南京大学
NANJING UNIVERSITY

Computer Networks

Wenzhong Li, Chen Tian
Nanjing University

Material with thanks to James F. Kurose, Mosharaf Chowdhury, and other colleagues.

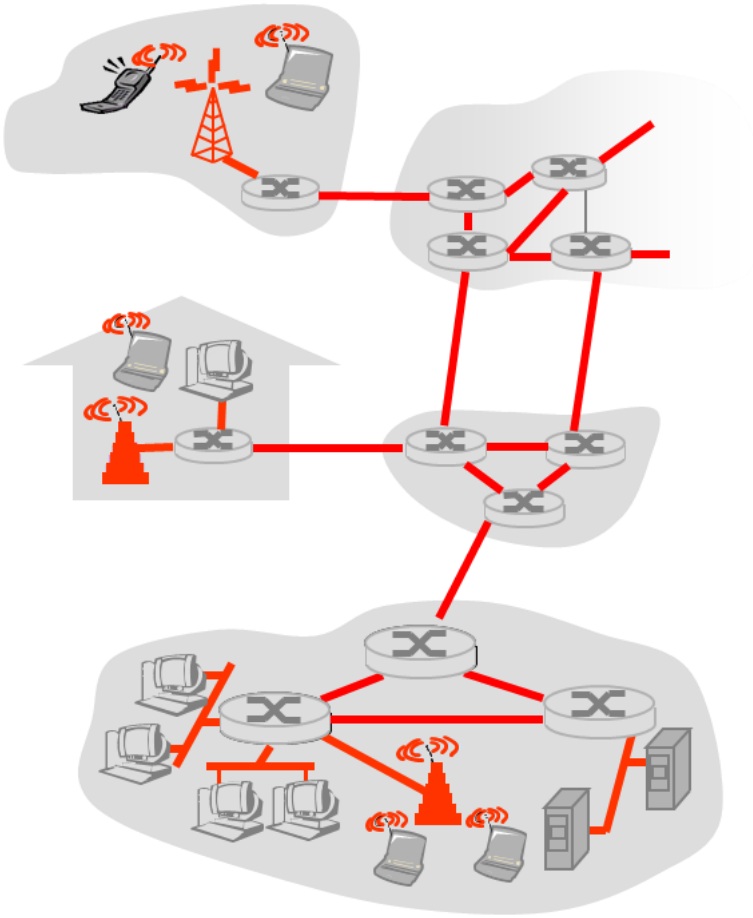
Chapter 2. Link Layer

- Link Layer Service
  - Framing
  - Link access
  - Reliable delivery
  - Error detection and correction
- Local Area Network (LAN)
  - Token Ring
  - Ethernet
- Medium access control (MAC)
- Bridges and Layer-2 switch
- Wireless Networks

Link Layer Service

Direct Links

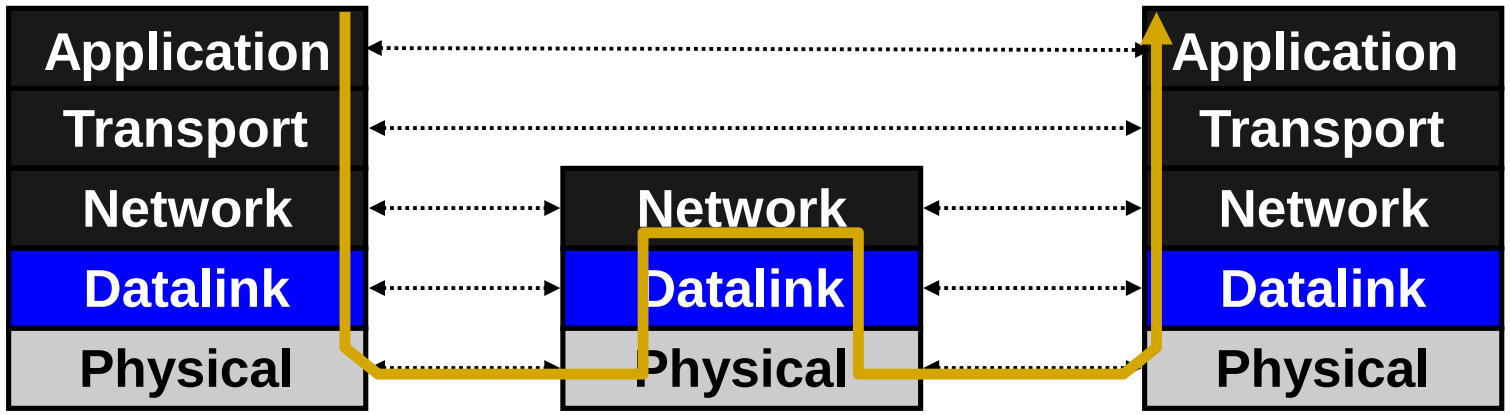
- Hosts and routers are nodes
- Communication channels that connect adjacent nodes are links
- Different types of links
  - Wired point-to-point links
  - Wired multiple access links (LANs)
  - Wireless links (WiFi)



link layer has responsibility of transferring datagram from one node to physically adjacent node over a link

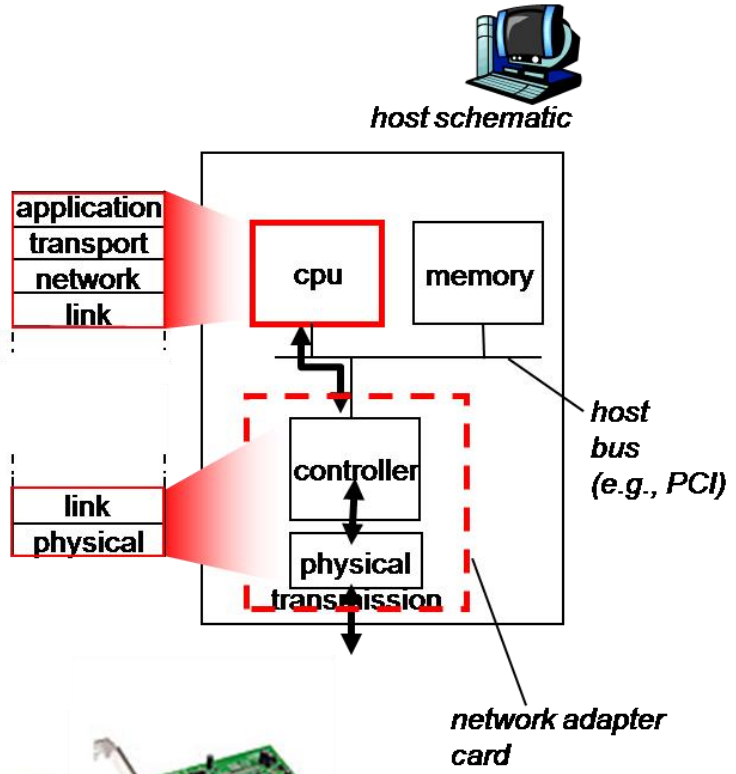
Data link layer

- layer-2 packet: frame, encapsulates datagram
- Transfers data between adjacent nodes or between nodes on the same local area network

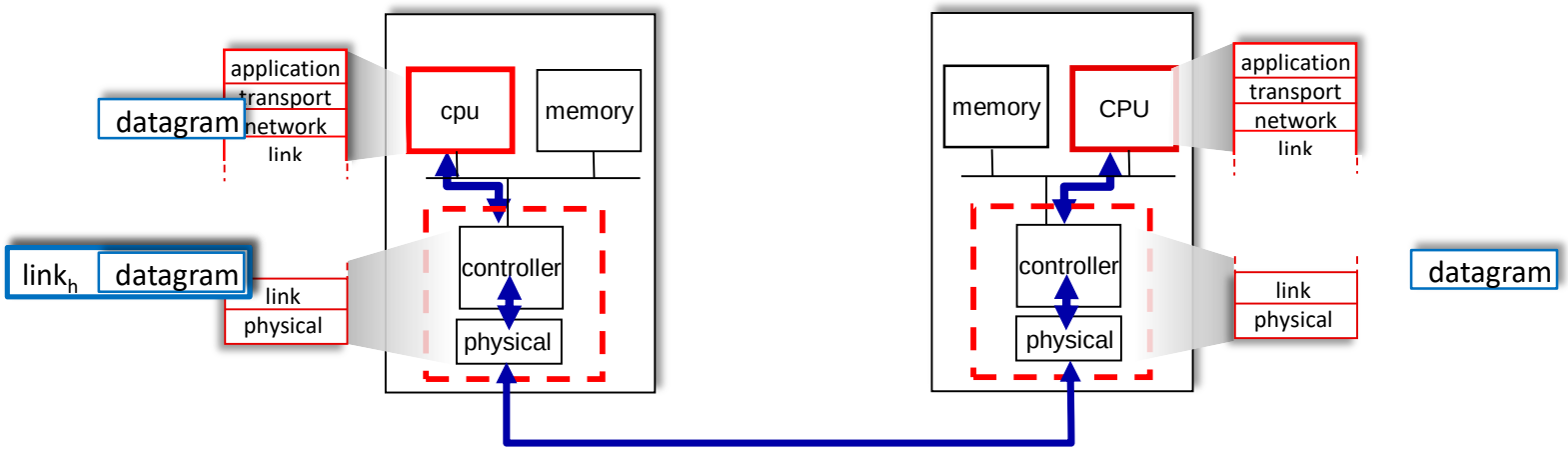


Implement the Link Layer

- In host and router (switch)
- Link layer implemented in "adaptor"
  - i.e. network interface card (NIC)
  - Ethernet card, 802.11 card
- Implements link, physical layer
- Attaches into host's system buses
- Combination of hardware, software, firmware



Interfaces Communicating



sending side:

- encapsulates datagram in frame
- adds error checking bits, reliable data transfer, flow control, etc.

receiving side:

- looks for errors, reliable data transfer, flow control, etc.
- extracts datagram, passes to upper layer at receiving side

Link Layer Services

- Provides four primary services
- Framing
  - Encapsulate upper-level data into frame, adding header and trailer
- Link access
  - Coordinate access for shared multiple access medium
  - "MAC" addresses used in frame headers to identify source and destination
  - Half-duplex and full-duplex: Whether transmit and receive at the same time

Link Layer Services

- Reliable delivery over the link
  - Seldom used on low bit-error link (e.g. fiber)
  - Wireless links: high error rates
  - Flow control： Pacing between adjacent sending and receiving nodes
- Error detection and correction
  - Handling errors caused by signal attenuation or noise
  - Receiver detects presence of errors
  - Signals sender for retransmission or drops frame

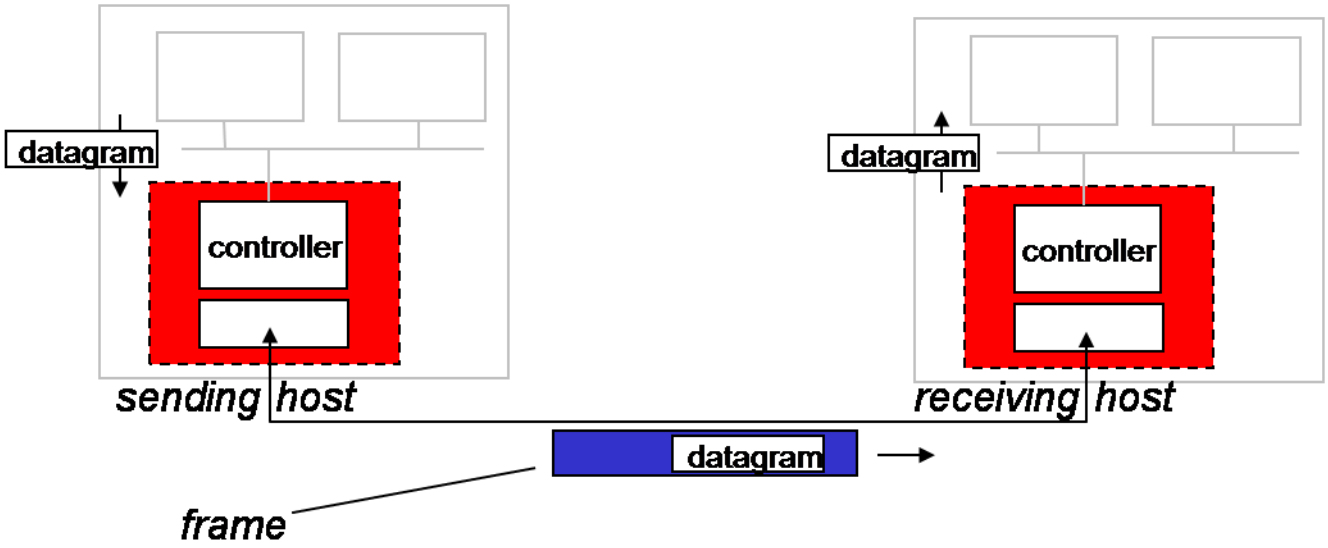
(1) Link Layer Framing

Sending side

- Encapsulates datagram in frame
- Adds error checking bits, flow control, etc.

Receiving side

- Looks for errors, flow control, etc.
- Extracts datagram, passes to upper layer

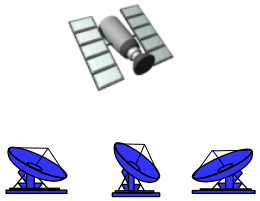


(2) Link Access

- Two types of "links":
- Point-to-point: dedicated pairwise communication
  - E.g., long-distance fiber link
  - E.g., Point-to-point link b/n Ethernet switch and host
- Broadcast: shared wire or medium
  - Traditional Ethernet (pre ~2000)
  - 802.11 wireless LAN

shared wire (e.g., cabled Ethernet)

shared RF (e.g., 802.11 WiFi)



shared RF (satellite)

humans at a cocktail party (shared air, acoustical)

Multiple Access Control (MAC)

- Context: a shared broadcast channel
  - Must avoid having multiple nodes speaking at once
  - Otherwise, collisions lead to garbled data
- Need distributed algorithm to determine which node can transmit

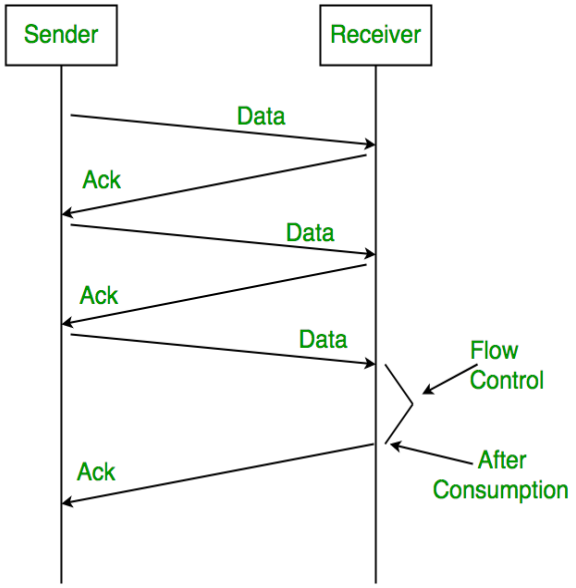
- Multiple access protocol
  - Distributed algorithm that determines how nodes share channel, i.e., determine when node can transmit
  - Communication about channel sharing must use channel itself!
    - No out-of-band channel for coordination

(3) Reliable delivery

- Flow Control over the link
- Ensuring the sender not overwhelm the receiver
  - Preventing buffer overflow
- Methods
  - Stop and Wait
  - Sliding window

Stop and Wait

- Source: transmits frame
- Destination: receives frame and replies with ACK
- Source: waits for ACK before sending next frame
- Destination can stop flow by not send ACK
- Work well for large frames



Sliding Window

- Allow multiple frames to be in transit
- Receiver has buffer (window) sized Win
- Sender can send up to Win frames without ACK
- Each frame is numbered
- ACK includes number of next frame expected
- Sequence number bounded by field of size (k)
  - Frames are numbered modulo 2^k
- Question: how to set k given Win?
- Answer: $Win <= 2^k$

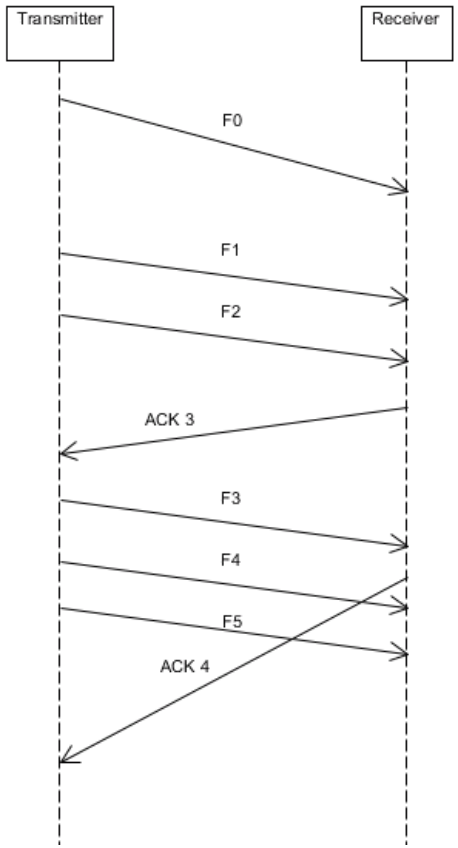


Illustration of Sliding Window

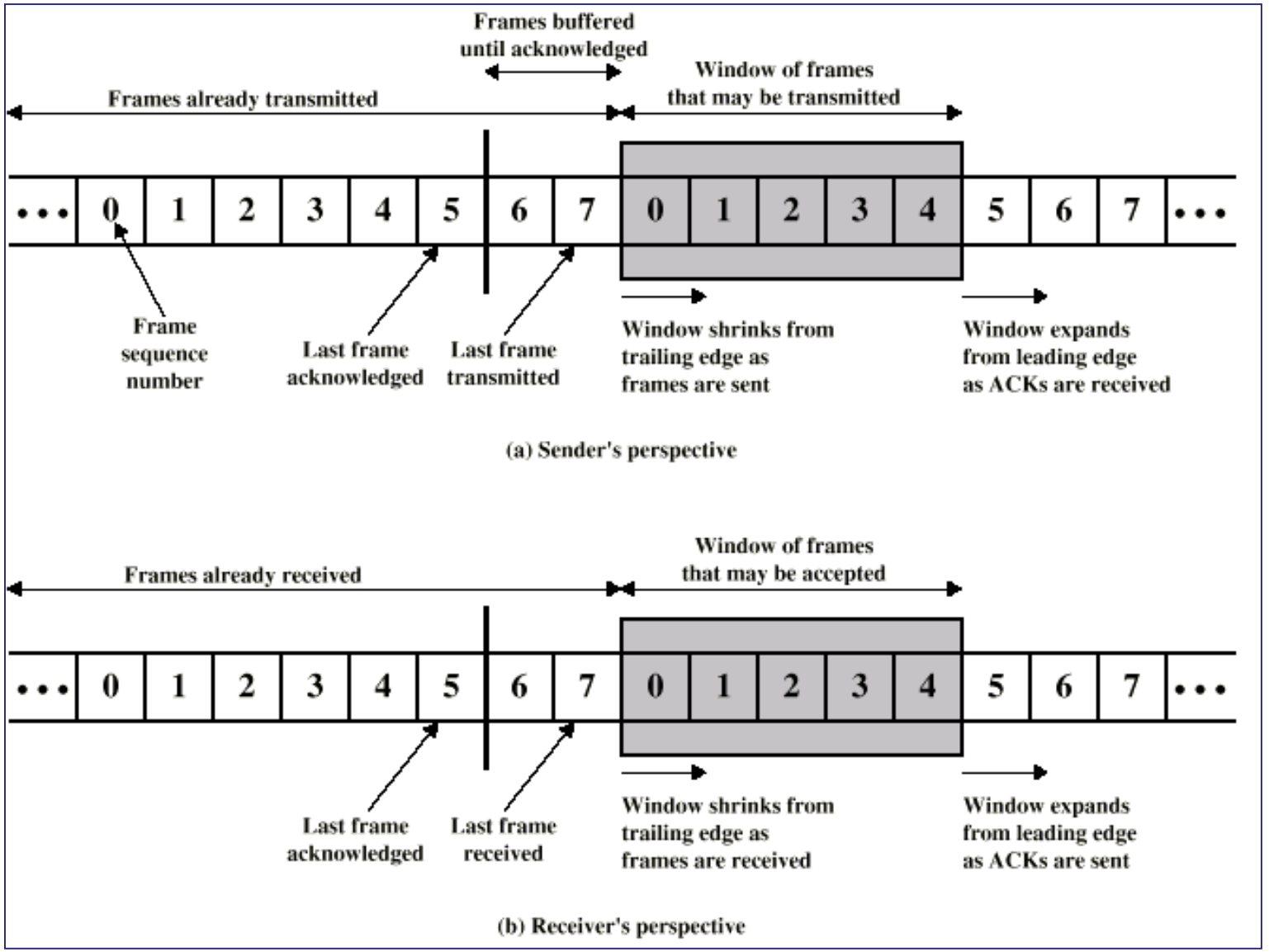
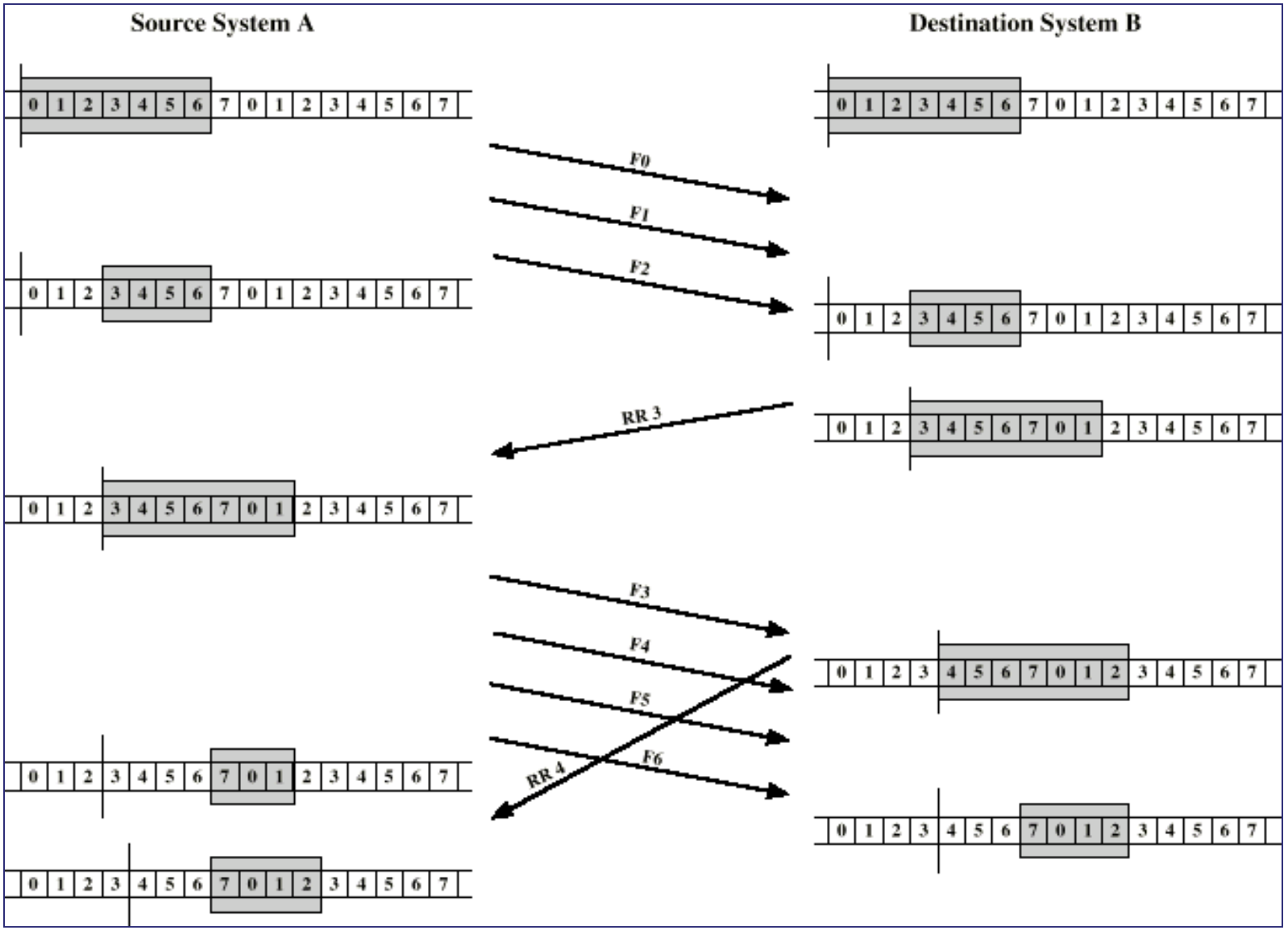


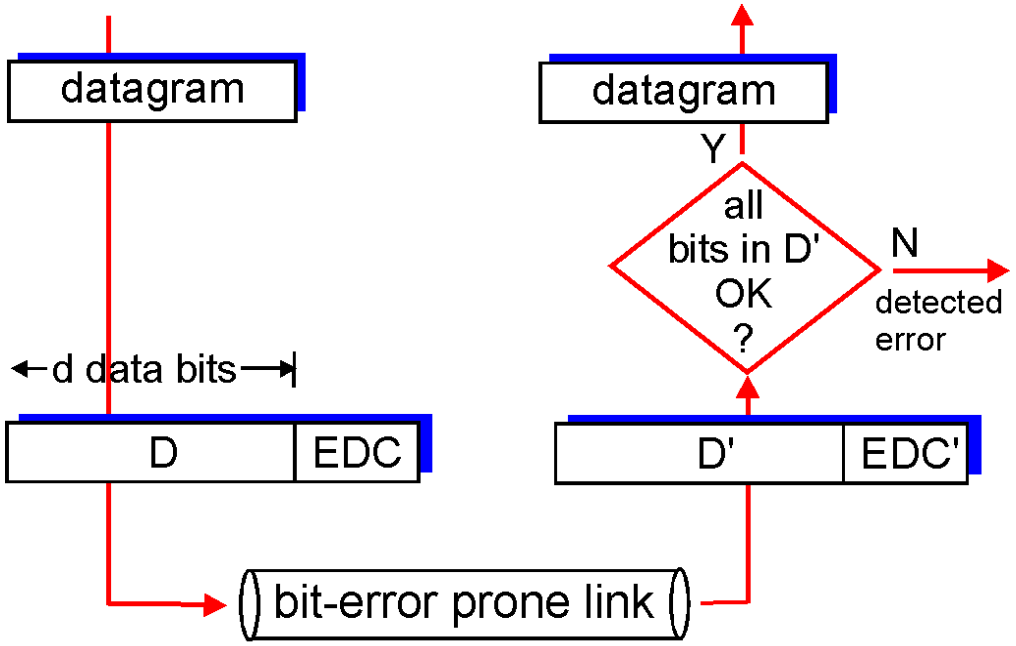
Illustration of Sliding Window



Error Handling in Sliding Window

- Go Back N
  - If error, reply with rejection (NAK)
  - The error frame and all future frames need be retransmitted
- Selective Reject
  - Only rejected error frames need be retransmitted
  - Receiver must maintain large enough buffer

(4) Error Detection and Correction



- EDC= Error Detection and Correction bits (redundancy)
- D = Data protected by error checking, may include header fields

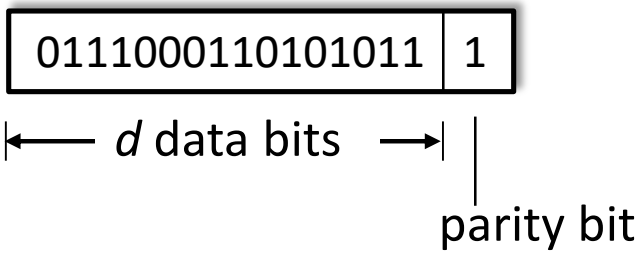
Note: error detection not 100% reliable!

- Larger EDC field yields better detection and correction

Parity Checking

single bit parity:

- detect single bit errors



Even/odd parity: set parity bit so there is an even/odd number of 1's

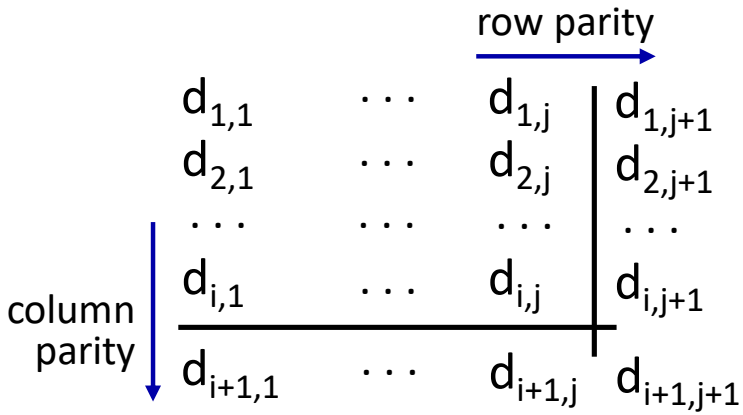
At receiver:

- compute parity of *d* received bits
- compare with received parity bit – if different than error detected

Parity Checking

Can detect and correct errors (without retransmission!)

- two-dimensional parity: detect *and correct* single bit errors



no errors: (even parity)

1 0 1 0 1	1
1 1 1 1 0	0
0 1 1 1 0	1
0 0 1 0 1	0

detected and correctable single-bit error:

1 0 1 0 1	1
1 0 1 1 0	0
0 1 1 1 0	1
0 0 1 0 1	0

parity error

parity error

Internet checksum (review, see section 3.3)

Goal: detect errors (*i.e.,* flipped bits) in transmitted segment

sender:

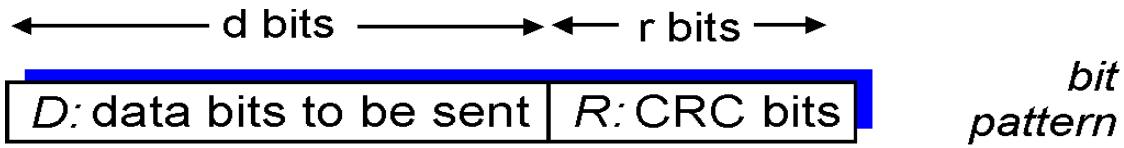
- treat contents of UDP segment (including UDP header fields and IP addresses) as sequence of 16-bit integers
- checksum: addition (one's complement sum) of segment content
- checksum value put into UDP checksum field

receiver:

- compute checksum of received segment
- check if computed checksum equals checksum field value:
  - not equal - error detected
  - equal - no error detected. *But maybe errors nonetheless?* More later ….

Cyclic Redundancy Check

- Widely used in hardware-based implementation
- View data bits, D, as a binary number
- Choose r+1 bit pattern (generator or polynomial), G
  - G is called a Key, which is known to both the sender and receiver ahead.
- Since $D * 2^r = a * G \oplus R$, so $D * 2^r \oplus R = a * G$
- Sender: send $D * 2^r \oplus R$, represented by <D,R>
- Receiver: when received <D,R>
  - If <D, R> exactly divisible by G (modulo 2) , no error
  - If divides <D,R> by G has non-zero remainder: error detected!
- Limit: Can detect burst errors less than r+1 bits



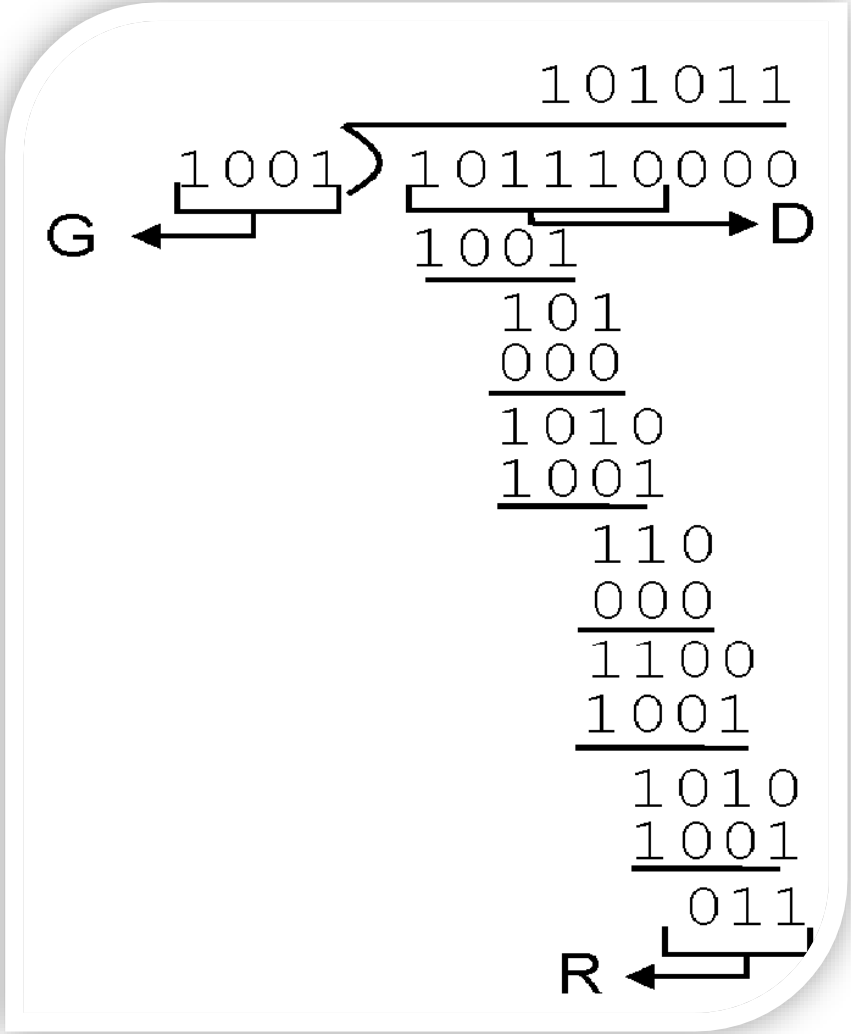
$D * 2^r$ XOR R — *mathematical formula*

Example of CRC

- Since $D * 2^r = a * G \oplus R$, so $D * 2^r \oplus R = a * G$
- Obtain R by:

$$R = \text{remainder}\left[\frac{D \cdot 2^r}{G}\right]$$

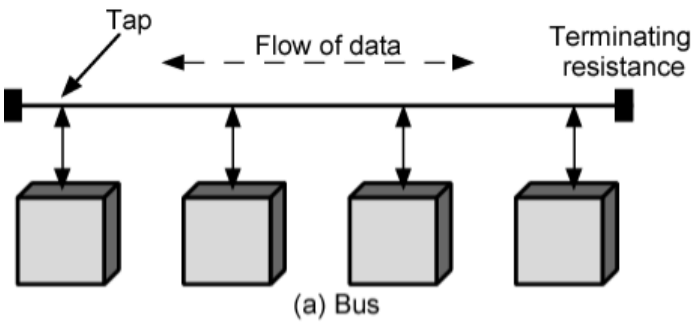
- Question：
- D=101110, r=3, G=1001
- R=?



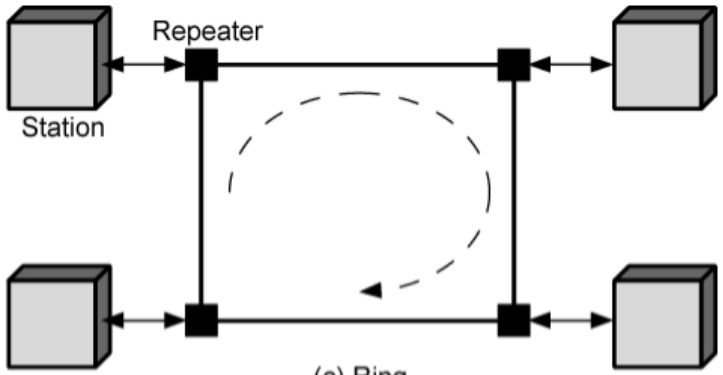
循环移位模2运算

Local Area Network (LAN)

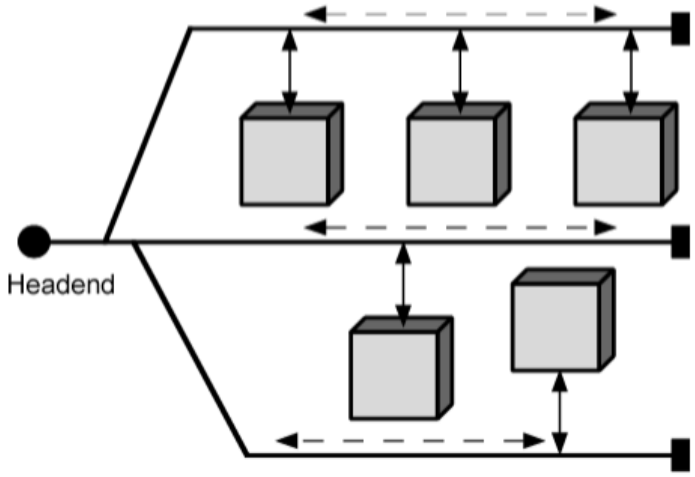
Different Topologies of LAN



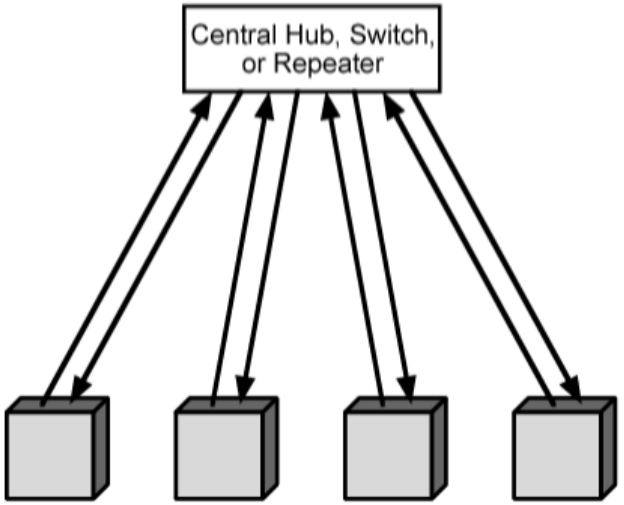
(a) Bus



(c) Ring



(b) Tree



(d) Star

Different Types of LANs

- Token Ring
- Ethernet

- Wireless LAN

Token Ring

Token Ring

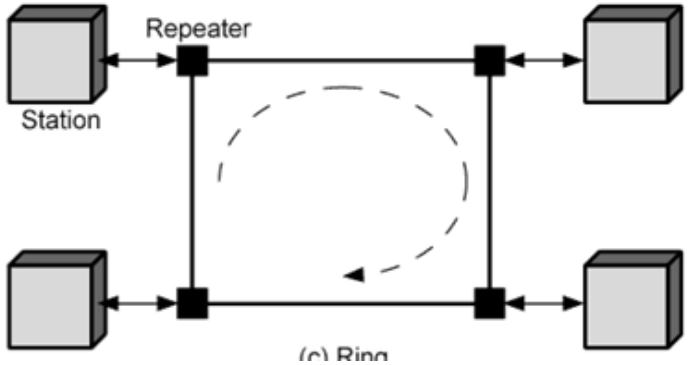
- A protocol for LAN, IEEE 802.5

- Developed from IBM's commercial token ring
- Because of IBM's presence, token ring has gained broad acceptance

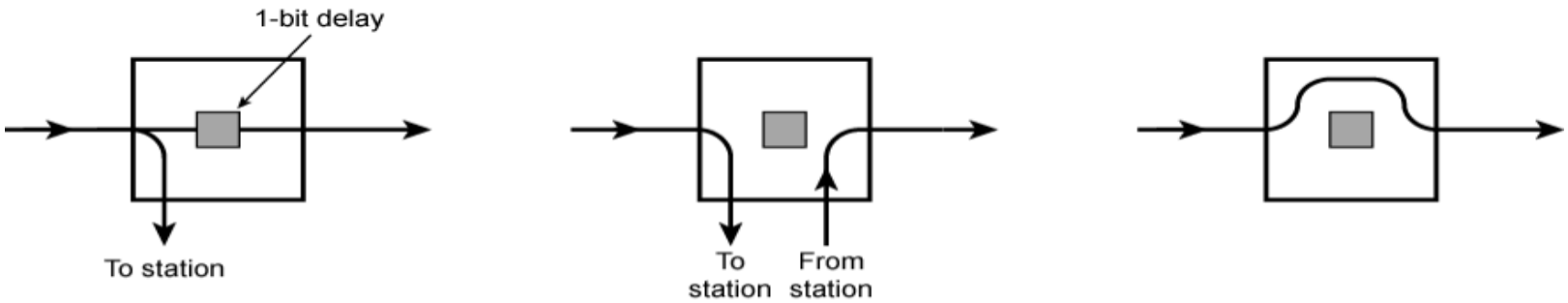
- Never achieved popularity of Ethernet

Ring Operation

- Each repeater connects to two others via unidirectional transmission links
- 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

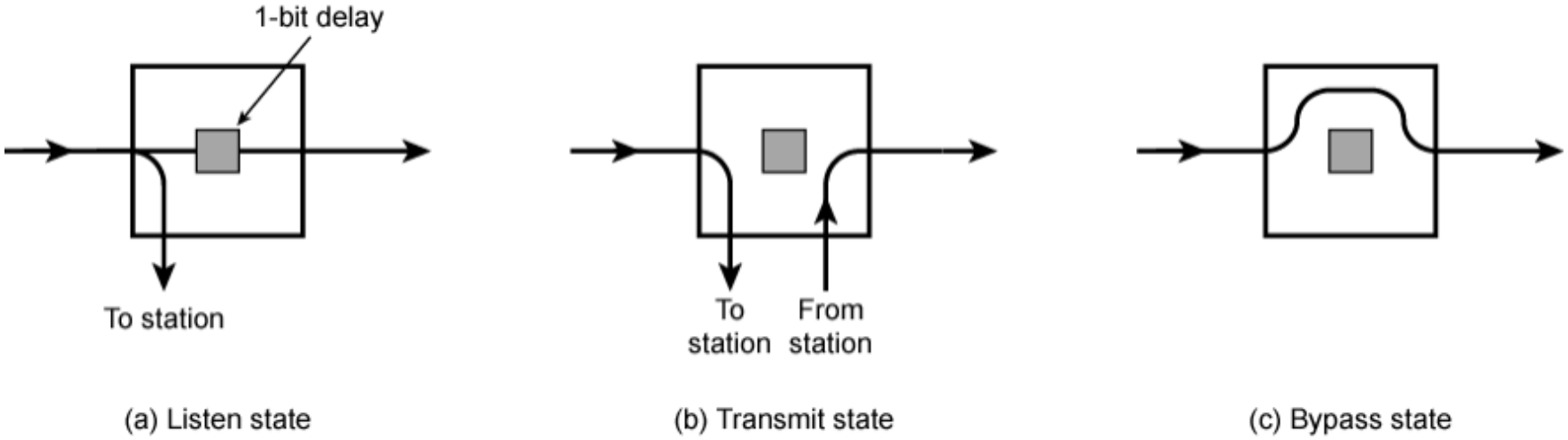


(a) Listen state (b) Transmit state (c) Bypass state

Listen State

- 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)
  - Or make reservation

Ring Repeater States



(a) Listen state (b) Transmit state (c) Bypass state

Transmit state

- Reclaim frame and pass back to station for checking (ACK)
- May buffer other's frame for retransmission later

Bypass state

- Do nothing more than a connector

802.5 MAC Protocol

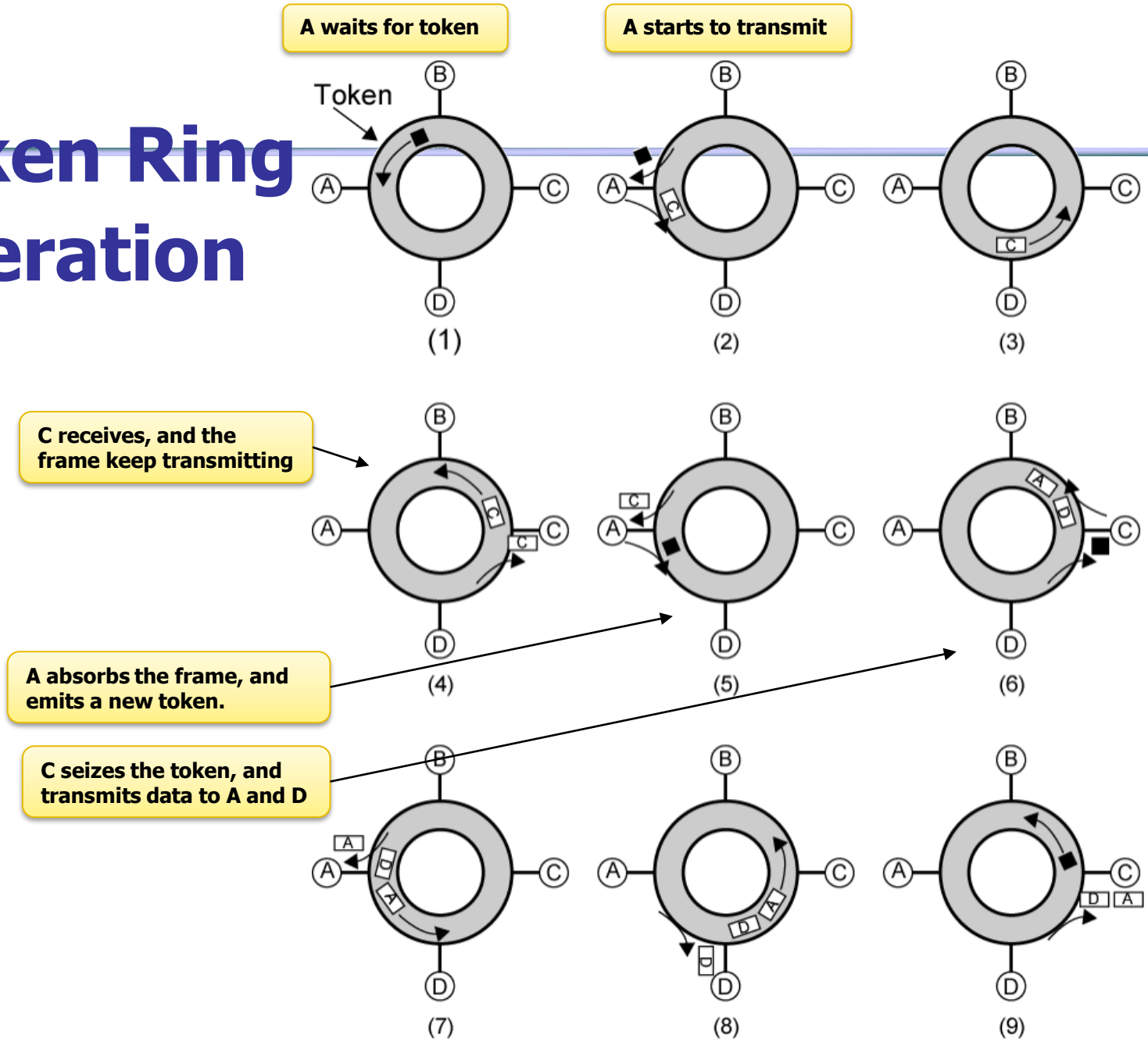
- Small frame (token) circulates when idle
- Station waits for token

- Changes one bit in token to make it SOF (Start of Frame) 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



802.5 Physical Layer

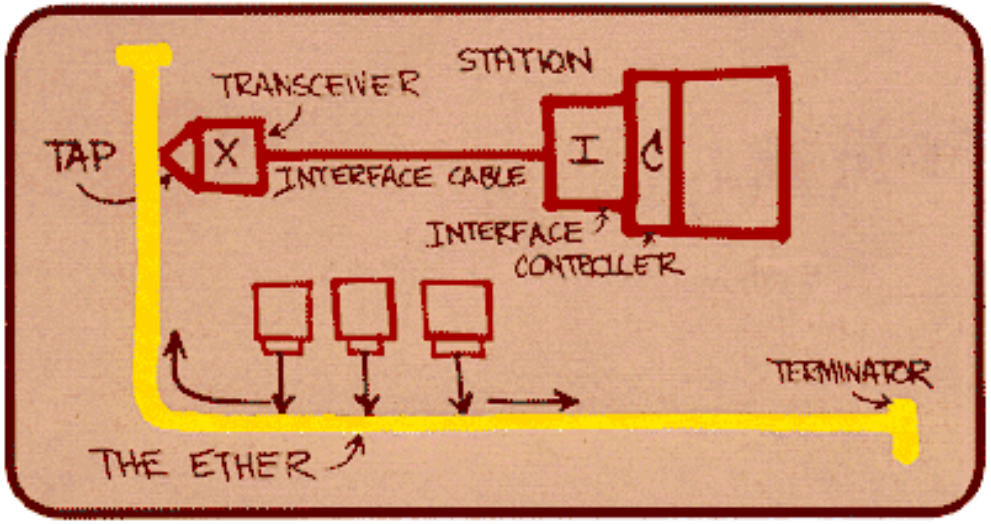
Data Rate (Mbps)	4	16	100	100	1000
Medium	UTP, STP, Fiber	UTP, STP, Fiber	UTP, STP	Fiber	Fiber
Signaling	Differential Manchester	Differential Manchester	MLT-3	4B5B NRZI	8B/10B
Max Frame Len	4,550	18,200	18,200	18,200	18,200
Access Control	TR or DTR	TR or DTR	DTR	DTR	DTR

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

Ethernet

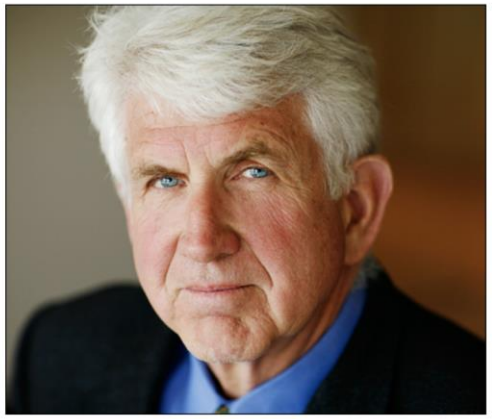
Ethernet

- "Dominant" wired LAN technology:
  - First widely used LAN technology
  - Simpler, cheap
  - Kept up with speed race: 10 Mbps – 400 Gbps
  - Single chip, multiple speeds (e.g., Broadcom BCM5761)



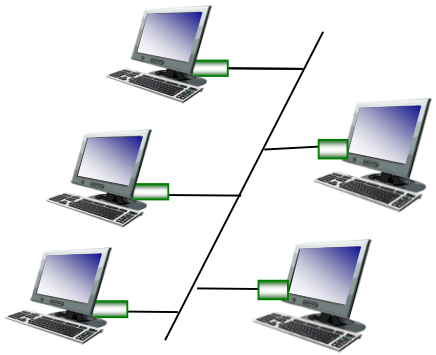
Metcalfe's Ethernet sketch

Bob Metcalfe: Ethernet co-inventor, 2022 ACM Turing Award recipient

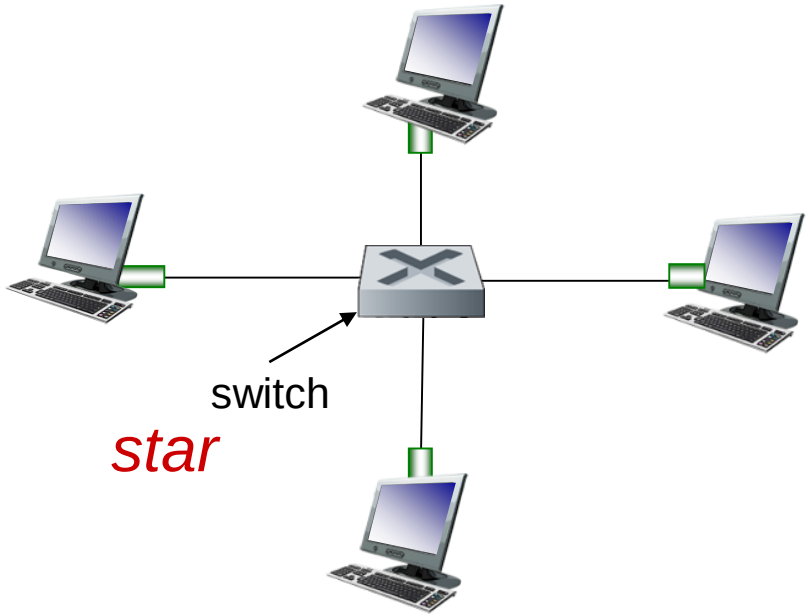


Ethernet: physical topology

- *bus:* popular through mid 90s
  - all nodes in same collision domain (can collide with each other)
- *star:* prevails today
  - active *switch* in center
  - each "spoke" runs a (separate) Ethernet protocol (nodes do not collide with each other)



bus: coaxial cable



star

Broadcast vs. switched Ethernet

- Invented as a broadcast technology
  - Hosts share channel
  - Each packet received by all attached hosts
  - CSMA/CD for media access control
- Modern Ethernets are "switched"
  - Point-to-point links between switches and between a host and switch
  - No sharing $\Rightarrow$ no CSMA/CD
    - Uses "self learning" and "spanning tree" algorithms for routing

The evolution of Ethernet

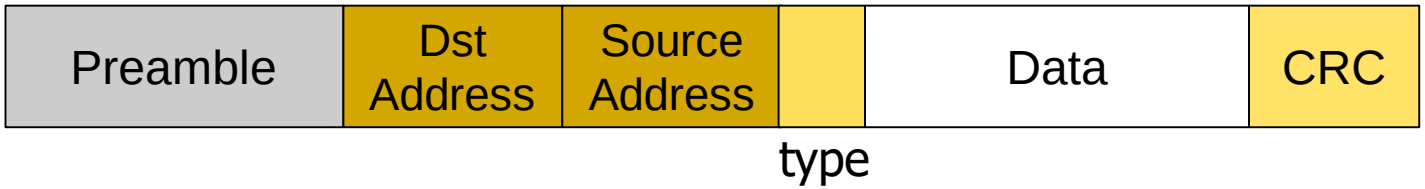
- Changed almost everything except the frame format
  - From the shared media coax cables to dedicated links
  - From 3 Mbit/s to 100 Gbit/s
  - From electrical signaling to optical
- Lesson: the right interface can accommodate many changes
  - Evolve the implementation while maintaining the interface (backward compatibility)

Ethernet: unreliable, connectionless

- *Connectionless:* no handshaking between sending and receiving NICs
- *Unreliable:* receiving NIC doesnt send acks or nacks to sending NIC
  - data in dropped frames recovered only if initial sender uses higher layer rdt (e.g., TCP), otherwise dropped data lost

Ethernet Frame Format

- Encapsulates IP datagram



- Preamble: 7 bytes for clock synchronization and 1 byte to indicate start of frame
- Addresses: 6 bytes
- Type: 2 bytes, higher-layer protocol (e.g., IP)
- Data payload: max 1500 bytes, min 46 bytes
- CRC: 4 bytes for error detection

Medium Access Control (MAC) Address

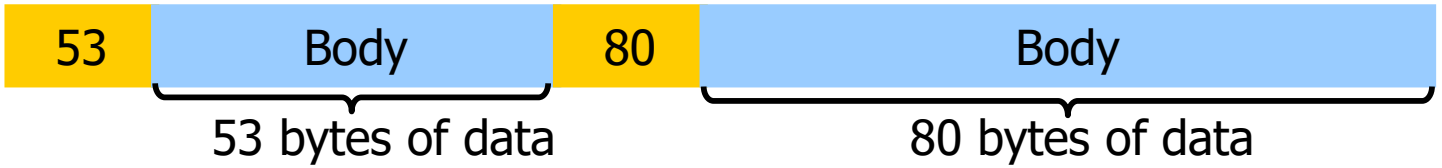
- MAC address
  - Numerical address associated with a network adapter
  - Flat name space of 48 bits (e.g., 00-15-C5-49-04-A9 in HEX)
  - Unique, hard-coded in the adapter when it is built
- Hierarchical Allocation
  - Blocks: assigned to vendors (e.g., Dell) by the IEEE
    - First 24 bits (e.g., 00-15-C5-**-**-**)
  - Adapter: assigned by the vendor from its block
    - Last 24 bits

Framing frames

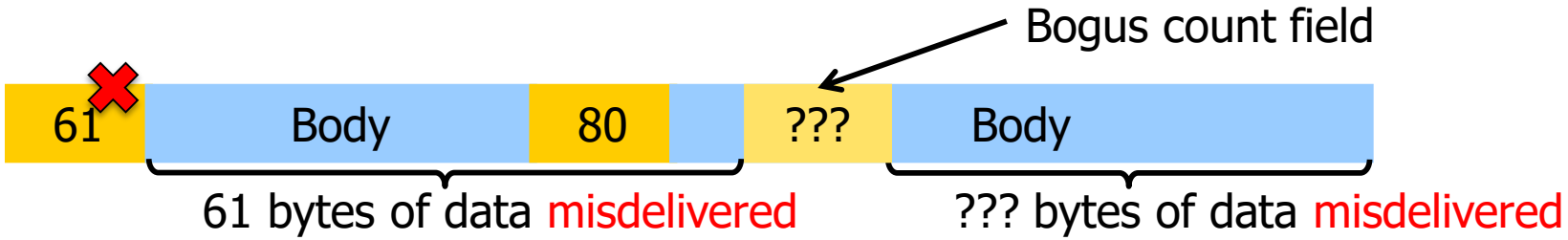
- Physical layer puts bits on a link
- But, two hosts connected on the same physical medium need to be able to exchange frames
  - Service provided by the link layer
  - Implemented by the network adaptor
- Framing problem: how does the link layer determine where each frame begins and ends?

Simple approach: Count bytes

- Sender includes number of bytes in header



  - Receiver extracts this number of bytes of body
- What if the Count field is corrupted?



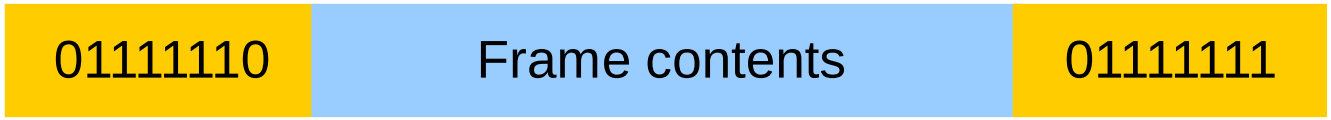
  - L2 will frame the wrong bytes → a framing error
  - CRC tells you to discard this frame, but what about the next one?

Desynchronization

- Once framing on a link is desynchronized, it can stay that way
- Need a method to resynchronize

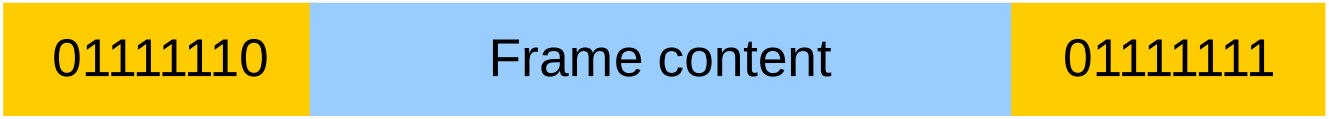
Framing with sentinel bits

- Delineate frame with special "sentinel" bit pattern
  - e.g., 01111110 ⇒ start, 01111111 ⇒ end



- What if sentinel occurs within frame?
- Solution: bit stuffing
  - Sender always inserts a 0 after five 1s in the frame contents
  - Receiver always removes a 0 appearing after five 1s

When receiver sees five 1s…

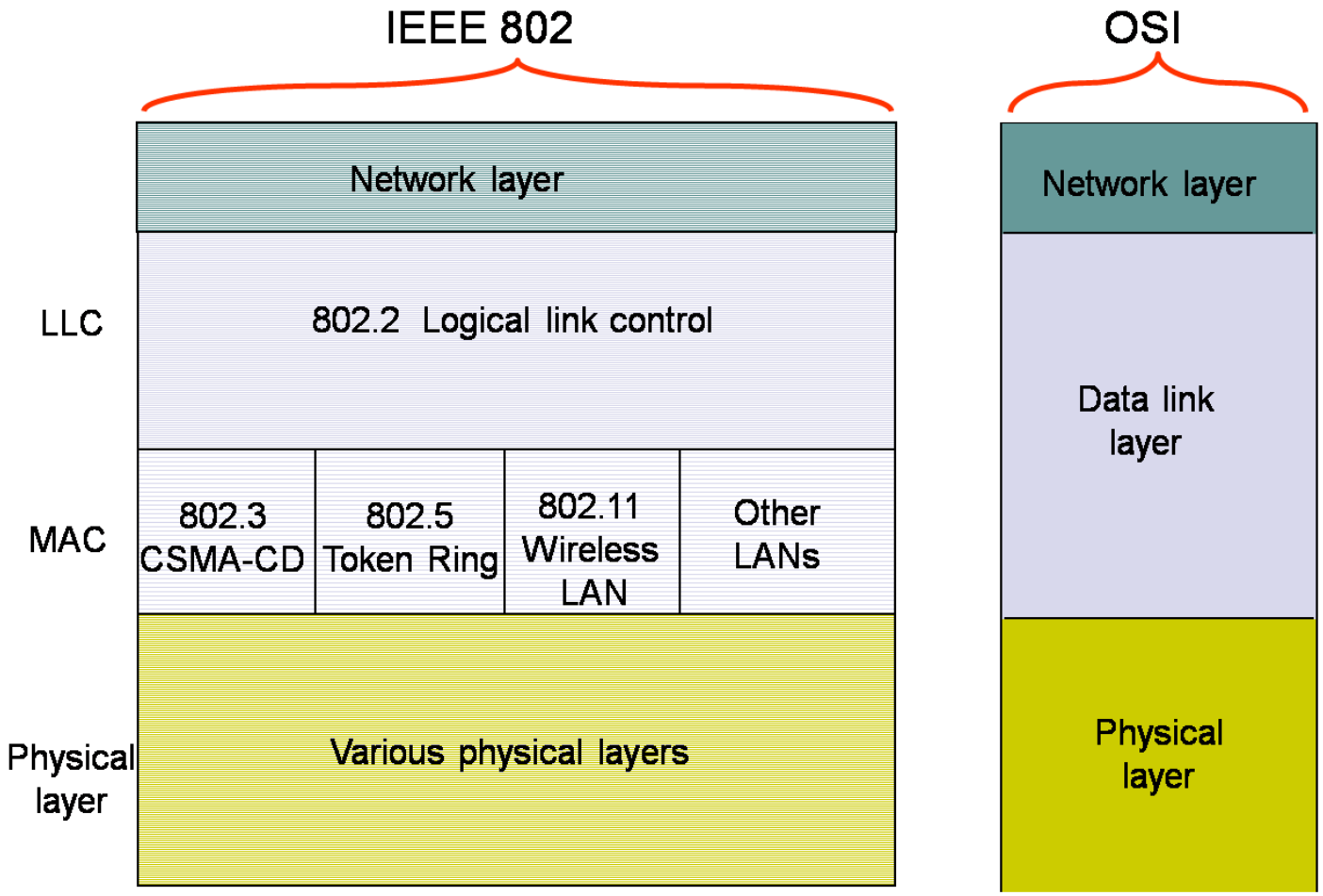


- If next bit 0, remove it; begin counting again
  - Because this must be a stuffed bit; we can't be at beginning/end of frame (those had six or seven 1s)
- If next bit 1 (i.e., we've seen six 1s) then:
  - If following bit is 0, this is start of frame
    - Because the receiver has seen 01111110
  - If following bit is 1, this is end of frame
    - Because the receiver has seen 01111111

Example: sentinel bits

- Original data, including start/end of frame:
  - 01111110011111101111101111100101111111
- Sender rule: five 1s → insert a 0
  - After bit stuffing at the sender:
  - 01111110011111010111110011111000101111111
- Receiver rule: five 1s and next bit 0 → remove 0
  - 01111110011111101111101111100101111111

IEEE 802.3


IEEE 802
OSI
Network layer
LLC
802.2 Logical link control
MAC
802.3 CSMA-CD
802.5 Token Ring
802.11 Wireless LAN
Other LANs
Physical layer
Various physical layers
Network layer
Data link layer
Physical layer

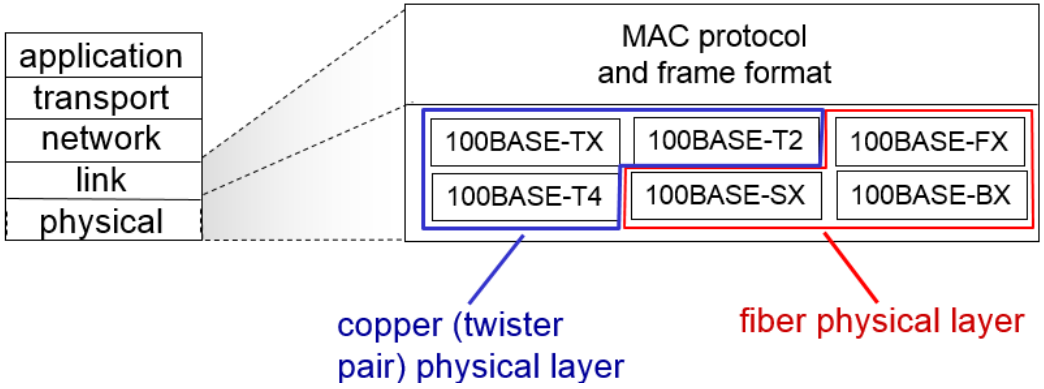
802.3 Physical Layer

- *many* different Ethernet standards
  - common MAC protocol and frame format
  - different speeds: 2 Mbps, 10 Mbps, 100 Mbps, 1Gbps, 10G bps
  - different physical layer media: fiber, cable

speed

medium

	10base5	10base2	10baseT	10baseFX
Medium	Thick coax	Thin coax	Twisted pair	Optical fiber
Max. Segment Length	500 m	200 m	100 m	2 km
Topology	Bus	Bus	Star	Point-to-point link



Summary

- 链路层服务
- 错误检测：奇偶校验，CRC的计算
- 流控制：Stop and Wait，Sliding Window
- 局域网
  - 令牌环
  - 以太网
    - 以太网帧格式