Lecture 3: Link Layer. Ethernet.

Acknowledgement: Materials presented in this lecture are predominantly based on slides from:

- *Computer Networking: A Top Down Approach,* J. Kurose, K. Ross, 7th ed., 2017, Pearson, Chapter 3
- J. FitzGerald, A. Dennis, A. Durcikova : Business Data Communications and Networking, 12th ed., 2014, John Wiley & Sons, Chapter 5
- **IEEE 802.3 Standard**

Lecture 3: Link Layer. Ethernet

Overview:

- Link layer services
- Revision of the 802.11 and LTE link layers
- About Ethernet
- Ethernet frame format
- Frame Check Sequence
- Shared Ethernet
- CSMA/CD MAC protocol
- Switched Ethernet
- Inside switches
- Forwarding table and learning procedure
- Modes of Switch Operations
- Connecting switches
- Unicast, multicast broadcast addresses
- IEEE802.3 Standards. Overview
  - Gigabit Ethernet

Link layer: introduction

- Data-link layer has responsibility of transferring IP packets from one node to physically adjacent node over a wired or wireless link
- The link layer is also used to connect a **group of** computers into a Local Area Network (LAN) creating a **subnet**.
- In FIT5187 you have discussed two examples of wireless **data link** layers,
  - WiFi – IEEE802.11
  - LTE
- The most common wired data link layer protocol is the Ethernet – IEEE 802.3 standard

Link layer services

All data link layers:

- Encapsulates IP (higher layer) packets into frames adding
  - **Headers** containing the source and destination MAC (physical) addresses
  - **Trailers** containing the Frame Check Sequence (FCS)
    - The FCS is typically not used in the high reliability wired links (Fibre optics, twisted pair cables),
    - FCS (**error detection**) and **error correction** are essential in the unreliable wireless links.
- Organizes access to **shared** media (Media Access Control), in wireless and the wired Ethernet, in particular

802.11 MAC and PHY layers

- The structure consists of three main parts: data part, management part and optional security/802.1X part
- Data Link and PHY layers are organized into sublayers.
- **MAC** (Medium Access Control)
- **PLCP** (Physical Layer Convergence Procedure)
- **PMD** (Physical Medium dependant)
- Sublayers and parts communicate through SAPs – Service Access Points

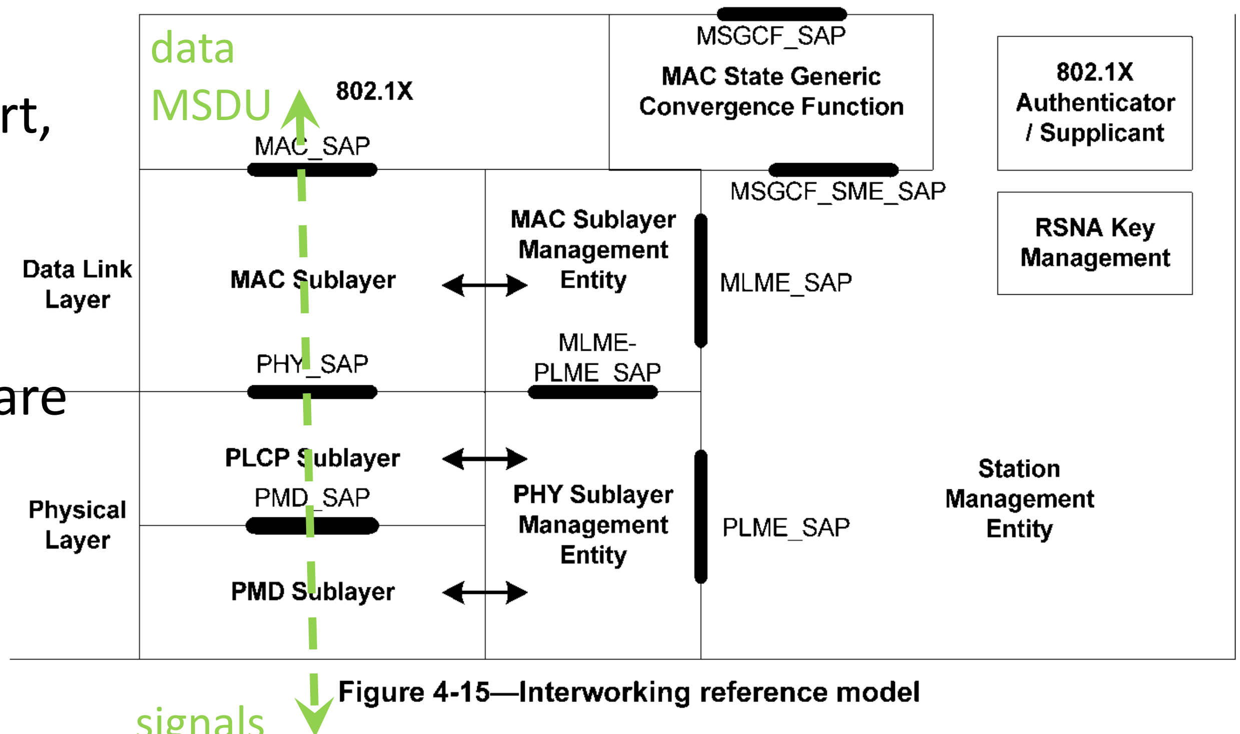


Figure 4-15—Interworking reference model

- Data arrives at the MAC_SAP organized into MSDU – MAC layer Service Data Unit
- **MSDU** flows through MAC and PHY sublayers and is converted into **PHY Signals**

LTE – Link Layer

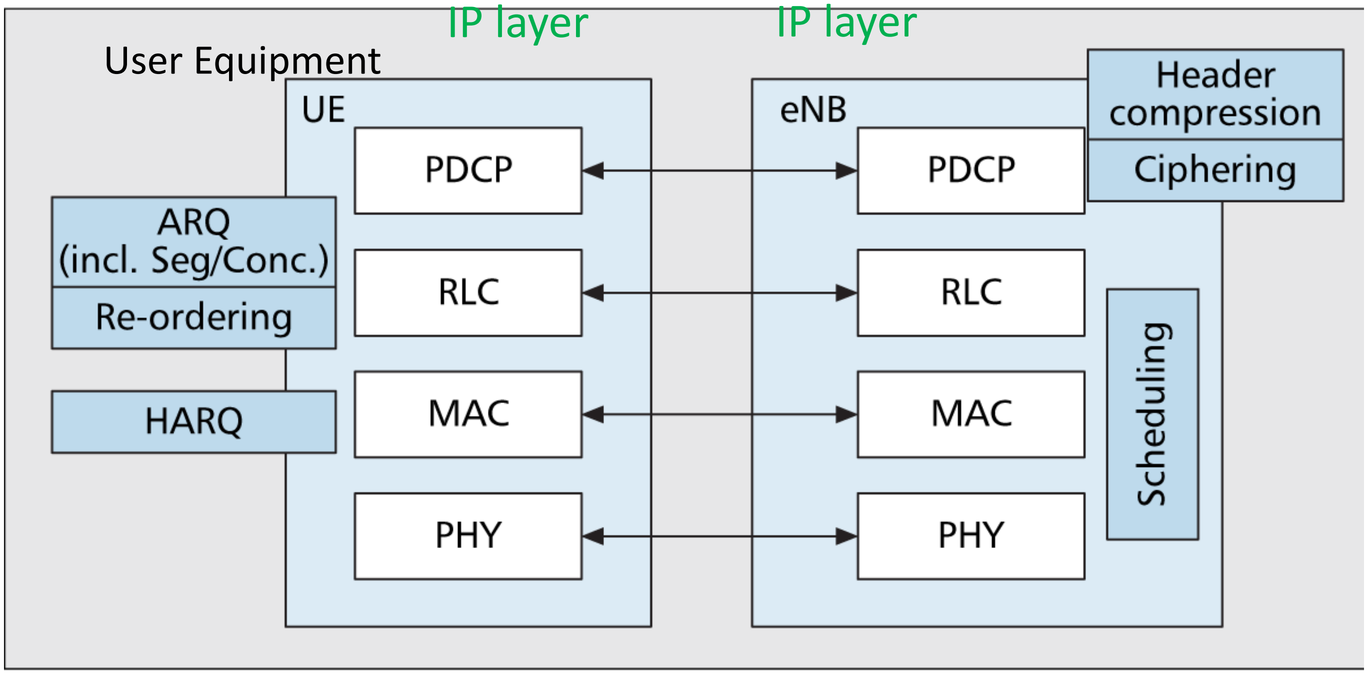


Figure 2. *User plane protocol stack.*

The LTE link layer consists of three sublayers:

- The Packet Data Convergence Protocol (PDCP)
- The radio link control (RLC)
- The medium access control (MAC)

PDCP is responsible for IP header compression and ciphering.

RLC comprises: **ARQ** (Automatic Repeat reQest) functionality and

- supports data segmentation and concatenation

MAC provides **HARQ** (Hybrid ARQ) and is responsible for

- medium access control
- scheduling operation and random access.

Ethernet: Introductory comments

- Used by all LANs today. Implemented in the Network Adapters aka Network Interface Cards (NIC)
- Originally (1983) developed by a consortium of Digital Equipment Corp., Intel and Xerox
- Standardized as IEEE 802.3
- Types of Ethernet
  - Original, Shared Ethernet (not used in its original form)
    - Used coaxial cables or (Ethernet) **hubs**. Computers in the LAN share the medium (circuit)
  - Switched Ethernet (dominant)
    - Uses (Ethernet) **switches**. Each computer in the LAN has its own circuit
  - Ethernet is used not only in LANs, but also in the Metropolitan and Wide Area Networks (MAN and WAN)

IEEE Std 802.3-2015: Ethernet Standards

- The current version of standards has been published in 2015.
- Clauses 3 and 4 of the **section 1** are on Moodle (intro: local copy)
- It consists of 6 volumes called sections
- Each section/volume contains a number of chapters called clauses and Appendixes
- **Section One**, Clauses 1 … 20, includes the specifications for **10 Mb/s operation** and **the MAC, frame formats and service interfaces used for all speeds of operation**.
- **Section Two**, Clauses 21 … 33, includes … also general information on **100 Mb/s operation** as well as most of the 100 Mb/s Physical Layer specifications.
- **Section Three**, Clauses 34 … 43, includes general information on **1000 Mb/s** operation as well as most of the 1000 Mb/s Physical Layer specifications.

Cont …

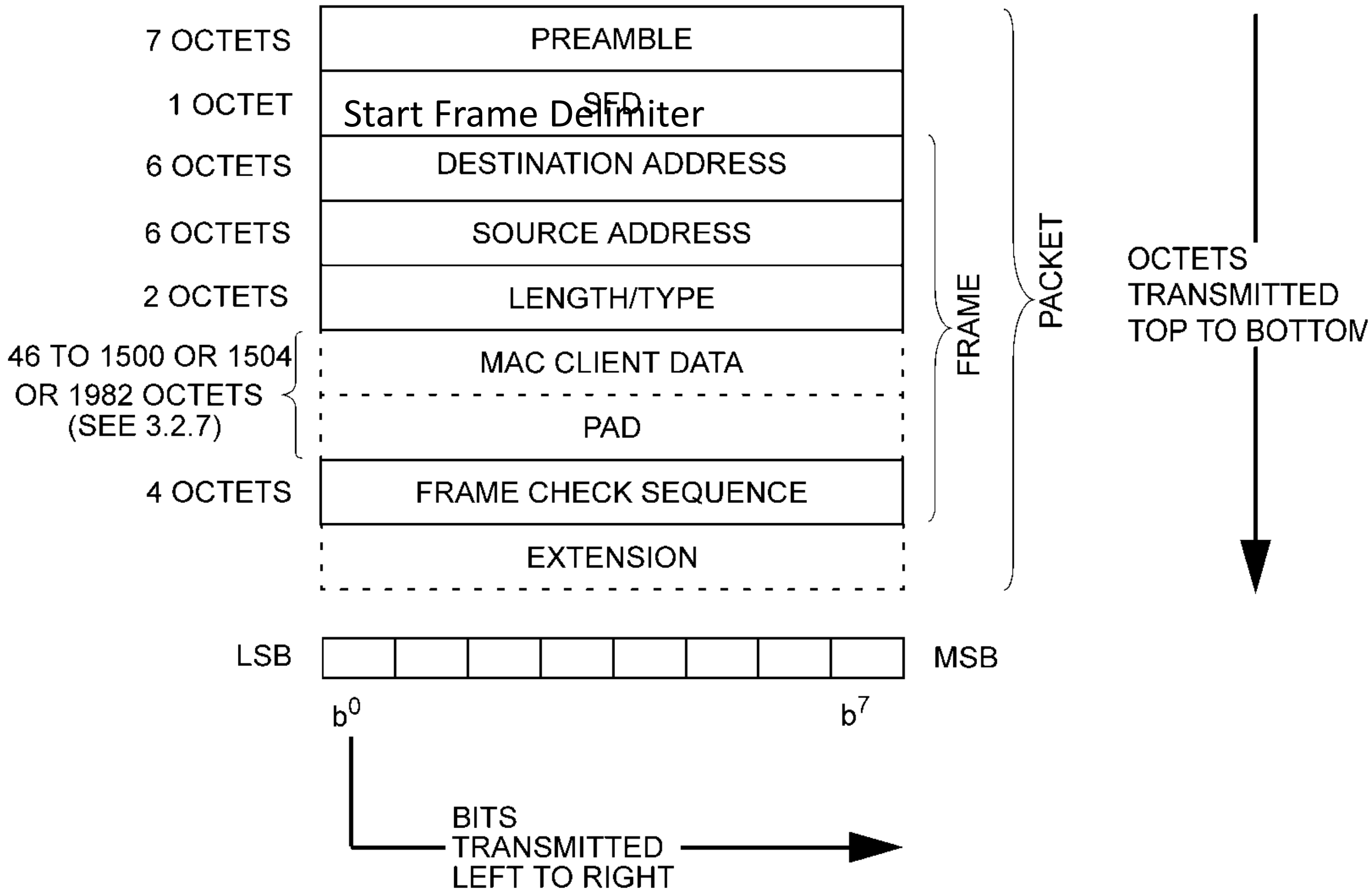
- **Section Four**, Clauses 44 … 55, includes general information on **10 Gb/s** operation as well as most of the 10 Gb/s Physical Layer specifications.
- **Section Five**, Clauses 56 … 77 specify ….

  Clause 68 specifies a **10 Gb/s** Physical Layer specification.

  …
- **Section Six**, Clauses 78 … 90, specifies …

  Clause 80 through Clause 89 and associated annexes includes general information on **40 Gb/s** and **100 Gb/s** operation as well the 40 Gb/s and 100 Gb/s Physical Layer specifications….
- IEEE 802.3 timeline

Ethernet packet and frame format

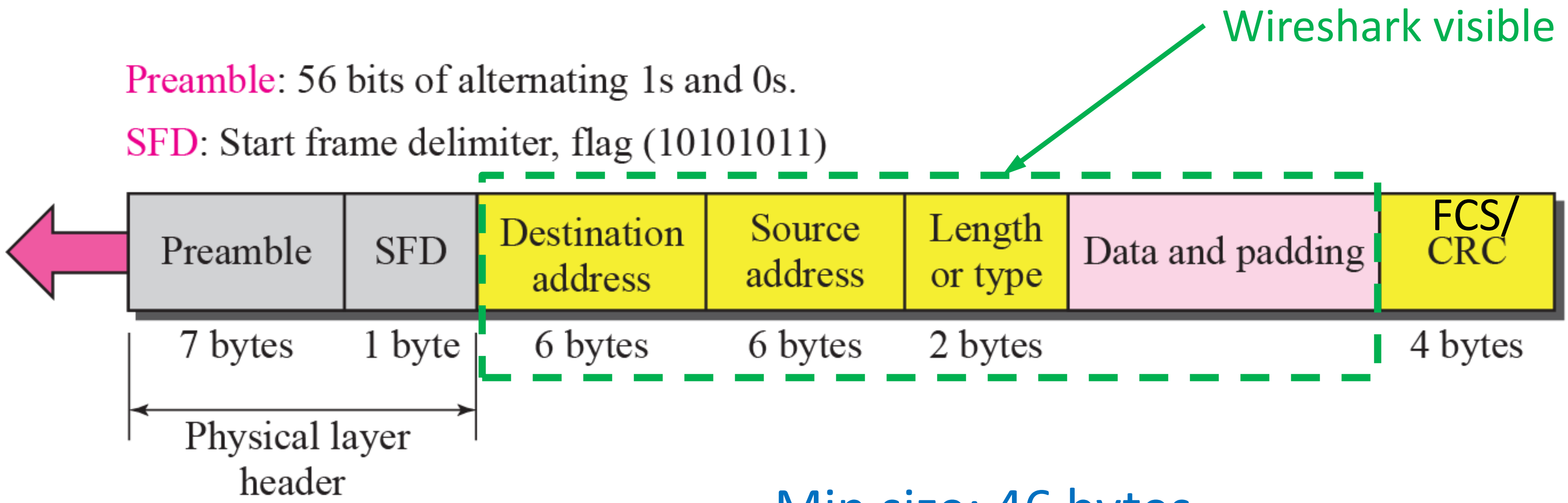


As described in Std IEEE 802.3, sec. 3.1.1 (Moodle)

(local copy)

Note the frame encapsulated in the packet

Ethernet frame. Another view



Min size: 46 bytes

PAD bits to make min 46 bytes

Be aware that bits in all bytes apart from FCS are being transmitted LSB first

max size in bytes:

1500 decimal—basic frames

1504 decimal—Q-tagged frames

1982 decimal—envelope frames

Length/type field

- **Length/type**, a 2-byte field, takes one of two meanings, depending on its numeric value.
- Length interpretation: If the value of this field is less than or equal to 1500 decimal (0x05DC hexadecimal), then the Length/Type field indicates the number of MAC client data octets contained in the subsequent MAC Client Data field of the basic frame .
- Type interpretation: If the value of this field is greater than or equal to 1536 decimal (0x0600 hexadecimal), then the Length/Type field indicates the Ethertype of the MAC client protocol.

  Examples of most common Ethertypes that you have to remember:
  - `0x`**`0800`**   the frame contains IPv4 packet.
  - `0x`**`86DD`**   the frame contains IPv6 packet.
  - `0x`**`0806`**   indicates an ARP frame (try a Wireshark example)

Address fields

- Each address field is 48 bits (6 bytes) in length.
- Each octet of each address field is transmitted LSB first.
- **The first bit** (LSB) in the **Destination Address** identify the address type:
  - 0 – an individual address (unicast)
  - 1 – a group address indicating that the address field contains a group address that identifies none, one or more, or all of the stations connected to the LAN (multicast)
- In the **Source Address** field, the first bit is reserved and set to 0 (unicast).
- **The second bit is used to** distinguish between locally or globally administered addresses:
  - 0 – for globally administered (or U, universal) addresses,
  - 1 – for locally assigned addresses the bit is set to 0.
- **Broadcast Destination Address contains all 1's**

Addresses Exercise

Define the type of the following **destination** addresses:

a. 4**A**:30:10:21:10:1A

b. 4**7**:20:1B:2E:08:EE

c. FF:FF:FF:FF:FF:FF

Solution

- To find the type of the address, we need to look at the second hexadecimal digit from the left:
- If it is even, the address is unicast (A source address is always unicast)
- If it is odd, the address is multicast.
- If all digits are F's, the address is broadcast.

Therefore, we have the following:

a. This is a unicast address because A in binary is 1010 (even).

b. This is a multicast address because 7 in binary is 0111 (odd).

c. This is a broadcast address because all digits are F's.

Frame Check Sequence (FCS)

- The FCS field contains a 4-byte (32-bit) CRC value. (Clause 3.2.9 , Moodle)
- FCS value is computed using the bits of the frame without FCS
- The encoding is defined by the following generating polynomial:

$$G(x) = x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^{8} + x^{7} + x^{5} + x^{4} + x^{2} + x + 1 = \text{0x}\mathbf{04C11DB7}$$

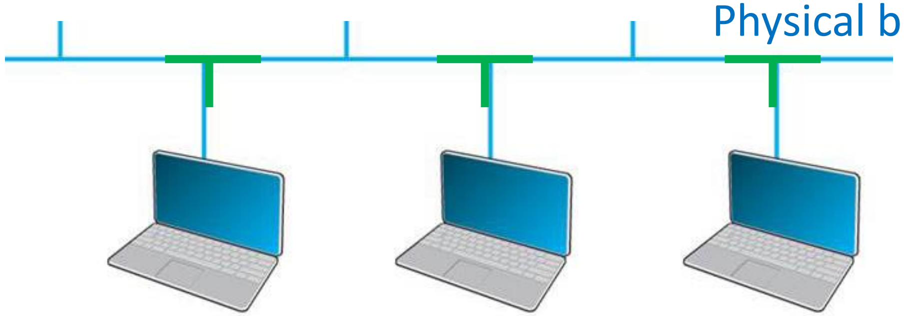
The CRC value is defined by the following procedure:

- The first 32 bits of the frame are complemented.
- The n bits of frame (without FCS) are then considered to be the coefficients of a polynomial $M(x)$ of degree $n - 1$.
- $M(x)$ is multiplied by x^{32} and divided by $G(x)$, producing a remainder $R(x)$ of degree ≤ 31.
- The coefficients of $R(x)$ are considered to be a 32-bit sequence.
- The bit sequence is complemented and the result is the CRC.

➢ **Compare with the 802.11 FCS!**

Coaxial Cable Shared Ethernet (historical)

Original, shared Ethernet, 10BASE5, used a coaxial cable to connect together NICs of the participating computers using a T-junction connectors



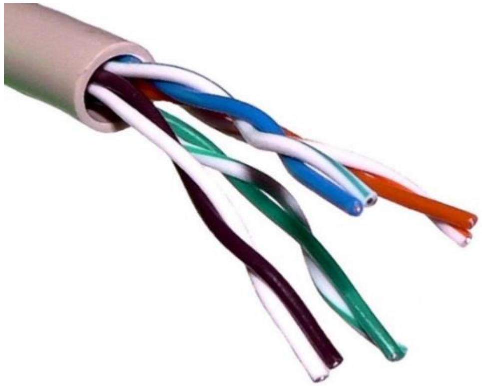
- All computers **share the link**, hence the need for the medium access control (MAC) protocol.
- Each computer receives messages from all other computers, whether the message is intended for it or not.

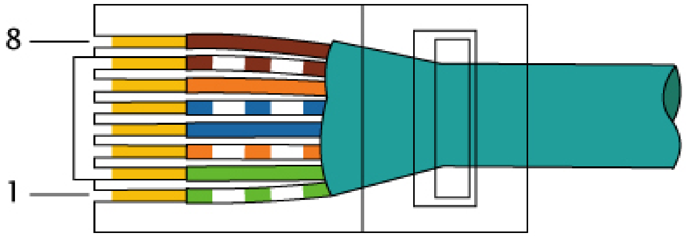
Ethernet uses a contention-based protocol known as

- CSMA/CD – Carrier Sense, Multiple Access, Collision Detection
- Bits on the link were encoded using the Manchester code.
- **10BASE5** – 10Mbps speed, BASE – baseband (from 0Hz, unmodulated) frequency range (unlike WiFi, LTA, …), 5 stands for 500 m total length of the cable.

Twisted Pair cables and connectors

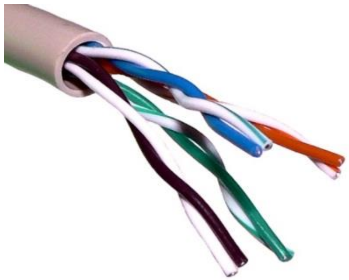
- Unshielded Twisted Pair (UTP) cables are used in the most popular Ethernet standards:
  - 10BASE-T, 100BASE-TX (Fast Ethernet), and 1000BASE-T (Gigabit Ethernet). Note the speed and `T' for twisted pair
- Depending on their electrical parameters the UTP cables are branded as: **cat**egory 5, 5e, 6 …
- Typically there are about 6 twists per 100mm
- The cable contains **4 twisted pairs** and the plugs/connectors are known as RJ45 or as 8P8C and the wiring method is known as T568A

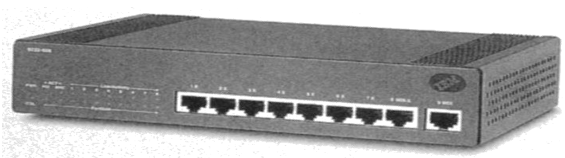


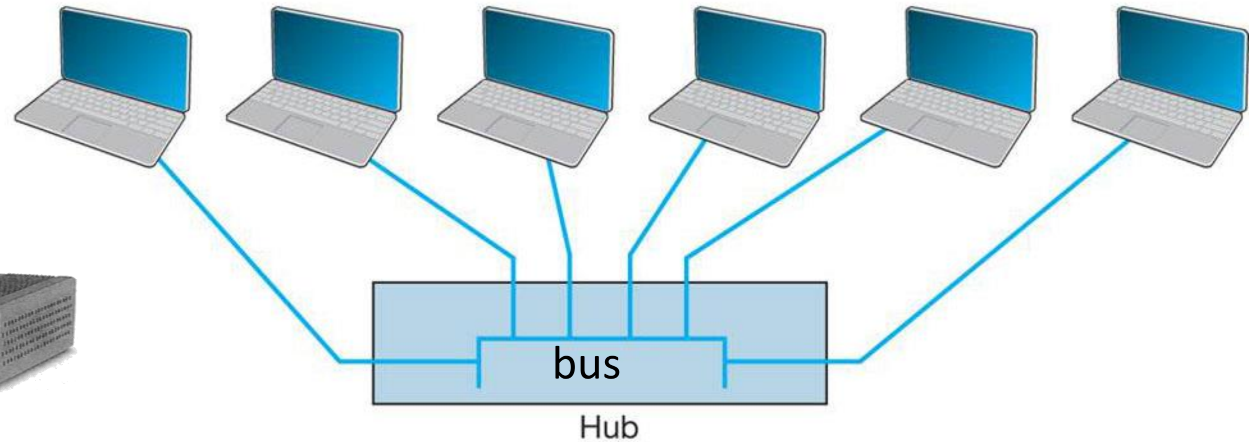


EIA/TIA-568A

UTP Shared Ethernet: Hubs and Repeaters







- The **Ethernet hub** is a **layer-1** device connecting data wires from each computers together thus forming a shared bus.
- The hub acts as a junction box, simplifying connections of the UTP cables
- The hub contains the **repeaters** which **regenerate** (reconstruct and strengthen) incoming signals that become weaker and distorted with the distance.
- As in the coaxial case, all computers can access the link **simultaneously** and receive messages from **all other computers**, whether the message is intended for it or not.

Optical fibre cables and connectors

- [Optical fiber/fibre aka fibre optics]() cables
- [Optical fiber connector]()

- Jump to slide 34 (PHY)
- Return to slide 21 if time permits

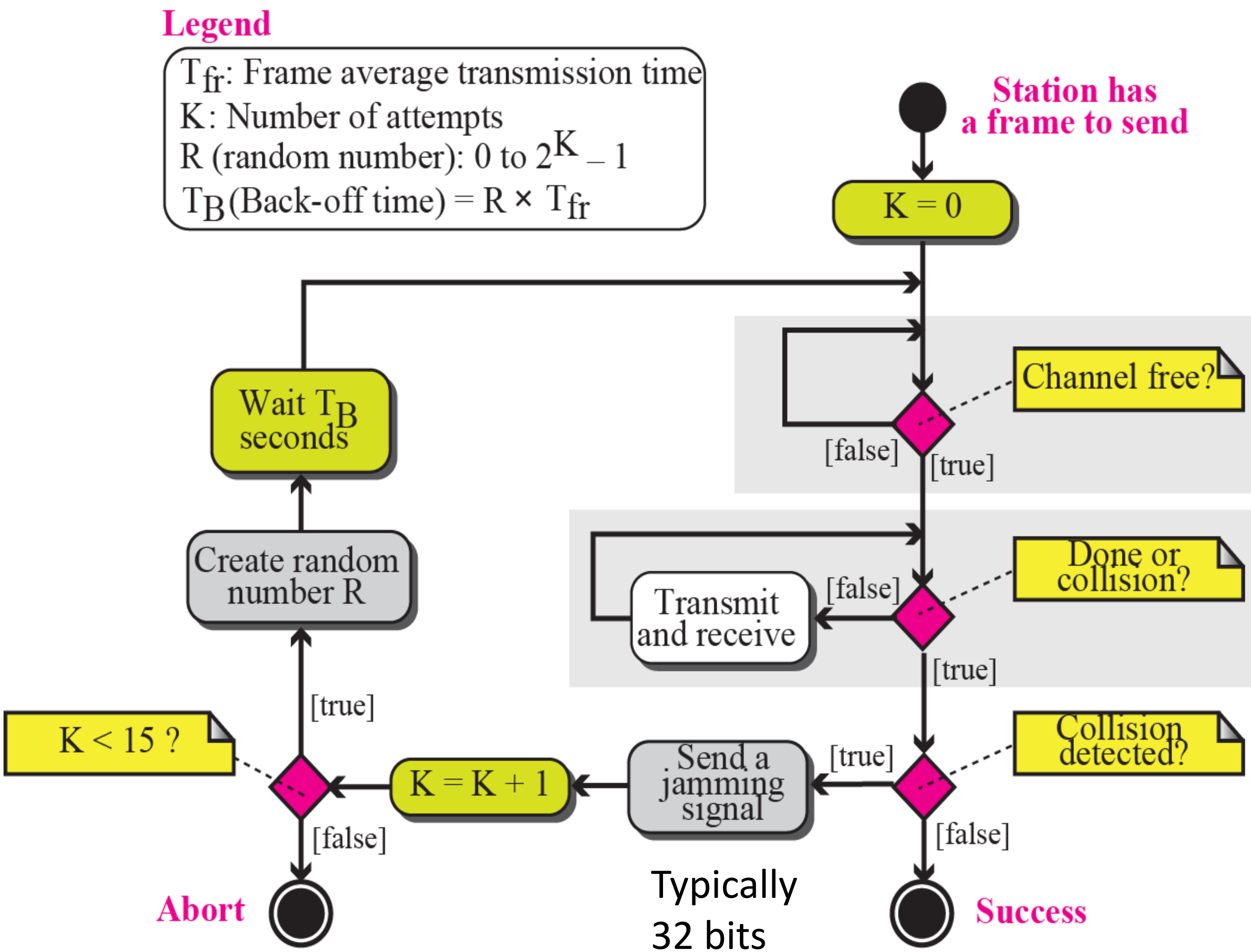
Media Access Control (MAC)

- If a single computer sends a frame, this frame is received by all computers sharing the circuit.
- The first task in each computer is to read the frame's destination address to see if the message is meant for it or not
- Frames can be sent by two computers on the same network/circuit at the same time
  - They will collide and destroy each other
  - The MAC protocol is designed to detect and resolve the contentions
- The contention-based protocol designed initially for the shared Ethernet is called **CSMA/CD** (Carrier Sense Multiple Access/Collision Detection) (Recall WiFi's CSMA/CA)

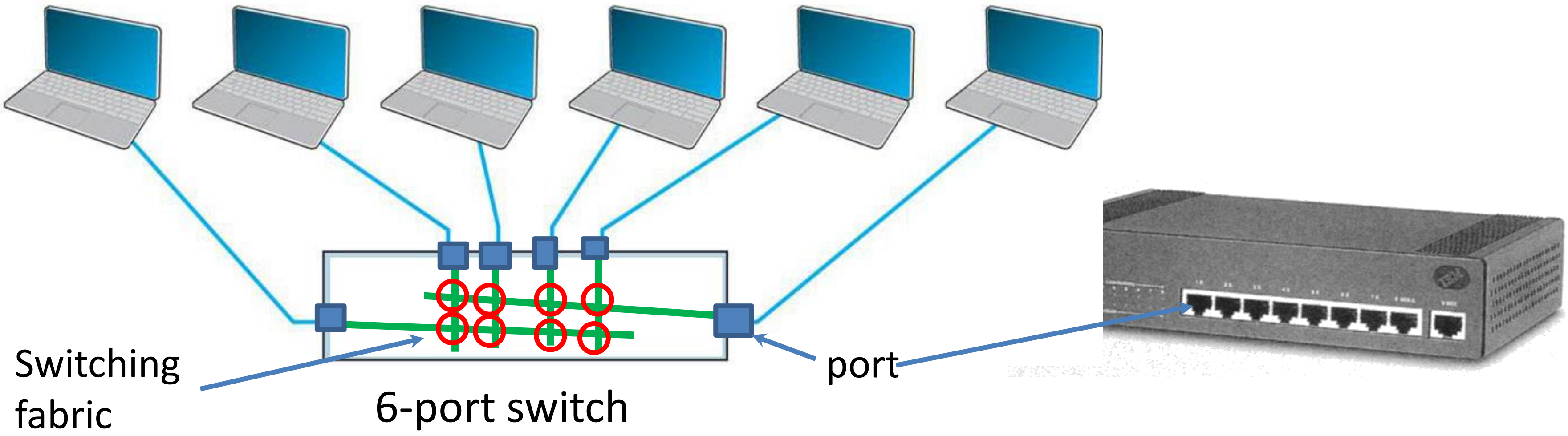
CSMA/CD MAC protocol

- Carrier Sense (CS):
  - Before sending anything, a computer listens to the bus to determine if another computer is transmitting
  - Transmit when no other computer is transmitting
- Multiple Access (MA):
  - All computers have access to the network medium
- Collision Detection (CD):
  - is declared when any signal other than its own is detected
  - If a collision is detected
    - To avoid a collision, all computer intending to transmit wait a random amount of time and then resend message
    - The computer with the smallest amount of the random time will transmit first.

CSMA/CD (simplified) flowchart

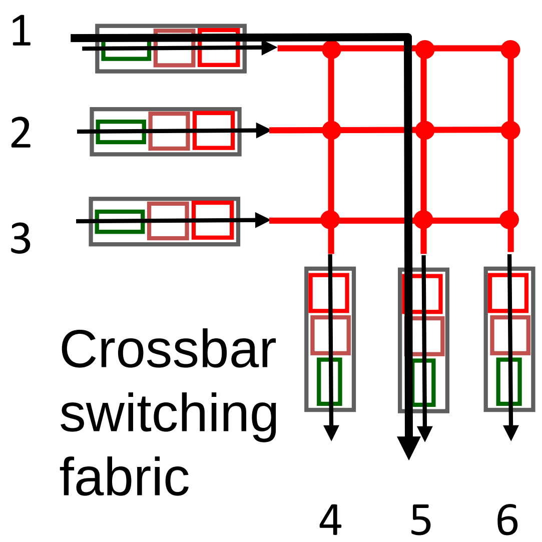


Switched Ethernet



- Switched Ethernet uses **switches** aka **multiport bridges**
  - An Ethernet Switch looks similar to a hub, but is very different inside
  - Designed to support a group of point-to-point circuits
    - No sharing of circuits
- The network has a **star topology** via the switch
- The **switch reads destination** address of the frame and only sends it to the corresponding port
  - while a **hub broadcasts** frames to all ports
- A switch is a **layer 2** device since it operates with the MAC addresses

Inside a switch

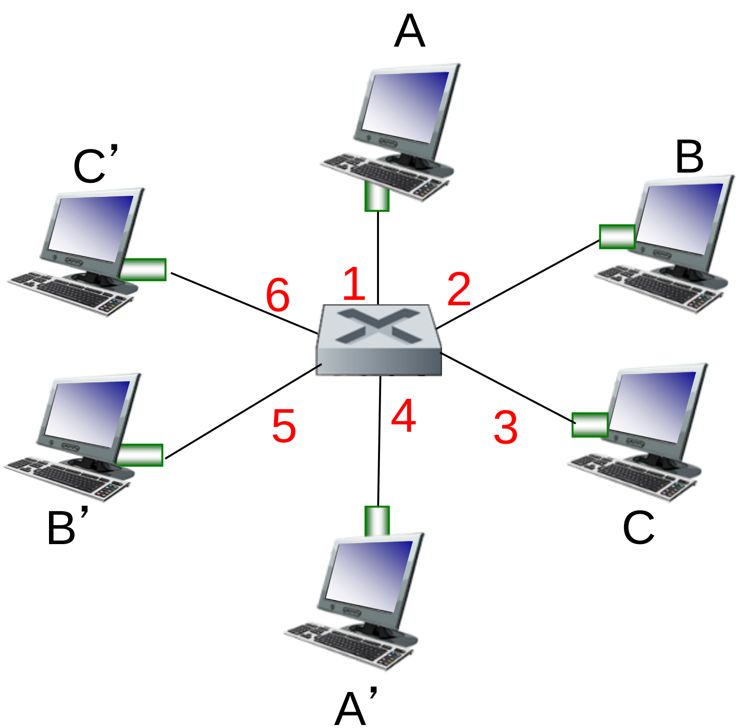


Crossbar switching fabric

- Internal structure of the switch is typically based on the crossbar switching fabric
- It allows simultaneous connections without collisions
- 1 can talk to 5, 2 to 3, 5 to 6 at the same time
- Note that the all the **ports** needs to be associated with the **source MAC** addresses

Switch: *multiple* simultaneous transmissions

- Same as previous slide
- **hosts** are connected to the switch through port aka interfaces
- The switch redirects packets according to MAC addresses
- Ethernet protocol used on *each* incoming link, but no collisions; full duplex
- *switching:* A-to-A' and B-to-B' can transmit simultaneously, without collisions



A switch with six interfaces/ports (1,2,3,4,5,6)

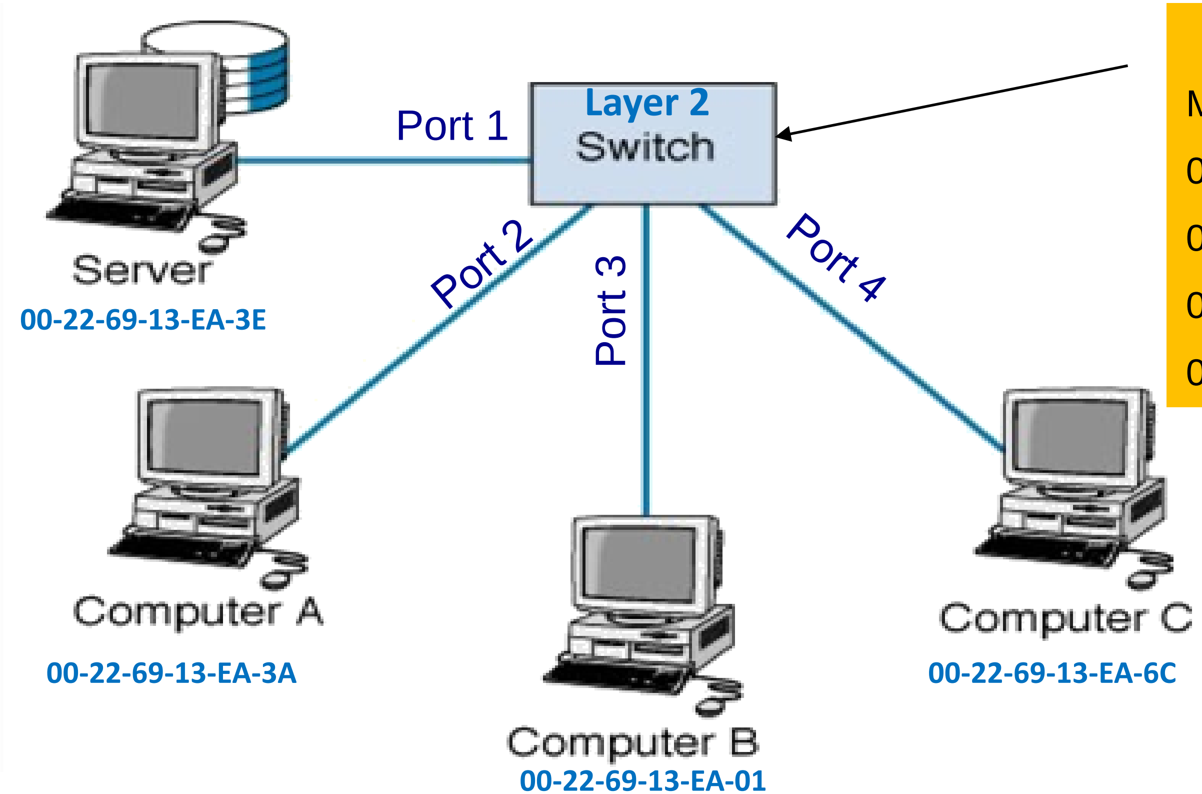
Forwarding Tables in the Switch

- The Ethernet switch creates and maintains the forwarding table
  - Lists the Ethernet address of computers connected to each port of the switch: e.g.

Port 1 ↔ 00-22-69-13-EA-3E

- When a frame is received, the switch reads its Layer 2 MAC destination address and sends the frame out to the corresponding port in its forwarding table.
- Recall that the Ethernet frame has: | DST_{PHY} | SRC_{PHY} | |

Forwarding table



Forwarding Table

MAC address	Port
00-22-69-13-EA-3E	1
00-22-69-13-EA-3A	2
00-22-69-13-EA-01	3
00-22-69-13-EA-6C	4

In order to work correctly the switch needs a Forwarding Table
The table associates the MAC addresses of each NIC (Network Interface Card) with the equivalent port number

Learning Switch Operation

Forwarding Table	
MAC address	Port
00-22-69-13-EA-3E	1
00-22-69-13-EA-3A	2
00-22-69-13-EA-01	3
00-22-69-13-EA-6C	4

- **A Switch starts** by working like a simple hub using the CSMA/CD MAC procedure with an **empty forwarding table**
- It gradually fills its forwarding table by learning about the nodes
  - Reads the source MAC address of the incoming frame and records it to the corresponding port number
  - Reads the destination MAC address. If not in the Table then it broadcasts the frame to all ports
  - Waits for the destination computers to respond, and repeats the first step
- *Switches are plug-and-play, self-learning* devices and do not need to be configured manually

Modes of Switch Operations

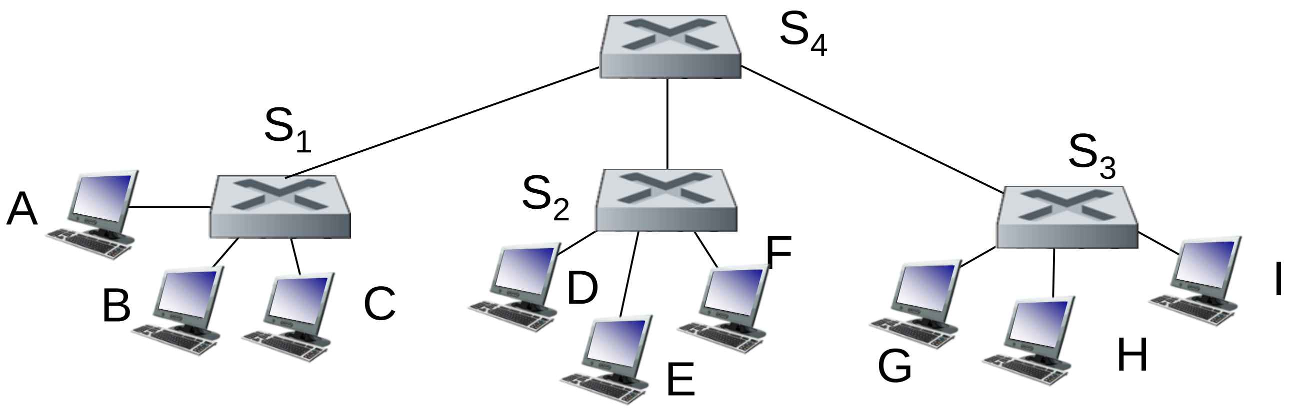
1. **Cut through switching**
   - Reads destination address and starts transmitting without waiting for the entire message to be received
   - Low latency; but may waste capacity (messages with errors)
   - Only on the same speed incoming and outgoing circuits
2. **Store and forward switching**
   - Waits until the entire frame is received, perform error control, and then transmit it
   - Less wasted capacity; slower network
   - Circuit speeds may be different
3. **Fragment free switching**
   - Reads the first 64 bytes (contains the header)
   - Performs error checking; if it is OK then begins transmitting
   - It is a compromise between previous two modes

MAC in Switched Ethernet

- Each circuit shared by a computer and the switch
- Still uses CSMA/CD media access control
  - Each device (computer or switch) listens before transmitting
- Multiple messages can be sent at the same time.
  - Computer A can send a message to computer B at the same time that computer C sends one to computer D
  - If two computers send frames to the same destination at the same time
  - Switch stores the second frame in memory until it finishes sending the first, then forwards the second

Interconnecting switches

- switches can be connected together

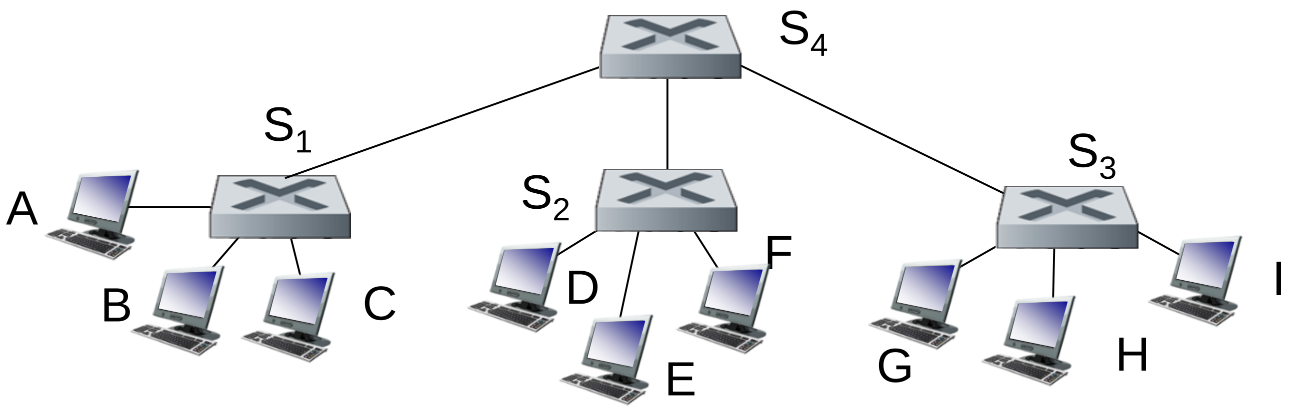


Q: sending from A to G – how does S_1 know to forward frame destined to F via S_4 and S_3?

A: self learning! (works exactly the same as in single-switch case!)

Self-learning multi-switch example

Suppose C sends frame to H, H responds to C



Q: show switch tables and packet forwarding in S_1, S_2, S_3, S_4

Popular Ethernet Types

Using UTP cables:

- 10BASE-T (Maximum data rate: 10 Mb/s, Baseband (no modulation) T – UTP cables, originally used **hubs**.
- 100BASE-T aka "Fast Ethernet": 100Mb/s, Baseband, UTP (Category 5), uses switches,
- 1000BASE-T aka "Gigabit Ethernet": 1Gb/s, Baseband, UTP (Category 5 or better), uses switches, uses **four pairs** of UTP cables
- 5GBASE-T similar to the above. Different details.

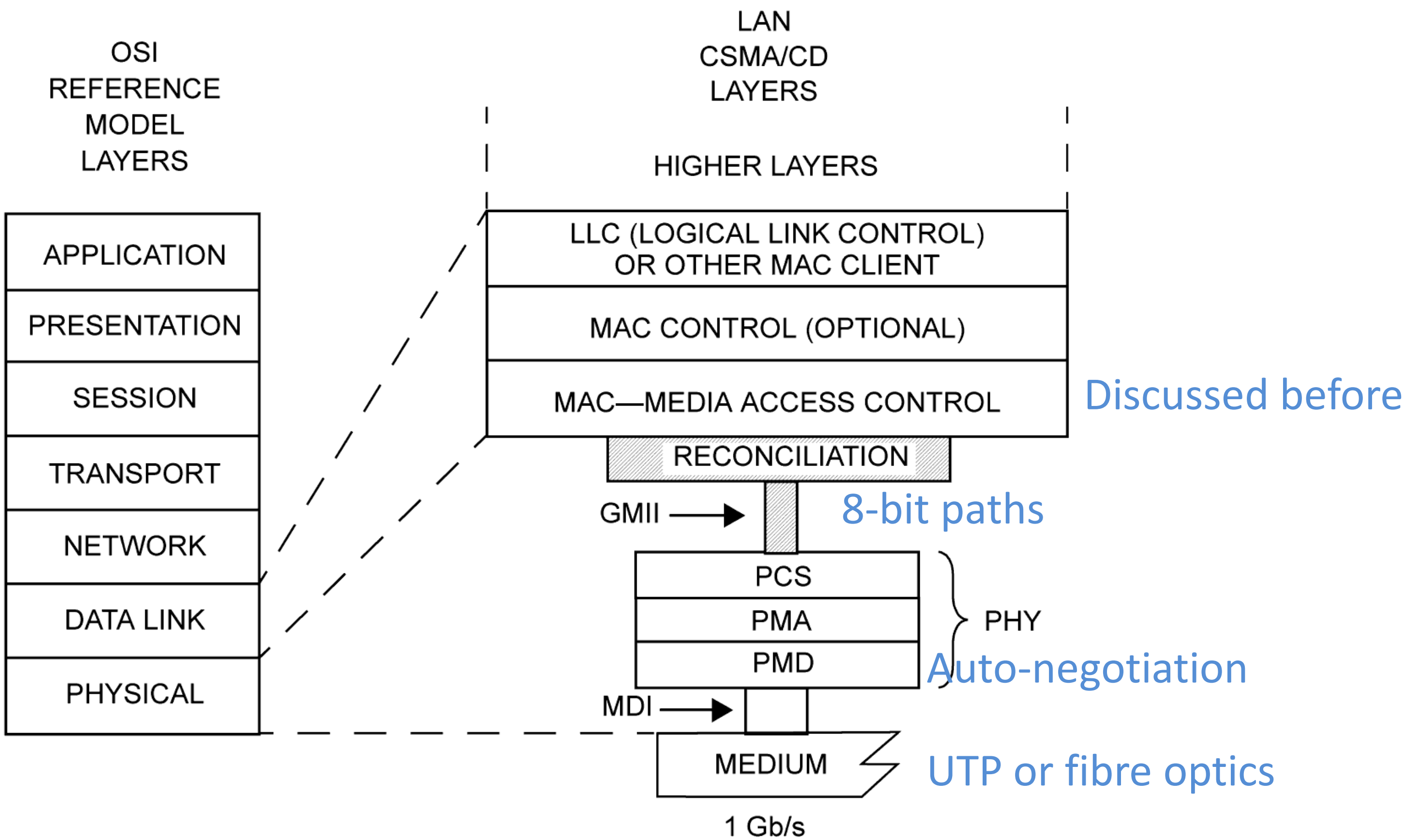
Gigabit Ethernet. General comments

Gigabit Ethernet uses

- the extended Ethernet MAC layer interface, connected through
- a Gigabit Media Independent Interface (GMII) layer to
- Physical Layer entities (PHY sublayers) such as
  - 1000BASE-LX, (a pair of fibres)
  - 1000BASE-SX, (a pair of fibres)
  - 1000BASE-CX, (a pair of fibres)
  - 1000BASE-T (**four** pairs of copper wires)

Gigabit Ethernet (section 3)

Data Link, Physical sublayers and related interfaces:



GMII = GIGABIT MEDIA INDEPENDENT INTERFACE
MDI = MEDIUM DEPENDENT INTERFACE
PCS = PHYSICAL CODING SUBLAYER

PHY = PHYSICAL LAYER DEVICE
PMA = PHYSICAL MEDIUM ATTACHMENT
PMD = PHYSICAL MEDIUM DEPENDENT

GMII and Reconciliation sublayers

- The Gigabit Media Independent Interface (GMII) provides an interconnection between
  - the Media Access Control (MAC) sublayer and
  - Physical Layer entities (PHY) and between
  - PHY Layer and Station Management (STA) entities.
- This GMII supports 1000 Mb/s operation through its eight bit wide transmit and receive paths.
- The Reconciliation sublayer provides
  - a mapping between the signals provided at the GMII and
  - the MAC/PLS service definition.

Auto-Negotiation

- Auto-Negotiation is used by 1000BASE-T/X devices to:
  - detect the abilities (modes of operation) supported by the device at the other end of a link segment,
  - determine common abilities, and configure for joint operation.
- Auto-Negotiation is performed upon link startup through the use of a special sequence of fast link pulses.

Physical Layer of 1000BASE-T: Coding

- The aggregate data rate of 1000 Mb/s is achieved by transmission at a data rate of 250 Mb/s over each wire pair (**four pairs** in the cable)
- Symbols (coded bits) can be transmitted and received on the same wire pairs at the same time enabling full duplex transmission
- **Each byte** is converted into **four quinary symbols** taken from the set

  {2, 1, 0, –1, –2}

- five-level Pulse Amplitude Modulation (PAM5)

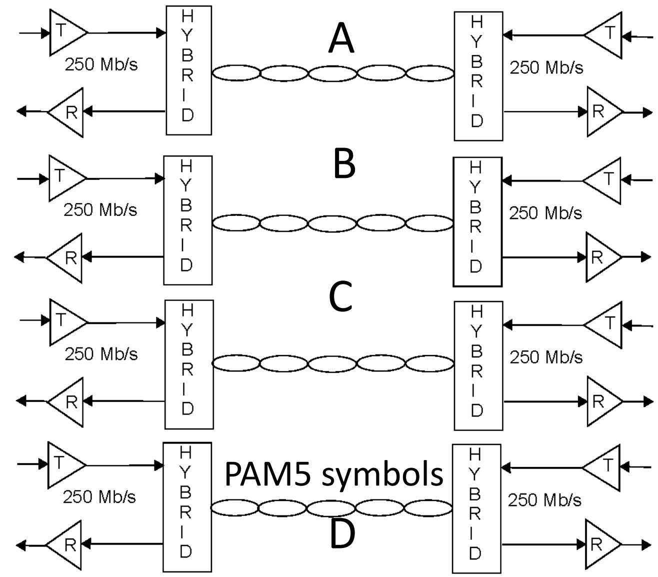


Figure 40–2—1000BASE-T topology

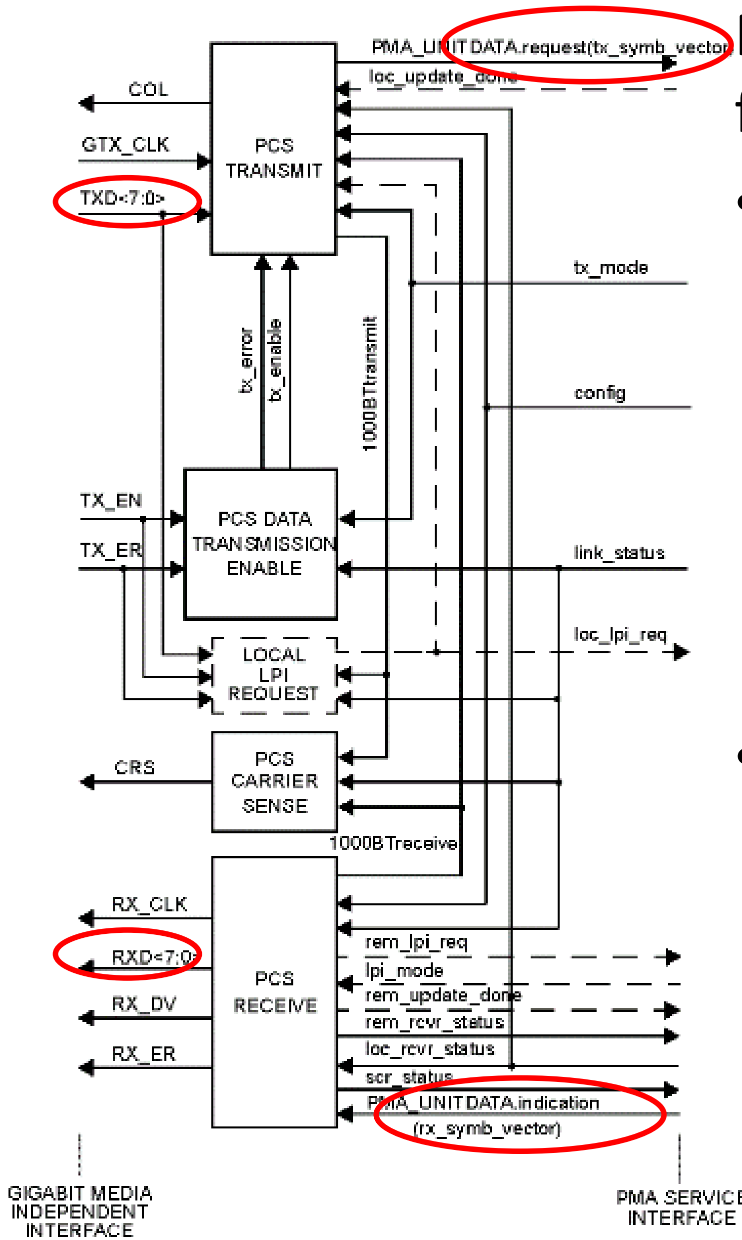
bytes ↔ symbols

$$(b_7 \ldots b_0) \leftrightarrow (A, B, C, D)$$

Physical Layer of 1000BASE-T

- 1000BASE-T uses a continuous signalling system; in the absence of data, Idle symbols are transmitted.
- Idle mode is a subset of code-groups in that each symbol is restricted to the set {2, 0, –2} to improve synchronization.
- A 1000BASE-T PHY can be configured either as a MASTER PHY or as a SLAVE PHY.
- The MASTER-SLAVE relationship between two stations sharing a link segment is established during Auto-Negotiation
- The MASTER PHY uses a local clock to determine the timing of transmitter operations.
- The SLAVE PHY recovers the clock from the received signal and uses it to determine the timing of transmitter operations

PCS Physical Coding Sublayer



Physical Coding Sublayer (PCS) has two fundamental functions:

- PCS **Transmit** function (clause 40.3.1.3) which converts **bits of each byte** TXD[7:0] received from the Reconciliation layer (through GMII) into **four** quinary symbols (TA,TB,TC,TD), each being sent on a pair of twisted copper wires
- PCS **Receive** function (clause 40.3.1.4) which converts the four-tuples of quinary symbols into the bytes RXD(7:0) sent to the Reconciliation layer

PCS Transmit Functions

(Based on section 3, clause 40, 40.3.1.3/4)

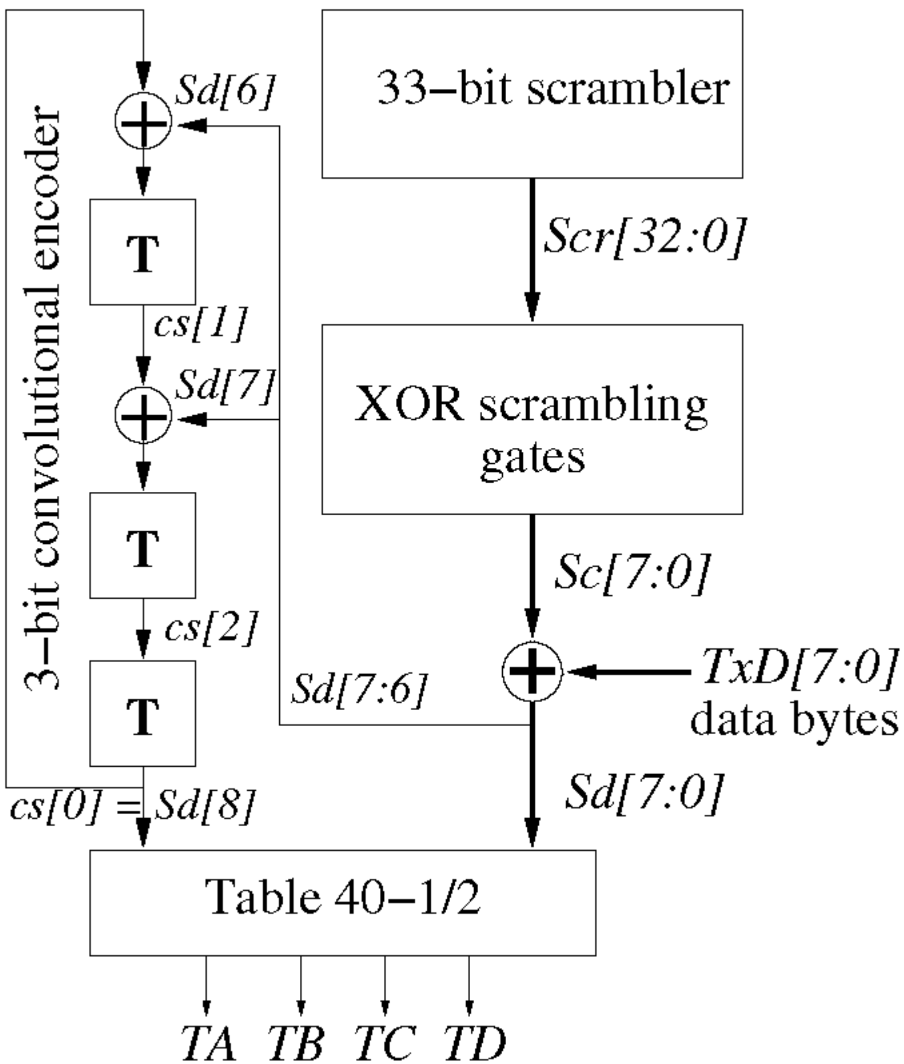
- In the normal mode of operation, the PCS Transmit function uses an **8B1Q4 coding technique** to generate at each symbol period (T = 8ns) code-groups that represent **data, control or idle** based on the code-groups defined in Table 40–1 and Table 40–2.

During transmission of data, the TXD[7:0] bits are

- **scrambled** by the PCS using a 33-bit side-stream scrambler, then
- **encoded** by a three-state convolutional encoder, then
- **converted** into a code-group of quinary symbols and transferred to the PMA.

PCS Transmit Function

Simplified block-diagram (ignoring control signals):



- In the tutorial exercise, you will be asked to give a more detailed description of the PCS transmit function based on section 3, clause 40, 40.3.1.3/4 of the 802.3 standard
- EXPLAIN DETAILS OF THE BLOCKS

Scramblers

- There are **two** transmitter side-stream scramblers described by the following polynomials, one for MASTER and one for SLAVE

$$g_M(x) = 1 + x^{13} + x^{33}; \;\; g_S(x) = 1 + x^{20} + x^{33}$$

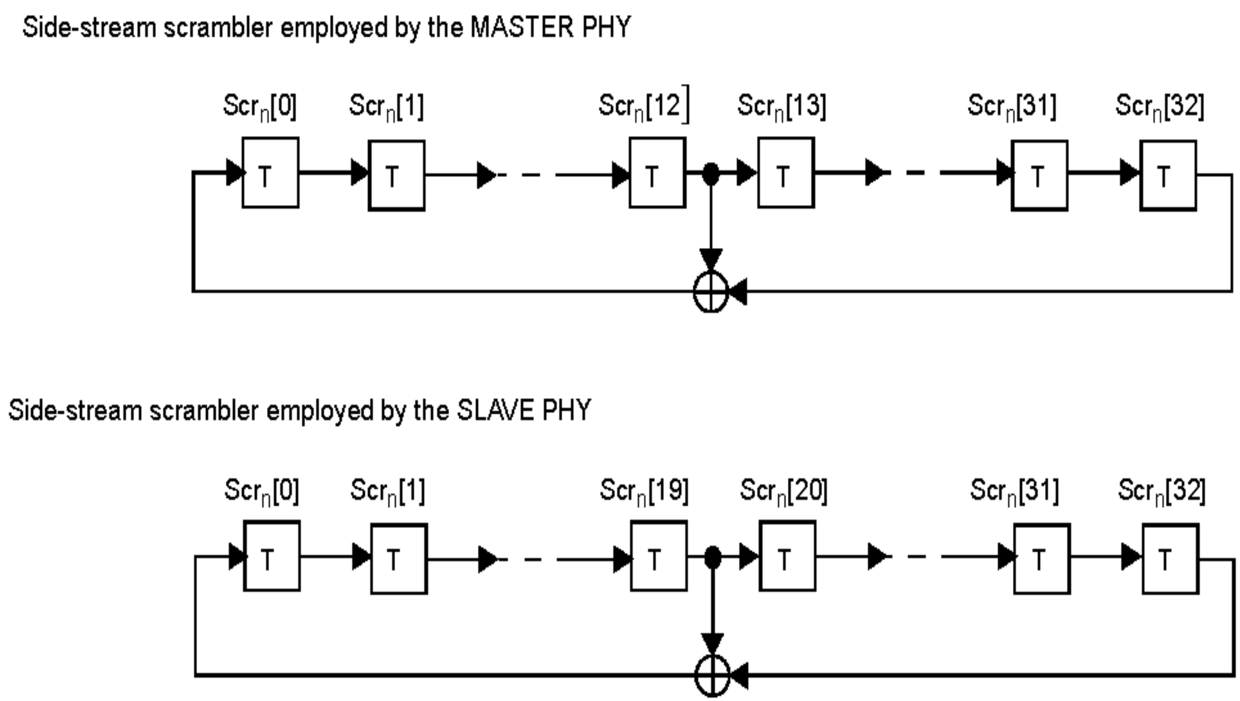


Figure 40–6—A realization of side-stream scramblers by linear feedback shift registers