Channel Physical Media

- Provides **range of options** for
 - Physical medium
 - Data rate on medium
 - Topology of network
- Shielded twisted pair, video coaxial cable, and optical fiber
- Data rates 100 Mbps to 3.2 Gbps
- Point-to-point from 33 m to 10 km



FC-0 Nomenclature

Speed—Medium—Transmitter-Distance



Commonly used:

- 100-TW-S-EL
- 100-M5-I-SL
- 100-SM-L-LL



Fibre Channel Fabric

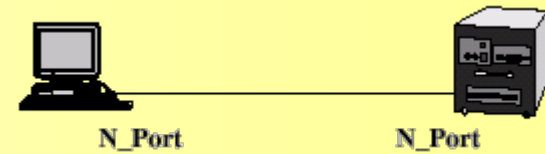
- General topology called fabric or **switched topology**
 - Arbitrary topology includes at least one switch to interconnect number of end systems
 - Or a switched network of many switches, some supporting end nodes
- **Routing transparent to nodes**
 - Each port has unique address
 - When data transmitted into fabric, edge switch uses destination port address to determine location
 - Either delivers frame to node attached to same switch
 - Or transfers frame to adjacent switch to begin routing to remote destination



Alternative Topologies (No Switches)

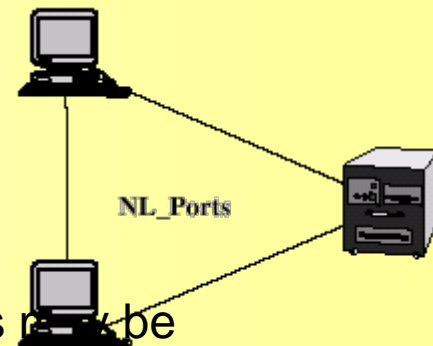
■ Point-to-point topology

- ☐ Only two ports
- ☐ Directly connected, with no intervening switches
- ☐ No routing



■ Arbitrated loop topology

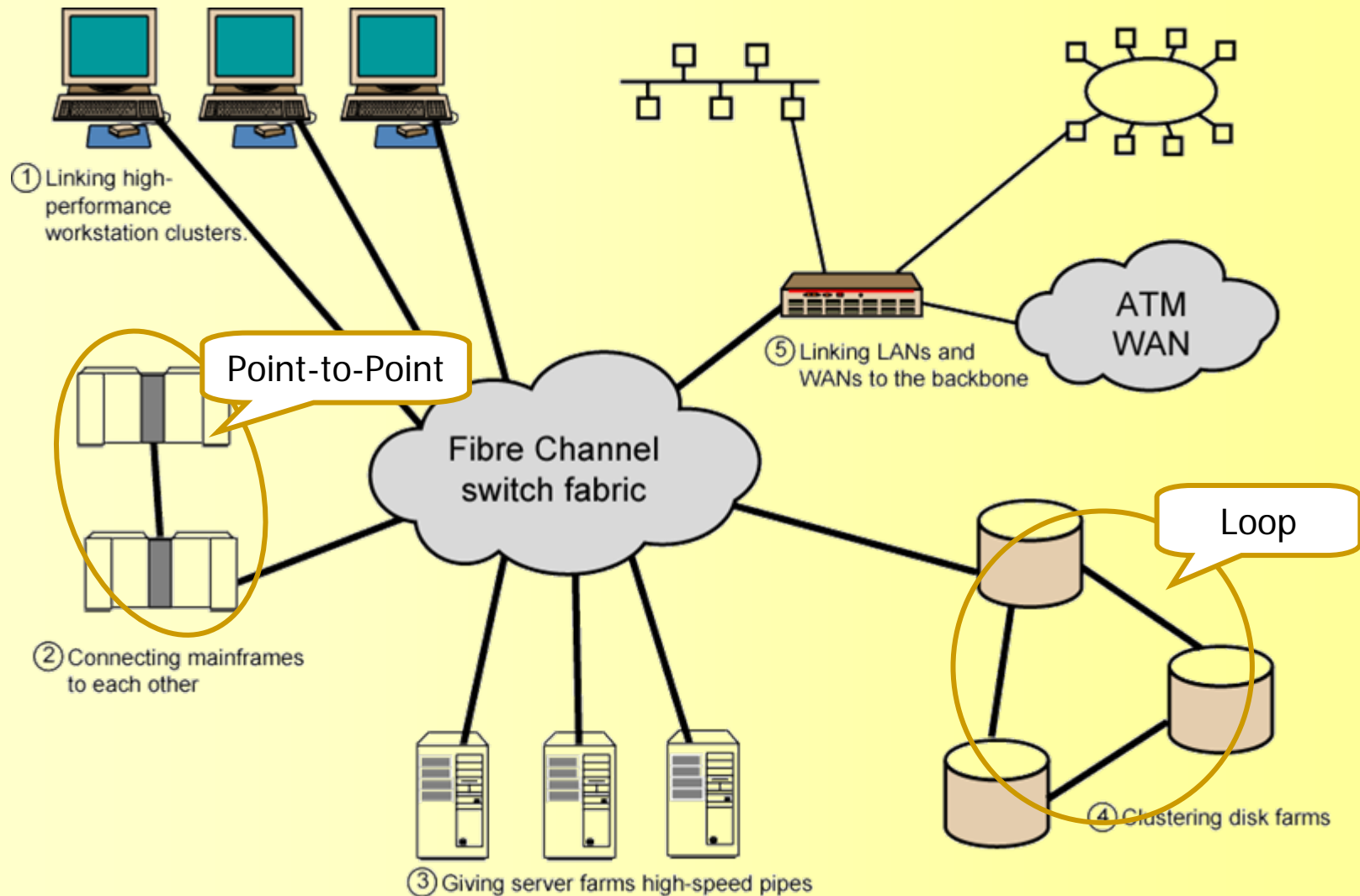
- ☐ Simple, low-cost topology
- ☐ Up to 126 nodes in loop
- ☐ Operates roughly equivalent to token ring



- Topologies, transmission media, and data rates may be combined



Five Applications of Fibre Channel





Fabric Advantages

- Scalability of capacity
 - As **additional ports added**, aggregate capacity of network increases
 - Minimizes congestion and contention
 - Increases throughput
- Protocol independent, Distance insensitive
- Flexibility
 - Switch and transmission link technologies may change without affecting overall configuration
- **Burden on nodes minimized**
 - Node responsible for managing point-to-point connection between itself and fabric
 - Fabric responsible for routing and error detection



Fibre Channel Prospects

- Backed by Fibre Channel Association
- Interface cards for different applications available
- Most widely accepted as peripheral device interconnect
 - To replace such schemes as SCSI
- Technically attractive to general high-speed LAN requirements
 - Must compete with Ethernet and ATM LANs
- Cost and performance issues should dominate the consideration of these competing technologies



作业

- 16.2
- 16.6
- 16.11