ES/EG 4546

Communications Engineering 2:

The Ethernet Local Area Network

G. Fairhurst

r57 2020

Course Overview

Course web site:

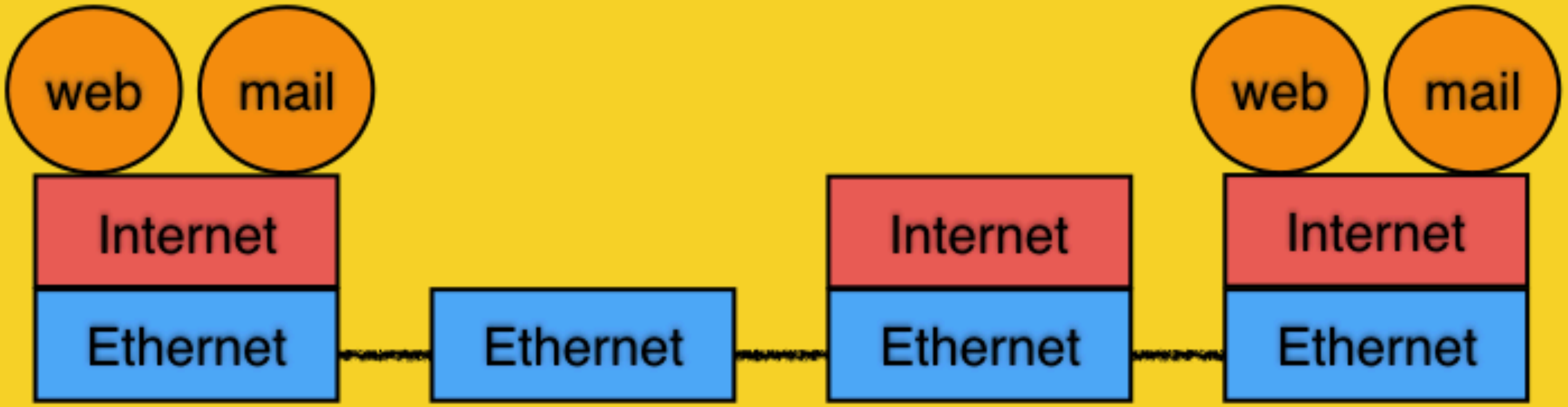
- *Syllabus for the entire course*
- *Copies of presentation material*
- *Notes for lectures and related topics*
- *Videos to support the course.*

Course Contents:

- *Lectures*
- *Tutorials*
- *Example Classes*
- *Practical Exercise (assessed)*
- *Examination*

ES/EG 4546

Communications Engineering 2:

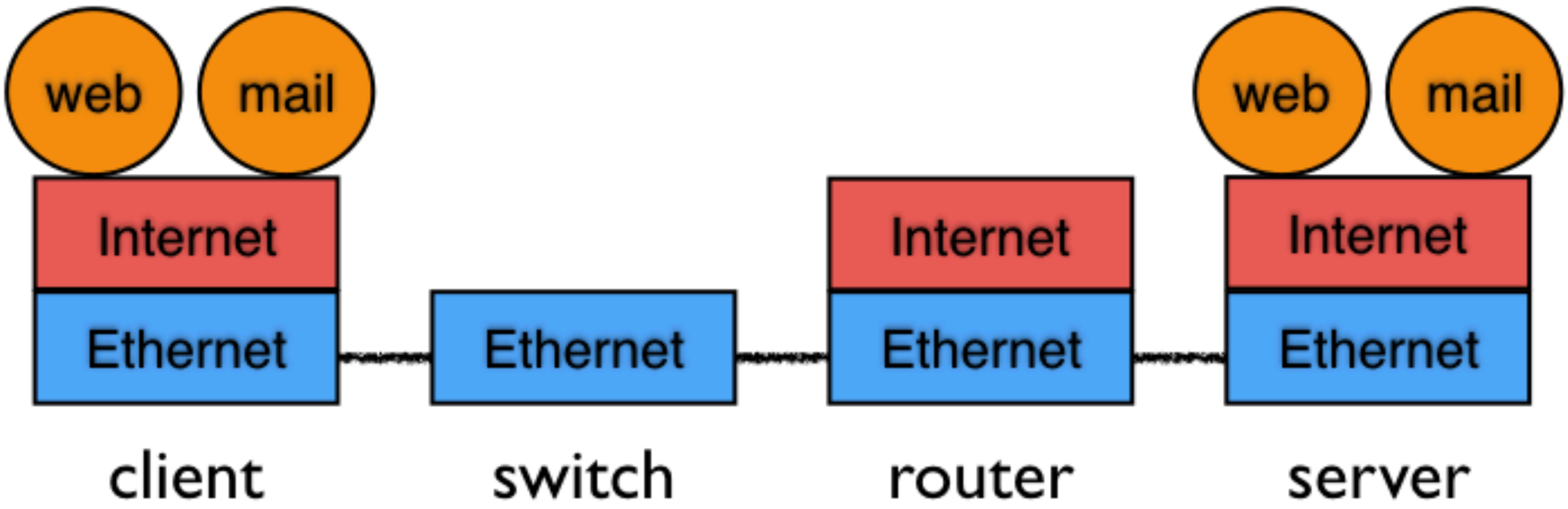


Communication between Systems

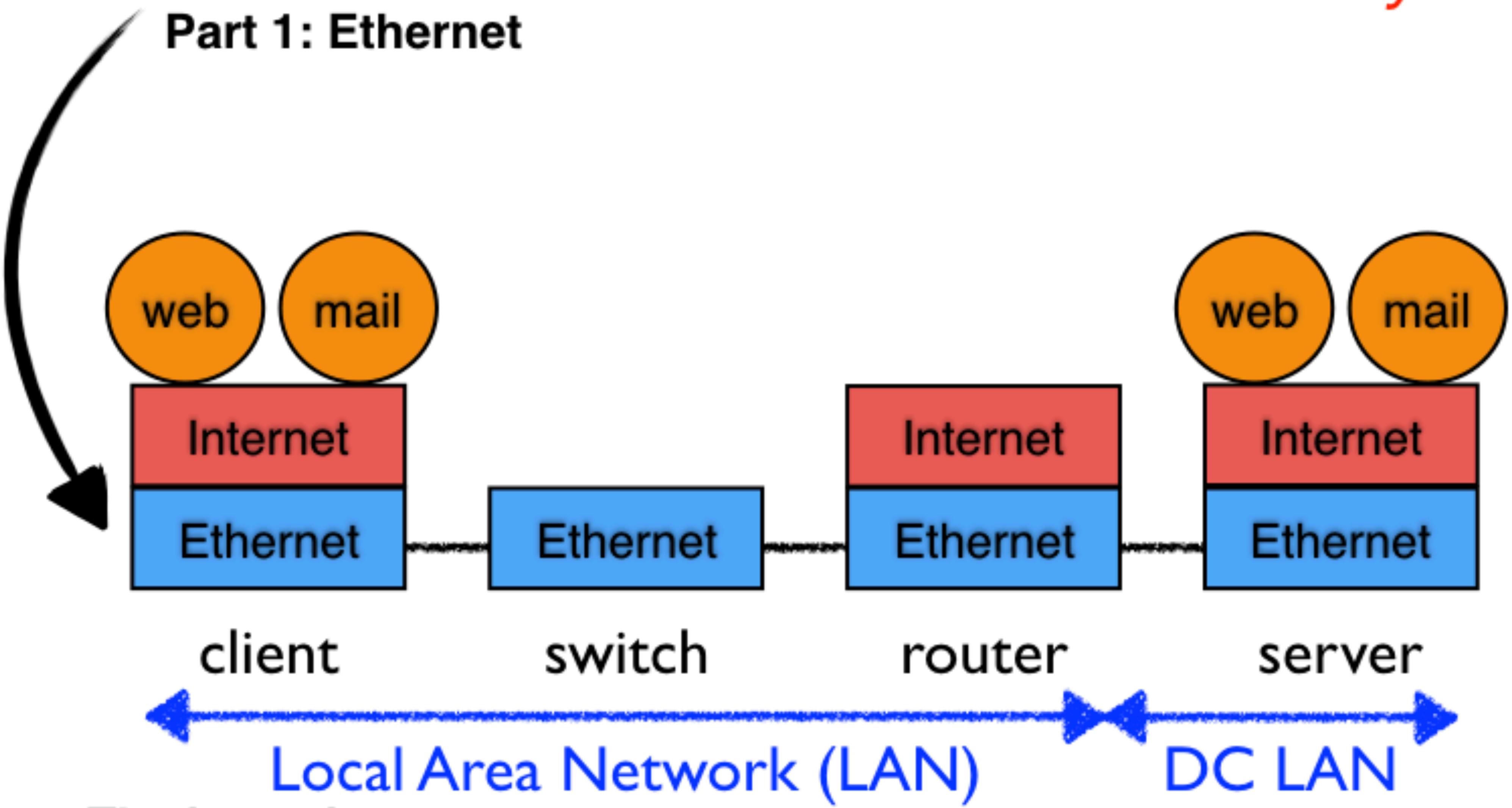
Networking is about communication between **systems**

- Examples of systems are PCs, Phones, Printers, Servers, etc
- Each system implements a ***Protocol Stack***

We divide our understanding of a stack into 2 layers:



Protocol Stacks and Layers



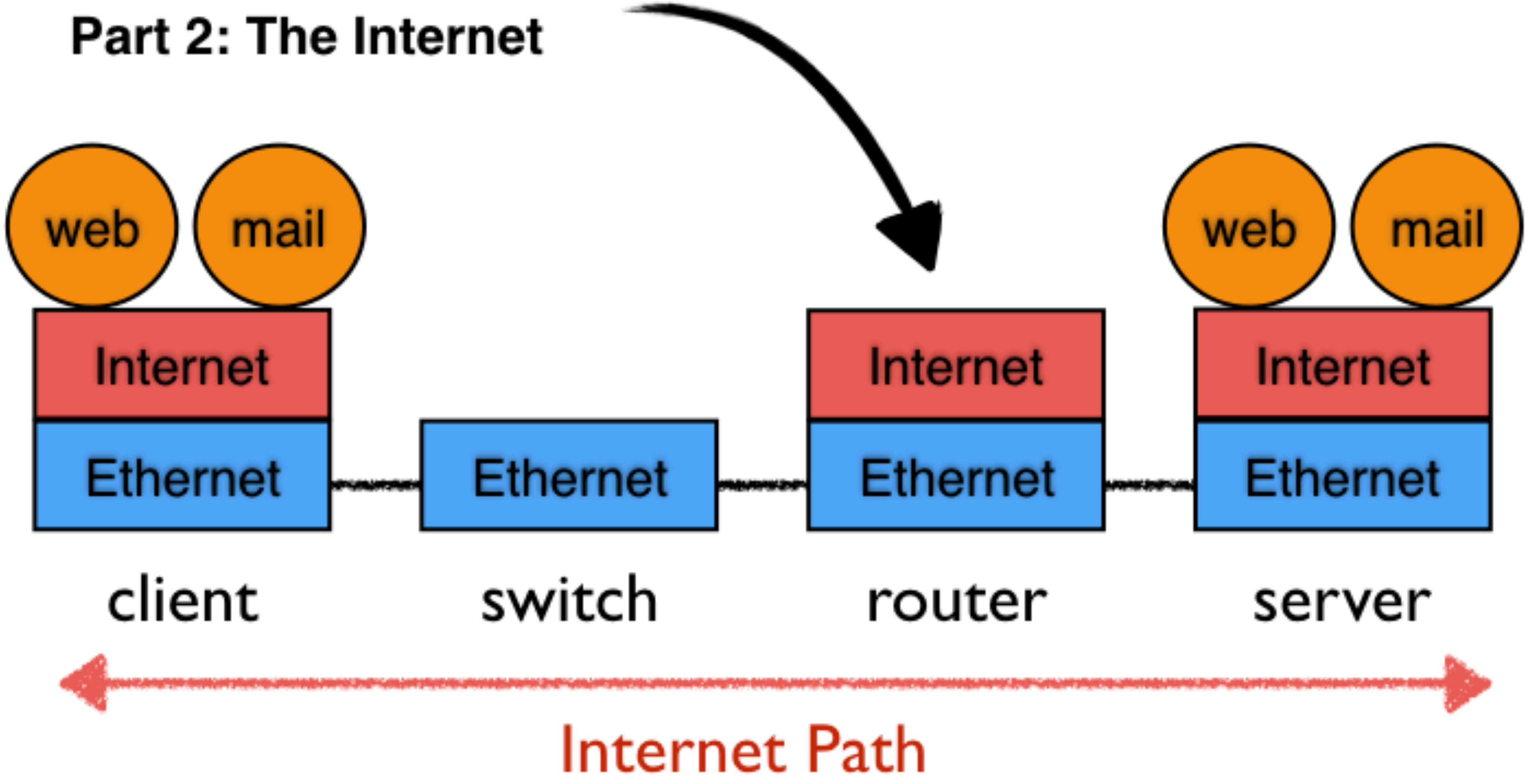
The lower layers

Communication across an Ethernet LAN

Link Layer - Medium Access Control (protocol)

Physical Layer - Transmission Control (cabling)

Protocol Stacks and Layers



The upper upper layers

Communication across an Internetwork of LANs

Transport Layer - Communicating end to end between systems

Internet Layer - Networking across an Internet path

Part 1: Ethernet

1. The Origins of the Ethernet LAN
  10B5 and 10B2 coaxial cables

2. Ethernet Frames
  Addressing (Multicast & Broadcast)
  A shared physical Medium & Medium Access Control

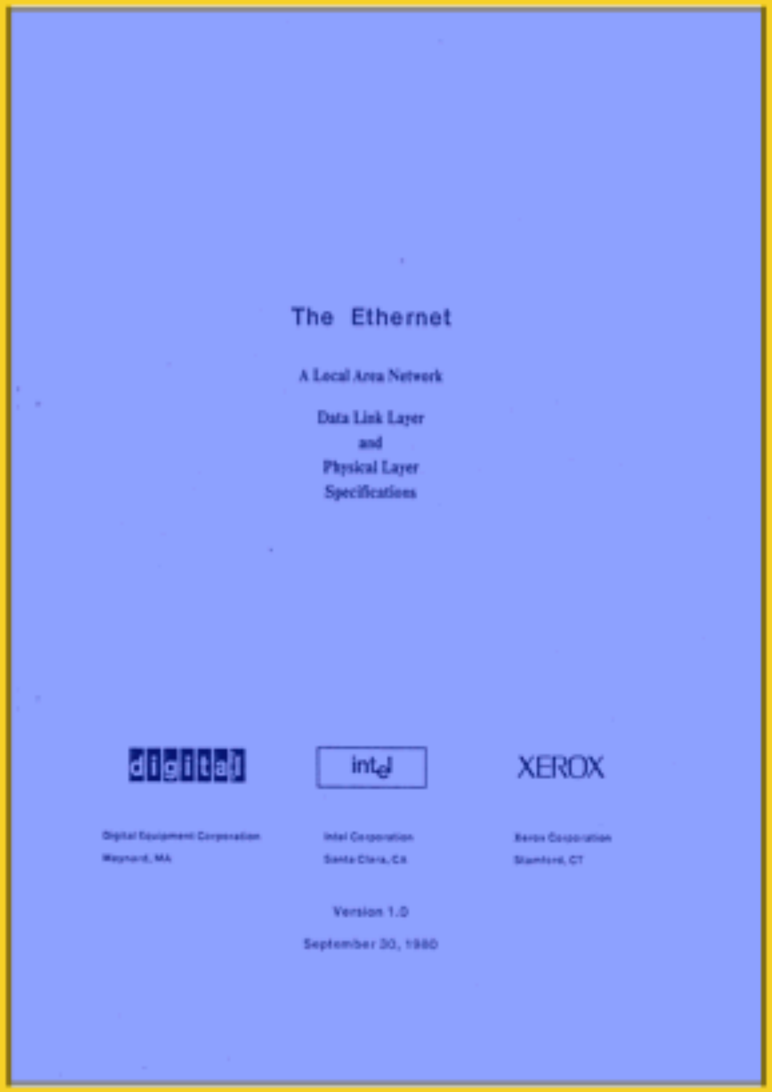
3. Ethernet Transmission
  Sending frames
  Frame reception

4. Connecting LAN Segments to form a Collision Domain
  Hubs, repeaters and the 5-4-3 Repeater Rule
  10BT and 10BF

5. Bridges and Switches
  Forwarding using address tables
  Dynamic learning

6. Faster Ethernet
  Fast Ethernet
  VLANs, Gigabit and 10GB Ethernet

The Origins of the Ethernet LAN



Blue Book

IEEE 802.3

10B5 Coaxial Cables

Module 1.1

A Local Area Network is....

sends ***packets*** of data in frames

local (one building, group of buildings, etc)

always controlled by ***one administrative authority***

usually high speed and always ***shared***

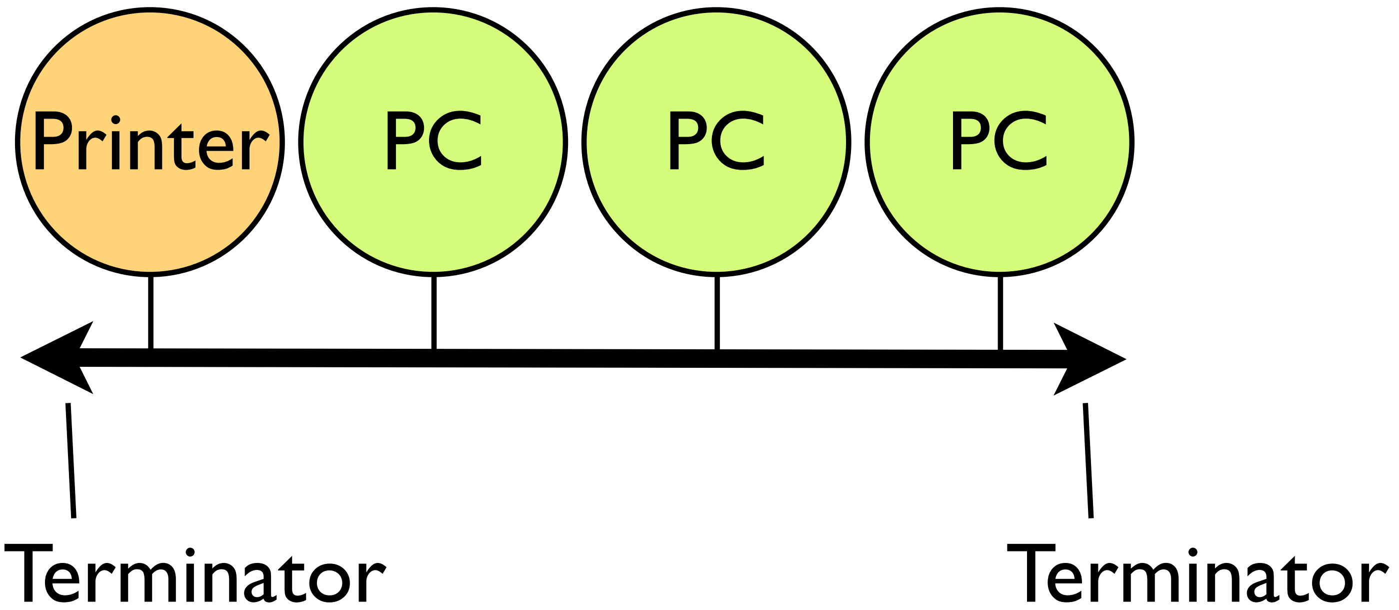
often assumes other users of the LAN are ***trusted***

either ***planned*** (structured) or ***unstructured***

What is Ethernet?

First LAN designed at Xerox "PARC" (1972)

- 2.94 Mbps 75 Ohm Coaxial cable
- A Bus Topology used share expensive laser printers
- File sharing followed later



Ethernet

*“**the ether**” the air, when it is thought of as the place in which radio or electronic communication takes place*, OED.

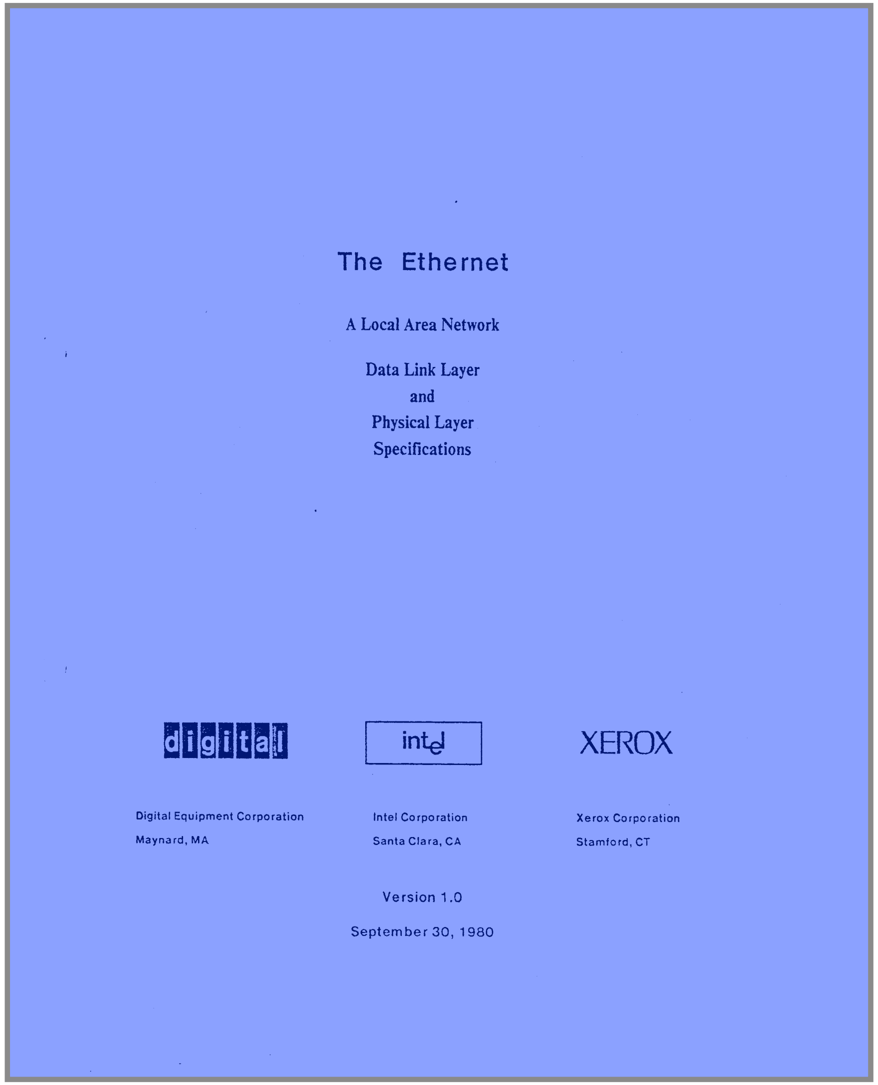
What is Ethernet?

Ethernet v2 - Blue Book

First published 1980, updated 1982
Digital, Intel, Xerox (DI
10 Mbps Speed
50 Ohm Coaxial cable

An Open Standard

The invention of Ethernet as an open, non-proprietary, industry-standard local network was perhaps even more significant than the invention of Ethernet technology itself.



The Ethernet

A Local Area Network

Data Link Layer
and
Physical Layer
Specifications

digital | intel | XEROX

Digital Equipment Corporation
Maynard, MA

Intel Corporation
Santa Clara, CA

Xerox Corporation
Stamford, CT

Version 1.0

September 30, 1980

LAN Topologies

Single Link

Bus

Shared Cable
(e.g. Coaxial Cable: 10B2, 10B5)

Star

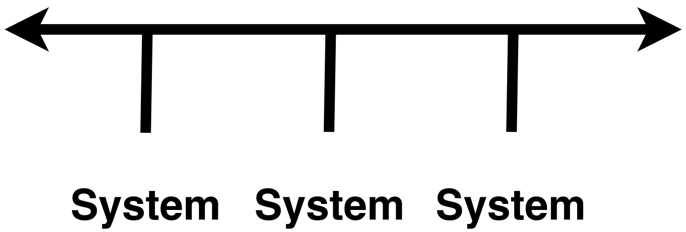
Connection to a Hub
(e.g. Twisted Pair Cable: 10BT)

Tree

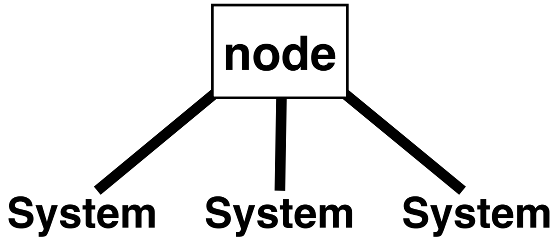
Connected Hubs/Switches
… network of routers ….
(Variety of media: 10B2,10B5, 10BT, 10BF)

Point-to-Point link

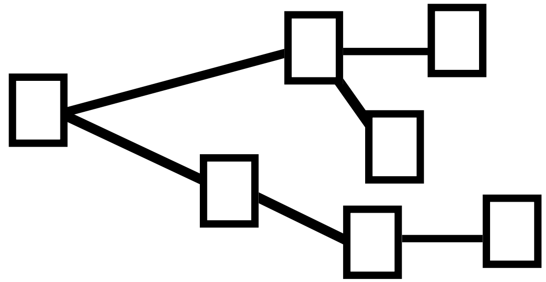
Bus



Star

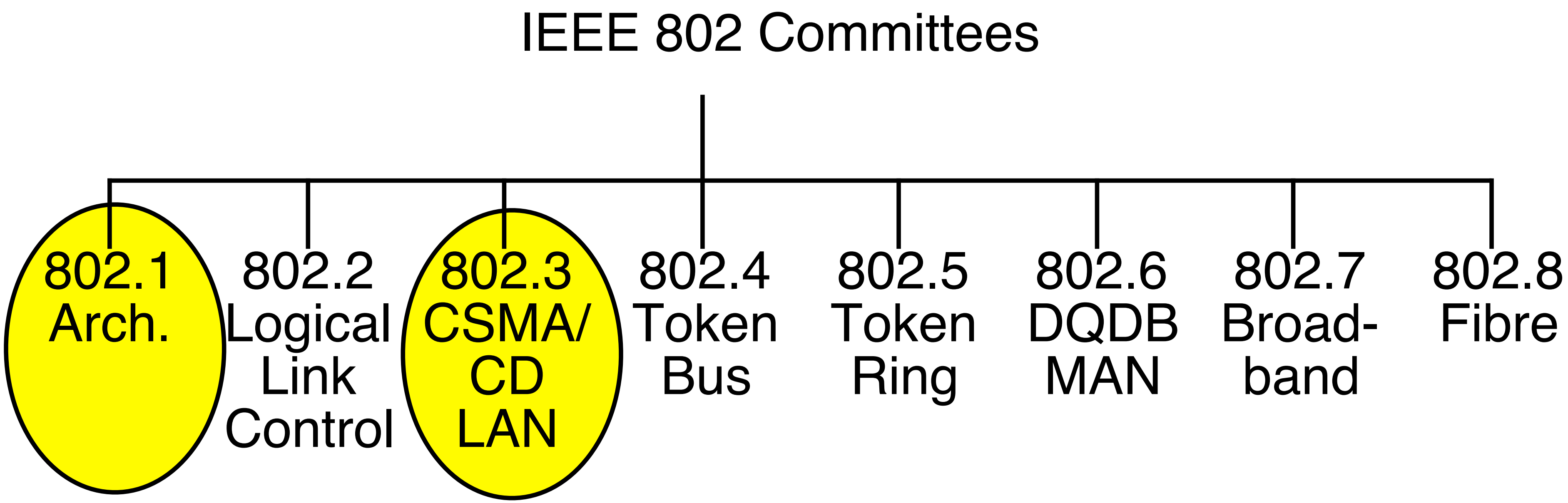


Tree



What is Ethernet?

Ethernet Standardised by IEEE in 1983:



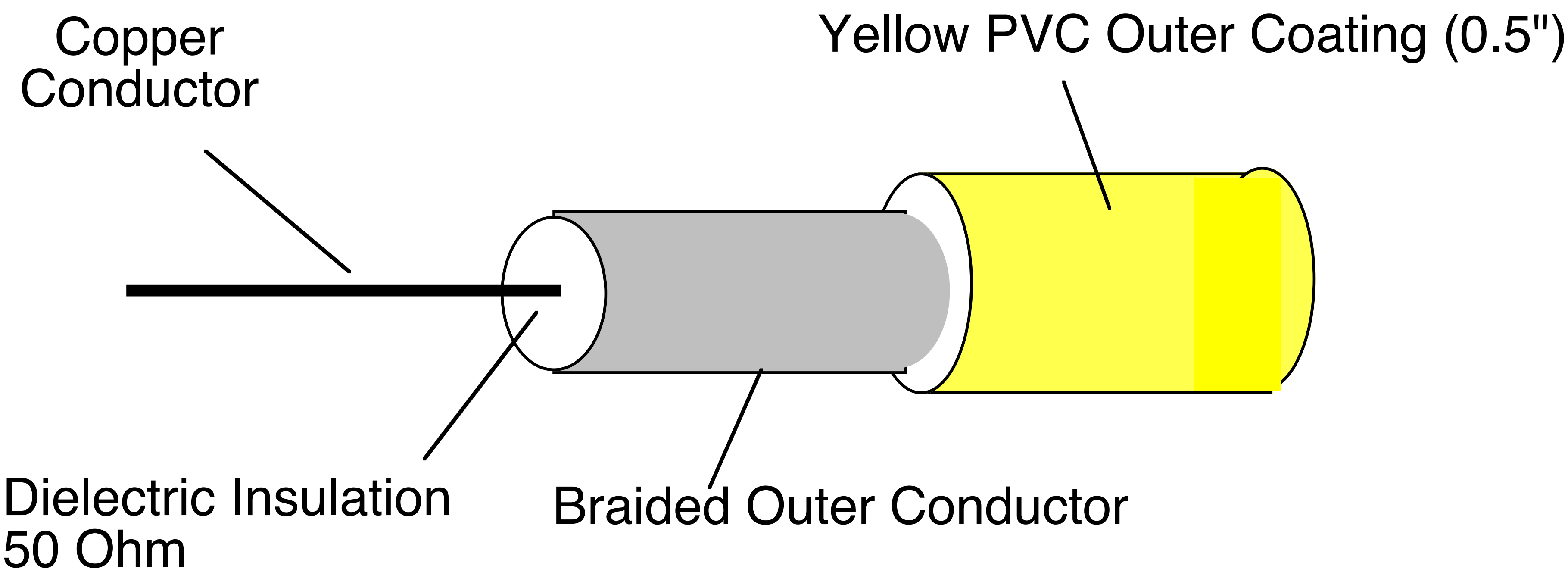
IEEE 802.3

Two original variants: Thick Ethernet and Thin Ethernet at 10 Mbps

Speeds now available:

- 100 Mbps (Fast Ethernet)
- 1000 Mbps (1 Gbps)
- 10000 Mbps (10 Gbps)
- 40 Gbps, 100 Gbps, …

10B5 Ethernet Media



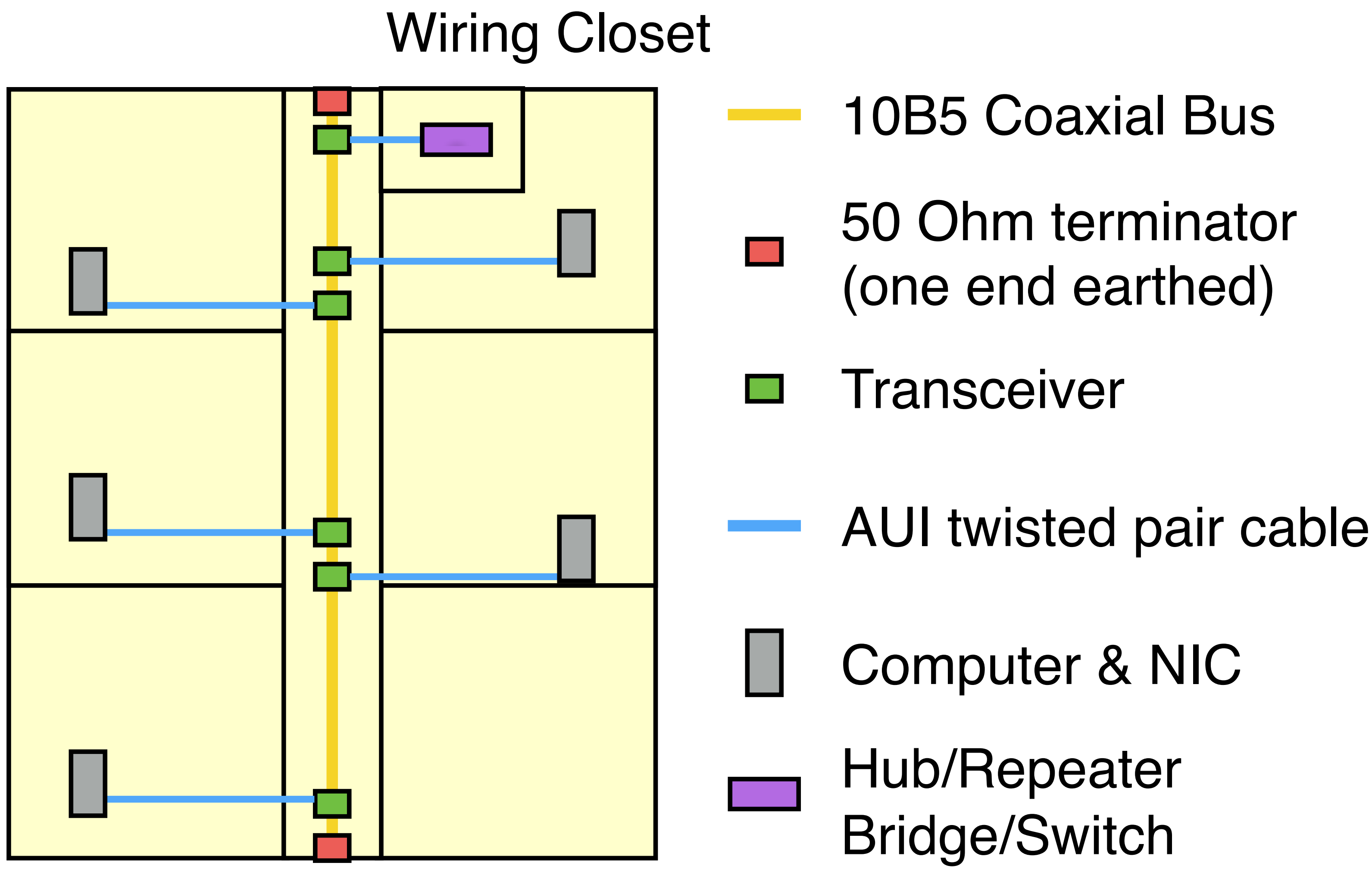
High performance co-axial cable

- Segment length ≤ 500m
- Good noise immunity
- N-Type connector at each end

1024 NICs attached to a single cable segment

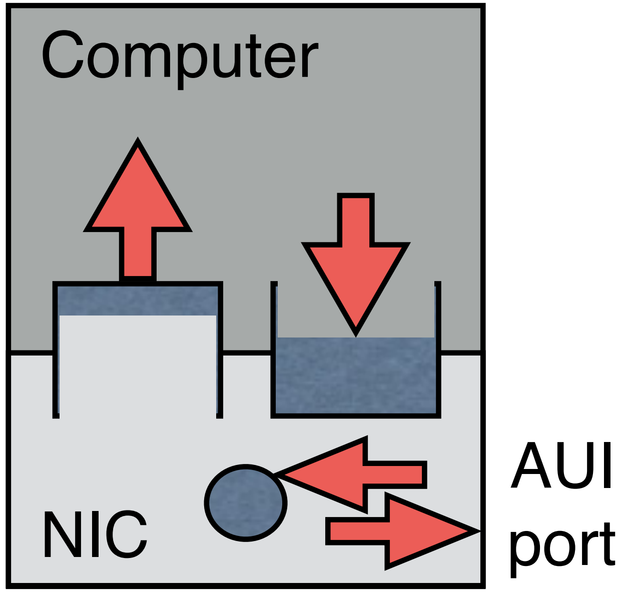
Ethernet 10B5 Cable Segment

Cable usually installed as a trunk running down corridor



Typical Use of 10B5 within an Office (max 500m segment)

Ethernet Network Interface Card (NIC)



Originally a card inserted in a PC or computer

Transmission and reception using the media

Input (receive) and Output (transmit) queues

A queue of frames for transmission

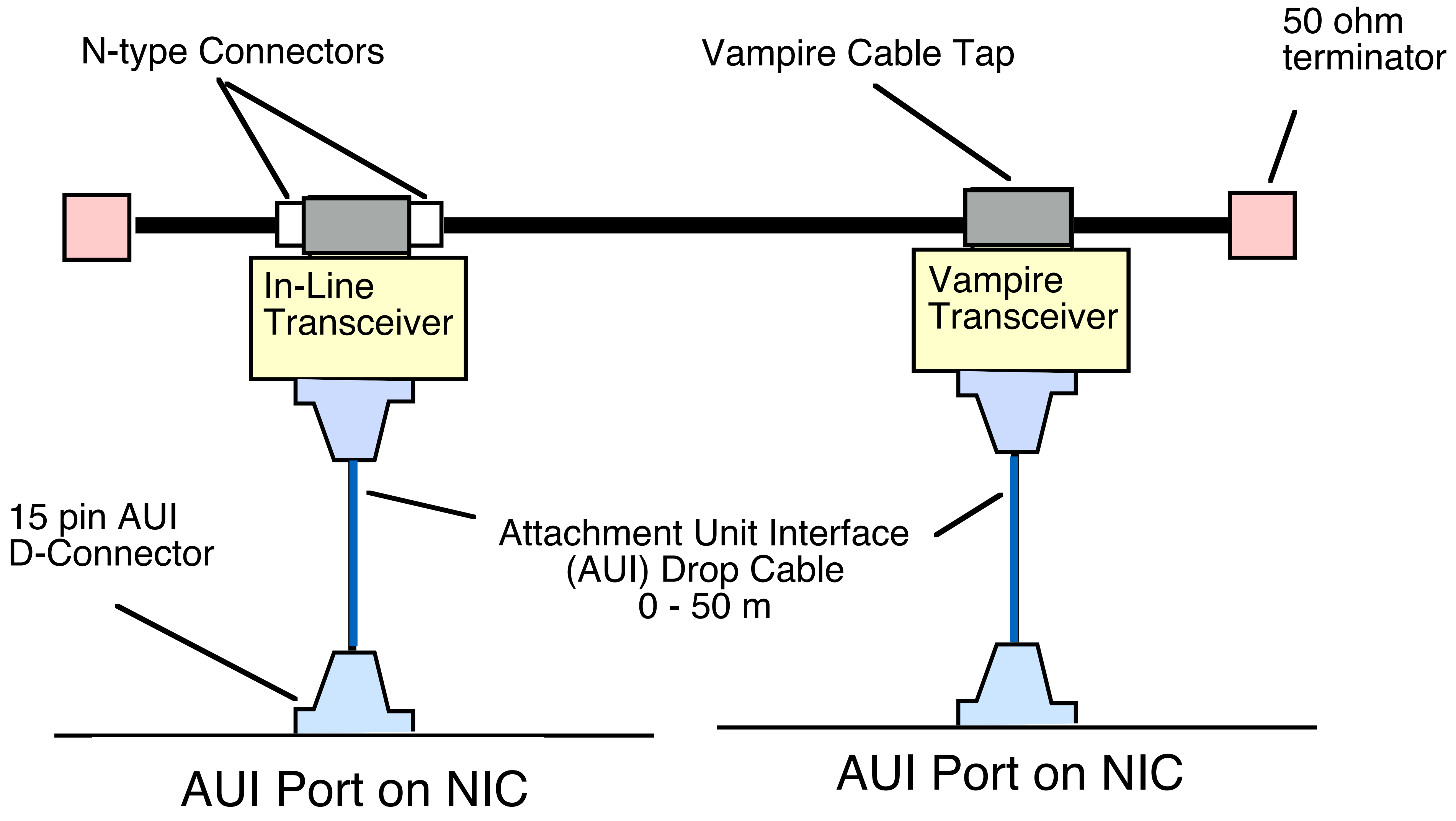
- Sender completes a Tx descriptor in the queue (location of data in memory, length of data, etc)
- Sender writes a register in the NIC to ask for this to be sent
- NIC then performs a DMA of the data, serialises data and adds information needed to transmit a frame on the cable

A queue to hold received frames

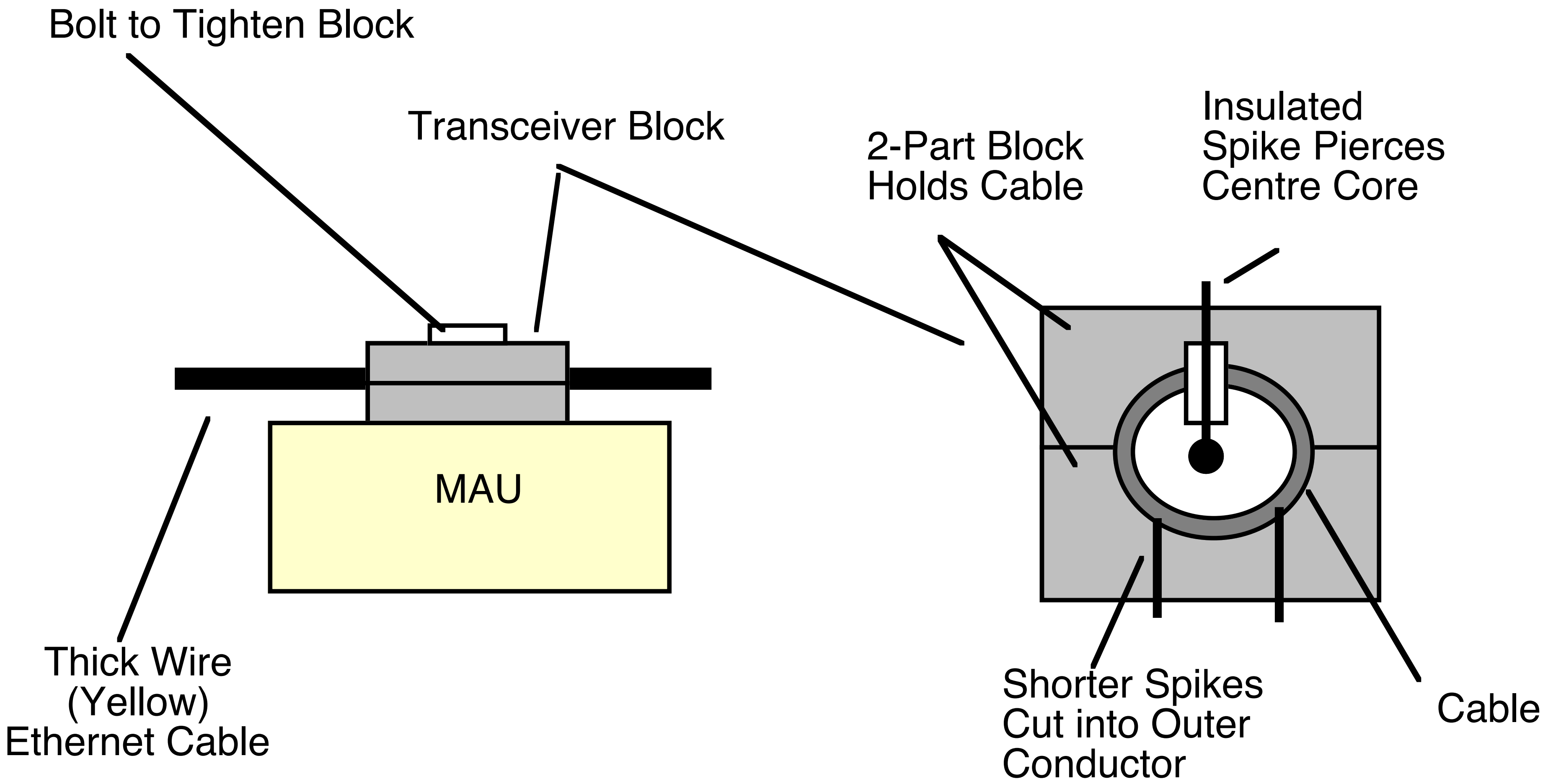
- The NIC processes a frame received on the cable
- The frame is stored internally and a Rx descriptor is created
- The data in valid frames is DMA'ed to computer memory
- The receiver is interrupted to say that frames have been received

10B5 (Thick Ethernet Transceiver)

Two types of transceiver are supported:
- In-Line (N-type screw connector as cable installed)
- Vampire transceiver (insulation displacement after cable installed)



10B5 (Thick Ethernet Vampire Transceiver)



Cable drilled; transceiver block tightened around the cable

This connects spikes to outer and inner cable conductors

Transceiver electronics (MAU) bolted to the transceiver block

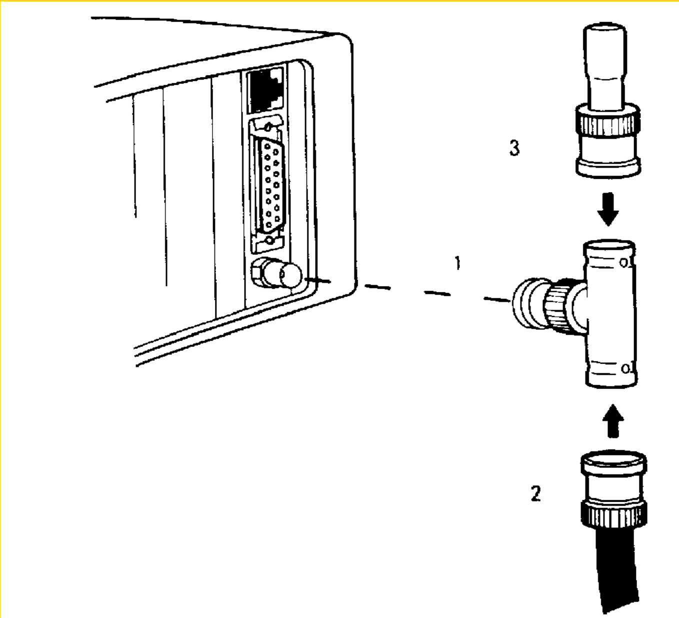
Medium Attachment Unit (MAU)

Ethernet Success Story

- Simple low cost LAN (compared to computers)
- Familiarity to customers !!!
- Wired networks are still the most common media
- Has become an standard for Internet LANs

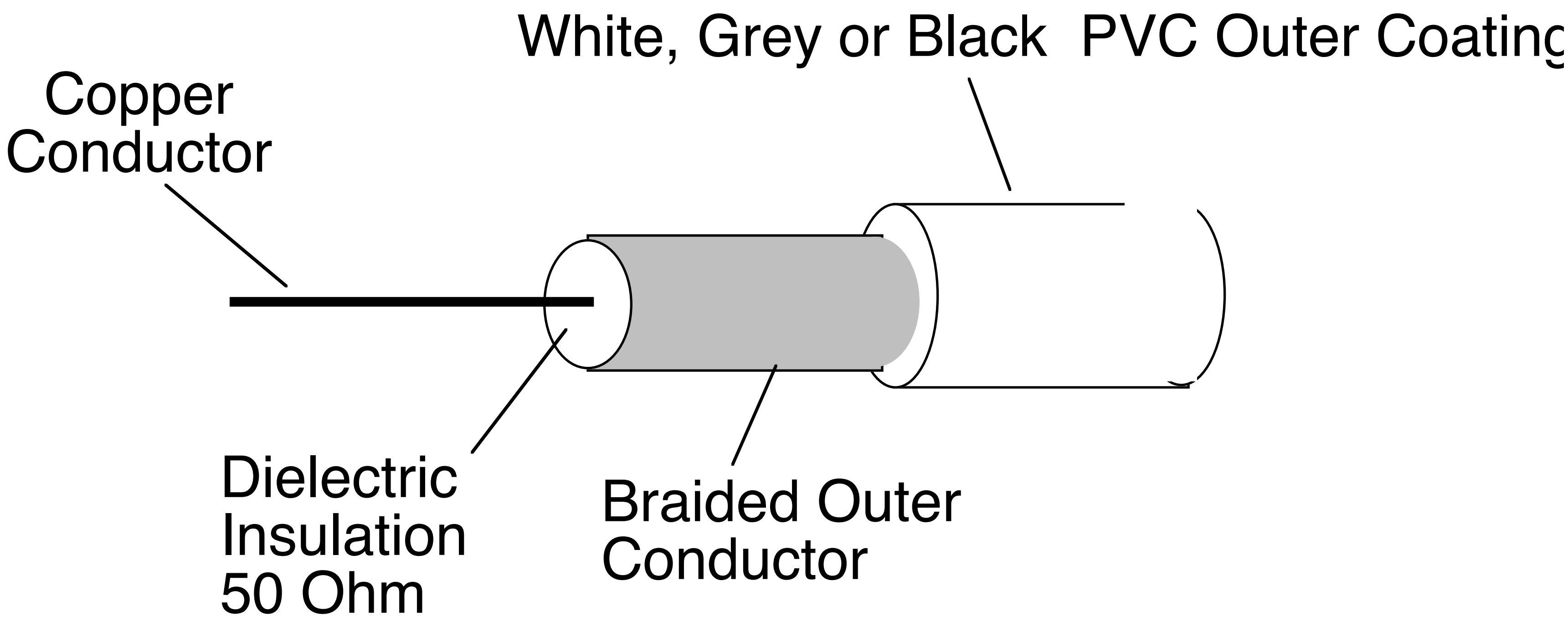
The Origins of the Ethernet LAN

10B2 Coaxial Cables



Module 1.2

10B2 (Thin Ethernet)



Low cost co-axial cable (RG58u)
- Segment length 0.5m ≤ 185 m
- Flexible, easy installation

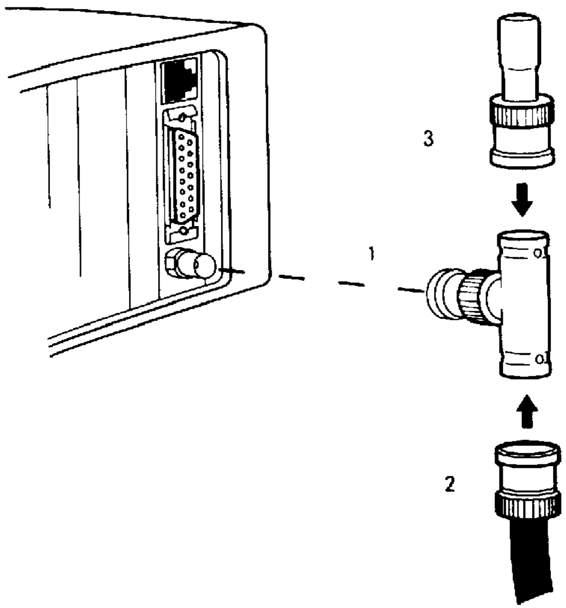
30 NICs allowed using one segment

NICs with In-built or external transceiver

10B2 (Thin Ethernet)

BNC connector at each end of cable

"T" joiner connects the NIC to two cables



BNC connector and "T" joiner

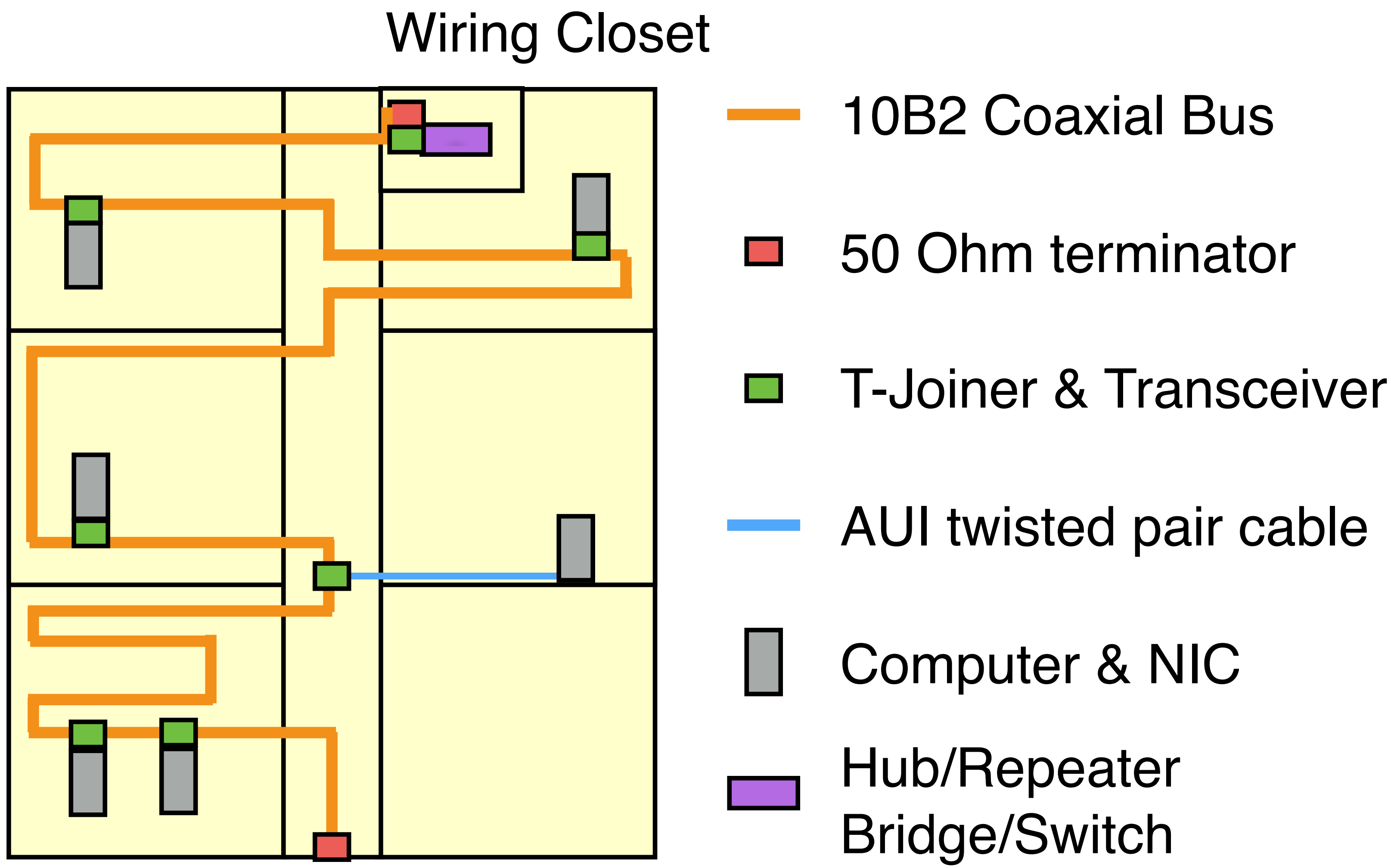
BNC 50 Ohm Terminator

NICs with In-built or external transceiver

Flexible lengths of cable with BNC plugs

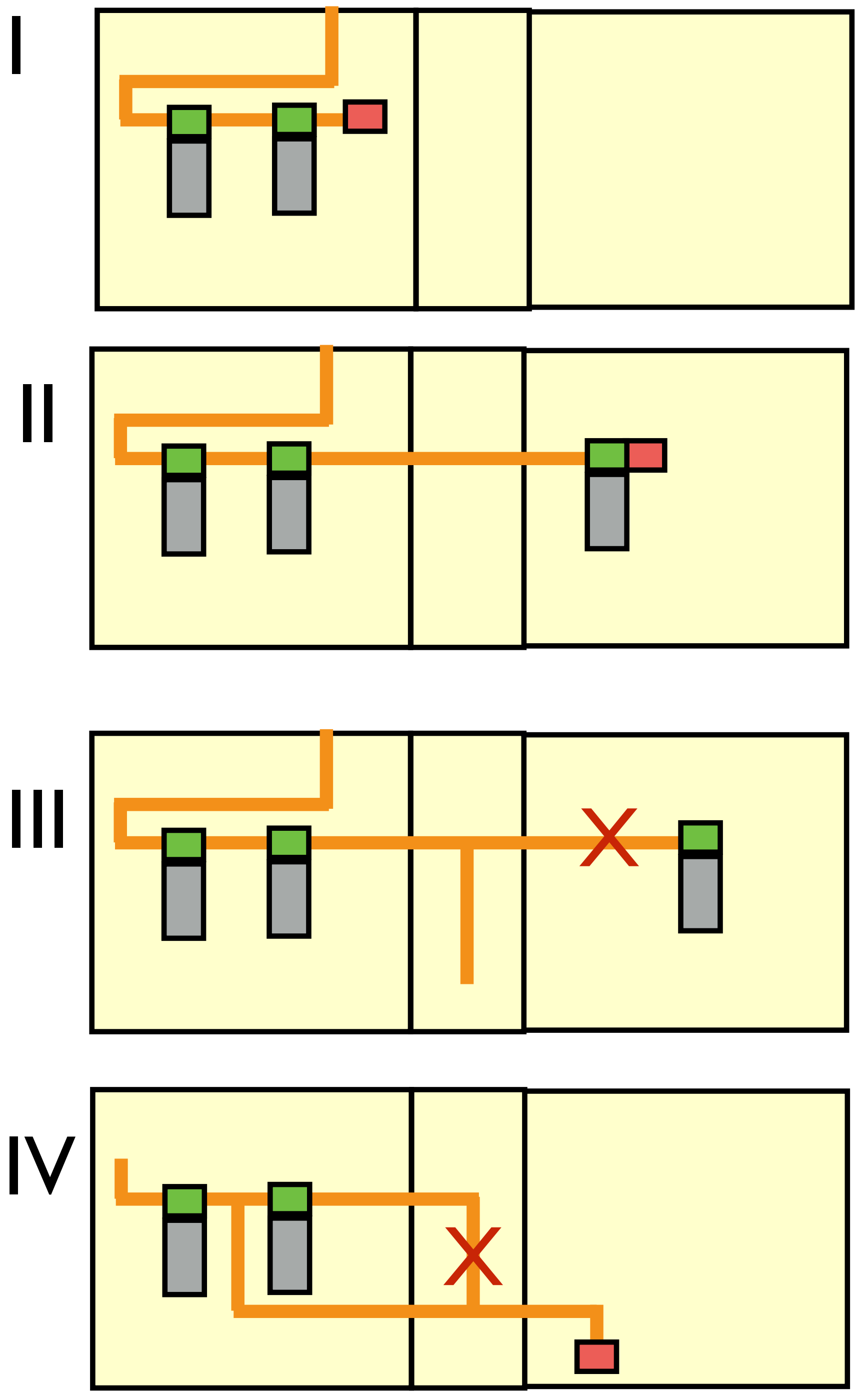
Ethernet 10B2 Cable Segment

Often cable connected device to device, rather than pre-installed



Typical Use of 10B2 within an Office (max 185m segment)
Maximum of 30 computers on a single segment

10B2 (BNC Connector)



Easy to install

Plug “T” into NIC and connect cable!

Unplug the BNC connector

add another "T" and a new cable

and connect another NIC

Must form one bus:

No loops

No stubs between “T” and NIC

Easy to extend....

... difficult to manage (unstructured)!

Ethernet Success Story

- **Ethernet already familiar with customers**
- **10B2 made the network even more Cost-Effective**
- **Very Easy to Install**

  Simple BNC twist connector

  Great for unstructured networks that can evolve
- **A larger LAN can use Repeaters**

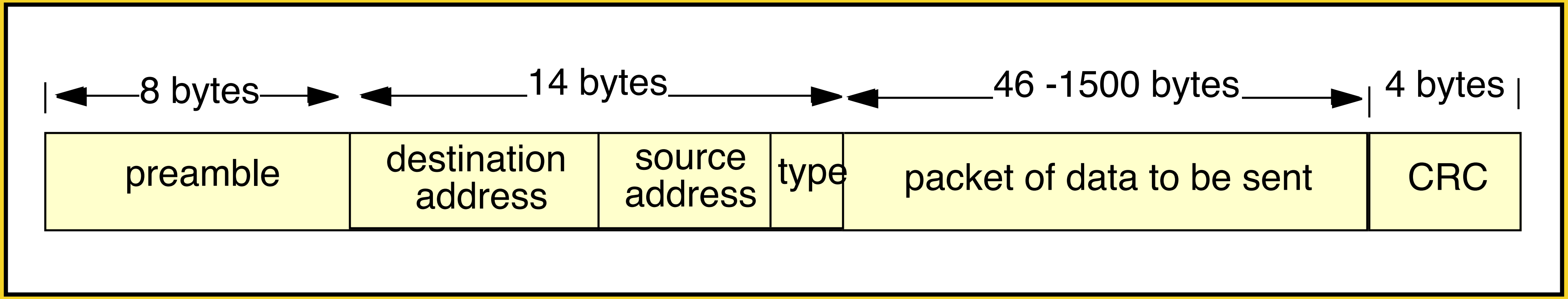
Check the web pages and notes for 10B2 and 10B5!

Coaxial cable Ethernet is now only used in special networks.

Ethernet Frames

Addressing
A shared physical medium
Medium access control
Sending frames
Frame reception
Multicast and Broadcast

Link Layer



Module 2

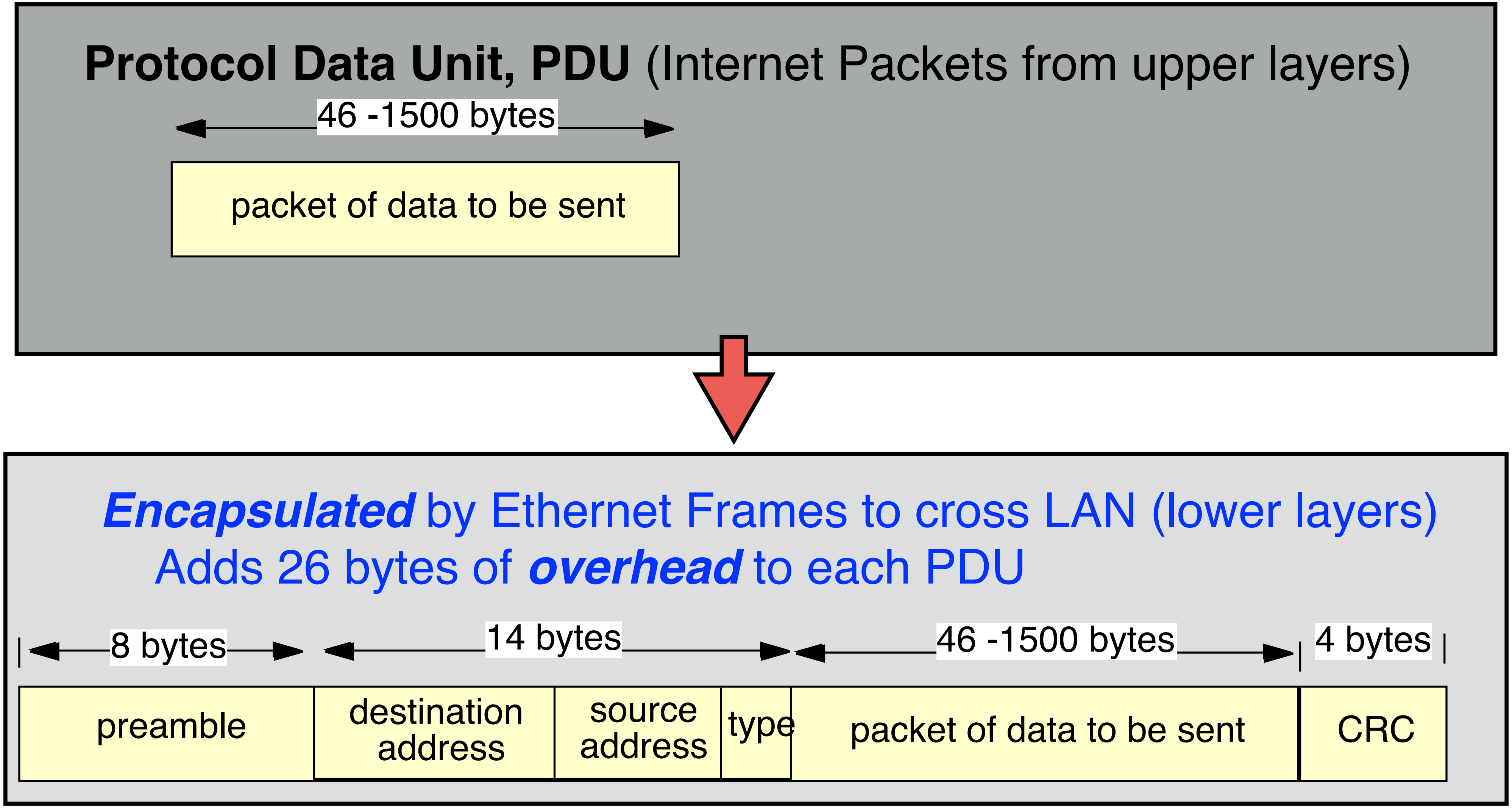
Ethernet Frames:
Addressing

Link Layer

FF:FF:FF:FF:FF:FF

Module 2.1

Ethernet Frames



e.g.

a 46 byte packet is carried in a 72 byte frame

a 1200 byte PDU is carried in a 1226 byte frame.

Ethernet MAC Address

08:00:20:00:00:01

A MAC Address is a 48-bit number

- Usually represented as 6 pairs of hexadecimal digits
- One hex digit corresponds to a value 0-15 (0x0 to 0xF)
- For ease of reading we separate each byte* by a colon.

We divide the 48-bit address into two parts:

- First 3 bytes: the organisationally unique identifier (OUI) - orange
- Second 3 bytes: the manufacturer-assigned address - yellow.

** In some documents an 8-bit byte is referred to as an octet.*

Ethernet MAC Address

08:00:20:00:00:01

Each Network Interface Card (NIC) has a unique MAC Address

- Held in a manufacturer-configured PROM

Addresses are globally unique

- A MAC Vendor Code (OUI) + Number
- About 1% of OUIs have been used

IEEE *sells* these blocks of addresses to *manufacturers*

- Each block has 256 cubed addresses
- That is 16 Million!!

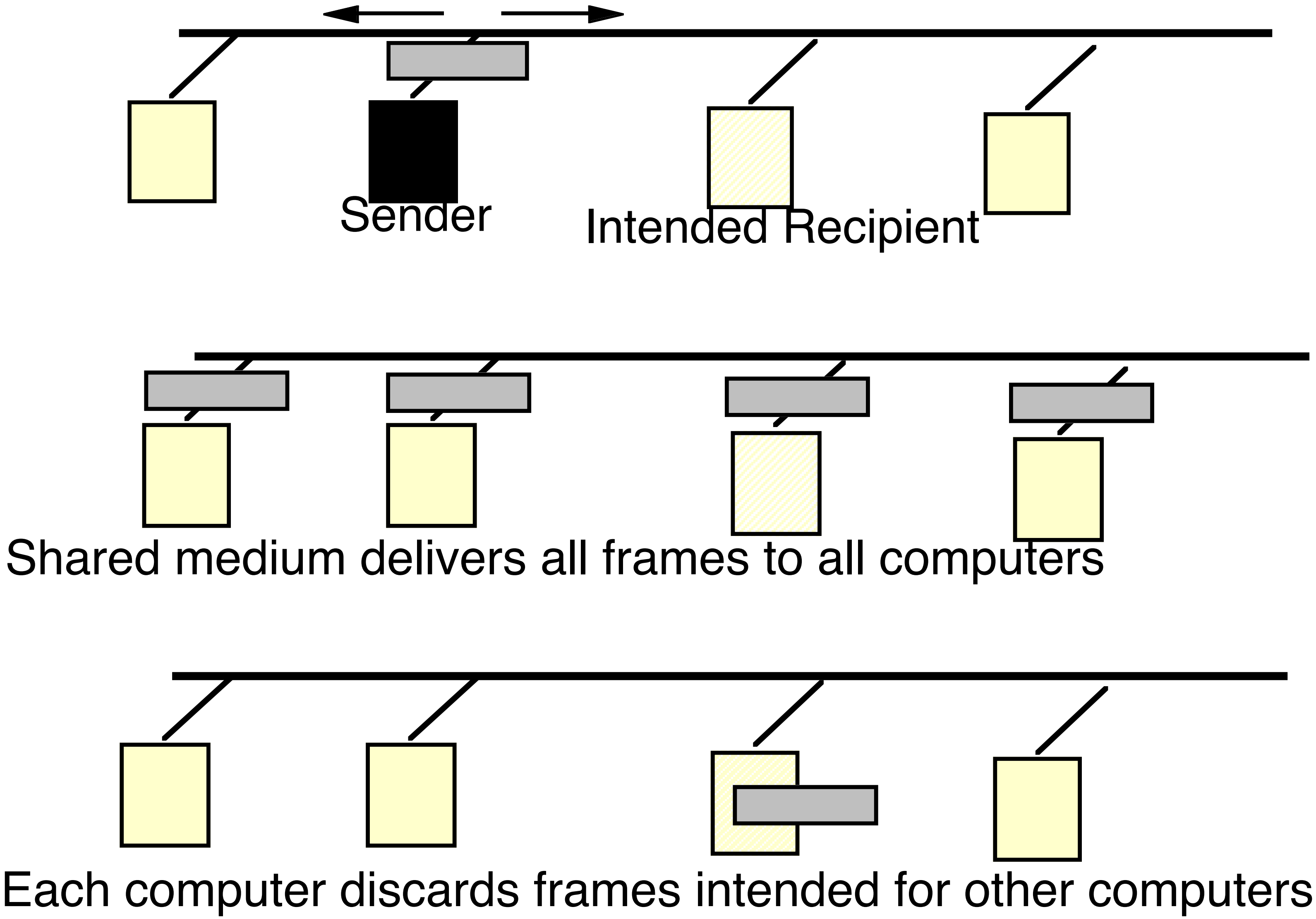
MAC Vendor Codes (OUIs)

08:00:20:00:00:01

The first 3B of address indicates the assigned manufacturer

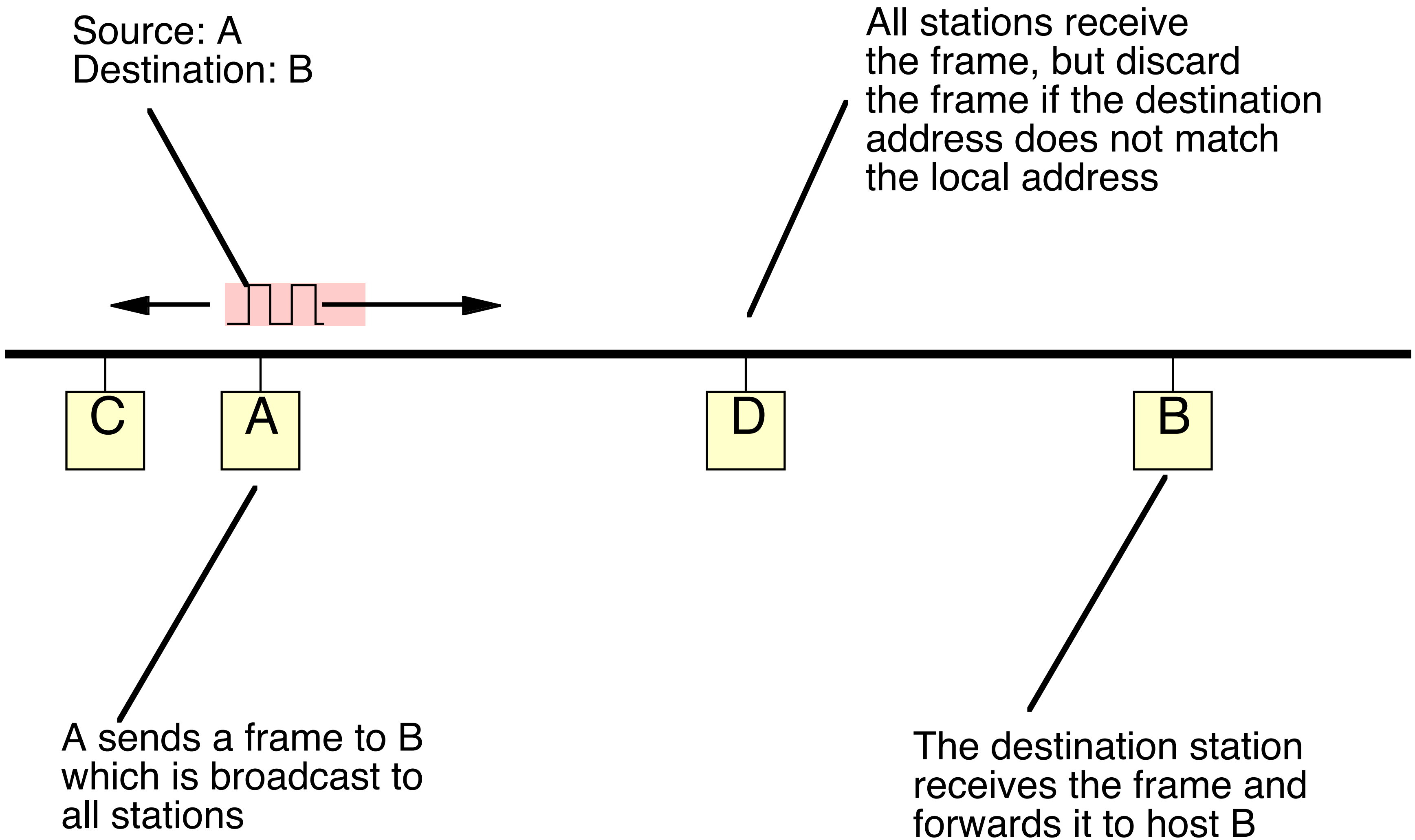
```
080002  3Com (Formerly Bridge)
080003  ACC (Advanced Computer Communications)
080005  Symbolics            Symbolics LISP machines
080008  BBN
080009  Hewlett-Packard
08000A  Nestar Systems
08000B  Unisys
080011  Tektronix, Inc.
080014  Excelan BBN Butterfly, Masscomp, Silicon Graphics
080017  NSC
08001A  Data General
08001B  Data General
08001E  Apollo
080020  Sun                  Sun machines
080022  NBI
080025  CDC
080026  Norsk Data (Nord)
```

Shared Access to Ethernet Medium



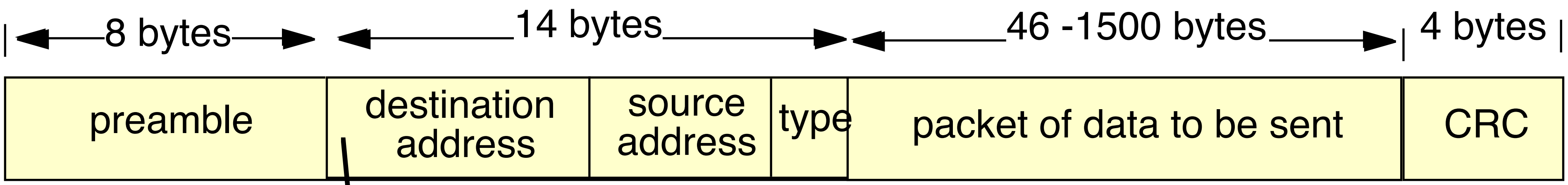
We can use the destination address to perform unicast communication, where frames are only received by a specific destination computer

Using the Destination MAC address



This assume a sender knows the value of the MAC address in the remote NIC's PROM (we'll find out how it does this later!)

Ethernet Frame Structure



LAN address of intended recipient

first bit = 0 indicates point to point
first bit = 1 indicates broadcast or multicast

48 bits, expressed as 12 hexadecimal digits
e.g., 12:34:56:78:9A:BC

A theoretical 200,000,000,000 addresses
Actually 70,000,000,000... (2 bits are used)
20,000 MAC addresses for each person on the planet!

Special MAC Addresses

FF:FF:FF:FF:FF:FF

The all 1's Address is used to send to all NICs

- Known as the **broadcast** destination address
- Only ever used as ***destination address***

00:00:00:00:00:00

The all 0's Address is special

- Known as the **unknown** address
- Only ever used as ***source address***

Use of Broadcast Frames by IPv4 ARP

FF:FF:FF:FF:FF:FF

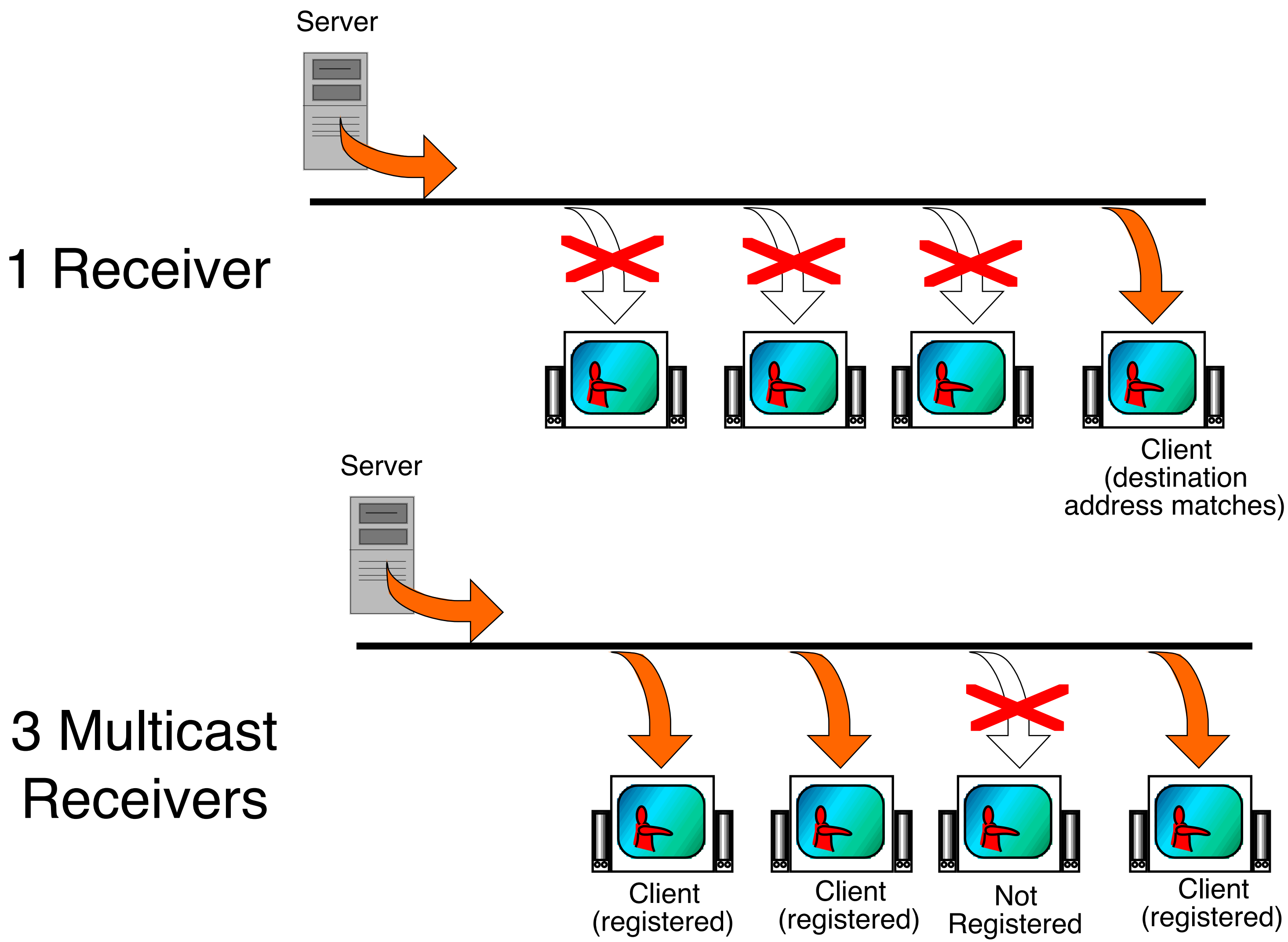
A sender using the IPv4 Address Resolution Protocol (ARP) sends an ARP request to discover the MAC address of the interface (NIC) with which it wishes to communicate.

The ARP request therefore is sent as a ***broadcast frame***.
This request is received by ***all systems*** on the same Ethernet LAN.

In contrast, the interface (NIC) responding to an ARP request already knows the address of the system sending the ARP request.

The ARP reply is sent in a ***unicast frame*** directed to the querier.
Only the querier receives this requested response.

Sending to multiple recipients: Multicast on Ethernet



TV/Radio/etc Transmission (can often have several receivers)

Also used by some protocols to deliver to multiple computers

IPv4 Group MAC Addresses

01:00:5E	:00:00:01

*

Groups addresses
- Have the ***least significant bit*** of the ***first byte*** to 1
- The remainder of the address carries the specific group address
- Last 23 bits of the IP group destination address, e.g., 224.0.0.1

Group addresses identify ***channels*** not Receivers
- Sender chooses a group address to use
- e.g. one channel may carry a specific Internet TV station
- another channel might be used to advertise DNS in a LAN

NICs need to ***register*** to receive from a group
- A computer may ***register*** none or more group addresses
- e.g. a multicast DNS client registers IP address 251.0.0.224
  - This registers for the MAC address of 01:00:5E:00:00:FB
- The NIC passes all frames that match a registered group address

** IPv4 Address mapping*

IPv6 Group MAC Addresses

33:33:xx:xx:xx:xx

Groups addresses

Have the ***least significant bit*** of the ***first byte*** to 1

The remainder of the address carries the specific group address copied from the last 32-bits of the IPv6 group destination address.

IPv6 doesn't use broadcast packets at all

Instead it uses multicast to send packets to groups of receivers

Some Layer 2 protocols also use multicast:

e.g. the Spanning Tree uses address 01-80-C2-00-00-00 to send control frames to the next adjacent Ethernet Switch.

The sender doesn't know the MAC address used by a switch, but does not want its frames to be received by other NICs.

Addressing Summary

- **All NICs have a MAC Address**

  Provides a handy income stream to the IEEE :-)

- **All NICs receive every frame with:**

  a ***broadcast*** MAC destination address ff:ff:ff:ff:ff:ff

  a destination address that matches its ***MAC address***

  a destination address that matches a ***registered*** multicast group address (i.e. used by a program on the computer)

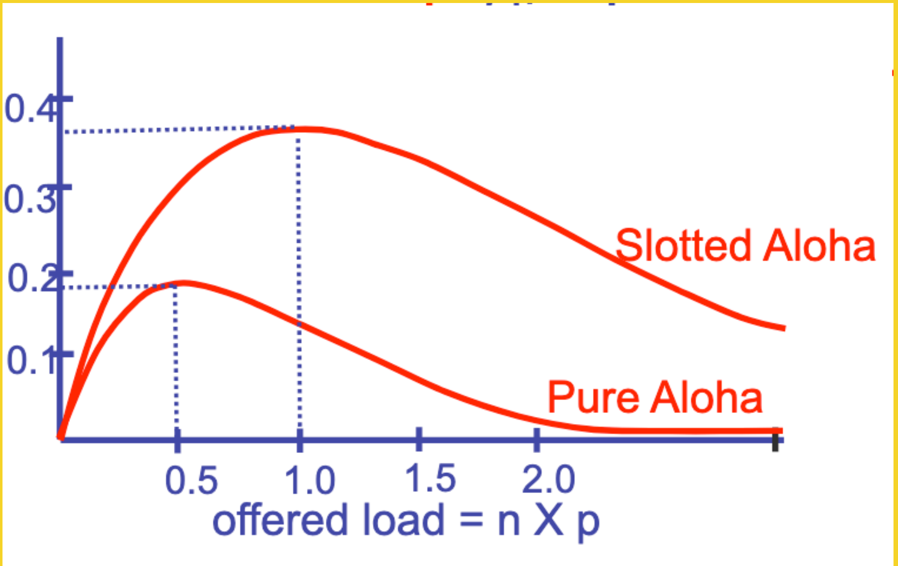
- **All filtering is performed within the NIC:**

  Computer does not know about discarded frames

  A computer can override filtering, by placing the NIC into ***promiscuous mode*** - where all frames are received

Ethernet Frames:

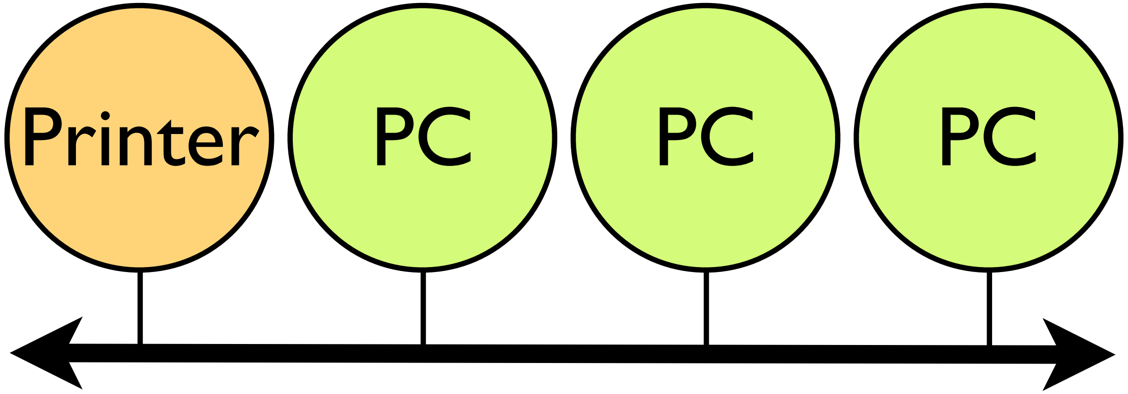
A shared physical medium



Link Layer

Module 2.2

Sharing the media



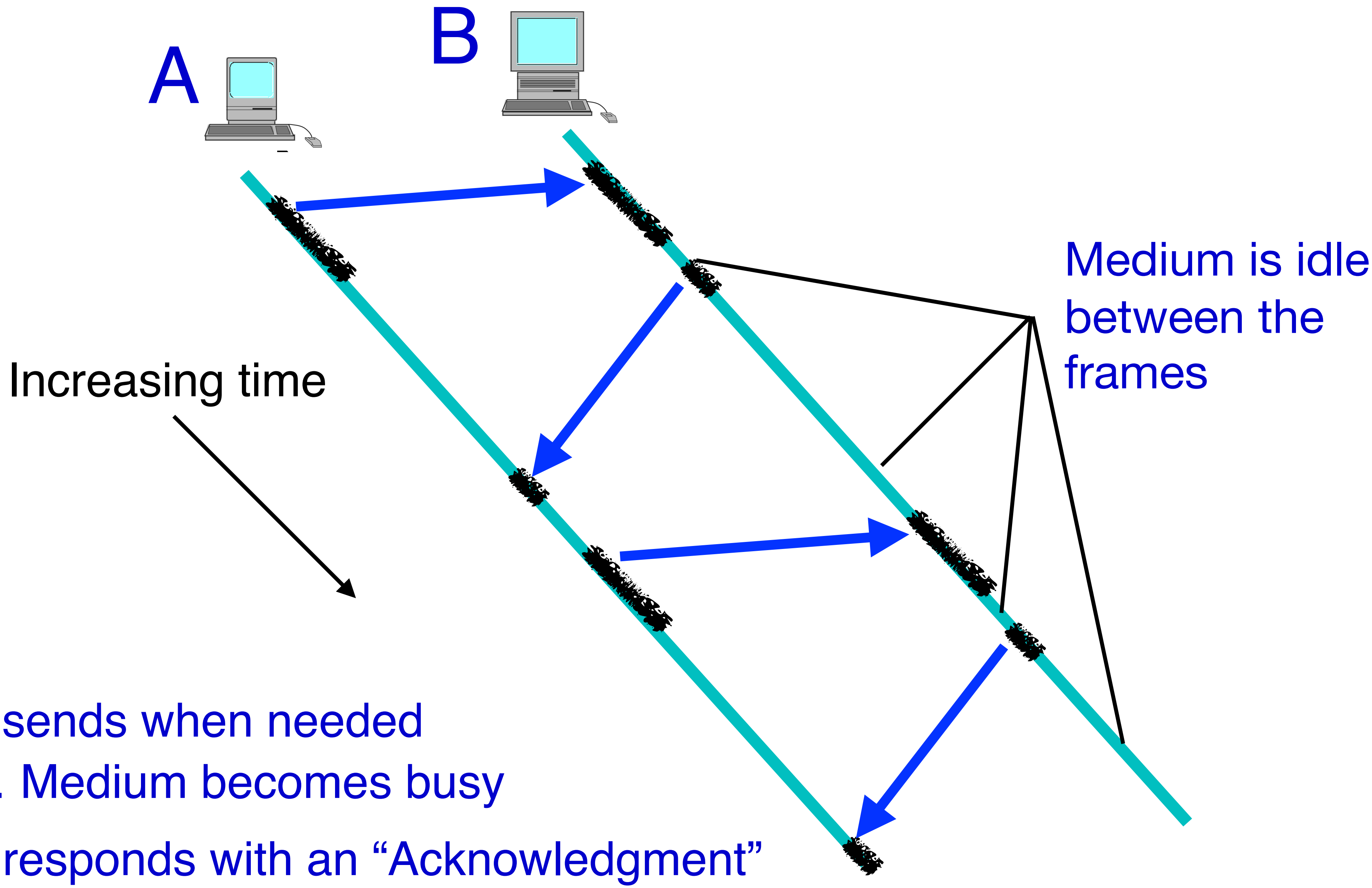
There is only one medium (the cable)

All NICs ***should*** be able to use this cable

Clearly only ***one*** should send at a time!

So, how does a NIC ***know*** if it may send?

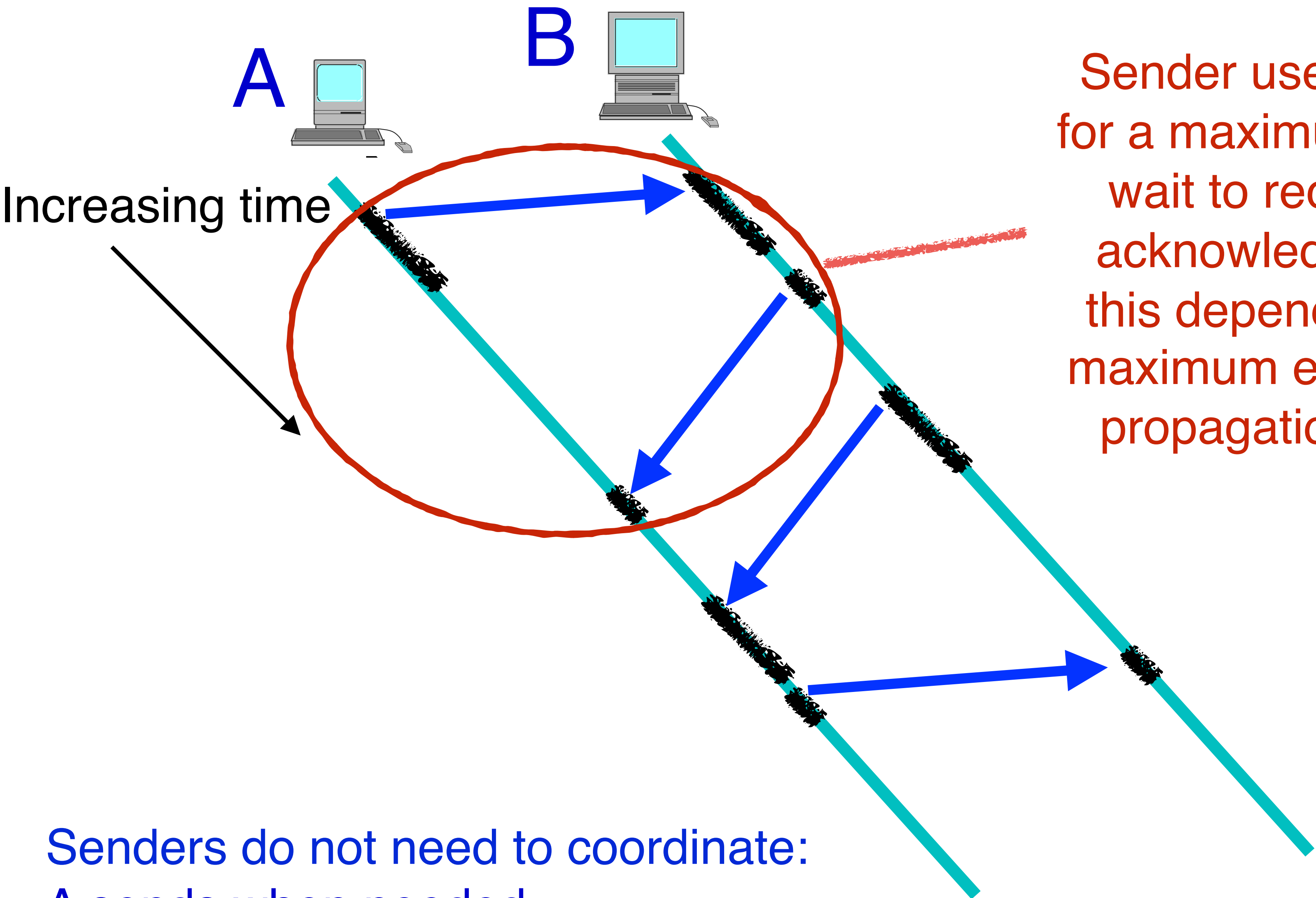
Medium Access using ALOHA



A sends when needed
… Medium becomes busy

B responds with an “Acknowledgment”
Either … A knows that B has received the data
… or ... it will resend the data again to B

Medium Access using ALOHA

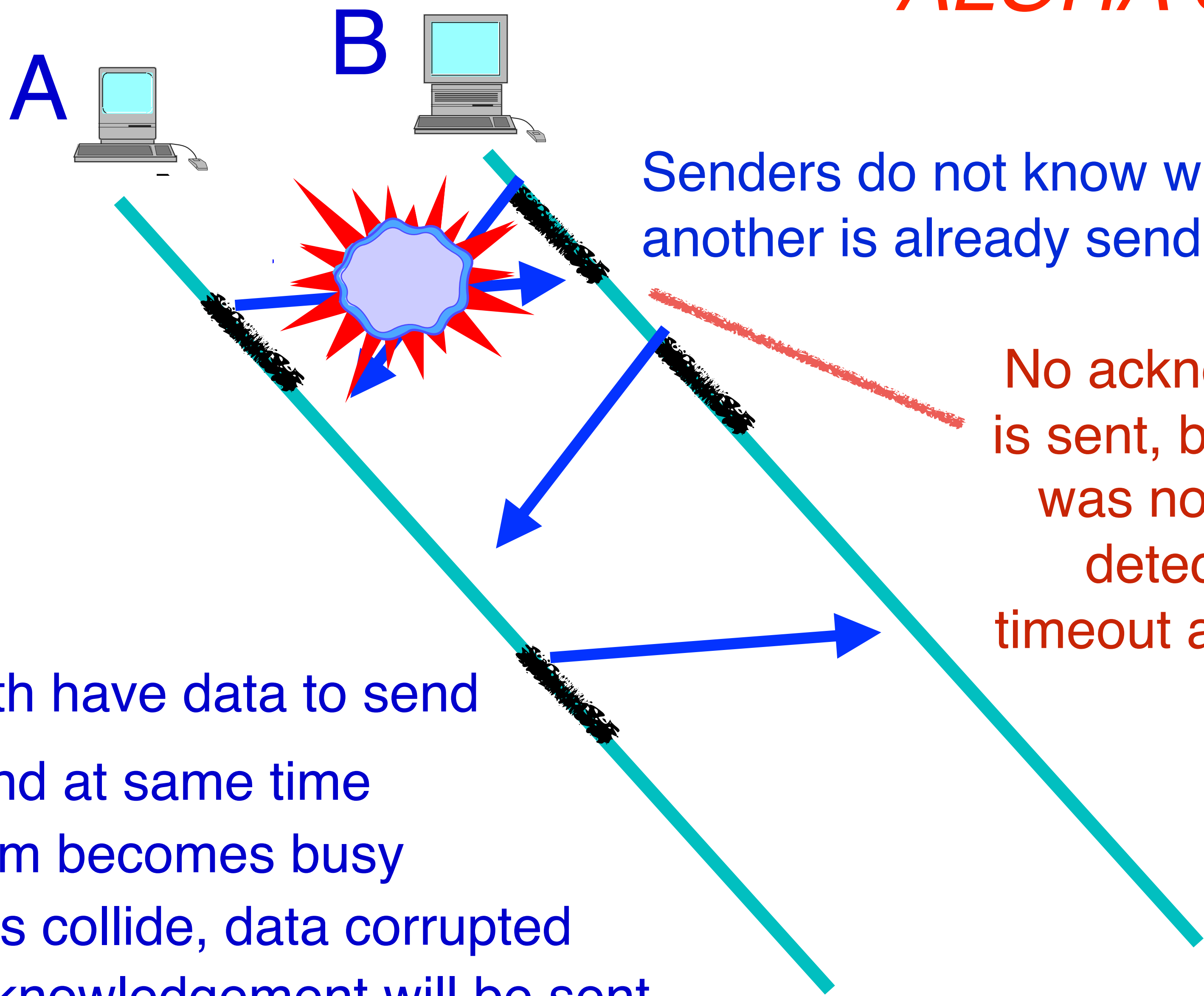


Sender uses a timer for a maximum time to wait to receive an acknowledgement, this depends on the maximum end-to-end propagation delay.

Senders do not need to coordinate:
A sends when needed
B sends when needed

Half Duplex Communication

ALOHA Collision



Senders do not know when another is already sending!

No acknowledgment is sent, because data was not received, detected by a timeout at the sender

A & B both have data to send

A & B send at same time

… Medium becomes busy

… Signals collide, data corrupted

… No acknowledgement will be sent

A & B will both need to send again later

As the load increases, the chances of collision also increases

Slotted ALOHA

If there is a common clock source we can divide time into slots.
All senders need to know the start of each timeslot
Senders only transmit a frame at the start of a timeslot

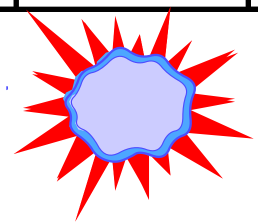
Timeslot	1	2	3	4	5	6	7
Sender 1							
Sender 2							
Sender 3							
Outcome	1	Empty	2	3	...		

Increasing time

Slotted ALOHA

If there is a common clock source we can divide time into slots.
All senders need to know the start of each timeslot
Senders only transmit a frame at the start of a timeslot
Timeslots with only one frame result in successful transmission

Timeslot	1	2	3	4	5	6	7
Sender 1							
Sender 2							
Sender 3							
Outcome	1	Empty	2	Collision	3	2	1



Efficiency of ALOHA

Suppose *n* senders have data to send with probability *p*
The probability of success for ALOHA,

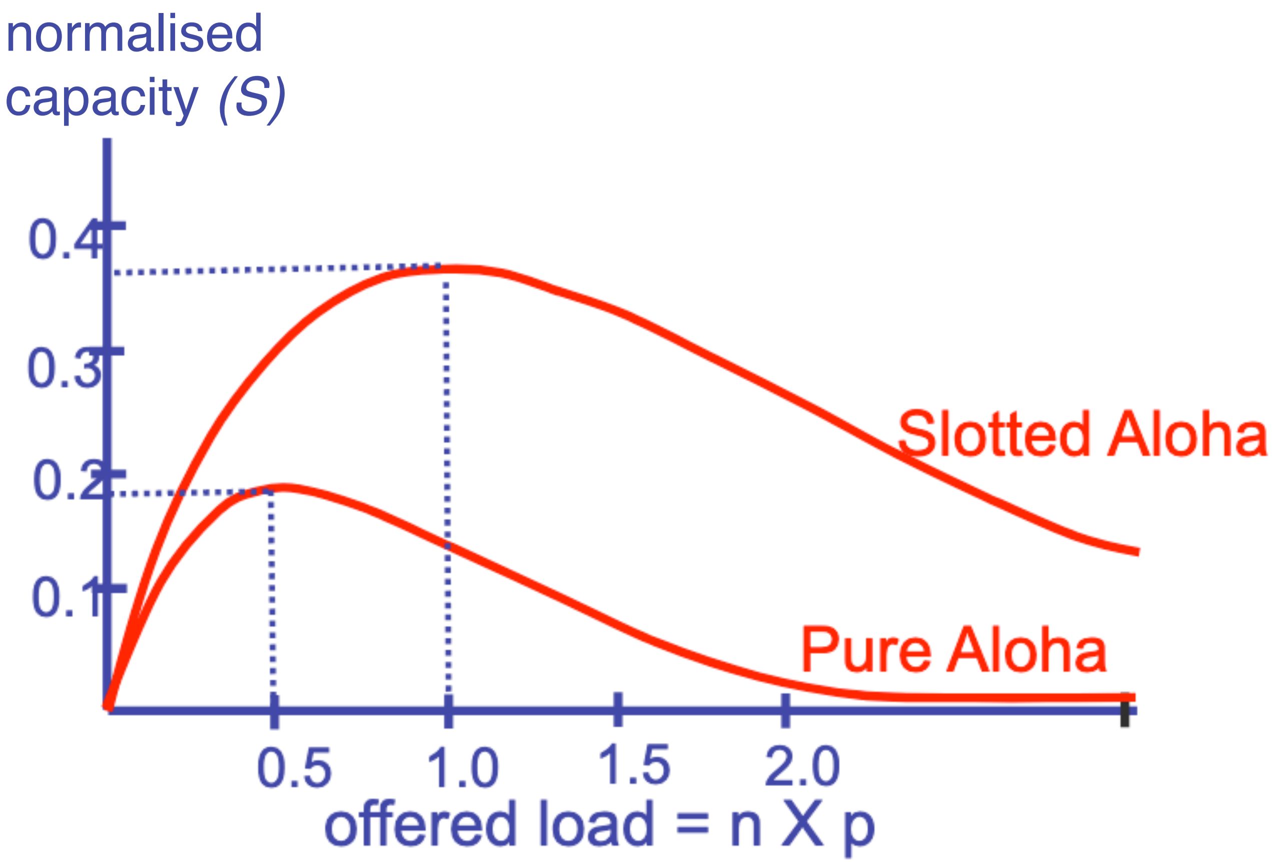
$S = p(1-p)^{(n-1)}$

Maximum capacity
S= 0.18

For slotted ALOHA,

$S = np(1-p)^{(n-1)}$

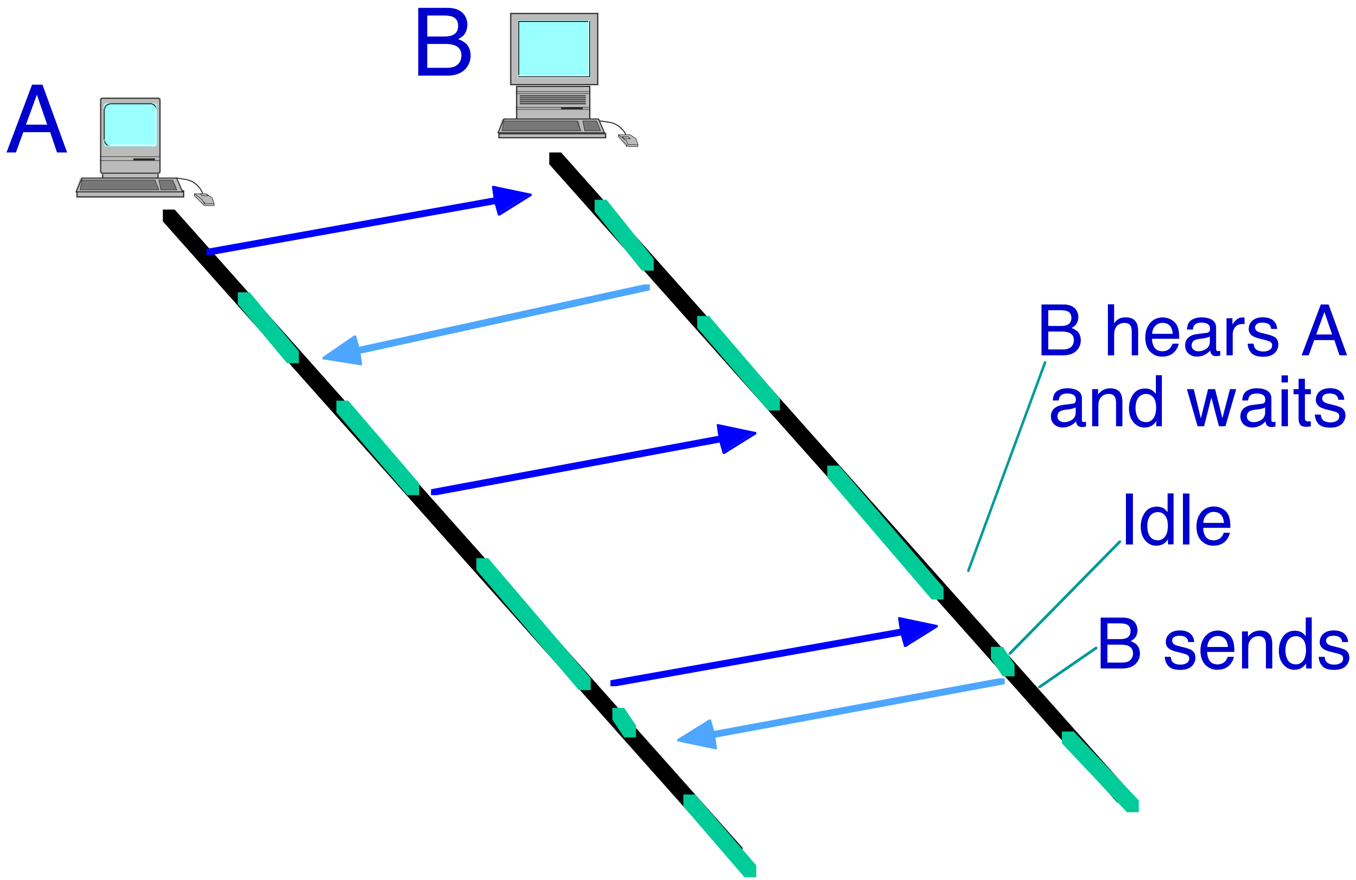
For optimal p,
S= 1/e = 0.37



Slotted ALOHA is much better than ALHOA, but still achieves <37%

Listen-Before-Talk

Listens for activity on the cable before sending
Requires Carrier Sense (CS) circuit



Also called *Carrier Sense Multiple Access* (CSMA)
Does not work well when one sender is a *long distance* from another

ALOHA Summary

- **ALOHA is really very simple**

  Requires setting a timer to detect loss of an acknowledgment

- **Slotted ALOHA**

  Slotted ALOHA more efficient than unspotted version

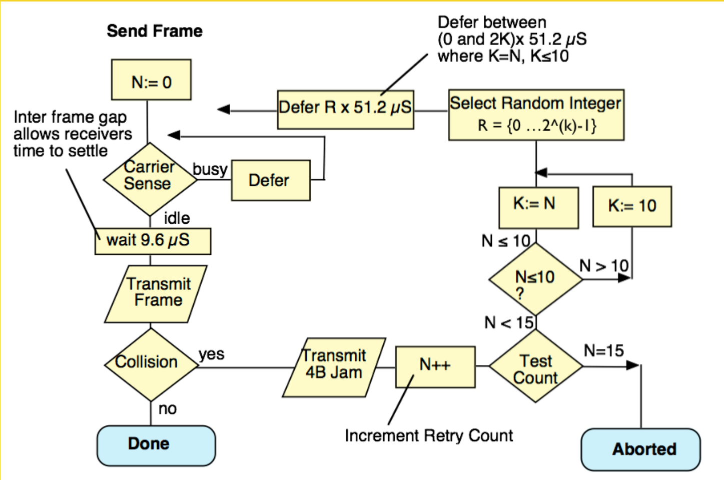
- **Carrier Sense or Listen Before Talk**

  Carrier Sensing improves efficiency

  Not the design chosen for Ethernet, but still used in other networks

Ethernet Frames:
Medium Access Control

Medium Access Control (MAC) needs to solve three challenges:
- how to be decentralised with no "master" controller
- how to scale to large numbers of active nodes
- how to deal with propagation delay

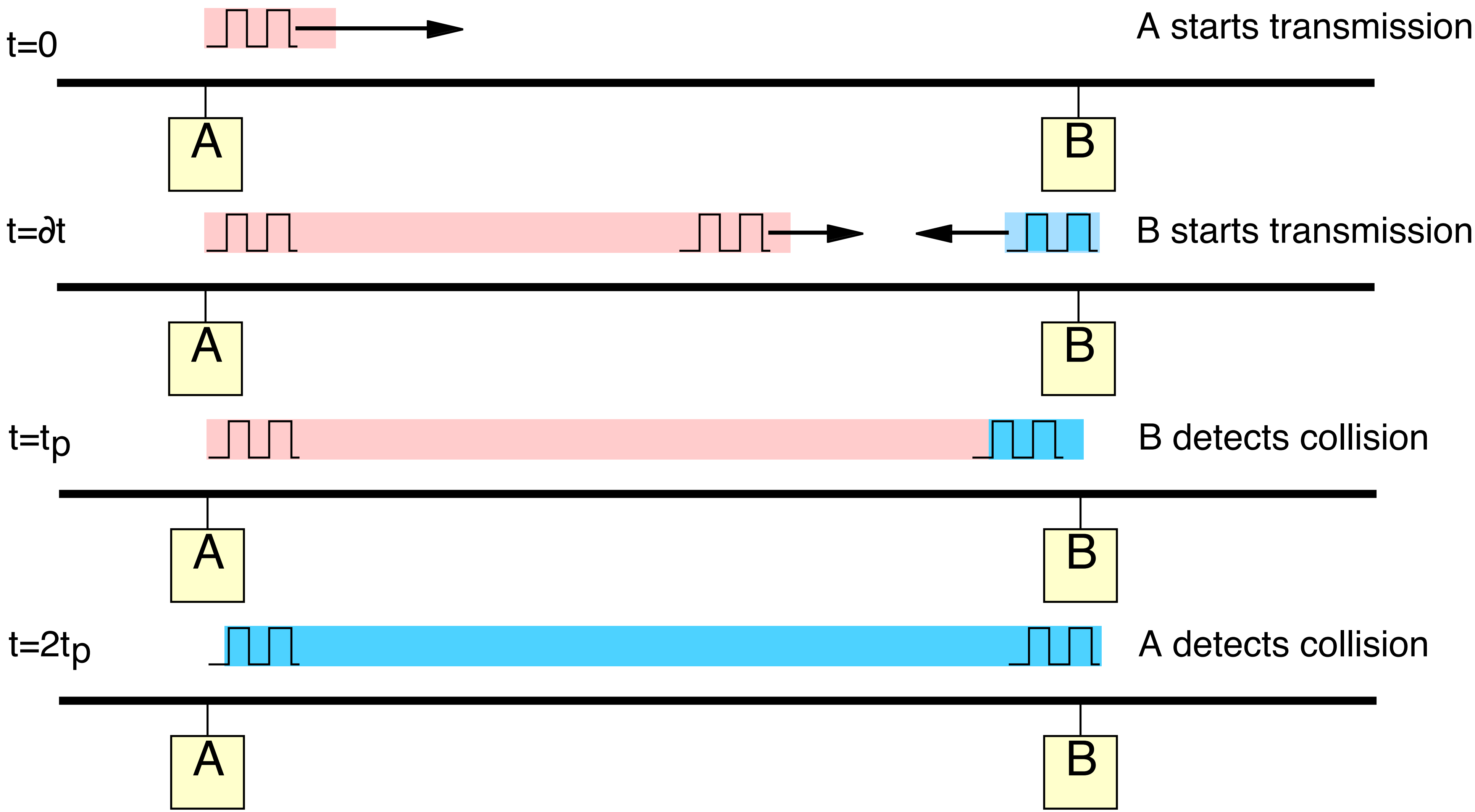


Link Layer

Module 2.3

Collisions and Collision Detection

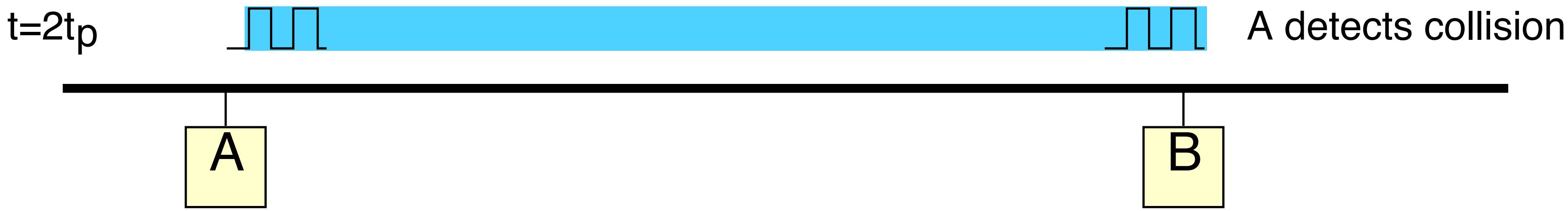
Nodes try to avoid collisions by using a Carrier Sense (CS) to detect when the medium is idle before they start sending



Requires Collision Detect (CD) to detect a collision

Slot Time

All senders need to know when any collision occurs



The time to detect a collision depends on the propagation delay
... and other delays

In a CSMA/CD system this is set by the ***slot time***

The slot time for IEEE 802.3 of **51.2** μs at 10 Mbps

This limits the maximum cable ***distance*** to 3km at 10 Mbps

The need to detect a collision sets the ***minimum*** frame size
The minimum Ethernet frame size is 64B (60 bytes+CRC32)

Parameters impacting the Slot Time

Component	Properties	Delay (microsec)
AUI Cable	6x 50m , 0.65c	3.08
Transceiver	3 transceivers (6x 1.2 micosec)	7.2
3xCoax Medium	e.g. 1500m, 0.77c	13
2xOther Media	e.g.1000m, 0.65c	10.26
Repeater delay	Propagation delay	2
Signal Rise Time		8.4
Elec Circuit	Propagation delay	1.05
Total		44.99

Total Slot Time of system < 51.2 μs

This is for informational only (not required in the exam)

Retransmission after Collision

The minimum frame size assures us that all nodes that
are sending will ***detect*** the collision.

After detecting a collision, sends a JAM and then stops sending.

The data has not been sent, and therefore needs retransmitted.

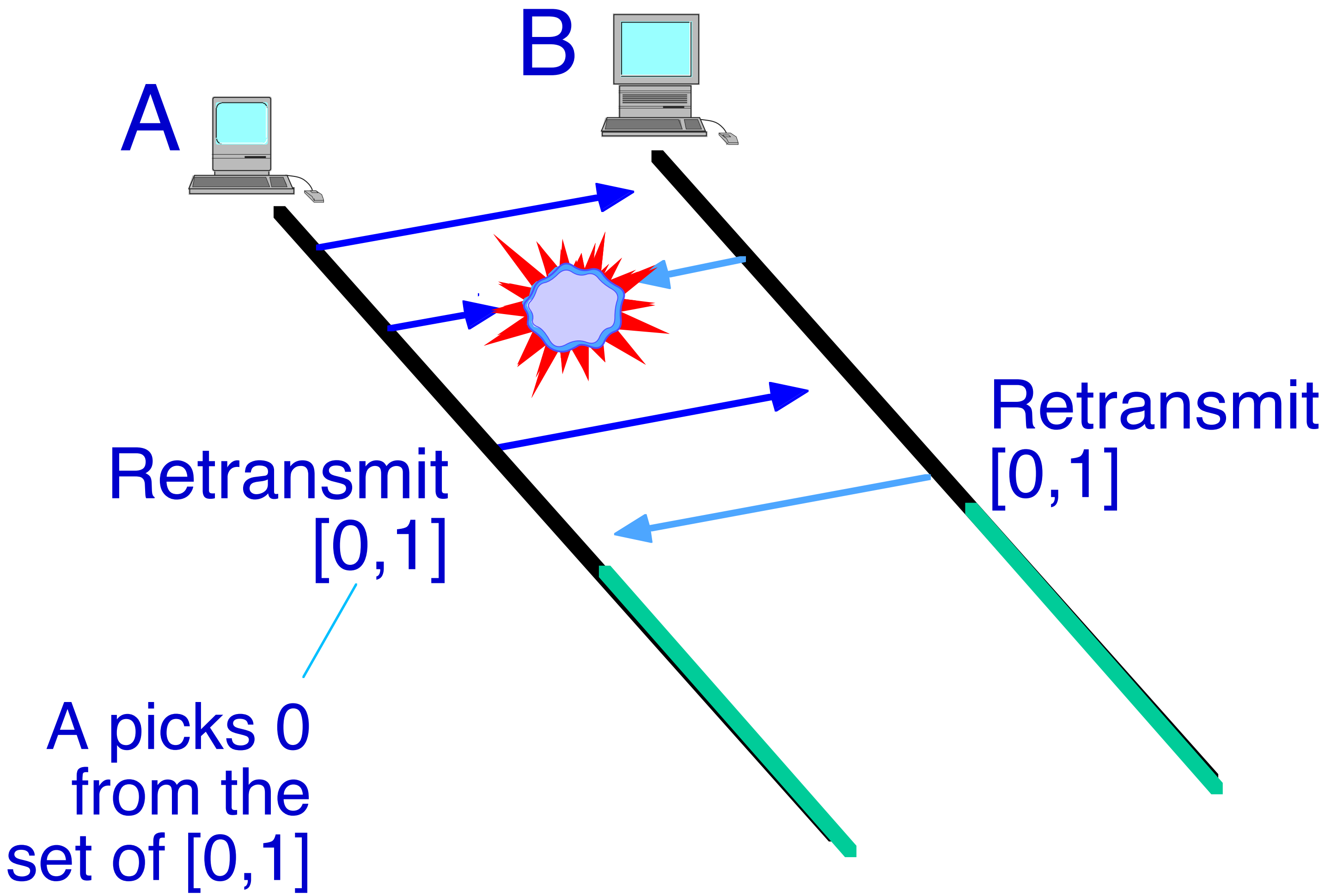
They need to send at different times to not suffer another collision.

A decentralised method has no way to know which node or
nodes experienced the collision.

The method therefore chooses to ***random backoff time***
to delay their own retransmission.

If they choose different random backoff times, they suceed.

Backoff and Retransmission



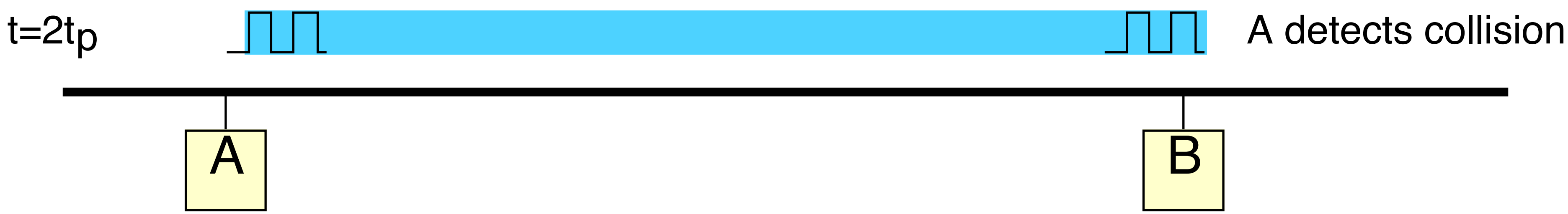
In this example, after the first collision k=1

A & B choose from a set of 2^k values: In this case: [0, or 1]

50% probability that A & B choose different retransmissions

A happens to choose [0], and so waits t x 0. Therefore it sends first

Detail of Random Exponential Backoff



If multiple NICs retransmit at the same time, a collision will occur again

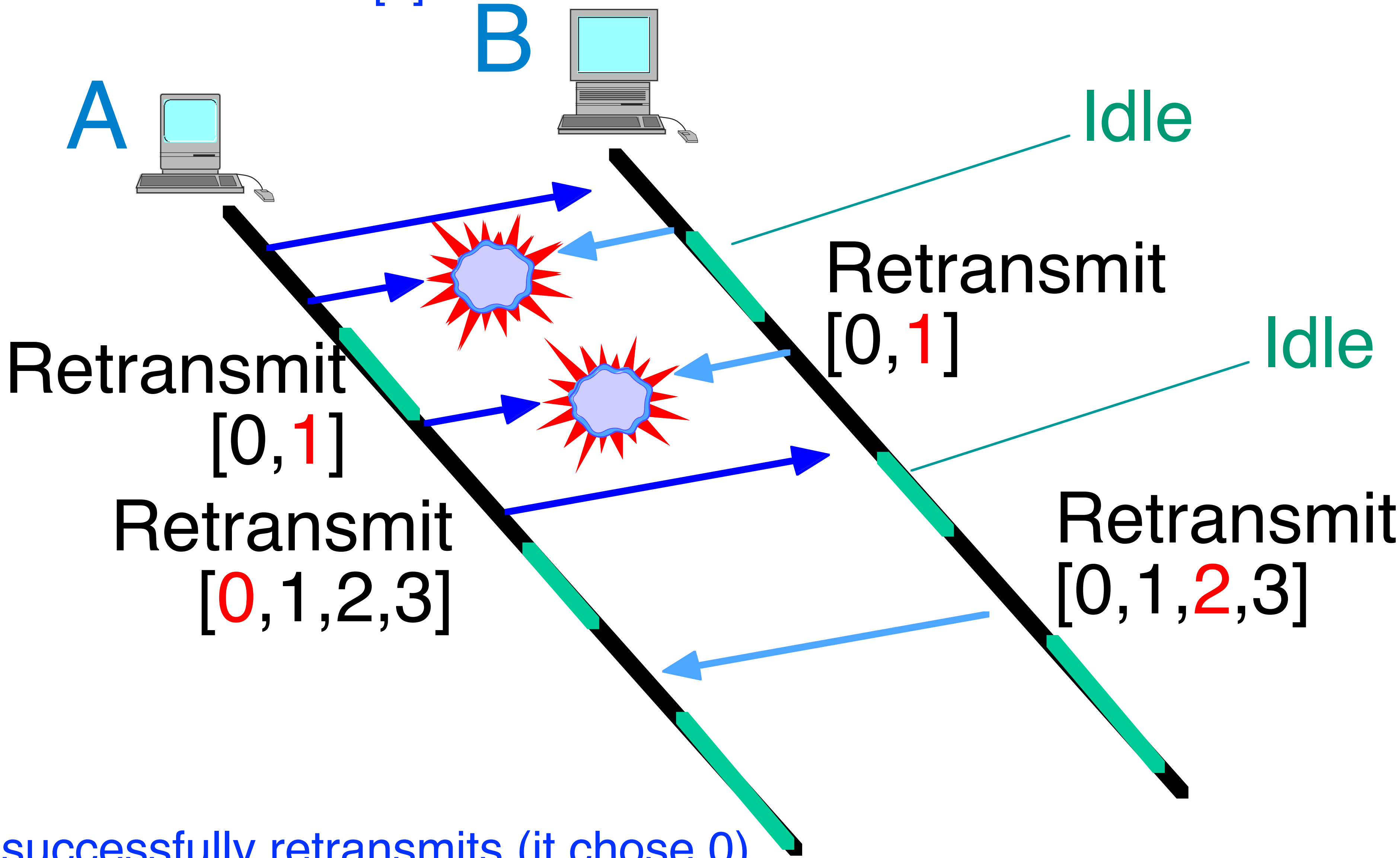
- Senders jam the medium and then back-off!

Each sender waits for a randomly chosen period of time

- k counts the set of values, increasing each retransmission, initially k=1
- Senders choose a random number from a set of values [0... (2^k-1)]
- Wait for the chosen value multiplied by Ethernet Slot Time (51.2μS)
- Each attempt increases k, so the set increases ([0,1], [0,1,2,3],[0..7]…)
- This exponentially increases the random backoff set

Exponential Back-Off

In this example there is a collision; both A,B happen to choose the same backoff time [1] and so a second collision must occur...



A successfully retransmits (it chose 0)
B defers one slot time (it chose 1)

Random Backoff

[0,1] First Retx		
Random number at A	Random number at B	Result
0	0	Collision
0	1	A sends first
1	0	B sends first
1	1	Collision after 1 slot time

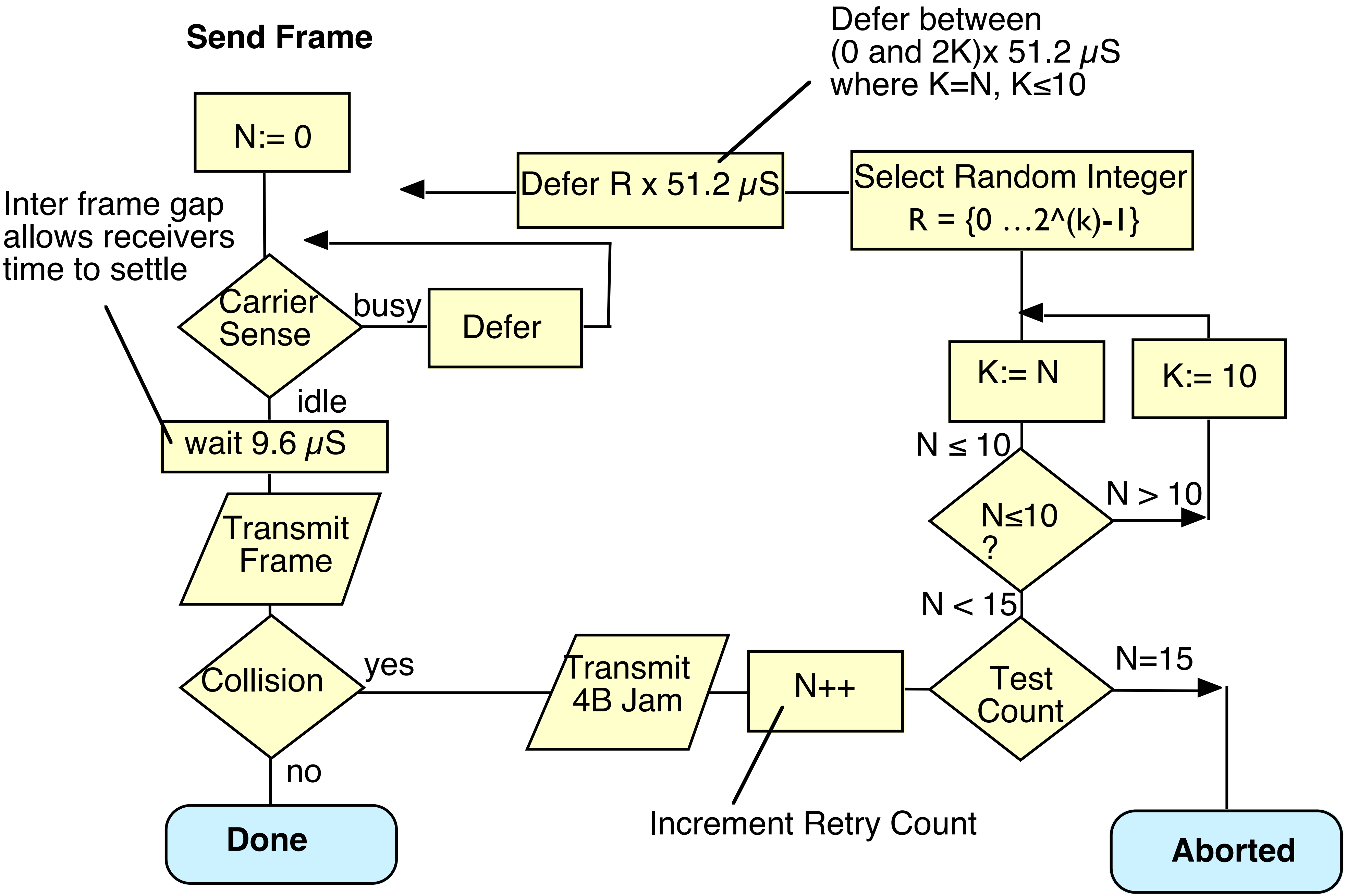
[0,1,2,3] Second Retx		
A	B	Result
0	0	Collision
0	1	A sends
0	2	A sends
0	3	A sends
1	0	B sends
1	1	Collision
1	2	A sends
1	3	A sends
2	0	B sends
2	1	B sends
2	2	Collision
2	3	A sends
3	0	B sends
3	1	B sends
3	2	B sends
3	3	Collision

1st attempt 50% chance of collision - fair to each sender

2nd attempt 25% chance of collision - fair to each sender

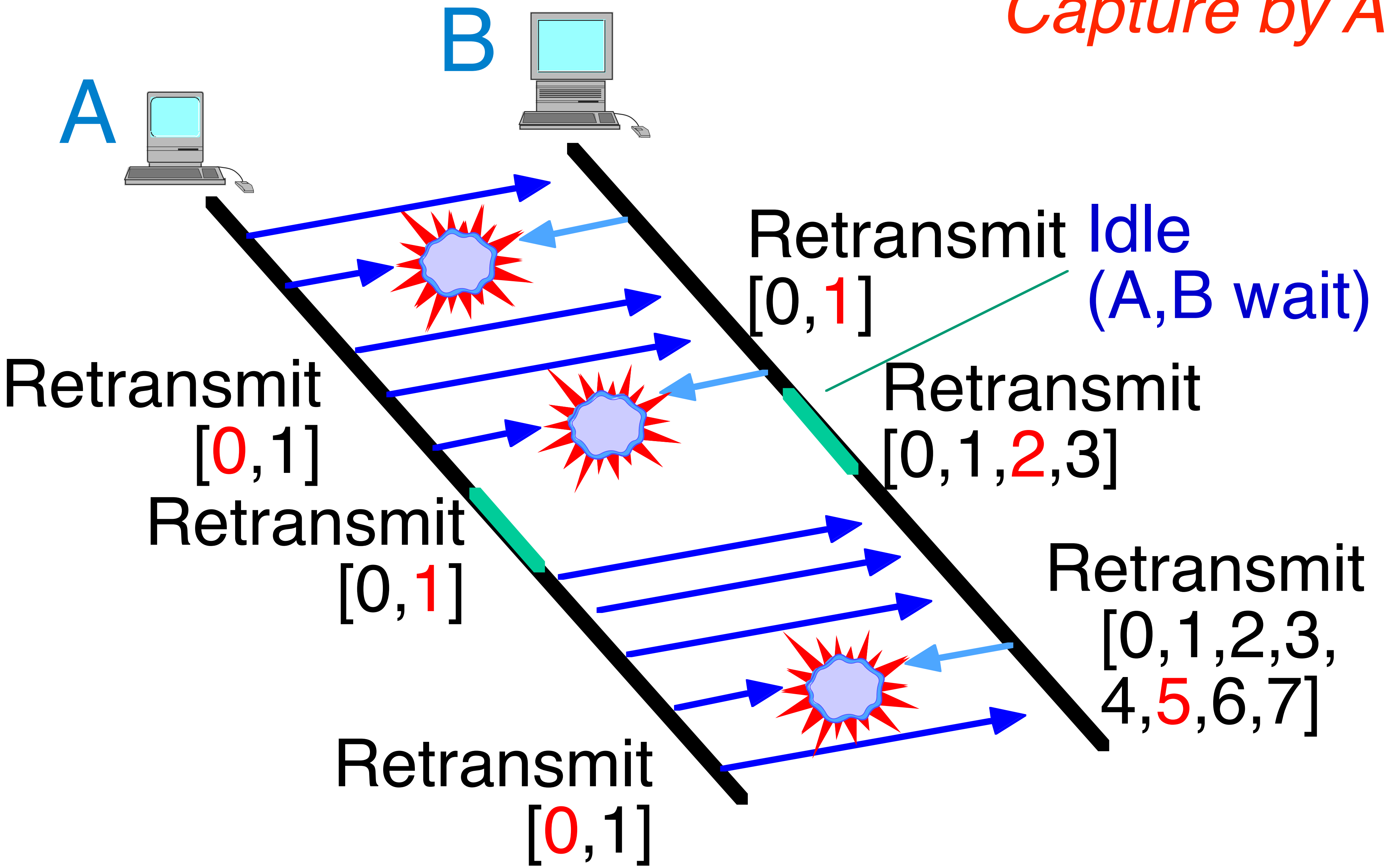
The collision probability halves each retransmission round <10

CSMA/CD



Send Frame
Defer between
(0 and 2K)x 51.2 μS
where K=N, K≤10
N:= 0
Defer R x 51.2 μS
Select Random Integer
R = {0 …2^(k)-1}
Inter frame gap
allows receivers
time to settle
Carrier
Sense
busy
Defer
idle
wait 9.6 μS
K:= N
K:= 10
N ≤ 10
N≤10
?
N > 10
Transmit
Frame
N < 15
Collision
yes
Transmit
4B Jam
N++
Test
Count
N=15
no
Done
Increment Retry Count
Aborted

Capture by A



The algorithm becomes unfair when one NIC sends more than others

This might be common for a router/WIFI Access Point

A brief idle period after sending many frames, restores the fairness

Multiple Access - Summary

Each of these techniques is in use in some form of network

ALOHA

Problem: Many collisions when many nodes

Efficiency: 100% (1 node) 18% (many)

S-ALOHA

Requires Timeslot Synchronisation

Problem: Still collisions when many nodes

Efficiency: 100% (1 node) 37% (many)

Listen-Before-Talk (CSMA)

Requires Carrier Sense (CS) circuit

Problem: Fewer collisions, but still possible

Collision Detection (CSMA/CD)

Requires Carrier Sense (CS) and Collision Detect (CD) circuits

Problem: Capture possible - benefits from limiting burst size

Efficiency: 100% (1 node) higher (many)

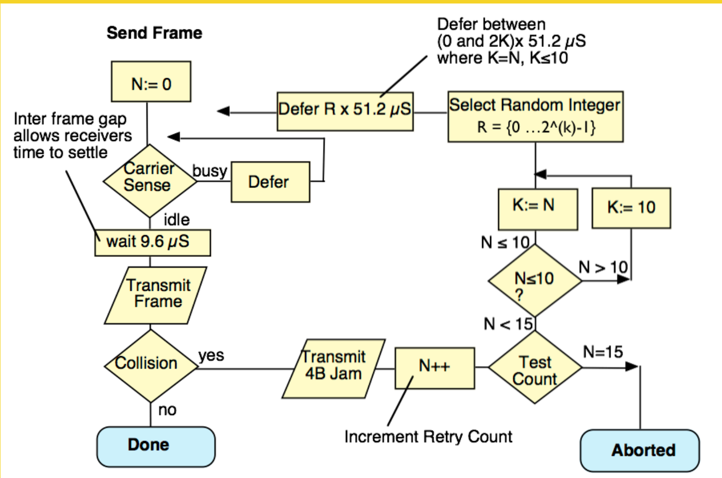
Recap: Strengths v Weakness of CSMA/CD

- **Strengths**
  No controlling system needed to coordinate use Ethernet
  Easy to add new systems (NICs) - just plug and play!
  Performance "reasonably fair"

- **Weakness**
  Performance degrades with increasing load
  One "busy" system can "capture" capacity
    - more of a problem for "upstream"
    (e.g., a WiFi base station, router)
    Could fix by limiting bursts of transmission

- **On balance, this was a good design!**

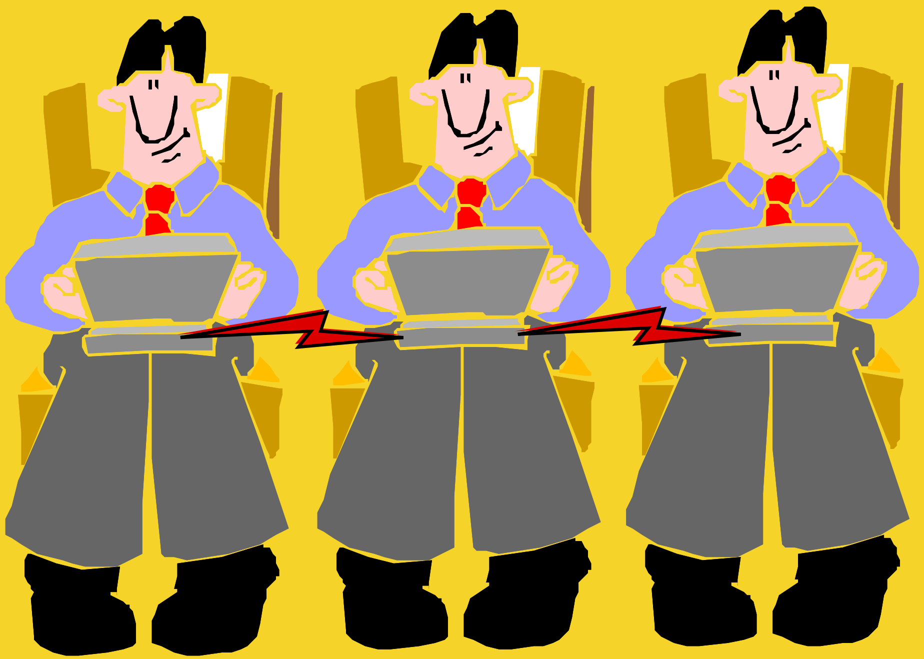
Ethernet Frames:
Medium Access Control



CSMA/CD

Module 2 - Additional Material

Wireless Ethernet



Wire-less physical layer
No cable

Module 2.4

Radio Link

2.4-2.485 GHz Industrial Science & Medicine (ISM) Band

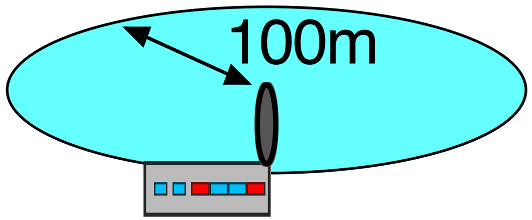
14 channels available worldwide
(fewer channels available in some countries)
Only 3 non-overlapping 20 MHz channels

Uses spread spectrum channels

First used by military ~ 50 years ago
Very high immunity to noise

RF Power

802.11b	100mW
Mobile Phone	600 mW
CB Radio	5W
Microwave Radio	2W



5.15-5.825 GHz Band also used for 802.11n (3 channels)

802.11 Success

WiFi deployment

- ~500,000 Hotspots in 144 countries!
- 1,000,000,000 chipsets since 2000
- 2.5 GHz, 5 GHz, 60 GHz

Speeds

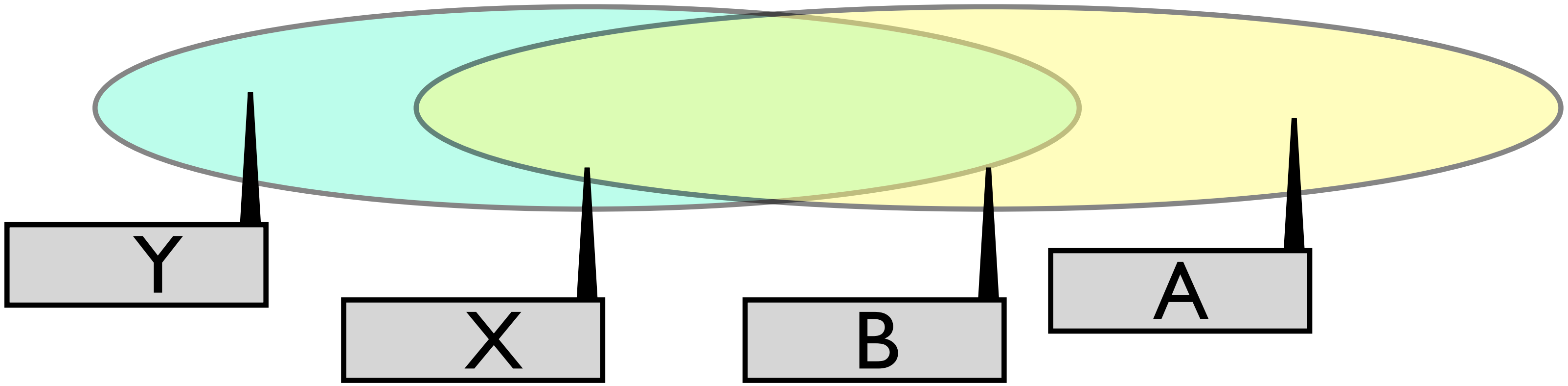
- Initial 11 Mbps
- Grew to 300 Mbps in a decade
- Since 2011, looking at 1 Gbps at short distances ~ 10m
- (rate reduces with distance at 100m or so, still only 11 Mbps)

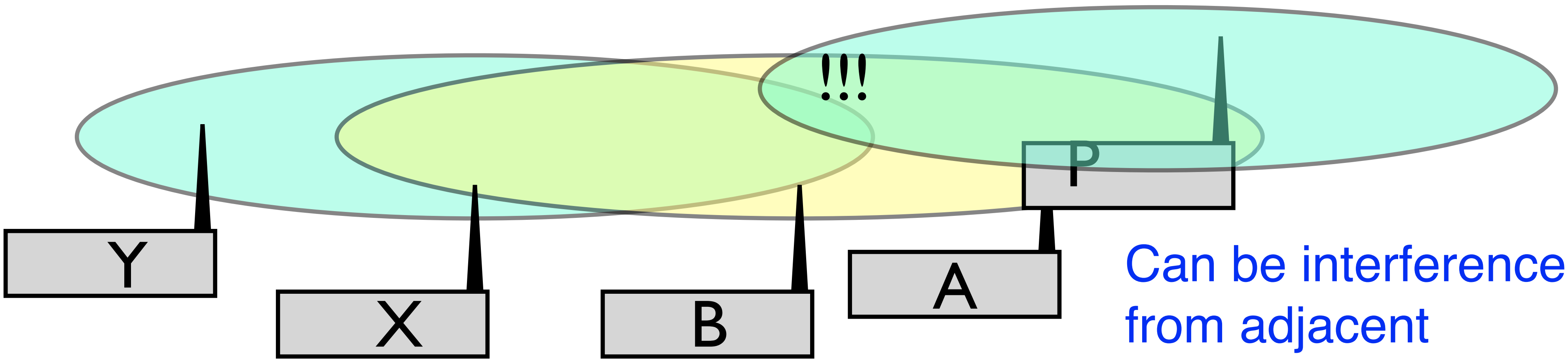
Frequency Channel Re-use

The ISM* frequency band allows several WiFi channels

All systems using a basestation use the same channel

This forms a logical network





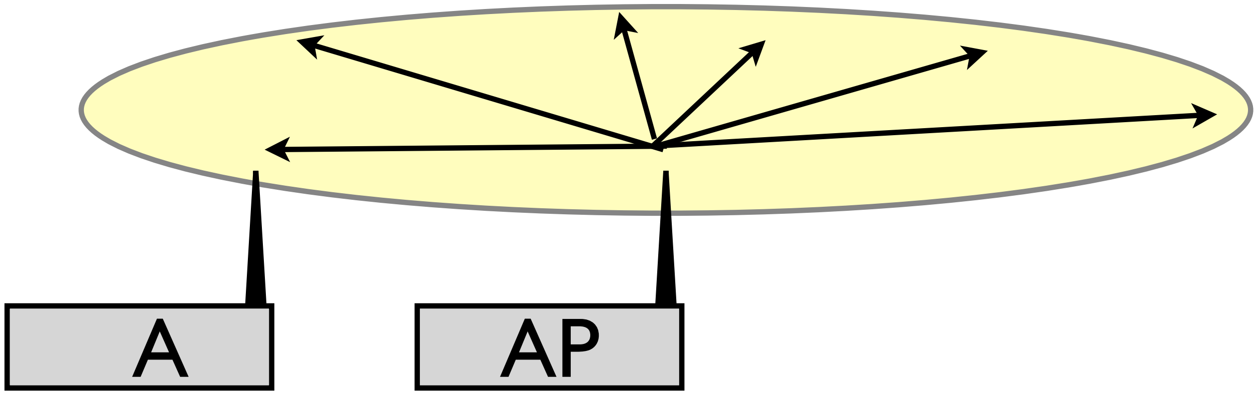
Can be interference from adjacent networks!

ISM - A frequency allocation for Industry, Science and Medical applications

Base Stations and Beacon Frame

How do you know which network you are using?

The WiFi access point (AP) broadcasts periodic beacon frames can also identify the network (SSID*)

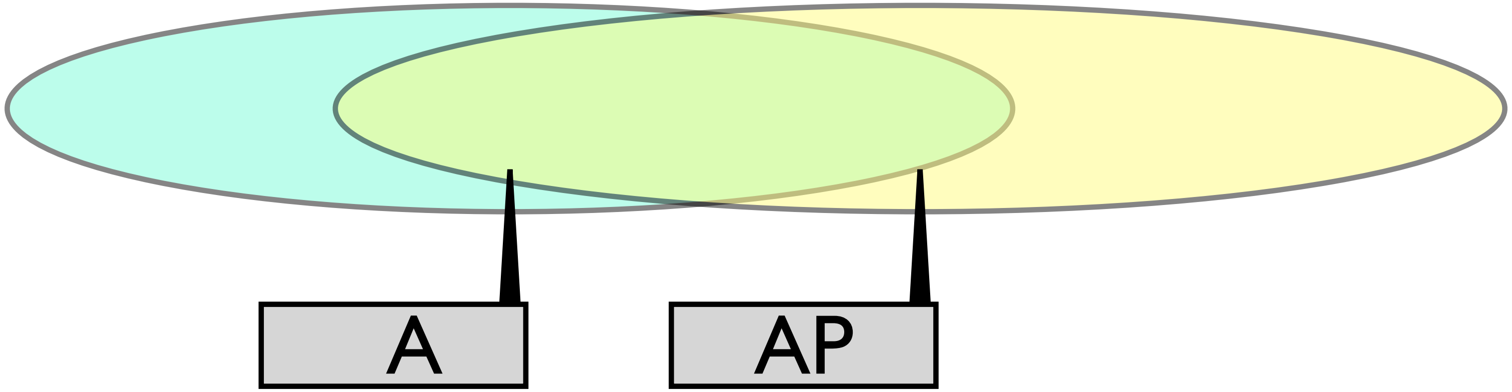


WiFi AP

The WiFi AP forms the logical centre of the WiFi network

SSID - Service Set Identifier, an Ethernet beacon frame
Beacon frames include the AP source MAC address
Beacon frames are sent to the broadcast MAC destination address

Wireless (802.11)

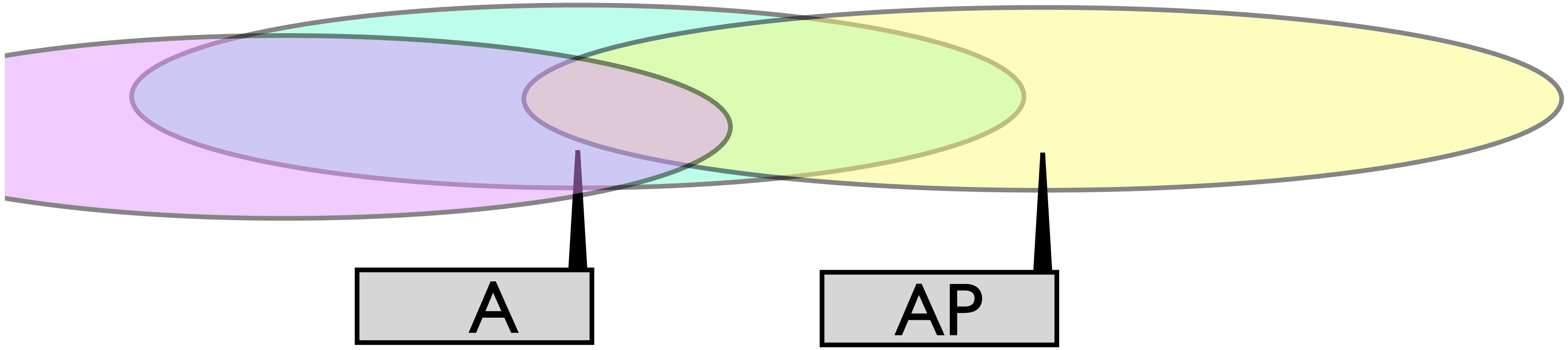


Each wireless node has a range
A is an end system; AP is a an access point

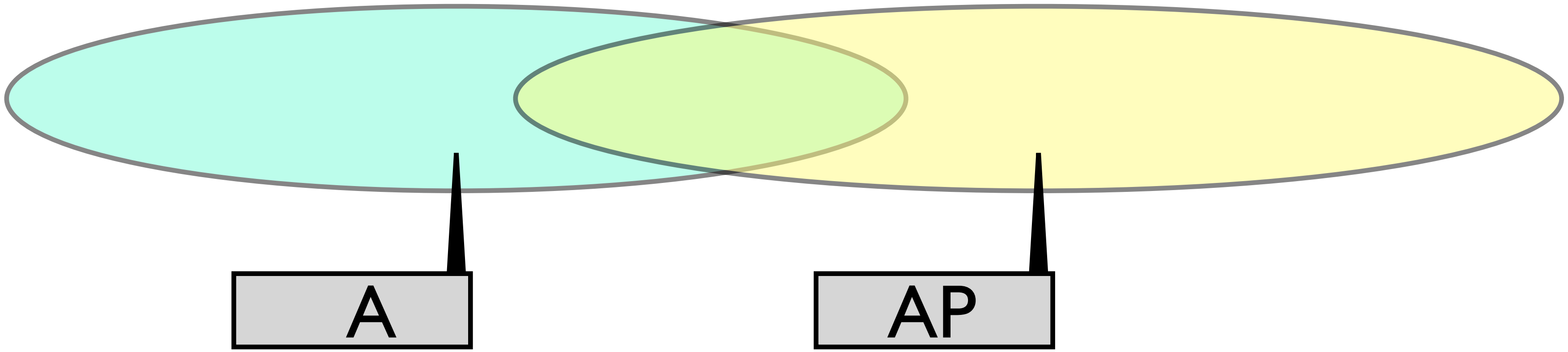
A needs to be able to receive signal from AP
(and AP from A)

When A sends to AP it can first sense the medium
(i.e. check if any system is sending)

Wireless (802.11)



A and AP can no longer communicate (interference)



A and AP can no longer communicate (signal strength)

Collision Avoidance

WiFi uses CSMA with ***Collision Avoidance***

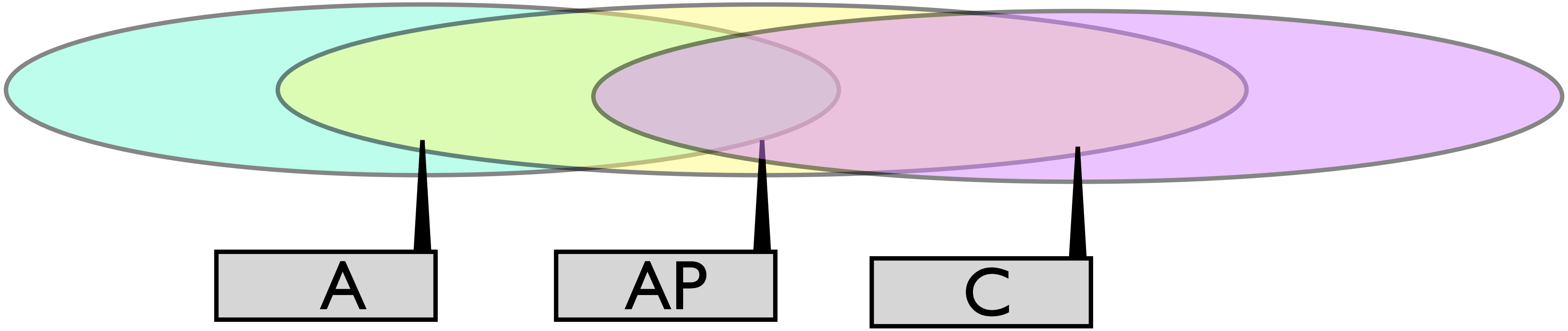
Three important changes:

1. A sender attempts to ***avoid*** causing a collision - it first listens to check the channel is idle (DCF Interframe Space). It then waits a randomly chosen time and if still idle, starts transmission.

2. A sender ***cannot monitor*** the entire wireless medium
   Receivers acknowledge (after a short delay) if a frame received. If no ACK is received within a timeout, the sender backs-off (as in CSMA/CD). Backoff increases with a limit of 5-7 attempts.

3. An optional procedure known as CTS/RTS detects hidden nodes.

Hidden Node Problem



Some nodes may not be able to “see” other transmissions

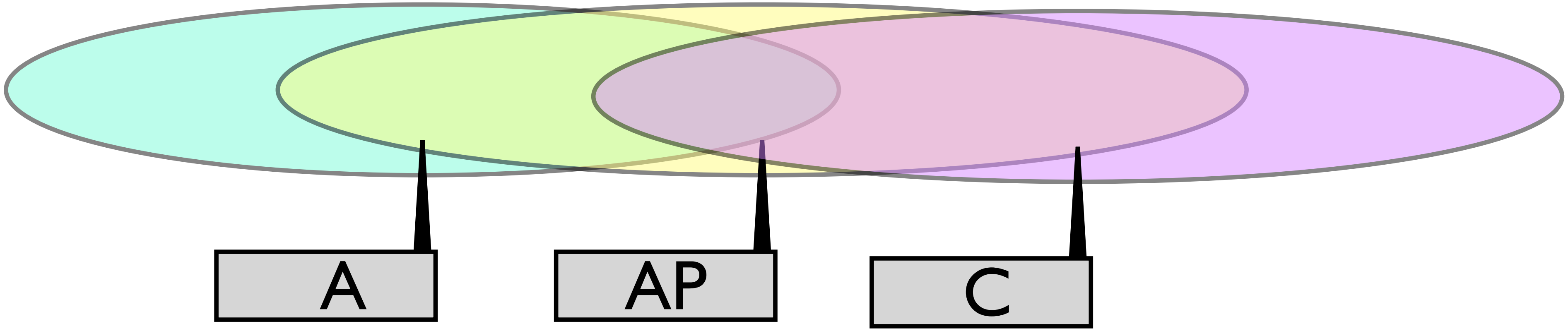
e.g.C does not know if A is sending

C may try to send to the AP (causing a collision)

Note 1: Wireless propagation can be very variable!

Note 2: By definition an AP sees signal from all nodes using AP

Virtual Carrier Detect



C first sends a ***Clear To Send*** frame to ask if it can transmit

- received by all nodes in range (i.e. Pink)

AP responds with an ***Ready To Send*** frame

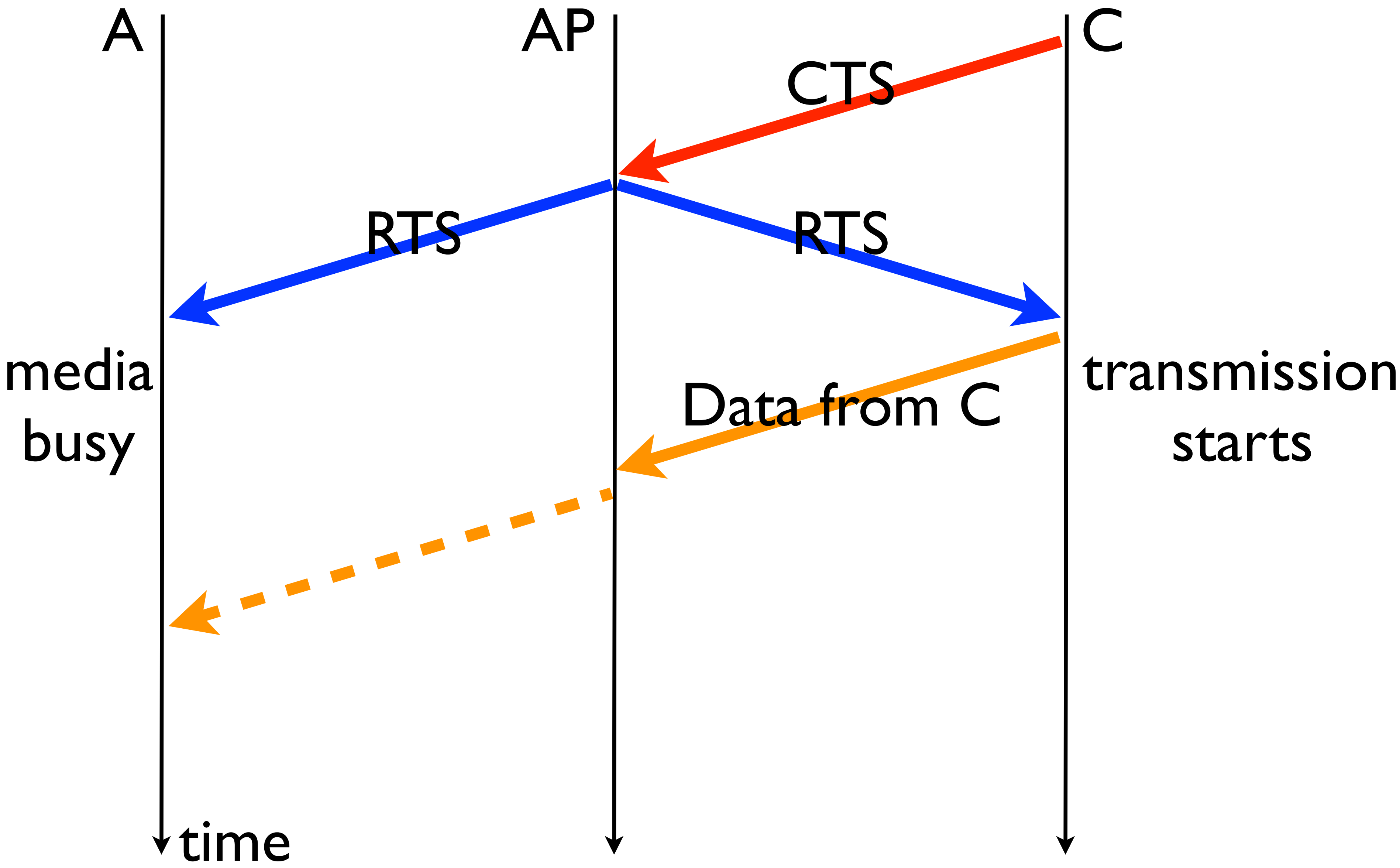
- received by all nodes in range (i.e. Pink & Yellow)

both now know the “channel is in use”

When Ready To Send is not received

sender must defer (“back-off”) before repeating Clear To Send

Hidden Node Problem and CTS/RTS



Note: If C needs to talk to A, it would rely on AP to relay (or repeat) the signal so that A can receive it.

Roaming

Several access points (APs) may form a LAN

APs connected together via a cabled LAN

Roaming between access points

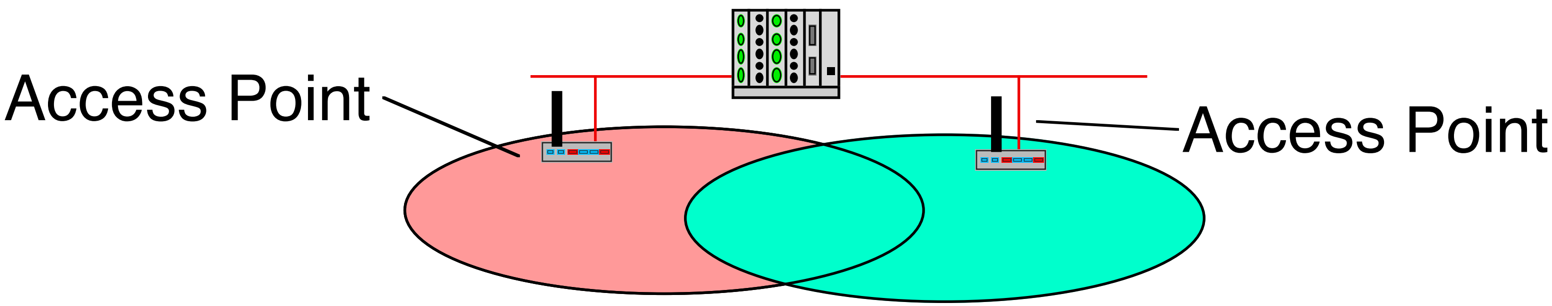
All APs sends a “beacon” signal to all nodes (SSID)

Multiple APs can advertise the same SSID

Nodes can select the AP with the best “beacon” signal

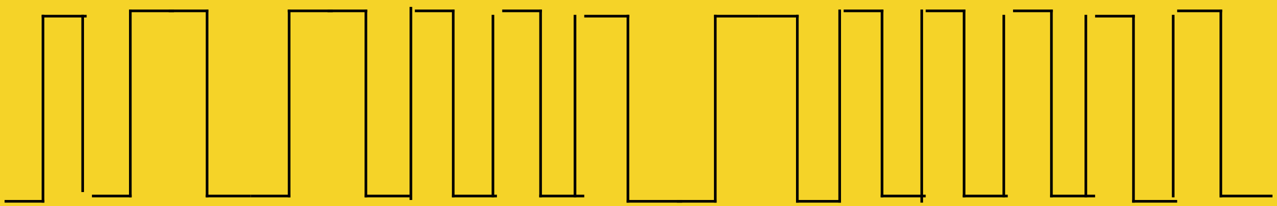
Wireless nodes keep the same MAC address

- users do not need to know the AP has changed!!



Ethernet Frames:

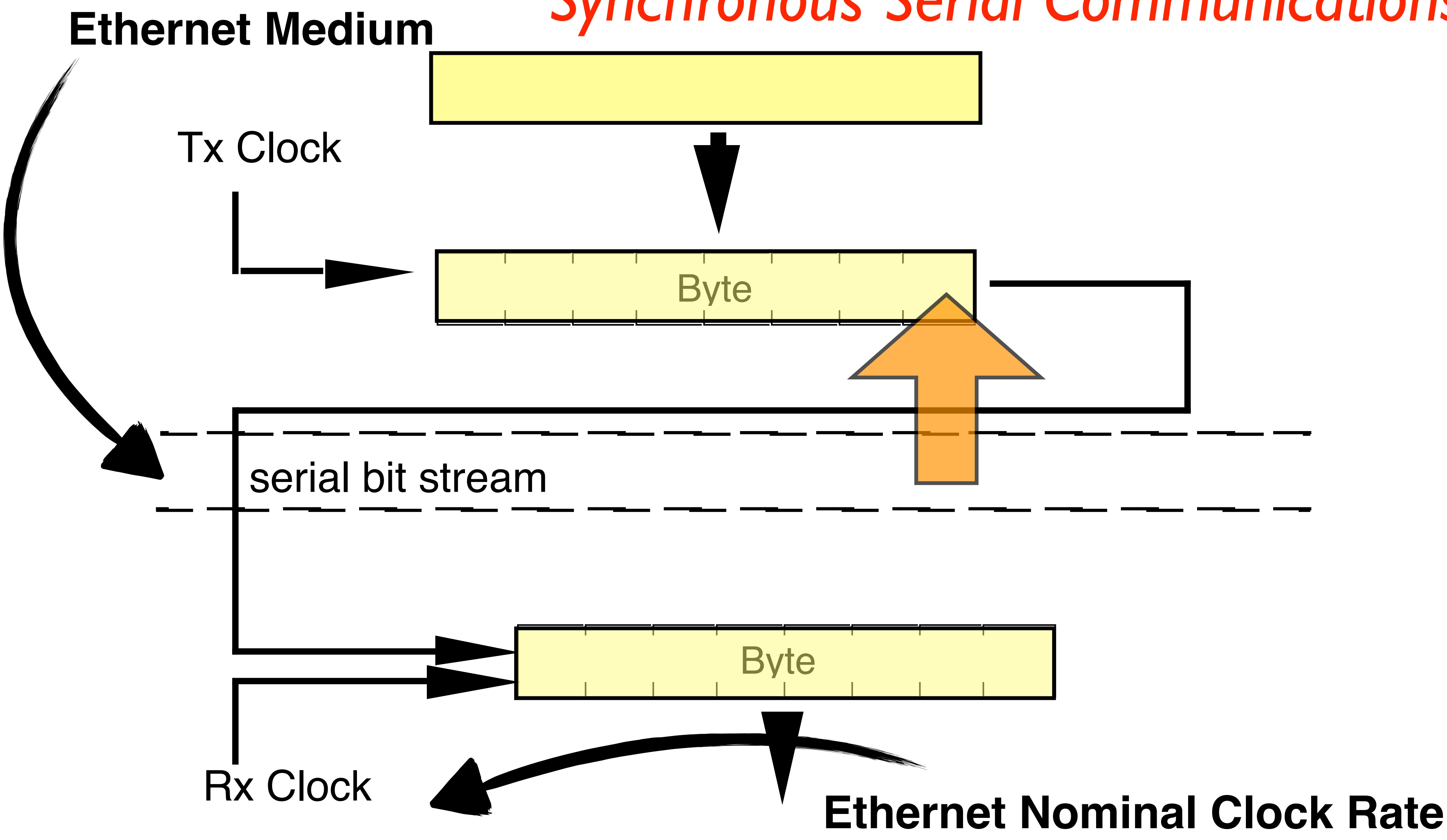
Sending Frames (Ethernet Transmit)



The Physical Layer

Module 3.1

Synchronous Serial Communications

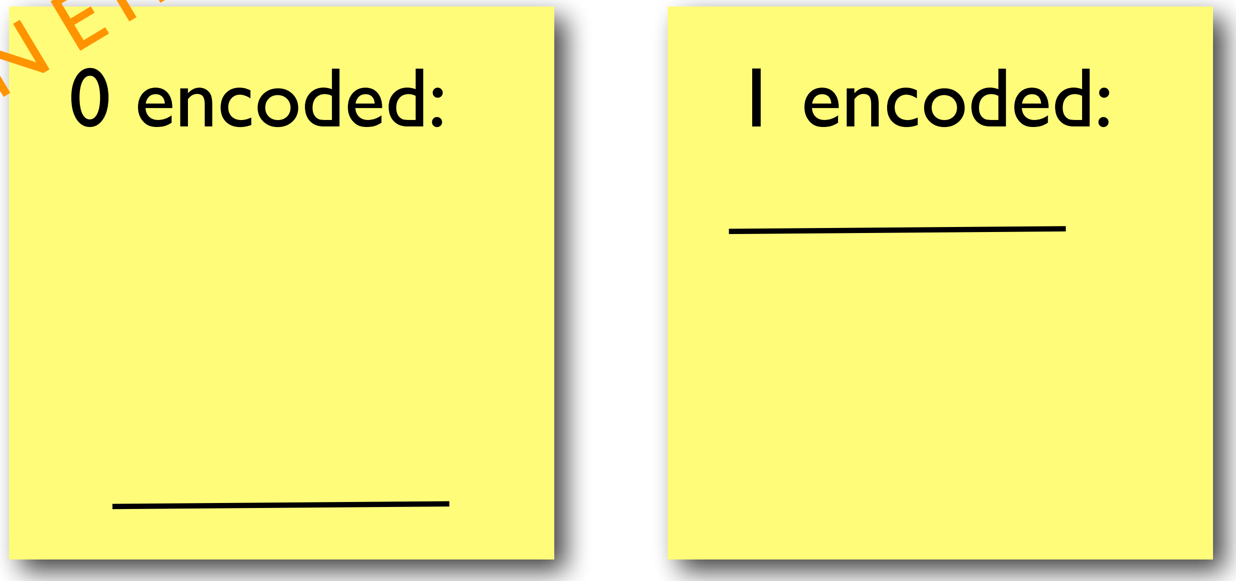


Uses two shift registers (both clocks must be the same rate)
- Note that bytes are sent l.s.b. first!

Recall the Ethernet broadcast/unicast address bit?

Non Return to Zero

NOT USED IN ETHERNET!

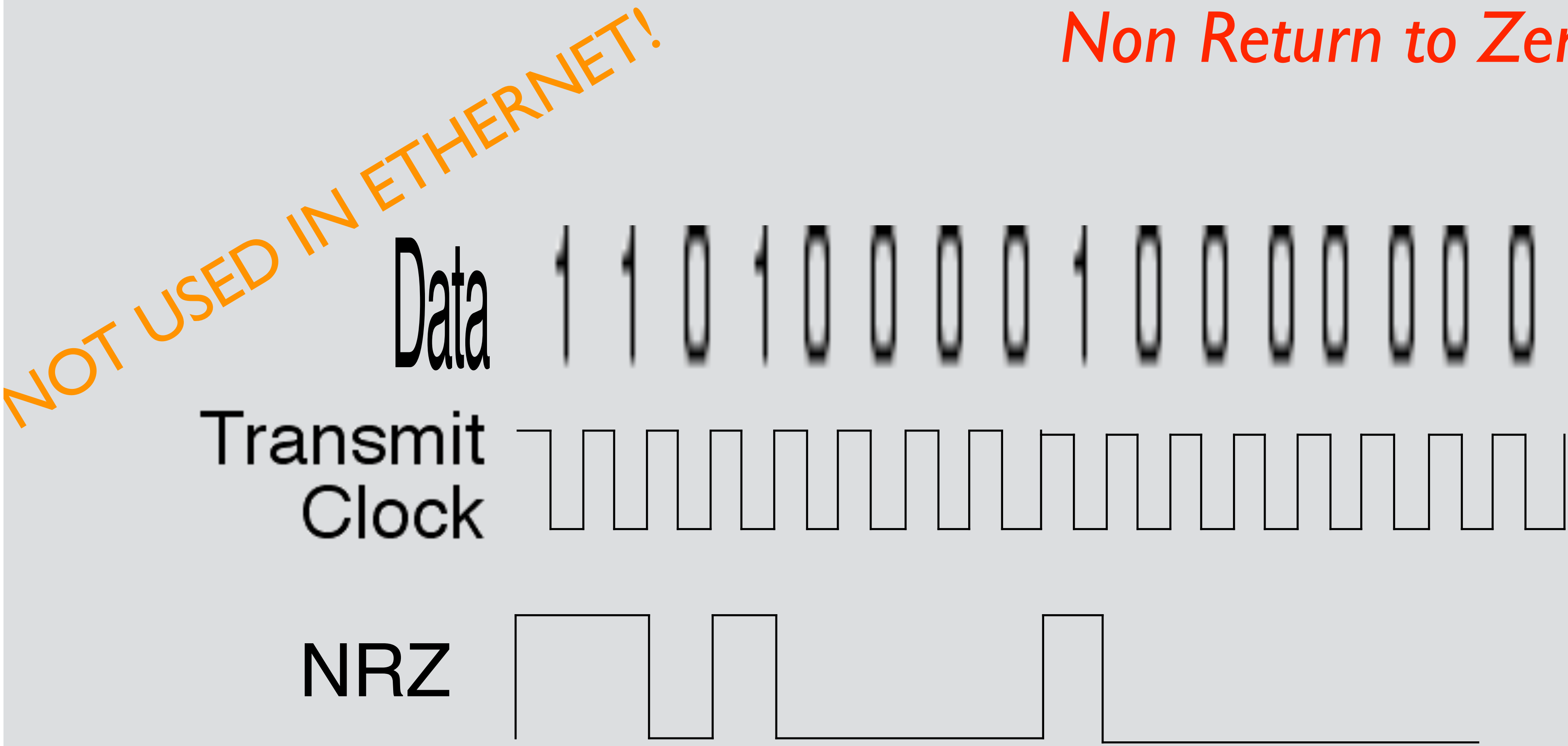


2 signal levels are used

The level of a baud indicates the value of each bit
- a low level indicates 0
- a high level indicates 1

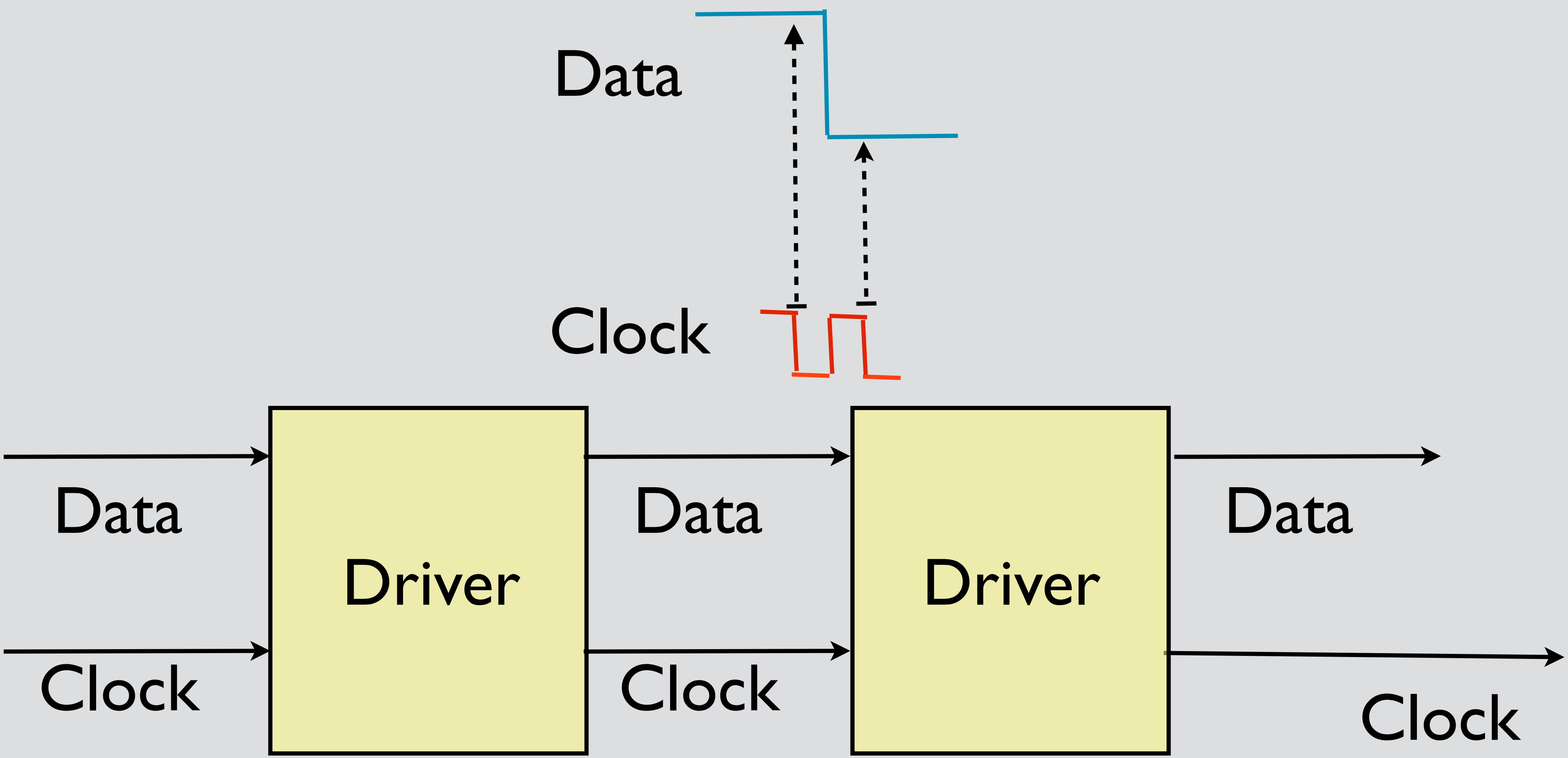
The bandwidth of NRZ is approx 1 Hz / bit

Non Return to Zero



The receiver needs some way of determining the clock transitions …
i.e. you can not just look at a NRZ encoded waveform to determine the sequence of 1 and 0 bits that it represents - you need to look at clock & data!

Traditional Synchronous Transmission

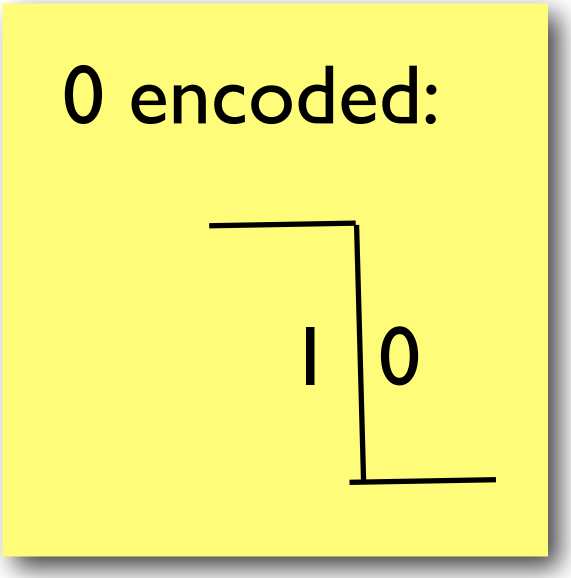


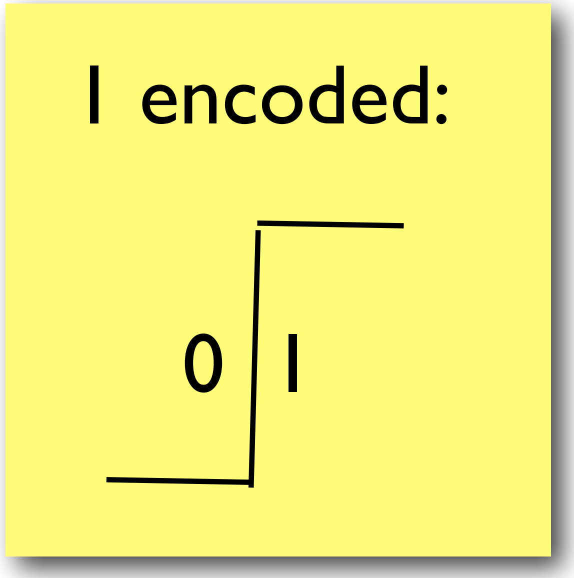
Clock signal transitions indicate centre of each bit

- Sender uses clock to time sending each bit
- Receiver uses the clock to detect the centre of each bit

Requires <u>two</u> sets of wires (clock & data + ground)

Manchester Encoding





2 signal levels used

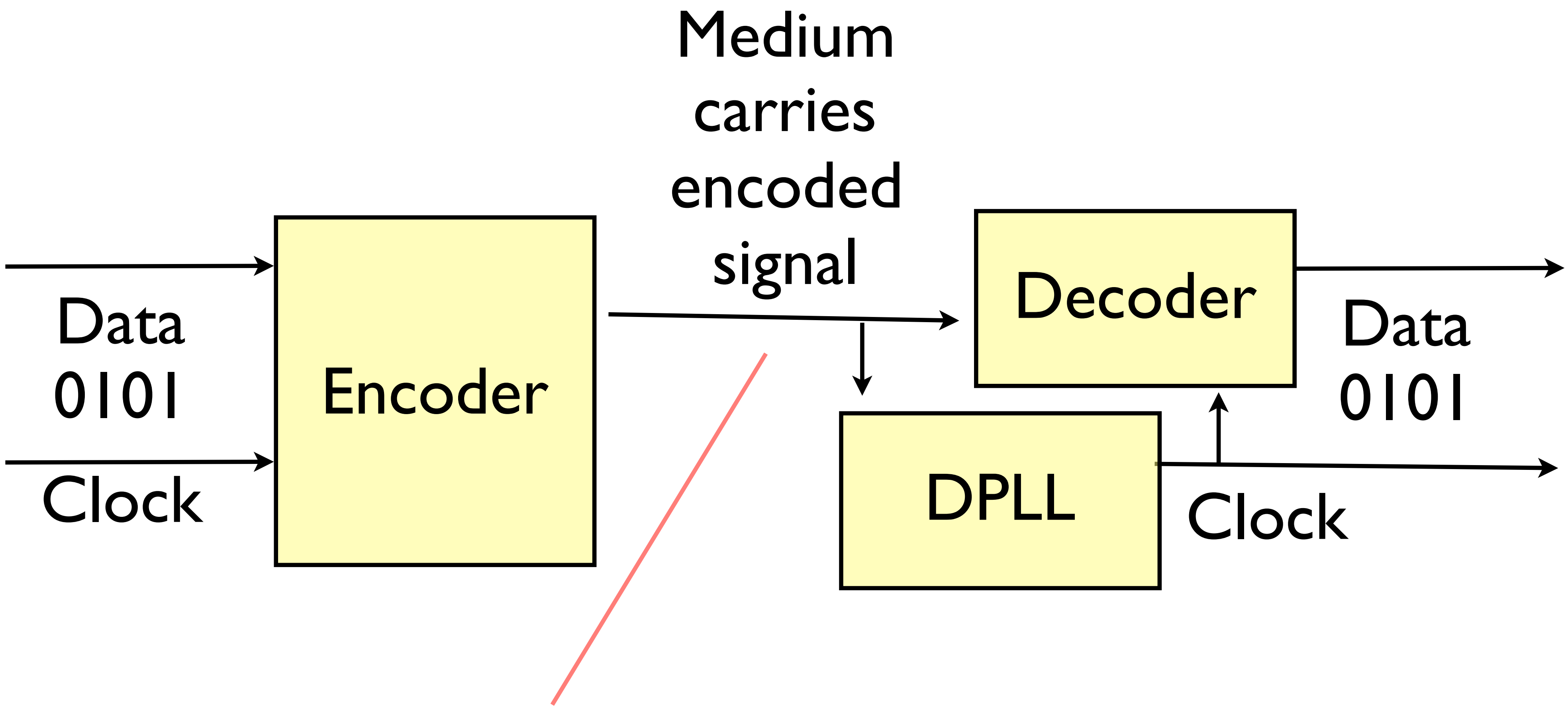
There is a transition in the centre of each bit

- a down-wards transition from a 1 baud to a 0 baud indicates a 0 bit
  - an up-wards transition indicates a 1 bit

The 2 bauds use double the cable bandwidth compared to NRZ!*

* 10B2/10B5 use high bandwidth RF cable, so this is not an issue.

Encoded Data



What no clock wire?

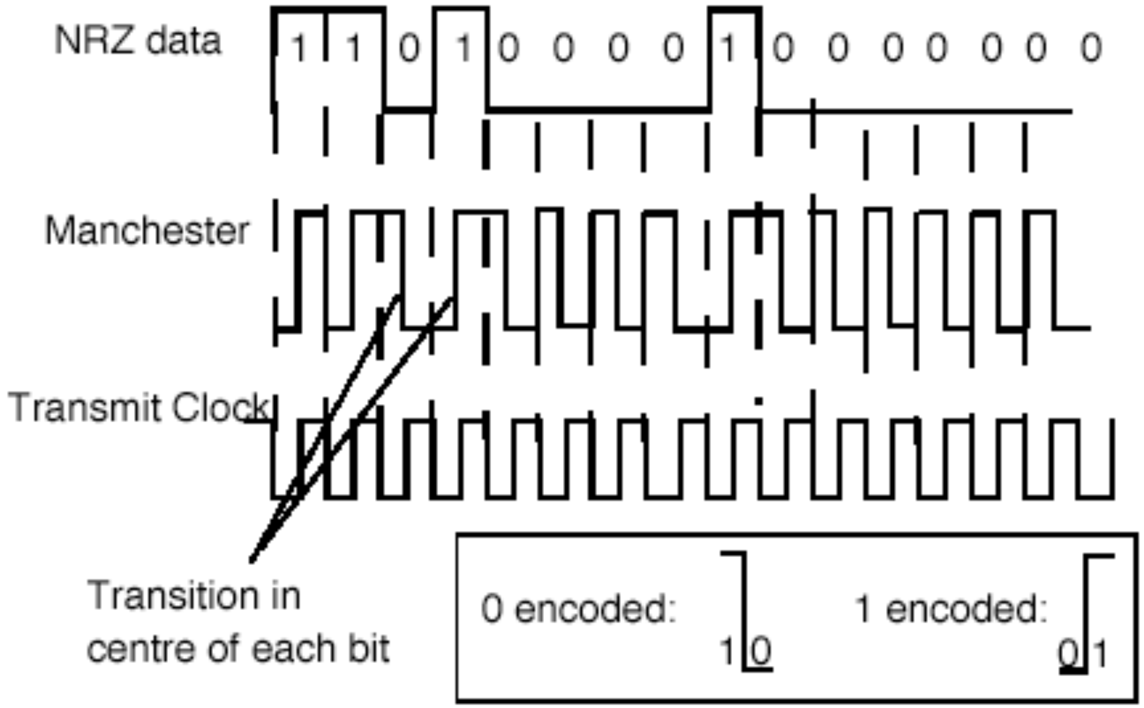
The sender encodes the clock and data as a waveform

The cable transmits this combined clock & data signal as pairs of bauds

This needs only one "wire"

At the receiver, a Digital Phase Locked Loop (DPLL) regenerates clock

Manchester Encoding

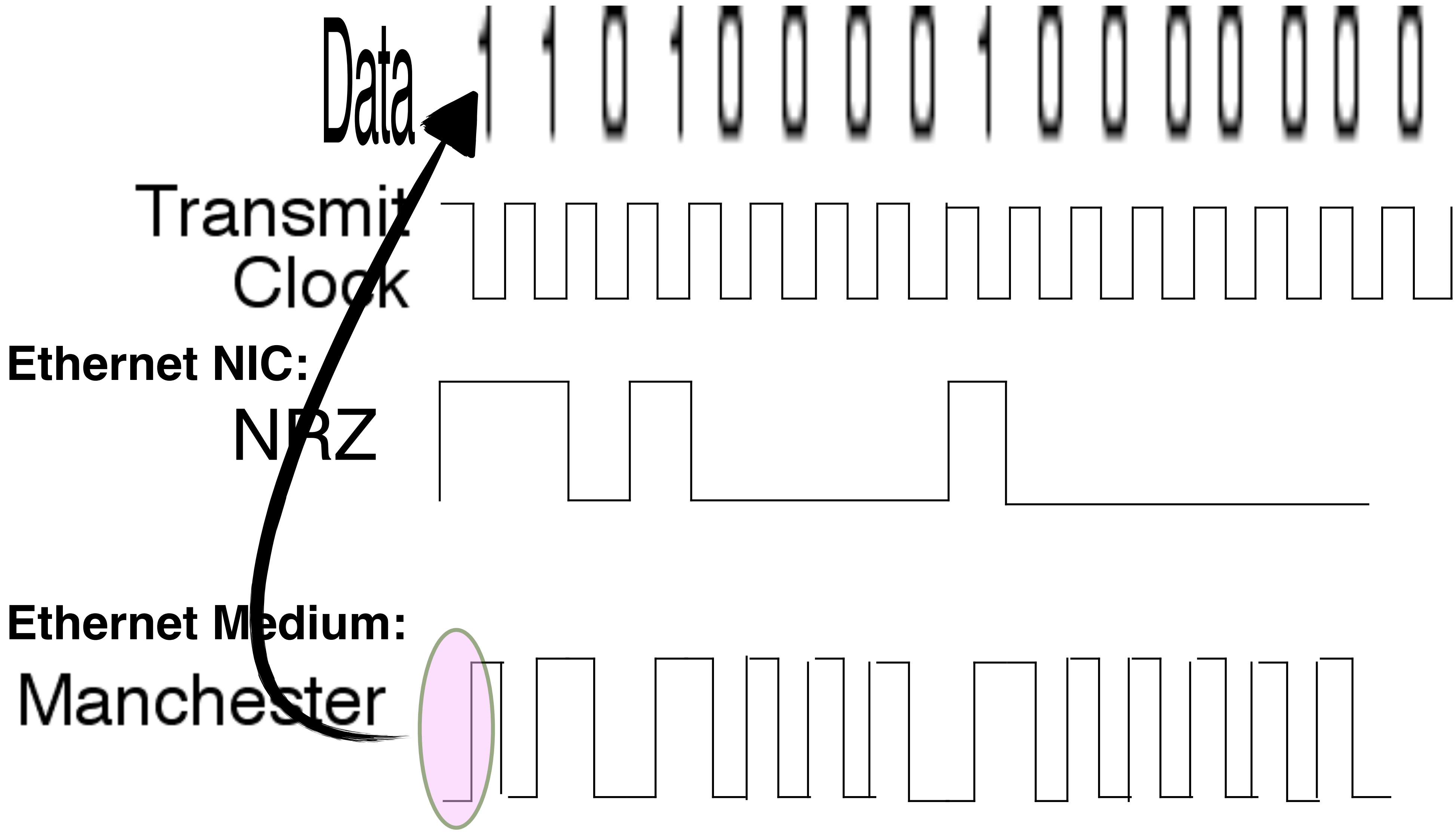


Looking at the waveform it is clear there is:

No DC component (even for long runs of 0's or 1's)

A timing component at the fundamental clock frequency (10 MHz)

Manchester Encoded Signal

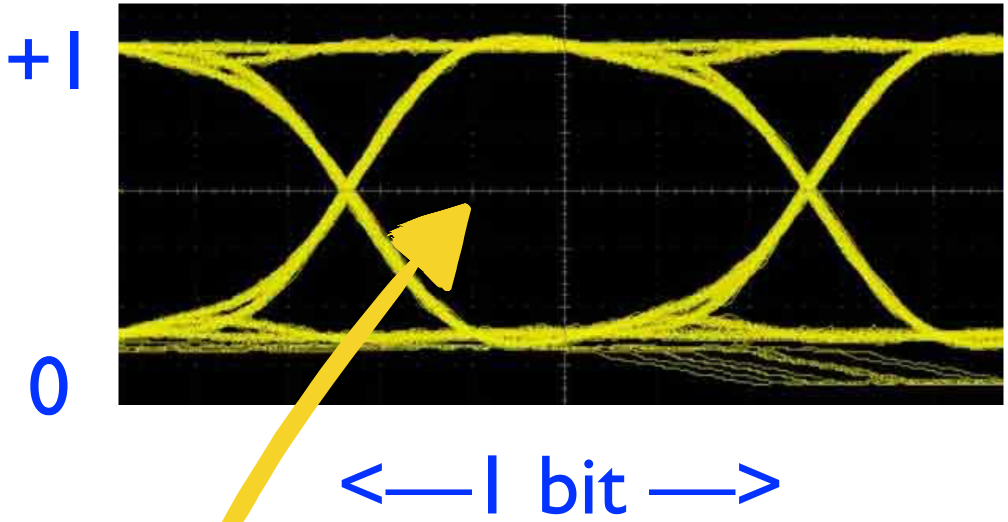


Eye Diagram for Manchester Encoded Signal

Oscilloscope plot using an eye diagram

The eye diagram plots voltage v. time

With a timebase trigger for ***multiple scans*** through the waveform



Two distinct levels are clear

Noise is evident causing blurring in the vertical axis

The slew rate is limited, i.e. the rise time for transitions

Transitions in level only occur at the edge of bits

Transitions never occur in the centre of the display!

Summary

- **Manchester Encoding**

  Encodes ***each*** data bit as a pair of bauds

  No net DC signal

  Uses double the baud rate

  Embedded clock

- **A DPLL is used at the receiver to decode the clock**

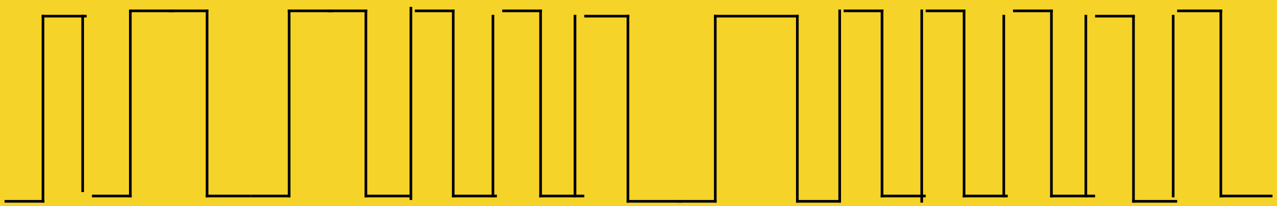
  This aligns the local clock with the received bauds

- **Data is decoded**

  2 bauds are read to decode ***each*** data bit

Ethernet Frames:

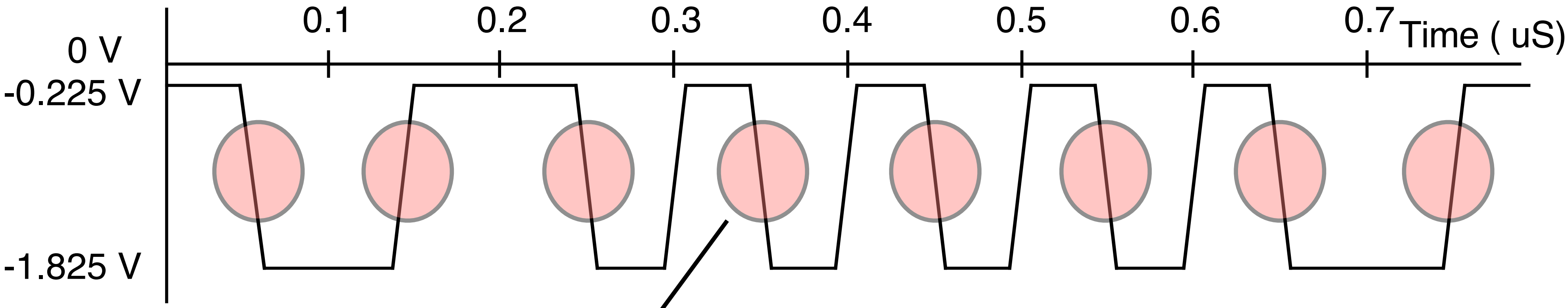
Receiving data (Ethernet Receive)



The Physical Layer

Module 3.2

Ethernet Waveform

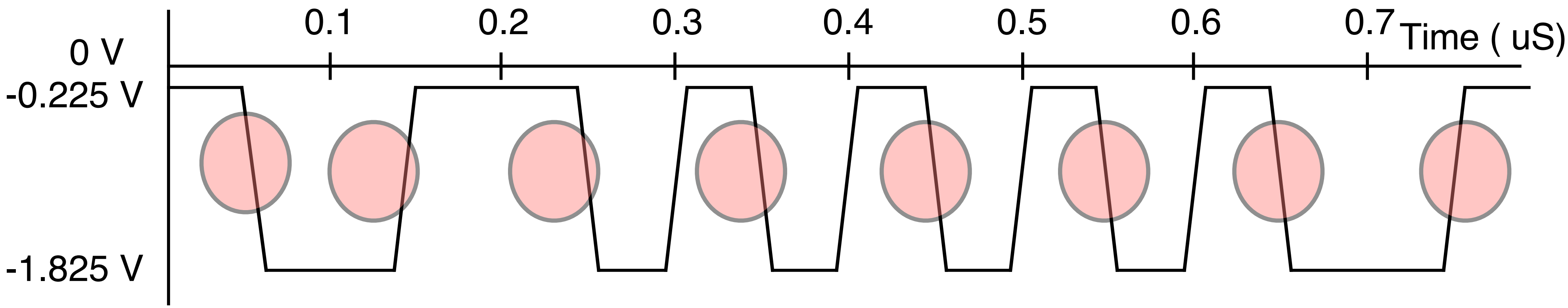


Transitions
at centre of bits

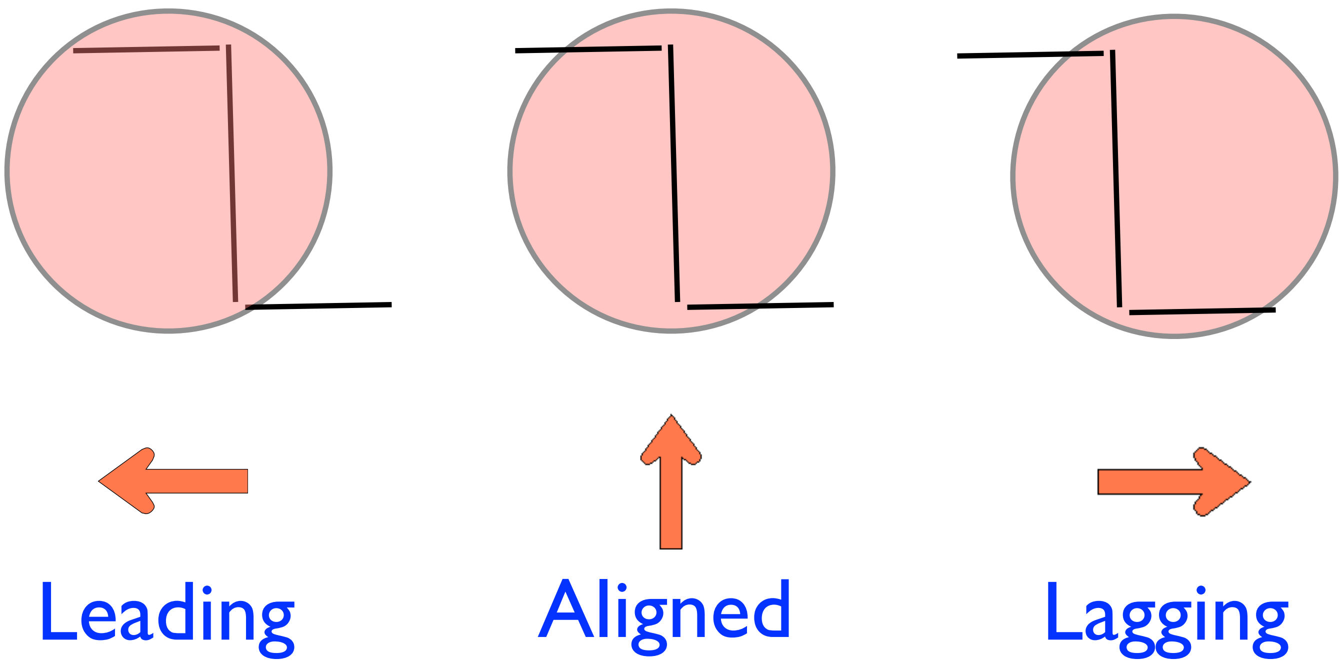
Can ***you*** decode this?

The signal isn't referenced to zero volts
Rise time ~ 25nS
The waveform as seen on an oscilloscope may be inverted!

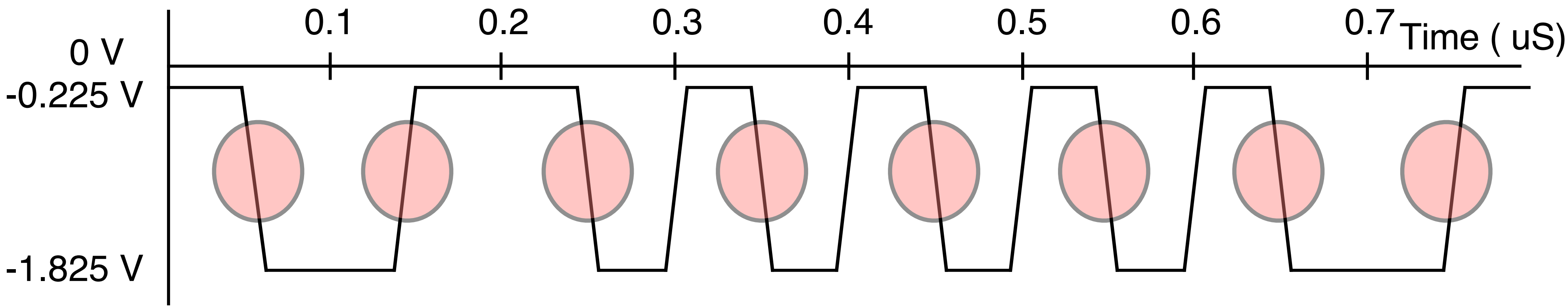
Sampling the Received Waveform

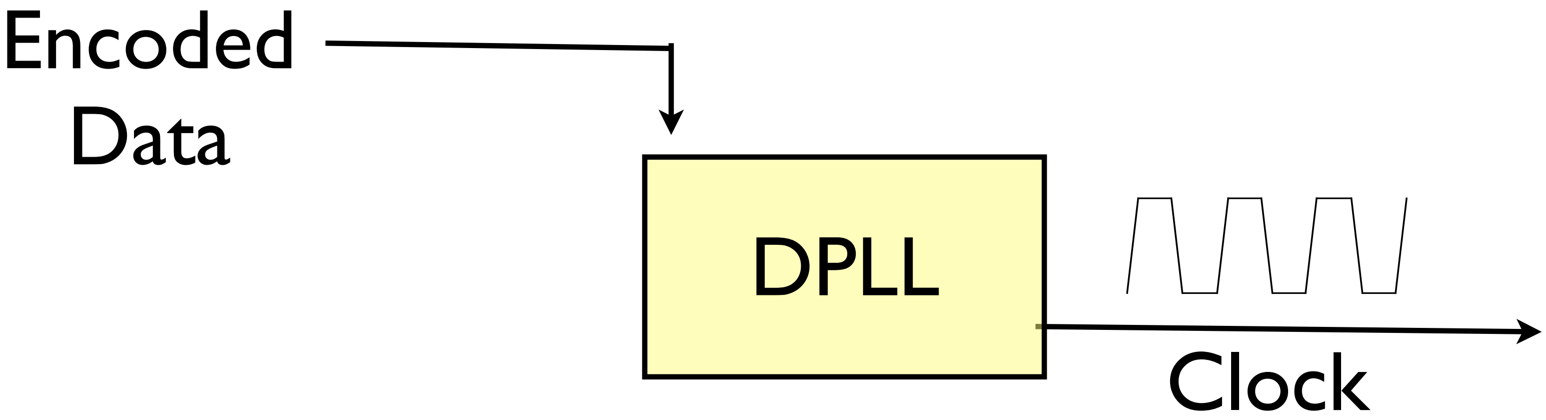


If we sample pairs of bauds, the waveform at receiver might result in one of three cases:



Ethernet Clock Recovery





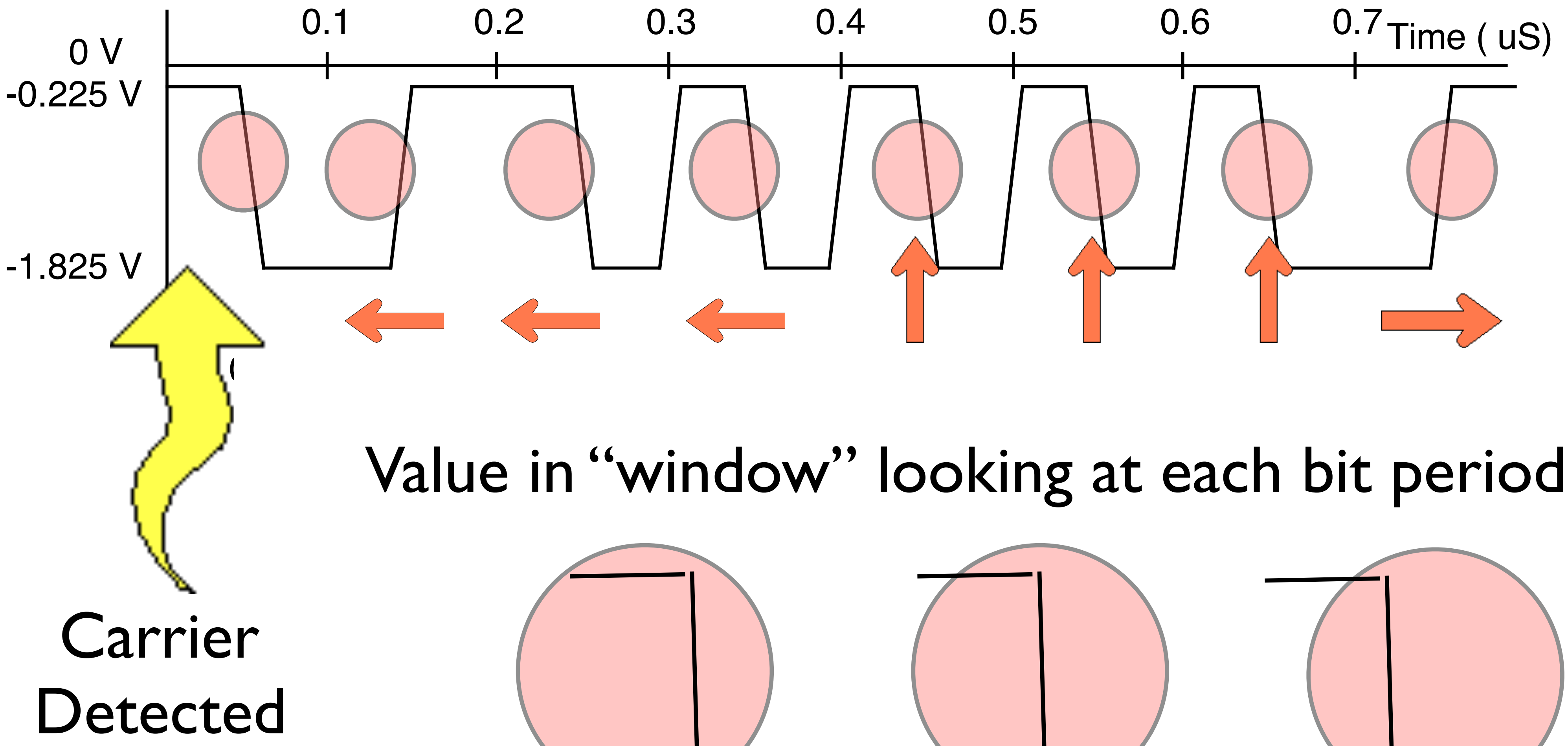
DPLL contains a clock (oscillator)

Uses phase transitions to lock the local receive oscillator frequency

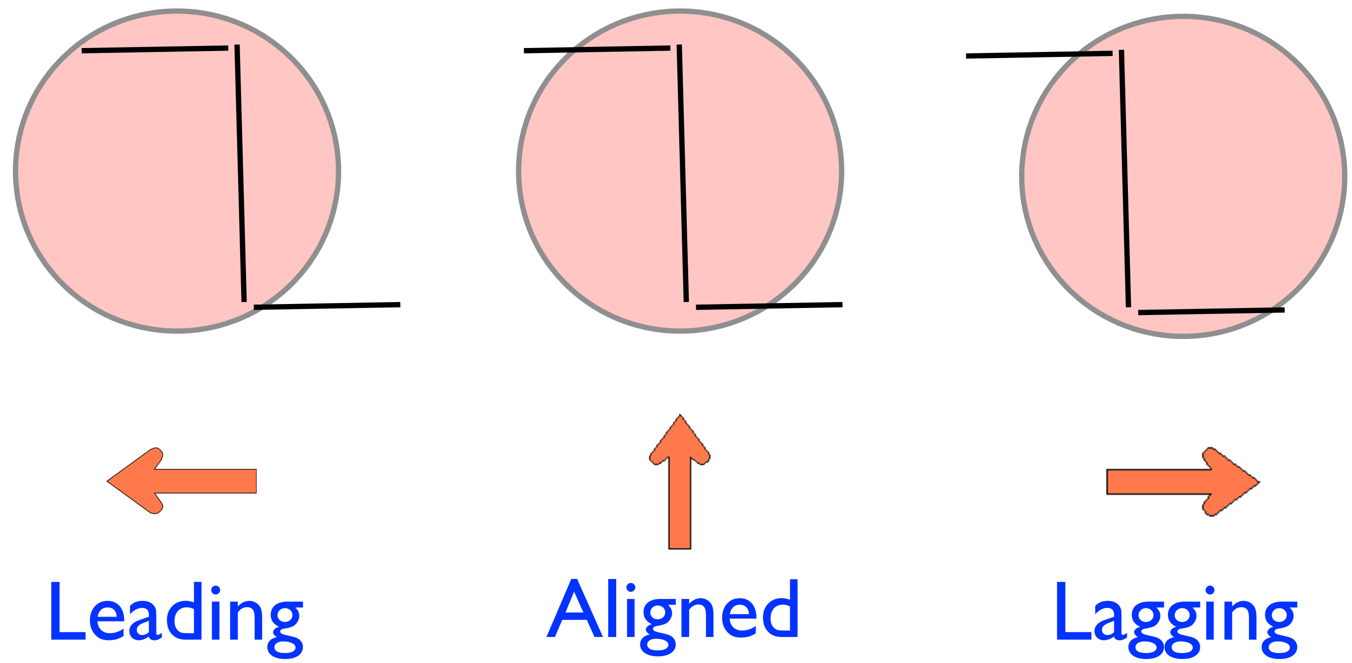
- If a transition is *lagging*, *decrease the clock* period (increase frequency)
- If *leading*, *increase the clock period* (decrease the frequency)

After many transitions, the recovered clock ***matches the encoded data***

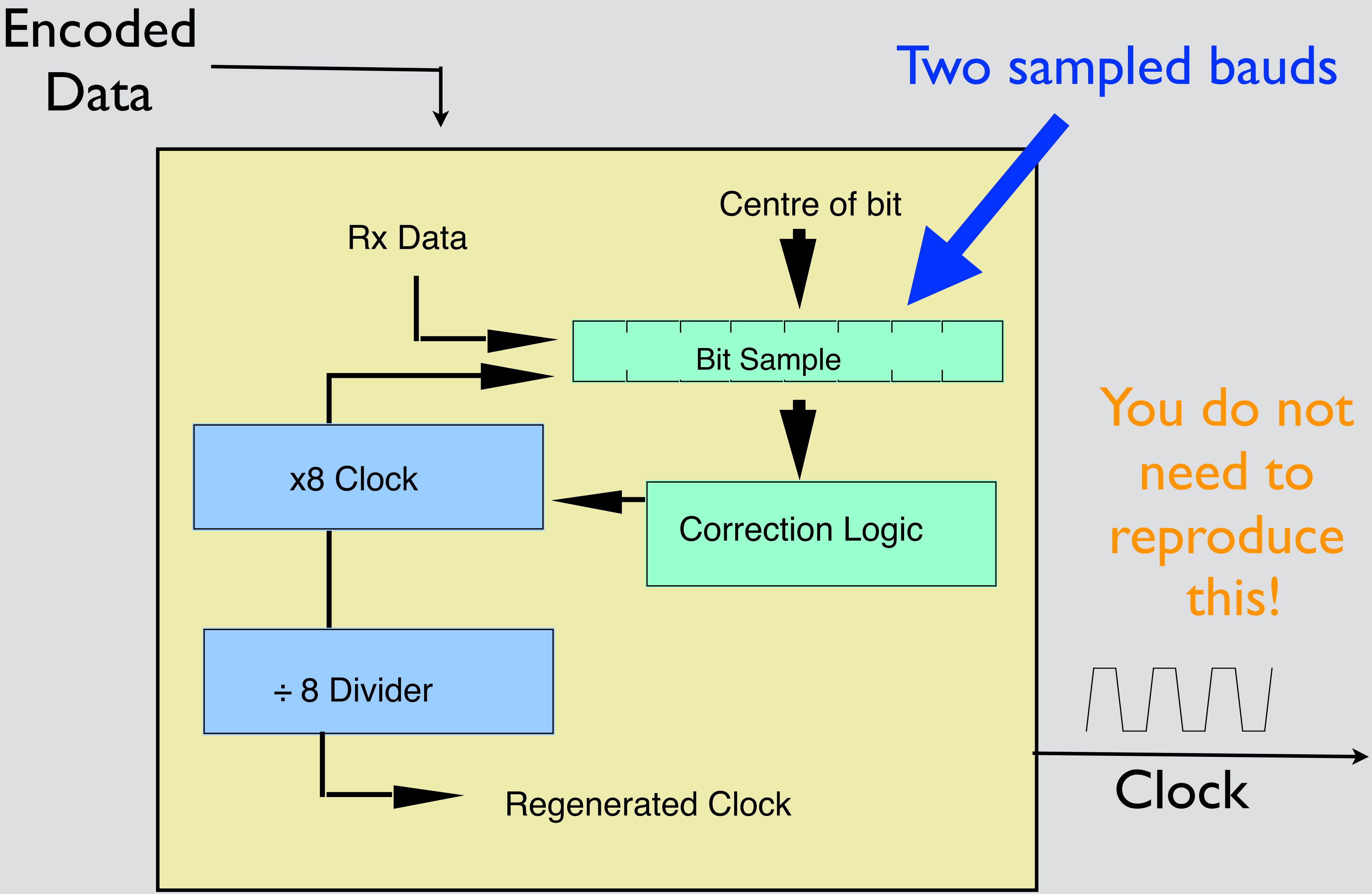
Ethernet Clock Recovery by DPLL



Value in "window" looking at each bit period:



Digital Phase-Locked Loop (DPLL)



You do not need to reproduce this!

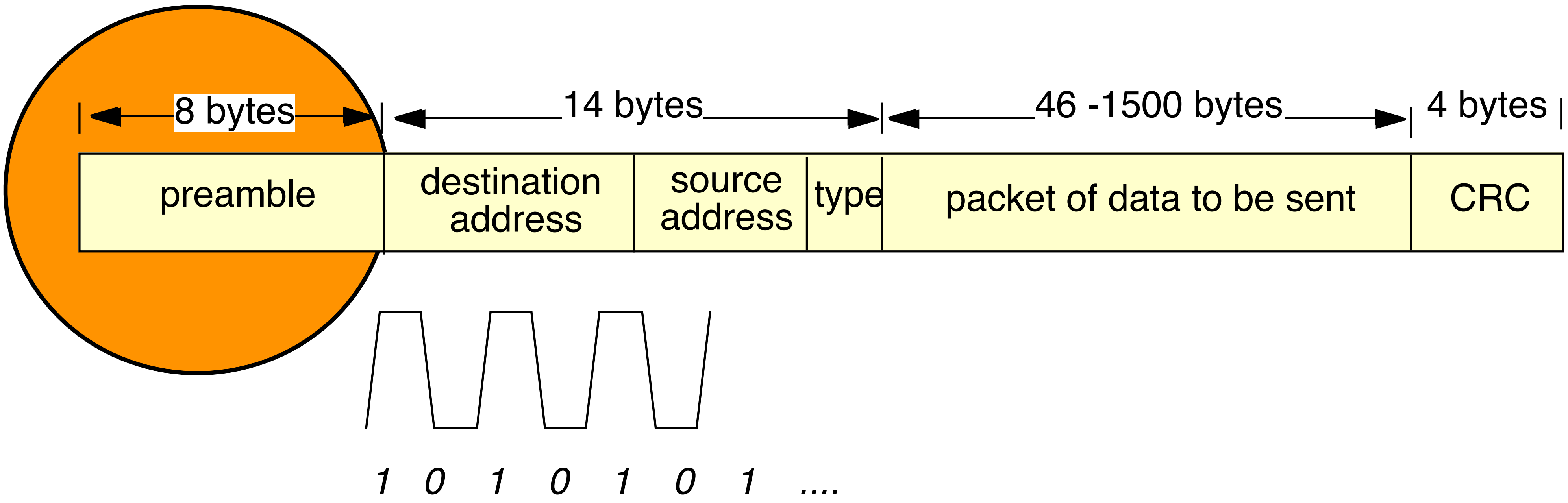
Preamble Sequence

Each frame starts with a fixed-format preamble

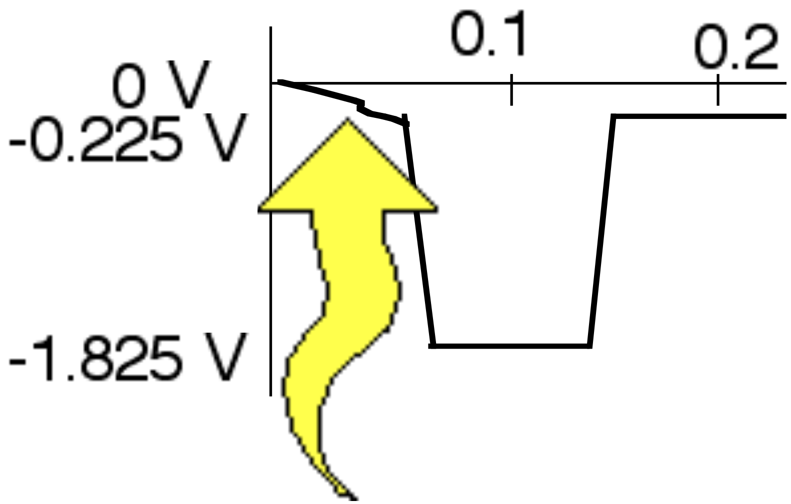
This has two functions:

(1) The format is chosen to help assist the DPLL achieve lock
This means the preamble uses an alternating '0' and '1' bit pattern

(2) The preamble is used to detect the start of frame delimiter (SFD)
The final 2 bits of the last byte (SFD) are set to '11'
This reveals the encoding rule for a '1'



Ethernet Inter-Frame Gap / Spacing



Carrier
Detected

A silent time between frames (no carrier on medium)

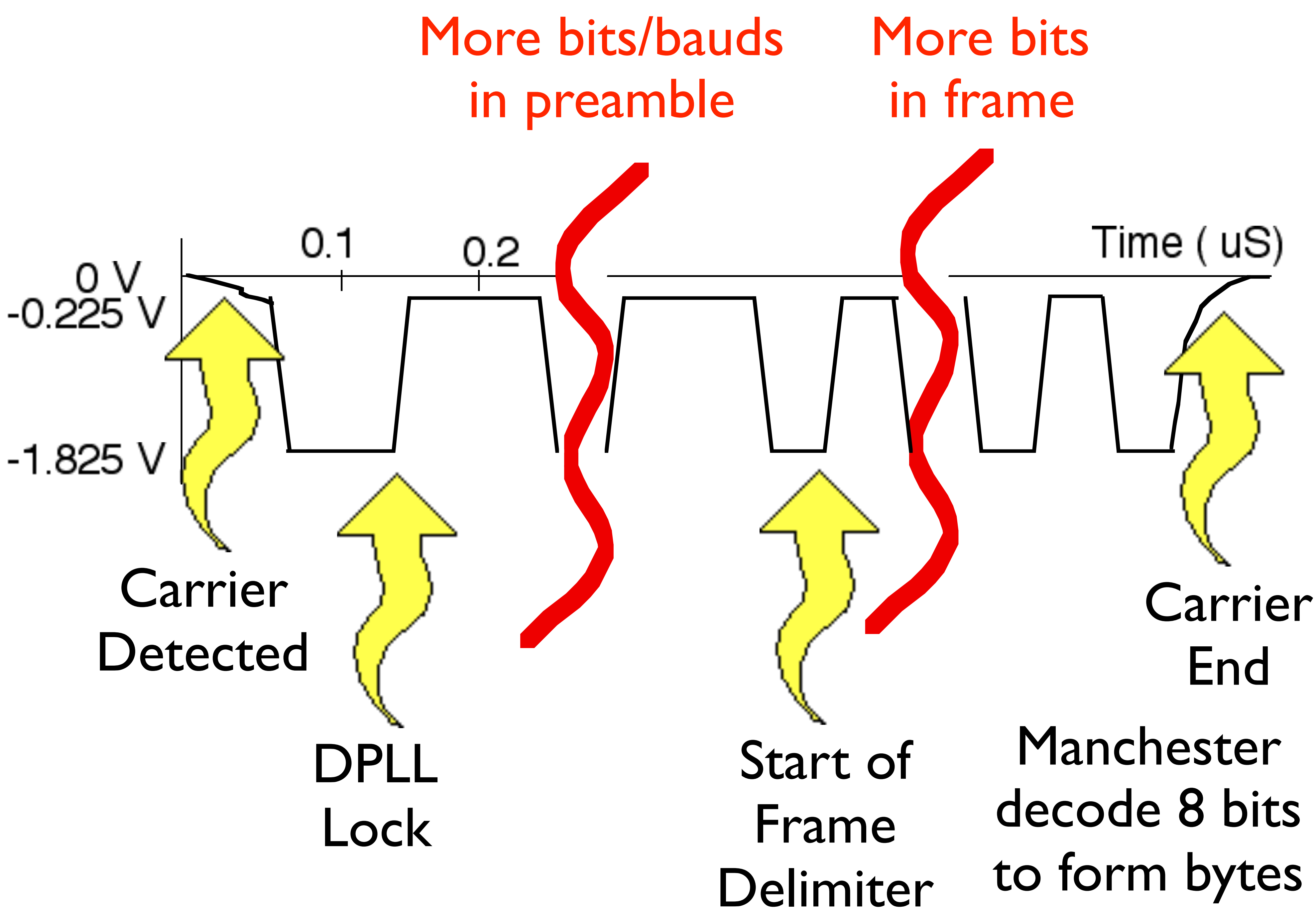
Allows transceiver electronics to recover after end of previous frame

20 byte periods (measured from end to next SFD)

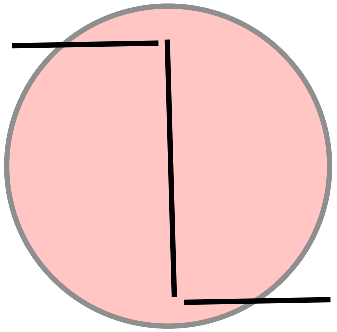
10 Mbps: > 9.6 microsecs between frames (at receiver)

(some descriptions say 10.4 microsecs at sender)

Ethernet Frame



Resolving ambiguity in the Received Polarity



Is this a '0' or a '1'
Is the waveform inverted?

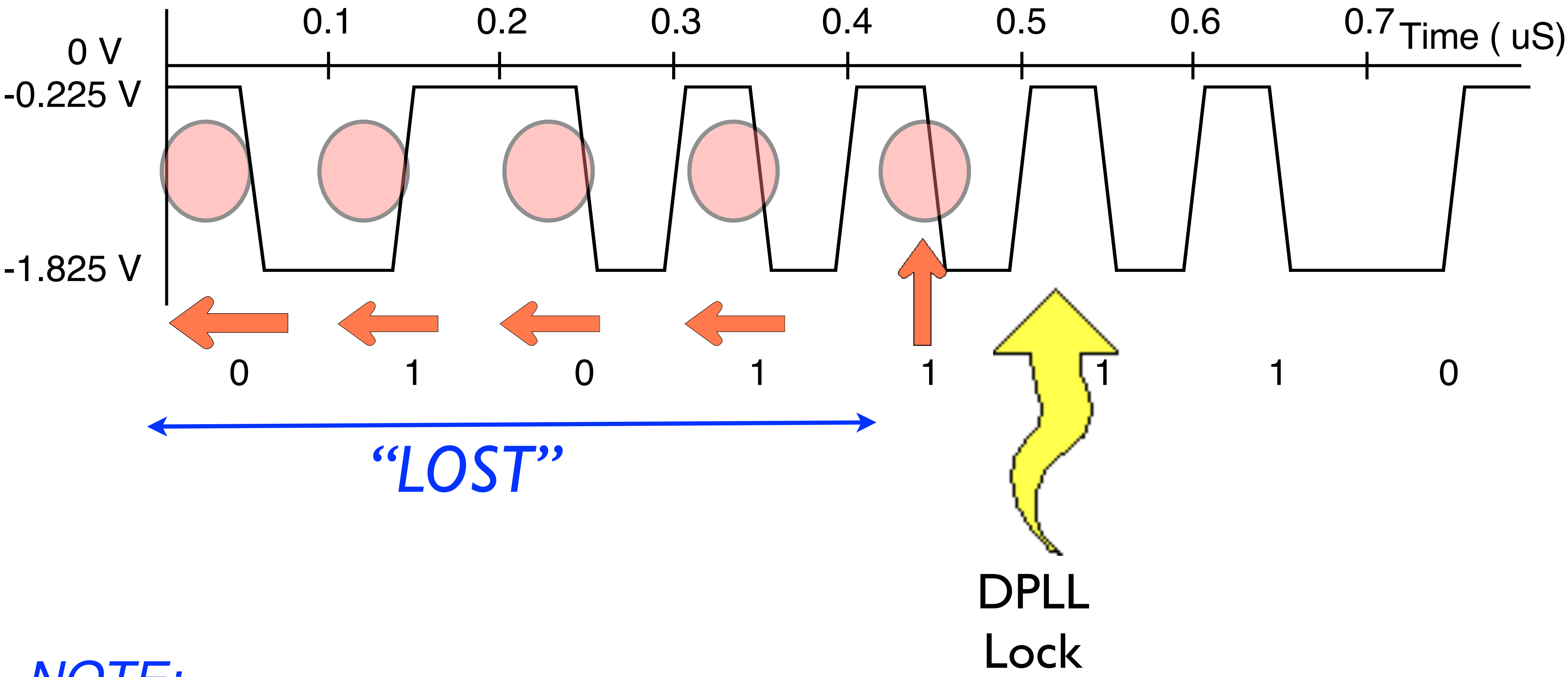
A waveform can be inverted ...

Ethernet has a trick that allows a received to discovery the polarity of the received signal bauds...

Recall that the SFD ends with the sequence '11'

When the decoder sees the end of the preamble it can unambiguously discover the pair of Manchester-encoded bauds used for a '1' bit.

Loss of the Start of the Preamble!!



NOTE:

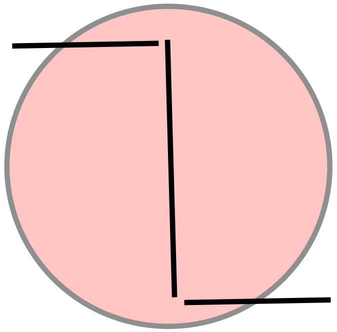
(1) Each sender will have a slightly different clock signal

*A receiver therefore has to **retrain** the DPLL to each new sender*

(2) Bauds received before the DPLL has lock may not be decoded

Not all bauds of the preamble are “therefore received” by the decoder

Summary: Four Steps to Reception



4 steps required to decode each frame

1) The start of a frame needs to be detected using the CS circuit

2) A clock signal is recovered at the receiver (using a DPLL)

3) The polarity and start of the data is determined from the SFD

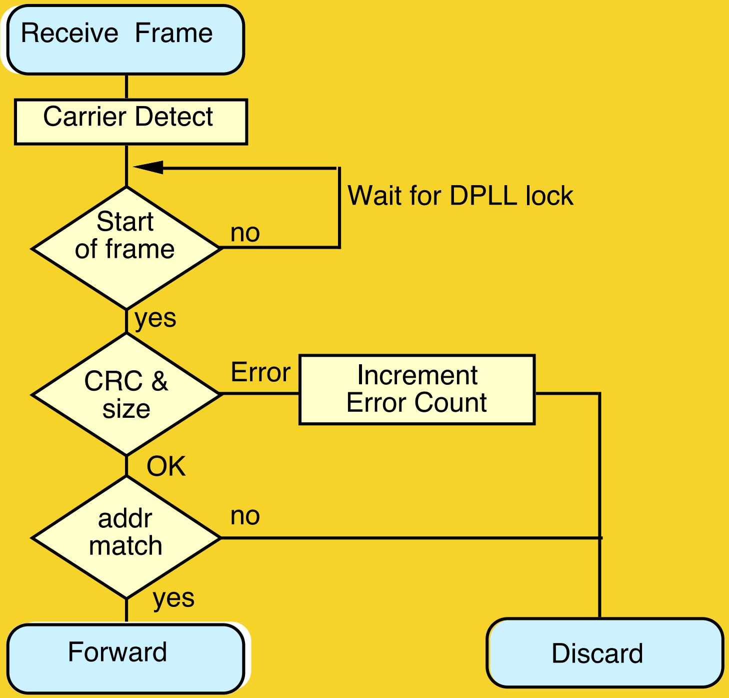
4) The end each frame is detected using the CS circuit

Summary

- **There is an Inter Frame Gap (IFG) between each frame**
- **All Ethernet frames have a preamble**
    - 62 bits have the pattern 10
    - The first baud triggers the carrier detect circuit to start listening
    - Remainder of the preamble helps gain DPLL lock (takes time)
    - Not all preamble bits are “received” by the decoder
- **End of preamble marked by the SFD**
    - Polarity detected by the 2 SFD bits, with value 11
- **The final bit of the frame is detected by absence of a carrier**
    - A CRC-32 is used to verify this process

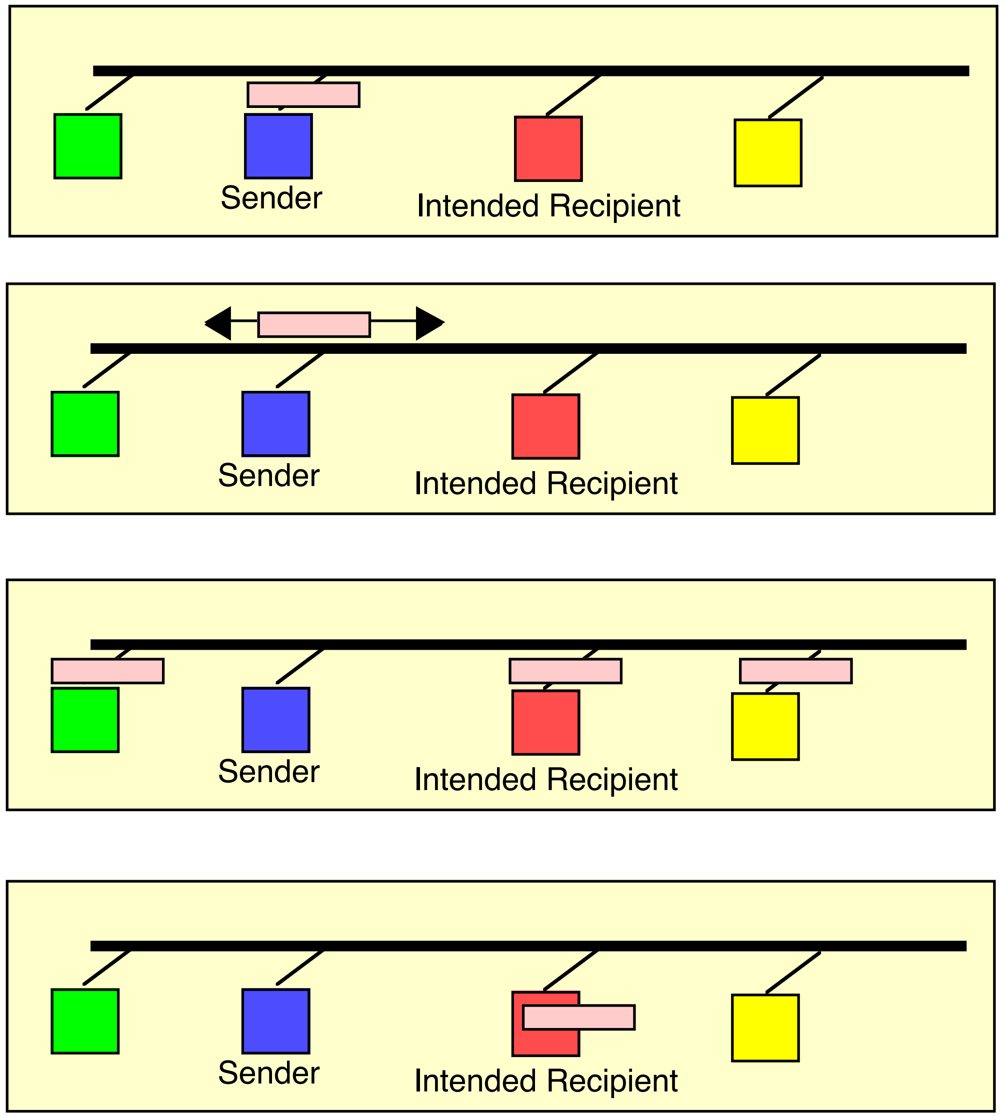
Ethernet Frames:

Frame Reception



Module 3.3

LAN (MAC) address



Cyclic Redundancy Check (CRC)

8 bytes	14 bytes			46 -1500 bytes	4 bytes
preamble	destination address	source address	type	packet of data to be sent	CRC

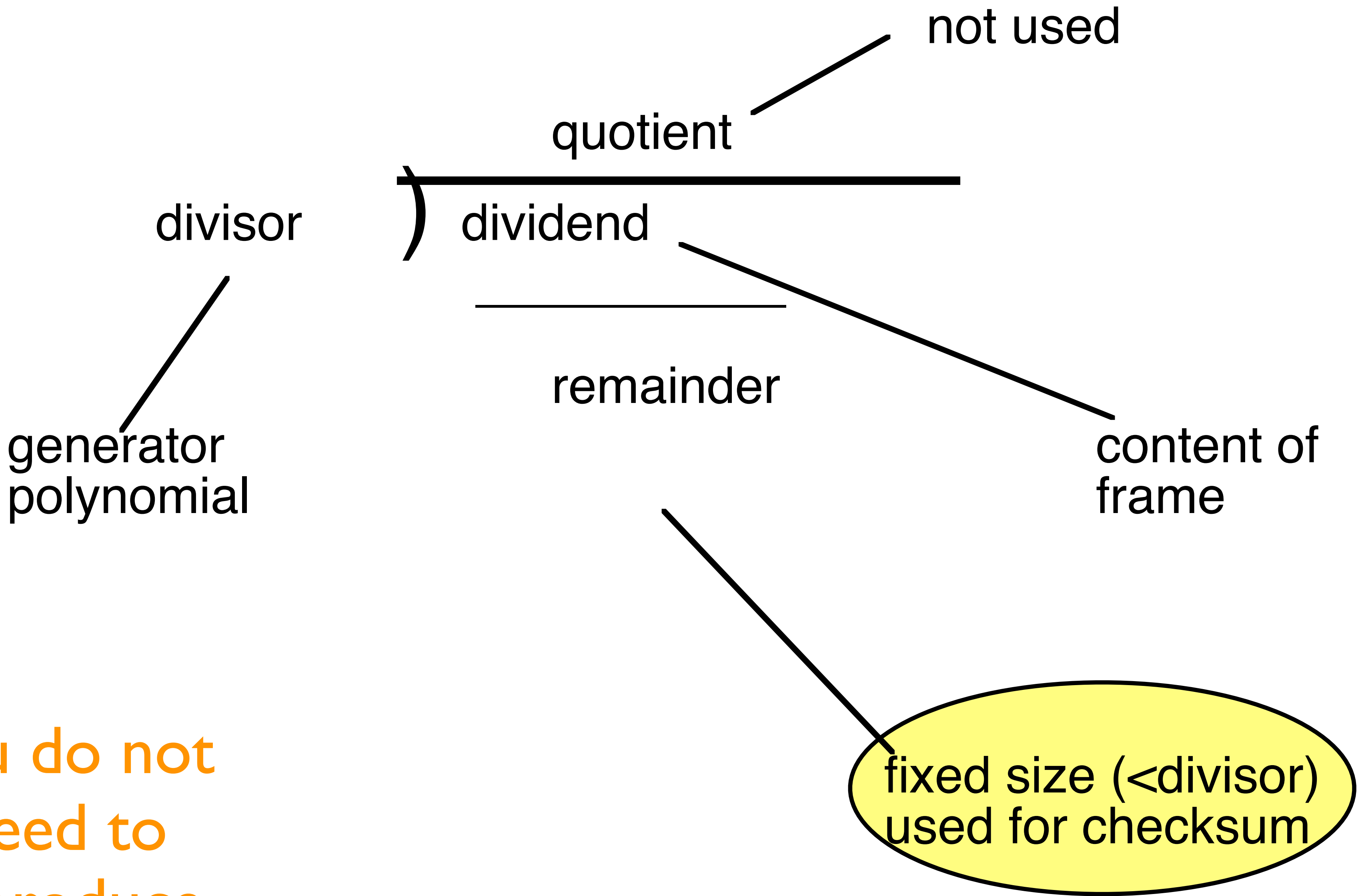
CRC-32 is a form of digital signature (32-bit hash of frame)

- Calculated at the sender & sent at end of each frame
- Re-calculated at the receiver
- Sent value is compared with received value at receiver
- This verifies the integrity of the data in the frame

The CRC-32 has a high probability of detecting:

- Any frames corrupted in transmission
- Frames where the DPLL failed to track the clock

Division



You do not need to reproduce this!

Why Modulo 2 Division?

Because the hardware solution is simple!!!!!

Truth Table for Modulo-2 Division (XOR)

$$0 \oplus 0 = 0$$
$$0 \oplus 1 = 1$$
$$1 \oplus 0 = 1$$
$$1 \oplus 1 = 0$$

You do not need to reproduce this!

- CRC calculations ignore the carry

Modulo 2 Division

Modulo 2 division
replaces addition
in BCC calculation

First digit
must be '1'

0's are appened
to the dividend
(flush bits)

```
                1
        _________________________________
11001  ) 1 1 1 0 0 1 0 1  0 0 0 0
       ⊕ 1 1 0 0 1
         ---------
         0|1 1 0 1
```

Divisor
(Generator Polynomial)

This digit must always be 0

You do not need to reproduce this!

Example simplified to generate a short (4 bit) CRC

Modulo Division

Revise your notes from Level 3 course!!!

```
                  1 0
11001 ) 1 1 1 0 0 1 0 1  0 0 0 0
      ⊕ 1 1 0 0 1|  ..
        0|0 1 0 1 1
         ⊕ 0 0 0 0 0
           0|1 0 1 1
```

1 Bring next digit of dividend down
2 Copy msb of value to quotient
3 Insert 0 (if quotient 0) or divisor (if quotient 1)
4 Calculate XOR sum
5 Discard msb of value (always 0)

You do not need to reproduce this!

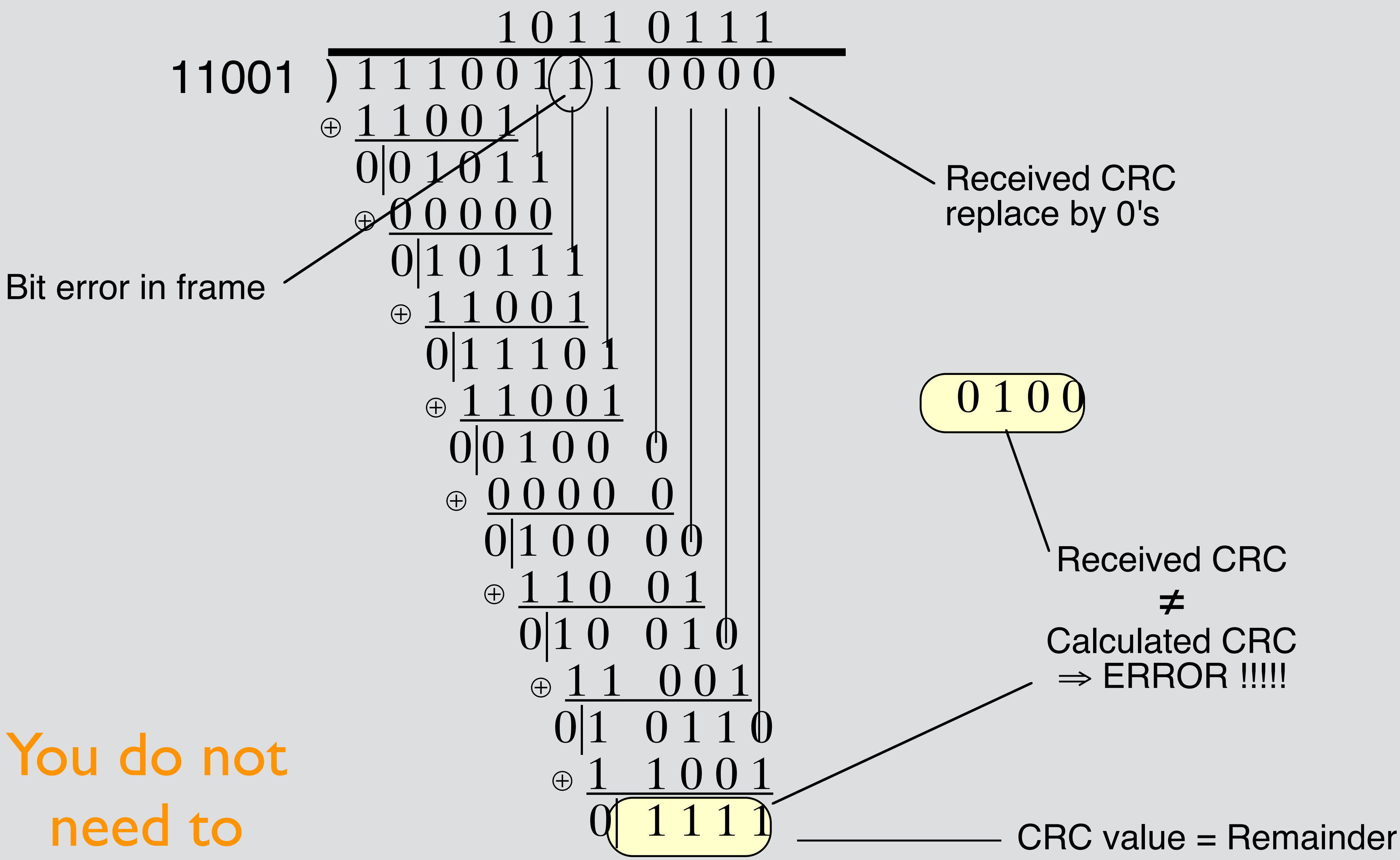
CRC Value

```
                          1 0 1 1 0 1 0 0
11001 ) 1 1 1 0 0 1 0 1 0 0 0 0
      ⊕ 1 1 0 0 1
        0|0 1 0 1 1
        ⊕ 0 0 0 0 0
          0|1 0 1 1 0
          ⊕ 1 1 0 0 1
            0|1 1 1 1 1
            ⊕ 1 1 0 0 1
              0|0 1 1 0 0
              ⊕ 0 0 0 0 0
                0|1 1 0 0 0
                ⊕ 1 1 0 0 1
                  0|0 0 0 1 0
                  ⊕ 0 0 0 0 0
                    0|0 0 1 0 0
                    ⊕ 0 0 0 0 0
                      0|0 1 0 0
```

CRC value = Remainder

You do not need to reproduce this!

CRC Value after an Error



You do not need to reproduce this!

Ethernet

*"**the ether**" the air, when it is thought of as the place in which radio or electronic communication takes place*, OED.

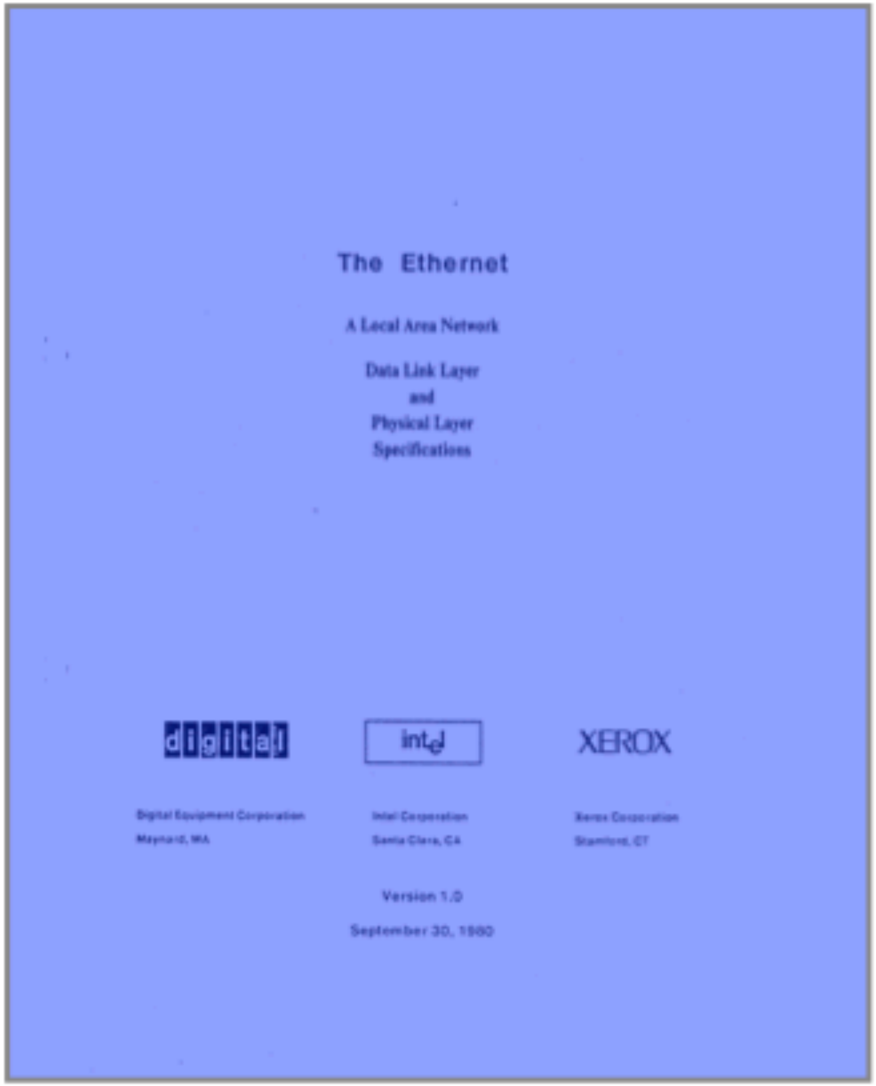
What is Ethernet?

Ethernet v2 - Blue Book

First published 1980, updated 1982
Digital, Intel, Xerox (DI
10 Mbps Speed
50 Ohm Coaxial cable

An Open Standard

The invention of Ethernet as an open, non-proprietary, industry-standard local network was perhaps even more significant than the invention of Ethernet technology itself.



LAN Topologies

Single Link

Bus

Shared Cable
(e.g. Coaxial Cable: 10B2, 10B5)

Star

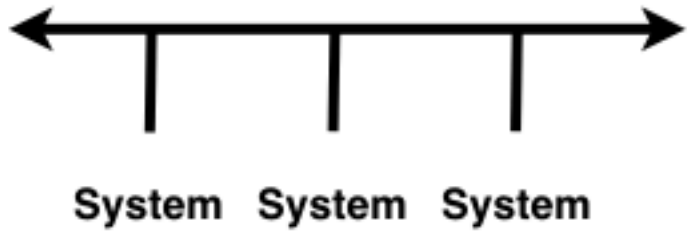
Connection to a Hub
(e.g. Twisted Pair Cable: 10BT)

Tree

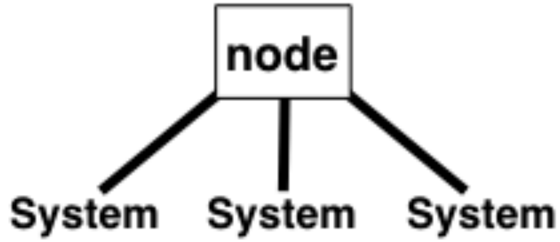
Connected Hubs/Switches
… network of routers ….
(Variety of media: 10B2,10B5, 10BT, 10BF)

Point-to-Point link

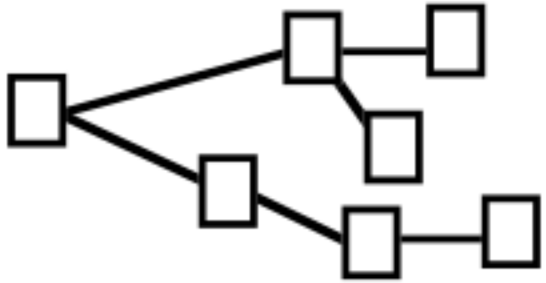
Bus



Star

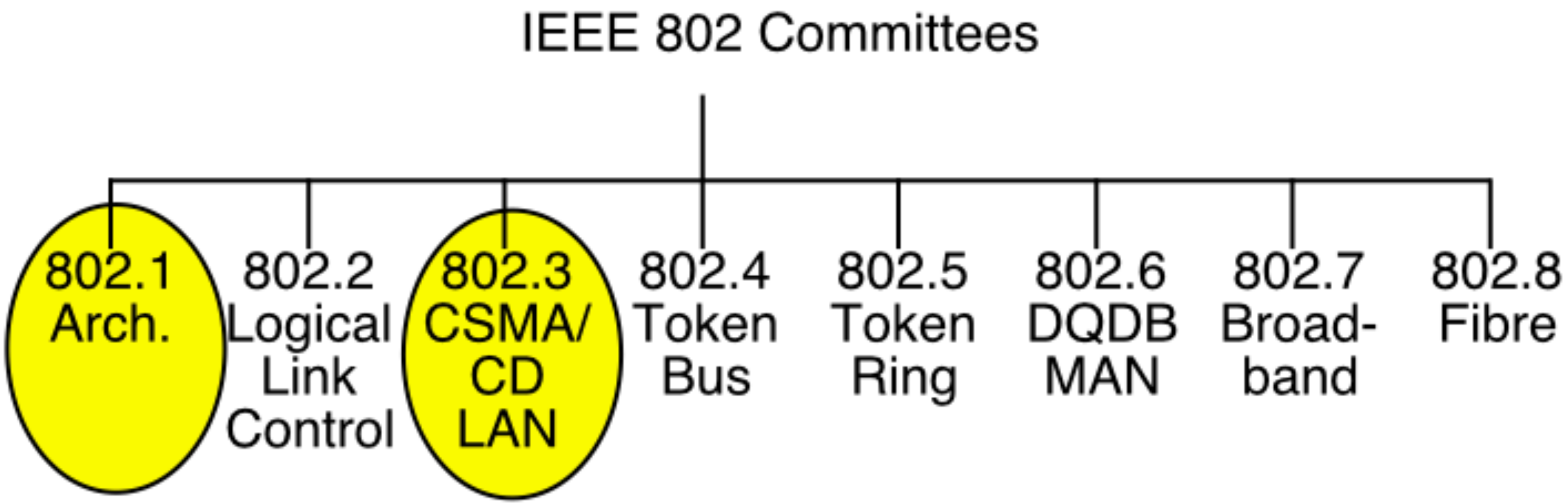


Tree



What is Ethernet?

Ethernet Standardised by IEEE in 1983:



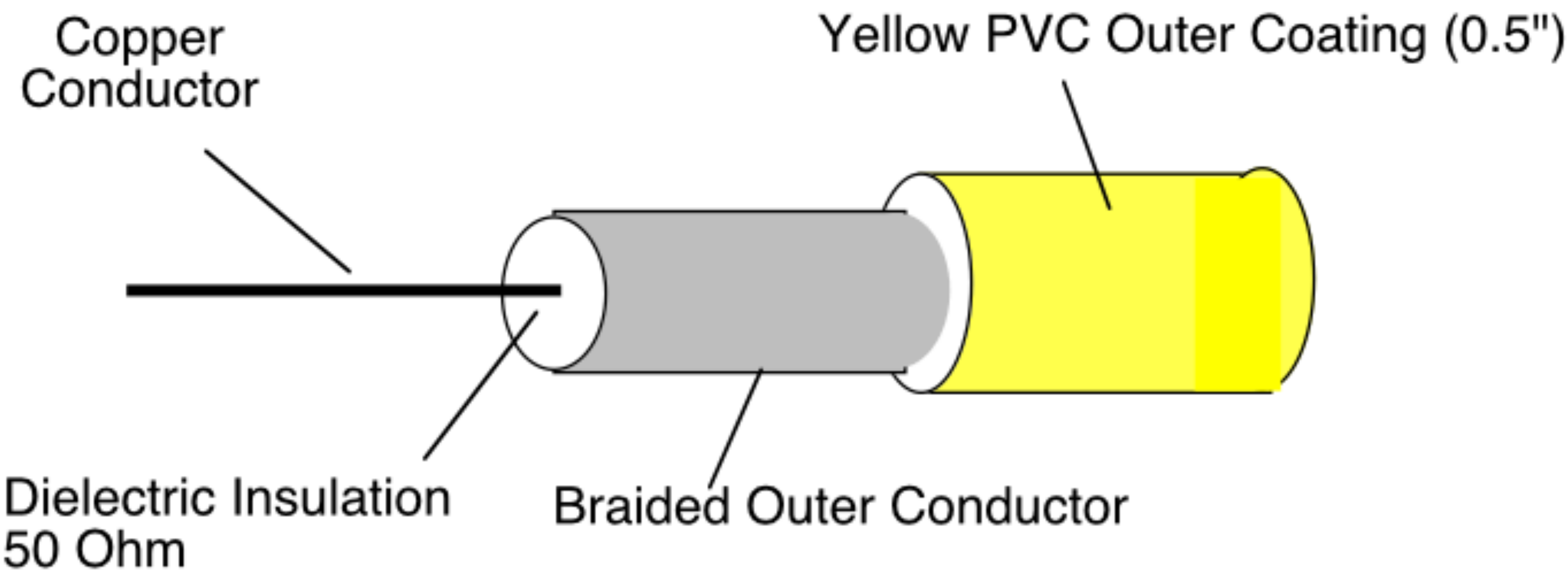
IEEE 802.3

Two original variants: Thick Ethernet and Thin Ethernet at 10 Mbps

Speeds now available:

- 100 Mbps (Fast Ethernet)
- 1000 Mbps (1 Gbps)
- 10000 Mbps (10 Gbps)
- 40 Gbps, 100 Gbps, …

10B5 Ethernet Media



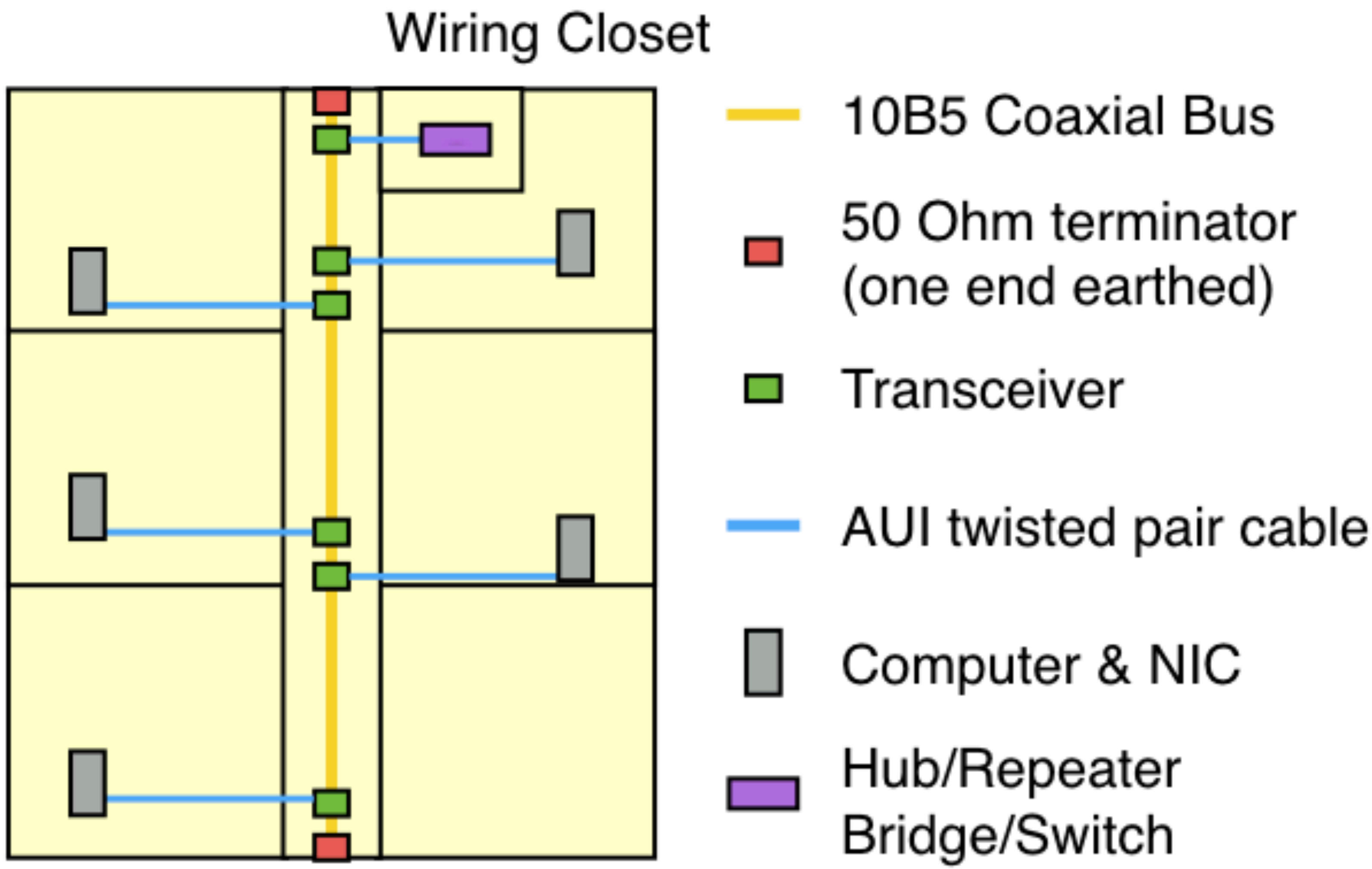
High performance co-axial cable

- Segment length ≤ 500m
- Good noise immunity
- N-Type connector at each end

1024 NICs attached to a single cable segment

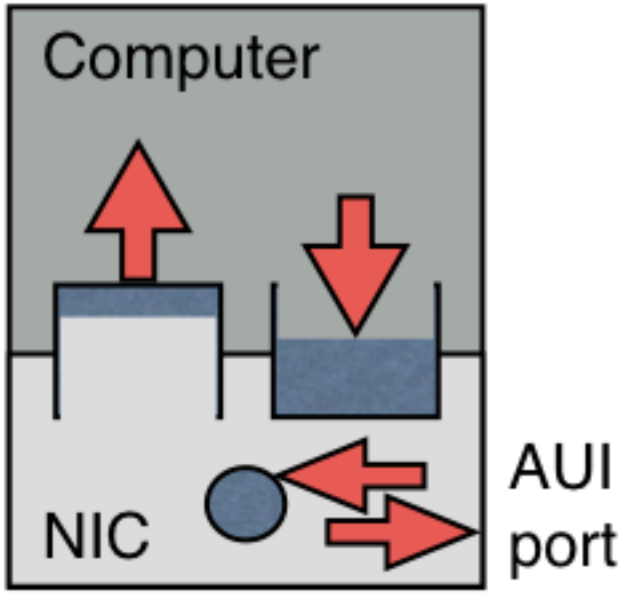
Ethernet 10B5 Cable Segment

Cable usually installed as a trunk running down corridor



Typical Use of 10B5 within an Office (max 500m segment)

Ethernet Network Interface Card (NIC)



Originally a card inserted in a PC or computer

Transmission and reception using the media

Input (receive) and Output (transmit) queues

A queue of frames for transmission

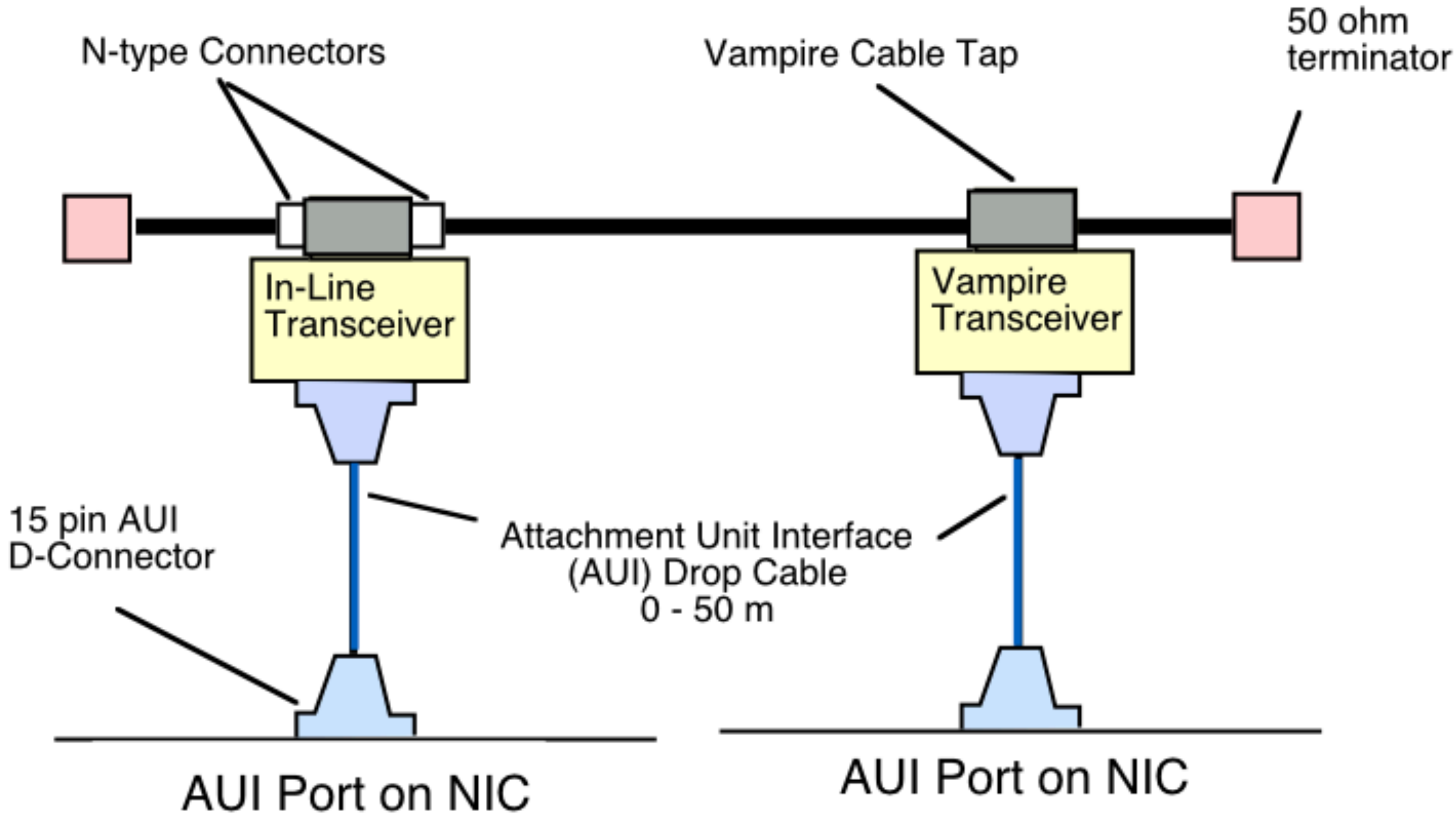
- Sender completes a Tx descriptor in the queue (location of data in memory, length of data, etc)
- Sender writes a register in the NIC to ask for this to be sent
- NIC then performs a DMA of the data, serialises data and adds information needed to transmit a frame on the cable

A queue to hold received frames

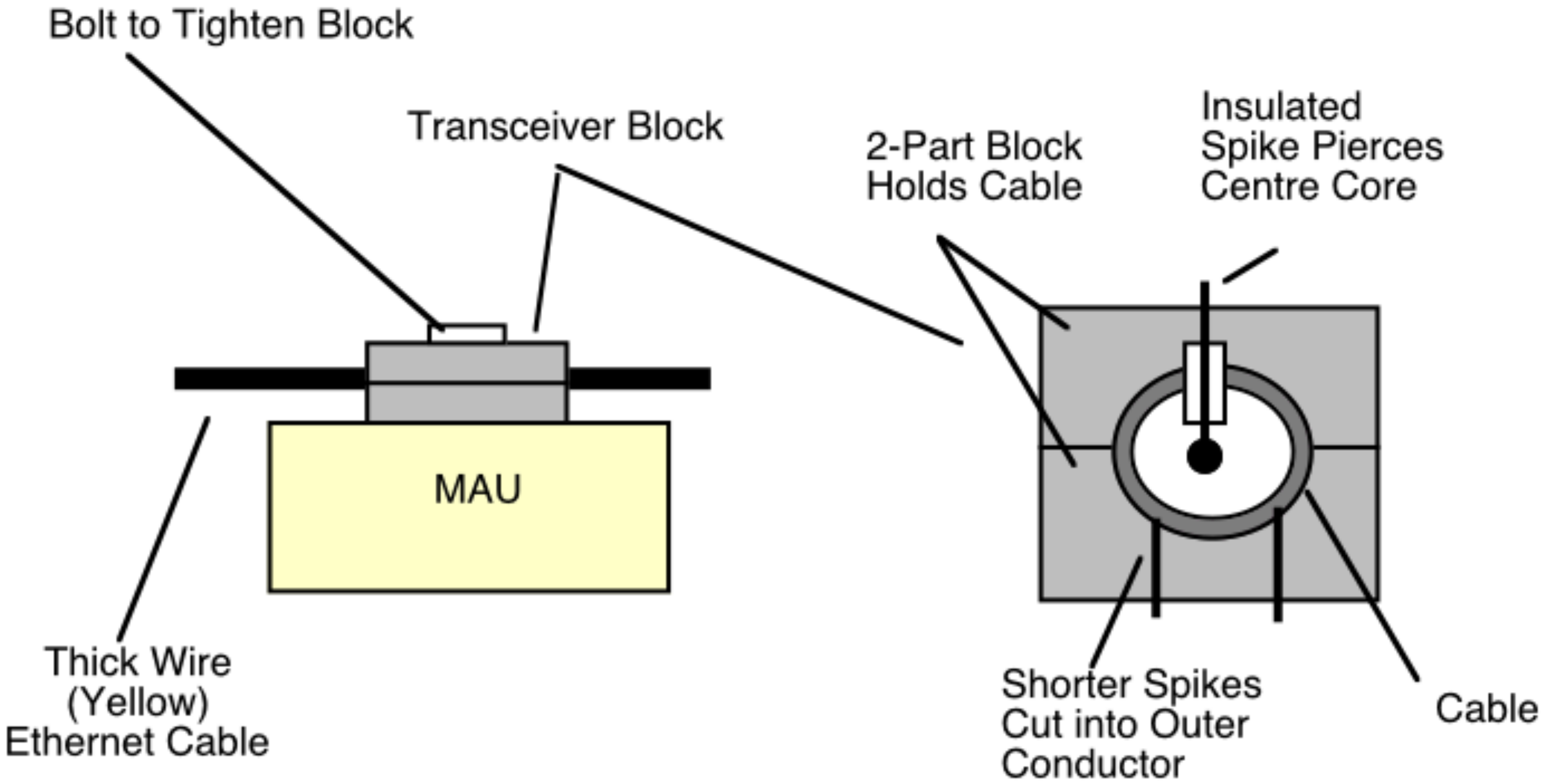
- The NIC processes a frame received on the cable
- The frame is stored internally and a Rx descriptor is created
- The data in valid frames is DMA'ed to computer memory
- The receiver is interrupted to say that frames have been received

10B5 (Thick Ethernet Transceiver)

Two types of transceiver are supported:
- In-Line (N-type screw connector as cable installed)
- Vampire transceiver (insulation displacement after cable installed)



10B5 (Thick Ethernet Vampire Transceiver)



Cable drilled; transceiver block tightened around the cable

This connects spikes to outer and inner cable conductors

Transceiver electronics (MAU) bolted to the transceiver block

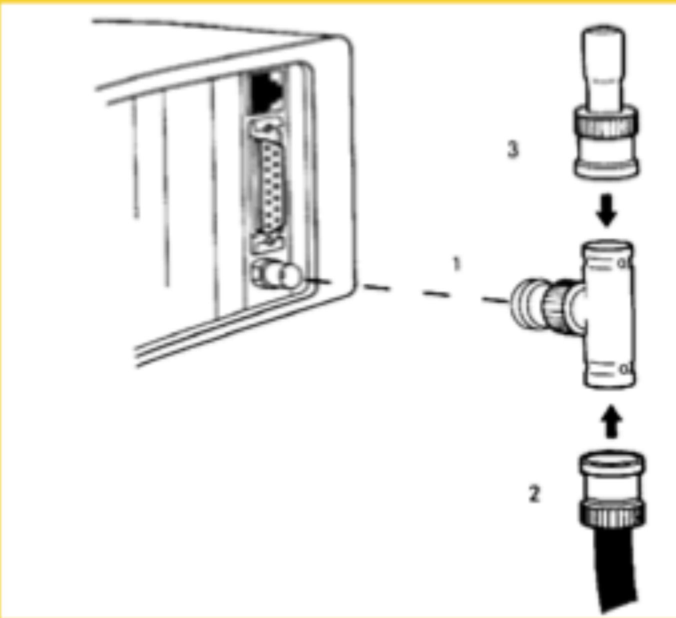
Medium Attachment Unit (MAU)

Ethernet Success Story

- **Simple low cost LAN (compared to computers)**
- **Familiarity to customers !!!**
- **Wired networks are still the most common media**
- **Has become an standard for Internet LANs**

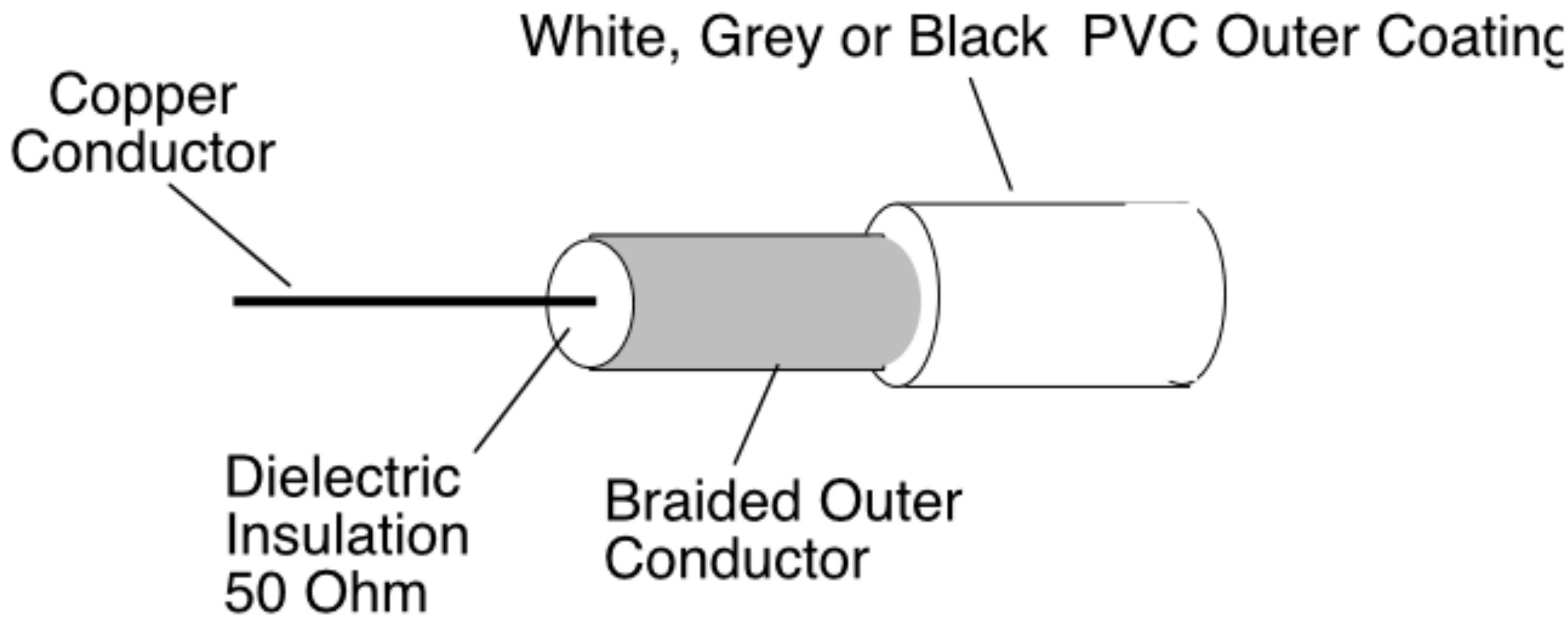
The Origins of the Ethernet LAN

10B2 Coaxial Cables



Module 1.2

10B2 (Thin Ethernet)



Low cost co-axial cable (RG58u)
- Segment length 0.5m ≤ 185 m
- Flexible, easy installation

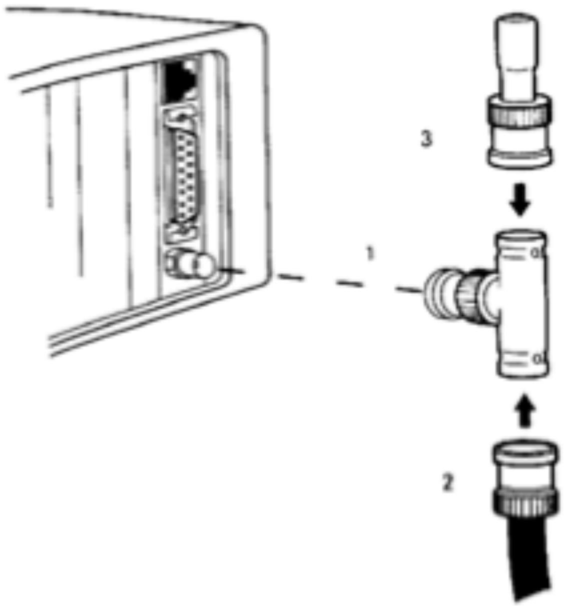
30 NICs allowed using one segment

NICs with In-built or external transceiver

10B2 (Thin Ethernet)

BNC connector at each end of cable

"T" joiner connects the NIC to two cables



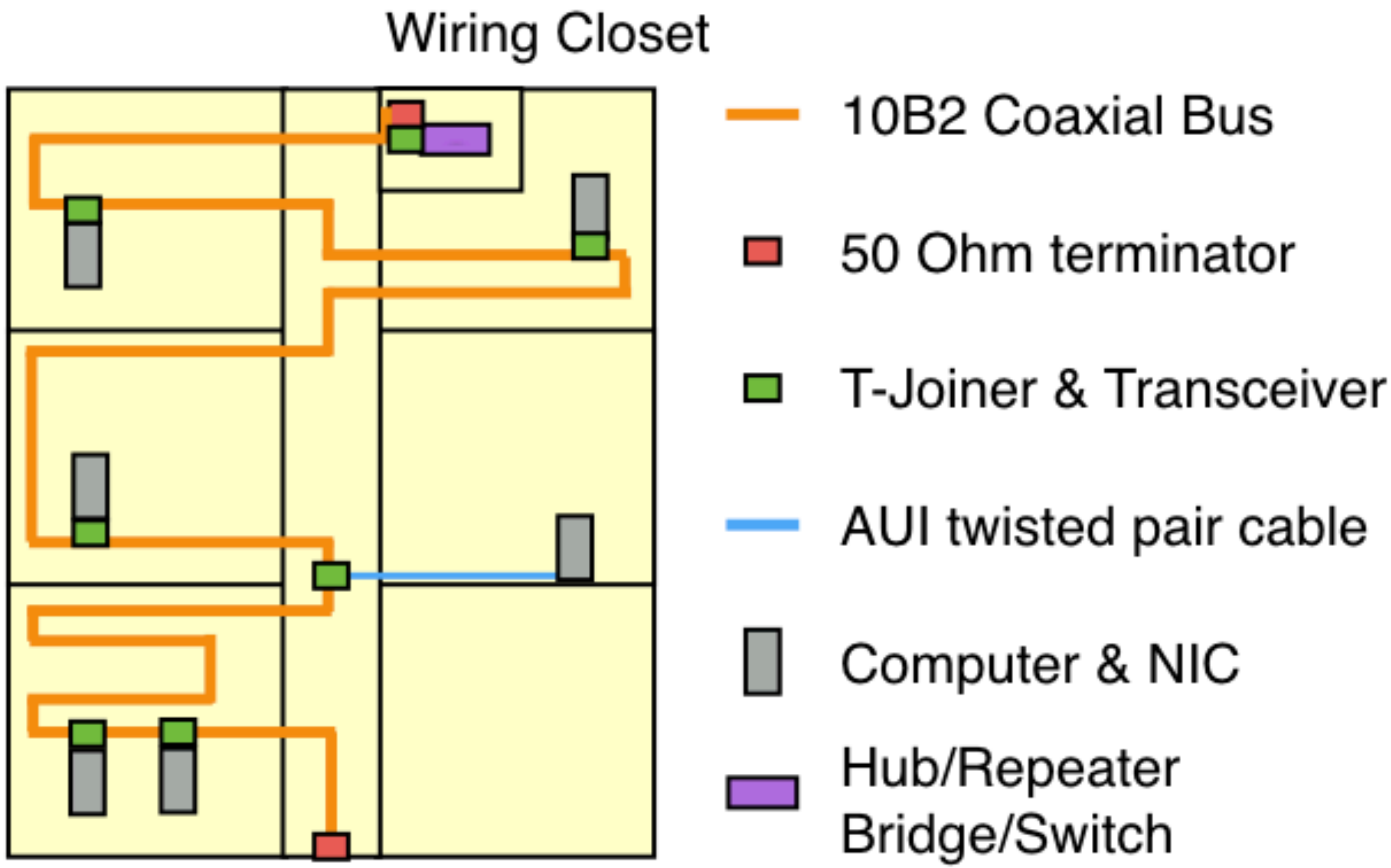
BNC connector and "T" joiner
BNC 50 Ohm Terminator

NICs with In-built or external transceiver

Flexible lengths of cable with BNC plugs

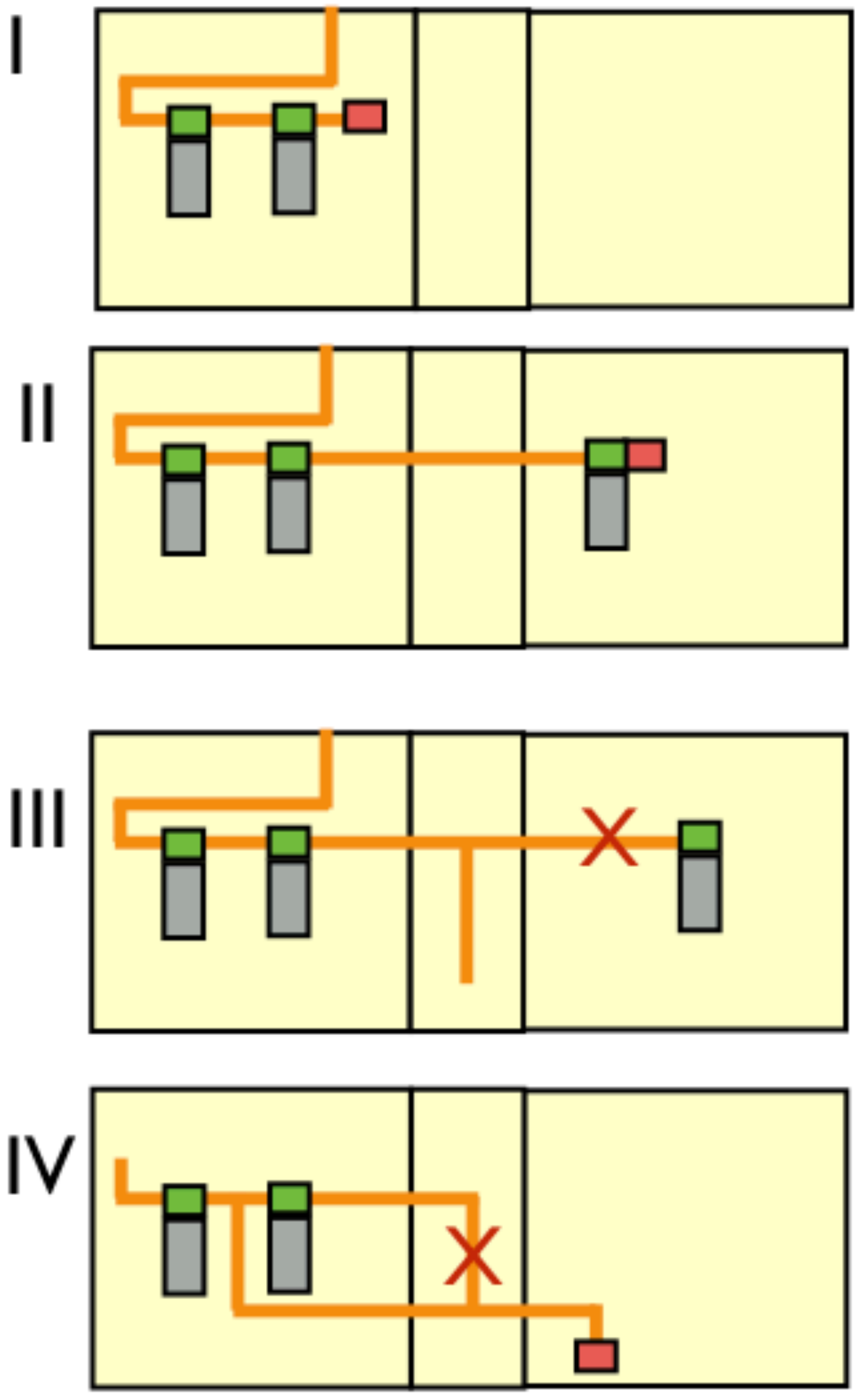
Ethernet 10B2 Cable Segment

Often cable connected device to device, rather than pre-installed



Typical Use of 10B2 within an Office (max 185m segment)
Maximum of 30 computers on a single segment

10B2 (BNC Connector)



Easy to install

Plug “T” into NIC and connect cable!

Unplug the BNC connector

add another "T" and a new cable

and connect another NIC

Must form one bus:

No loops

No stubs between “T” and NIC

Easy to extend....

... difficult to manage (unstructured)!

Ethernet Success Story

- **Ethernet already familiar with customers**
- **10B2 made the network even more Cost-Effective**
- **Very Easy to Install**

  Simple BNC twist connector

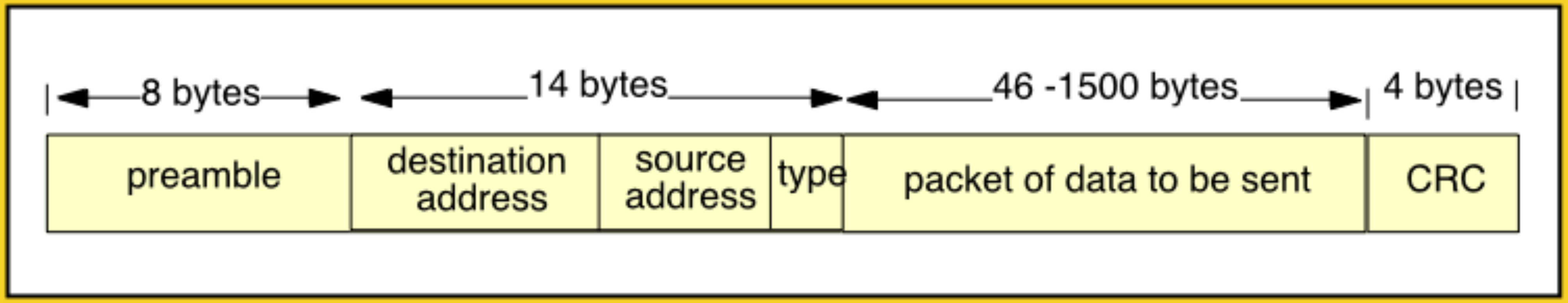
  Great for unstructured networks that can evolve
- **A larger LAN can use Repeaters**

Check the web pages and notes for 10B2 and 10B5!
Coaxial cable Ethernet is now only used in special networks.

Ethernet Frames

Addressing
A shared physical medium
Medium access control
Sending frames
Frame reception
Multicast and Broadcast

Link Layer



Module 2

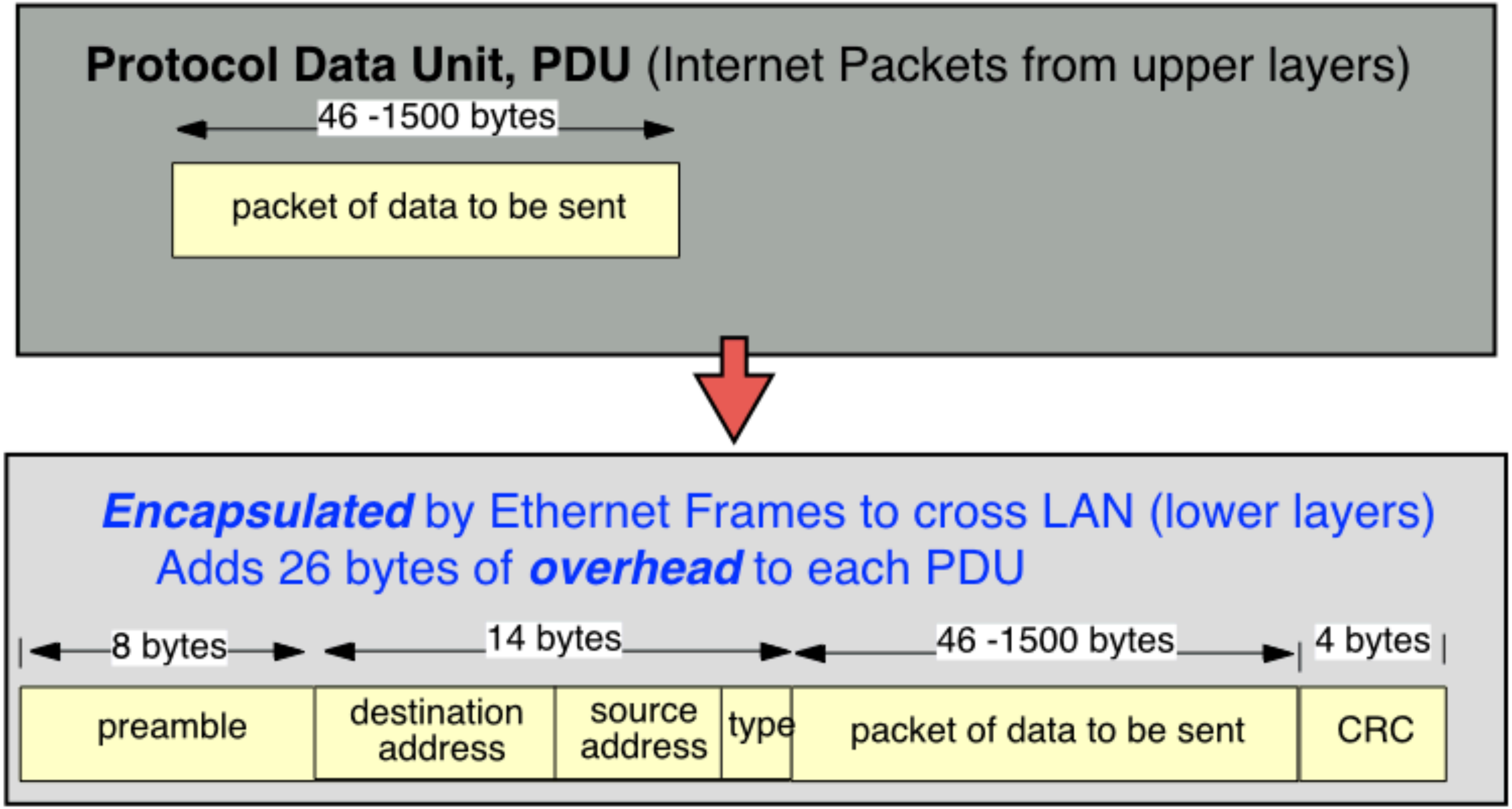
Ethernet Frames:
Addressing

Link Layer

FF:FF:FF:FF:FF:FF

Module 2.1

Ethernet Frames



e.g.

a 46 byte packet is carried in a 72 byte frame

a 1200 byte PDU is carried in a 1226 byte frame.

Ethernet MAC Address

08:00:20:00:00:01

A MAC Address is a 48-bit number

- Usually represented as 6 pairs of hexadecimal digits
- One hex digit corresponds to a value 0-15 (0x0 to 0xF)
- For ease of reading we separate each byte* by a colon.

We divide the 48-bit address into two parts:

- First 3 bytes: the organisationally unique identifier (OUI) - orange
- Second 3 bytes: the manufacturer-assigned address - yellow.

** In some documents an 8-bit byte is referred to as an octet.*

Ethernet MAC Address

08:00:20	:00:00:01

- Each Network Interface Card (NIC) has a unique MAC Address
  - Held in a manufacturer-configured PROM
- Addresses are globally unique
  - A MAC Vendor Code (OUI) + Number
  - About 1% of OUIs have been used
- IEEE *sells* these blocks of addresses to *manufacturers*
  - Each block has 256 cubed addresses
  - That is 16 Million!!

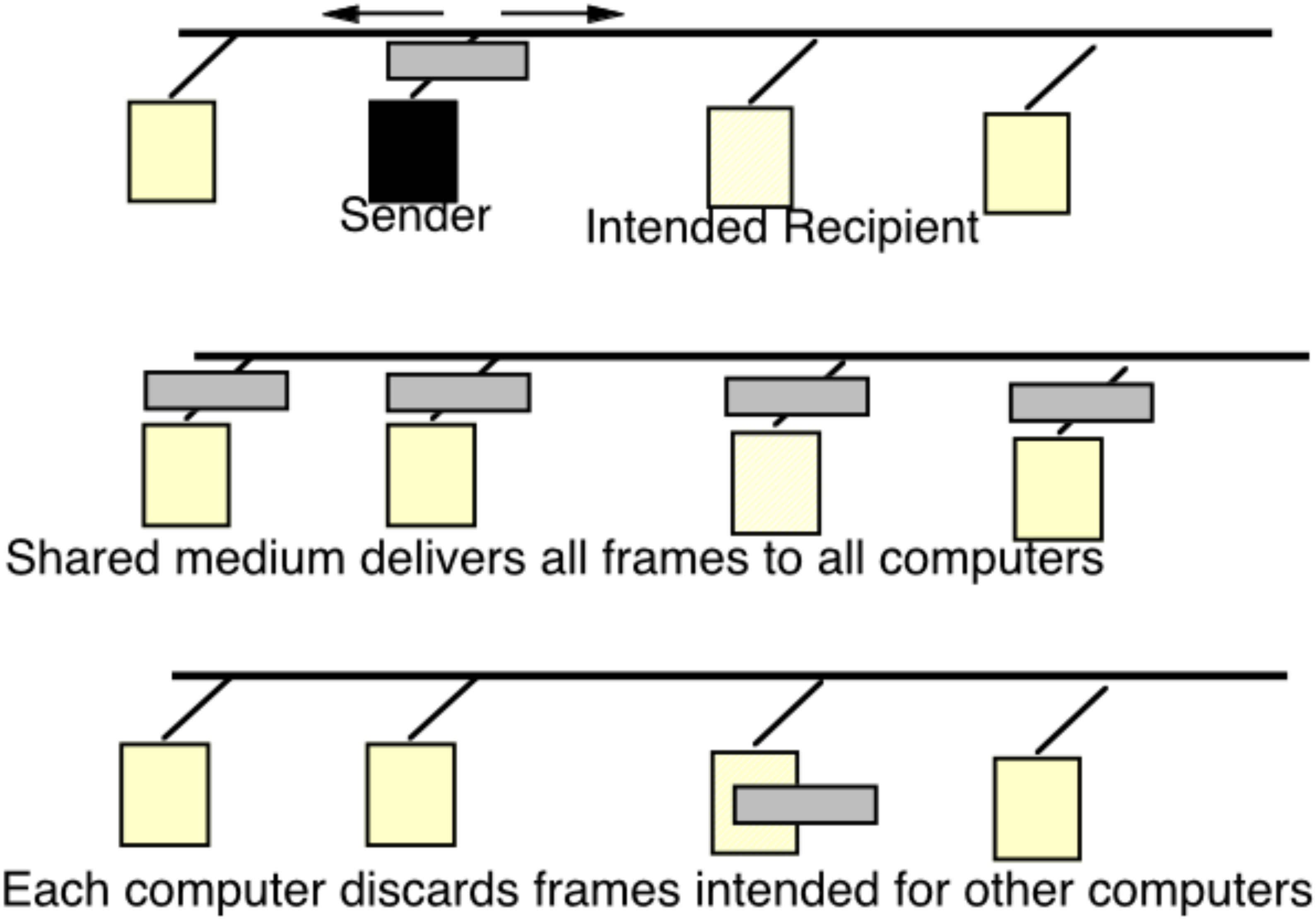
MAC Vendor Codes (OUIs)

08:00:20:00:00:01

The first 3B of address indicates the assigned manufacturer

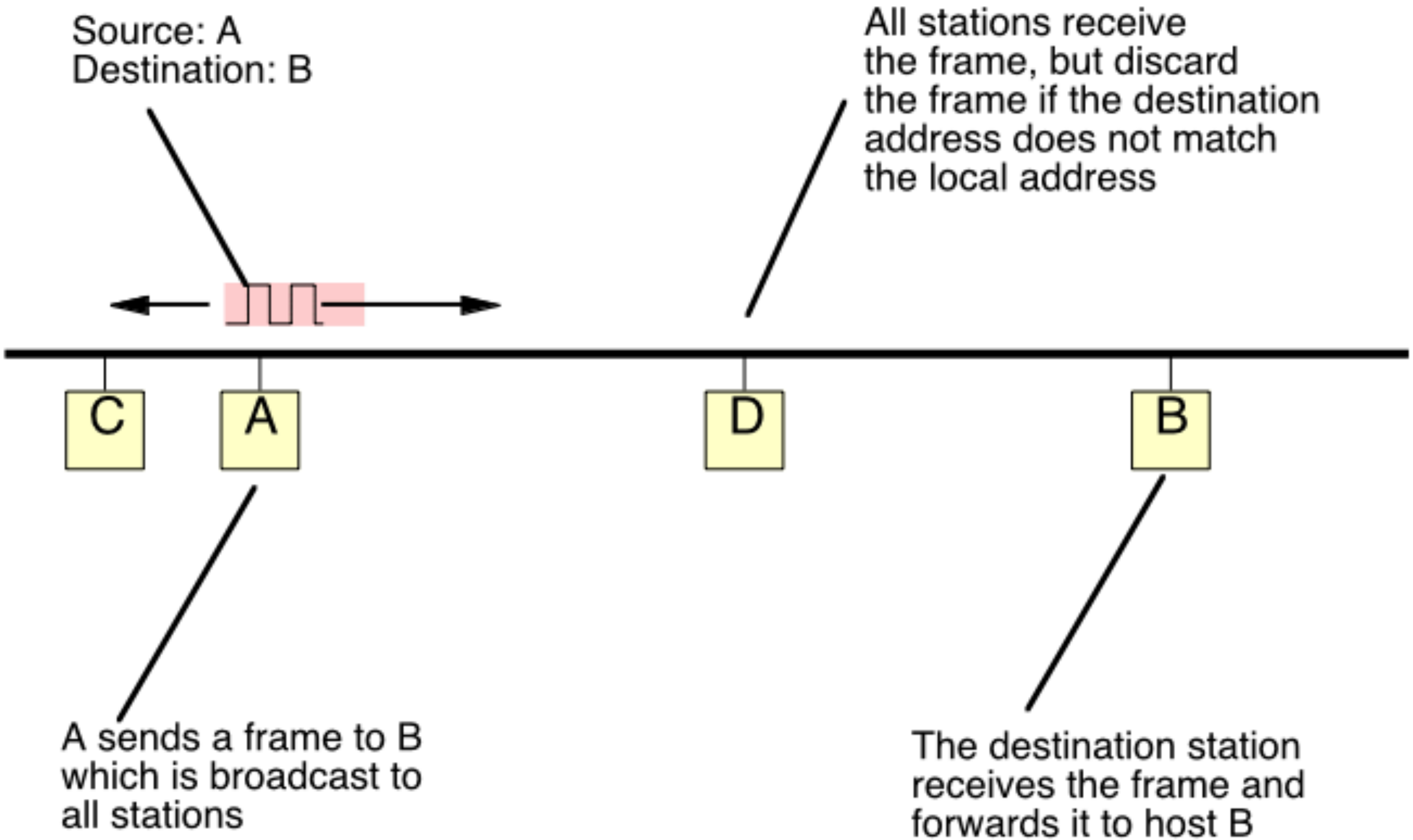
080002	3Com (Formerly Bridge)	
080003	ACC (Advanced Computer Communications)	
080005	Symbolics	Symbolics LISP machines
080008	BBN	
080009	Hewlett-Packard	
08000A	Nestar Systems	
08000B	Unisys	
080011	Tektronix, Inc.	
080014	Excelan BBN Butterfly, Masscomp, Silicon Graphics	
080017	NSC	
08001A	Data General	
08001B	Data General	
08001E	Apollo	
080020	Sun	Sun machines
080022	NBI	
080025	CDC	
080026	Norsk Data (Nord)	

Shared Access to Ethernet Medium



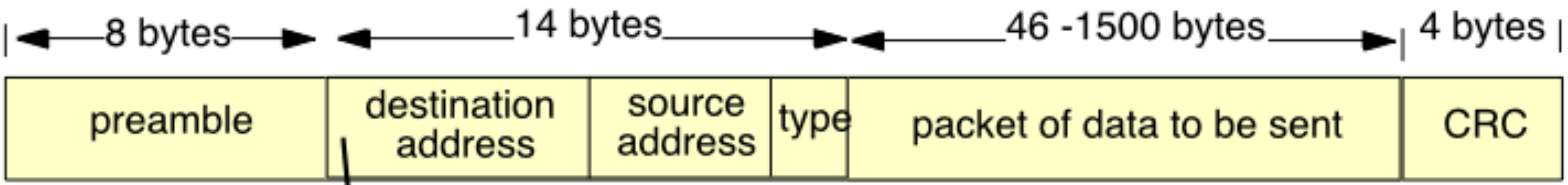
We can use the destination address to perform unicast communication, where frames are only received by a specific destination computer

Using the Destination MAC address



This assume a sender knows the value of the MAC address in the remote NIC's PROM (we'll find out how it does this later!)

Ethernet Frame Structure



LAN address of intended recipient

first bit = 0 indicates point to point
first bit = 1 indicates broadcast or multicast

48 bits, expressed as 12 hexadecimal digits
e.g., 12:34:56:78:9A:BC

A theoretical 200,000,000,000 addresses
Actually 70,000,000,000... (2 bits are used)
20,000 MAC addresses for each person on the planet!

Special MAC Addresses

FF:FF:FF:FF:FF:FF

The all 1's Address is used to send to all NICs

- Known as the **broadcast** destination address
- Only ever used as ***destination address***

00:00:00:00:00:00

The all 0's Address is special

- Known as the **unknown** address
- Only ever used as ***source address***

Use of Broadcast Frames by IPv4 ARP

FF:FF:FF	FF:FF:FF

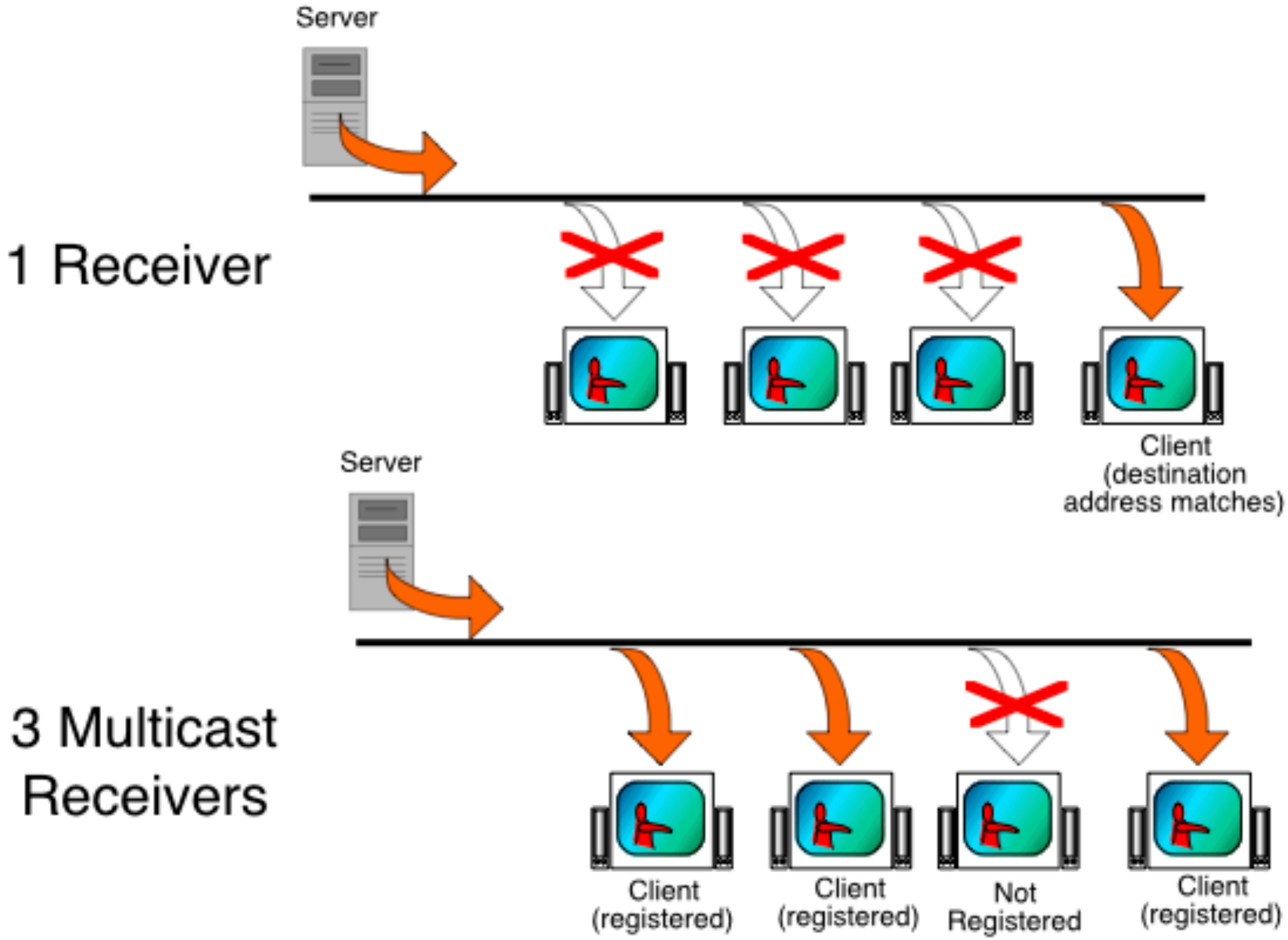
A sender using the IPv4 Address Resolution Protocol (ARP) sends an ARP request to discover the MAC address of the interface (NIC) with which it wishes to communicate.

The ARP request therefore is sent as a ***broadcast frame***.
This request is received by ***all systems*** on the same Ethernet LAN.

In contrast, the interface (NIC) responding to an ARP request already knows the address of the system sending the ARP request.

The ARP reply is sent in a ***unicast frame*** directed to the querier.
Only the querier receives this requested response.

Sending to multiple recipients: Multicast on Ethernet



TV/Radio/etc Transmission (can often have several receivers)

Also used by some protocols to deliver to multiple computers

IPv4 Group MAC Addresses

01:00:5E	00:00:01

*

Groups addresses

- Have the ***least significant bit*** of the ***first byte*** to 1
- The remainder of the address carries the specific group address
- Last 23 bits of the IP group destination address, e.g., 224.0.0.1

Group addresses identify ***channels*** not Receivers

- Sender chooses a group address to use
- e.g. one channel may carry a specific Internet TV station
- another channel might be used to advertise DNS in a LAN

NICs need to ***register*** to receive from a group

- A computer may ***register*** none or more group addresses
- e.g. a multicast DNS client registers IP address 251.0.0.224
  - This registers for the MAC address of 01:00:5E:00:00:FB
- The NIC passes all frames that match a registered group address

** IPv4 Address mapping*

IPv6 Group MAC Addresses

33:33:xx	xx:xx:xx

Groups addresses

Have the ***least significant bit*** of the ***first byte*** to 1

The remainder of the address carries the specific group address copied from the last 32-bits of the IPv6 group destination address.

IPv6 doesn't use broadcast packets at all

Instead it uses multicast to send packets to groups of receivers

Some Layer 2 protocols also use multicast:

e.g. the Spanning Tree uses address 01-80-C2-00-00-00 to send control frames to the next adjacent Ethernet Switch.

The sender doesn't know the MAC address used by a switch, but does not want its frames to be received by other NICs.

Addressing Summary

- **All NICs have a MAC Address**

  Provides a handy income stream to the IEEE :-)

- **All NICs receive every frame with:**

  a ***broadcast*** MAC destination address ff:ff:ff:ff:ff:ff

  a destination address that matches its ***MAC address***

  a destination address that matches a ***registered*** multicast group address (i.e. used by a program on the computer)

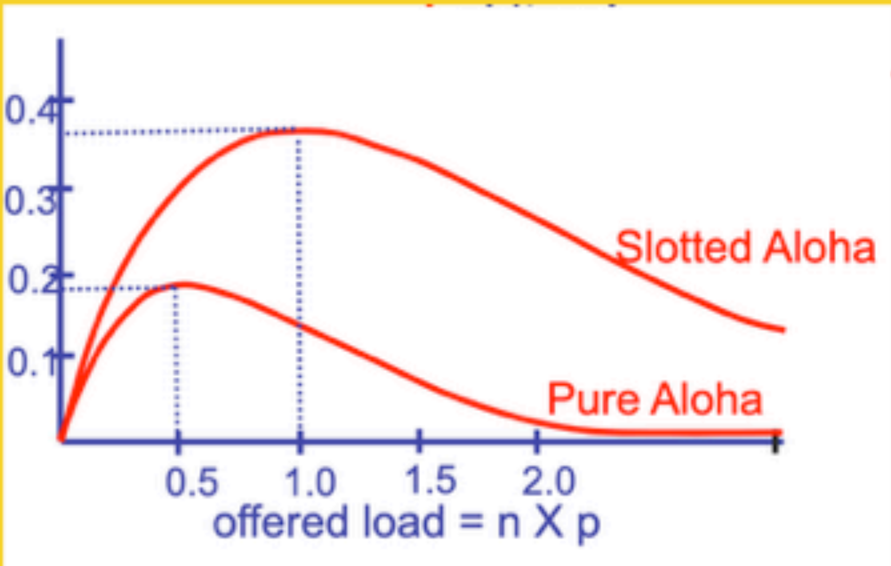
- **All filtering is performed within the NIC:**

  Computer does not know about discarded frames

  A computer can override filtering, by placing the NIC into ***promiscuous mode*** - where all frames are received

Ethernet Frames:

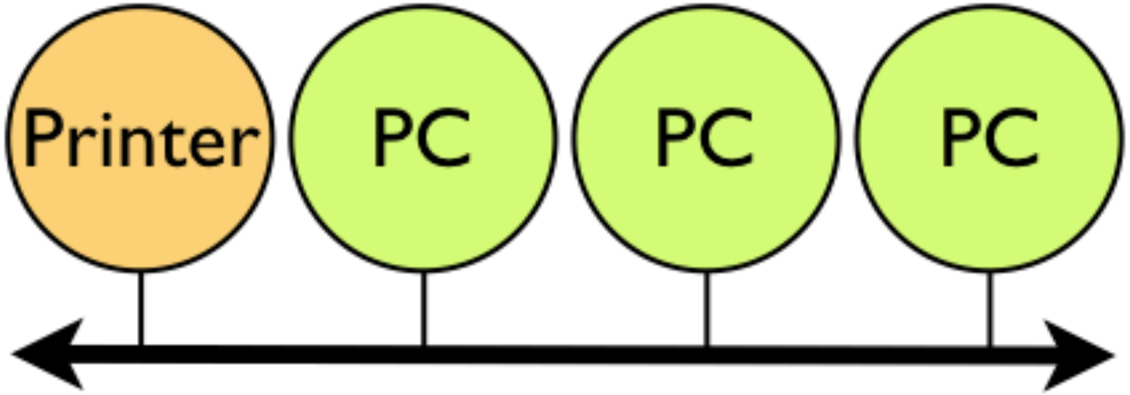
A shared physical medium



Link Layer

Module 2.2

Sharing the media



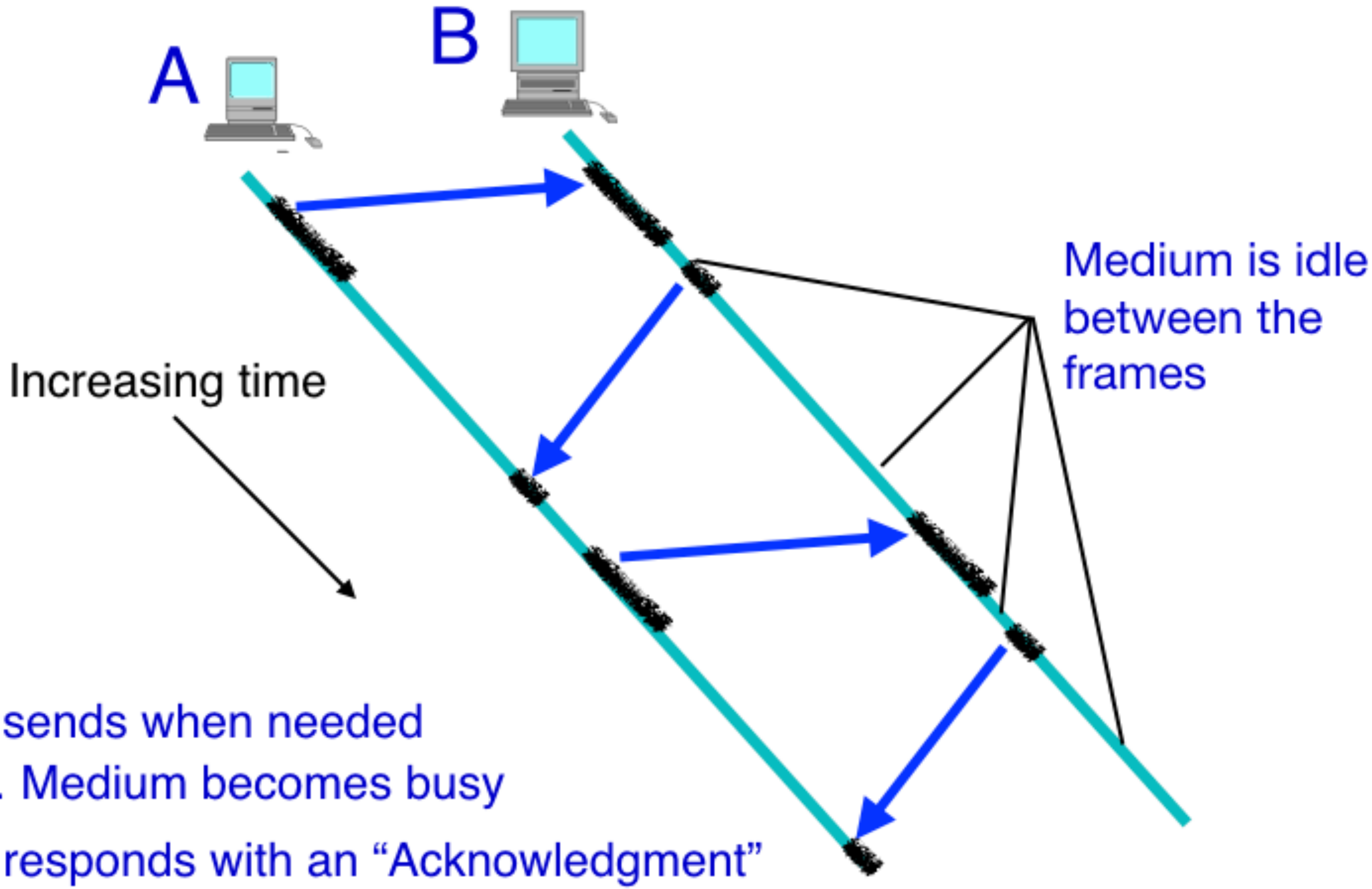
There is only one medium (the cable)

All NICs ***should*** be able to use this cable

Clearly only ***one*** should send at a time!

So, how does a NIC ***know*** if it may send?

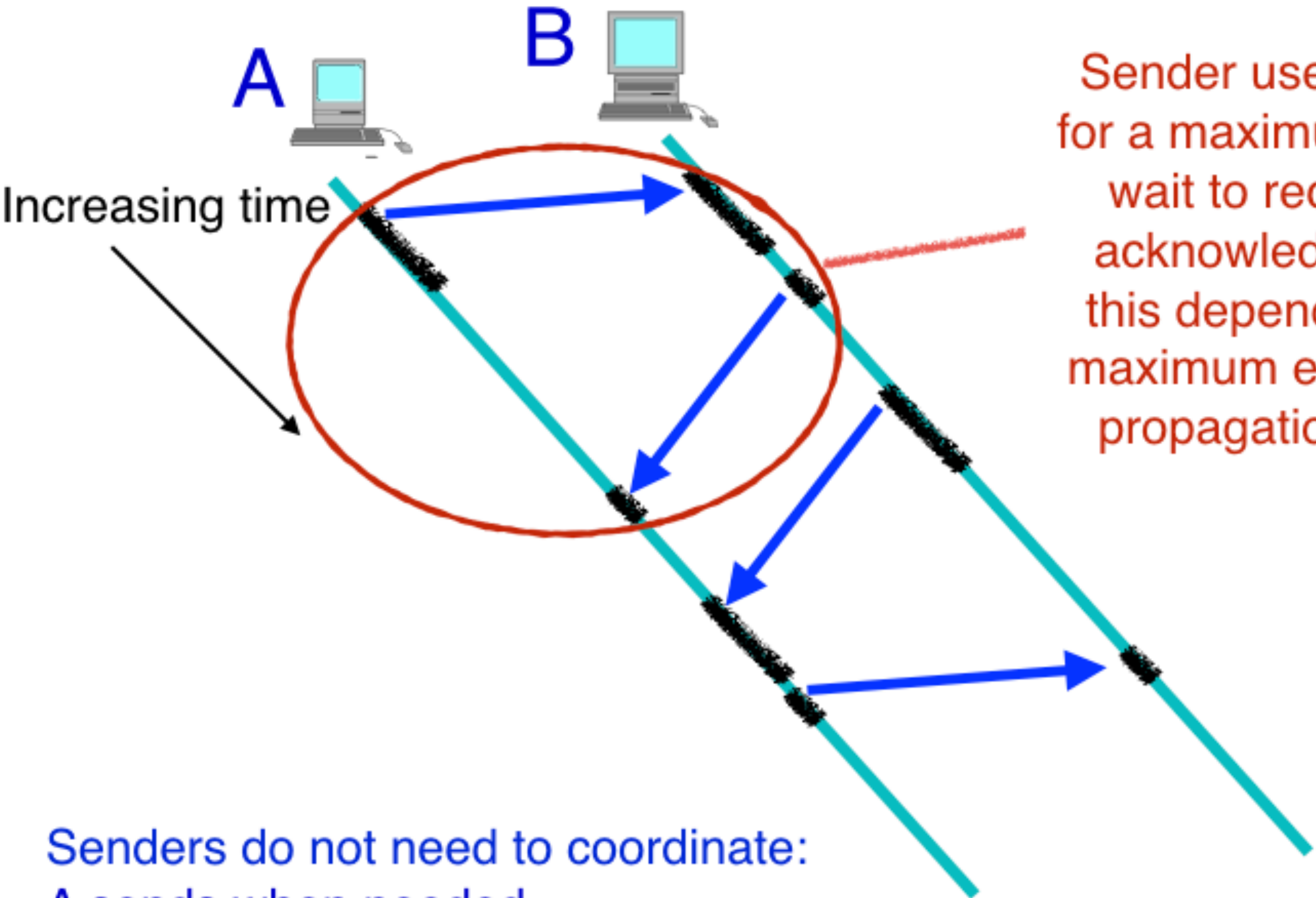
Medium Access using ALOHA



A sends when needed
… Medium becomes busy

B responds with an “Acknowledgment”
Either … A knows that B has received the data
… or ... it will resend the data again to B

Medium Access using ALOHA



Sender uses a timer for a maximum time to wait to receive an acknowledgement, this depends on the maximum end-to-end propagation delay.

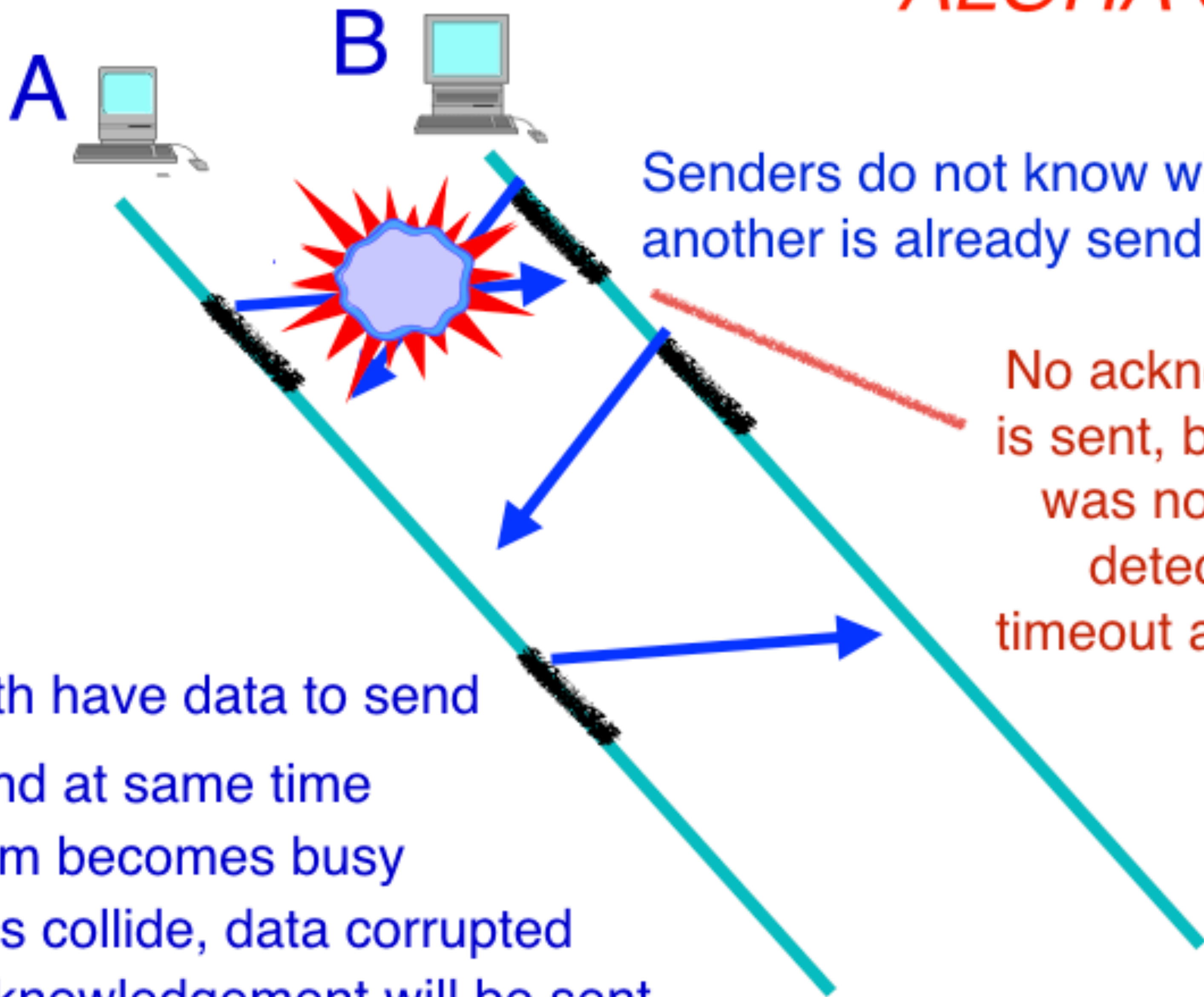
Senders do not need to coordinate:

A sends when needed

B sends when needed

Half Duplex Communication

ALOHA Collision



A & B both have data to send

A & B send at same time

… Medium becomes busy

… Signals collide, data corrupted

… No acknowledgement will be sent

A & B will both need to send again later

As the load increases, the chances of collision also increases

Slotted ALOHA

If there is a common clock source we can divide time into slots.
All senders need to know the start of each timeslot
Senders only transmit a frame at the start of a timeslot

Timeslot	1	2	3	4	5	6	7
Sender 1							
Sender 2							
Sender 3							
Outcome	1	Empty	2	3	. . .		

Increasing time →

Slotted ALOHA

If there is a common clock source we can divide time into slots.
All senders need to know the start of each timeslot
Senders only transmit a frame at the start of a timeslot
Timeslots with only one frame result in successful transmission

Timeslot	1	2	3	4	5	6	7
Sender 1							
Sender 2							
Sender 3							
Outcome	1	Empty	2	Collision	3	2	1

Efficiency of ALOHA

Suppose *n* senders have data to send with probability *p*
The probability of success for ALOHA,

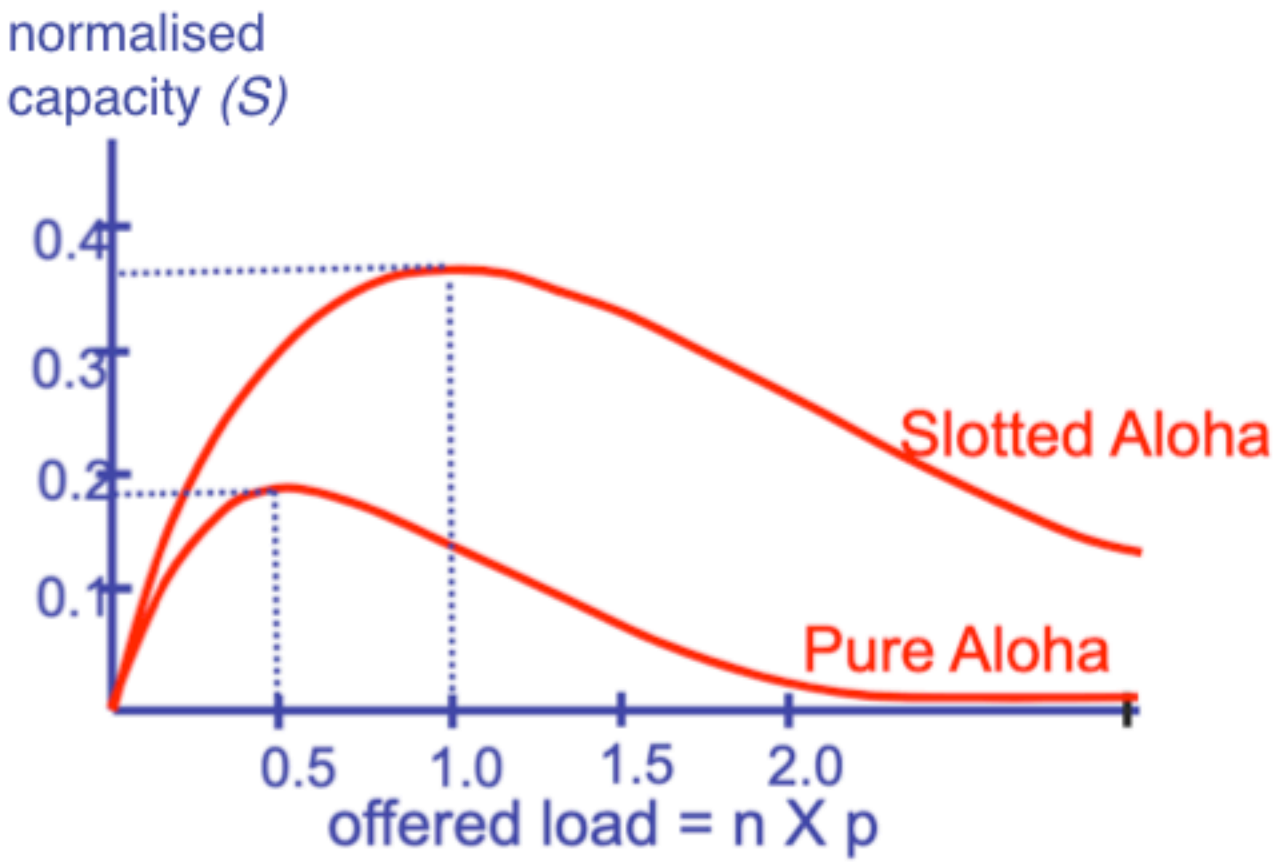
$S = p\,(1-p)^{(n-1)}$

Maximum capacity
S= 0.18

For slotted ALOHA,

$S = np(1-p)^{(n-1)}$

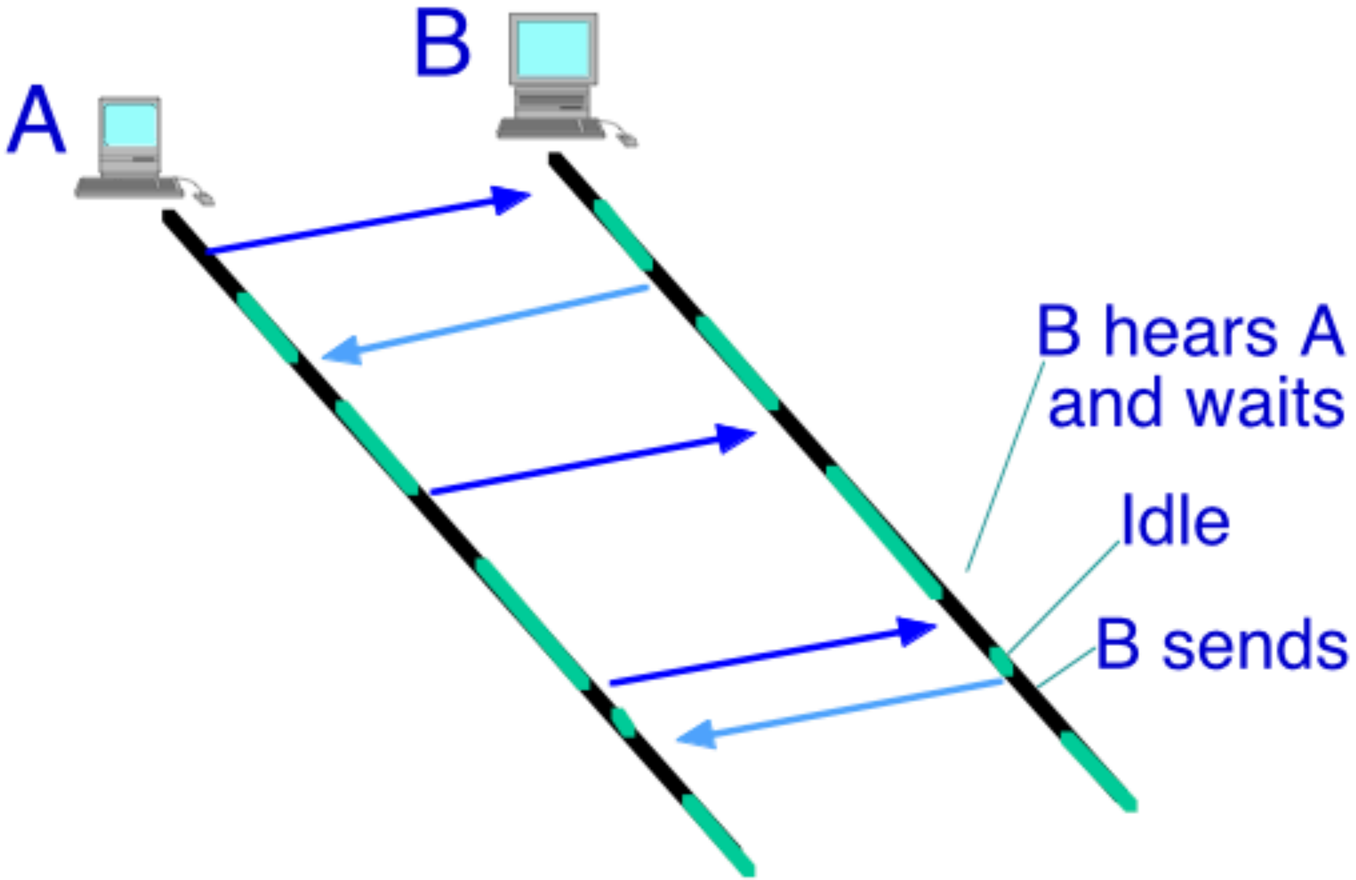
For optimal p,
S= 1/e = 0.37



Slotted ALOHA is much better than ALHOA, but still achieves <37%

Listen-Before-Talk

Listens for activity on the cable before sending
Requires Carrier Sense (CS) circuit



Also called *Carrier Sense Multiple Access* (CSMA)
Does not work well when one sender is a *long distance* from another

ALOHA Summary

- **ALOHA is really very simple**

  Requires setting a timer to detect loss of an acknowledgment

- **Slotted ALOHA**

  Slotted ALOHA more efficient than unspotted version

- **Carrier Sense or Listen Before Talk**

  Carrier Sensing improves efficiency

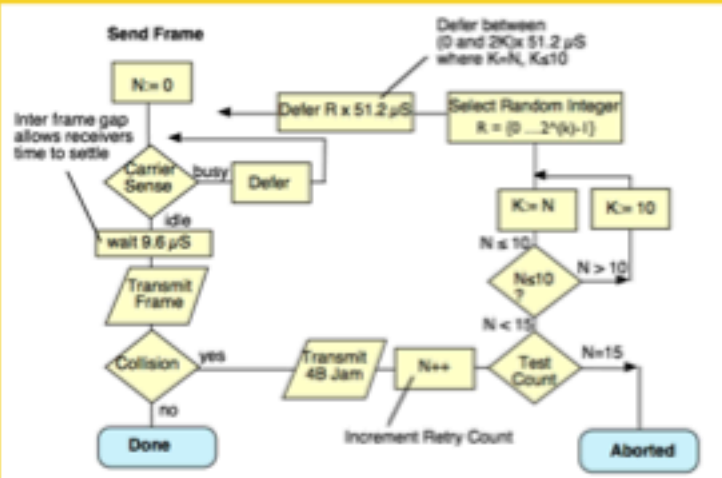
  Not the design chosen for Ethernet, but still used in other networks

Ethernet Frames:

Medium Access Control

Medium Access Control (MAC) needs to solve three challenges:

- how to be decentralised with no "master" controller
- how to scale to large numbers of active nodes
- how to deal with propagation delay

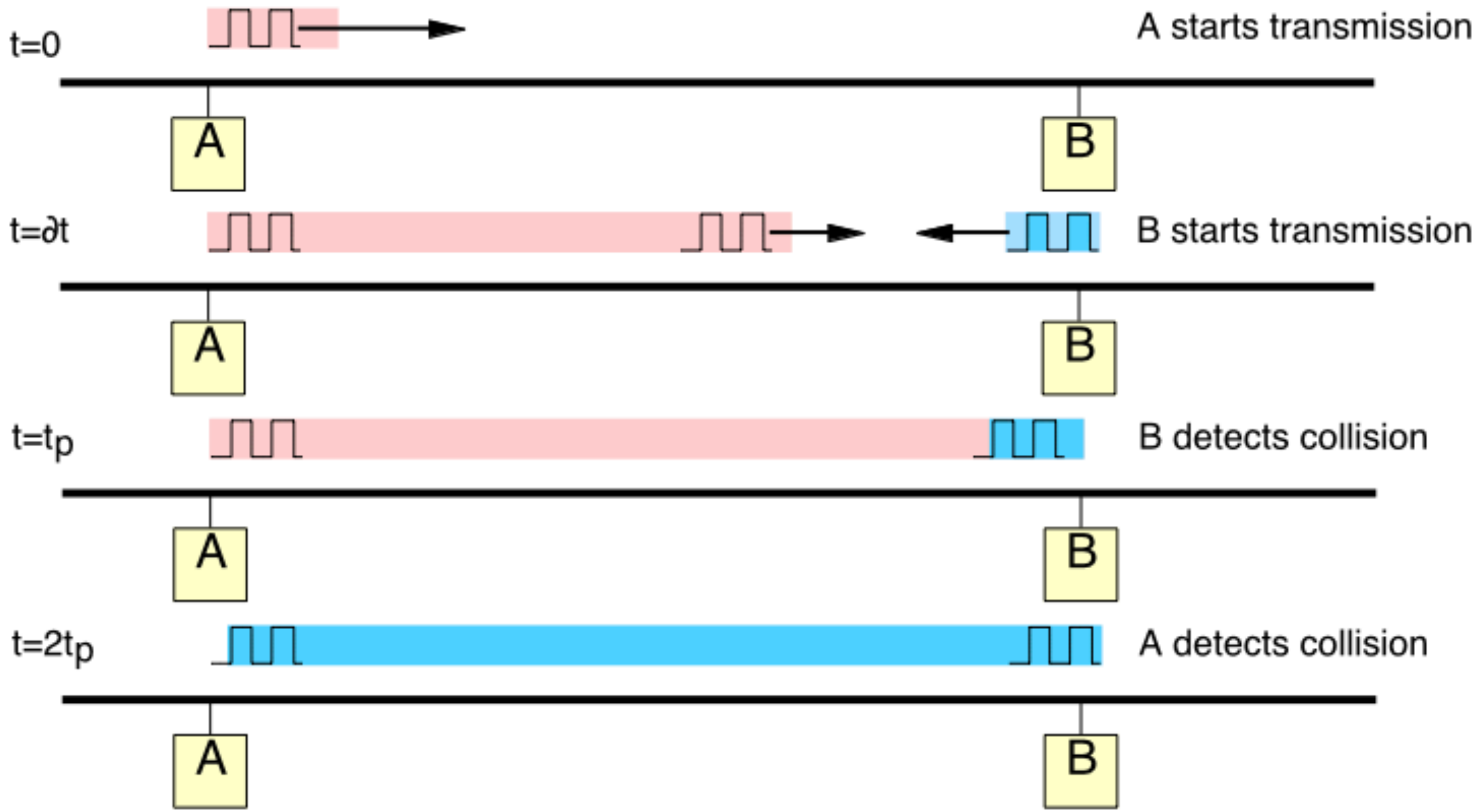


Link Layer

Module 2.3

Collisions and Collision Detection

Nodes try to avoid collisions by using a Carrier Sense (CS) to detect when the medium is idle before they start sending



Requires Collision Detect (CD) to detect a collision

Slot Time

All senders need to know when any collision occurs

The time to detect a collision depends on the propagation delay
... and other delays

In a CSMA/CD system this is set by the ***slot time***

The slot time for IEEE 802.3 of **51.2** μs at 10 Mbps

This limits the maximum cable ***distance*** to 3km at 10 Mbps

The need to detect a collision sets the ***minimum*** frame size
The minimum Ethernet frame size is 64B (60 bytes+CRC32)

Parameters impacting the Slot Time

Component	Properties	Delay (microsec)
AUI Cable	6x 50m , 0.65c	3.08
Transceiver	3 transceivers (6x 1.2 micosec)	7.2
3xCoax Medium	e.g. 1500m, 0.77c	13
2xOther Media	e.g.1000m, 0.65c	10.26
Repeater delay	Propagation delay	2
Signal Rise Time		8.4
Elec Circuit	Propagation delay	1.05
Total		44.99

Total Slot Time of system < 51.2 μs

This is for informational only (not required in the exam)

Retransmission after Collision

The minimum frame size assures us that all nodes that are sending will ***detect*** the collision.

After detecting a collision, sends a JAM and then stops sending.

The data has not been sent, and therefore needs retransmitted.

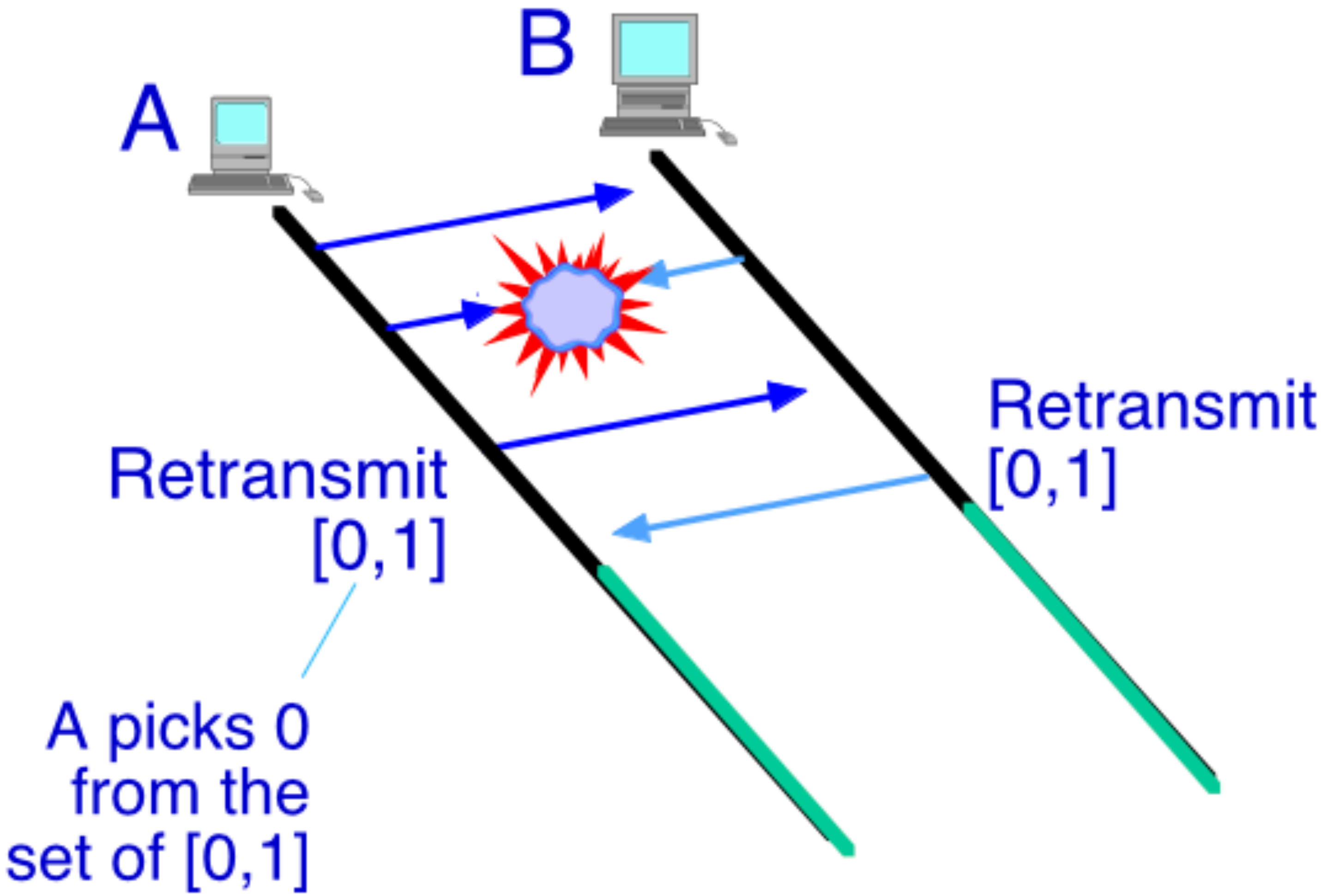
They need to send at different times to not suffer another collision.

A decentralised method has no way to know which node or nodes experienced the collision.

The method therefore chooses to ***random backoff time*** to delay their own retransmission.

If they choose different random backoff times, they suceed.

Backoff and Retransmission



In this example, after the first collision k=1

A & B choose from a set of 2^k values: In this case: [0, or 1]

50% probability that A & B choose different retransmissions

A happens to choose [0], and so waits t x 0. Therefore it sends first

Detail of Random Exponential Backoff

If multiple NICs retransmit at the same time, a collision will occur again

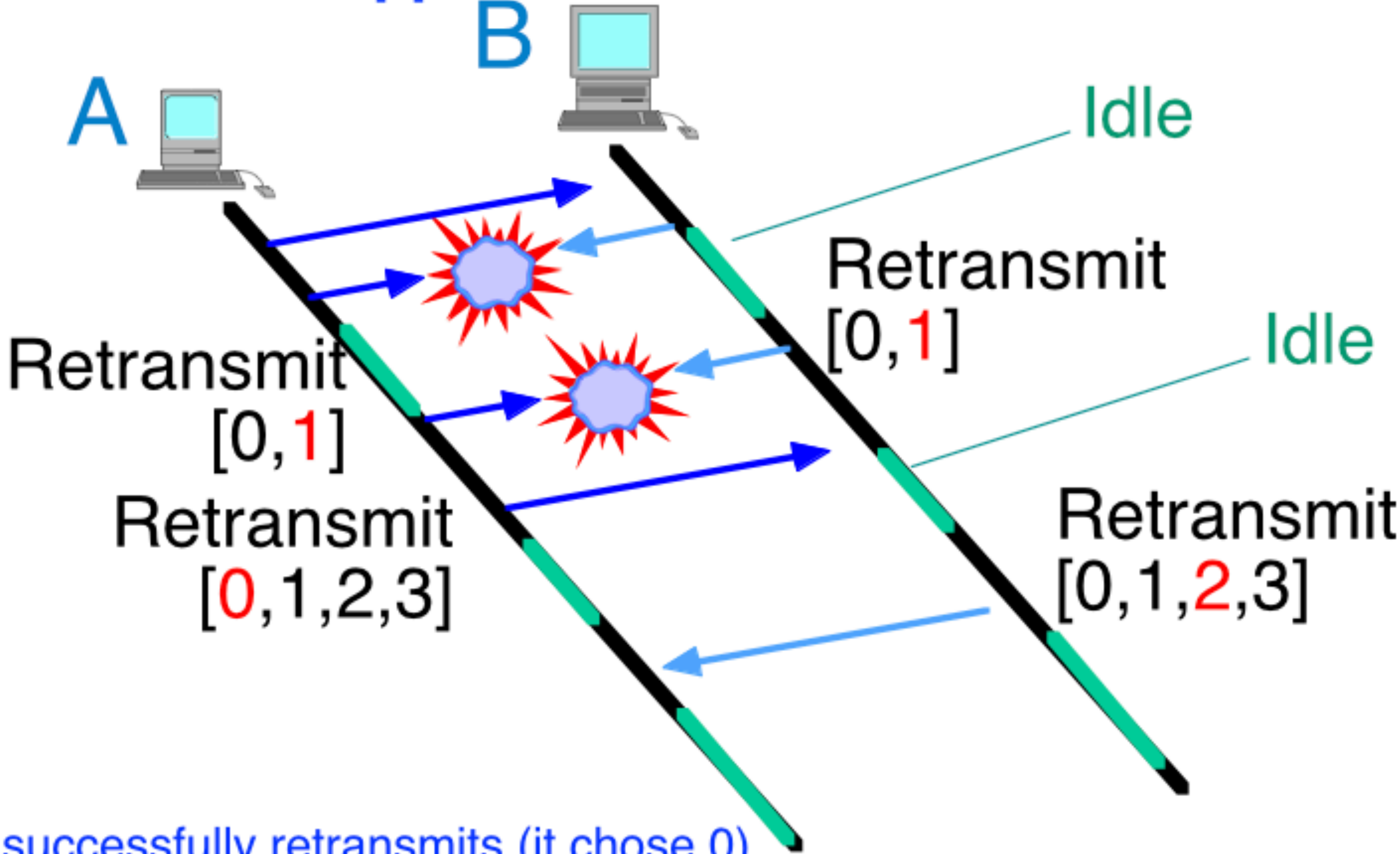
- Senders jam the medium and then back-off!

Each sender waits for a randomly chosen period of time

- k counts the set of values, increasing each retransmission, initially k=1
- Senders choose a random number from a set of values [0... (2^k-1)]
- Wait for the chosen value multiplied by Ethernet Slot Time ($51.2\mu S$)
- Each attempt increases k, so the set increases ([0,1], [0,1,2,3],[0..7]…)
- This exponentially increases the random backoff set

Exponential Back-Off

In this example there is a collision; both A,B happen to choose the same backoff time [1] and so a second collision must occur...



A successfully retransmits (it chose 0)
B defers one slot time (it chose 1)

Random Backoff

[0,1] First Retx		
Random number at A	Random number at B	Result
0	0	Collision
0	1	A sends first
1	0	B sends first
1	1	Collision after 1 slot time

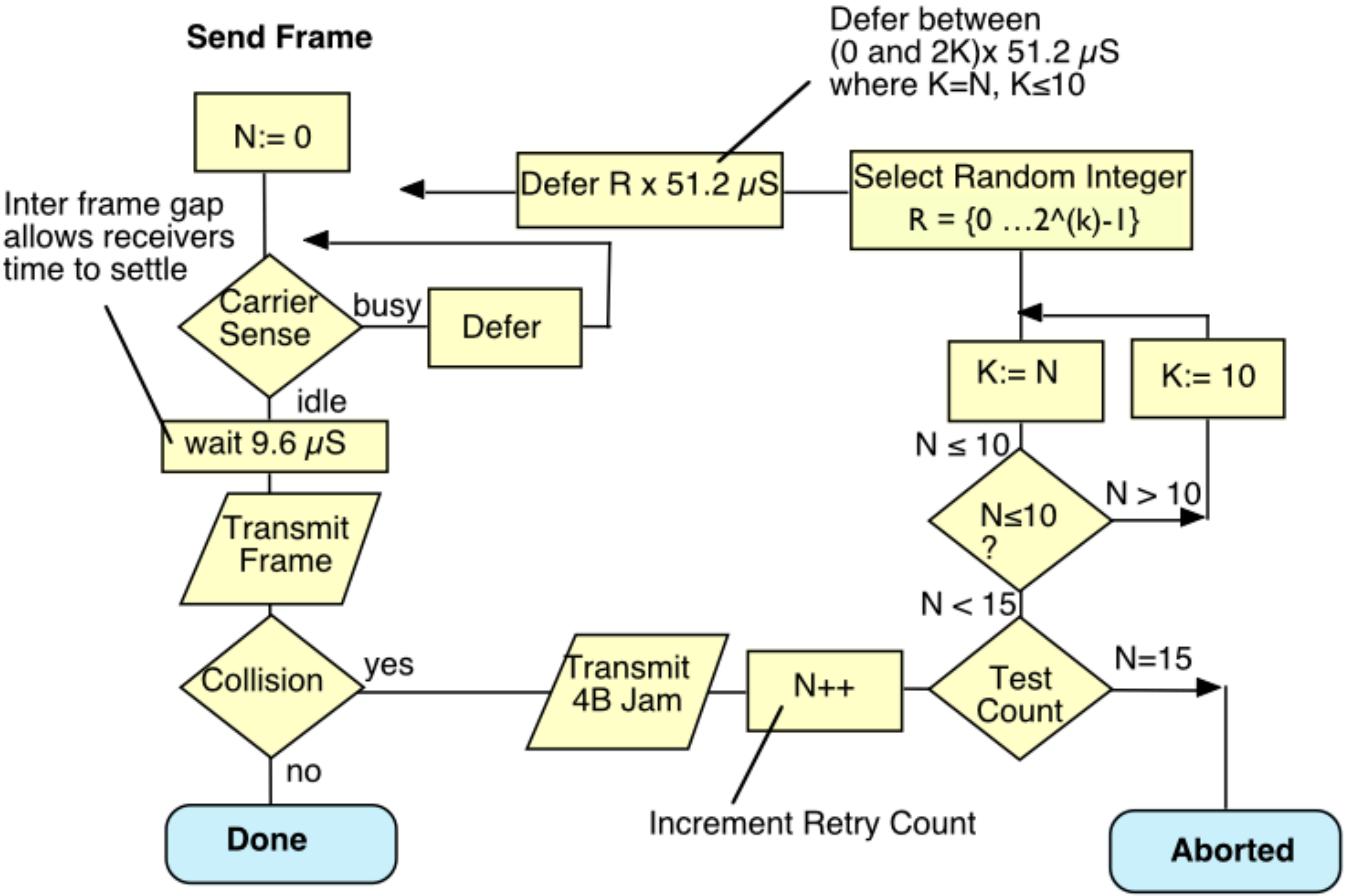
[0,1,2,3] Second Retx		
A	B	Result
0	0	Collision
0	1	A sends
0	2	A sends
0	3	A sends
1	0	B sends
1	1	Collision
1	2	A sends
1	3	A sends
2	0	B sends
2	1	B sends
2	2	Collision
2	3	A sends
3	0	B sends
3	1	B sends
3	2	B sends
3	3	Collision

1st attempt 50% chance of collision - fair to each sender

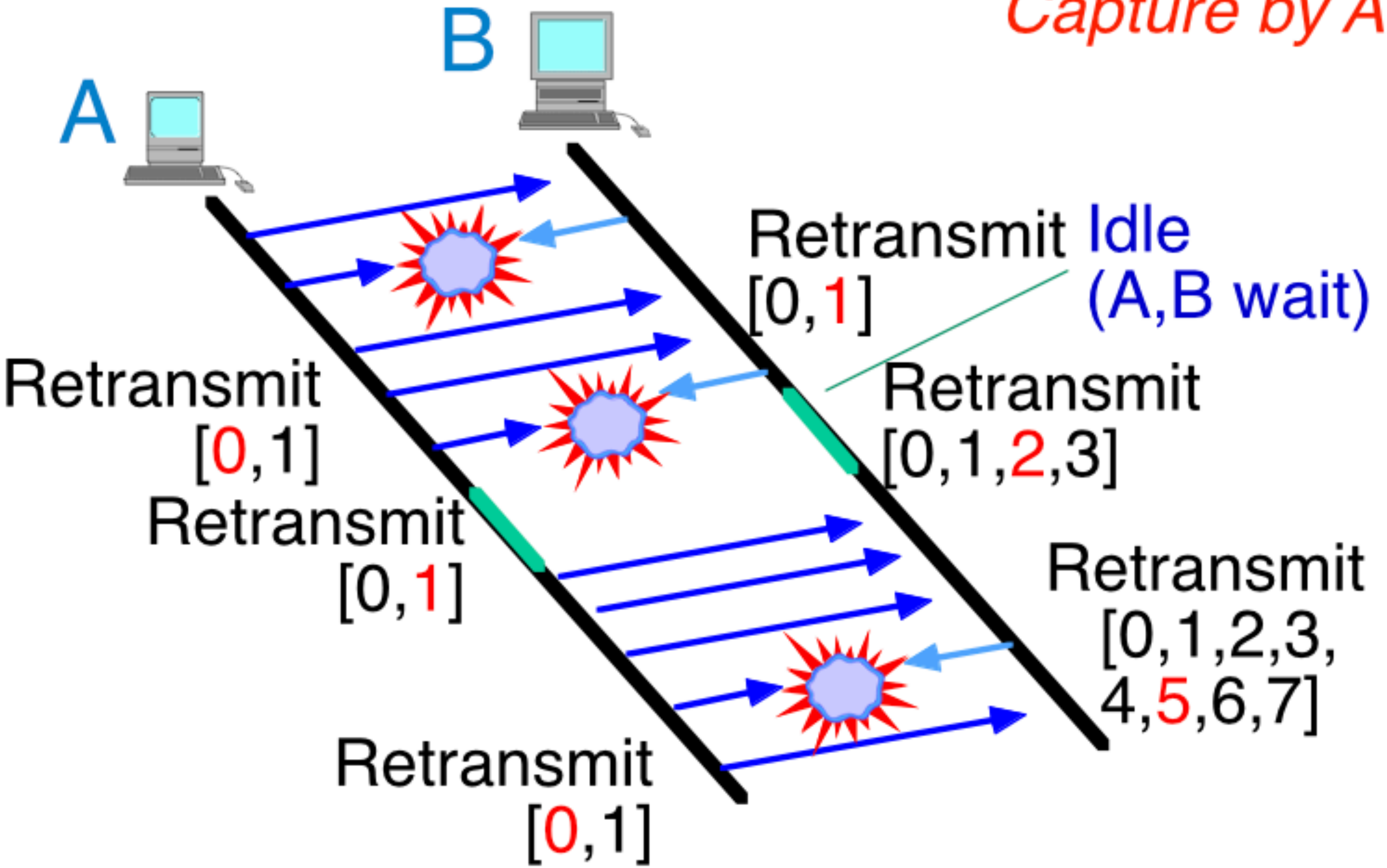
2nd attempt 25% chance of collision - fair to each sender

The collision probability halves each retransmission round <10

CSMA/CD


Send Frame
N:= 0
Defer R x 51.2 μS
Select Random Integer
R = {0 …2^(k)-1}
Defer between
(0 and 2K)x 51.2 μS
where K=N, K≤10
Inter frame gap
allows receivers
time to settle
Carrier
Sense
busy
Defer
idle
wait 9.6 μS
Transmit
Frame
K:= N
K:= 10
N ≤ 10
N≤10
?
N > 10
N < 15
Collision
yes
Transmit
4B Jam
N++
Test
Count
N=15
no
Done
Increment Retry Count
Aborted

Capture by A



The algorithm becomes unfair when one NIC sends more than others
This might be common for a router/WIFI Access Point
A brief idle period after sending many frames, restores the fairness

Multiple Access - Summary

Each of these techniques is in use in some form of network

ALOHA

Problem: Many collisions when many nodes

Efficiency: 100% (1 node) 18% (many)

S-ALOHA

Requires Timeslot Synchronisation

Problem: Still collisions when many nodes

Efficiency: 100% (1 node) 37% (many)

Listen-Before-Talk (CSMA)

Requires Carrier Sense (CS) circuit

Problem: Fewer collisions, but still possible

Collision Detection (CSMA/CD)

Requires Carrier Sense (CS) and Collision Detect (CD) circuits

Problem: Capture possible - benefits from limiting burst size

Efficiency: 100% (1 node) higher (many)

Recap: Strengths v Weakness of CSMA/CD

- **Strengths**
  No controlling system needed to coordinate use Ethernet
  Easy to add new systems (NICs) - just plug and play!
  Performance “reasonably fair”

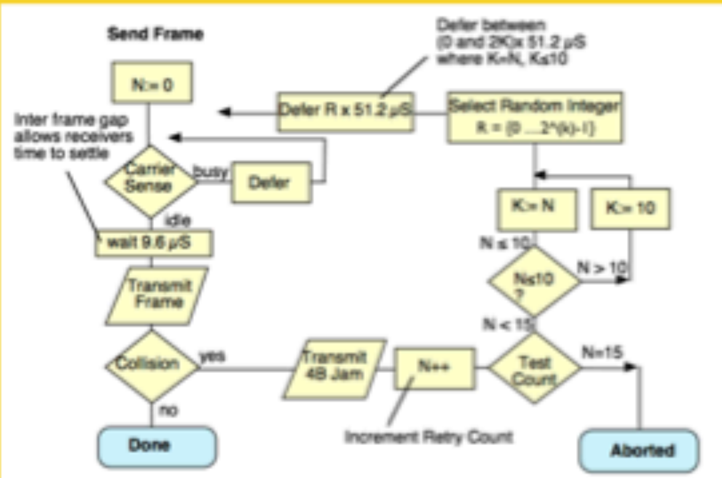
- **Weakness**
  Performance degrades with increasing load
  One “busy” system can “capture” capacity
  - more of a problem for “upstream”
  (e.g., a WiFi base station, router)
  Could fix by limiting bursts of transmission

- **On balance, this was a good design!**

Ethernet Frames:

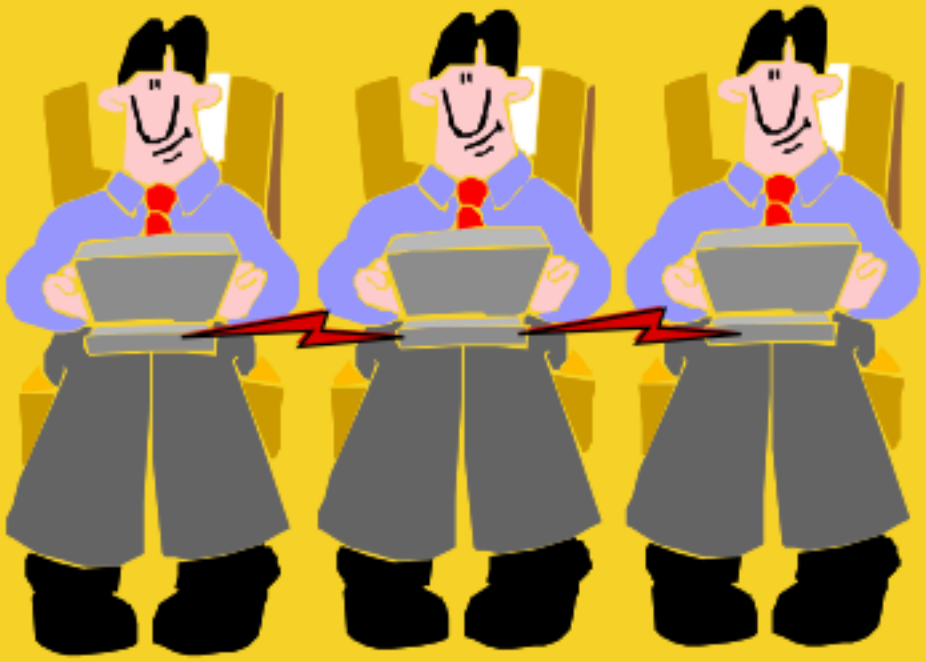
Medium Access Control

CSMA/CD



Module 2 - Additional Material

Wireless Ethernet



Wire-less physical layer
No cable

Module 2.4

Radio Link

2.4-2.485 GHz Industrial Science & Medicine (ISM) Band

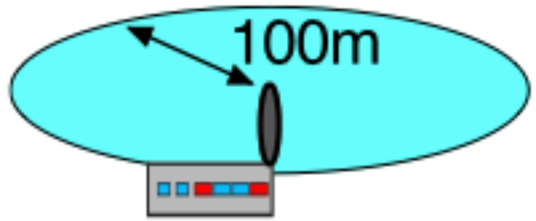
14 channels available worldwide
(fewer channels available in some countries)
Only 3 non-overlapping 20 MHz channels

Uses spread spectrum channels

First used by military ~ 50 years ago
Very high immunity to noise

RF Power

802.11b	100mW
Mobile Phone	600 mW
CB Radio	5W
Microwave Radio	2W



5.15-5.825 GHz Band also used for 802.11n (3 channels)

802.11 Success

WiFi deployment

~500,000 Hotspots in 144 countries!
1,000,000,000 chipsets since 2000
2.5 GHz, 5 GHz, 60 GHz

Speeds

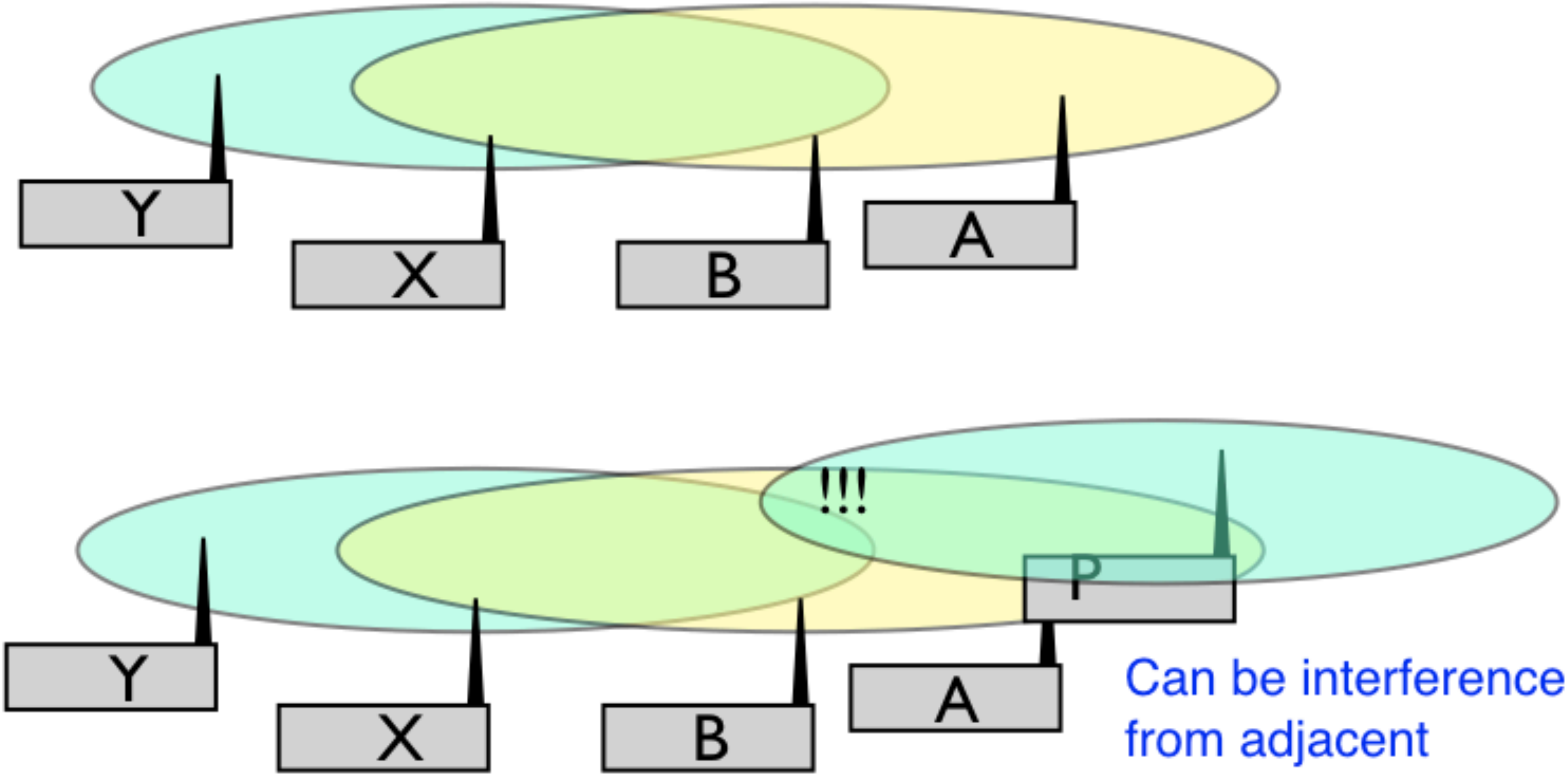
Initial 11 Mbps
Grew to 300 Mbps in a decade
Since 2011, looking at 1 Gbps at short distances ~ 10m
(rate reduces with distance at 100m or so, still only 11 Mbps)

Frequency Channel Re-use

The ISM* frequency band allows several WiFi channels

All systems using a basestation use the same channel

This forms a logical network



Can be interference from adjacent networks!

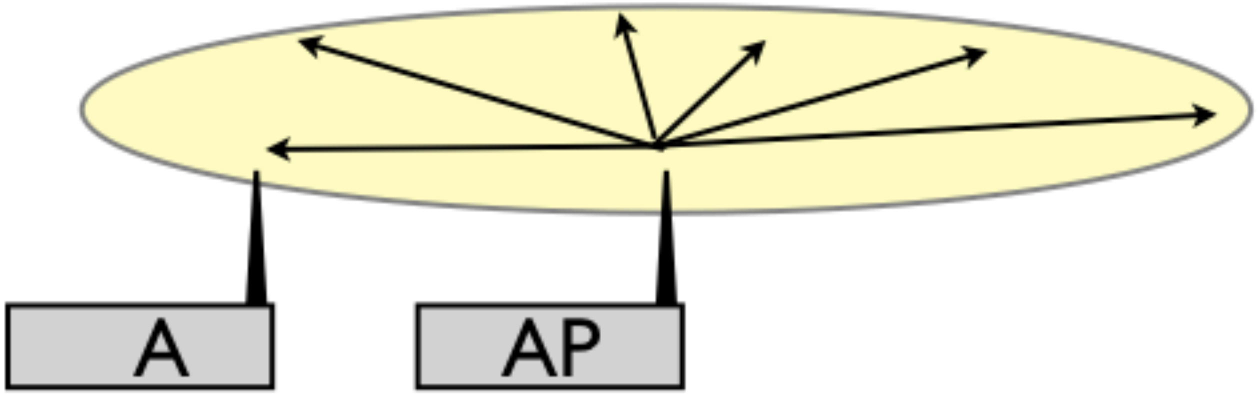
ISM - A frequency allocation for Industry, Science and Medical applications

Base Stations and Beacon Frame

How do you know which network you are using?

The WiFi access point (AP) broadcasts periodic beacon frames

can also identify the network (SSID*)



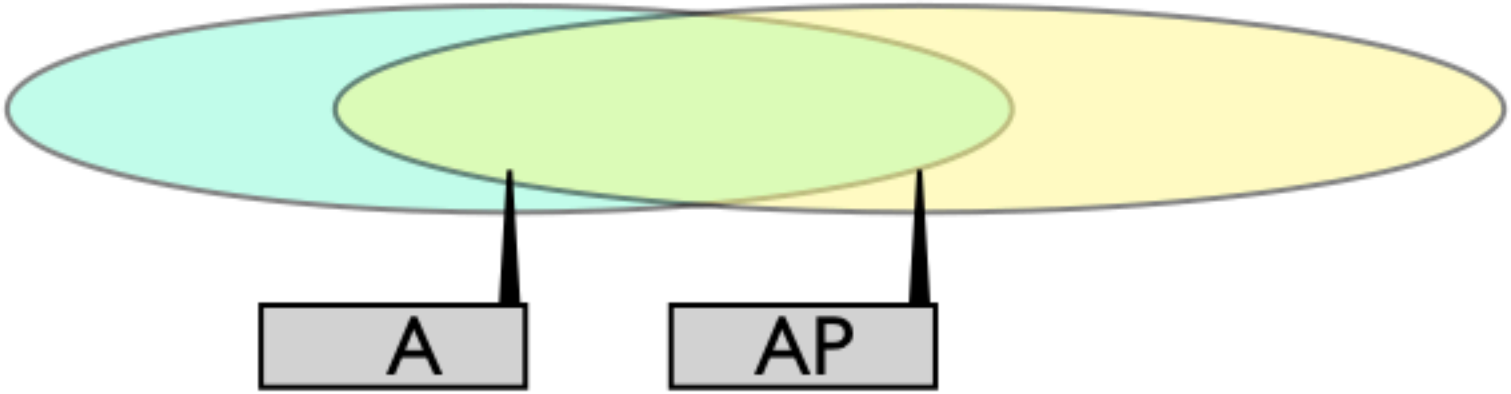
WiFi AP

The WiFi AP forms the logical centre of the WiFi network

SSID - Service Set Identifier, an Ethernet beacon frame

Beacon frames include the AP source MAC address

Beacon frames are sent to the broadcast MAC destination address

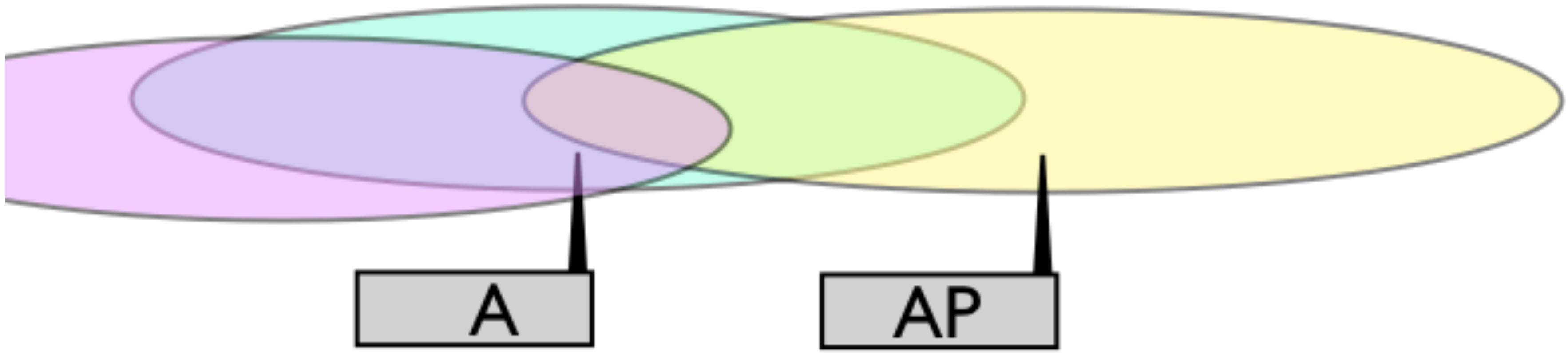


Each wireless node has a range
A is an end system; AP is a an access point

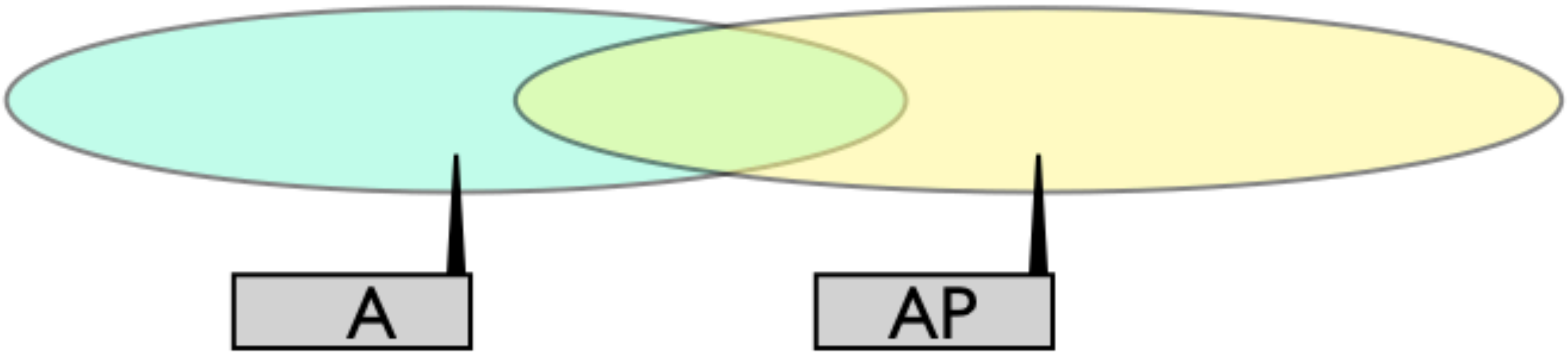
A needs to be able to receive signal from AP
(and AP from A)

When A sends to AP it can first sense the medium
(i.e. check if any system is sending)

Wireless (802.11)



A and AP can no longer communicate (interference)



A and AP can no longer communicate (signal strength)

Collision Avoidance

WiFi uses CSMA with ***Collision Avoidance***

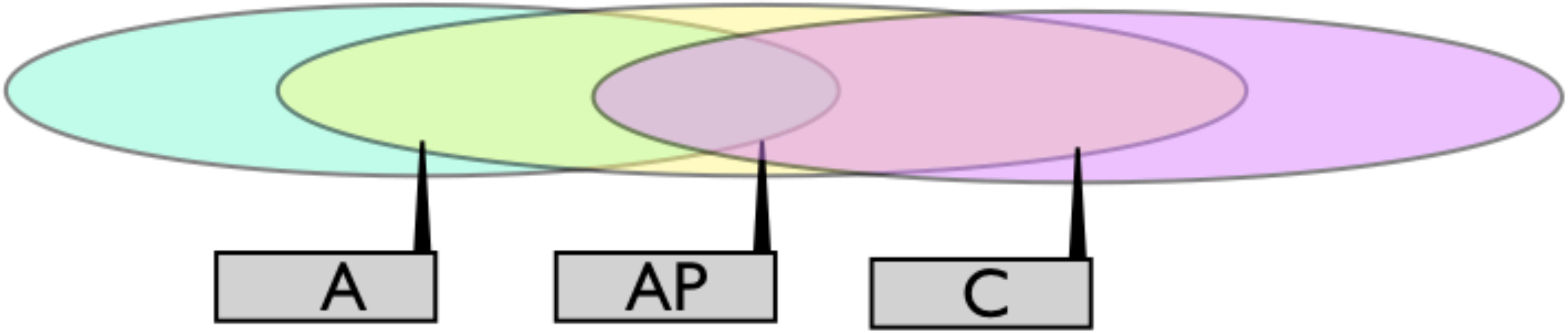
Three important changes:

1. A sender attempts to ***avoid*** causing a collision - it first listens to check the channel is idle (DCF Interframe Space). It then waits a randomly chosen time and if still idle, starts transmission.

2. A sender ***cannot monitor*** the entire wireless medium
   Receivers acknowledge (after a short delay) if a frame received.
   If no ACK is received within a timeout, the sender backs-off (as in CSMA/CD). Backoff increases with a limit of 5-7 attempts.

3. An optional procedure known as CTS/RTS detects hidden nodes.

Hidden Node Problem



Some nodes may not be able to “see” other transmissions

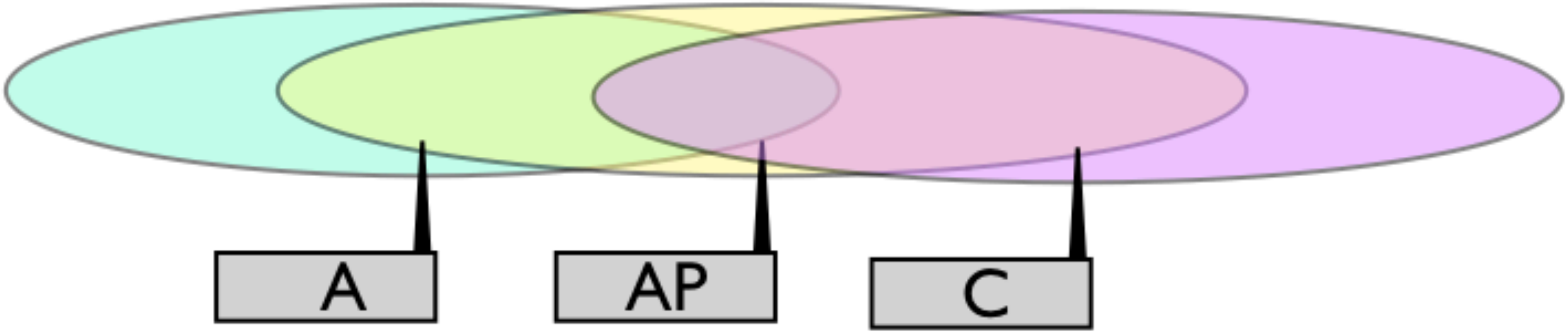
e.g.C does not know if A is sending

C may try to send to the AP (causing a collision)

Note 1: Wireless propagation can be very variable!

Note 2: By definition an AP sees signal from all nodes using AP

Virtual Carrier Detect

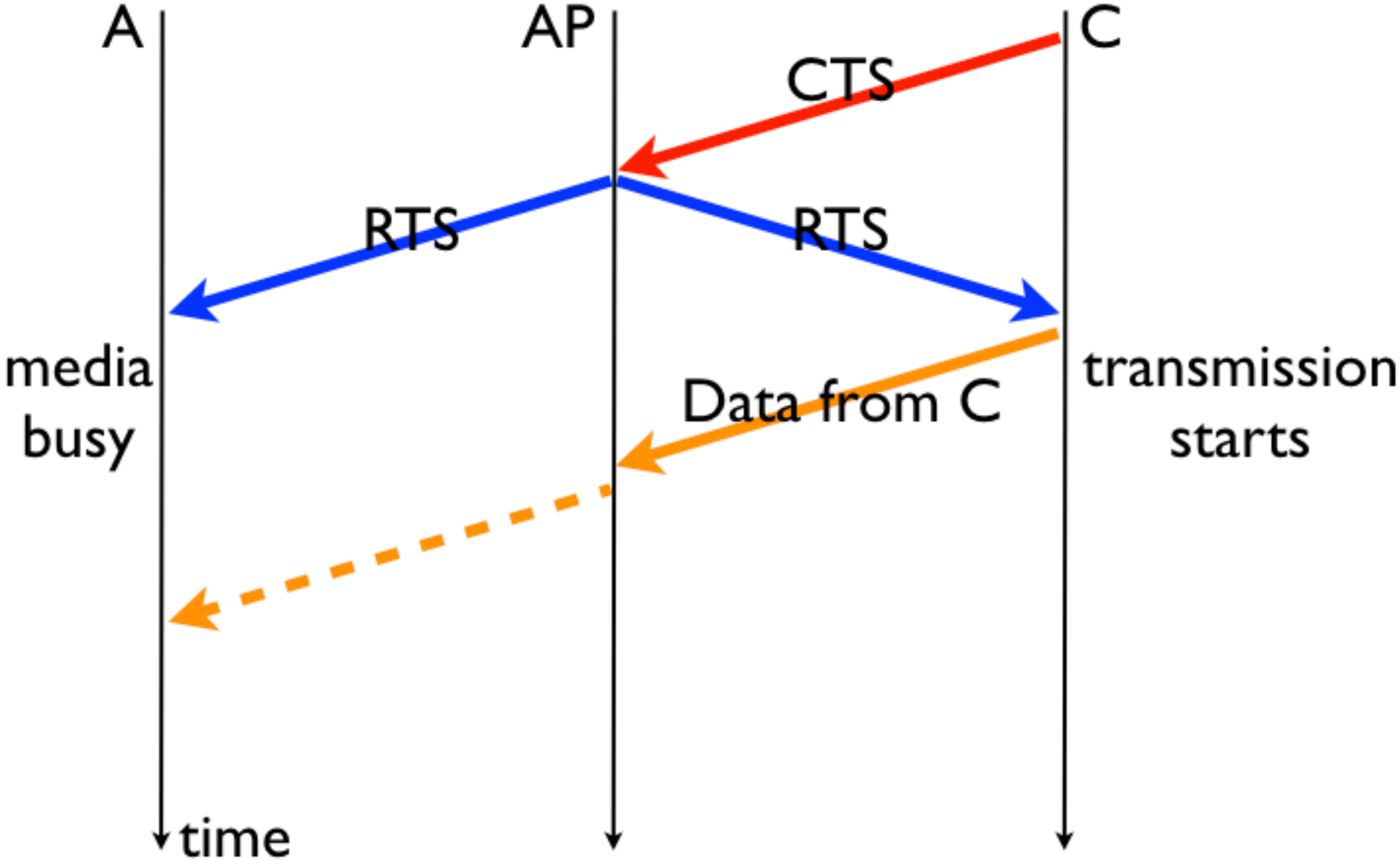


C first sends a ***Clear To Send*** frame to ask if it can transmit
- received by all nodes in range (i.e. Pink)

AP responds with an ***Ready To Send*** frame
- received by all nodes in range (i.e. Pink & Yellow)
both now know the "channel is in use"

When Ready To Send is not received
sender must defer ("back-off") before repeating Clear To Send

Hidden Node Problem and CTS/RTS



Note: If C needs to talk to A, it would rely on AP to relay (or repeat) the signal so that A can receive it.

Roaming

Several access points (APs) may form a LAN

APs connected together via a cabled LAN

Roaming between access points

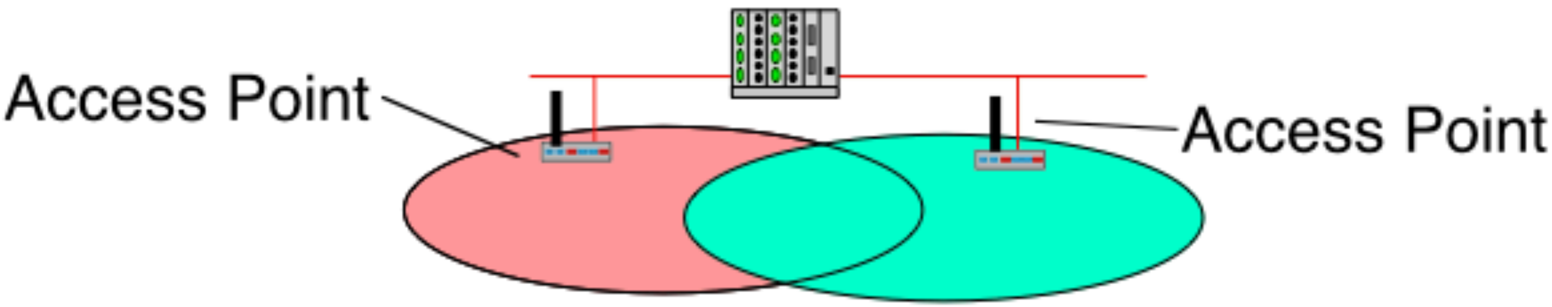
All APs sends a “beacon” signal to all nodes (SSID)

Multiple APs can advertise the same SSID

Nodes can select the AP with the best “beacon” signal

Wireless nodes keep the same MAC address

- users do not need to know the AP has changed!!



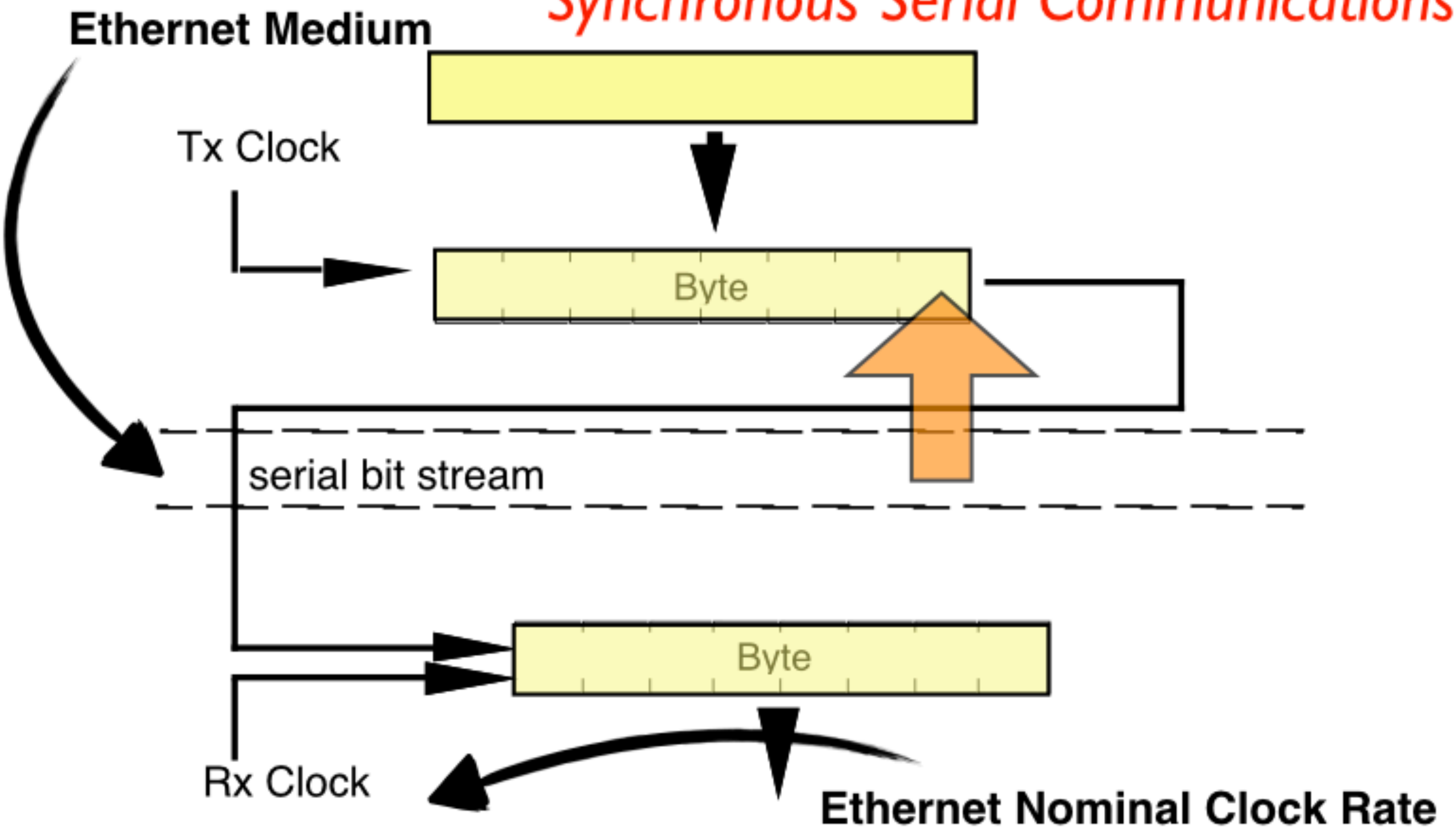
Ethernet Frames:

Sending Frames (Ethernet Transmit)

The Physical Layer

Module 3.1

Synchronous Serial Communications

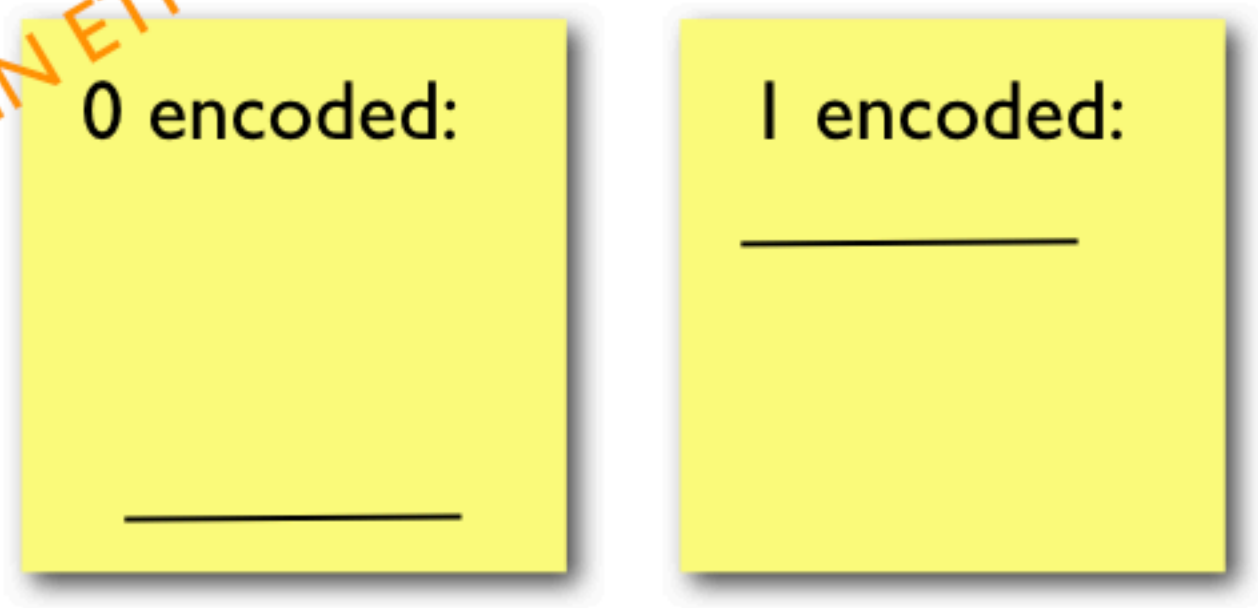


Uses two shift registers (both clocks must be the same rate)
- Note that bytes are sent l.s.b. first!

Recall the Ethernet broadcast/unicast address bit?

Non Return to Zero

NOT USED IN ETHERNET!



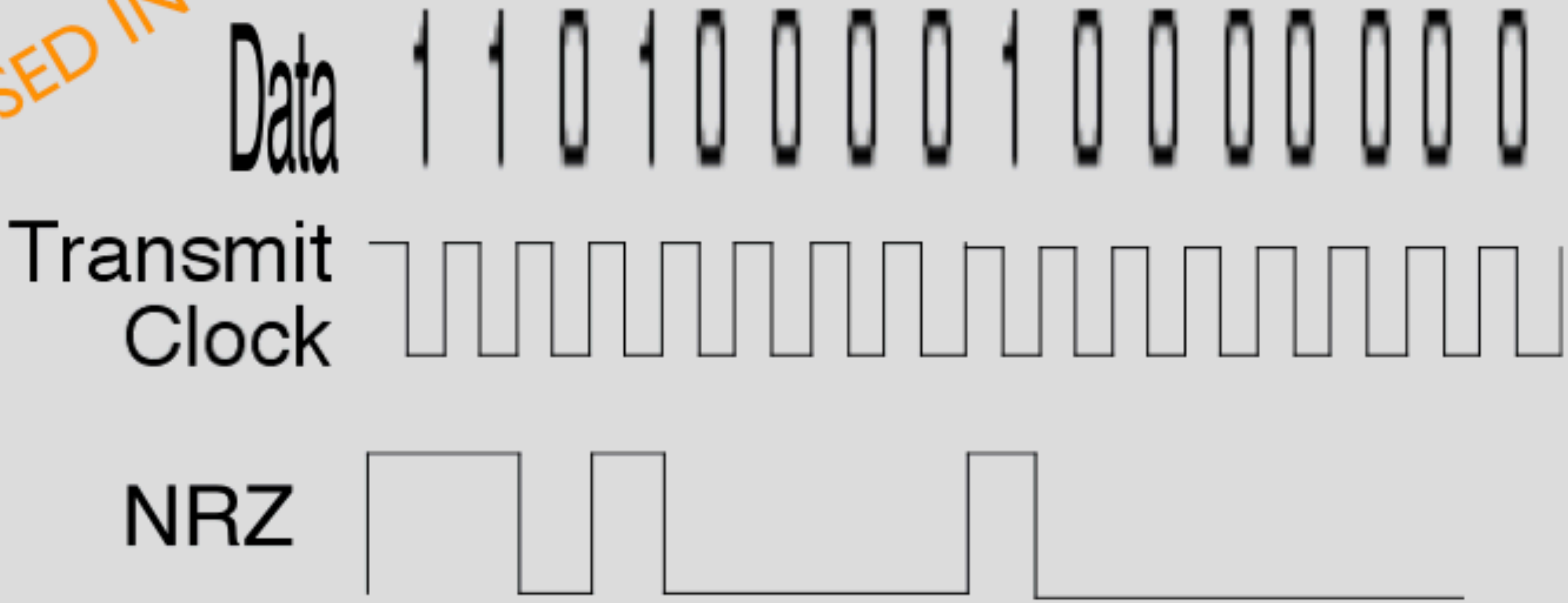
2 signal levels are used

The level of a baud indicates the value of each bit
- a low level indicates 0
- a high level indicates 1

The bandwidth of NRZ is approx 1 Hz / bit

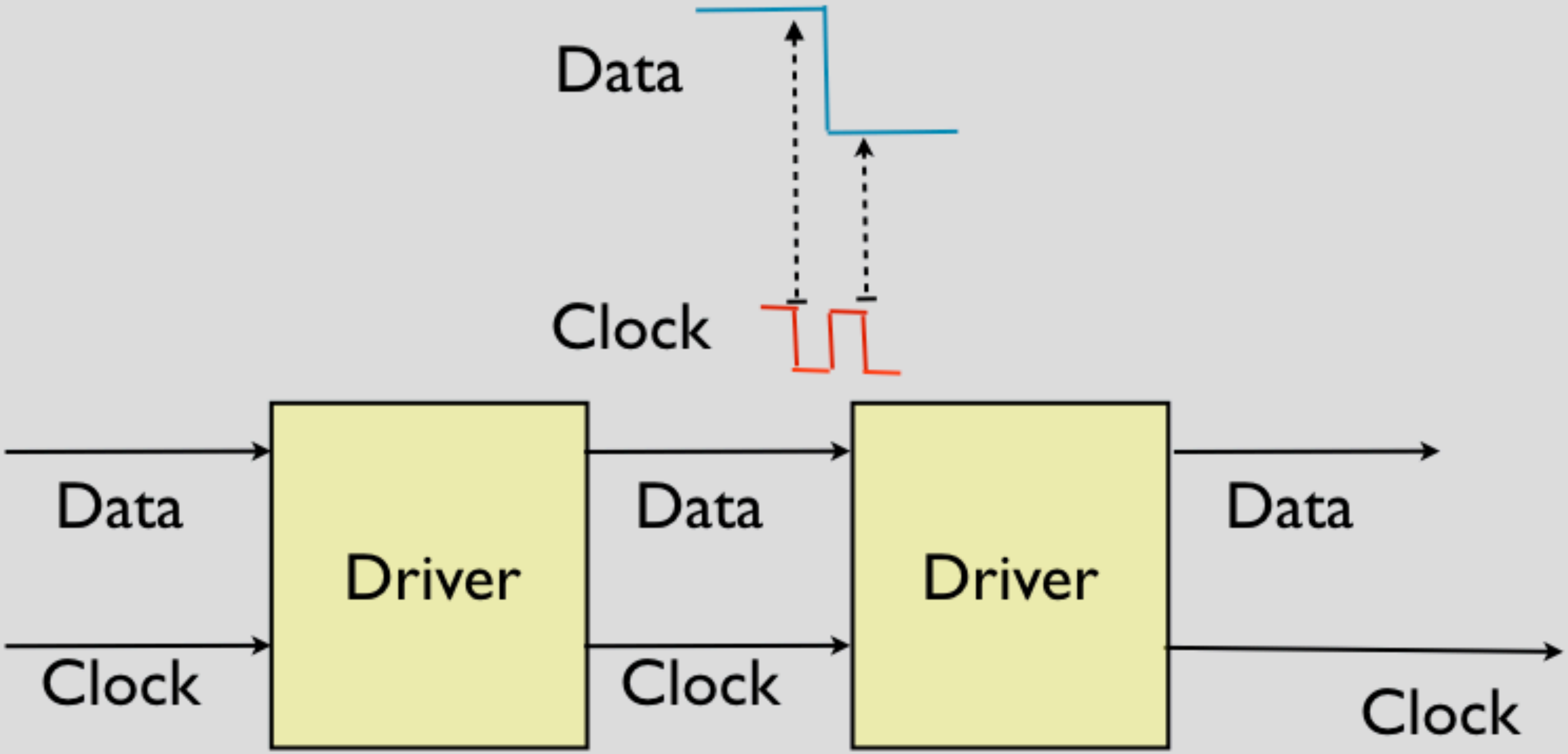
Non Return to Zero

NOT USED IN ETHERNET!



The receiver needs some way of determining the clock transitions …
i.e. you can not just look at a NRZ encoded waveform to determine the sequence of 1 and 0 bits that it represents - you need to look at clock & data!

Traditional Synchronous Transmission

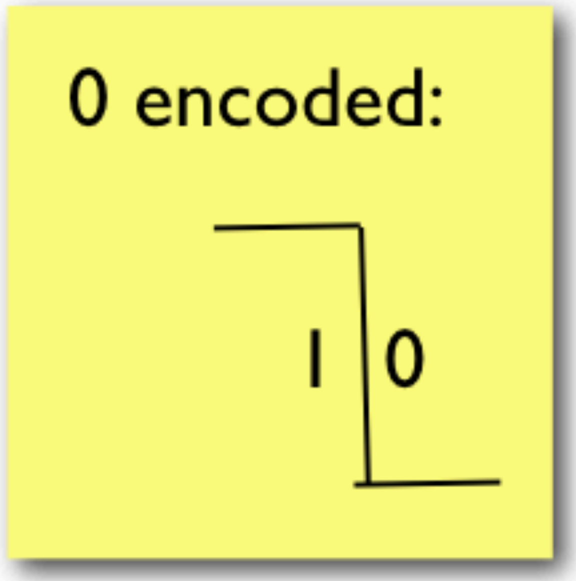


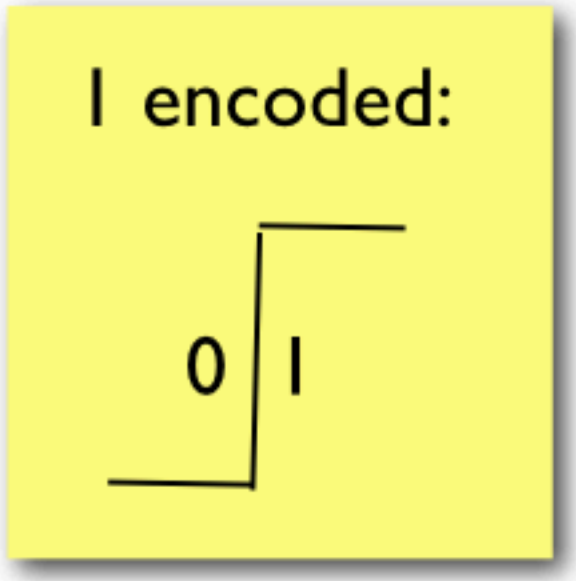
Clock signal transitions indicate centre of each bit

- Sender uses clock to time sending each bit
- Receiver uses the clock to detect the centre of each bit

Requires <u>two</u> sets of wires (clock & data + ground)

Manchester Encoding



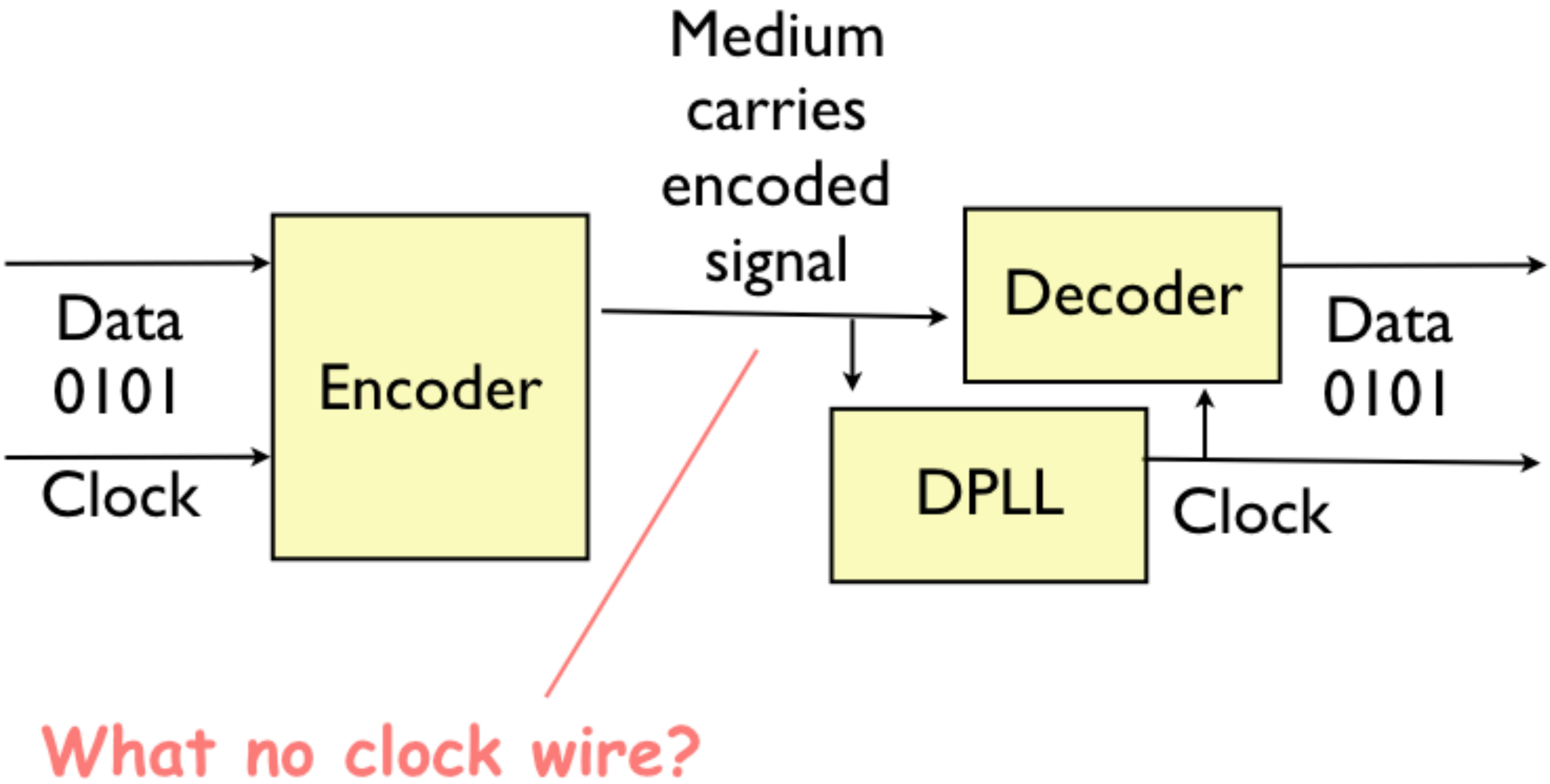


2 signal levels used
There is a transition in the centre of each bit
- a down-wards transition from a 1 baud to a 0 baud indicates a 0 bit
  - an up-wards transition indicates a 1 bit

The 2 bauds use double the cable bandwidth compared to NRZ!*

* 10B2/10B5 use high bandwidth RF cable, so this is not an issue.

Encoded Data

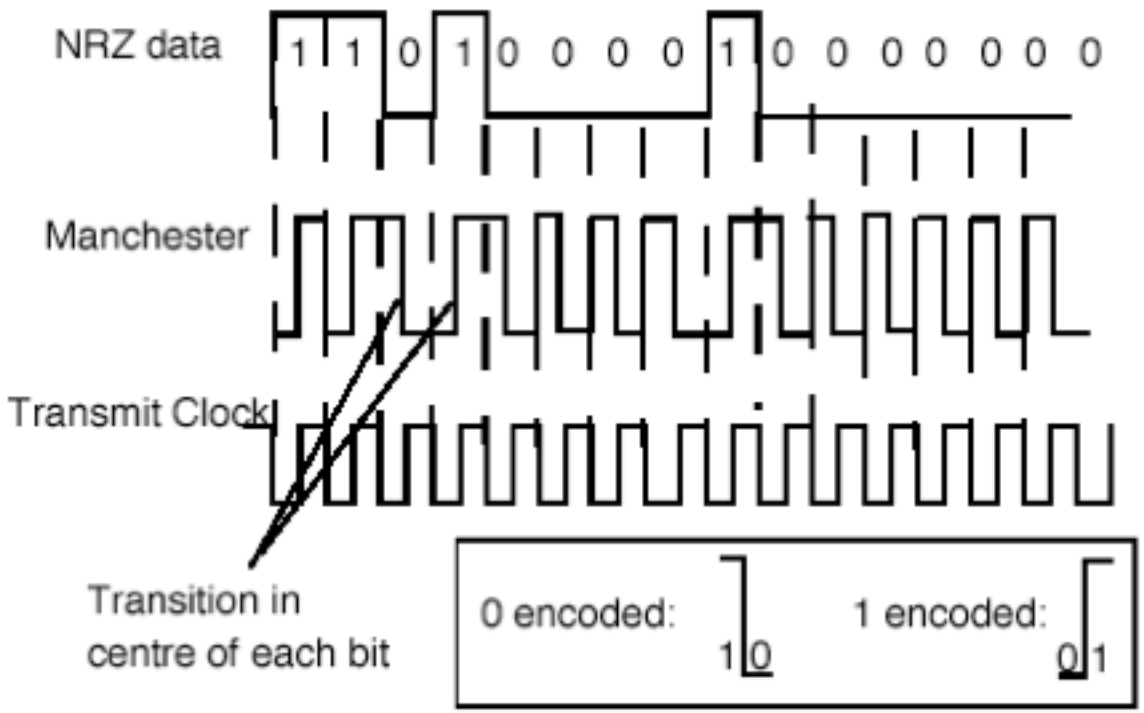


The sender encodes the clock and data as a waveform

- The cable transmits this combined clock & data signal as pairs of bauds
- This needs only one "wire"

At the receiver, a Digital Phase Locked Loop (DPLL) regenerates clock

Manchester Encoding

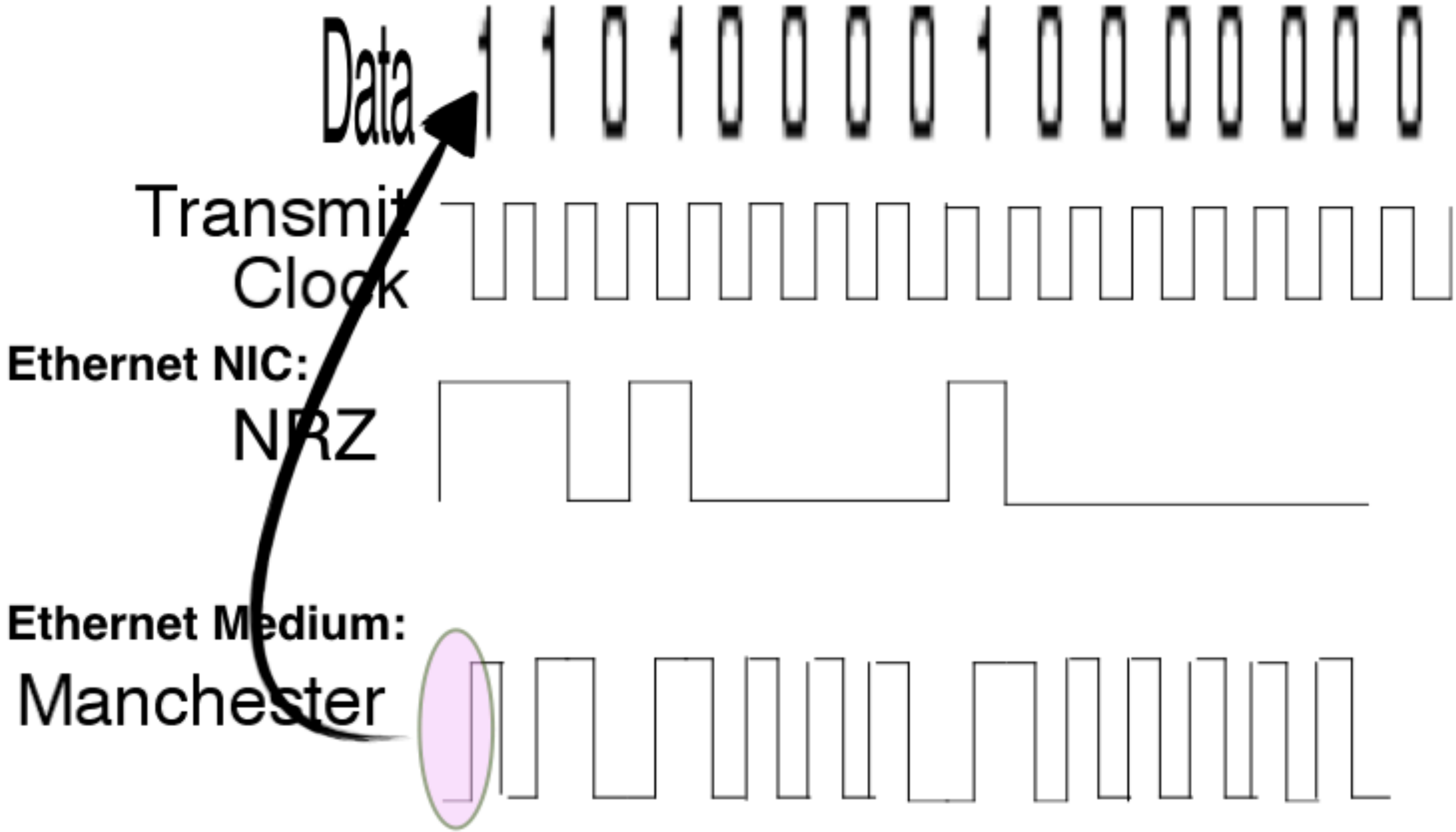


Looking at the waveform it is clear there is:

No DC component (even for long runs of 0‘s or 1’s)

A timing component at the fundamental clock frequency (10 MHz)

Manchester Encoded Signal

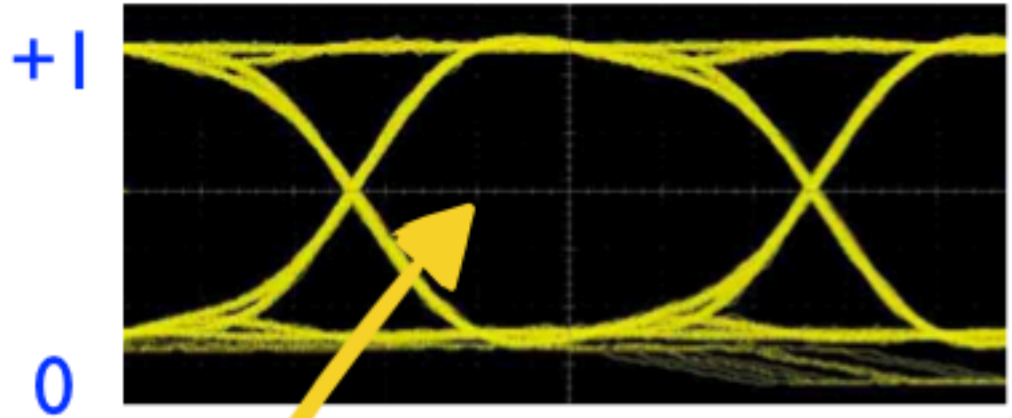


Eye Diagram for Manchester Encoded Signal

Oscilloscope plot using an eye diagram

The eye diagram plots voltage v. time

With a timebase trigger for ***multiple scans*** through the waveform



<—1 bit —>

Two distinct levels are clear

Noise is evident causing blurring in the vertical axis

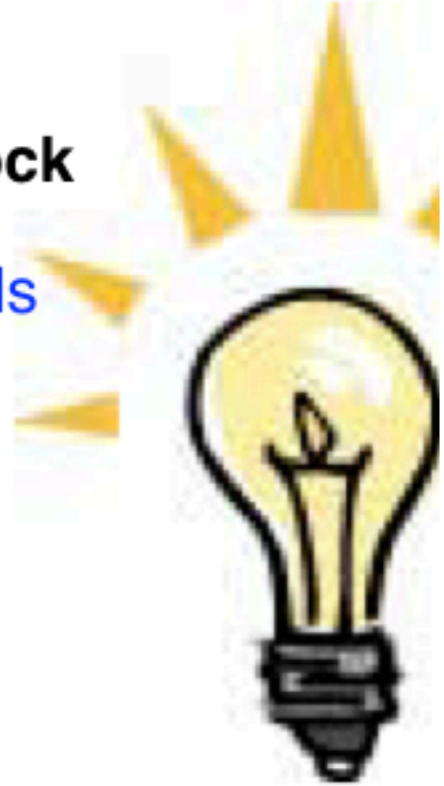
The slew rate is limited, i.e. the rise time for transitions

Transitions in level only occur at the edge of bits

Transitions never occur in the centre of the display!

Summary

- **Manchester Encoding**
  - Encodes ***each*** data bit as a pair of bauds
  - No net DC signal
  - Uses double the baud rate
  - Embedded clock
- **A DPLL is used at the receiver to decode the clock**
  - This aligns the local clock with the received bauds
- **Data is decoded**
  - 2 bauds are read to decode ***each*** data bit



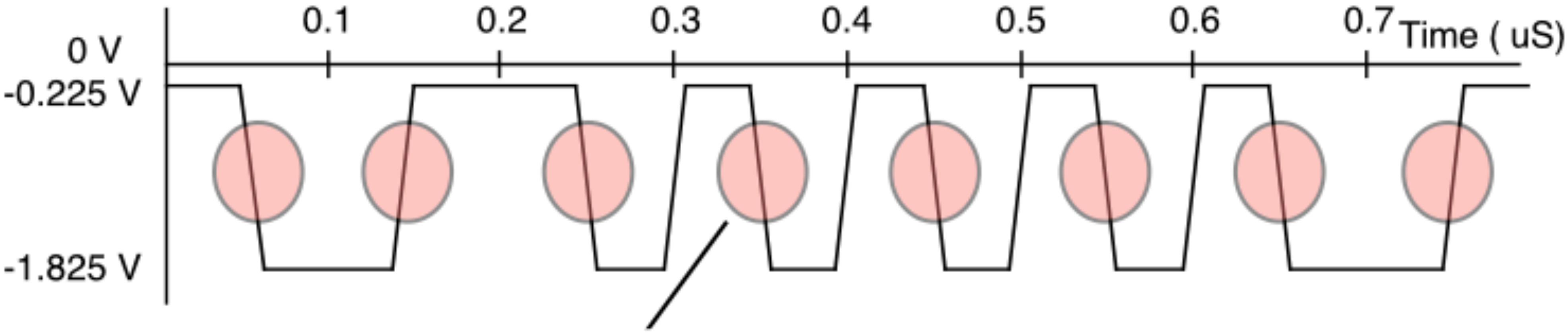
Ethernet Frames:

Receiving data (Ethernet Receive)

The Physical Layer

Module 3.2

Ethernet Waveform

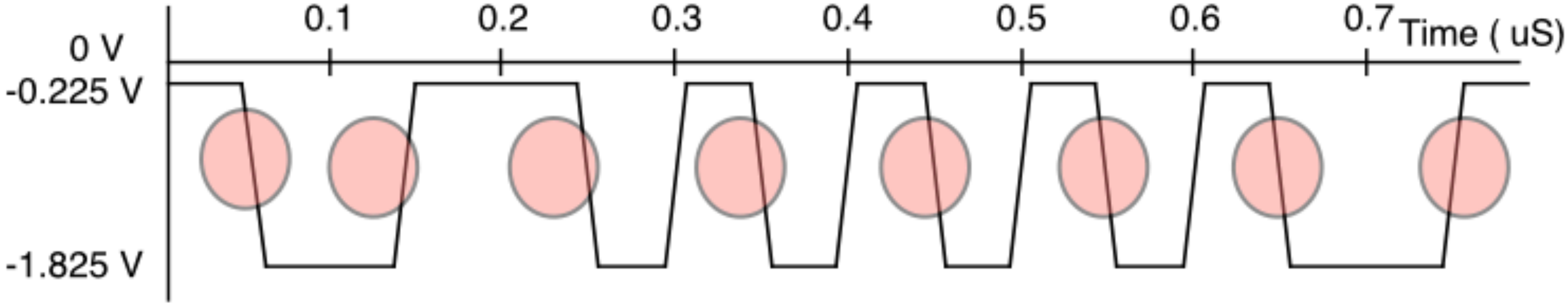


Can ***you*** decode this?

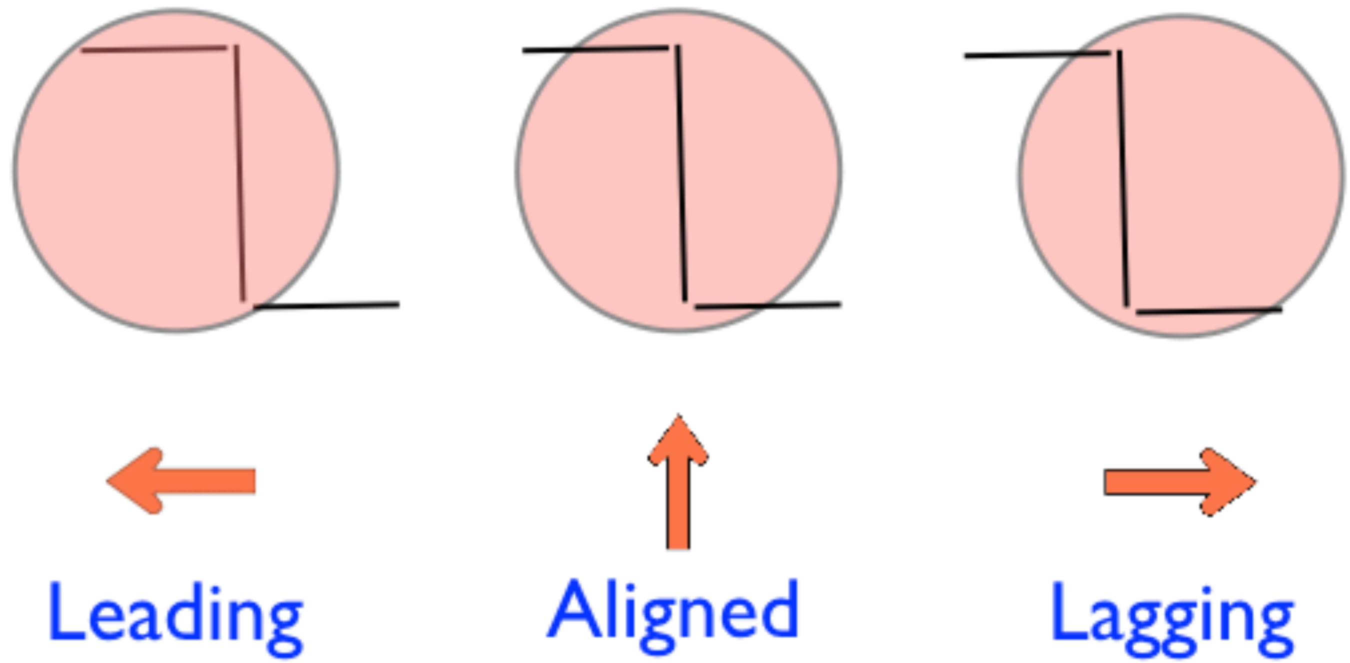
Transitions
at centre of bits

The signal isn't referenced to zero volts
Rise time ~ 25nS
The waveform as seen on an oscilloscope may be inverted!

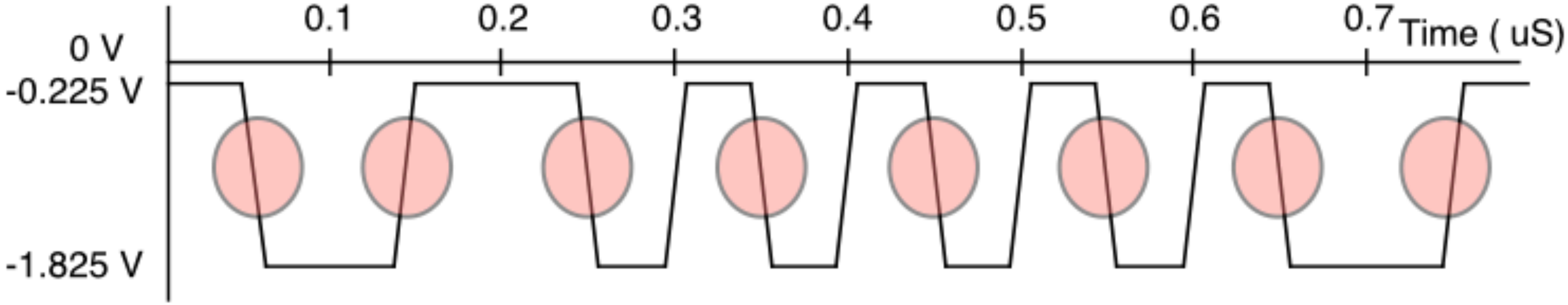
Sampling the Received Waveform

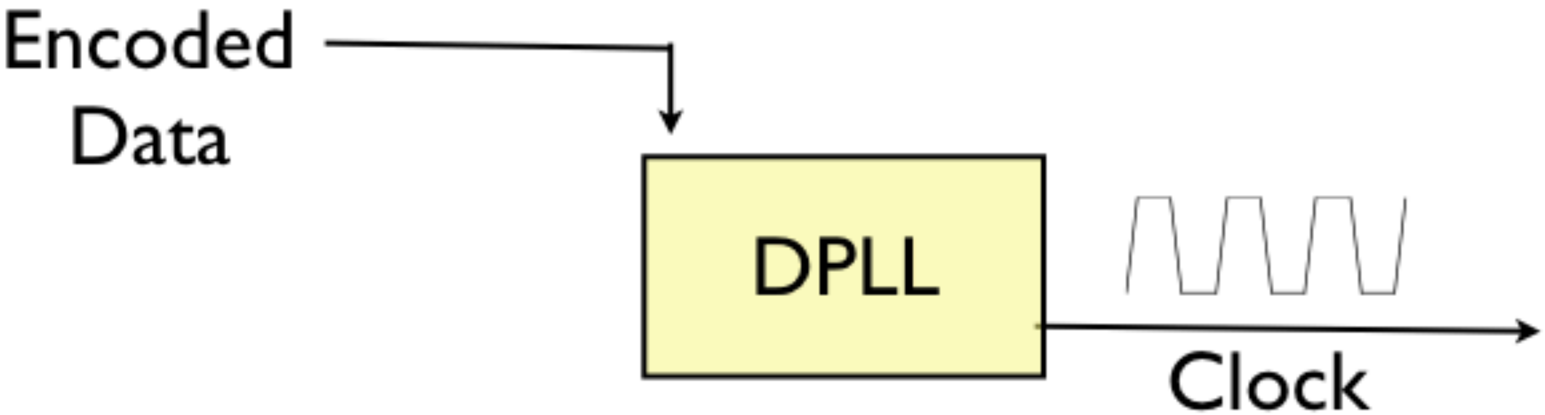


If we sample pairs of bauds, the waveform at receiver might result in one of three cases:



Ethernet Clock Recovery





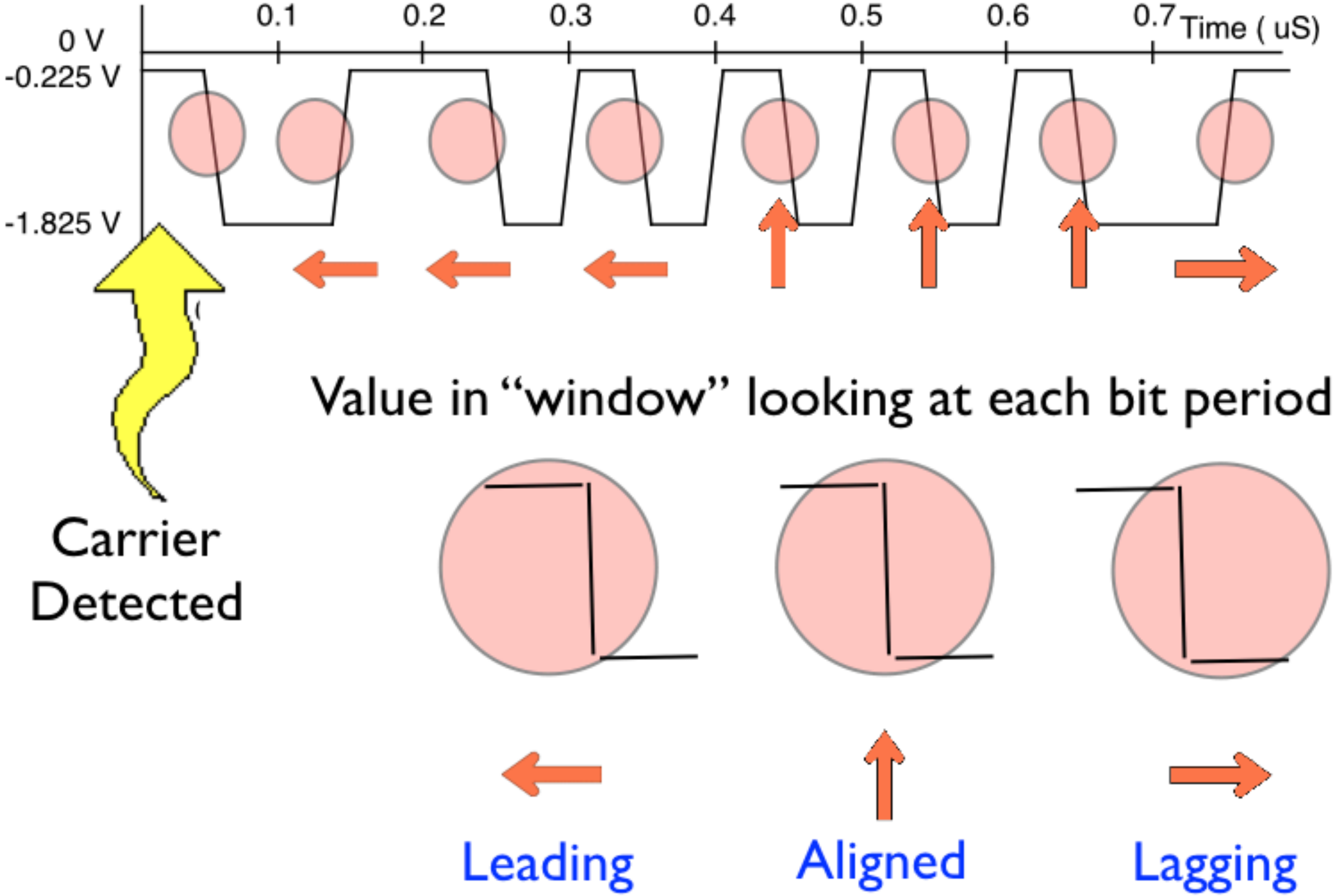
DPLL contains a clock (oscillator)

Uses phase transitions to lock the local receive oscillator frequency

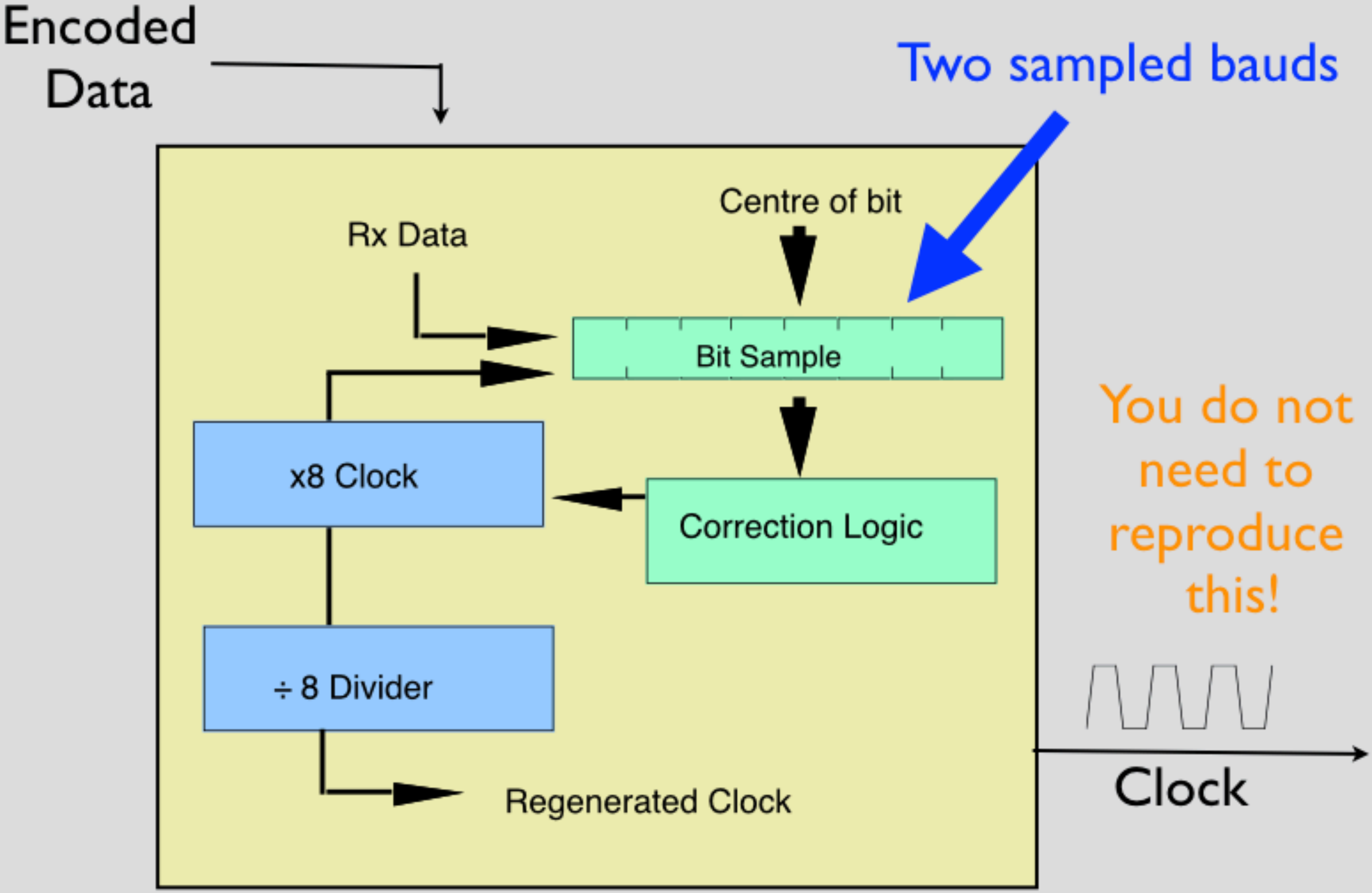
- If a transition is *lagging*, *decrease the clock* period (increase frequency)
- If *leading*, *increase the clock period* (decrease the frequency)

After many transitions, the recovered clock ***matches the encoded data***

Ethernet Clock Recovery by DPLL



Digital Phase-Locked Loop (DPLL)



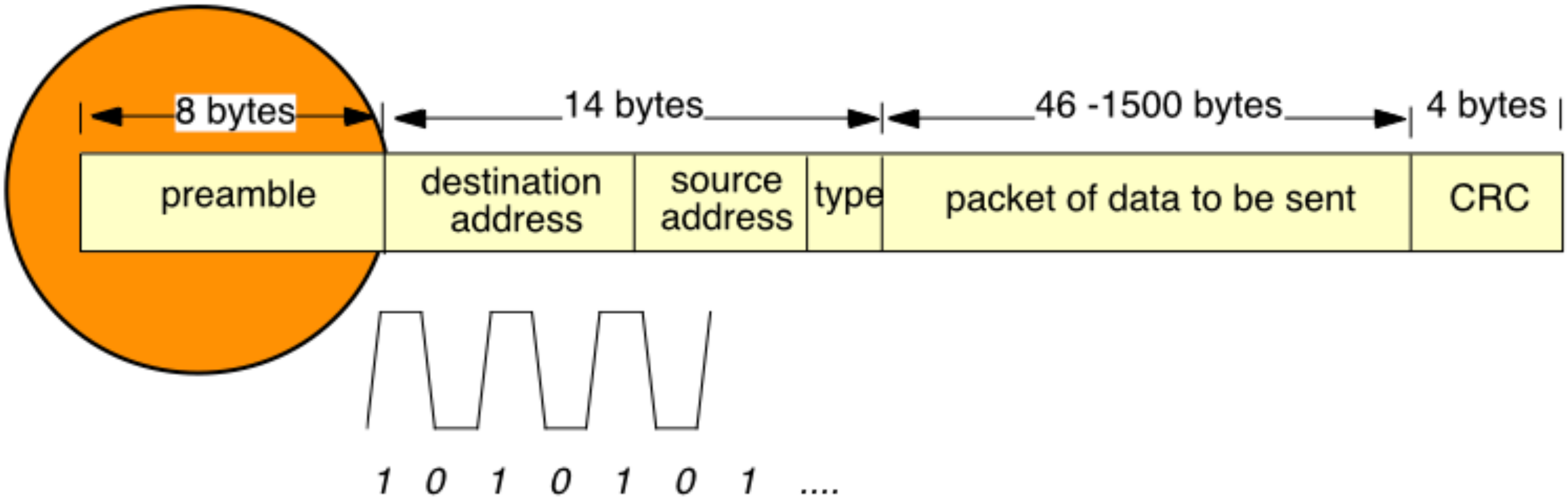
Preamble Sequence

Each frame starts with a fixed-format preamble

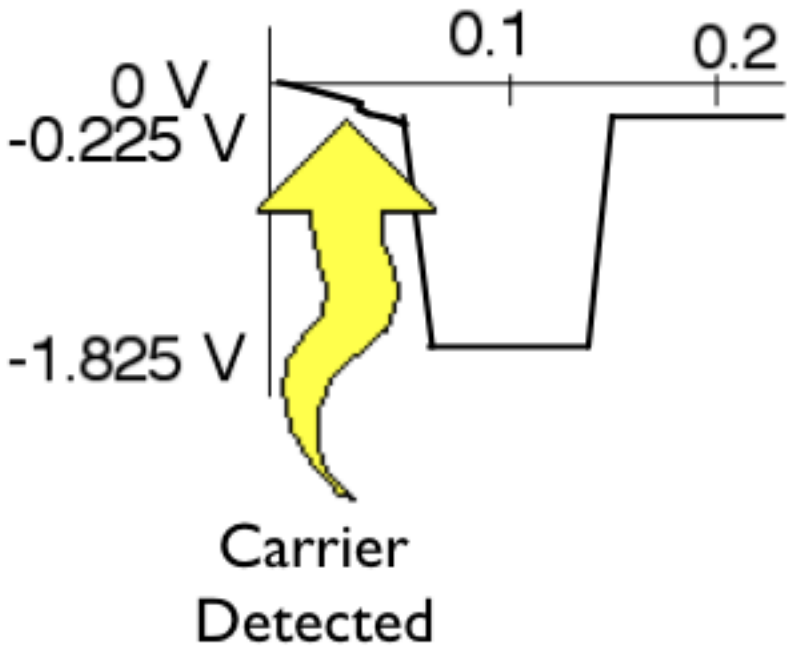
This has two functions:

(1) The format is chosen to help assist the DPLL achieve lock
This means the preamble uses an alternating '0' and '1' bit pattern

(2) The preamble is used to detect the start of frame delimiter (SFD)
The final 2 bits of the last byte (SFD) are set to '11'
This reveals the encoding rule for a '1'



Ethernet Inter-Frame Gap / Spacing



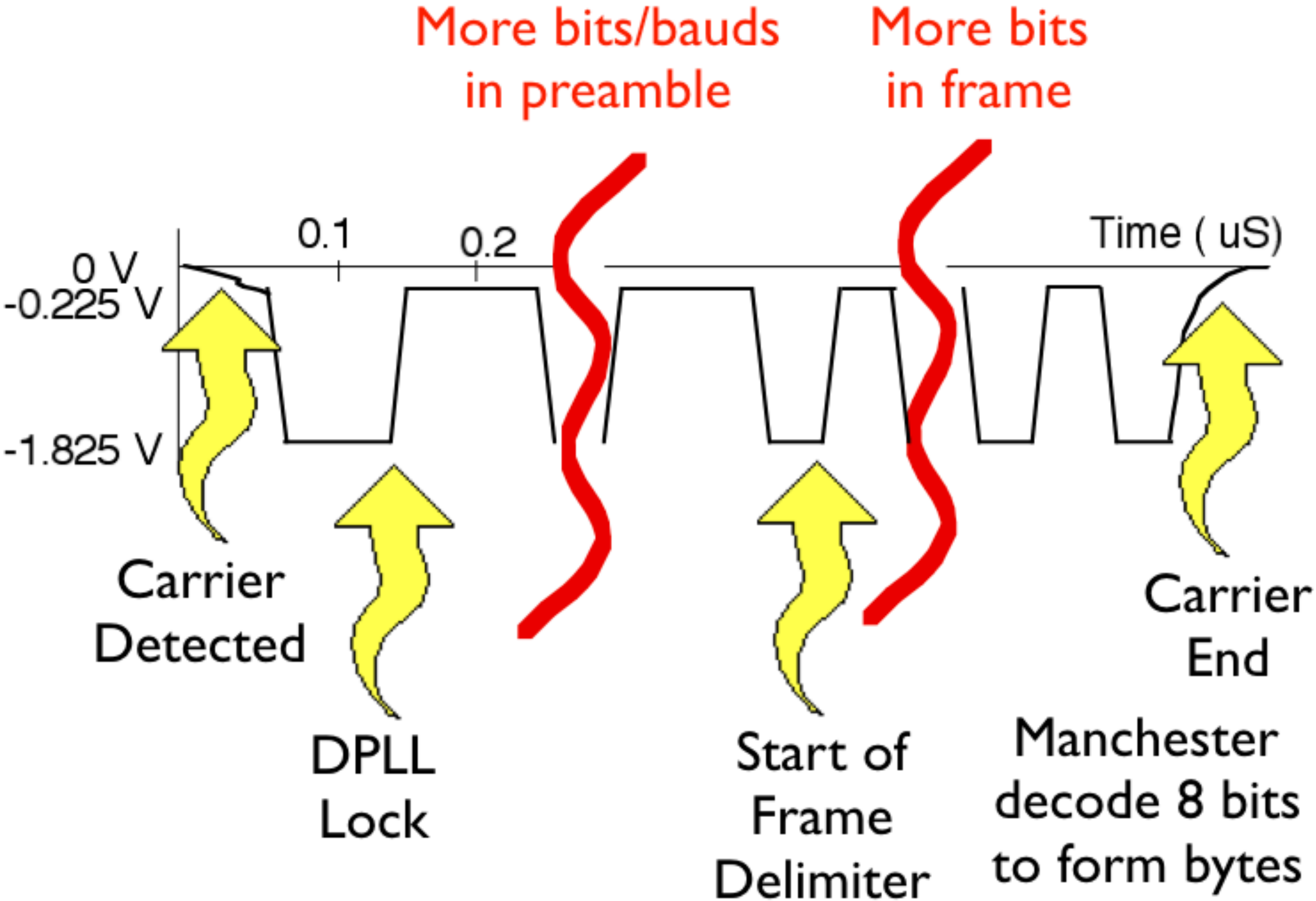
A silent time between frames (no carrier on medium)

Allows transceiver electronics to recover after end of previous frame

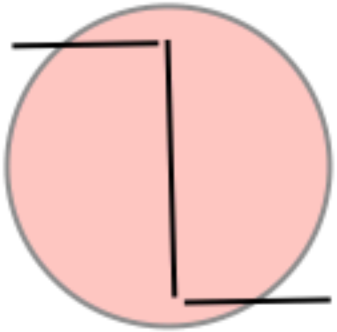
20 byte periods (measured from end to next SFD)

- 10 Mbps: > 9.6 microsecs between frames (at receiver)
- (some descriptions say 10.4 microsecs at sender)

Ethernet Frame



Resolving ambiguity in the Received Polarity



Is this a '0' or a '1'
Is the waveform inverted?

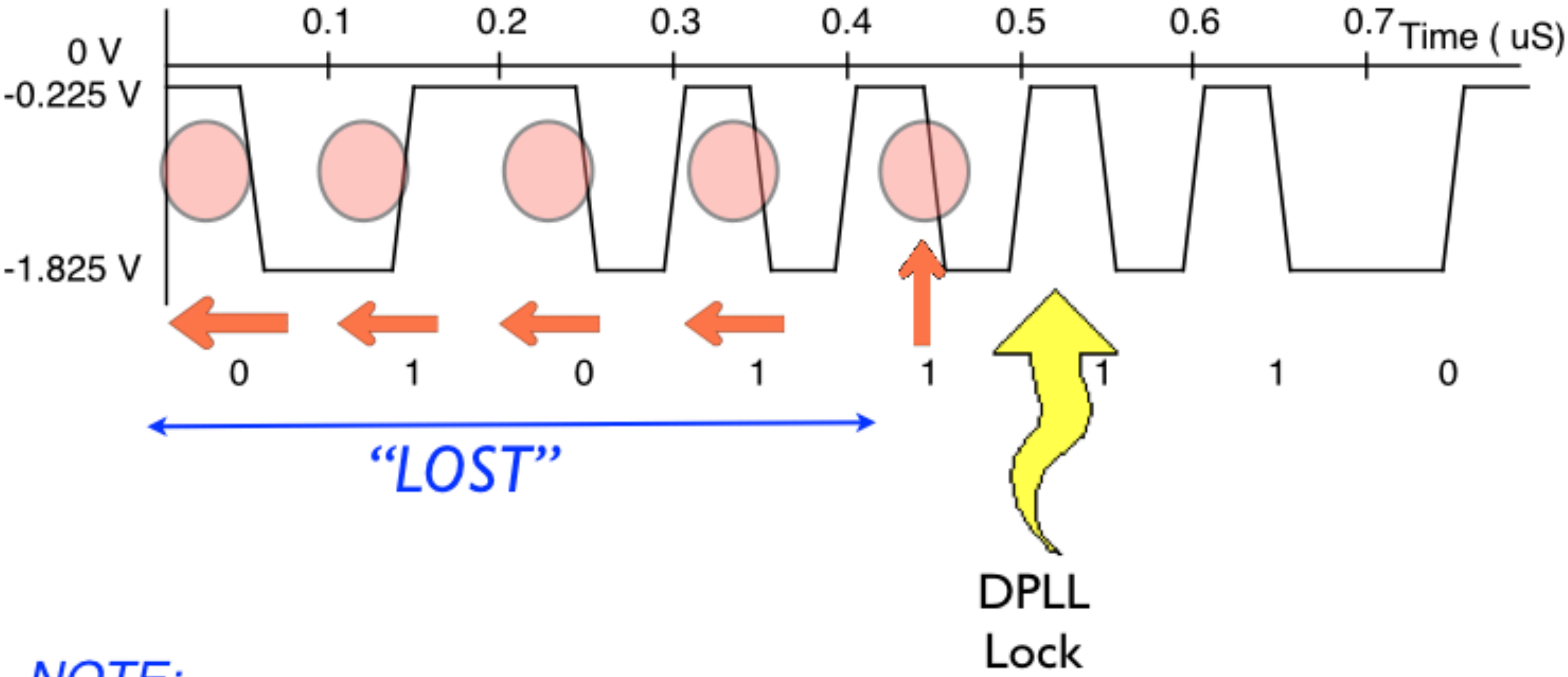
A waveform can be inverted ...

Ethernet has a trick that allows a received to discovery the polarity of the received signal bauds...

Recall that the SFD ends with the sequence '11'

When the decoder sees the end of the preamble it can unambiguously discover the pair of Manchester-encoded bauds used for a '1' bit.

Loss of the Start of the Preamble!!

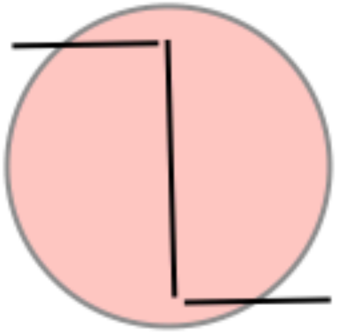


NOTE:

(1) Each sender will have a slightly different clock signal
A receiver therefore has to ***retrain*** *the DPLL to each new sender*

(2) Bauds received before the DPLL has lock may not be decoded
Not all bauds of the preamble are "therefore received" by the decoder

Summary: Four Steps to Reception



4 steps required to decode each frame

1) The start of a frame needs to be detected using the CS circuit

2) A clock signal is recovered at the receiver (using a DPLL)

3) The polarity and start of the data is determined from the SFD

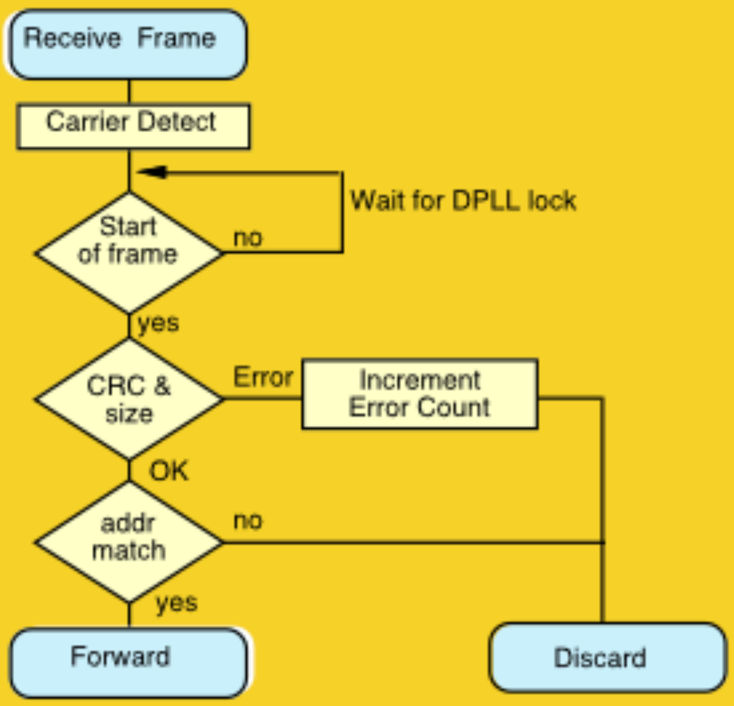
4) The end each frame is detected using the CS circuit

Summary

- **There is an Inter Frame Gap (IFG) between each frame**
- **All Ethernet frames have a preamble**
  - 62 bits have the pattern 10
  - The first baud triggers the carrier detect circuit to start listening
  - Remainder of the preamble helps gain DPLL lock (takes time)
  - Not all preamble bits are "received" by the decoder
- **End of preamble marked by the SFD**
  - Polarity detected by the 2 SFD bits, with value 11
- **The final bit of the frame is detected by absence of a carrier**
  - A CRC-32 is used to verify this process

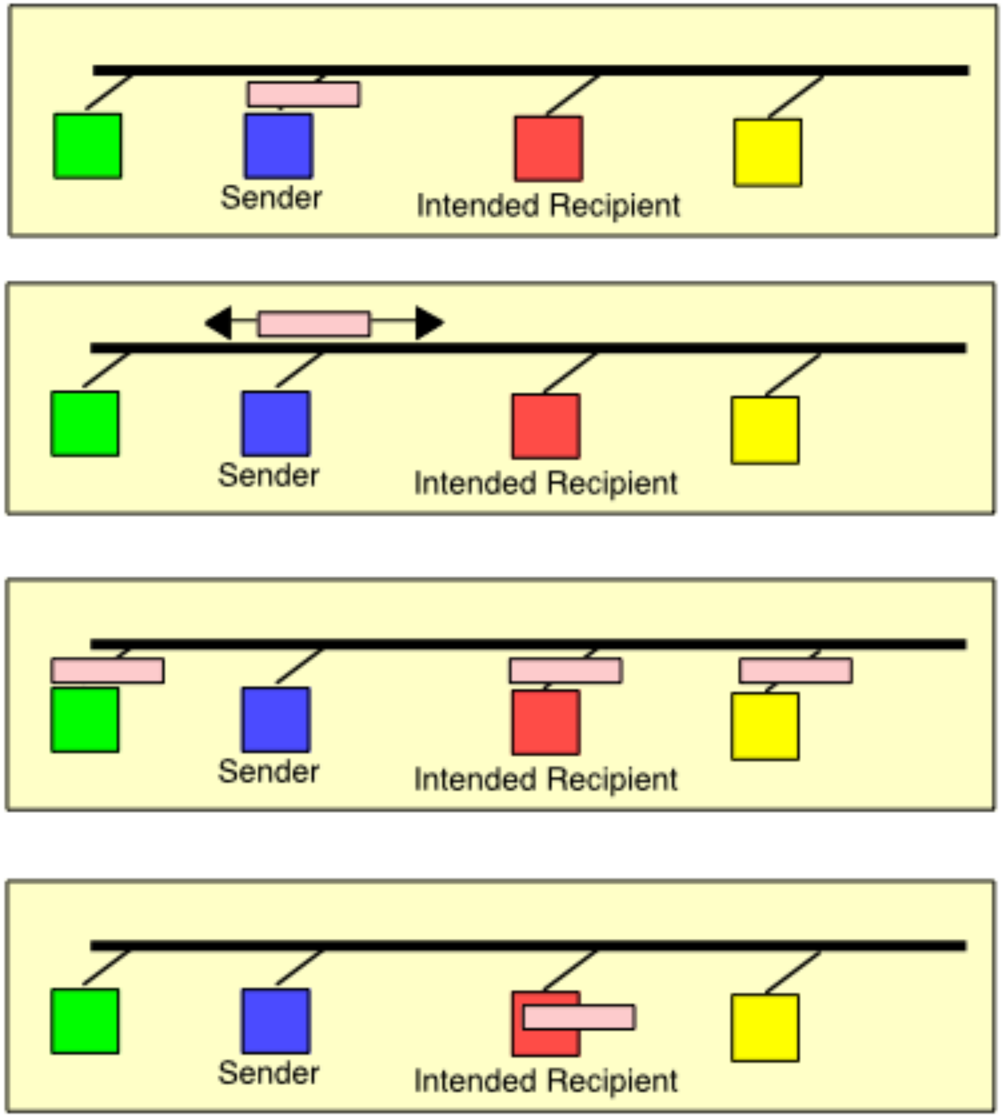
Ethernet Frames:

Frame Reception

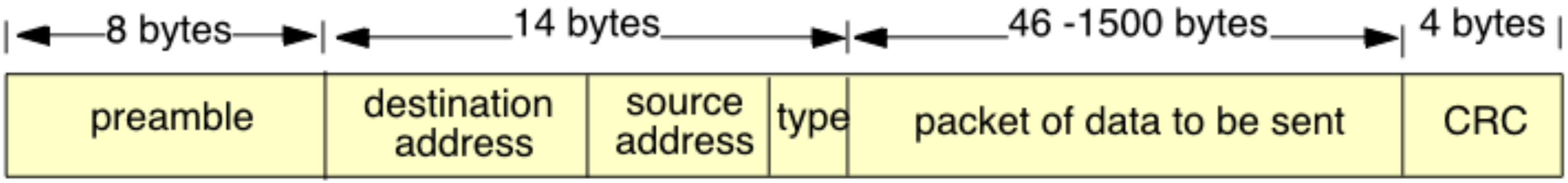


Module 3.3

LAN (MAC) address



Cyclic Redundancy Check (CRC)



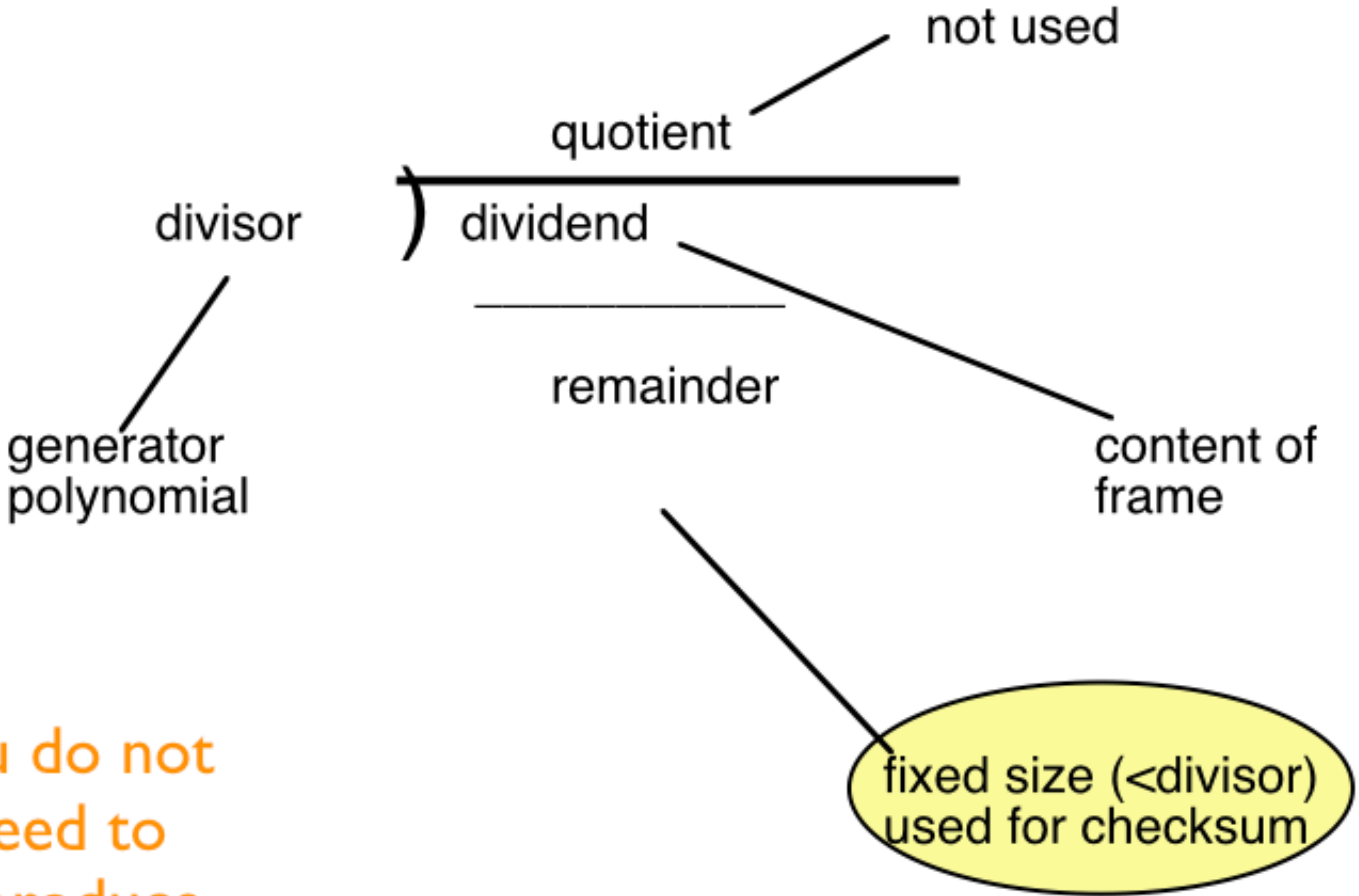
CRC-32 is a form of digital signature (32-bit hash of frame)

- Calculated at the sender & sent at end of each frame
- Re-calculated at the receiver
- Sent value is compared with received value at receiver
- This verifies the integrity of the data in the frame

The CRC-32 has a high probability of detecting:

- Any frames corrupted in transmission
- Frames where the DPLL failed to track the clock

Division



You do not need to reproduce this!

Why Modulo 2 Division?

Because the hardware solution is simple!!!!!

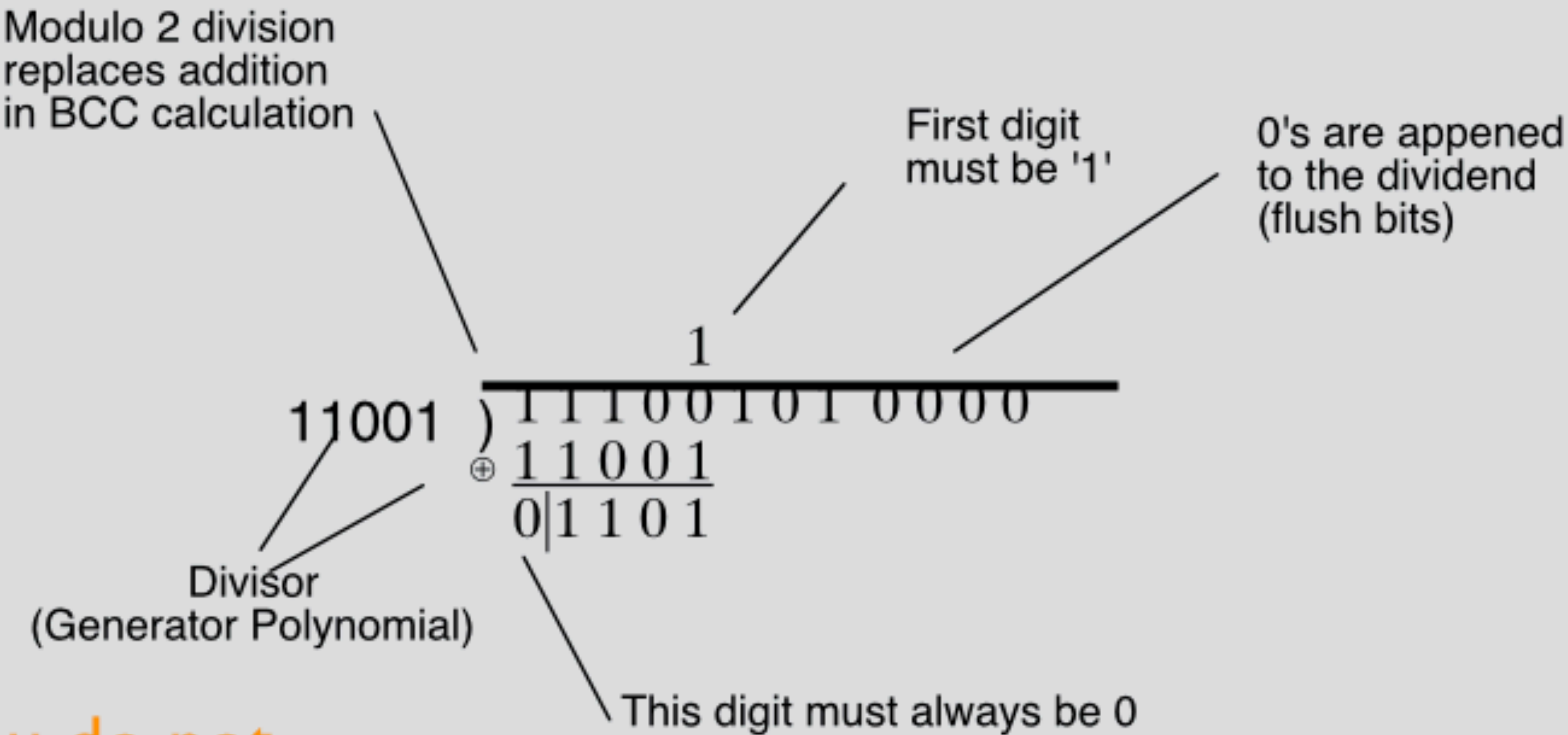
Truth Table for Modulo-2 Division (XOR)

$$0 \oplus 0 = 0$$
$$0 \oplus 1 = 1$$
$$1 \oplus 0 = 1$$
$$1 \oplus 1 = 0$$

You do not need to reproduce this!

- CRC calculations ignore the carry

Modulo 2 Division



You do not need to reproduce this!

Example simplified to generate a short (4 bit) CRC

Revise your notes from Level 3 course!!!

Modulo Division

```
                    1 0
11001 ) 1 1 1 0 0 1 0 1 0 0 0 0
      ⊕ 1 1 0 0 1|  ..
        0|0 1 0 1 1
        ⊕ 0 0 0 0 0
           0|1 0 1 1
```

1. Bring next digit of dividend down
2. Copy msb of value to quotient
3. Insert 0 (if quotient 0) or divisor (if quotient 1)
4. Calculate XOR sum
5. Discard msb of value (always 0)

You do not need to reproduce this!

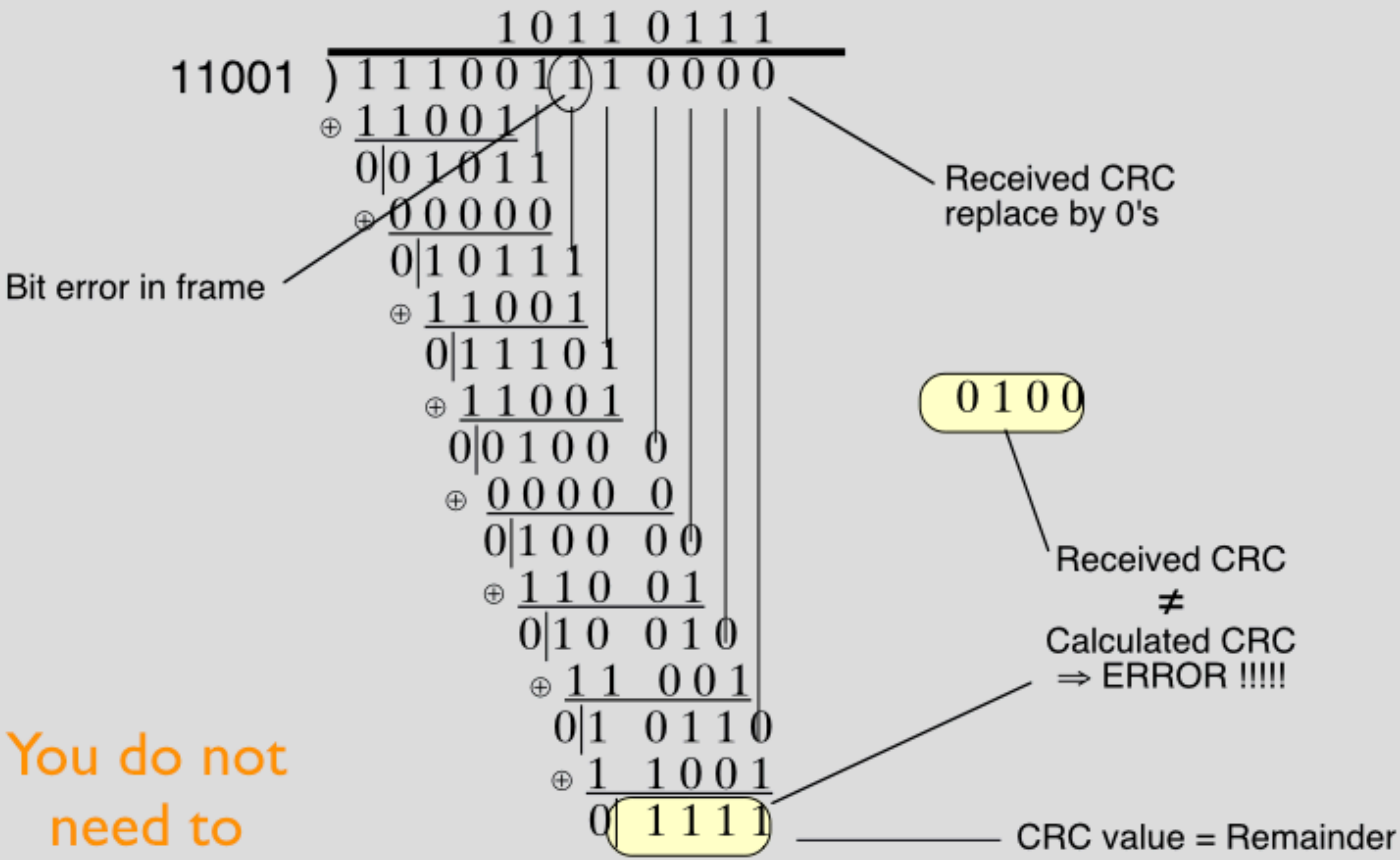
CRC Value

```
                 1 0 1 1 0 1 0 0
11001 ) 1 1 1 0 0 1 0 1 0 0 0 0
      ⊕ 1 1 0 0 1
        0|0 1 0 1 1
        ⊕ 0 0 0 0 0
          0|1 0 1 1 0
          ⊕ 1 1 0 0 1
            0|1 1 1 1 1
            ⊕ 1 1 0 0 1
              0|0 1 1 0 0
              ⊕ 0 0 0 0 0
                0|1 1 0 0 0
                ⊕ 1 1 0 0 1
                  0|0 0 0 1 0
                  ⊕ 0 0 0 0 0
                    0|0 0 1 0 0
                    ⊕ 0 0 0 0 0
                      0|0 1 0 0
```

CRC value = Remainder

You do not need to reproduce this!

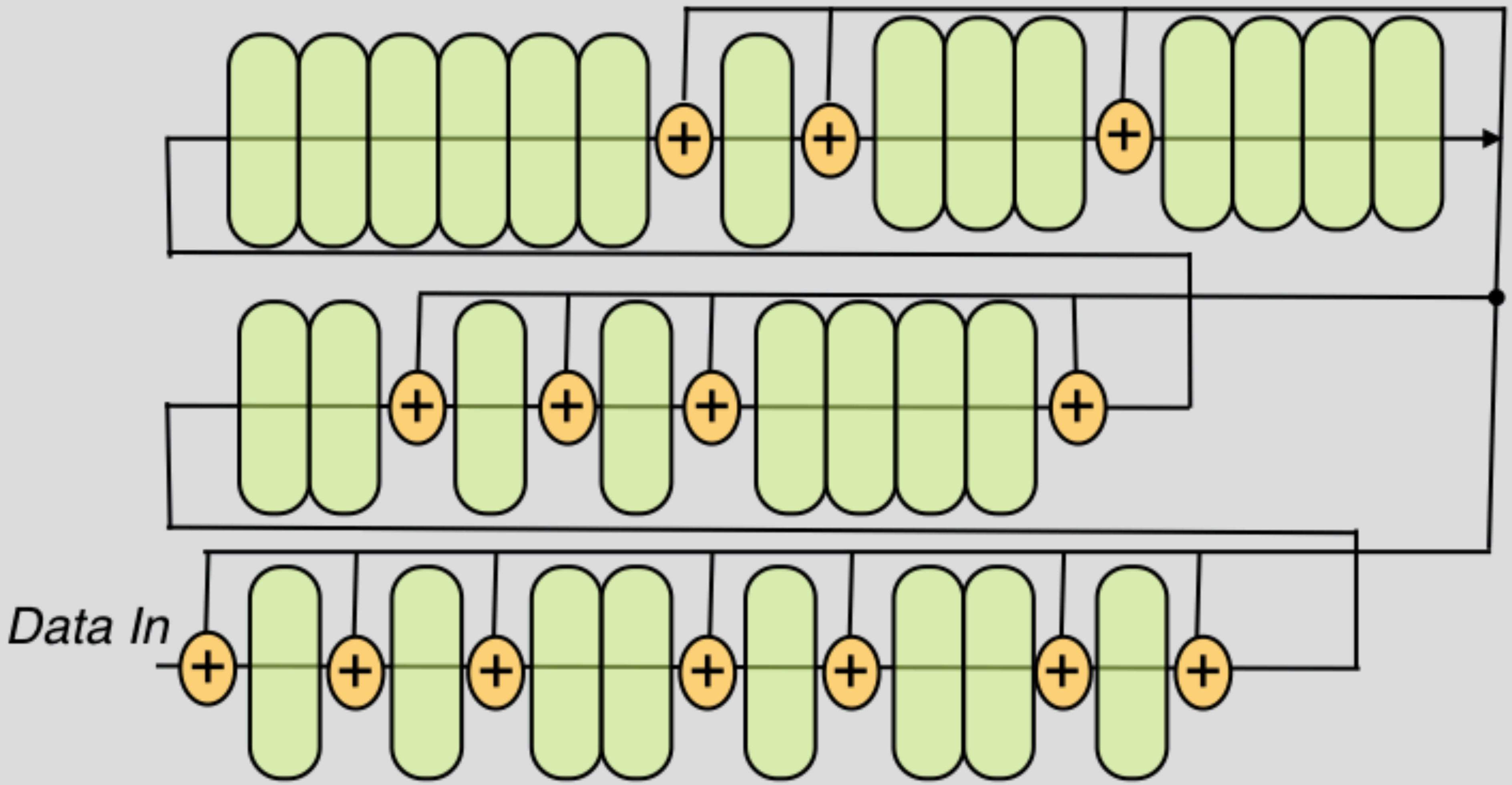
CRC Value after an Error



You do not need to reproduce this!

Hardware Example: CRC-32

$$\begin{aligned} Sum = {} & x^{32} + x^{26} + x^{23} + x^{22} \\ & + x^{16} + x^{12} + x^{11} + x^{10} \\ & + x^{8} + x^{7} + x^{5} + x^{4} + x^{2} + x + 1 \end{aligned}$$



32 1-bit shift register elements

What have we learned?

A mathematical check can detect errors

A good method can multiple (any) errors

- ***almost*** as good a $1/2^n$ chance of failure for an n-bit check

Can be implemented in hardware

A more complex generator that depends on initial state and data result in cyphers that cryptograhically authenticate

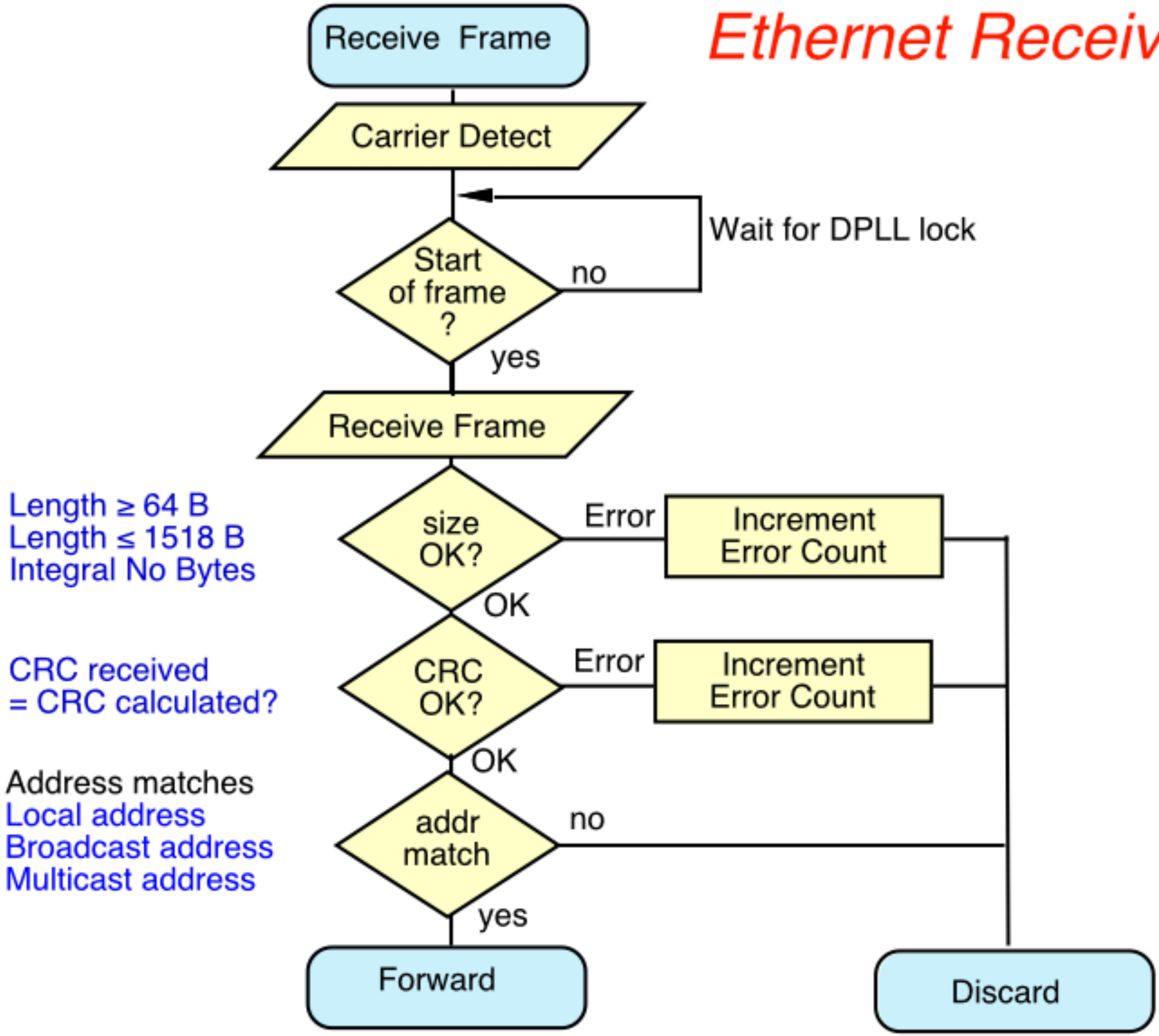
So…

how much protection do we need?

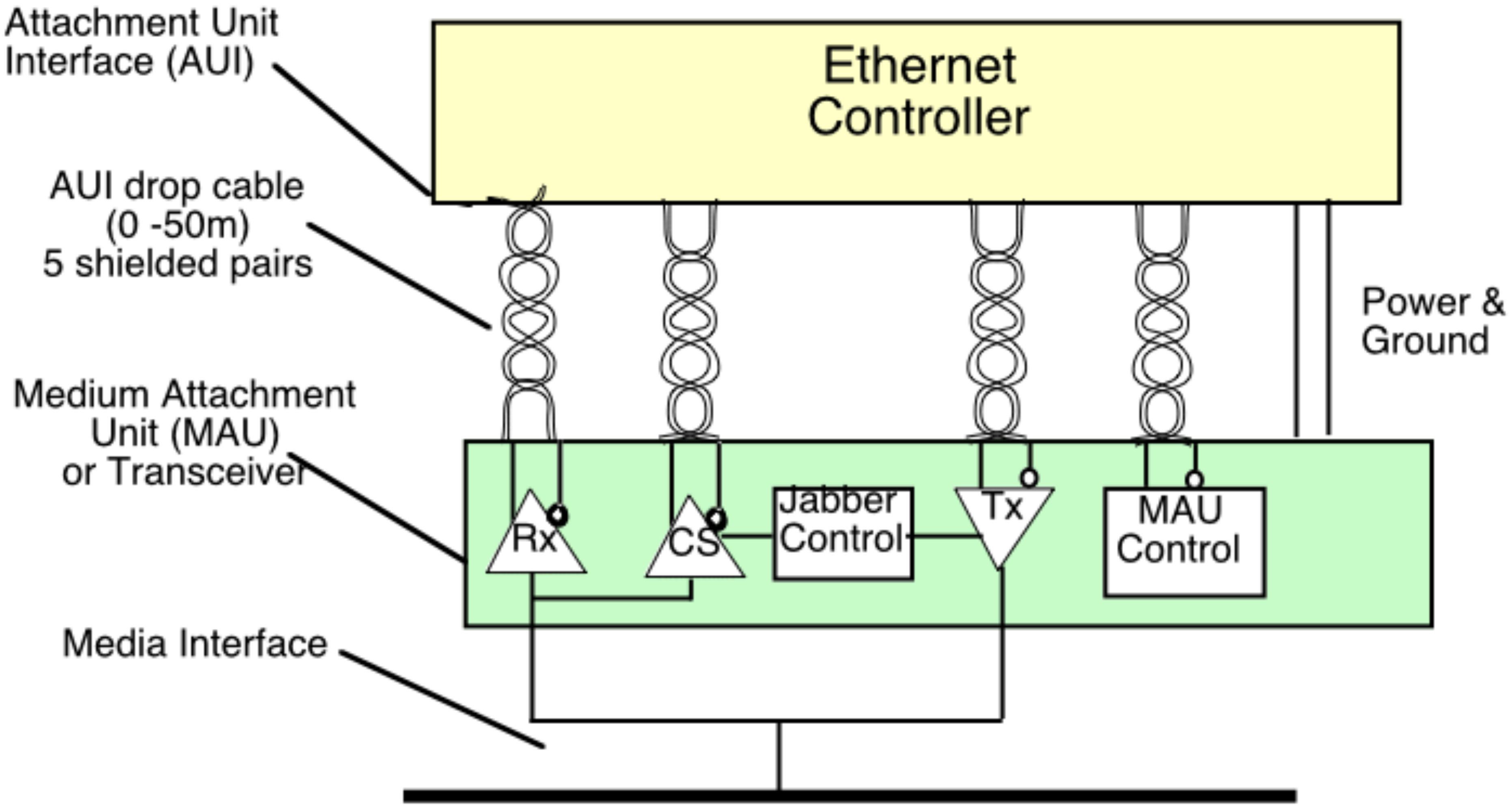
What size of Cyclic Redundancy Check?

- A 1-bit check (parity)
  - Detects 50% of errors, 50% are undetected.
  - Used for individual bytes … runs of errors indicate failure
- An 8-bit check (longitudinal parity)
  - 1/256 chance of a value being accidentally valid
  - Good for short frames, as in NMEA.
- A 16-bit check (Internet Checksum)
  - $1/(256)^2$ chance of a value being accidentally valid
  - Good for some types of packet errors (such as byte swap)
- A 32-bit check (Ethernet FCS)
  - $1/(256)^4$ chance of a value being accidentally valid
  - Good for bit errors in frames unto 10 KB
  - Detects byte errors in messages
- A 256-it check (SHA-256, AES-256)
  - $1/(256)^{32}$ chance of accidentally valid (1/4,294,967,295)
  - Needed for higher assurance of message integrity

Ethernet Receiver



Transceiver AUI Interface



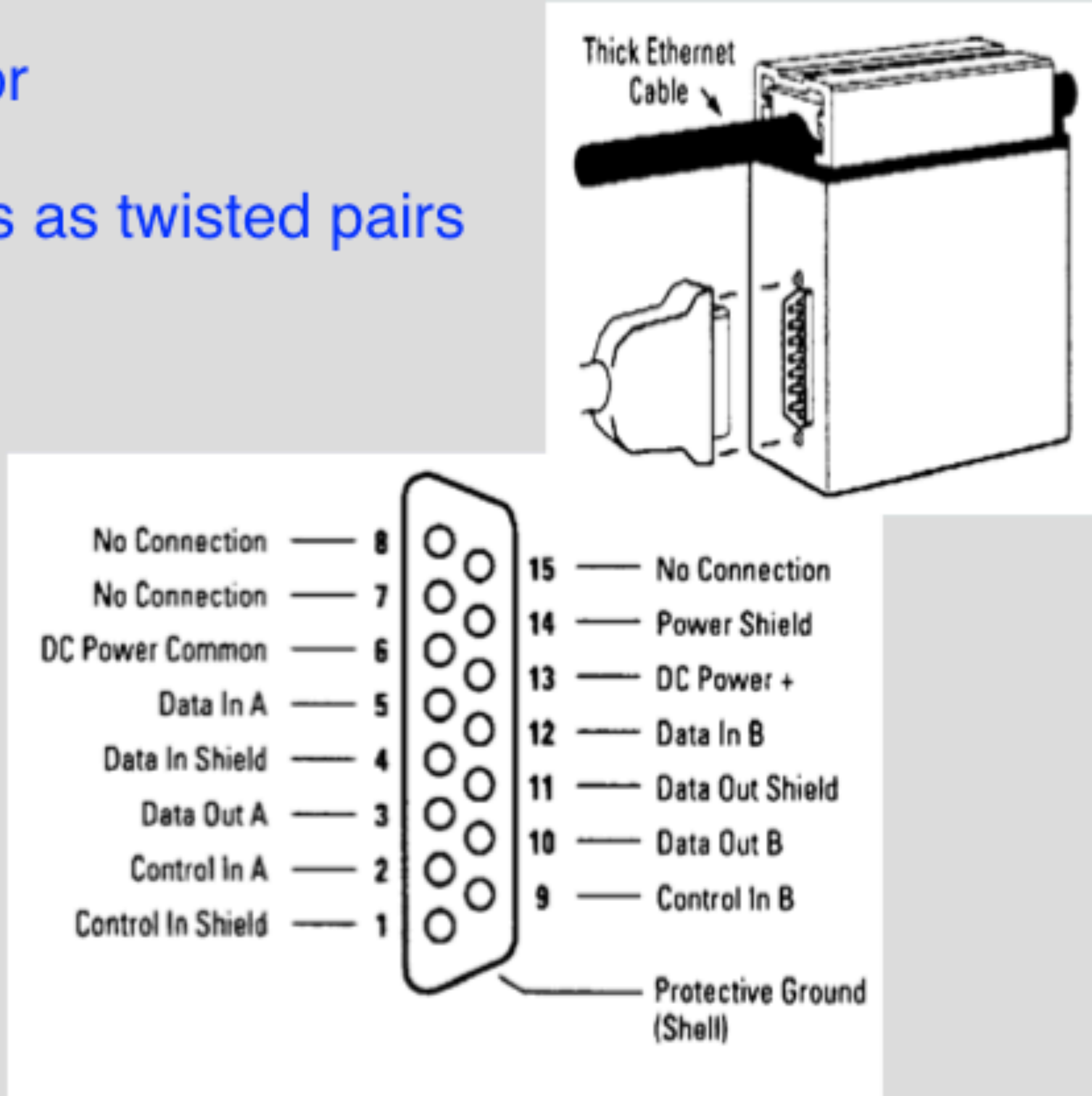
* Jabber is transmission of a frame longer than the maximum allowed.

10Mbps AUI Interface

15-Way D-Type Connector

Carries Tx and Rx Signals as twisted pairs

+12V Transceiver Power



This slide for additional information only

MAC Functions

- **Gain access to medium (Transceiver electronics)**
- **Co-ordinate sharing of the medium between users (CSMA/CD)**
- **Address single and groups of stations**
  - (i) Static address for each computer (copied from PROM)
  - (ii) 1 or more dynamic group addresses (e.g. multicast)
  - (iii) A single broadcast address to send to every computer in the LAN
- **Can detect some failures**
  - (i) Transmission errors (e.g. CRC-32)
  - (i) Protocol errors (e.g. jabber (too long) , runt (too short))
  - (ii) Cabling faults (e.g. loss of carrier, reflection from a break)

Notes:
The course does not require you to reproduce CRC calculations.
The course will not ask about the AUI interface or cable

Questions to think about

(i) What is the Ethernet ***destination address*** used for?
(ii) Why is the first bit never set in an Ethernet ***source address***?

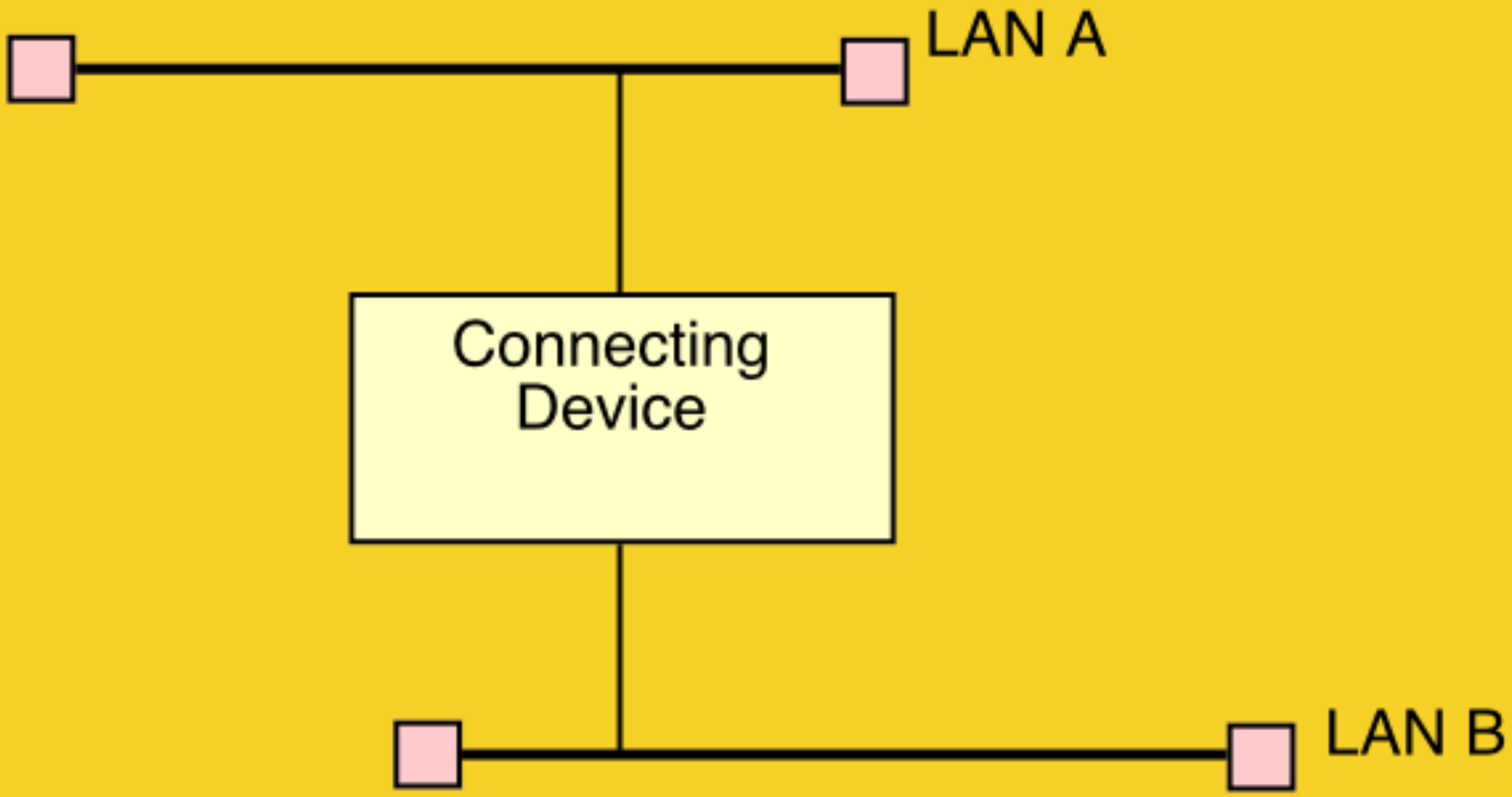
There are two types of anti-social Ethernet frame: Jabber and Runts

(i) What are the minimum and maximum Ethernet **frame sizes***?
(ii) Why does **Jabber** impact performance?
(iii) Why do **Runts** impact reliability?

*Note: * When Ethernet is used for an Internet link,*
the maximum frame size limits the largest size of Internet packet.
*The Upper Layers called this the **Maximum Transmission Unit (MTU)**.*

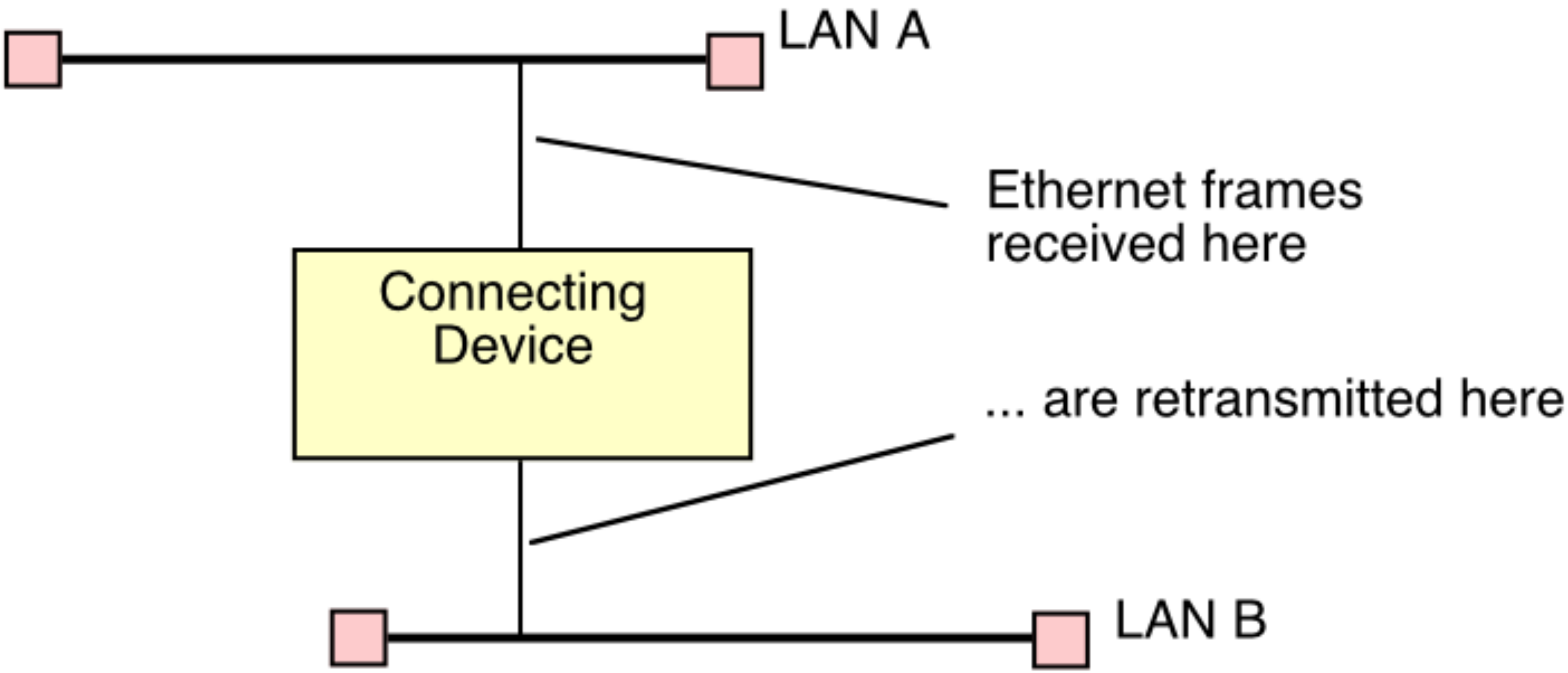
Connecting LAN Segments to form a Collision Domain

Repeaters forward frames



Module 4.1

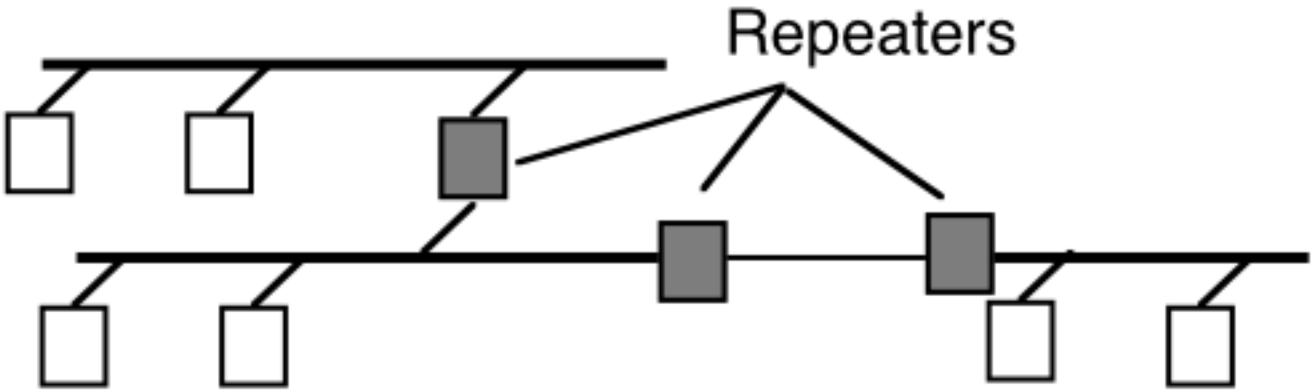
Repeater



Repeater regenerate the same signal to all output ports

- One port acts as an input
- The signal is then cleaned (regenerated) and sent to all outputs

Repeater



Uses:

- Extends media length and number of NICs
  - Recall 10B2 permits only 30 NICs per cable segment
- Allows conversion between media types
  - 10B5 to 10B2, or to 10BF, 10BT
- Allows for more flexible cable routing rather than a single bus

Function:

- Connect segments and regenerate signal

Part 1 Regeneration of Clock and Data

The output is not just a “better” signal, it is a perfect waveform!

- The clock at the input port is extracted using a DPLL
- Each bit is decoded using a Manchester Decoder

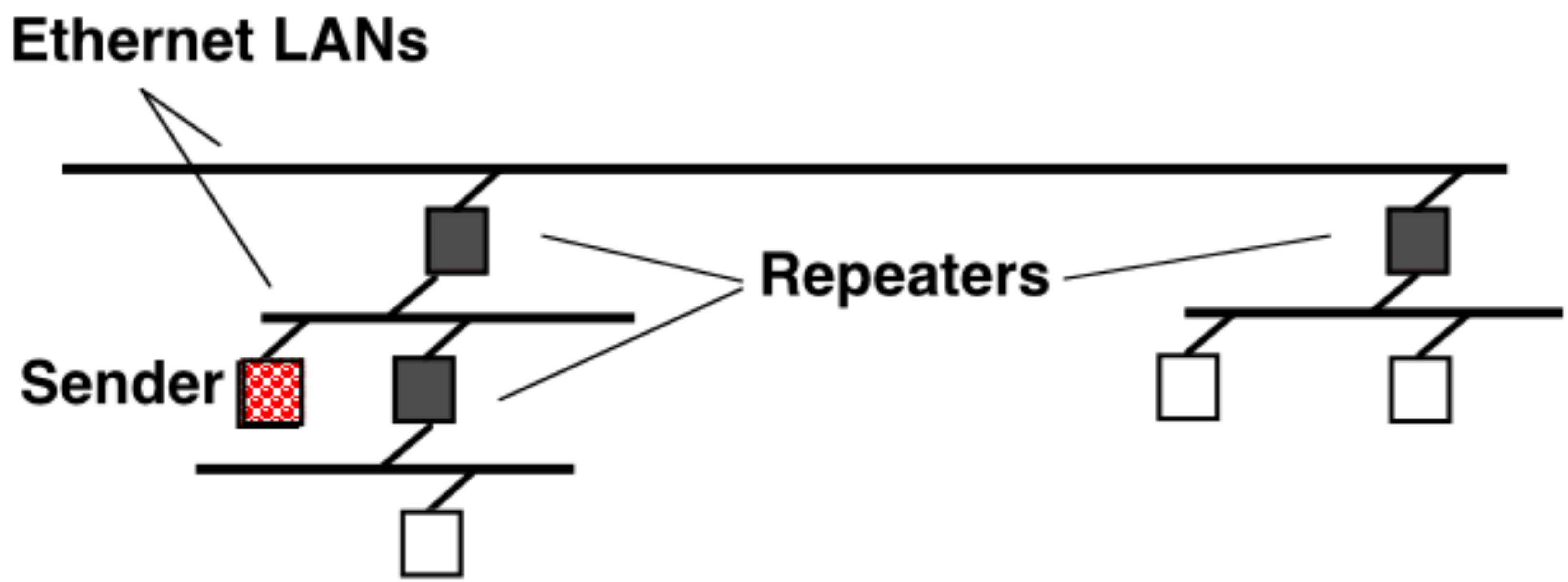
The timing of the signal is regenerated

- A stable clock is used to send each ‘1’ and ‘0’

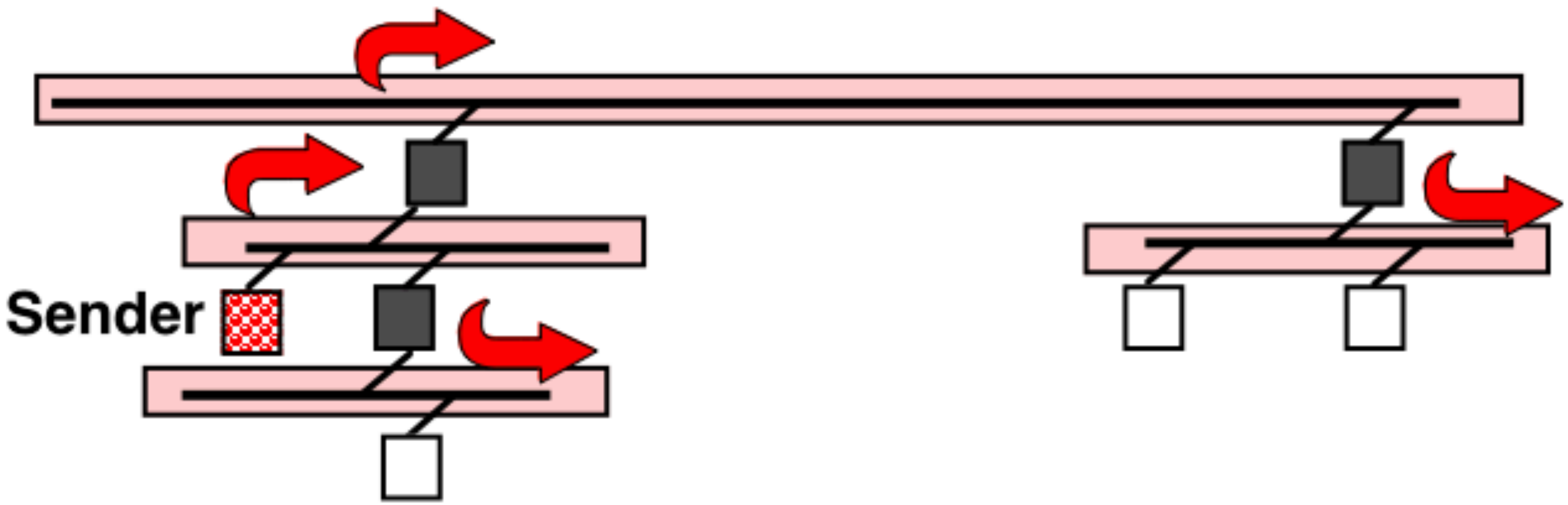
Reconstruct 0’s and 1’s of frame

- Each-output bit is re-encoded using a Manchester Encoder
- The repeater MUST also regenerate the ***full*** preamble

Regeneration to all parts of the LAN



Assume one NIC sends



In fact, the repeater must also generate “JAM” signals when needed.

Repeaters: Summary

- **Two cable segments can be connected by a repeater**

  The repeater regenerates the signals on each port except received

- **Repeaters can connect different media segments**

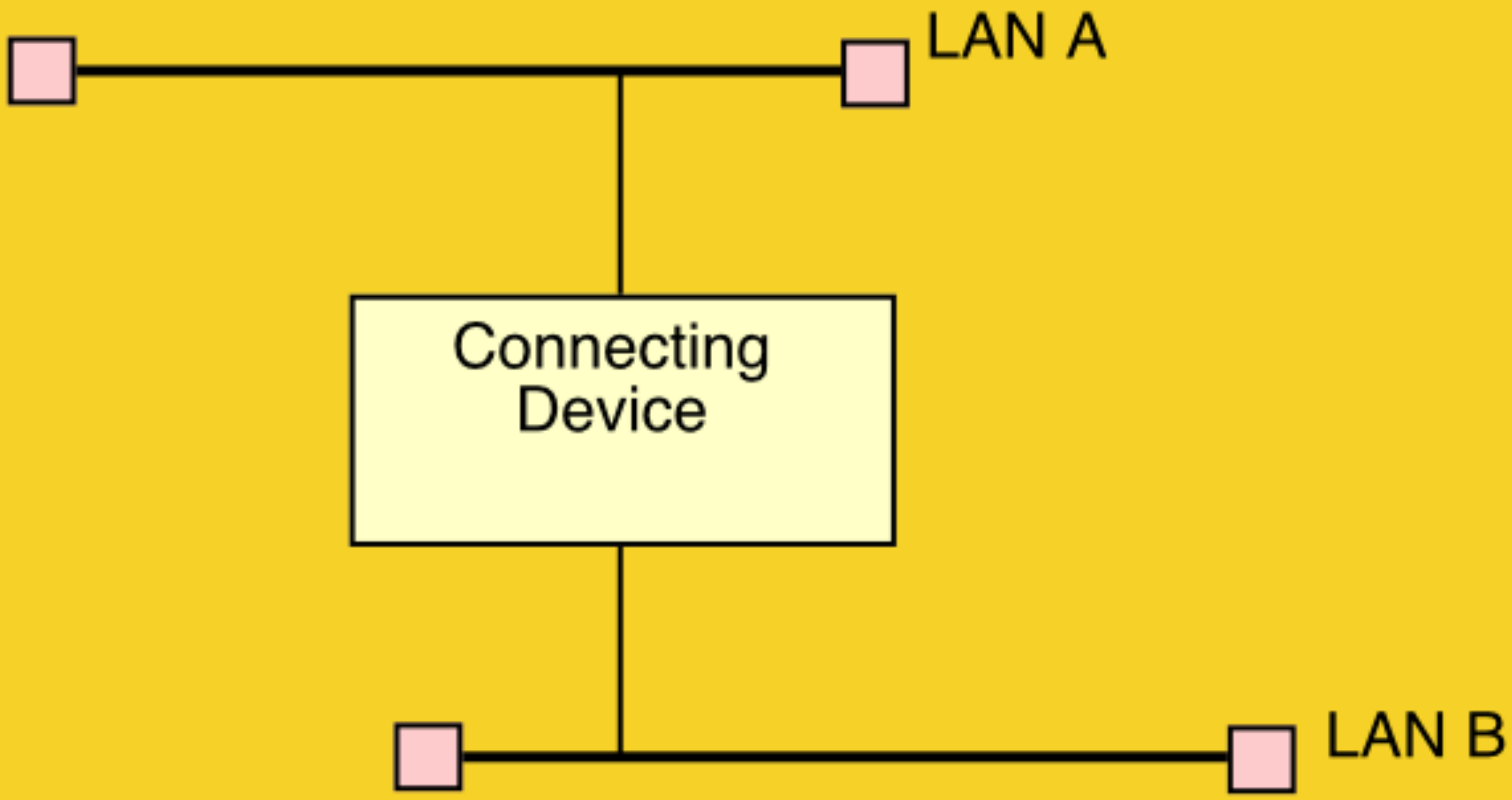
  Any ports can have any media

  All ports have to operate at the **same speed**

  Repeaters understand CSMA/CD
  …. more information about repeaters in the next part….

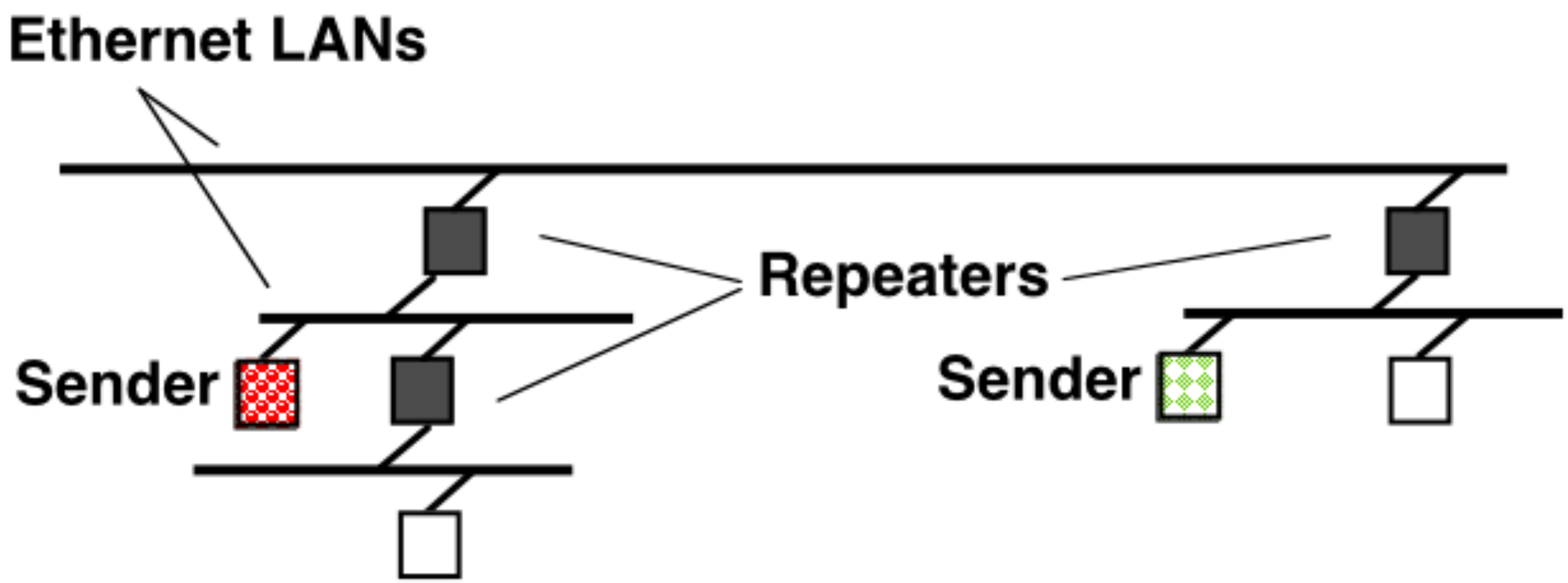
Connecting LAN Segments to form a Collision Domain

Repeaters must participate in CSMA/CD



Module 4.2

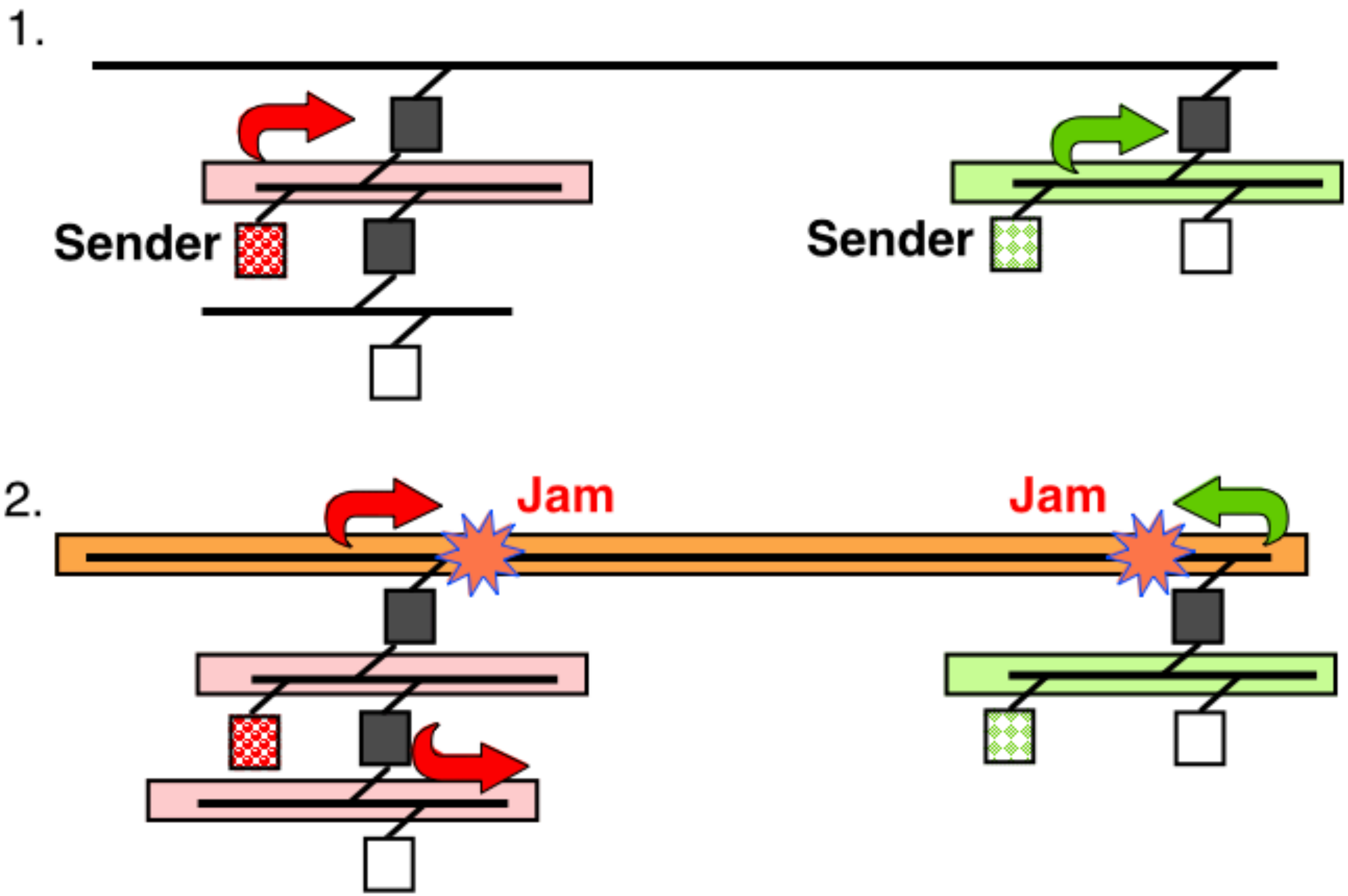
Participation in CSMA/CD



Assume both senders transmit at same time

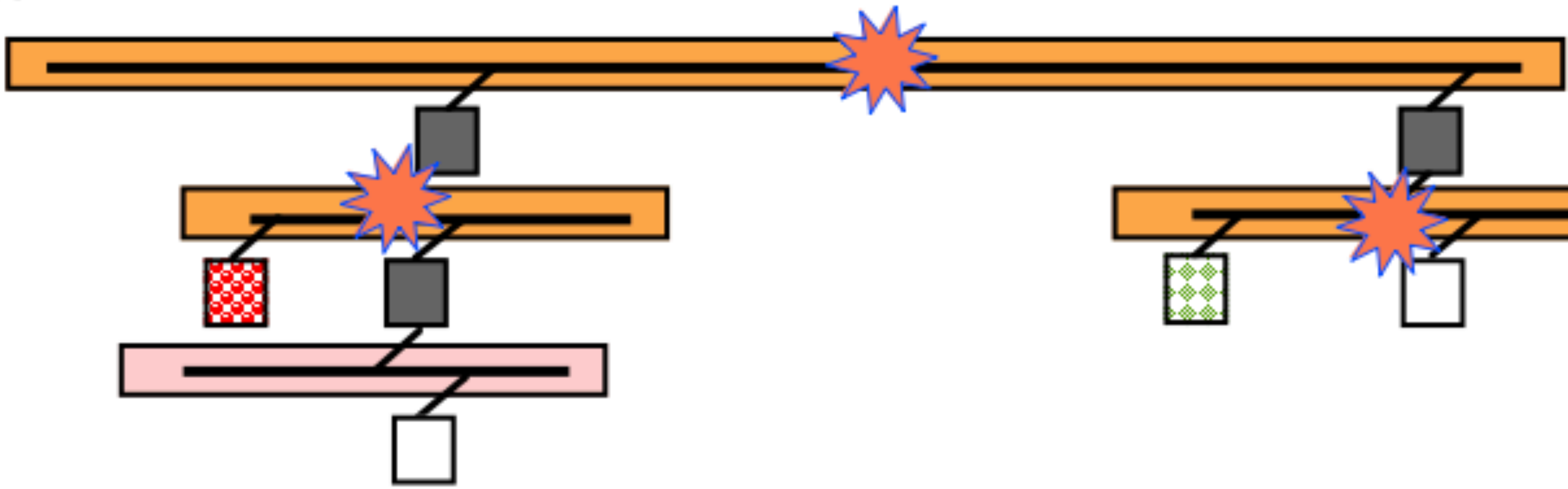
All need to see each other signals

Repeater Network (1)

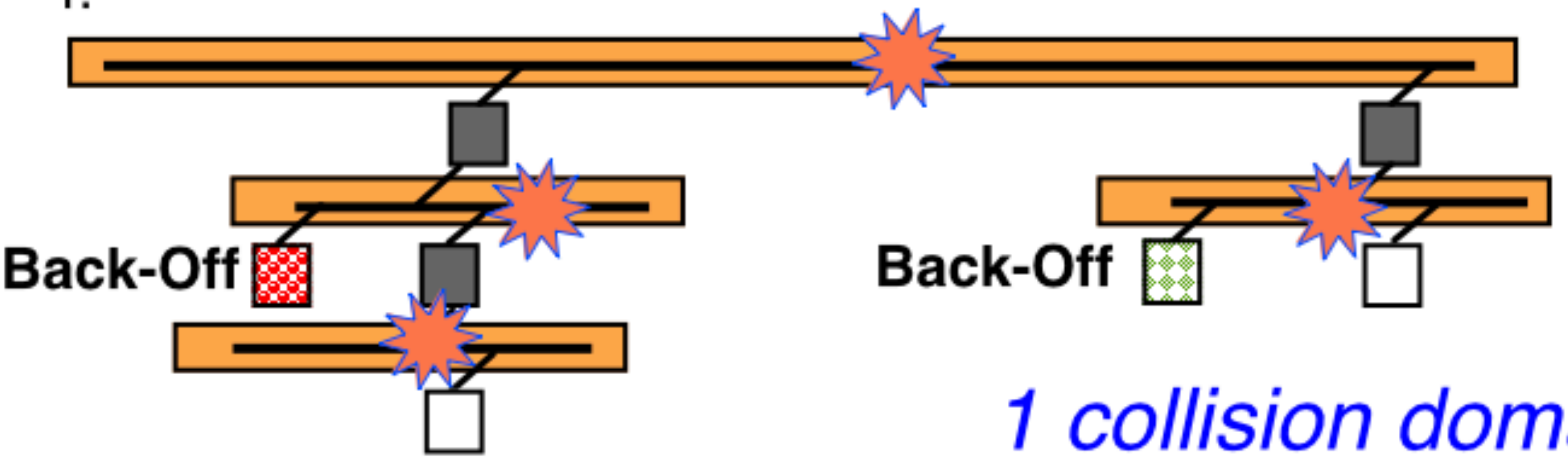


Repeater Network (2)

3.



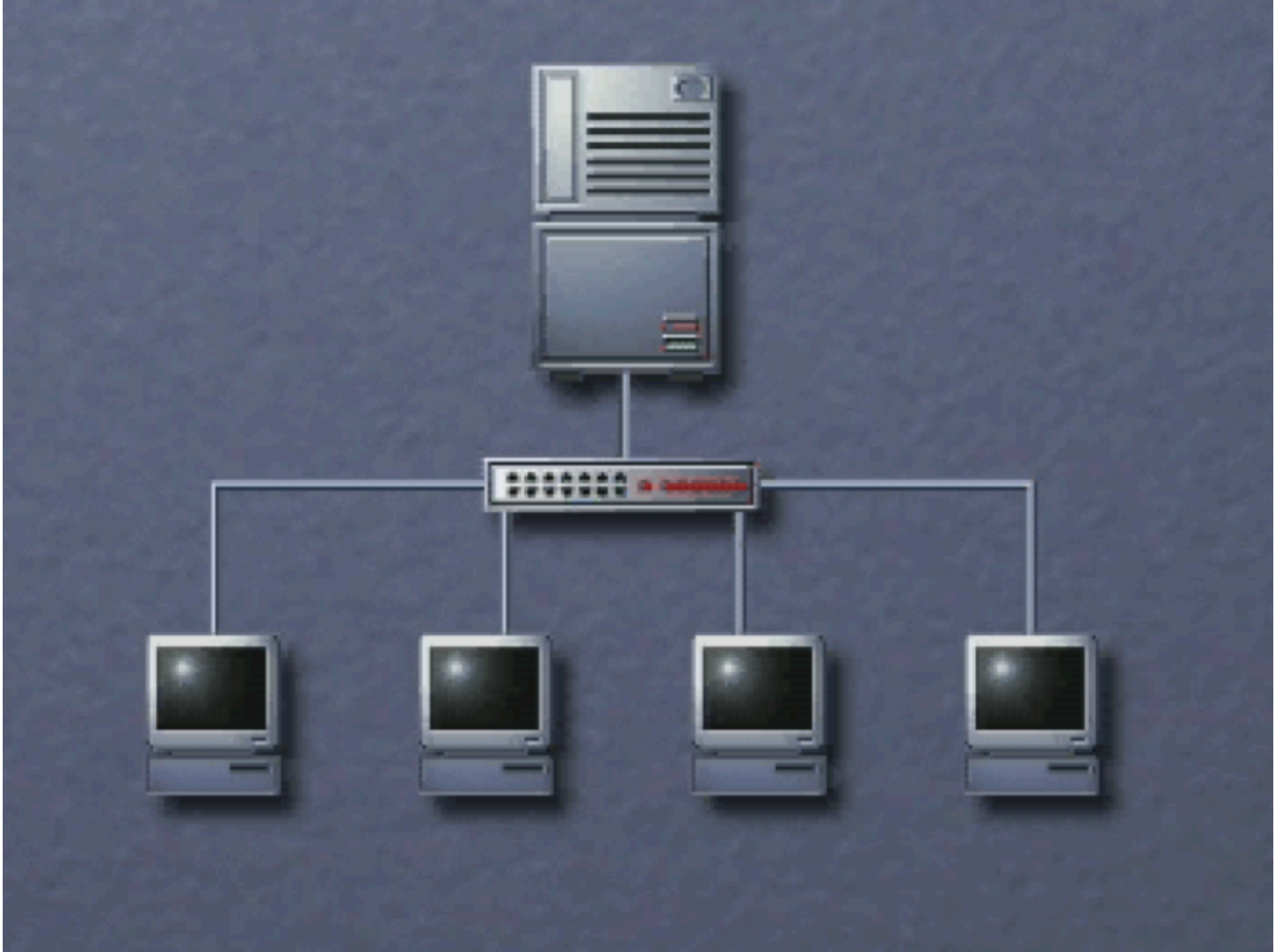
4.



1 collision domain

CSMA/CD and repeaters

- **Repeaters need to:**
  - Detect Collisions
    - “regenerate” collisions on all output ports
    - This takes <u>time</u>...
  - Regeneration limits maximum number of repeaters in <u>series</u>
    - A network can use as many repeaters as it wishes,
    - ... but the designer needs to carefully consider the topology.



5-4-3 Rule

- **LANs can use Hubs and Repeaters**

  Hubs and Repeaters are functionally the same
  Constructs a larger LAN forming a single larger collision domain
  Quite different to a bridge/switch (see later)

- **Any number of repeaters can be used in total providing:**

  Not more than 5 segments in series
  Not more than 4 repeaters on the path between any 2 systems
  Not more than 3 ***active**** segments in series

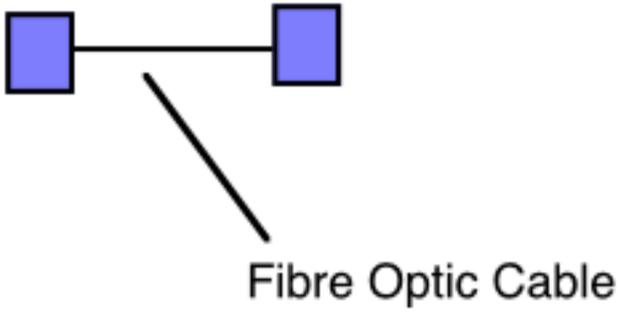
** Active cable segments connect more than two interfaces, which means they have to retrain the DPLL/clock for each frame This adds delay to operation and constrains the timing*

Connecting LAN Segments

Fibre Links

Module 4.3

10 Base Fibre



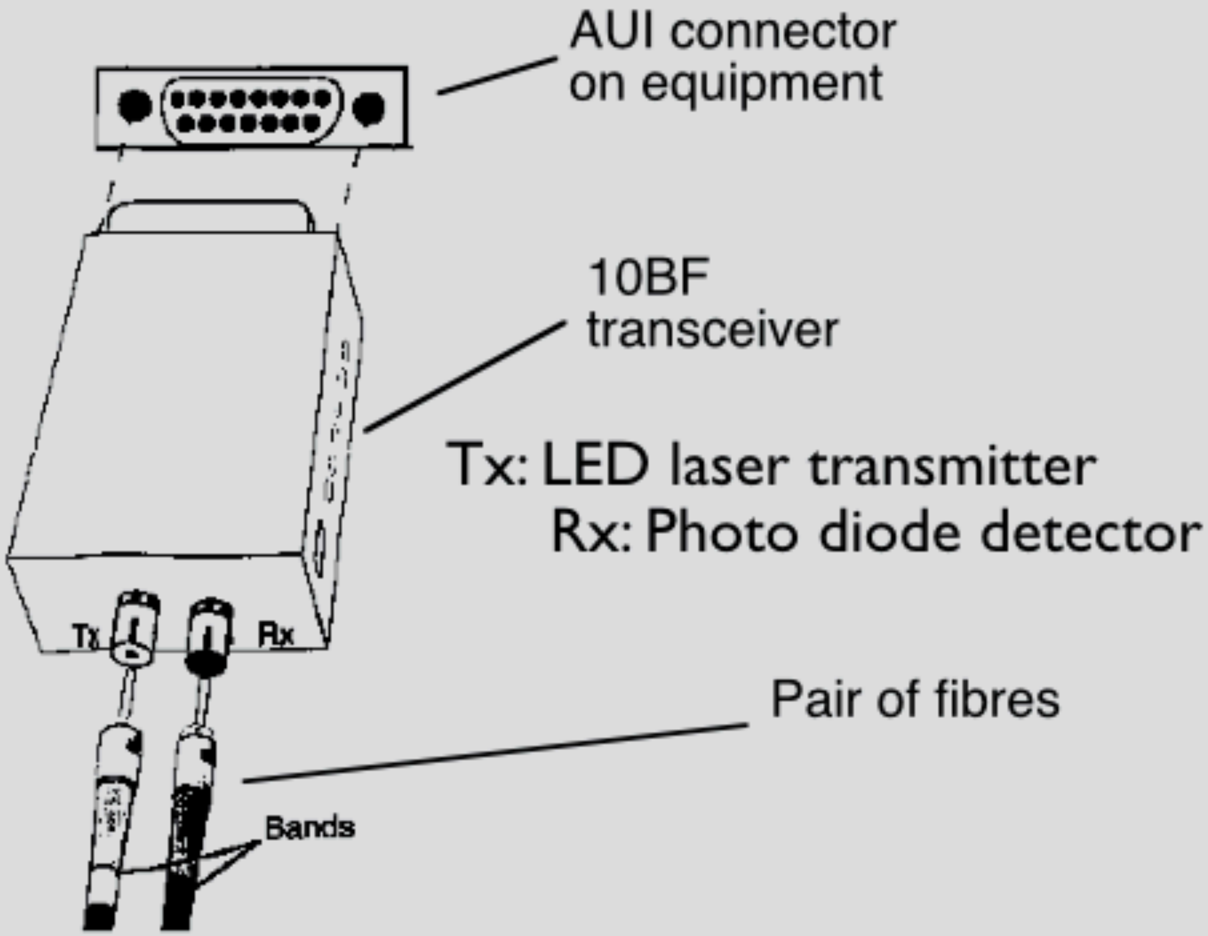
Used for pt-to-pt links
- Segment length ≤1 km (or more)

Fibre provides:
- High noise immunity
- No electrical path
  - (protected from lightning)
  - (secure, hard to tap-into)

Uses external transceiver
- (i.e. connects a pair of repeaters)
- Easy to upgrade transceiver speed

Typical Use of 10B2 to connect two pieces of equipment
A maximum of 2 NICs per cable segment

10 Base Fibre



There are different designs of fibre:

Longer distance cables have lower dispersion loss, but higher cost.

The transceiver needs to matched to the type of fibre.

This slide for additional information only

How large a network can be constructed using 10B2 and 10BF using repeaters?

10B5 ”thick” cable segments may be joined to 500m

- AUI cable up to 50m at each transceiver

“Repeaters” needed to get further

- 3 Copper segments (“ACTIVE”) end-to-end
- 1 fibre segment (“INACTIVE”) 1km

Total = 0.5 x 3+1+.05 x8 = 2.9 km !!!

Actually the size become 5.1m when using 10BF to connect segments

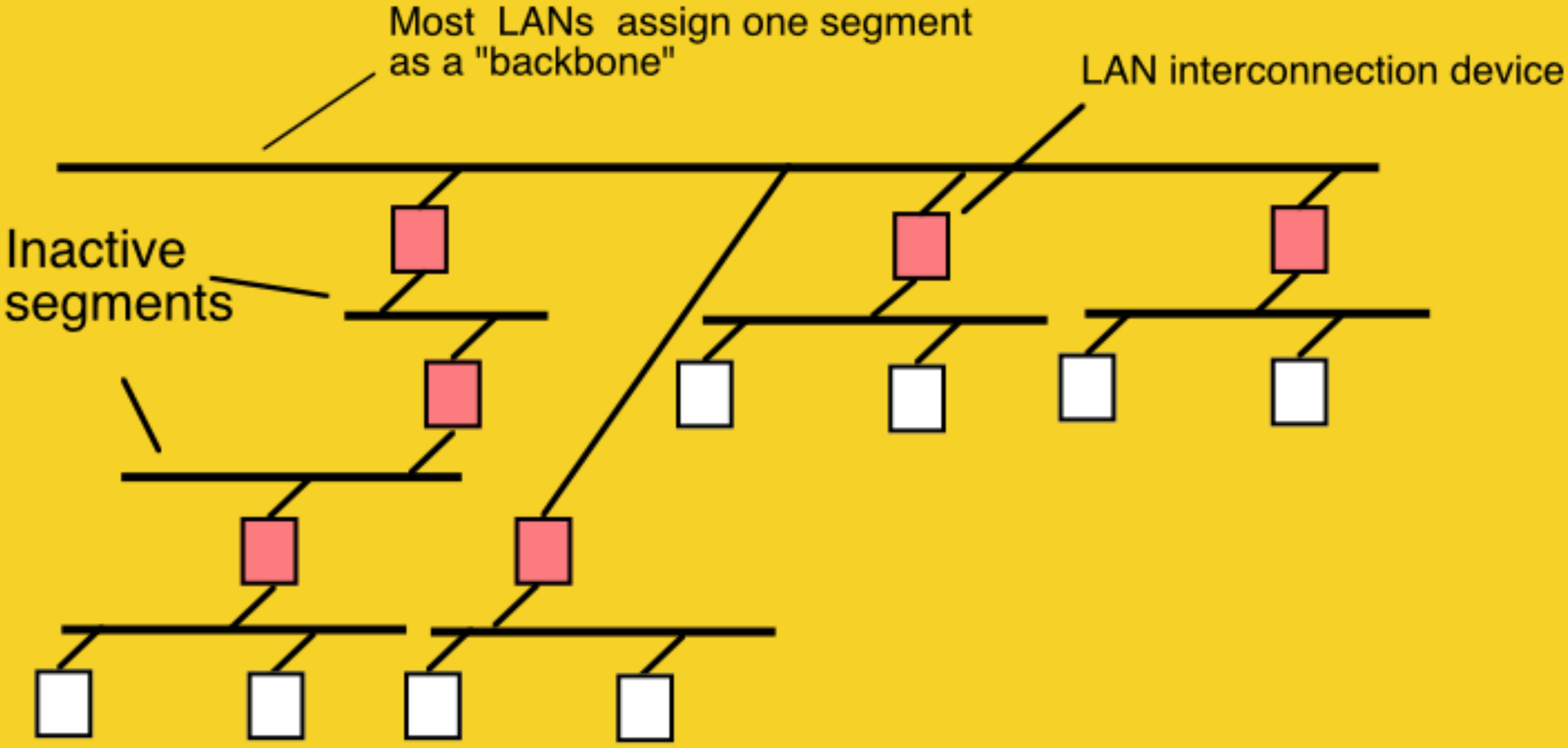
Fibre Media

- **A fibre cable segment connects 2 NICs**
  - Fibre can cover long distances
  - It costs more to buy and install
- **Fibre has additional useful properties**
  - Upgradable to higher speeds and other uses
  - Less easy to "tap"
  - Electrical isolation

…. We'll look more at fibre later

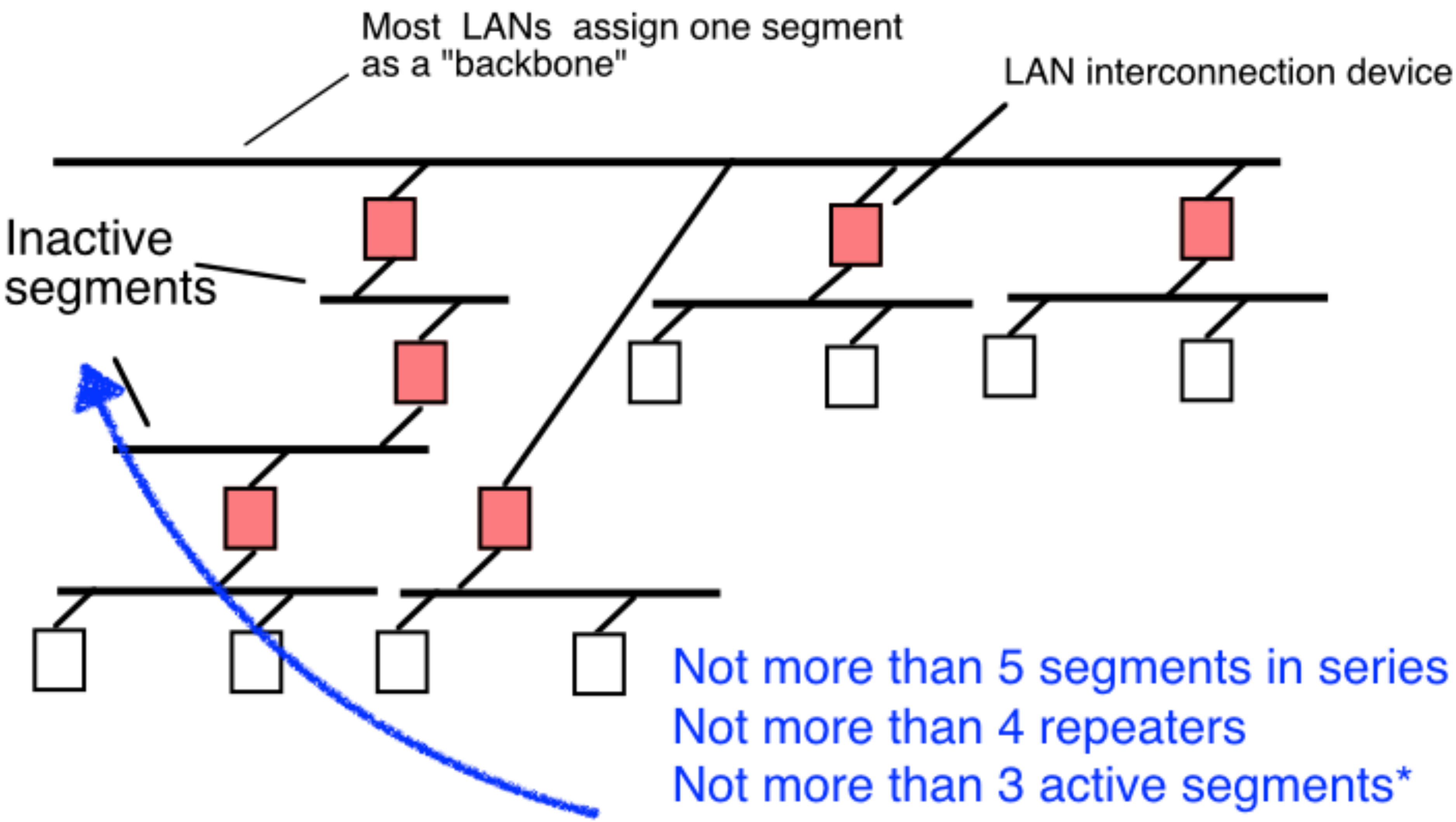
Connecting LAN Segments

5-4-3 Repeater Rule



Module 4.4

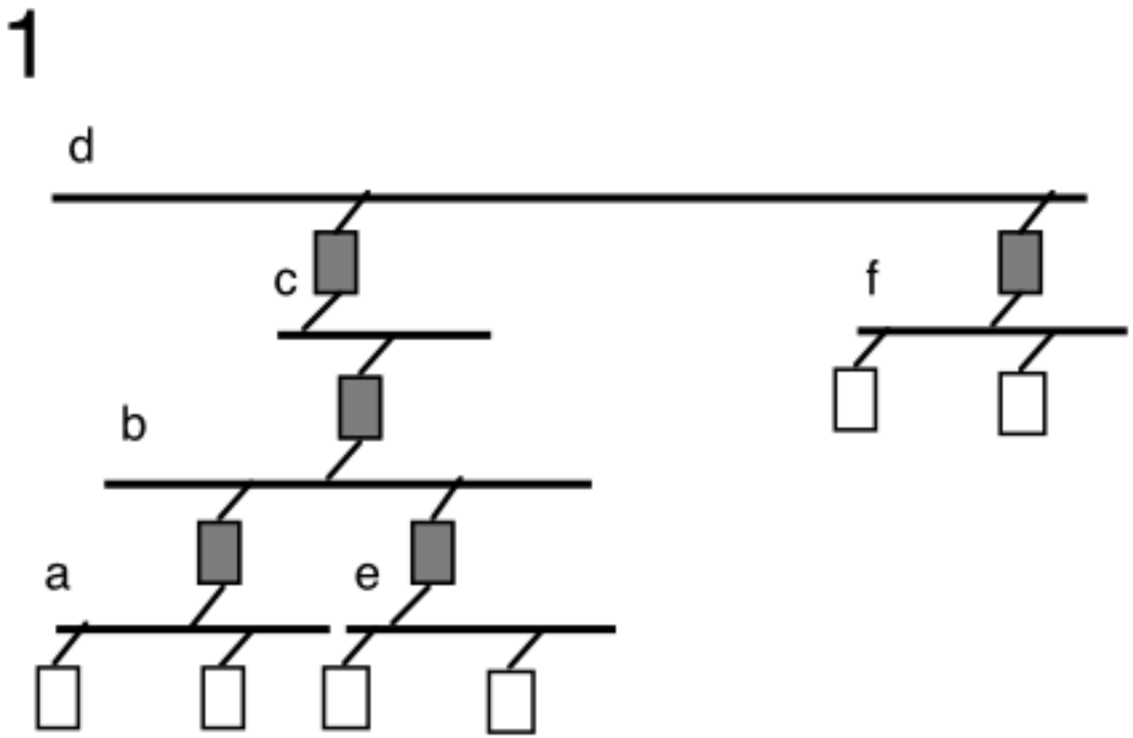
5-4-3 Rule for Networks using Repeaters



Not more than 5 segments in series
Not more than 4 repeaters
Not more than 3 active segments*

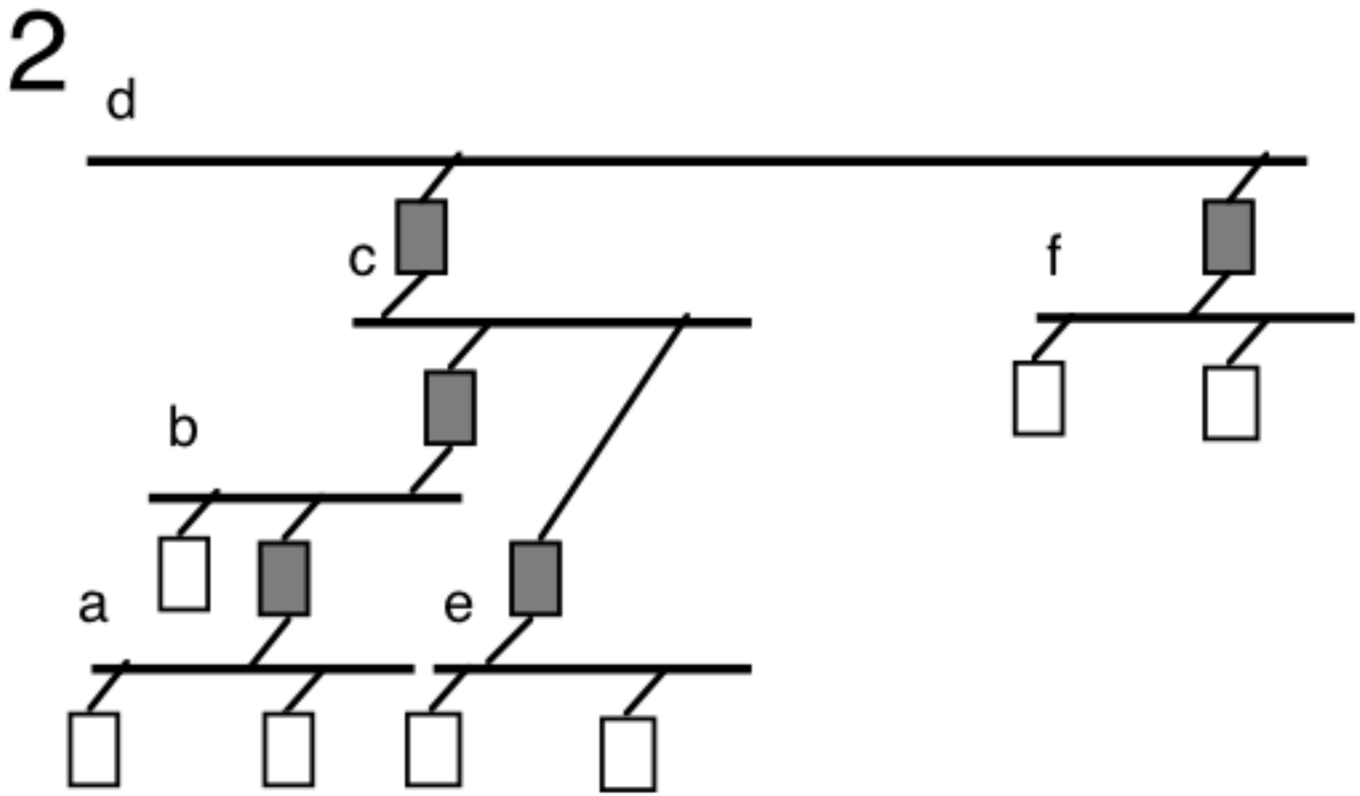
** Inactive cable segments connect just two interfaces, which means they have do not retrain the DPLL/clock for each frame*

Repeater Network



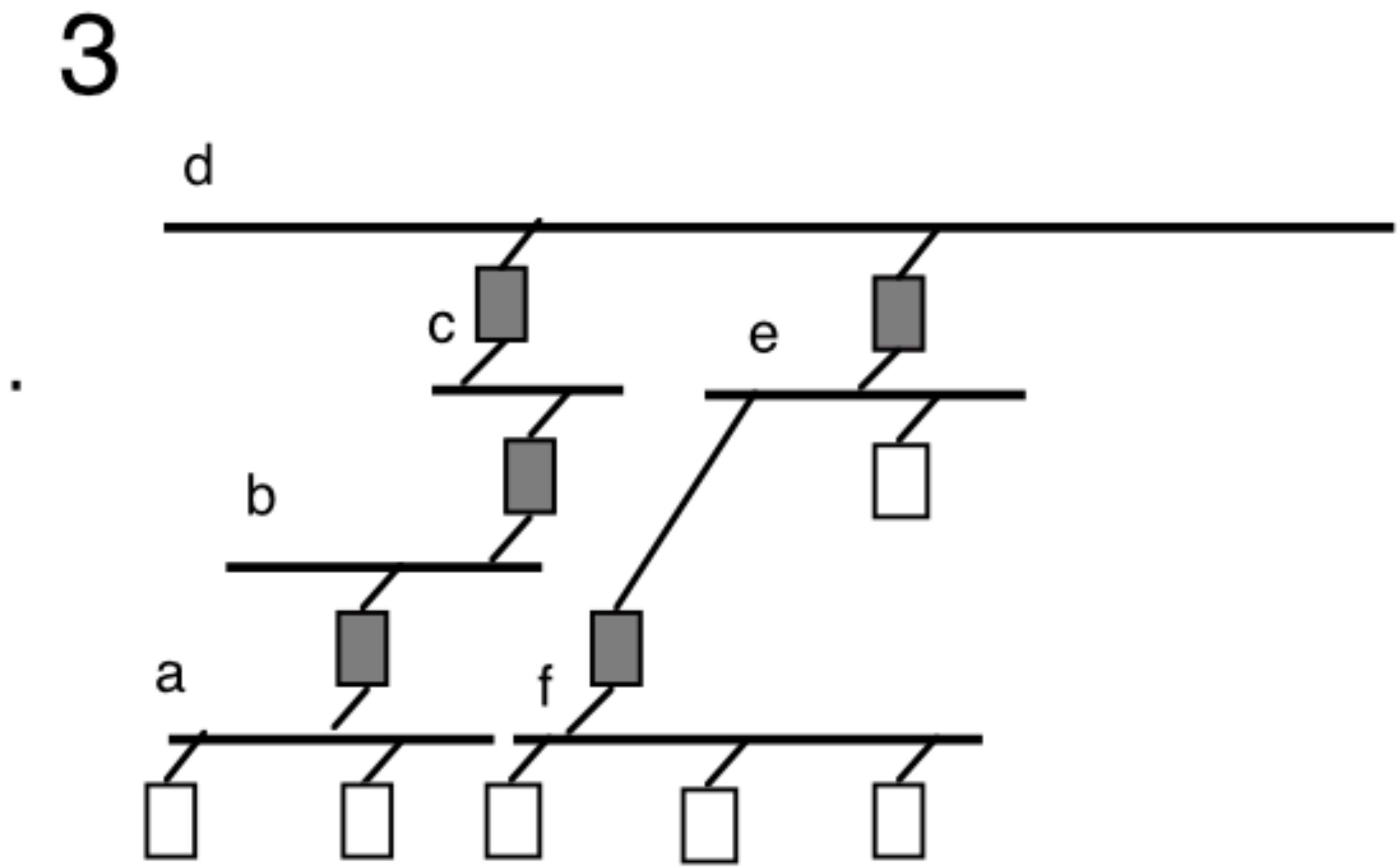
Not more than 5 segments in series
Not more than 4 repeaters
Not more than 3 active segments

Repeater Network



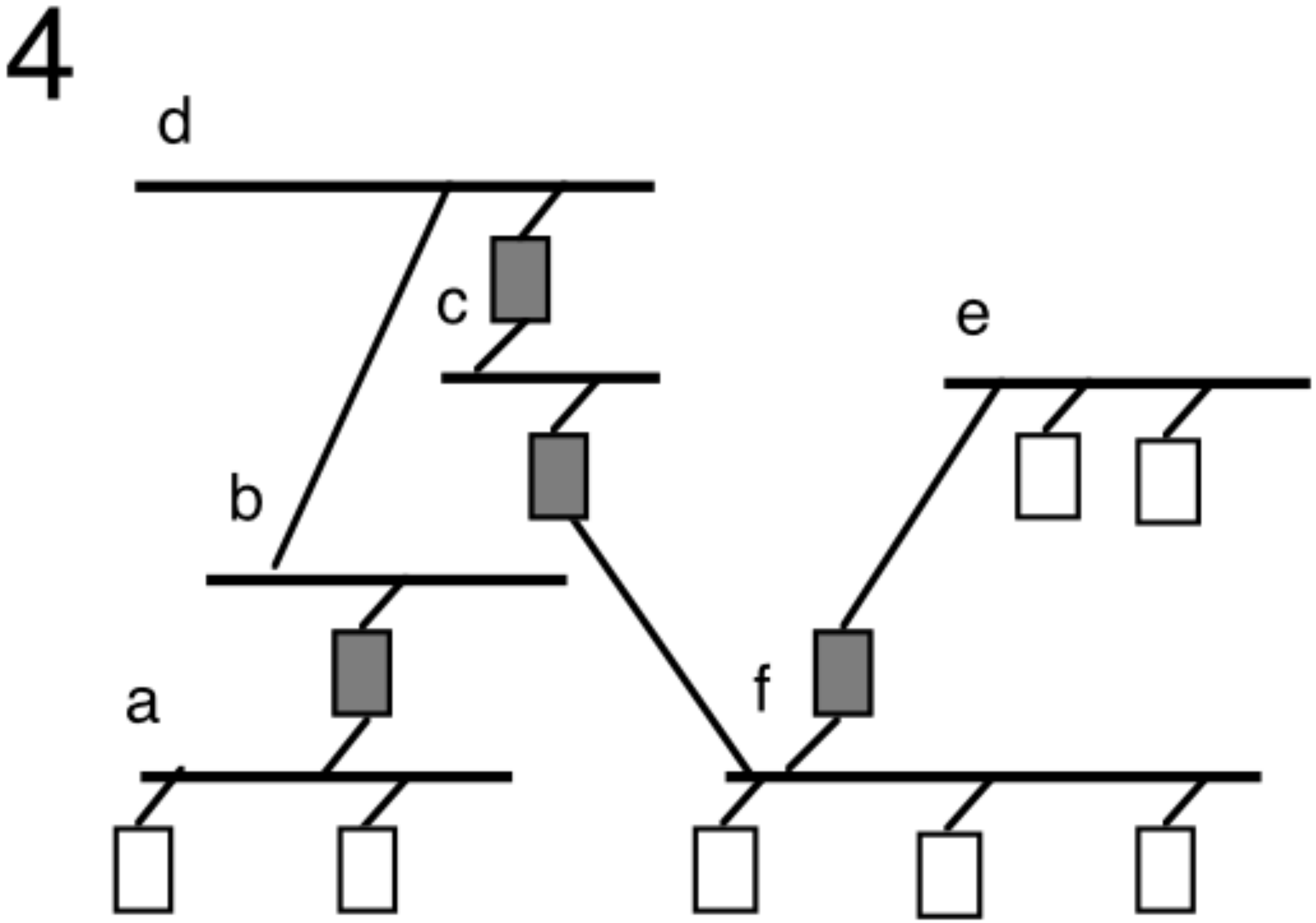
Not more than 5 segments in series
Not more than 4 repeaters
Not more than 3 active segments

Repeater Network



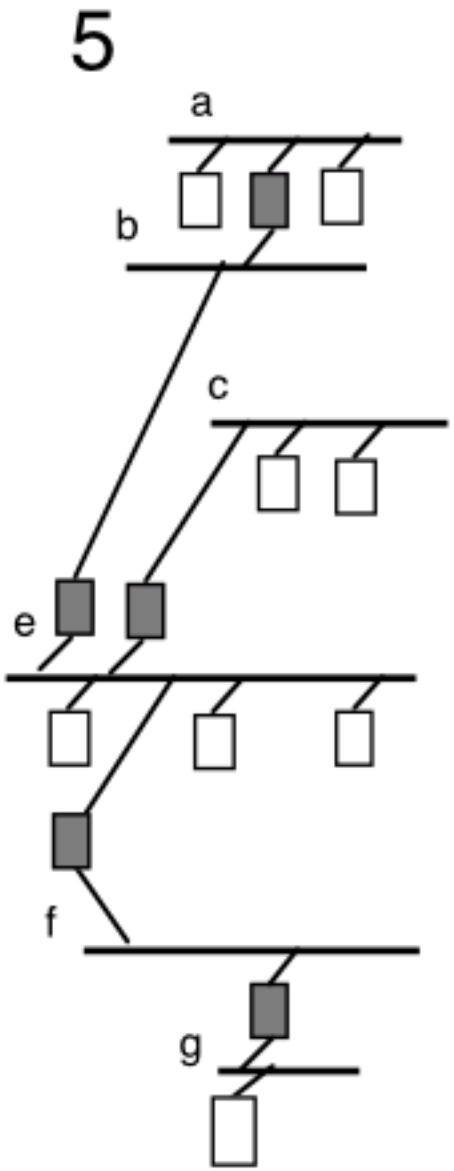
Not more than 5 segments in series
Not more than 4 repeaters
Not more than 3 active segments

Repeater Network



Not more than 5 segments in series
Not more than 4 repeaters
Not more than 3 active segments

Repeater Network



Not more than 5 segments in series
Not more than 4 repeaters
Not more than 3 active segments

5-4-3 Rule

- **LANs can use Hubs and Repeaters**

  Hubs and Repeaters are functionally the same
  Constructs a larger LAN forming a single larger collision domain
  Quite different to a bridge/switch (see later)

- **Any number of hub/routers can be used in total providing:**

  Not more than 5 segments in series
  Not more than 4 repeaters
  Not more than 3 ***active**** segments in series

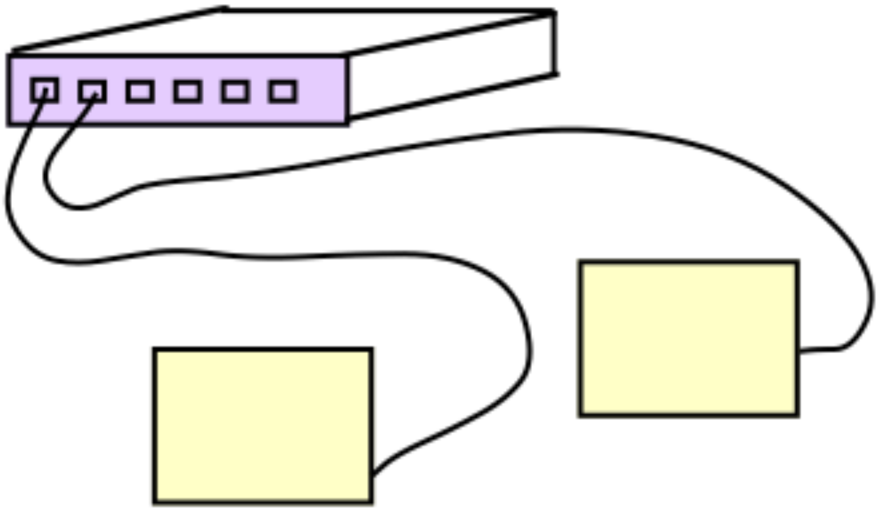
- **Needed to connect point-to-point cable segments:**

  Some technologies are inactive: 10BF and 10BT

Unshielded Twisted Pair Cabling

Module 4.5

10BT or UTP (Unshielded Twisted Pair)

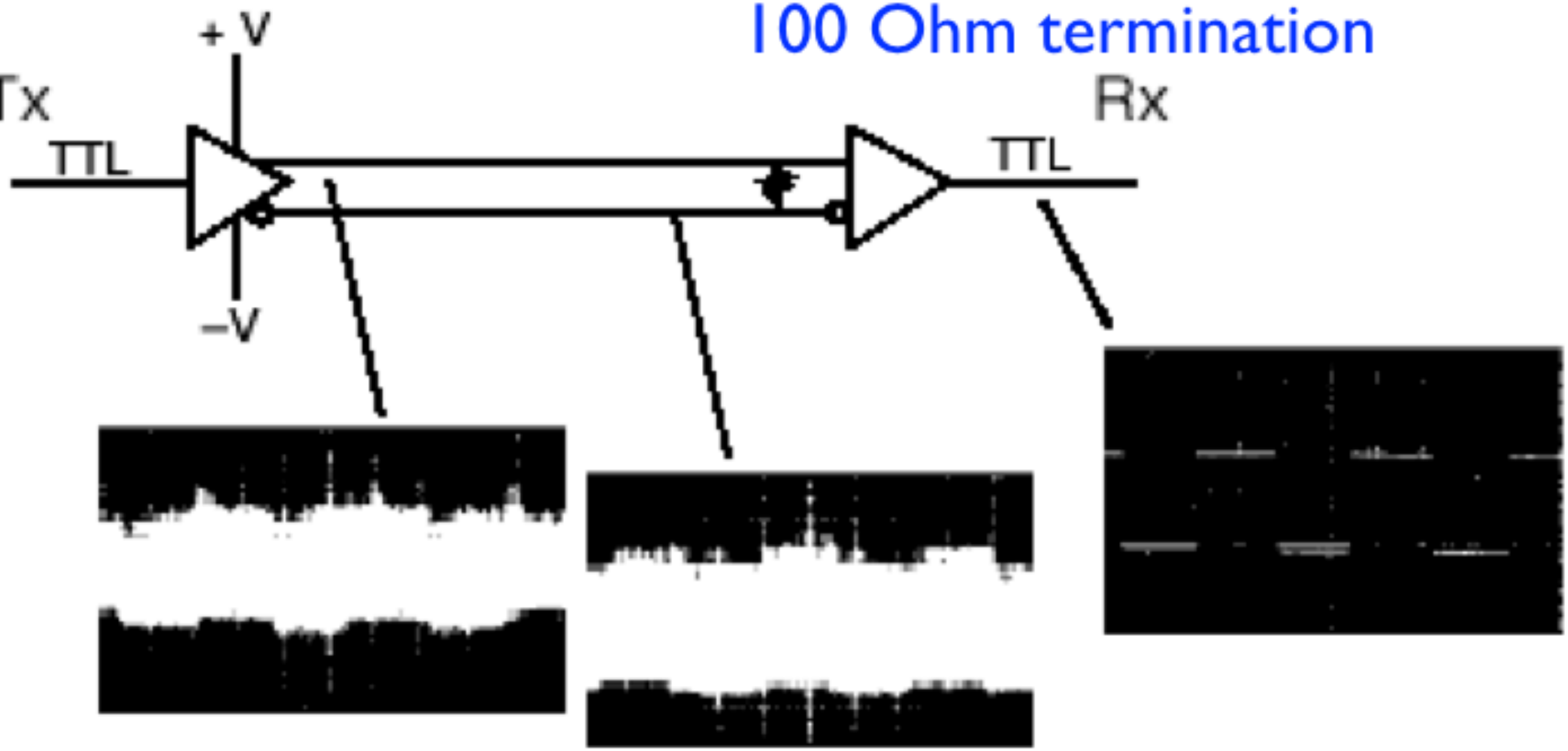


IEEE 802.3i standard (1990)
Segment length 0.6m – 100m
Cable flexible and very cheap
Easy to manage / install

Integrated or external transceiver

Differential Transmission

100 Ohm termination

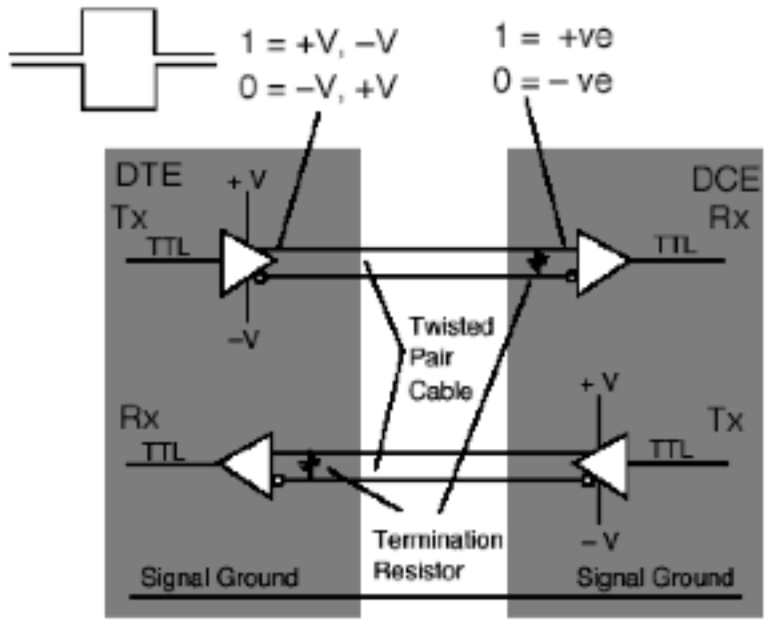


Each direction uses 2 wires TWISTED to form a PAIR

0 Signal sent +ve on one wire, -ve on other

1 Signal sent -ve on one wire, +ve on other

10 BT just two pairs



A UTP cable has four colour-codes twisted pairs

One pair is used for transmission

Pins 1,2 (white+orange/orange)

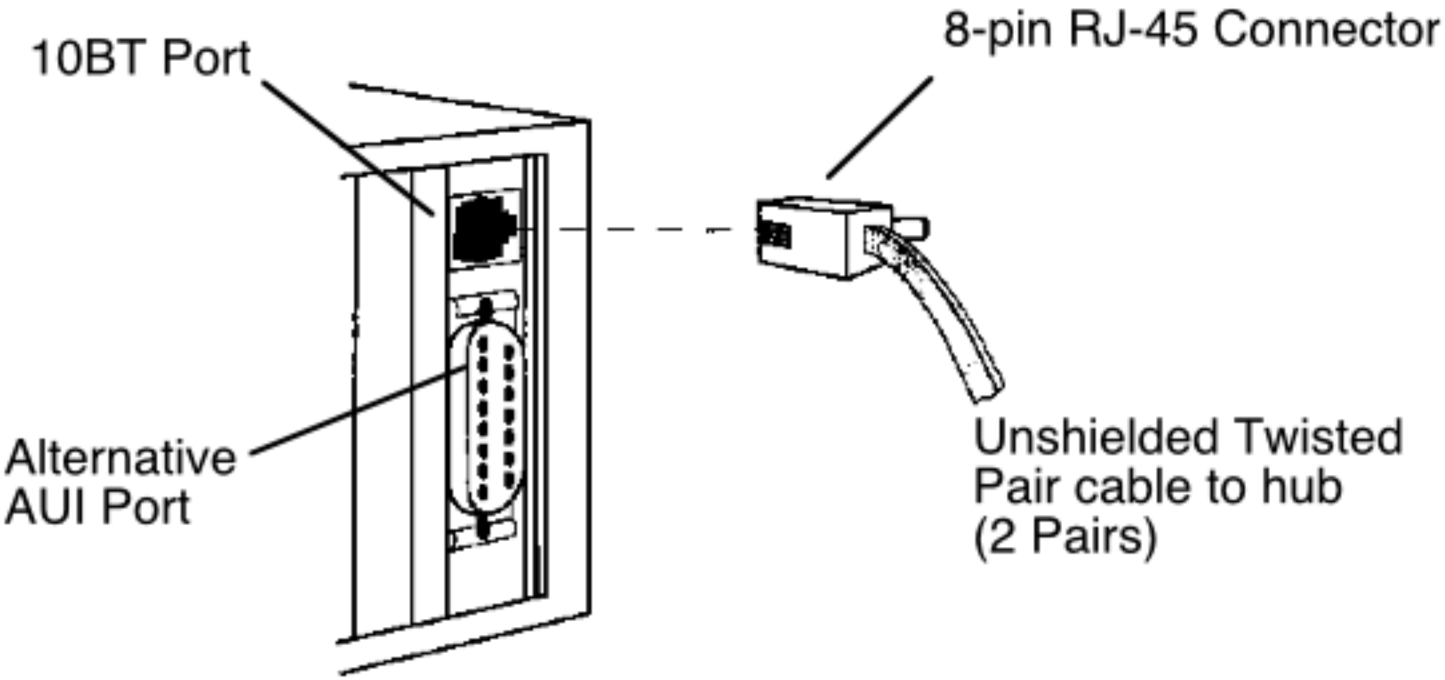
One pair is used for reception

Pins 3,6 (white+green/green)

CSMA/CD means one direction used at a time

Two pairs are not used in 10BT (or could be used for other purposes)

RJ-45 Connectors

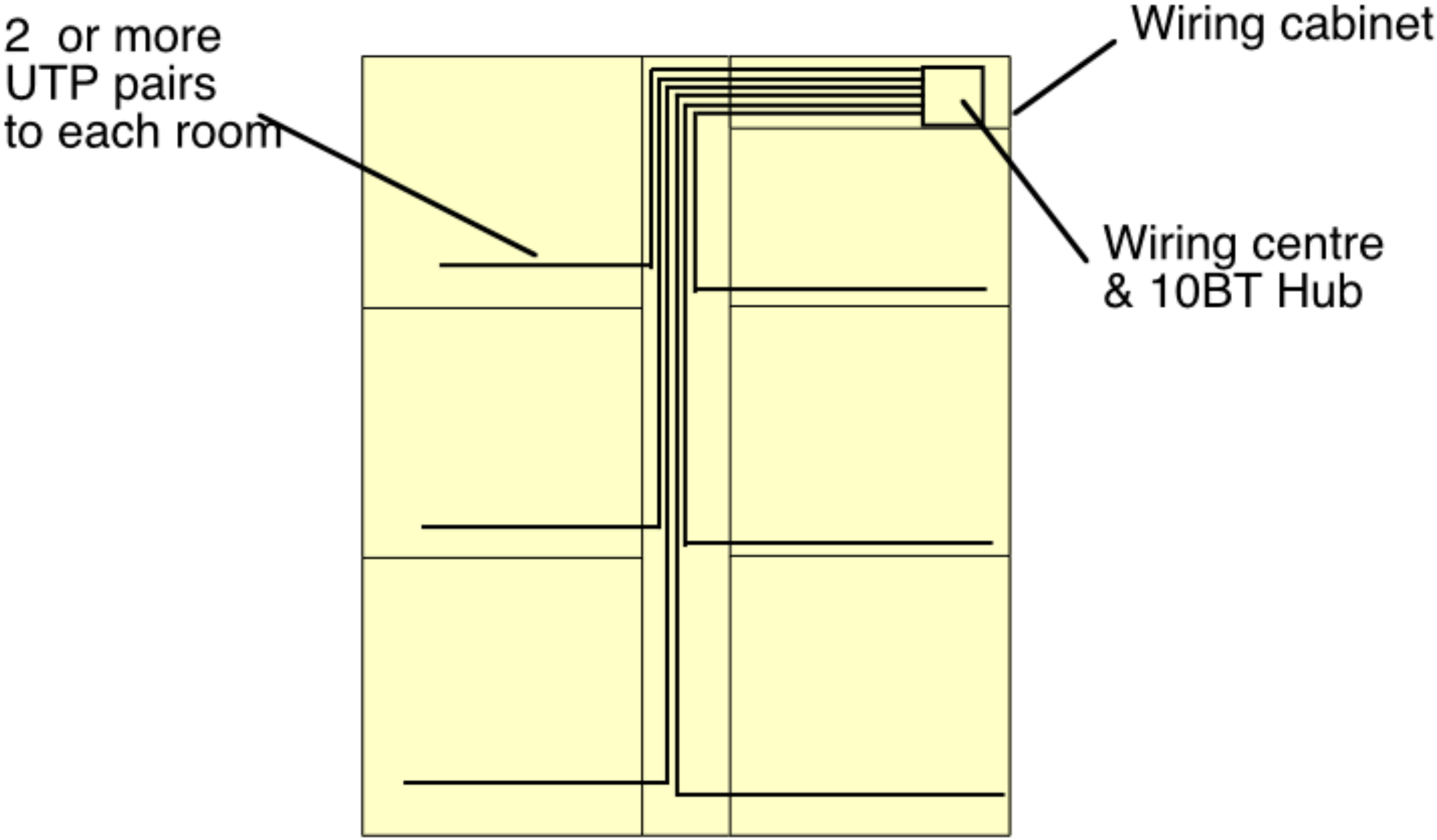


Punch-down tool for installing sockets

Crimp tool for terminating cables

Ethernet 10BT Cabling

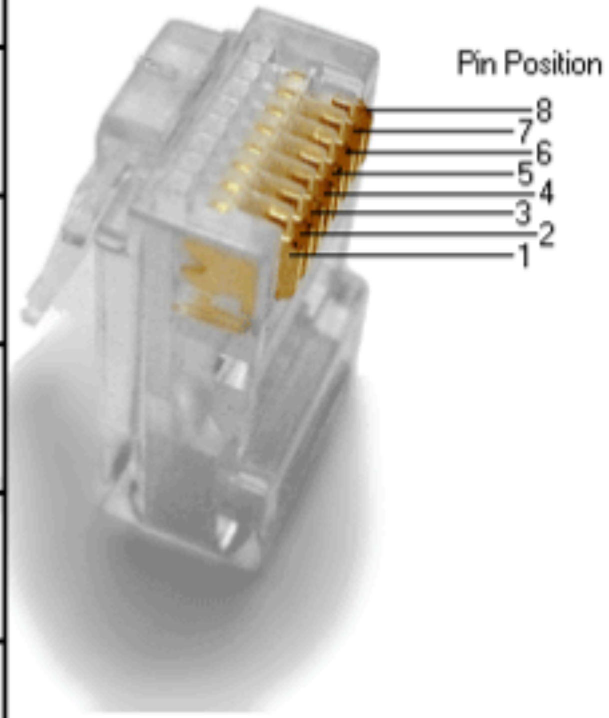
Often cable is pre-installed to many places in the office



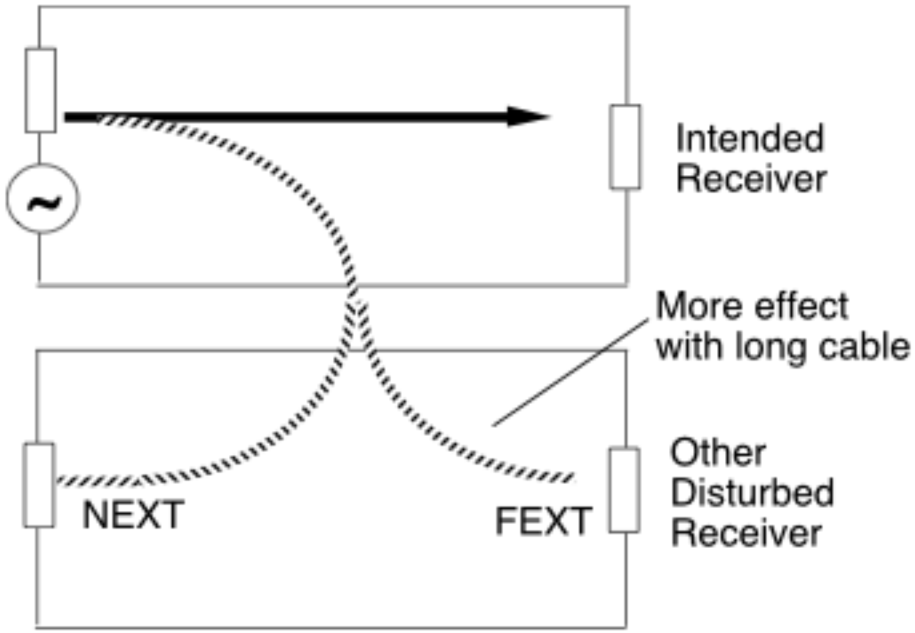
Typical Use of 10BT within an Office (max 100m each segment)
A maximum of 2 NICs per cable segment (repeaters/hubs are needed)

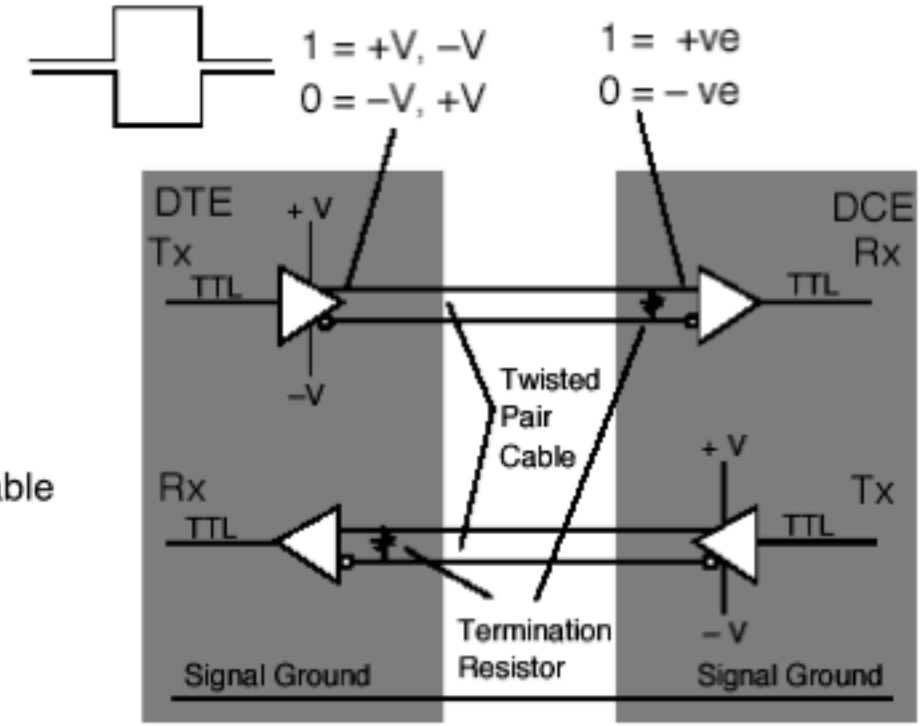
EIA/TIA TS 568 wiring

Pin	T568A Pair	T568B Pair	Signal	T568A Colour OLD	T568B/C Colour NEW
1	3	2	+	white/green	white/orange
2	3	2	-	green	orange
3	2	3	+	white/orange	white/green
4	1	1	-	blue	blue
5	1	1	+	white/blue	white/blue
6	2	3	-	orange	green
7	4	4	+	white/brown	white/brown
8	4	4	-	brown	brown



Cross Talk between cable pairs



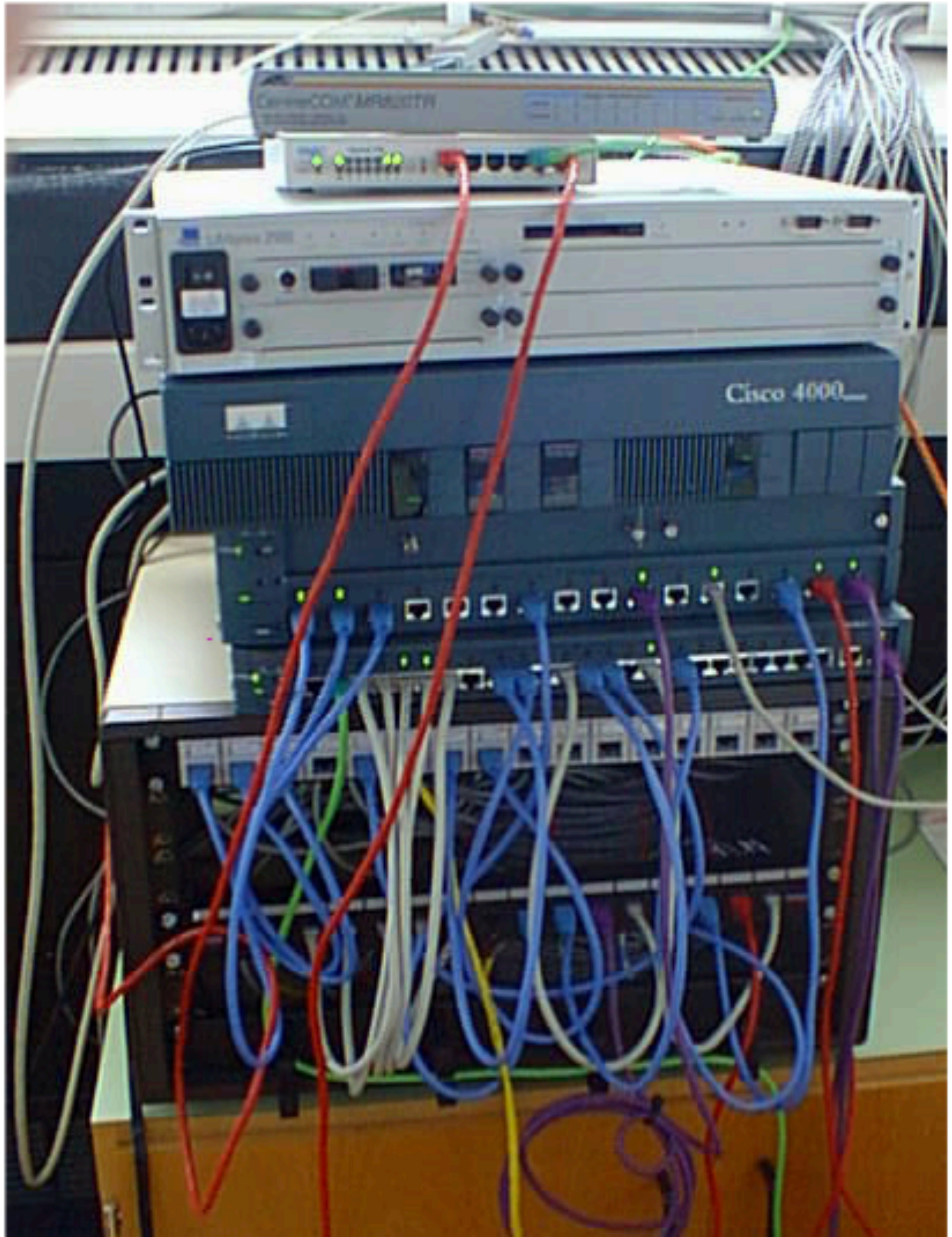


Often installed as a part of a bundle of cables

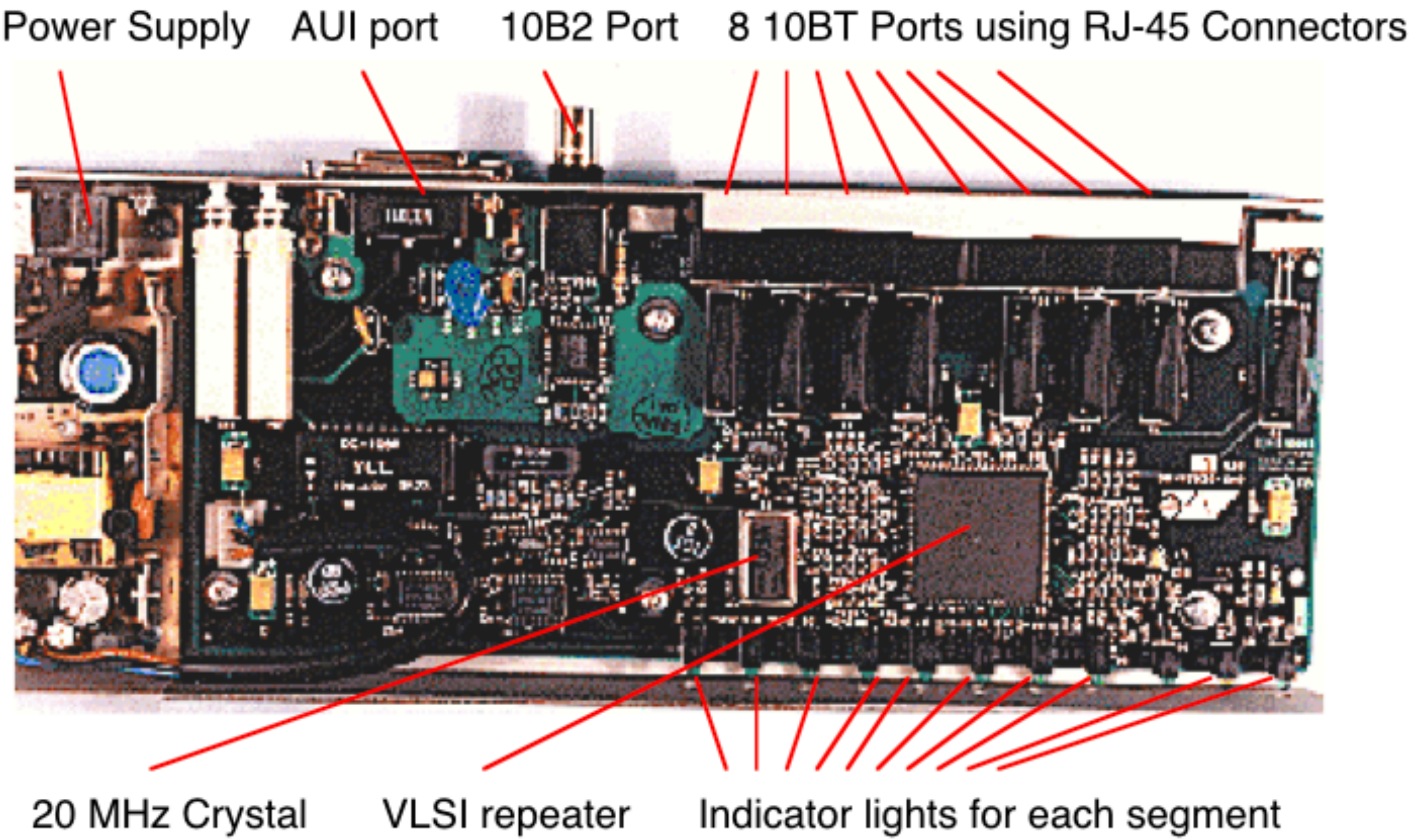
One pair radiates power to other pairs in the cabling budle

- NEXT - Mainly effect of near cable (design of RJ-45 connector)
- FEXT - Increases with cable length (100m maximum length)

10BT Equipment



10BT Hub


Power Supply
AUI port
10B2 Port
8 10BT Ports using RJ-45 Connectors
20 MHz Crystal
VLSI repeater
Indicator lights for each segment

10BT

- **Differential transmission using a balanced cable**

  RJ-45 connector
  Easy to manage / install

- **Cable flexible and very cheap**

  Unshielded Twisted Pair (UTP) specified by CAT 5

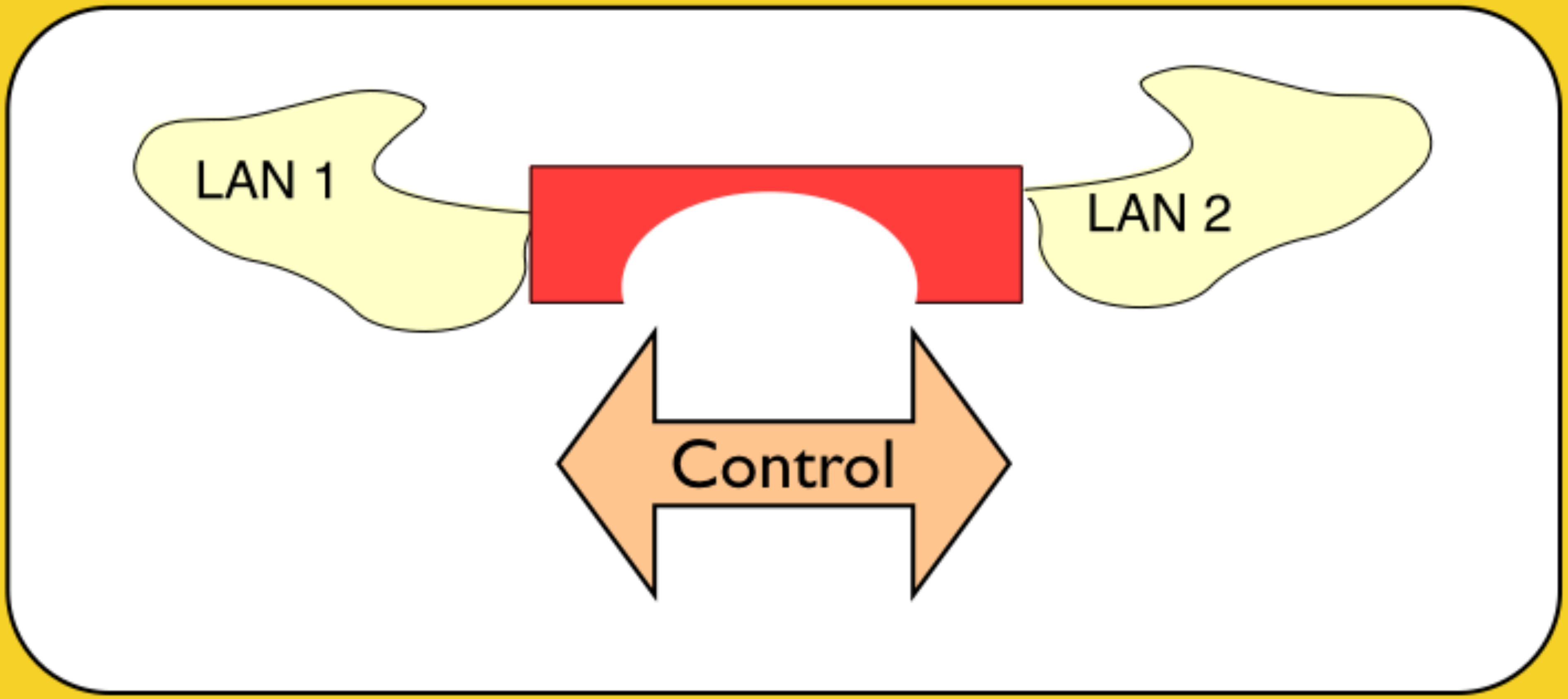
  More about other types of cable later….

[redacted]

Unshielded Twisted Pair Cabling

Module 4 Additional Video

Bridges & Switches:

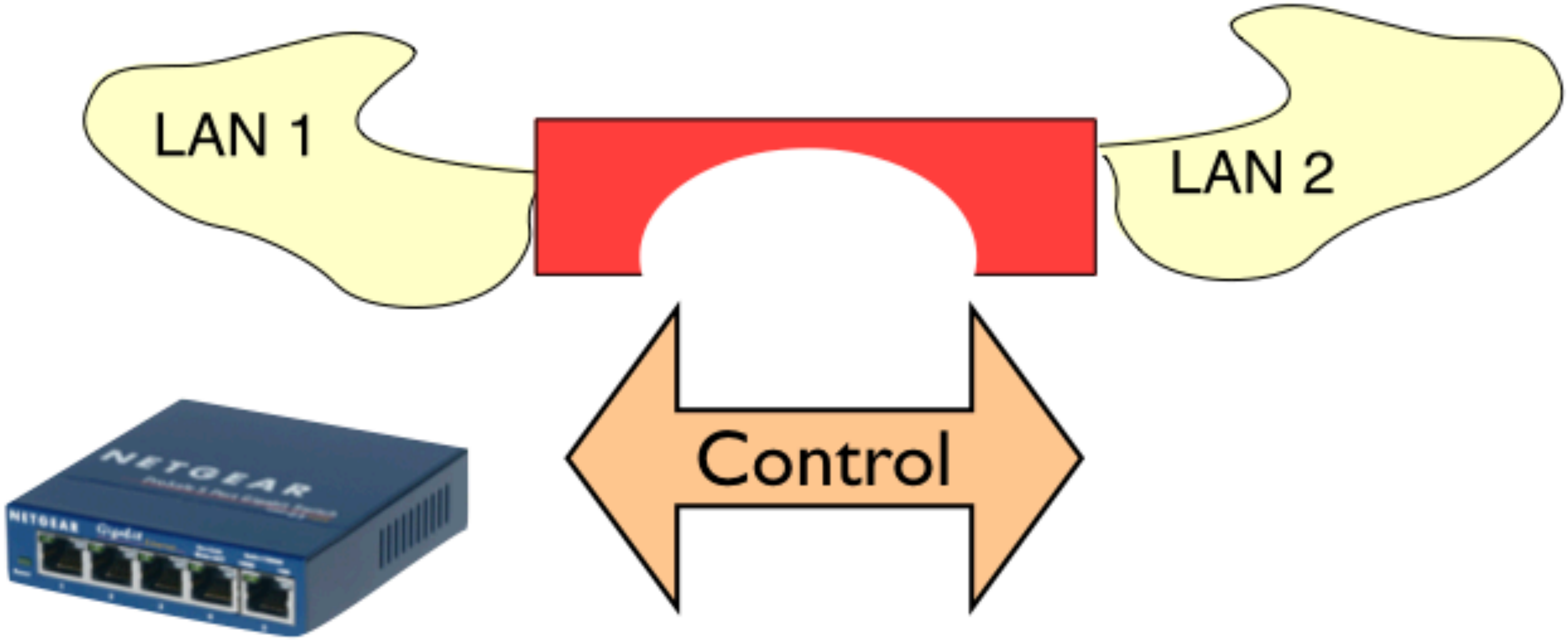


Module 5

Bridges & Switches:

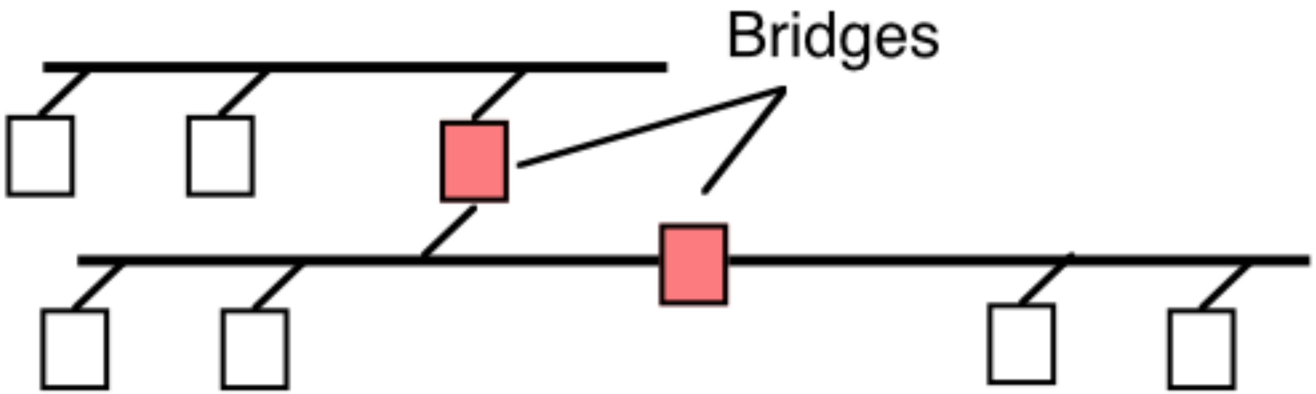
Building a Broadcast Domain from multiple Collision Domains

Forwarding using Address Tables



Module 5.1

When Do We Need A Bridge?



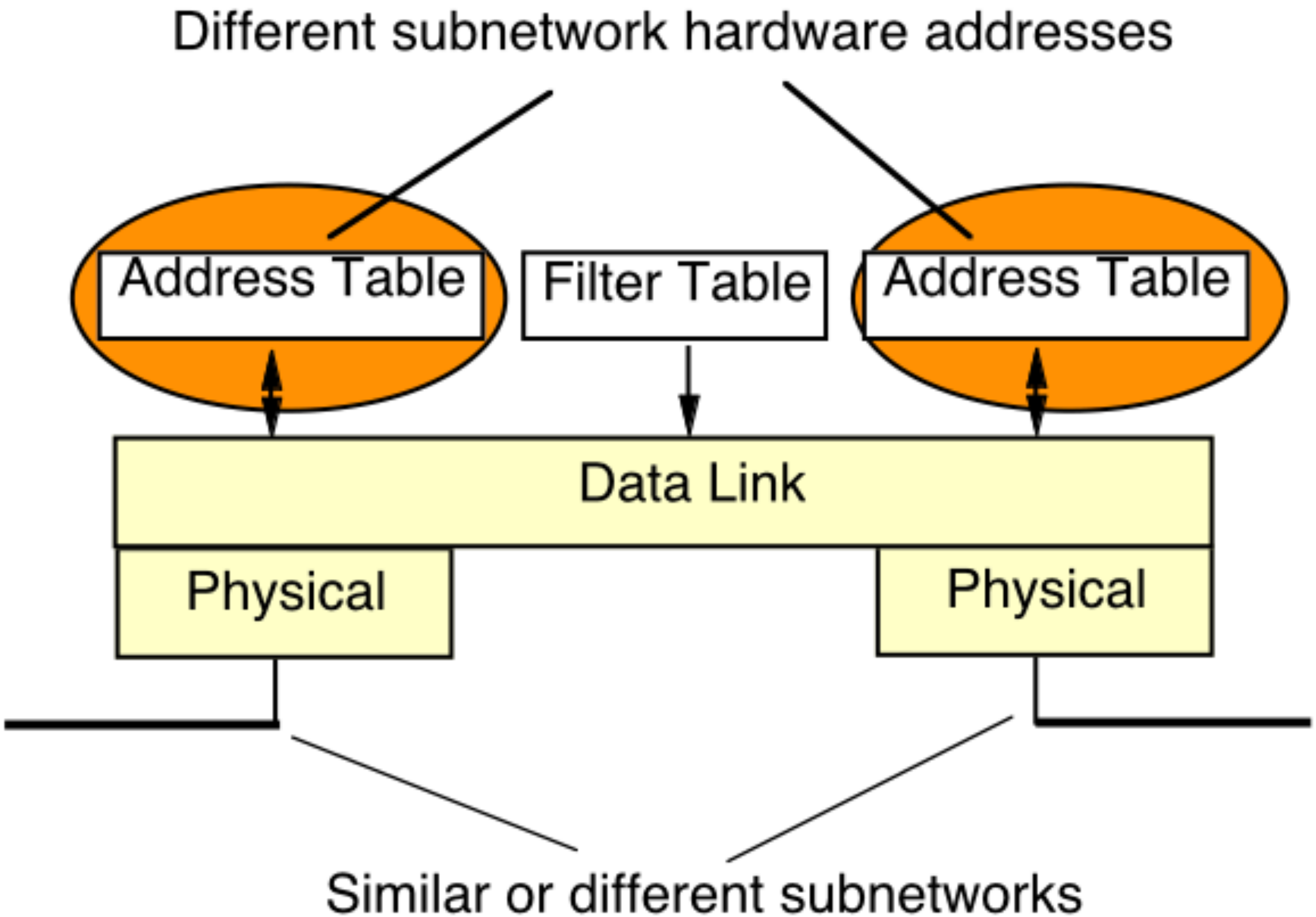
Bridges are needed to:

Connect > 1024 nodes
Extend total network diameter
Connect more than 5 segments in series

Bridges connect collision domains:

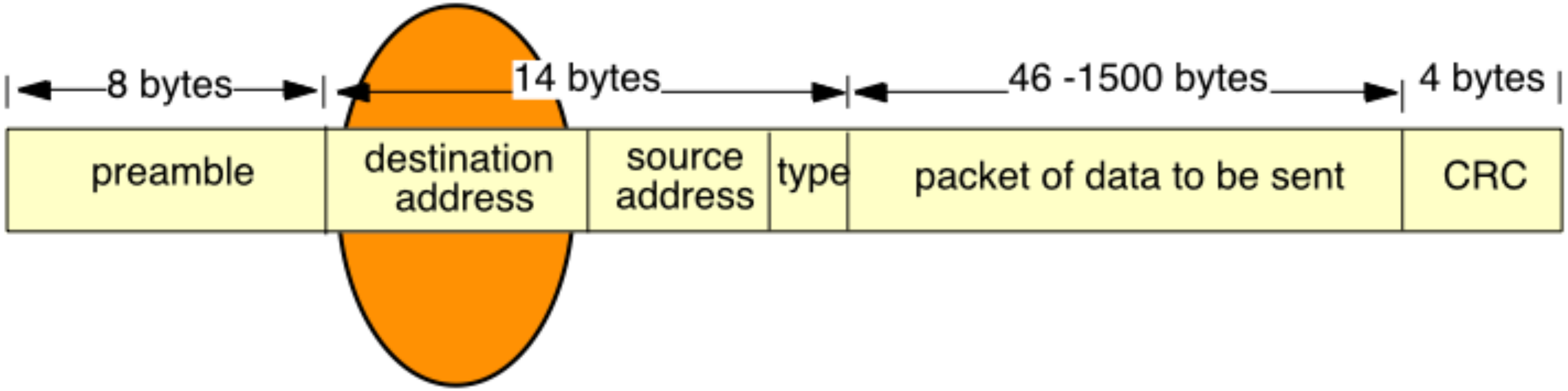
Increase maximum capacity of network
Deny unauthorised use of the network

Bridge



Different subnetwork hardware addresses
Address Table
Filter Table
Address Table
Data Link
Physical
Physical
Similar or different subnetworks

Use of the Ethernet Destination Address



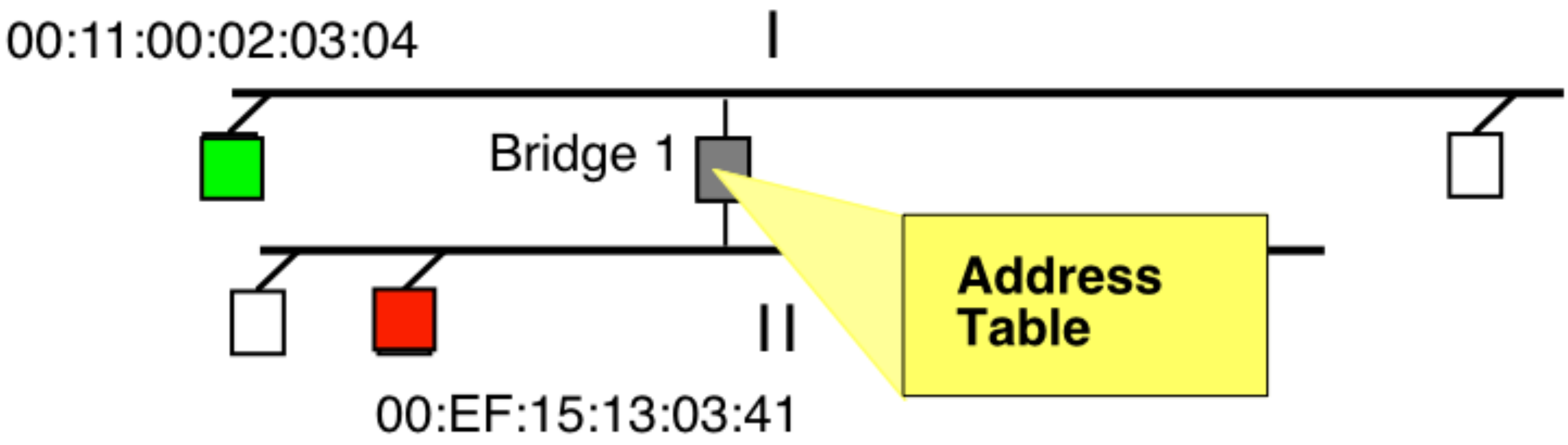
NIC inserts a destination address in each frame

Switches *"decide"* where and whether to forward the frame

- i.e. use the "topology" information in the address table

- switches decide this for ***each*** frame

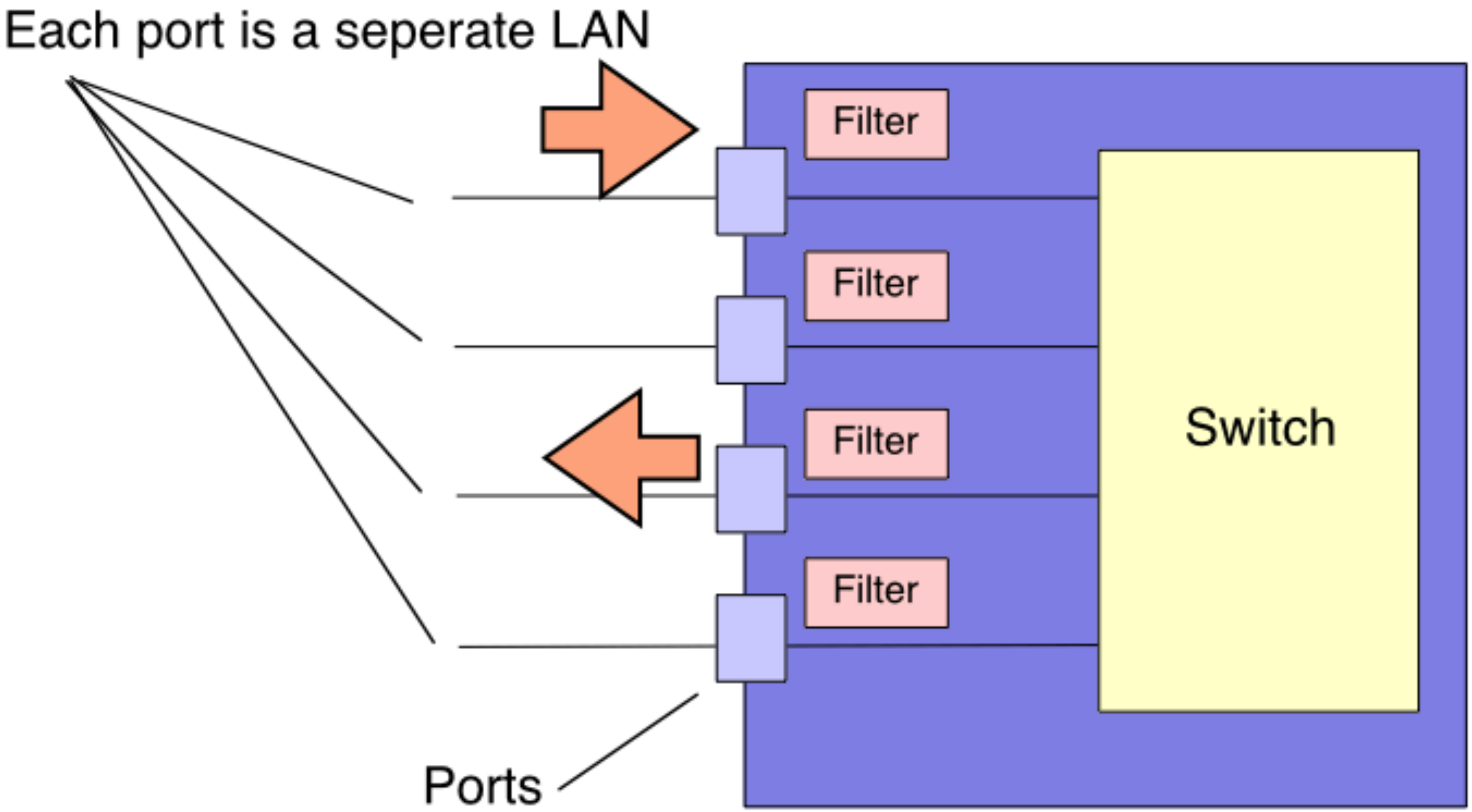
Address Table



One entry for each MAC Address, indicating port used

MAC Address	Static	Port
00:11:00:02:03:04	Yes	I
00:EF:15:13:03:41	Yes	II

Frames sent to "correct" port

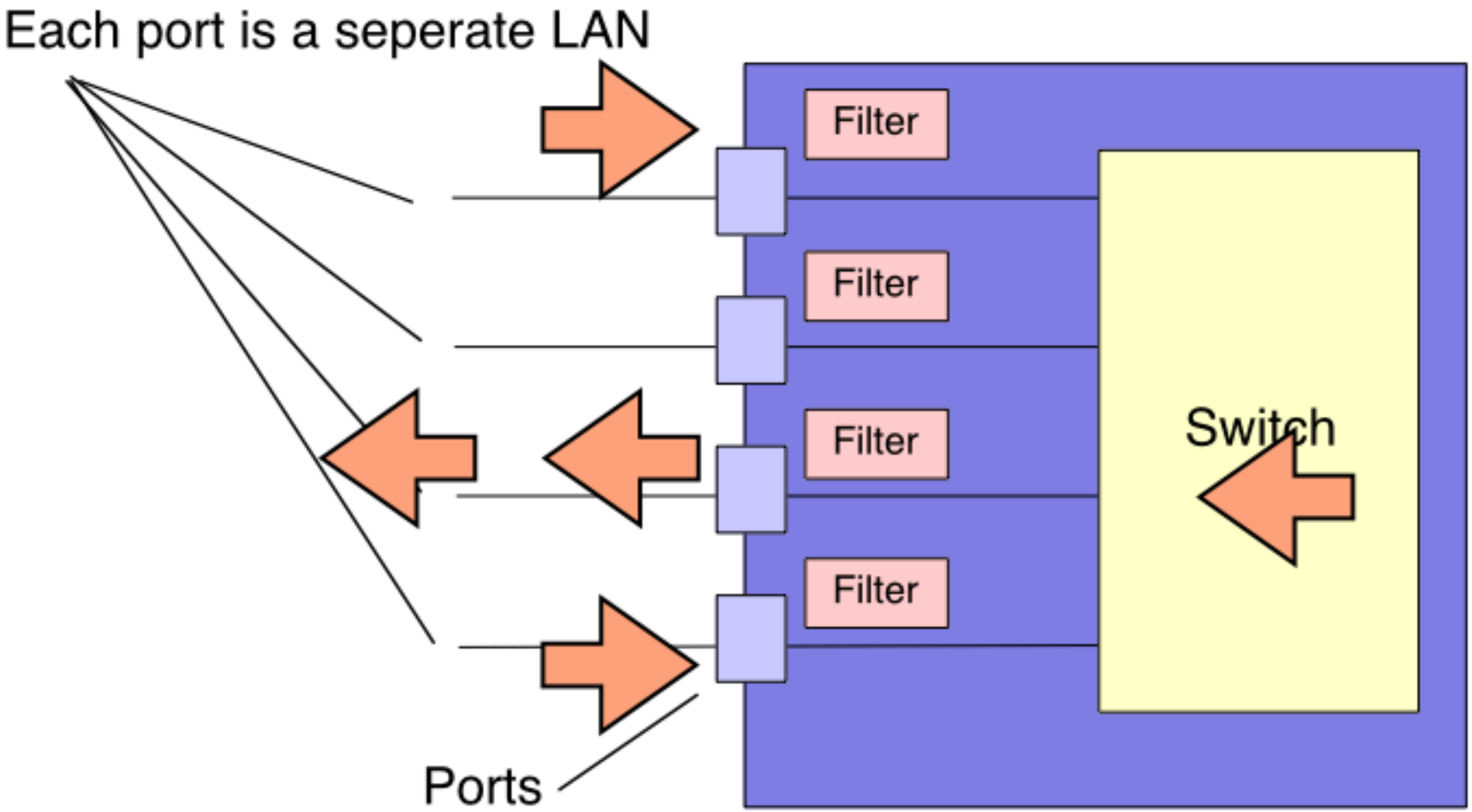


Frames are forwarded based on MAC destination address

They only need to be sent to the port that connects to a destination

The connected networks are called a "broadcast domain"

Frames are buffered within a Switch



Frames are buffered until they can be forwarded.

Forwarding (I)

Frames with an unknown destination are flooded

(when frame destination address is not in address table)

Sent to all ports except the receiving port

Broadcast frames are also “flooded”

Multicast also “flooded” (unless configured group addresses)

Unicast frames sent only to a destination is on another port

(when frame destination address is in the address table)

Sent only to the specific port listed in the table

(unless same as received)

Forwarding (II)

MAC Address	Static	Port
00:11:00:02:03:04	Yes	I
00:EF:15:13:03:41	Yes	II

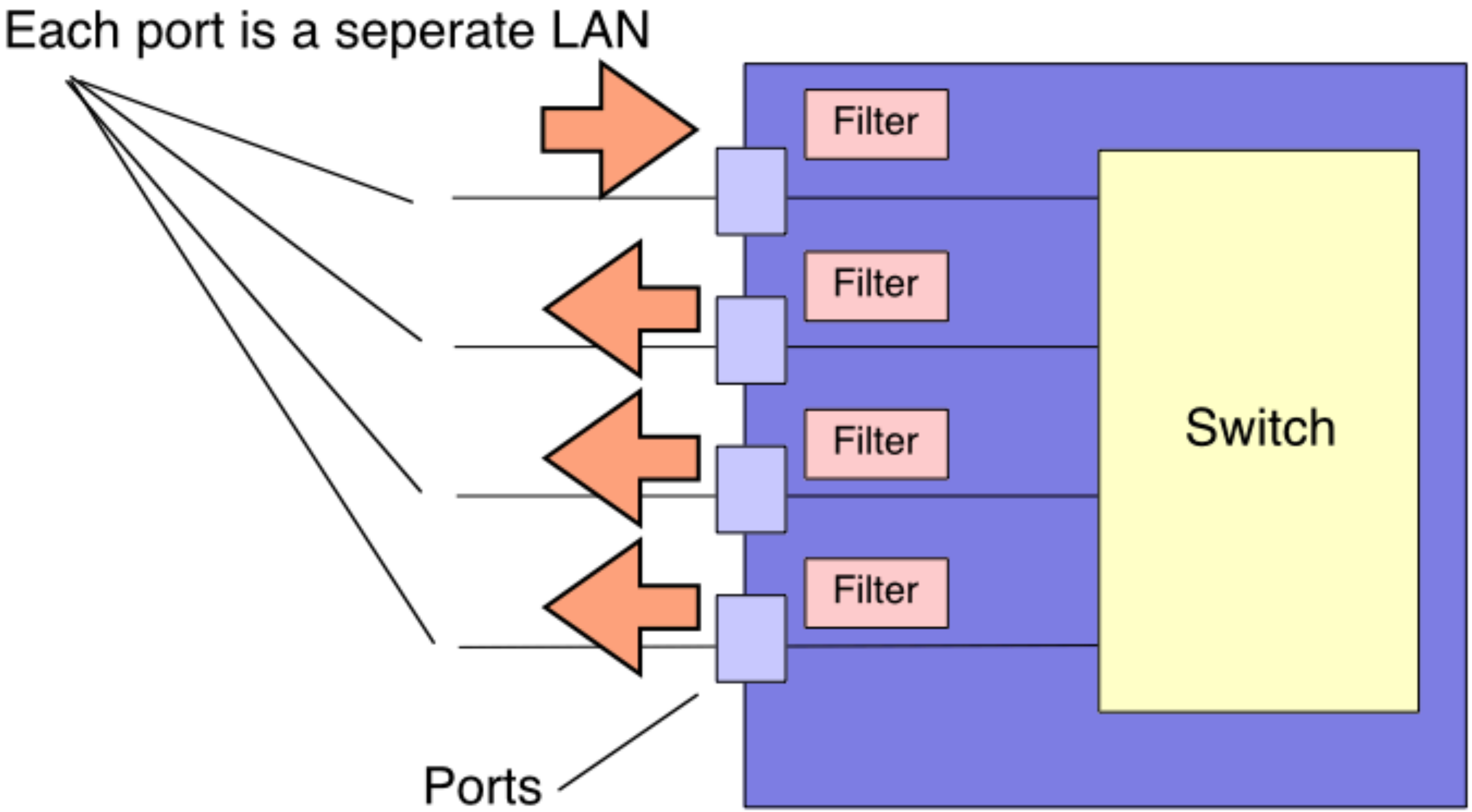
Is frame <u>destination</u> address in table?

- NO - forward to all ports EXCEPT incoming port (flood)
- YES - Look-up address and find table port

  Is table port == incoming port?

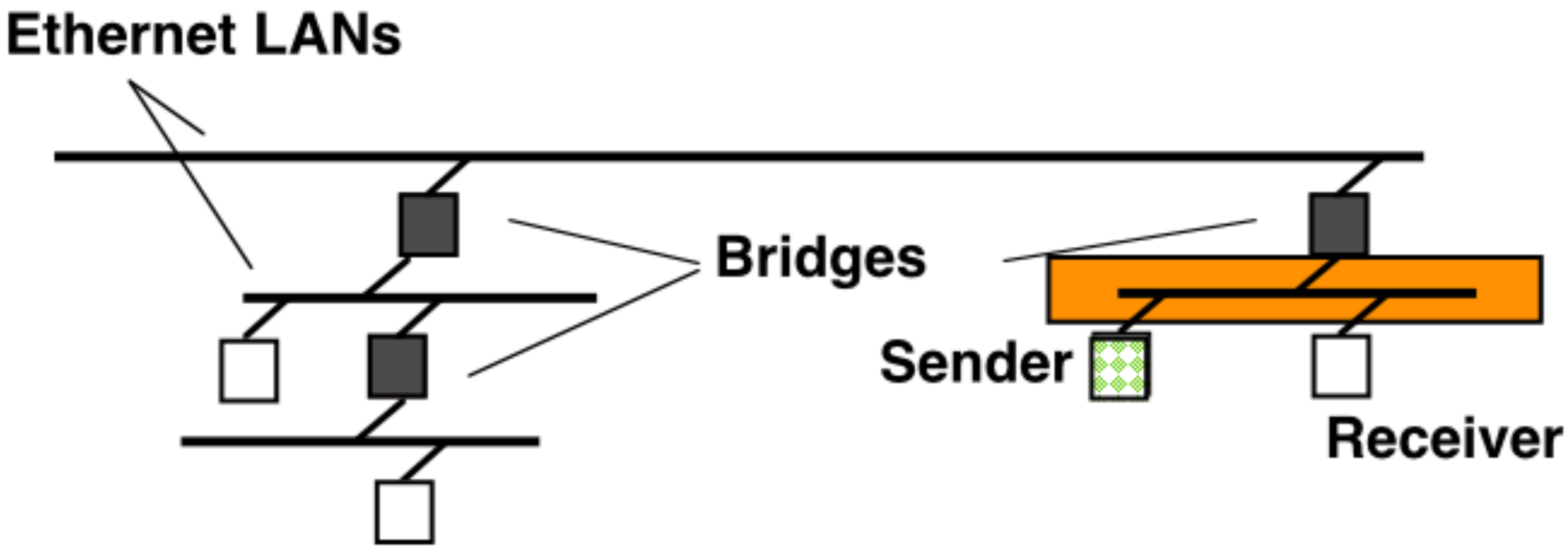
  - NO - forward only to table port
  - YES - discard the frame

Flooding addresses not in Address Table



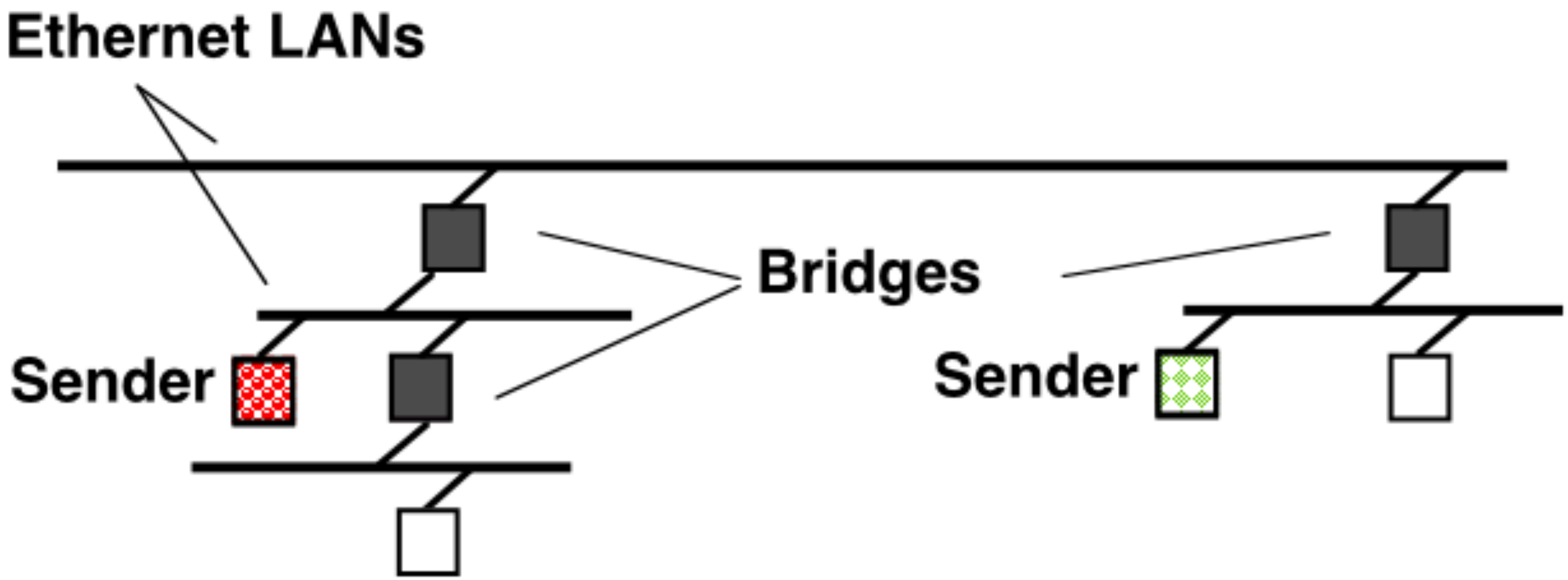
Frames are flooded if destination is not found in the address table
"Flooding" sends to all ports *except* the received port
(Almost the same as a "repeater/hub"!)

Example Network



Sender and Receiver on the same LAN segment

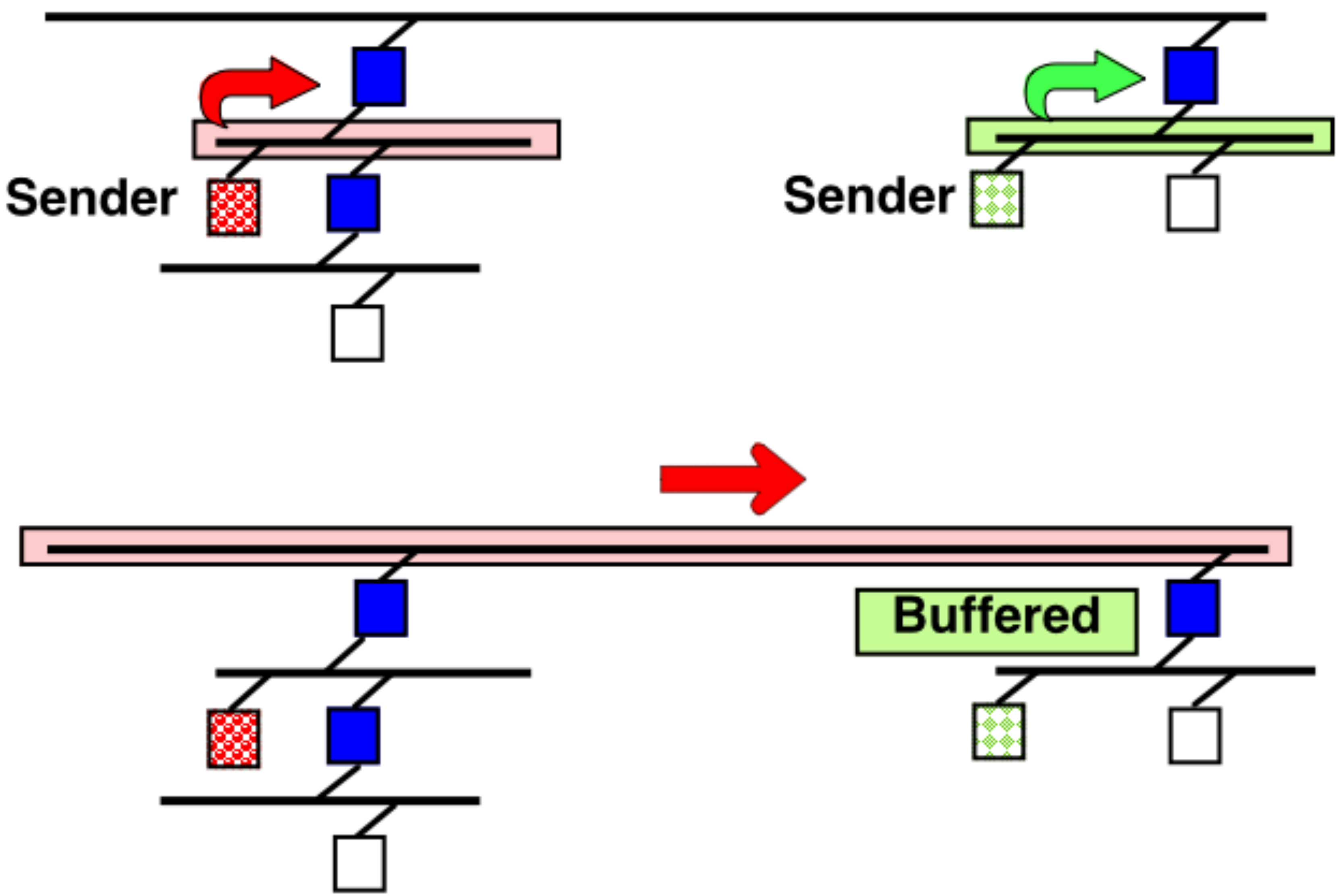
Example Network



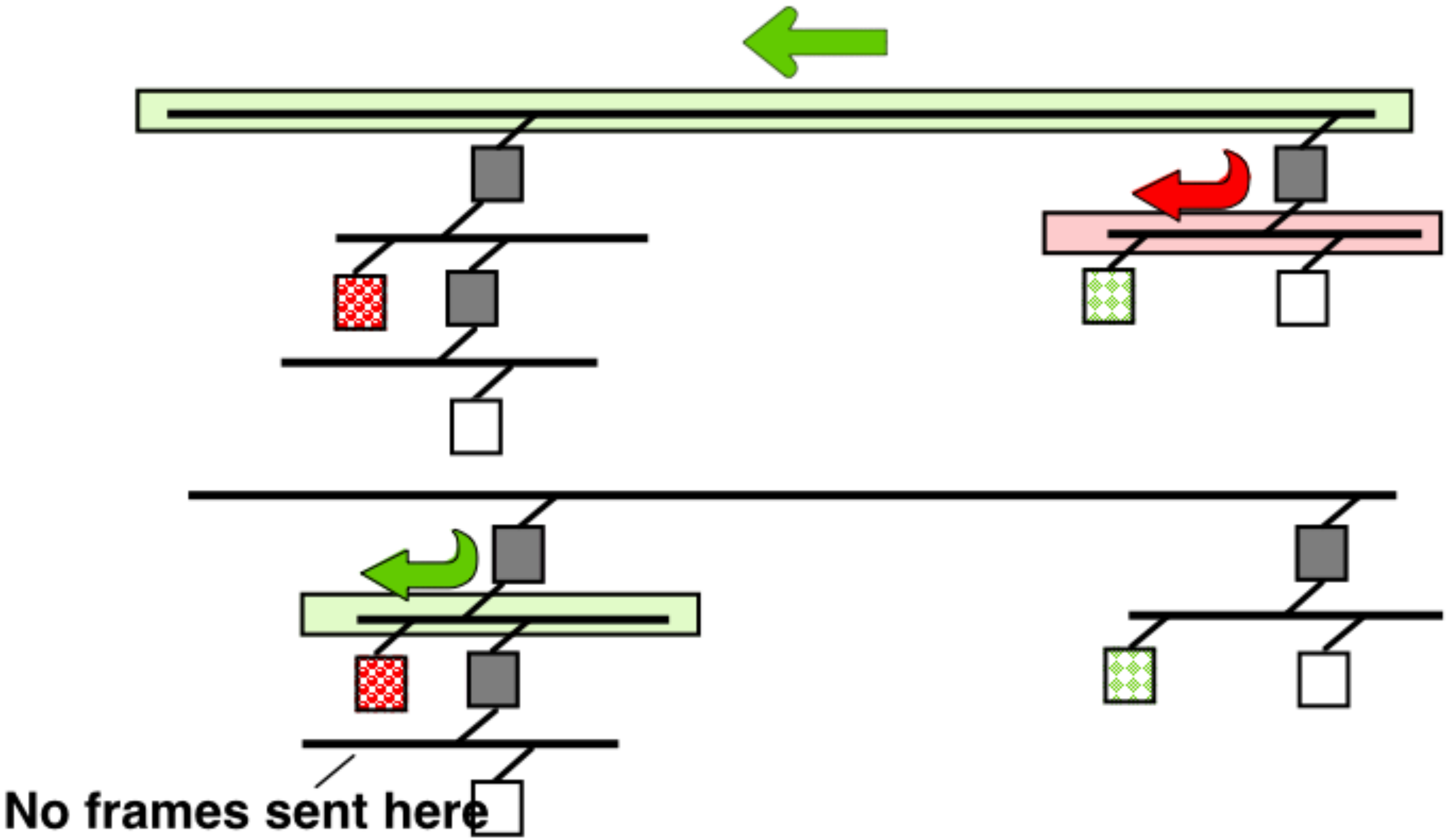
Assume both senders transmit at same time

red sends to green
green sends to red

Bridged Network (1)

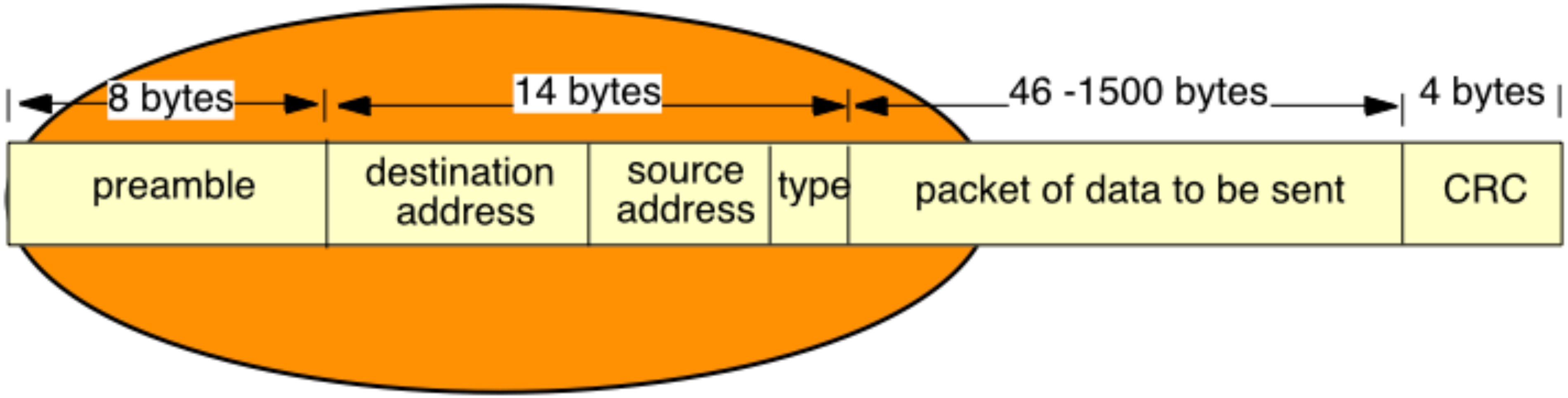


Bridged Network (2)



Separate collision domains

How many bytes need to be read?



The first 6 bytes identify the destination!

However, it is important to read at least first 64B

- collisions, result in frames etc less than 64B
- “runt” frames MUST NOT be forwarded

Cut-Through Forwarding

Simple bridges receive a frame in full before forwarding

This lets the bridge check the frame is valid

Frame Header contains all addresses

Could start to forward as soon as 64 bytes are received

This eliminates some of the delay in storing data

1.2 ms lower transit delay!

Disadvantages

Could start to forward an oversize frame :-(

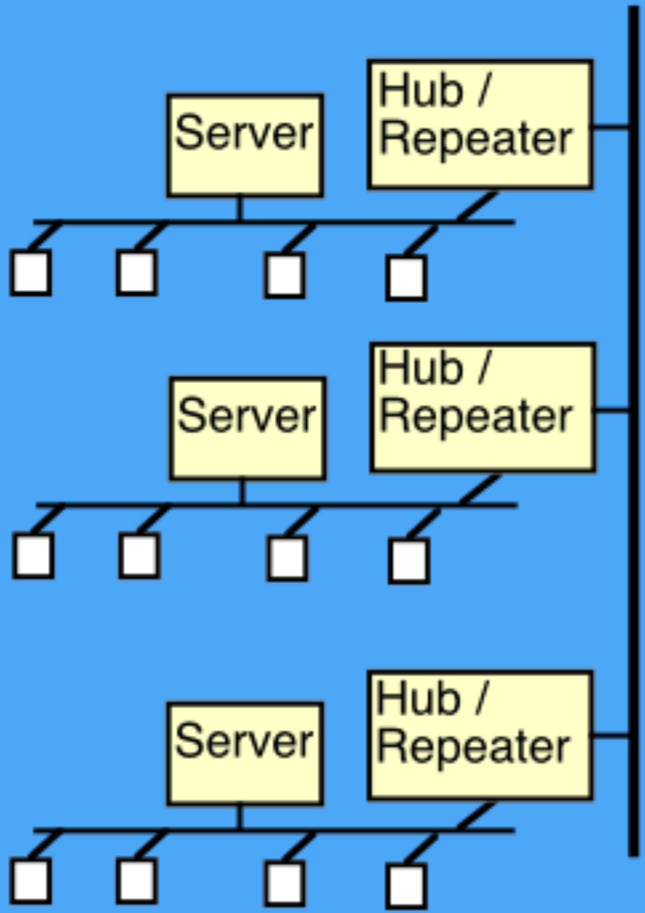
Could start to forward a frame with a bad CRC :-(

These frames are forwarded but the CRC is invalidated.

The destination will process and discard these as CRC errors

Known as "cut-through"

Enterprise Stage 1



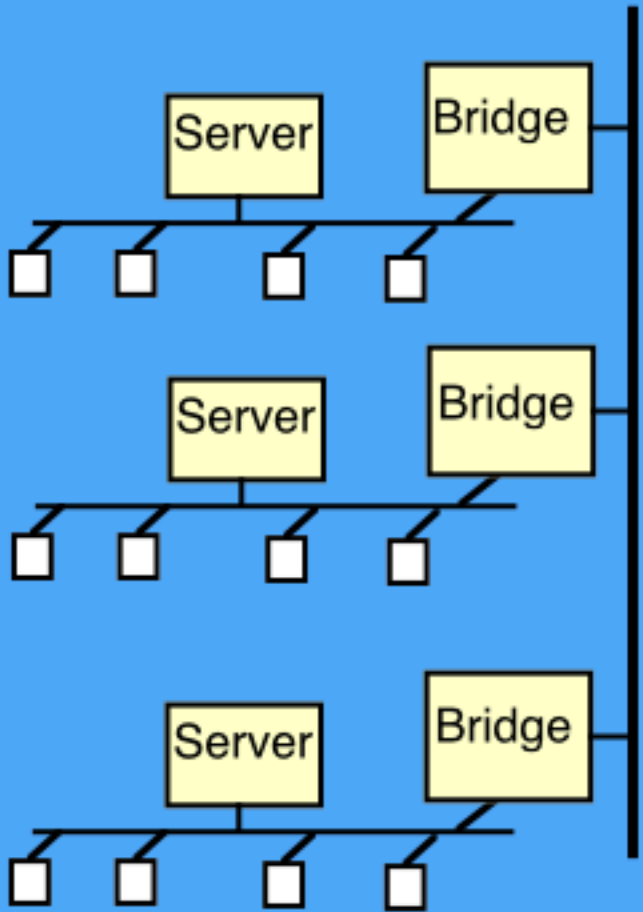
Each workgroup has own server

All connected via backbone
to form one network
(10B5 or 10BF with repeaters)

Repeaters / Hubs isolate faults

Enterprise Stage 2

Stage 2



Each workgroup has own server

Most traffic only local

LANs connected via backbone (typically 10B5 or 10BF)

Summary of Bridge Forwarding

- **A Bridge receiver operates in *promiscuous mode***
  (receives all frames ignoring destination address)

  A Bridge checks each received *frame*
  - Check length and CRC
  - Stores in internal memory
  - Cut-Through can forward before receiving CRC !
    - A bad CRC is forwarded if received - resulting in final discard

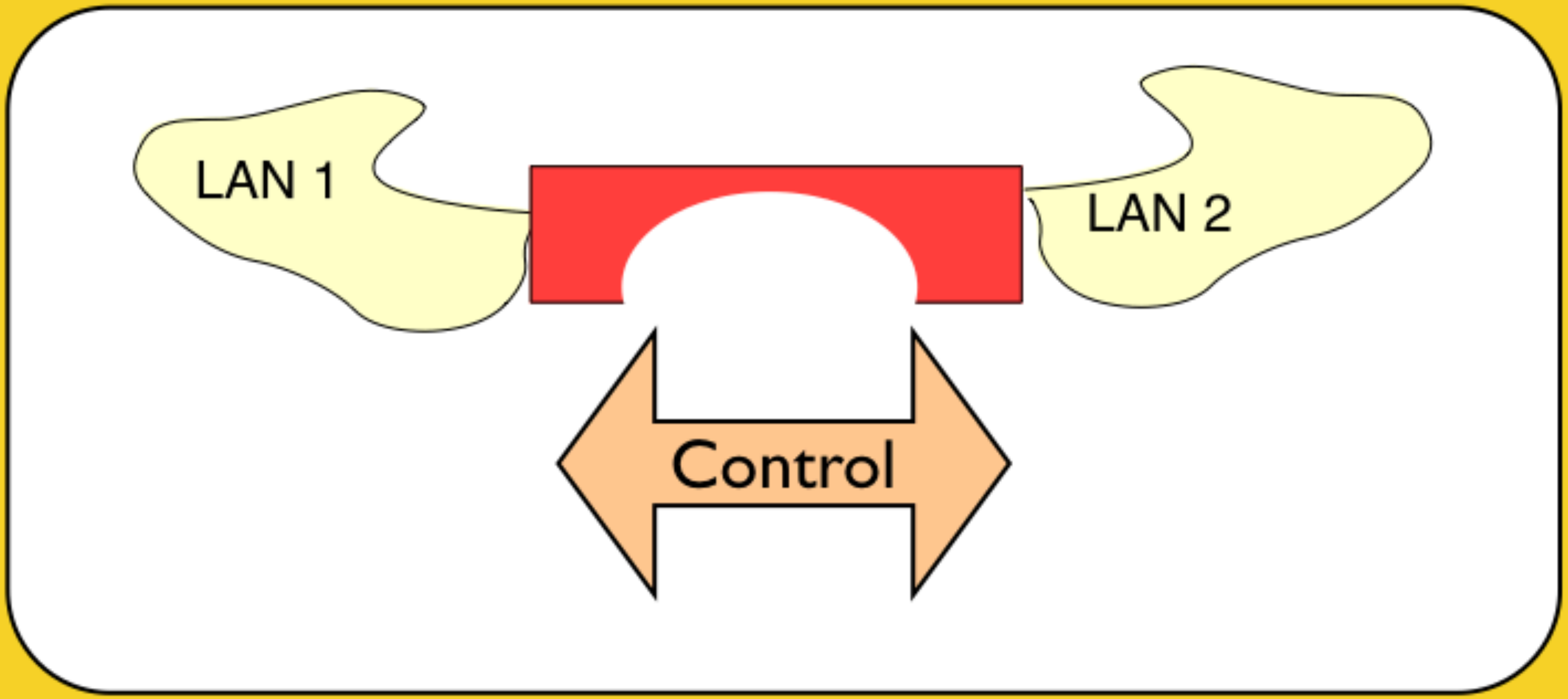
  Examines the *address table* for <u>destination</u> address
  - Forward if address matches <u>and</u> different port to output port
  - Discard if address matches <u>and</u> same port as output port
  - Otherwise, <u>flood</u> to all ports (except input)

  Examines the *filter table* for an address match
  - Discard if an address matches filter table
  - Could also send “traps” to alert a network manager

Bridges are “smart”

Bridges & Switches:

Dynamic Learning of Addresses



Module 5.2

Static Entries in tables

Static Tables are fine....

Can also fix the MAC address to a specific port

(e.g., useful in “public areas” to prevent hacking)

BUT!

Someone needs to keep address tables correct

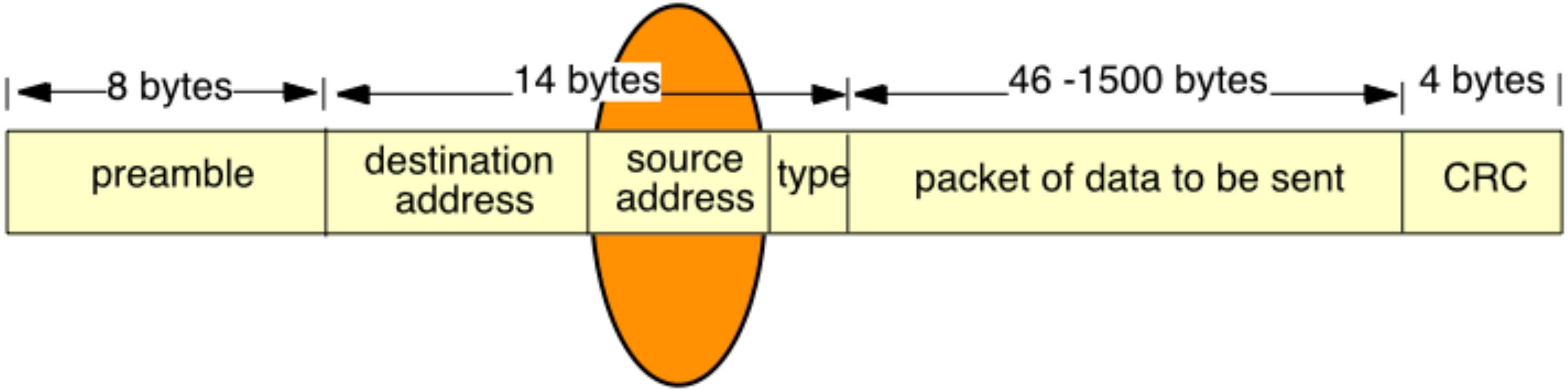
Address Table usually generated automatically

Makes bridges “Plug & Play

Difficult to track 100’s, 1000’s of addresses

An automated method is required...

Use of the Ethernet Source Address



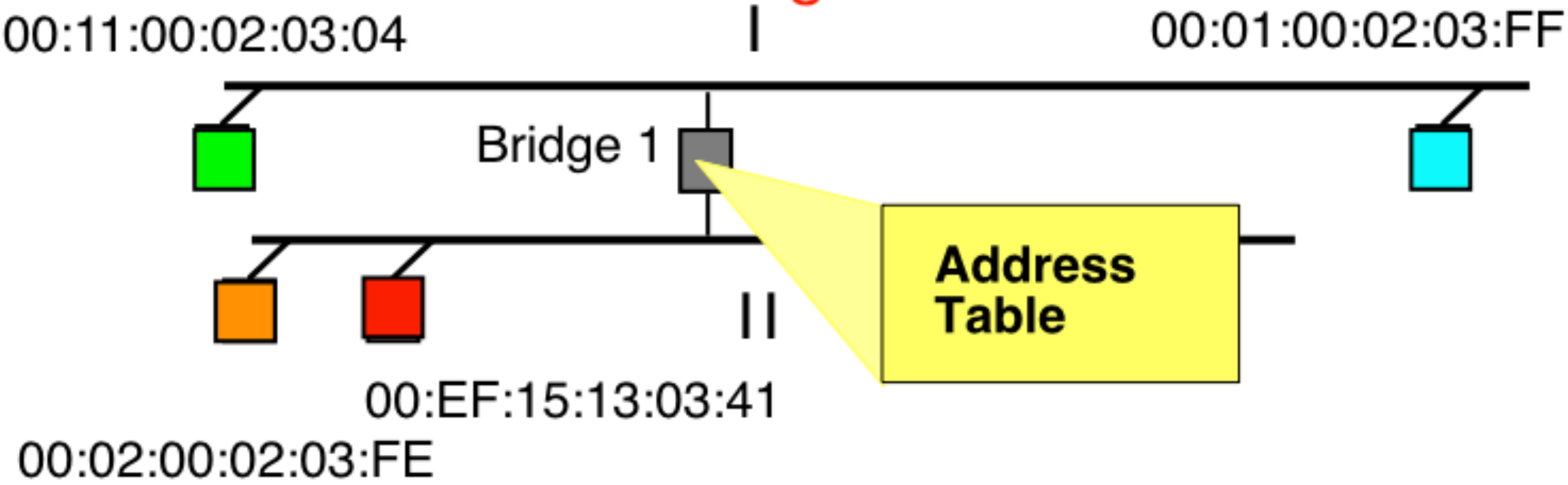
NIC inserts its *own* address in each frame!

Switches can now *"see"* where a source is

- i.e. dynamically assign port & MAC in address table

- actually switches do this for ***every*** frame

"Learning" entries in the Address Table



Entries made for each new (unicast) MAC Address

MAC Address	Static	Port
00:11:00:02:03:04	Yes	I
00:EF:15:13:03:41	Yes	II
00:01:00:02:03:FF	No	I
00:02:00:02:03:FE	No	II

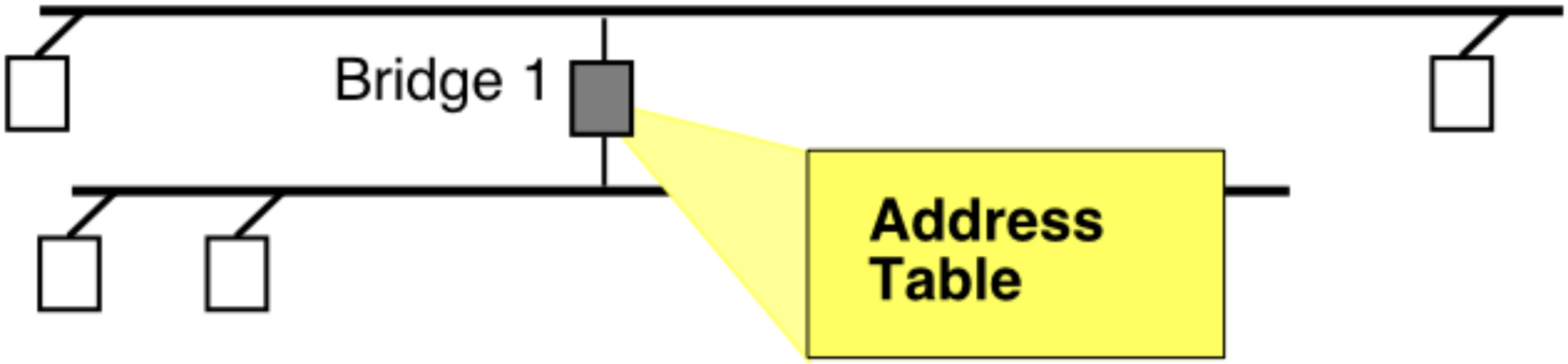
Dynamic Learning of Addresses in the Table

MAC Address	Static	Port	Expires
00:11:00:02:03:04	Yes	I	never
00:EF:15:13:03:41	Yes	II	never
00:01:00:02:03:FF	No	I	2 secs
00:02:00:02:03:FE	No	II	3 mins

Each entry is "aged"
old entries are deleted.

Age updated as frames arrive from a src address
Each second, all ages reduce
Zero entries are deleted

Denial Attack on the Address Table

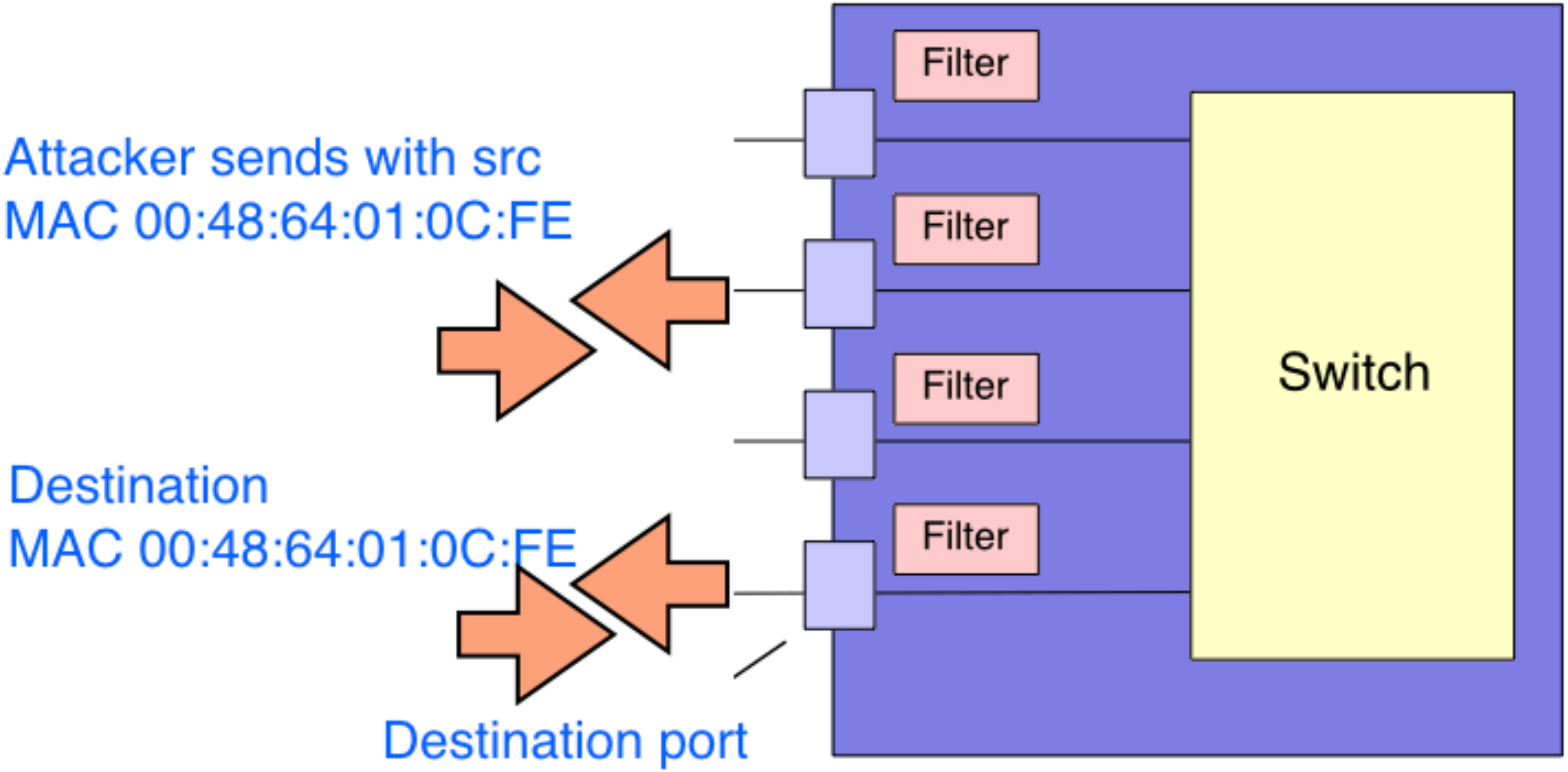


A denial-of-service attack could be made against a MAC address

Suppose a malicious computer sends packets with another computer's source address (there are programs that do this)

Updates address table (preventing destination receiving traffic)

Denial attack on Address Table



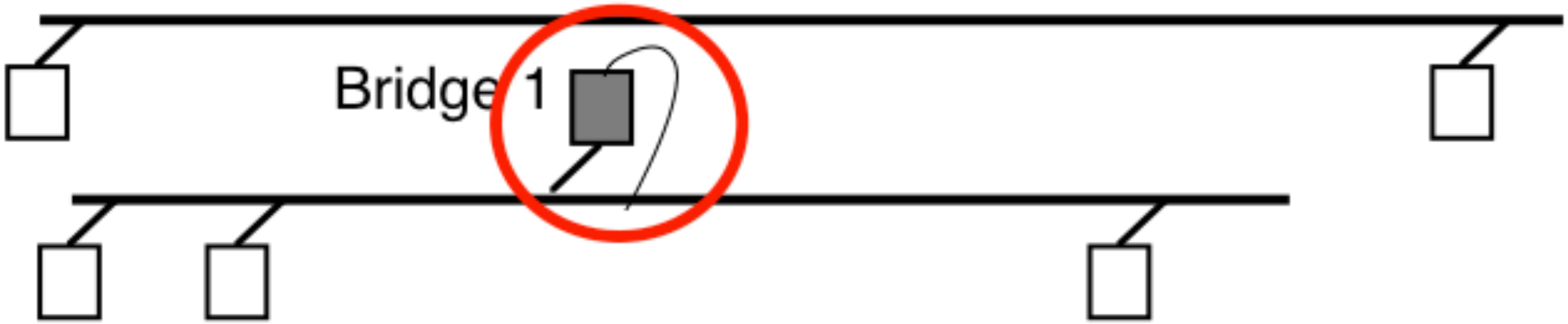
Attack updates Address Table, stealing traffic from intended destination

Attacker must keep doing this,

(the real destination will also update the table next time it sends)

Managed switches can detect this attack

Idiot-proof plug&play?

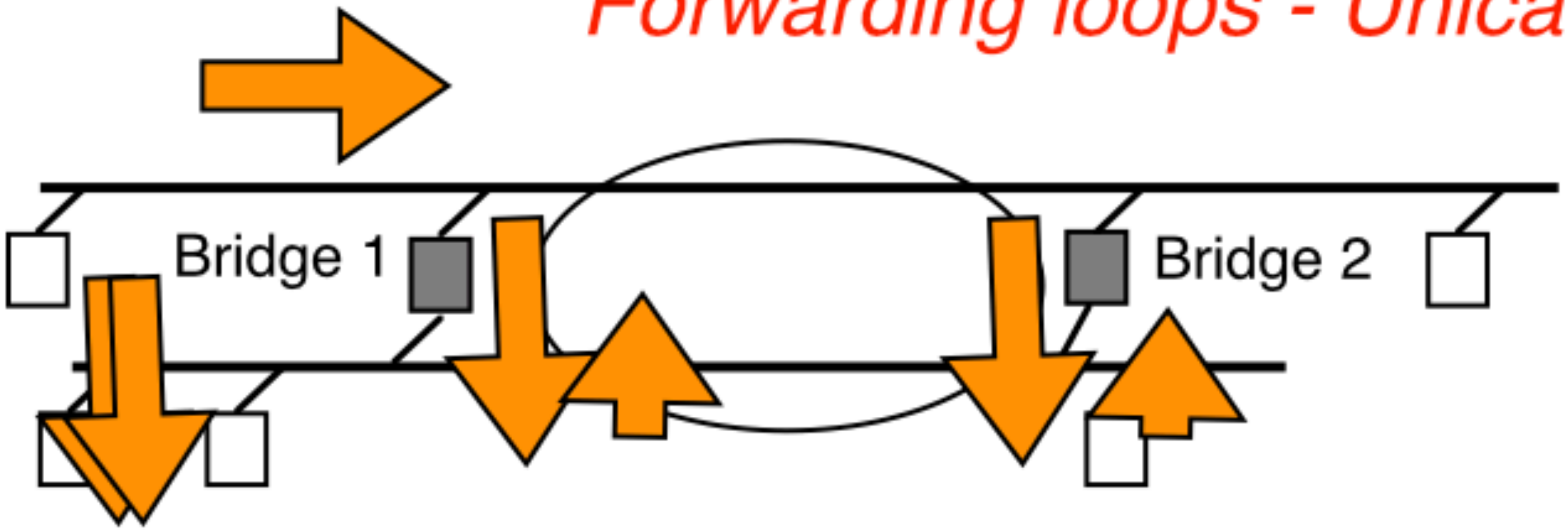


Connecting two networks needs a bridge

First deployed bridge did not work :-(

You need to connect a port to each network :-)

Forwarding loops - Unicast

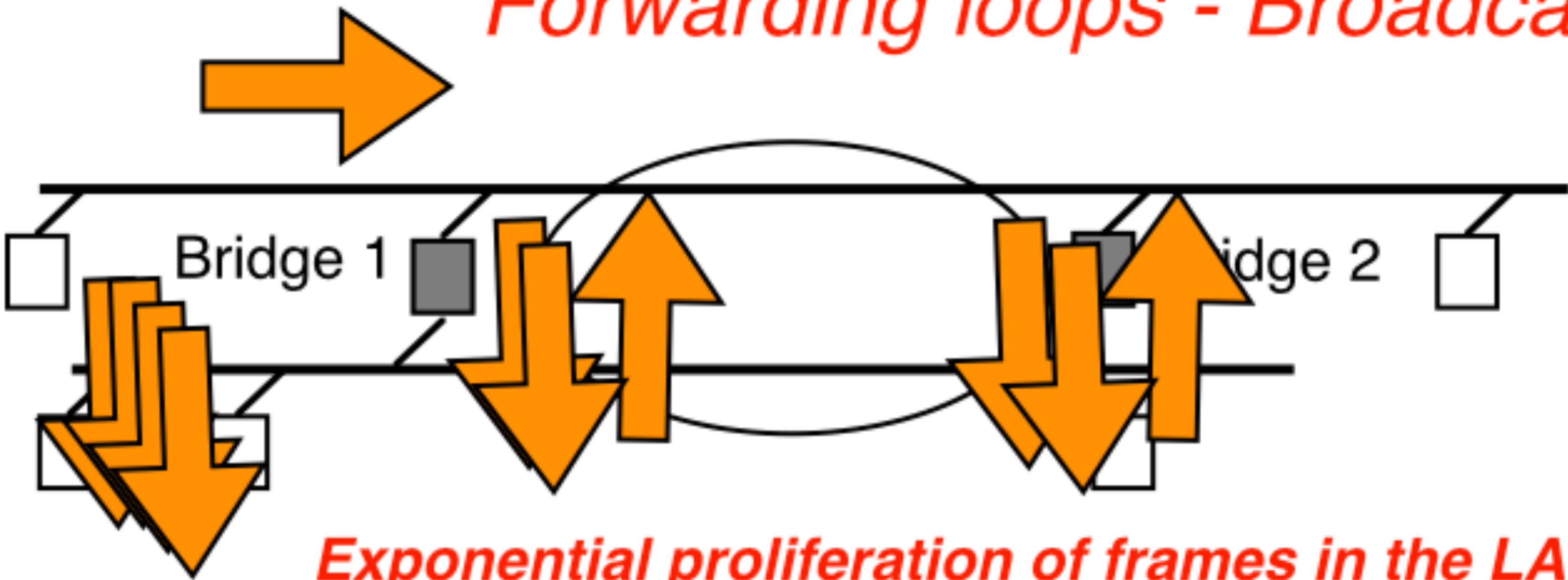


Connecting two bridges in parallel may cause duplication of unicast frames

Can cause incorrect learning of source address - and black-holing of frames.

Bridges MUST NOT forward in loops!

Forwarding loops - Broadcast



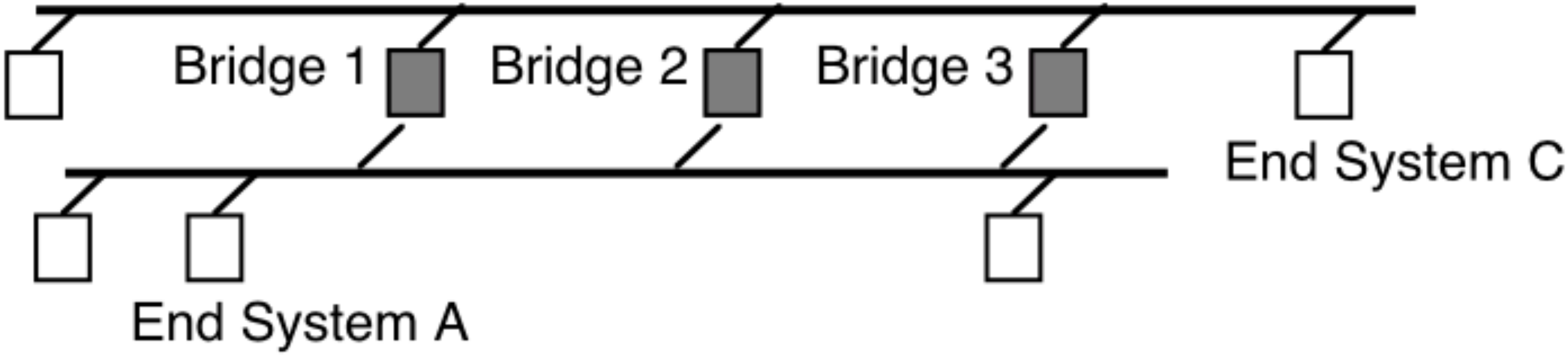
Exponential proliferation of frames in the LAN

Connecting two bridges in parallel may cause looping of broadcast (or flooded) frames

Bridges MUST NOT forward in loops!

The Spanning Tree Algorithm (STA) provides
an automatic way to ensure this (not in current course!).

Loops between bridges/switches?



A sends to C

Bridges 1,2,3 receive the frame

Bridges 1 forwards the frame, Bridges 2,3 receive the frame

Bridges 2,3 also forward a copy of the frame

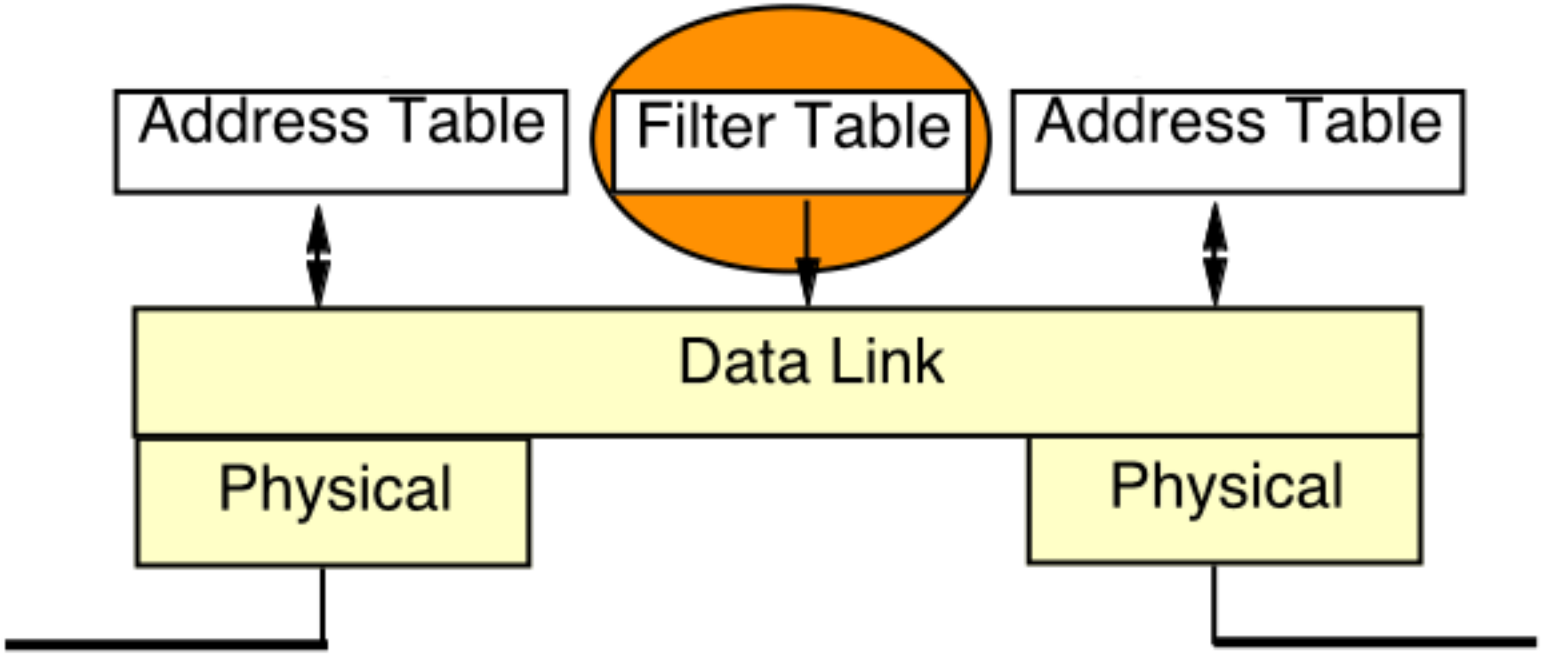
There are now three frames that have been forwarded

Bridge/Switch Filter Table

Filter Table can be used to set policies

Prevent frames from being forwarded to specific ports

Log/track users as they use the network



Bridges also check filter table BEFORE forwarding

Discard if address matches a filter table entry

May also send “traps” to alert network manager

Thinking about the Address Table

Things to think about:

An end system that ***only listens*** (never sends)
- Frames are broadcast to all ports
- Could configure a static entry

An end system is ***turned off***
- Address entry will age and be deleted

An end system ***moves*** to another collision domain
- Bridge will have learned the wrong port
- End system will not receive unicast frames
- Entry updated when end system sends

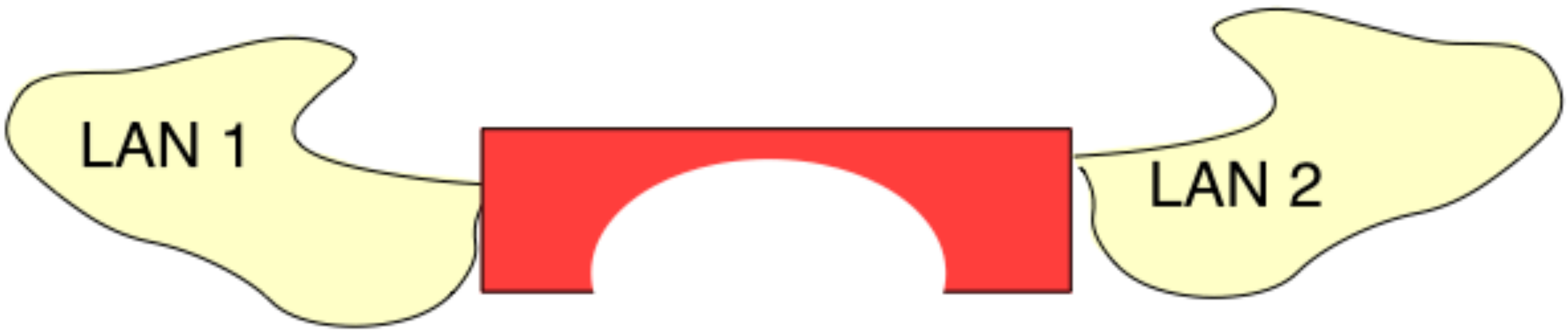
Summary of Bridge Learning

- **Bridges Learns form *Source Address* of Frames**
  - Receiving frames creates a dynamic entry in Address Table
  - Address is associated with ***port on which frame received***
  - Dynamic **a**ddress entries ***aged*** (old entries will be deleted)
  - Unknown destination addresses are flooded
- **Simple Plug and Play**
  - Must not be connected to form loops!
- **Examines *filter table* for an address match**
  - Discard frame if it matches an entry in the filter table
  - May also send "traps" to alert network manager
  - Can also send the "frame contents!"

Bridges are "smart"

Bridges & Switches:

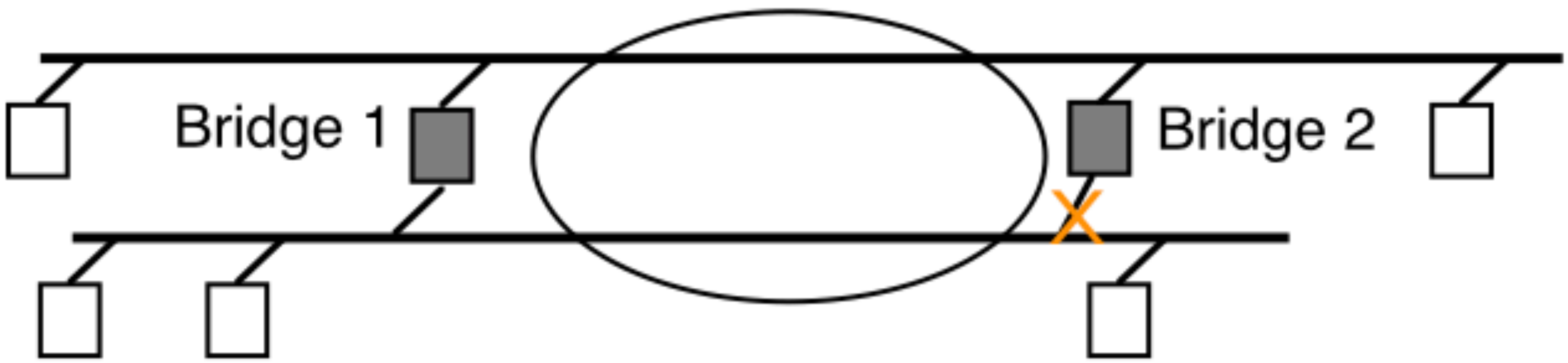
The Spanning Tree Algorithm



IEEE 802.1D - a method for managed switches to detect parallel paths and preventing looping

Module 5.3

The Spanning Tree Algorithm



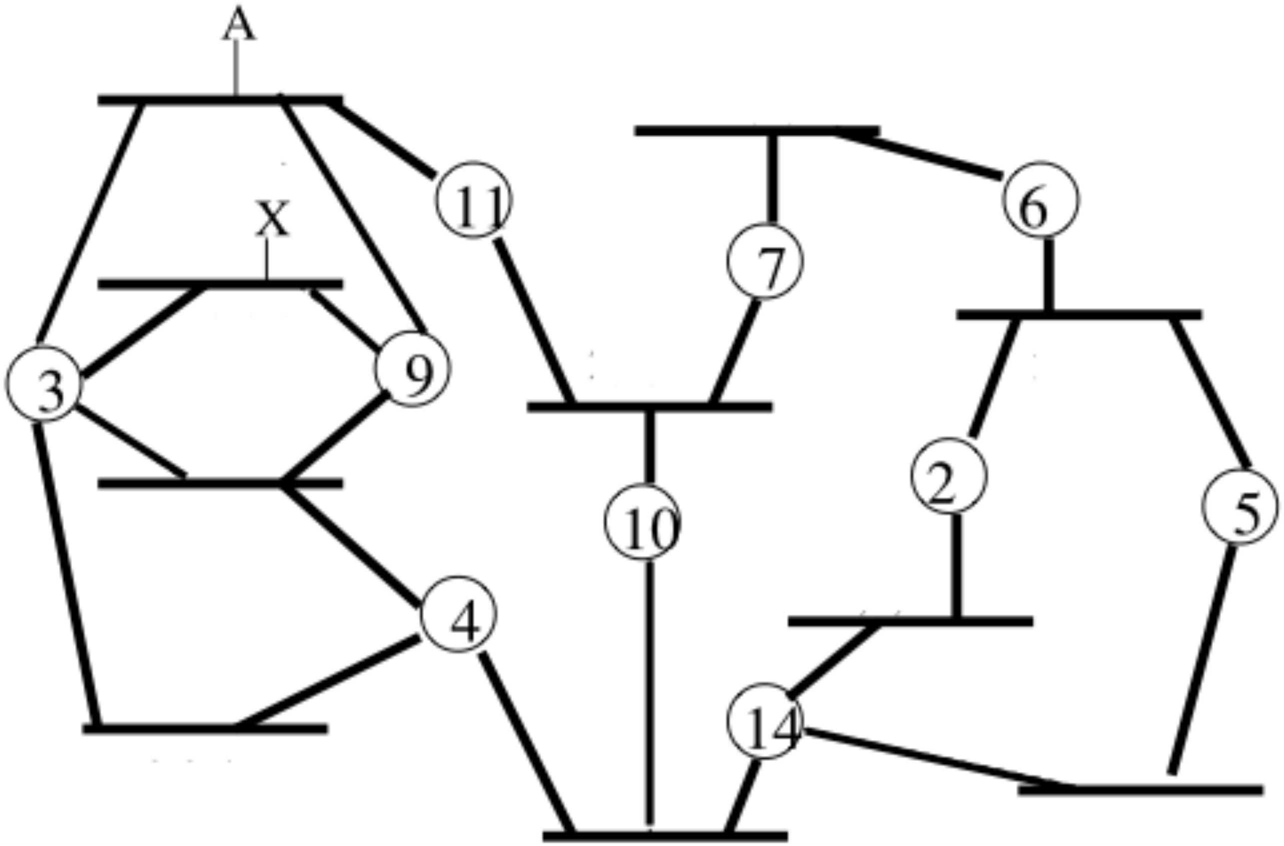
Connecting two bridges in parallel may cause looping

- You can avoid this by design

However, sometimes it is nice to have a "backup" link(s)

- It would be nice to do this automatically!

How do you choose a Forwarding Path?



Radia Perlman proposed the Spanning Tree Algorithm (STA)

It automatically elects one bridge as the root of the tree

STA then co-ordinates the other bridges to form a tree

Sending Bridge PDUs

Each bridge has a unique identifier

Bridge ID = {Priority : 2 bytes; MAC address: 6 bytes}

When a switch enables STP it ***multicasts**** BPDUs with the ID

e.g. once every 2 seconds by default on all ports

Any bridge running STA can receive these on its ports

*STA uses multicast group address 01:80:C2:00:00:00

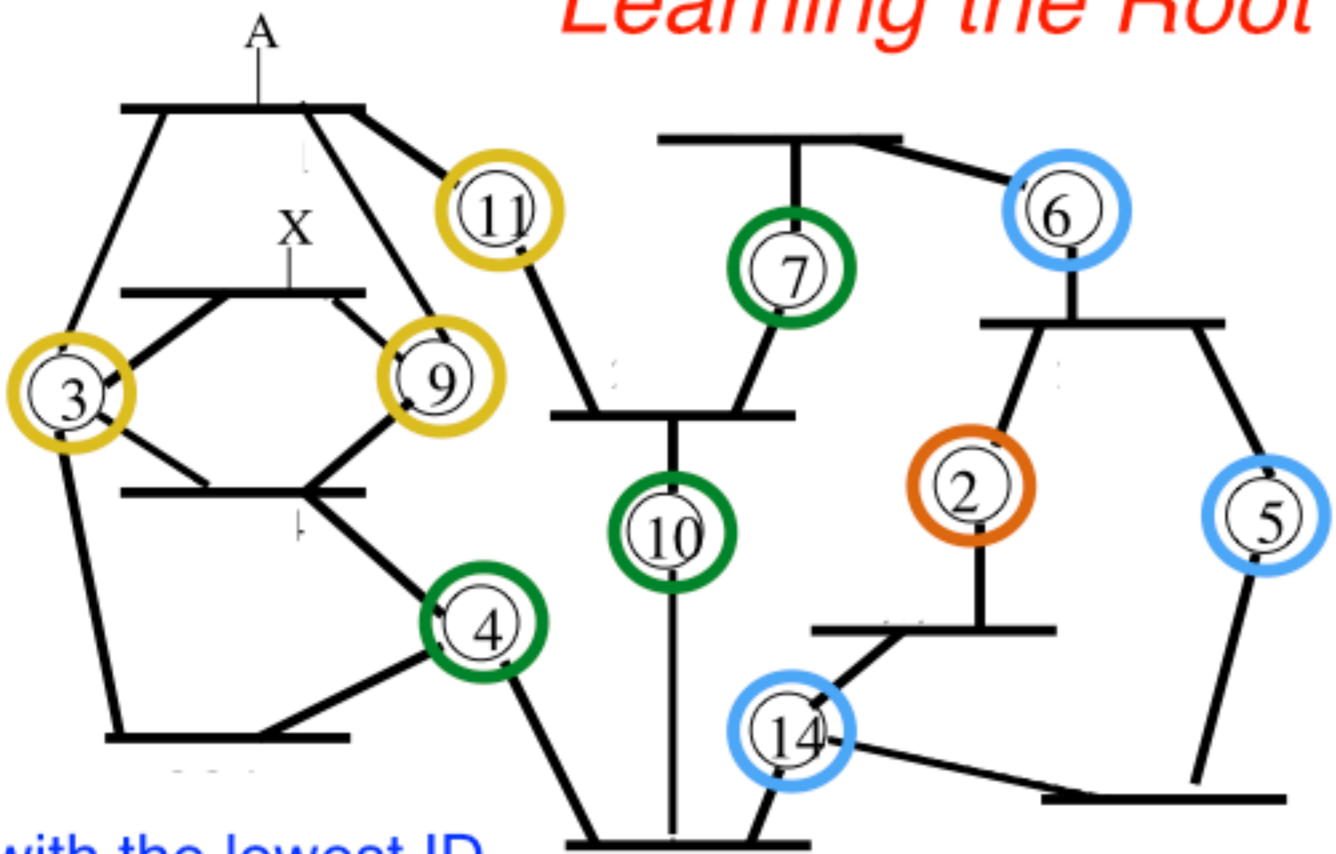
Receiving Bridge PDUs

Each bridge looks at the multicast BPDUs received on all ports

All bridges record for each port:
(The ***lowest ID*** seen;
the ***number of hops*** to the lowest ID;
the ***ID*** of the bridge sending the lowest ID)

This requires a bridge to keep about 50 bytes of data per port.

Learning the Root Bridge



The **root** is 2 with the lowest ID

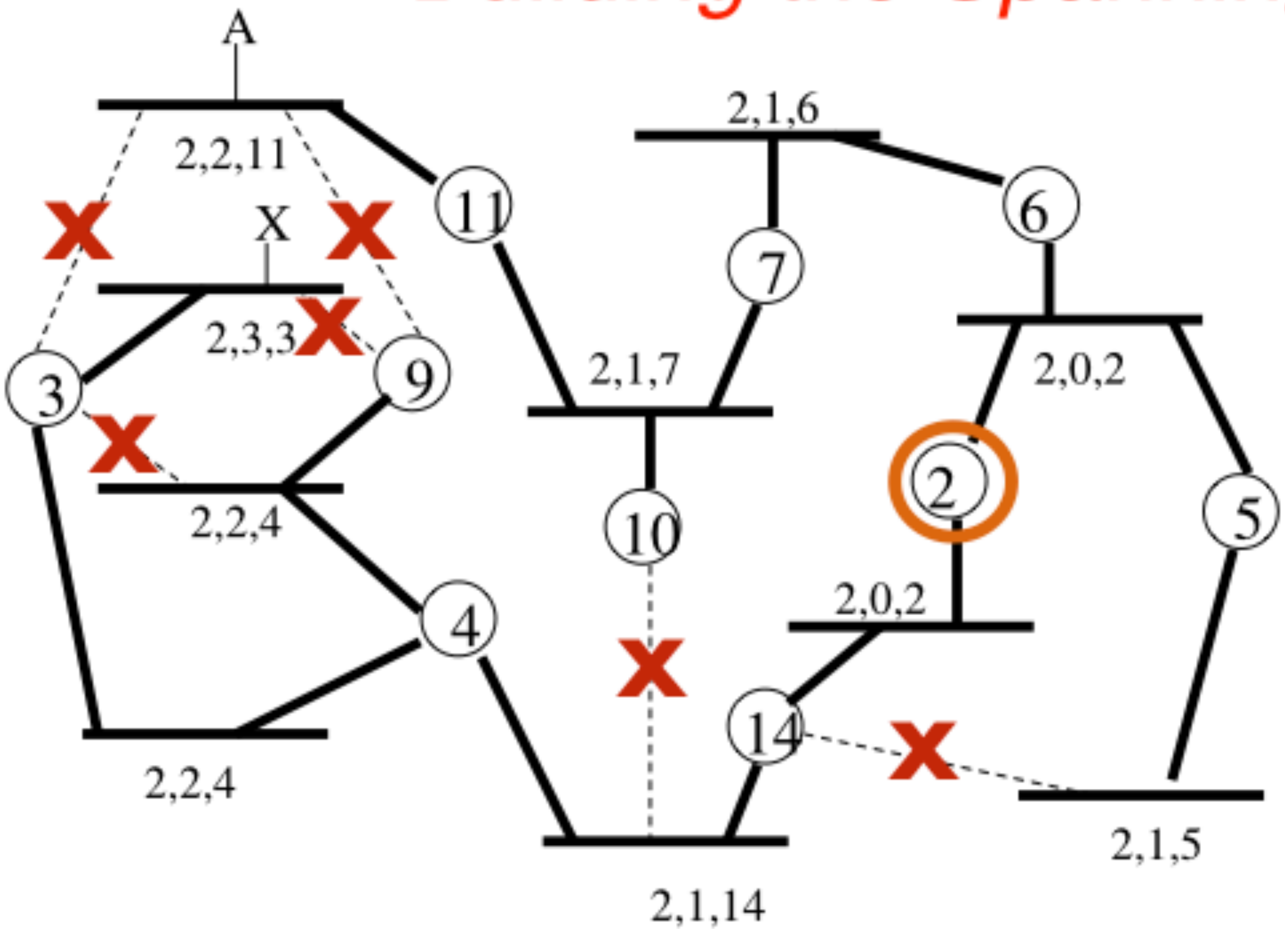
5, 6,14 are **directly connected** to the root (2)

4, 5, 7, 10 are **one hop** from the root (2)
4 via 14; 10 via 14; 7 via 6
and 5 via 14; but also via 2 (the lower ID wins - so record 2)

3, 9,11 are **two hops** from the root (2)
3 via 4 (2 ports, use lowest port); 9 via 4; 11 via 7 (with 2 ports)

Bridges also find other paths with higher counts - these are Blocked

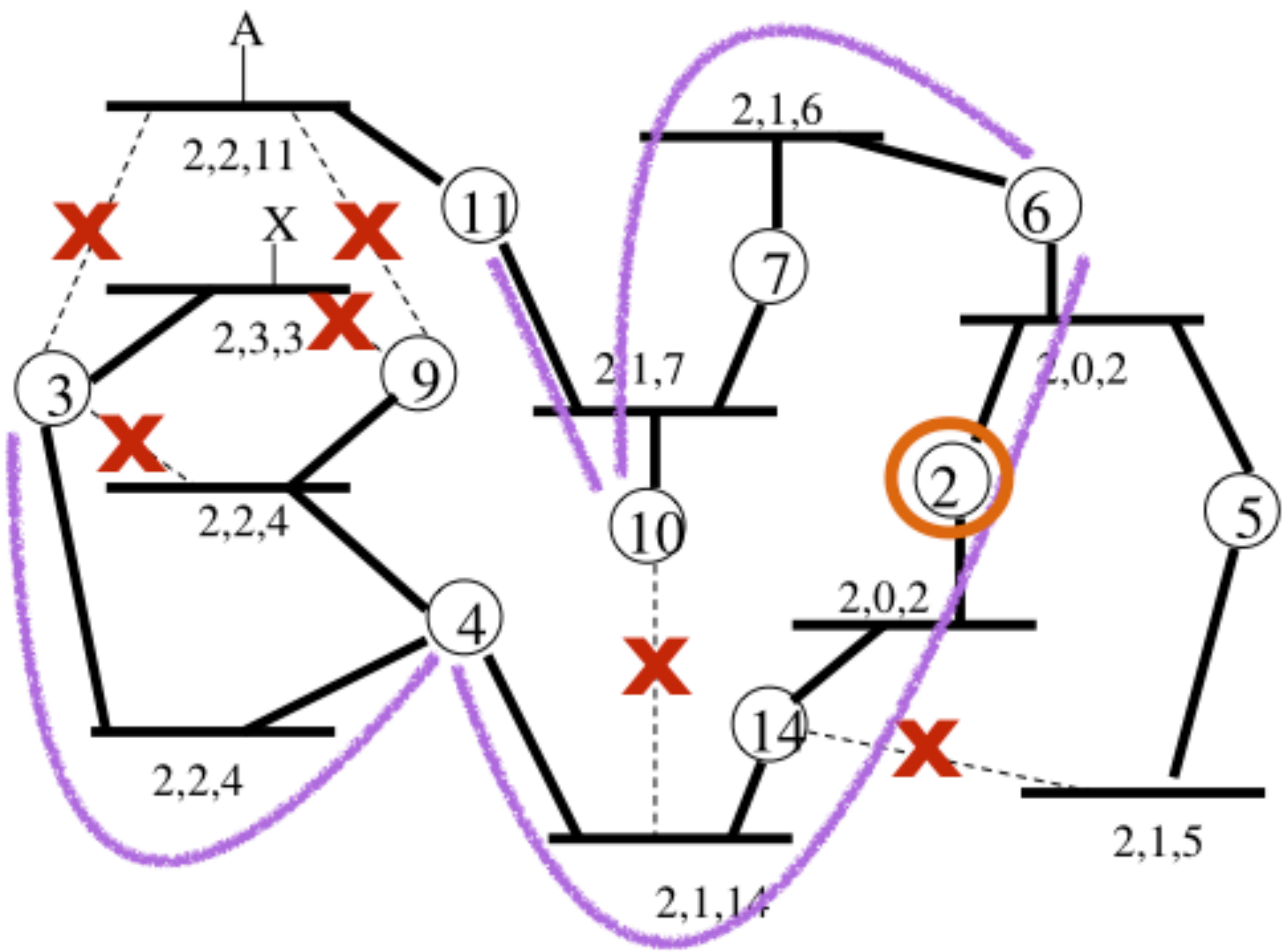
Building the Spanning Tree



Once a root bridge has been found:

- Other Bridges set the ***root port***, as the port closest to the root
- They ***Block*** all ports on ***all but the shortest path to root***
- Blocked ports **do not forward packets**
- Blocked ports ***do*** continue to exchange BPDU frames

The Lowest-Cost Tree Rooted on the Lowest ID



Suppose X connected to 3 needs to send to a system A connected to 11

The least-cost tree links all segments - but in this case uses 6 hops !!

Frames are not necessarily forwarded along an ***optimal path***

An administrator can configure the priority to set the default root

(Note: IP Routers do more optimal forwarding)

Algorhyme – the Spanning Tree Poem

I think that I shall never see
A graph more lovely than a tree.
A tree whose crucial property
Is loop-free connectivity.

A tree that must be sure to span
So packets can reach every LAN.
First, the root must be selected.
By ID, it is elected.

Least-cost paths from root are traced.
In the tree, these paths are placed.
A mesh is made by folks like me,
Then bridges find a spanning tree.

—Radia Perlman

See Radia and her daughter sing this: https://youtu.be/iE_AbM8ZykI

Detecting a port/cable/bridge Failure

If one or more links fail after the STA has built the tree
one set of bridges may be unable to forward to the other bridges.

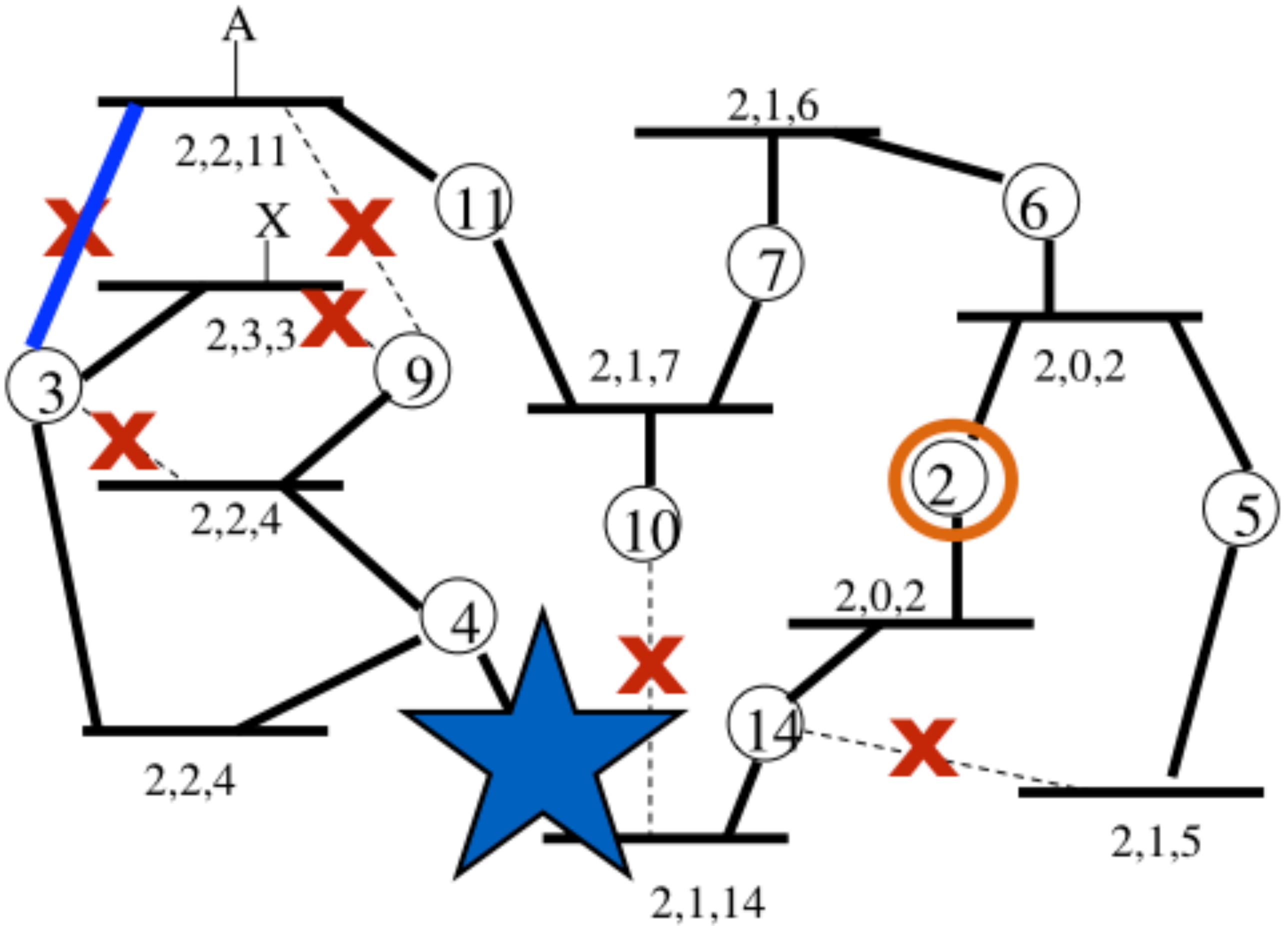
Eventually, the switch information about the root will time-out in some bridges - this takes time.

The network is still passing BPDUs on all its active links

The bridges then discover the new lowest cost path to the root

It doesn't matter what failure, STA finds a new path if this exits

A Link Failure Resulting in re-forming the ST

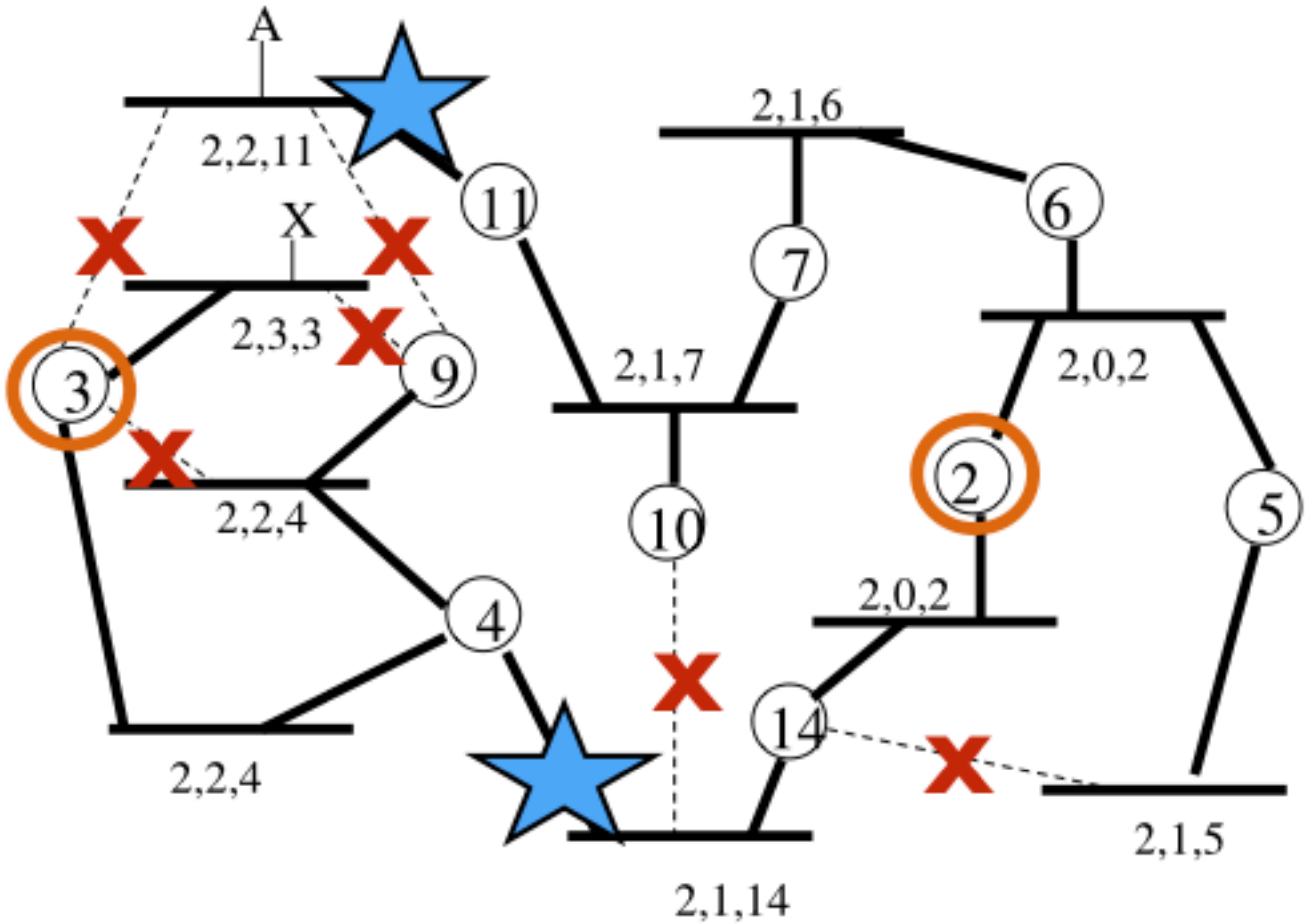


Suppose the link from 4 to 14 fails

The lack of connectivity from 4 to 14 is discovered
The shortest path to the root (2), is via 3
This is unblocked and the tree is re-formed using this port

This reconfiguration takes a few seconds, but fixes the problem

A Link Failure Resulting in Network Partition



Suppose two links fail - 4 to 14 and 3 to 11, X can no longer reach A

STA discovers the lack of connectivity to the root from 3,4,9
In the absence of a path to the root, STA elects 3 as a new root.
Two independent trees have been created.

The Network is said to have "Partitioned"

Detecting Partition

If one or more links fail after the STA has built the tree
one set of bridges may be unable to reach the root bridge

The network is said to have ***partitioned***.

Eventually, the switch information about the root will time-out

The partitioned part of the network will then elect a new root.

A pair of spanning trees will form, each rooted on a bridge in their part of the partitioned network.
These two trees function as two independent networks.

If links later become usable between the two spanning trees,
BPDUs will allow STA to discover the new root
STA will reconfigure around a single root, healing the partition

Summary of IEEE 802.1D STA

Connecting two bridges in parallel will cause looping

Loops can be avoided by design

Loops can be avoided using the Spanning Tree Algorithm (STA)

STA elects a root switch per (V)LAN
- Bridge with the **lowest MAC** becomes the Root

STA then disables all but one path through the LAN

Each bridge port is either: Blocked, Learning or Forwarding
With STA there is only ***one*** active forwarding path to each LAN

This is safe and automatic, but not optimised!

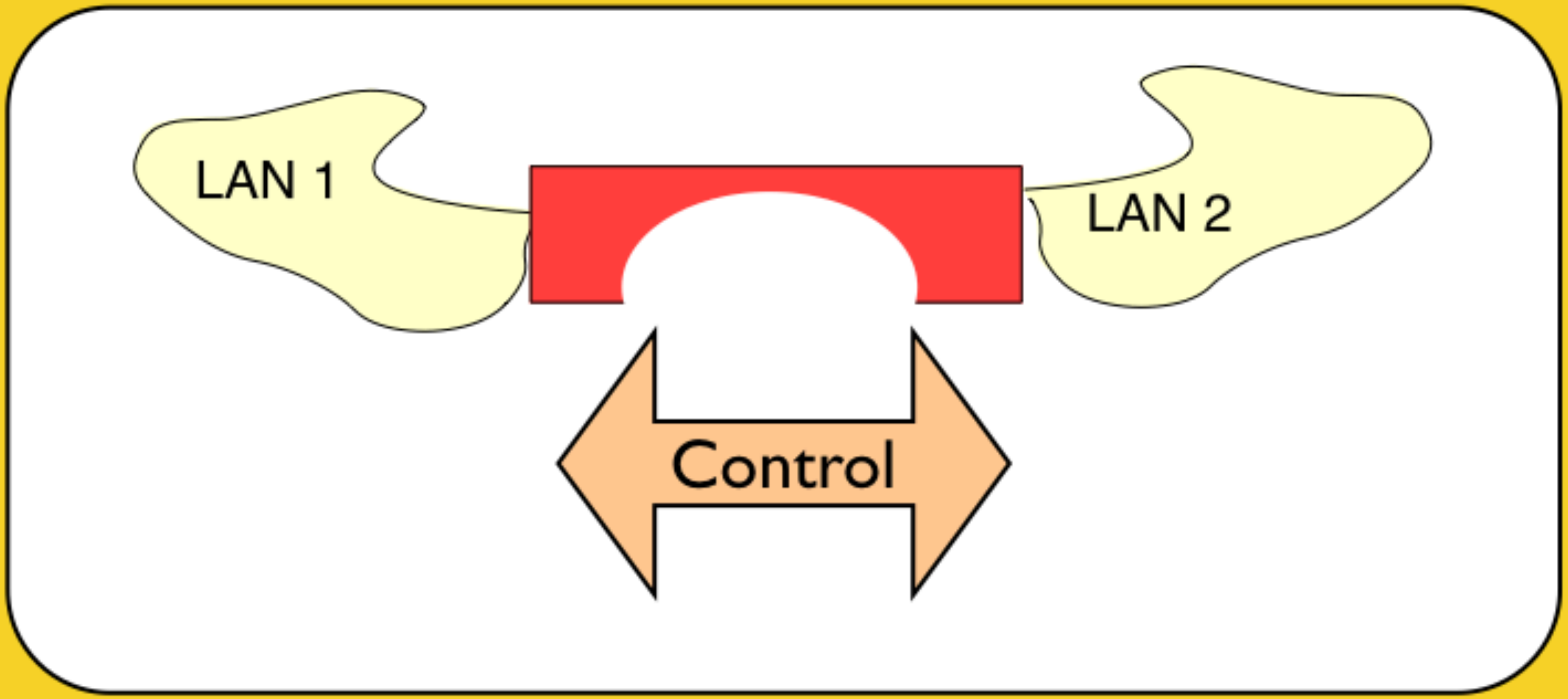
Priority can be used. There is also rapid spanning tree (802.1w)

See Radia talks about this: https://youtu.be/LgbhWUVx_ts

Bridges are "smart"

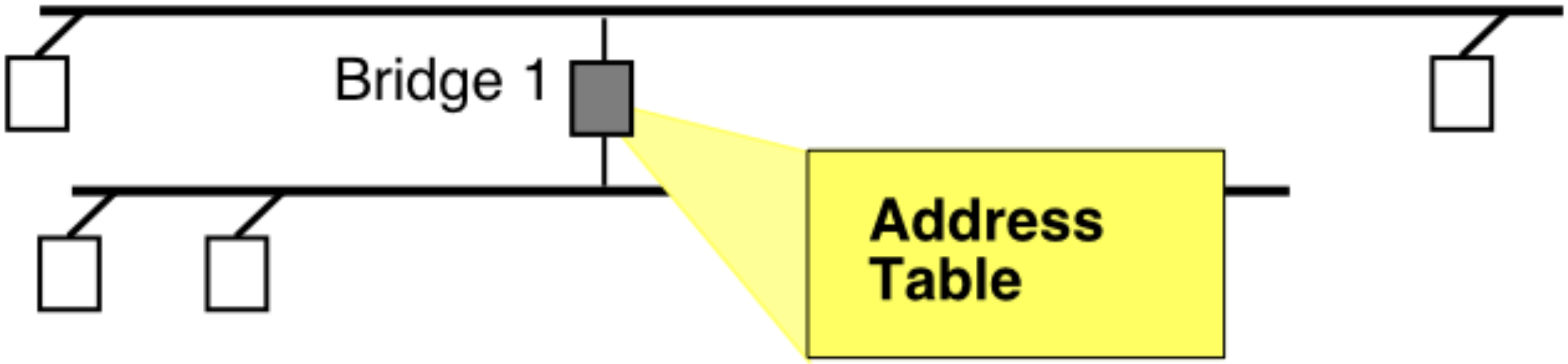
Bridges & Switches:

Hardware Tables using Contents Addressable Memory



Module 5.4

Address Table



Two types of table entry:

Static addresses to forward (set by administrator)

Learned address to forward (dynamic entries)

Table ***COULD*** be implemented as an array

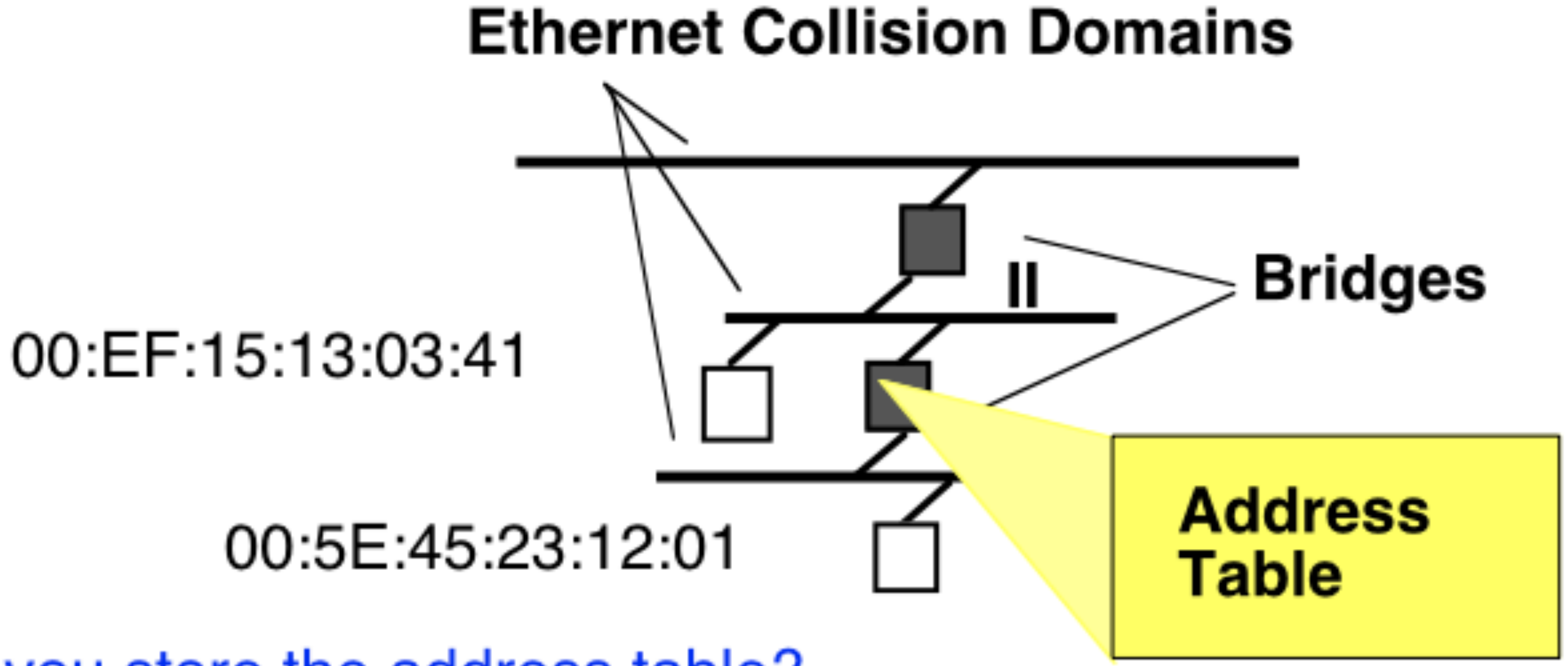
- may be a software "Tree" structure
- as tables grow this still become expensive to implement
- forwarding requires
  - (i) search for a match of the source address to update
  - (ii) search for a match of the destination address too forward

This will not work for high-speed switches!

```
sh mac add
          Mac Address Table
---------å----------------------------------

Vlan    Mac Address       Type        Ports
----    -----------       --------    -----
 All    0016.4718.e680    STATIC      CPU
 All    0100.0ccc.cccc    STATIC      CPU
 All    0100.0ccc.cccd    STATIC      CPU
 All    0100.0cdd.dddd    STATIC      CPU
   1    0002.b302.72b9    DYNAMIC     Gi0/1
   1    0003.ba9a.8c9b    DYNAMIC     Gi0/1
   1    0004.23b5.9b36    DYNAMIC     Gi0/1
   1    0004.76dd.bb0a    DYNAMIC     Gi0/1
   1    0007.e9bd.5d1f    DYNAMIC     Gi0/1
   1    0008.a334.7018    DYNAMIC     Fa0/24
   1    000e.0cea.1ff8    DYNAMIC     Gi0/1
   1    0010.6026.1436    DYNAMIC     Gi0/1
   1    0011.43e1.9fdf    DYNAMIC     Gi0/1
   1    0013.80b1.e216    DYNAMIC     Gi0/1
```

A Software Address Table



How do you store the address table?

Using a ***linear list*** takes (n) attempts at max to find a match, or (n/2) on average

An alternate is a ***well-balance binary tree*** This can decrease the search to LOG2(n).

Still, a software solution for 100s of entries is still computationally expensive

MAC address	Static	Port
00:5E:45:23:12:01:03	YES	I
00:EF:15:13:03:41:55	YES	II

Challenges in the design of the Address Table

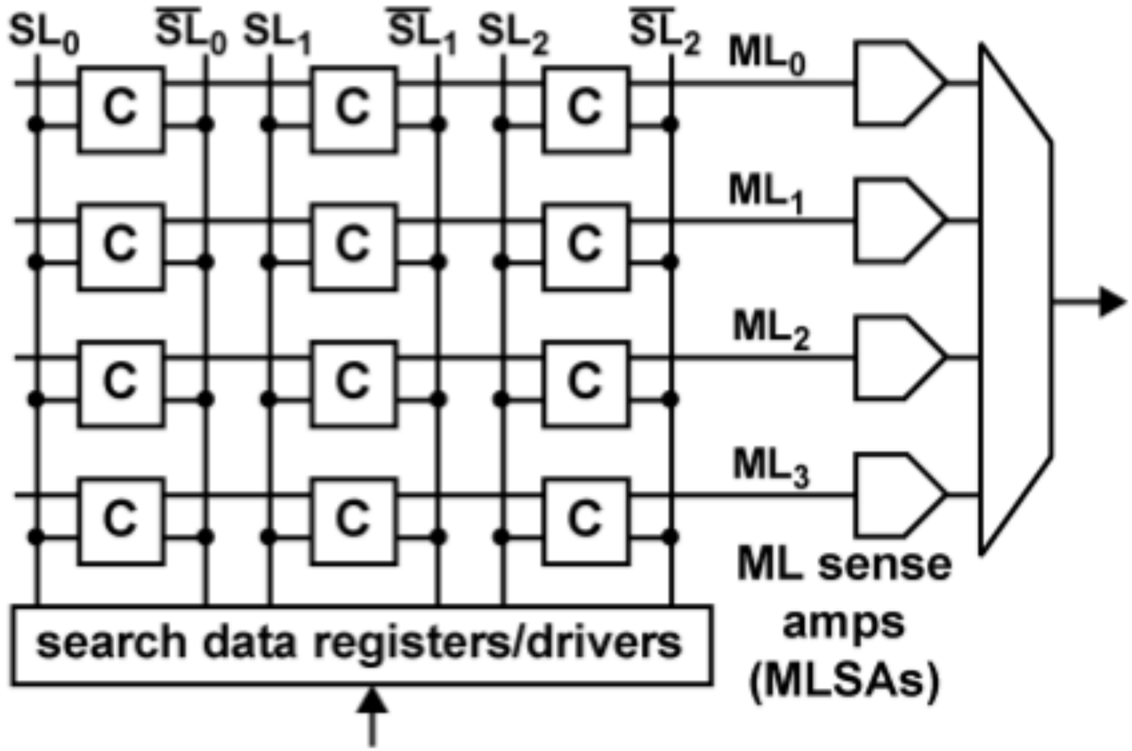
An enterprise-grade switch can track 1000's+ of addresses

- 0ften more than 10,000 addresses are needed

- Each lookup needs time to complete and all lookups need to be completed in the time taken to process a single frame.

- As speeds ***increase***, so also the lookup time ***reduces***!

Simplified Content Addressable Memory Design



An alternate is to use a Content Addressable Memory (CAM)

- A basic CAM consists of cells, and sense amps to detect a match
- Each read access is performed in one cycle (e.g 50 ns)

CAMs ~twice as complex as SRAM (2x area on chip)

- This means they are much more expensive than static RAM

Reading an Address Table entry from the CAM

A CAM contains a set of cells (one cell for each entry)

Each CAM cell stores one “**value**” together with an **“index”**

The value for an address table contains the MAC address

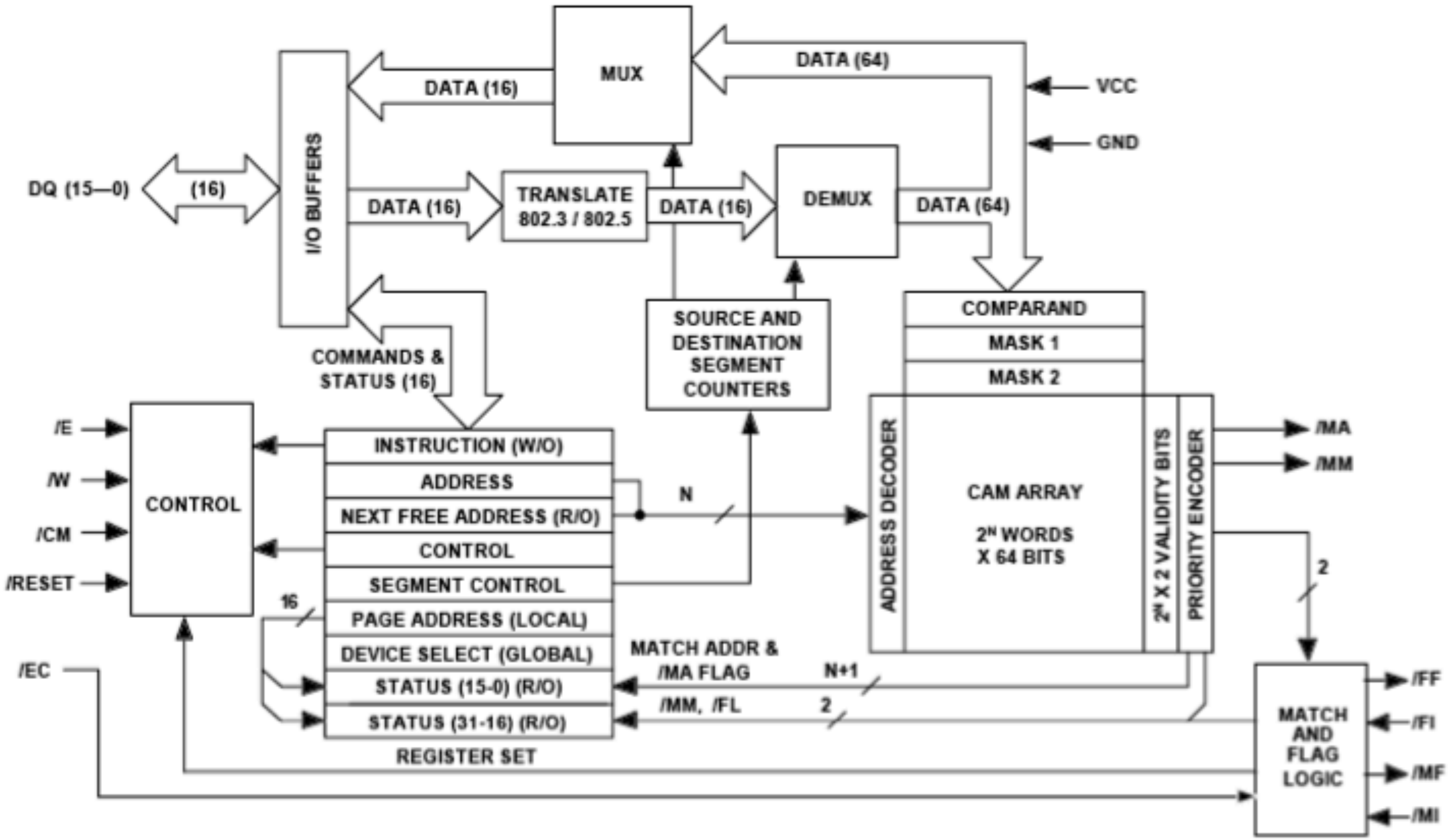
It *might* also contain other information, such as VLAN-ID.

Switches use the CAM to search for a match of the MAC address

To Read the CAM, a “**value**” is applied to the search input:

- If there is no entry the CAM returns “none”
- If it matches a stored value it returns the associated “index”
- The index can be used too determine the port, etc.

Practical CAM Design



CAM design (often implemented inside an ASIC)

Address Table

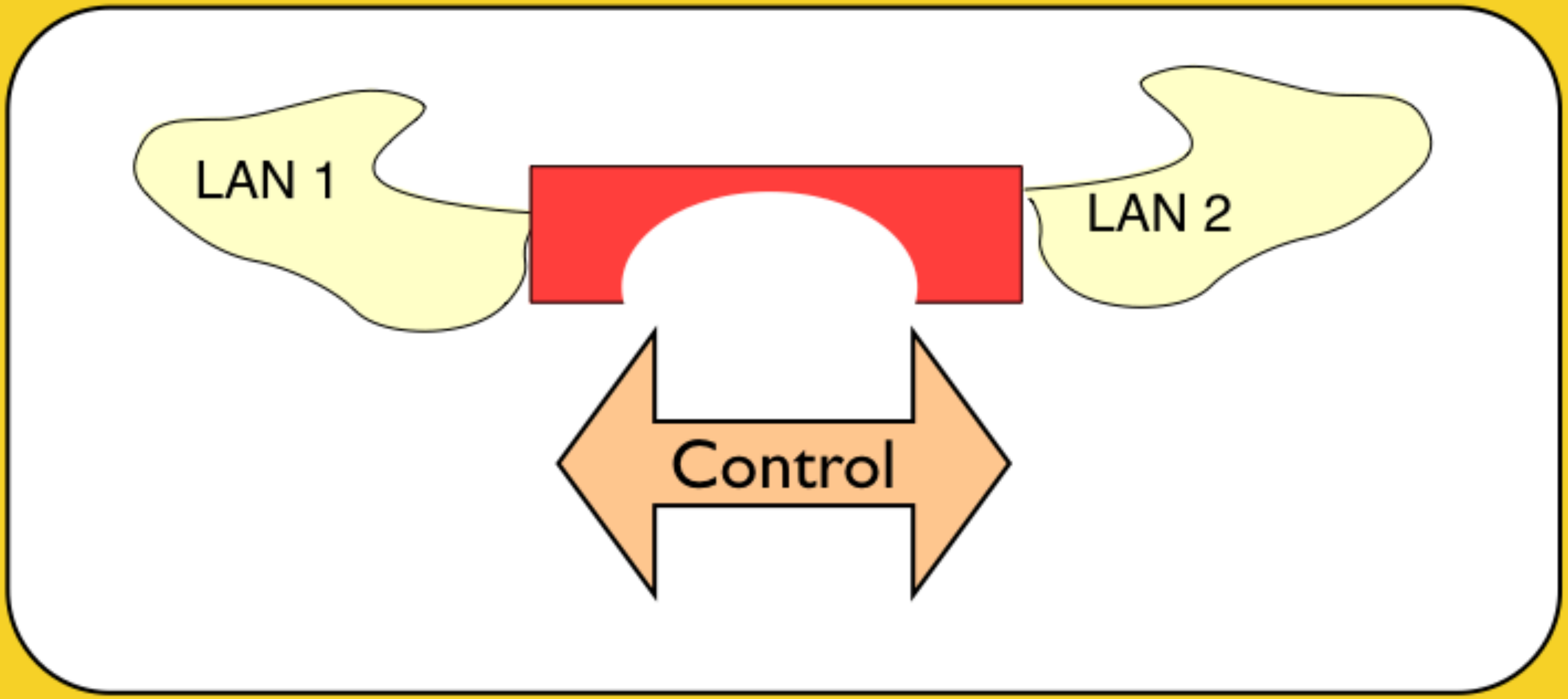
- **Different ways to store the address and filter tables:**
  - Array in memory - simple limited software bridge
  - Tree in memory - better software bridge but still limited
  - Enterprise switches use a CAM
  - More sophisticated can match multiple fields using a TCAM

- **Larger tables (e.g. 10K) are needed in larger LANs**
  - Don't be tempted to use a home switch!

- **Content Addressable Memory*:***
  - Faster lookup and store
  - But is also much more expensive
  - Are a limited resource (Enterprise switches protect this asset)

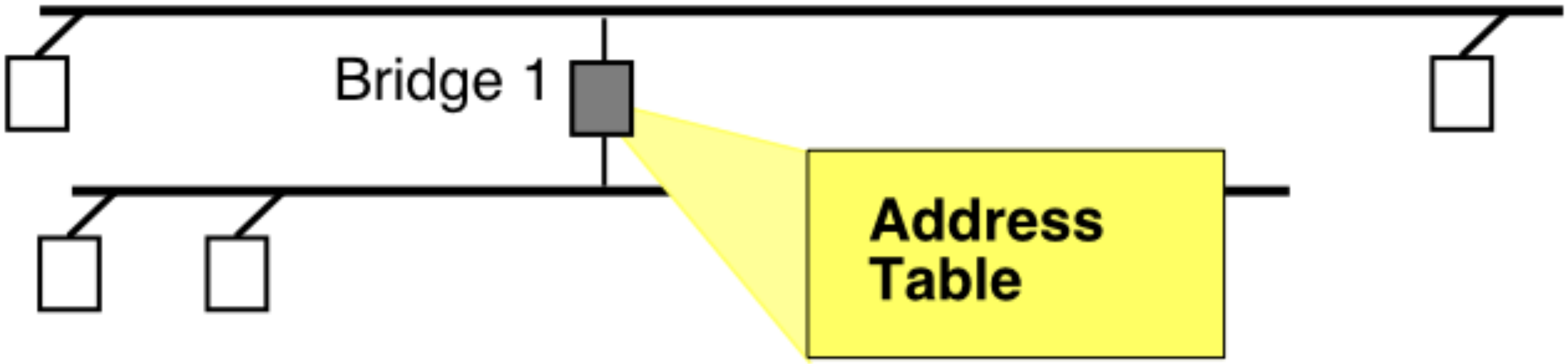
Bridges & Switches:

Attacks on Address Tables



Module 5.5

Attacks on the Address Table



Network engineers need to avoid the network from be attacked

- A malicious attacker might wish to reduce network performance

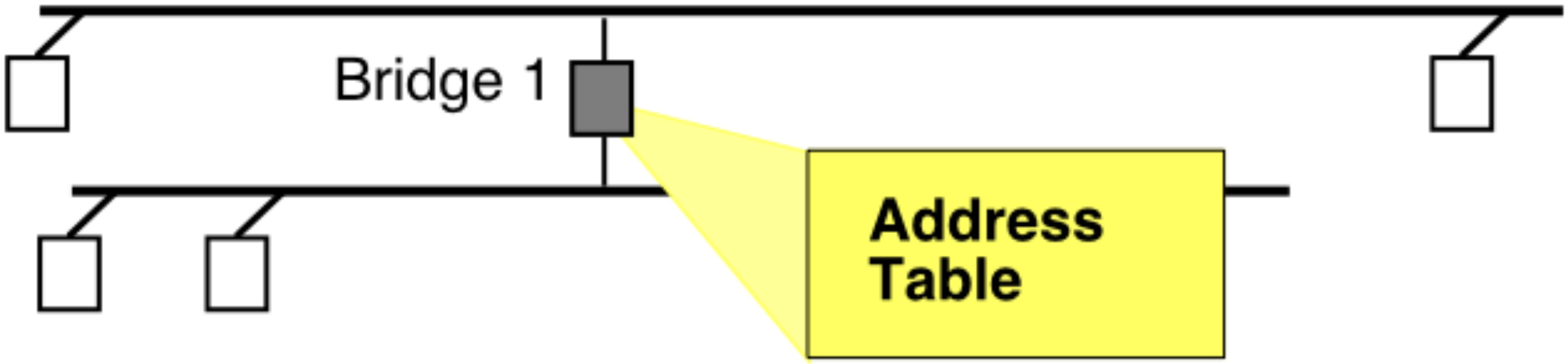
The Address Table is an expensive resource

- In a software-based design it consumes processing resource
- In a hardware-based design the CAM is of finite size

An attacker can utilise these limits to mount a denial of service attack

- This attack attempts to reduce the forwarding capability of a switch

Overflow Attack on the Address Table



A flooding denial-of-service attack could be made against the table

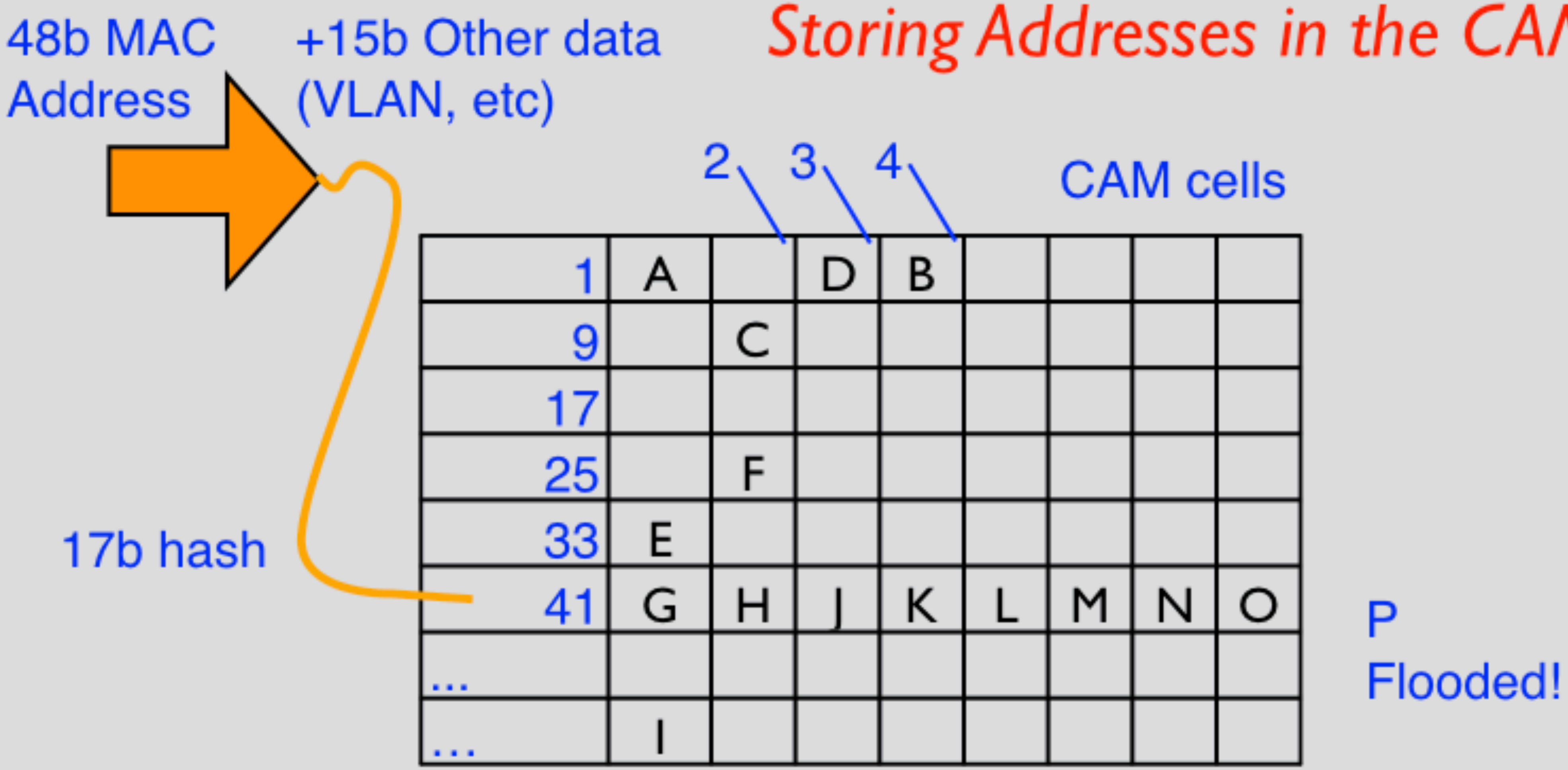
Suppose a malicious computer sends frames with lots of source addresses (there are programs that override the source address)

- The table ***could*** overflow

Once full, all other traffic is flooded to ***all ports***

The hash used to store in the CAM increases the vulnerability
- requires only small number of carefully chosen addresses

Storing Addresses in the CAM



What if values hash to same cell?

H=41;J=41; K=41 etc … P=41

If cell at the hash value is already used, use next cell

If ***next 8 cells*** are all full, then the switch floods the frame

Ternary CAM (TCAM)

A TCAM can match more than one set of non-contiguous bits

- i.e. matches 0,1 and don't care (X)
- 100XX matches 10000, 10010 etc

More than one cell can match the input (unlike CAM)

- Often uses a priority encoder to determine a single output

Mid-high-range switches and routers use TCAMs to implement:

- Access Control Lists (filter tables)
- QoS classification
- IP forwarding (when functioning as a router)

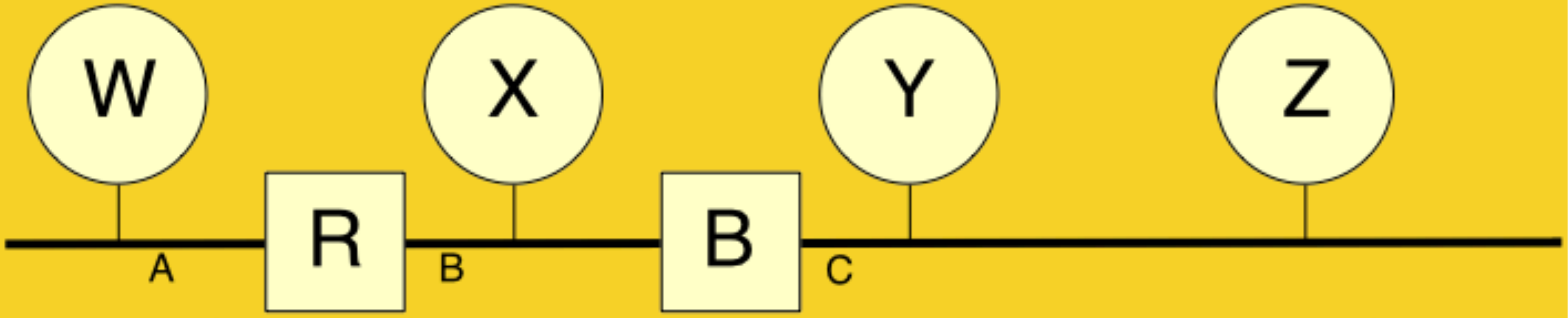
Using Address Tables

- **There are ways that the address table can be attacked**

    This helps understand the operation of a switch

    The course does not expect students to reproduce these examples of attacks

Bridge/Switch Question



Four computers (W,X,Y,Z) are connected by 3 Ethernet segments (A,B,C) using a Repeater (R) and a Bridge (B).

(a) Which computers receive (at the network level) the following frames (show also which LAN segments carry each frame)

W -> Broadcast

X -> Z

Y -> Z

Y -> Broadcast

(b) W, X are members of the multicast group 0x23.

W = 0x00102030 and X = 0x00102040.

Sketch the MAC header for a multicast frame sent from X.

Which segments carry this frame?

Thinking about the Address Table

Things to think about:

An end system that ***only listens*** (never sends)
- Frames are broadcast to all ports
- Could configure a static entry

An end system is ***turned off***
- Address entry will age and be deleted

An end system ***moves*** to another collision domain
- Bridge will have learned the wrong port
- End system will not receive unicast frames
- Entry updated when end system sends

Faster Ethernet

10/100/1000/10000

Fast Ethernet

Virtual LANs

Gigabit Ethernet

Data Centres

Module 6

Fast Ethernet

100 Mbps

Collision Domains

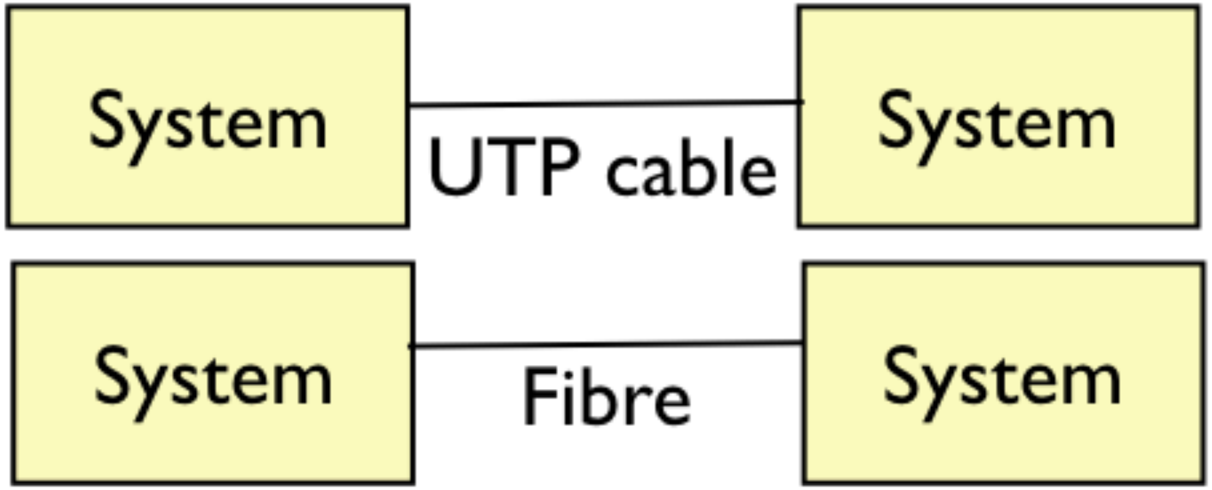
Broadcast Domains

Faster Transmission Speeds

100B-FX

Module 6.1

Fast (100 Mbps) Ethernet



Two Media:

100 Mbps Copper (UTP)

100 Mbps Fibre

Full Duplex*

* Historical note: the original spec included half Duplex (allowing CSMA/CD and 1 Hub). This was little used, and hubs were more complex than for 10 Mbps. The falling price of switches meant that were often cheaper and much more flexible than hubs. The half-duplex mode was seldom used.

Physical Layer for Twisted Pair Transmission

Copper (Unshielded Twisted Pair)

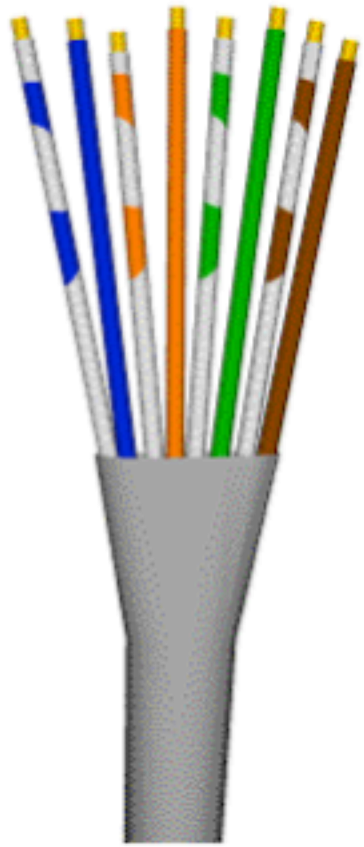
Uses 2 of the 4 twisted pairs in in CAT5 UTP

Pins 1 & 2 for Transmit; Pins 3,6 for Receive

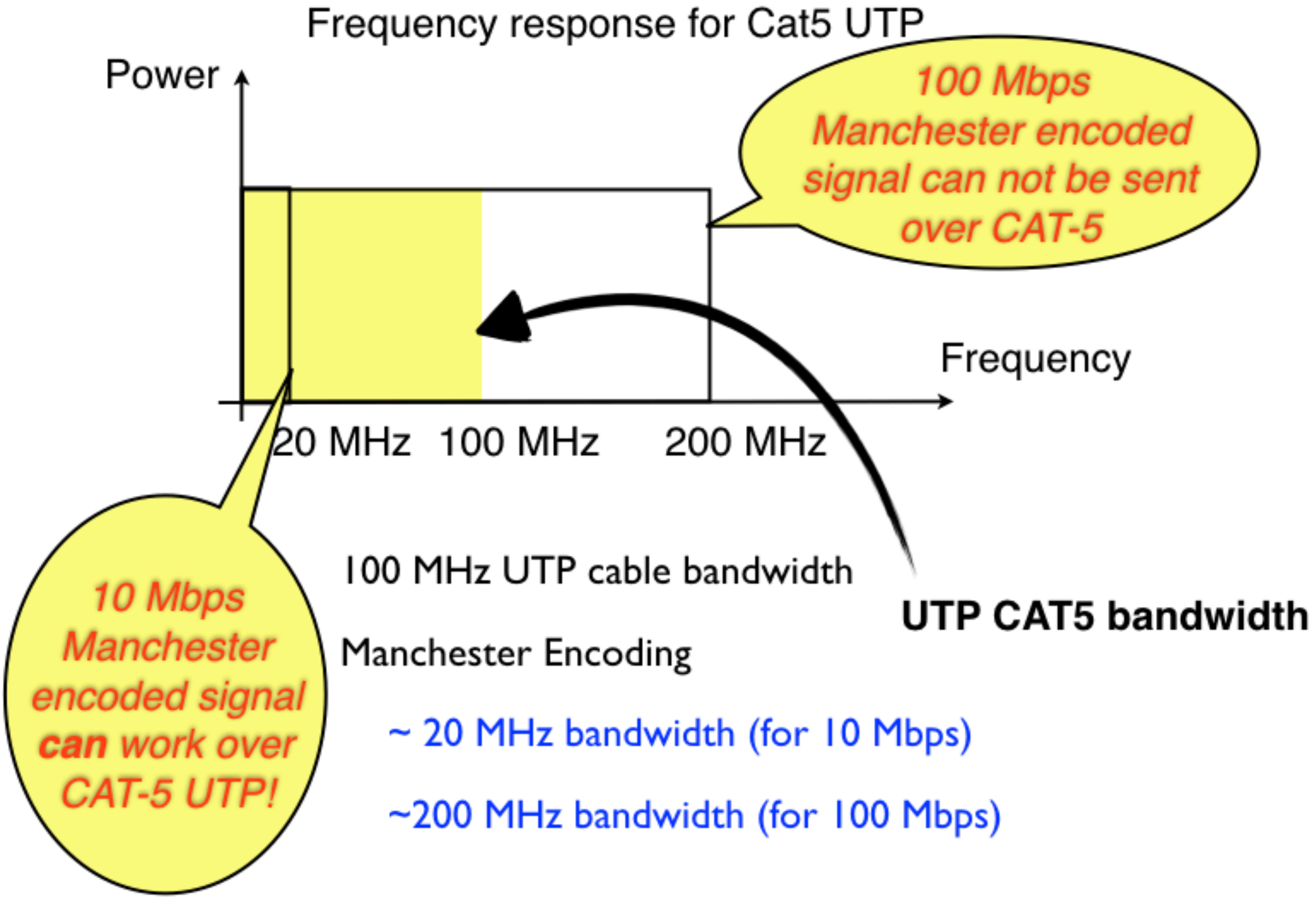
Cable properties

CAT 5 UTP has a bandwidth of 100 MHz

CAT 5e UTP has a bandwidth of 125 MHz



100Mbps Manchester Encoded Waveform



Manchester Encoding

~ 20 MHz bandwidth (for 10 Mbps)

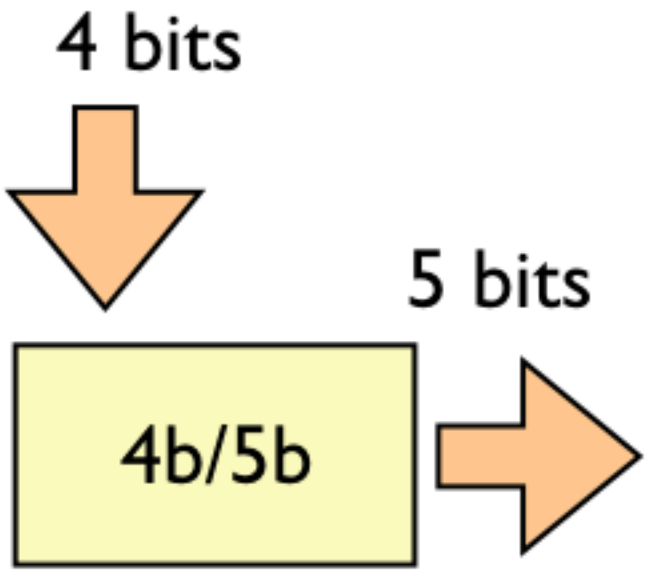
~200 MHz bandwidth (for 100 Mbps)

4b/5b Encoding

Two key goals:

No net DC current

Embedded clock signal (sufficient transitions for DPLL lock)



4b/5b encoding

4 input bits have 2^4 (16) values

5 output bits have 2^5 (32) values

Encode and send least significant 4b first

Encode and send most significant 4b next4

4b/5b Encoding

Decimal	Binary	Encoded
0	0000	11110
1	0001	01001
2	0010	10100
3	0011	10101
4	0100	01010
5	0101	01011
6	0110	01110
7	0111	01111
8	1000	10010
9	1001	10011
A	1010	10110
B	1011	10111
C	1100	11010
D	1101	11011
E	1110	11100
F	1111	11101

No constant level

≤ a sequence of 3 bits changed in 5 bits

Signaling Codes

The encoding rule uses 16 values for data, with16 unused

Some unused values denote signalling special events:

- Quiet (00000) Idle (11111) Halt (00100)
- Starting delimiters J (11000) K (10001)
- Ending Delimiter T (01101)
- Control Reset (00111) Set (11001)

The remaining values should never be sent

- Reception of these indicates an error

4b5b Encoded Waveform

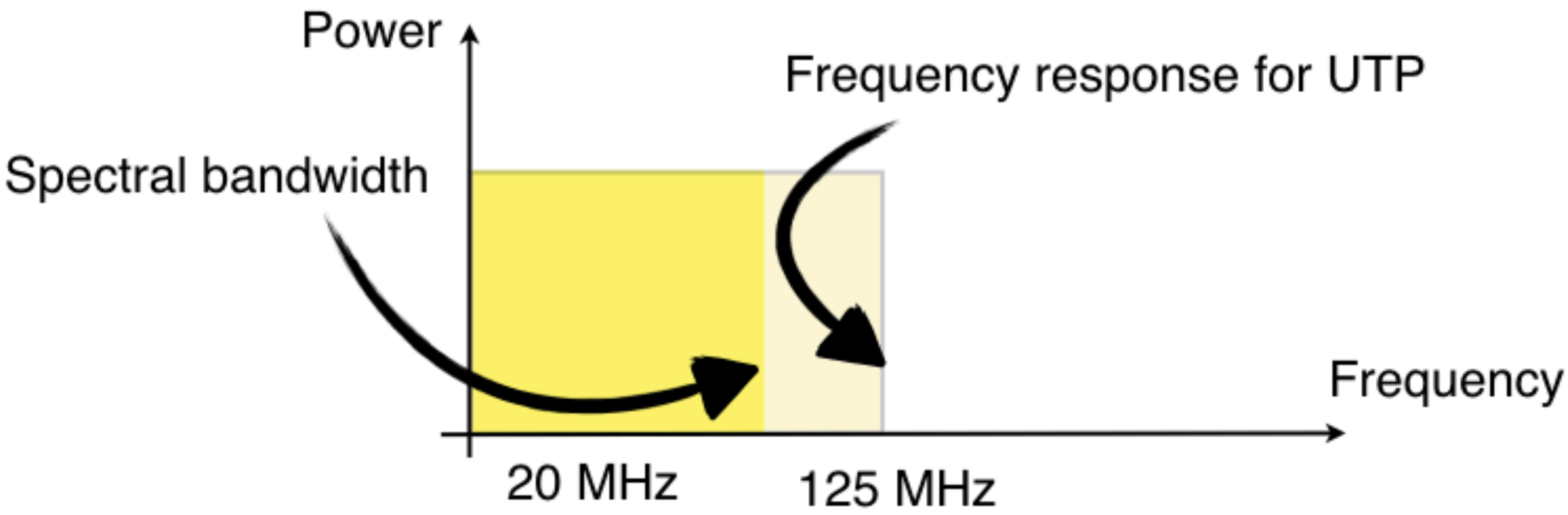
Encodes the clock with the data

- Includes transitions needed for receiver DPLL

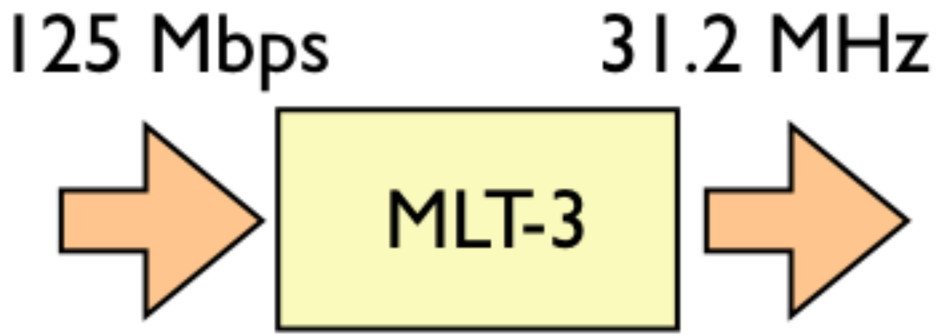
Encodes data patterns

- Stream contains start, end and other control signals

However, spectral bandwidth is 125 Mbaud, needing > 100 MHz!



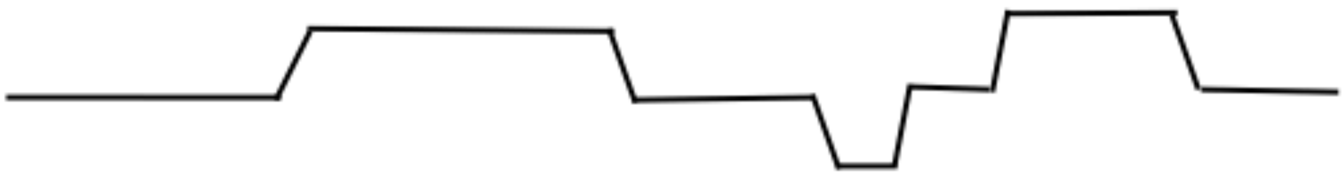
MLT-3 Encoding



MLT-3 Line Encoding

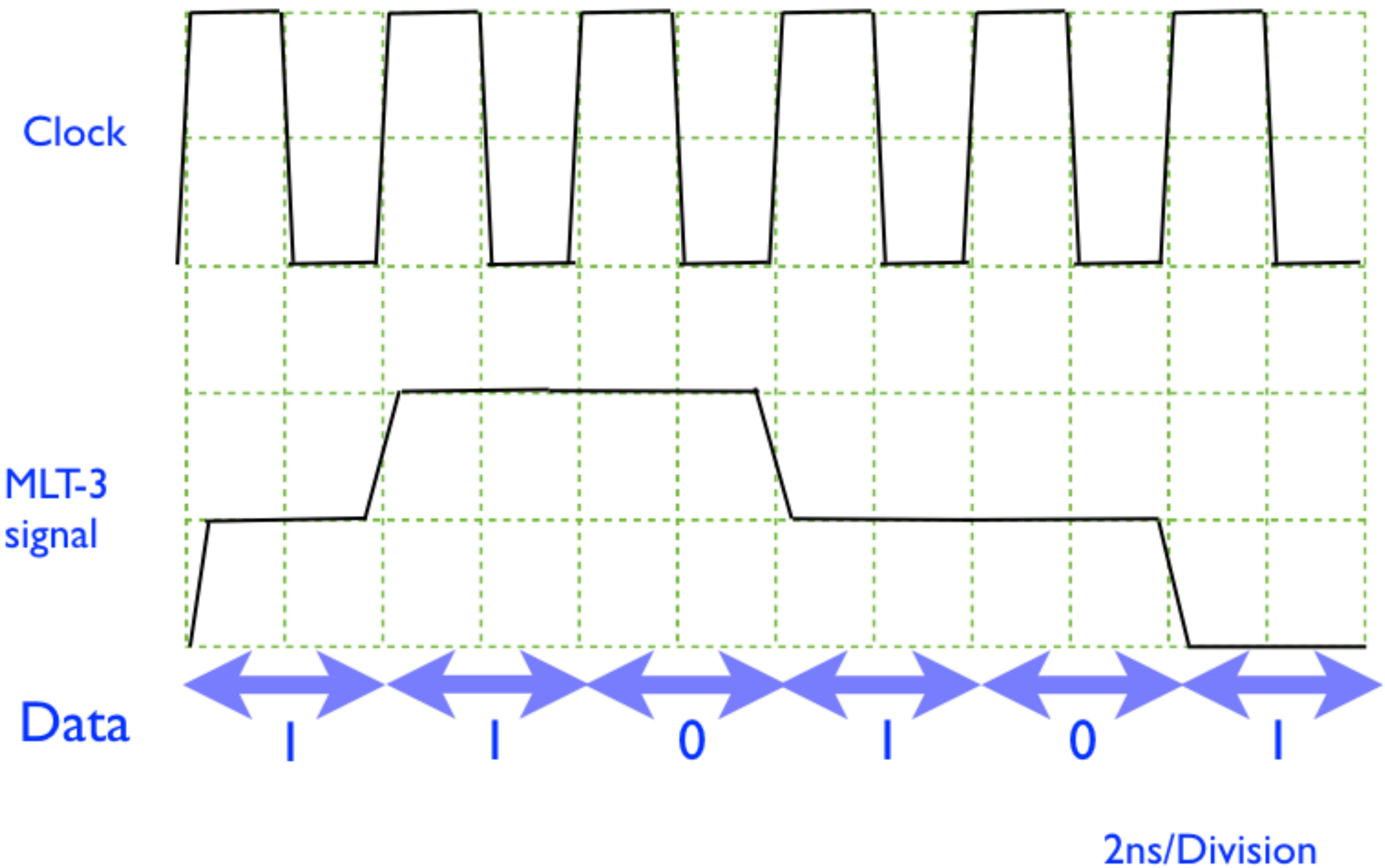
- Levels -V, 0, +V
- 0 data sent as no change in the level
- 1 data sent as a change to the next level, following a sequence:
  - (0) -> (+V) -> (0) -> (-V) -> (0) ...

NRZ	0	0	0	1	0	0	1	0	1	1	1	0	1	0
MLT-3	0	0	0	+	+	+	0	0	-	0	+	+	0	0



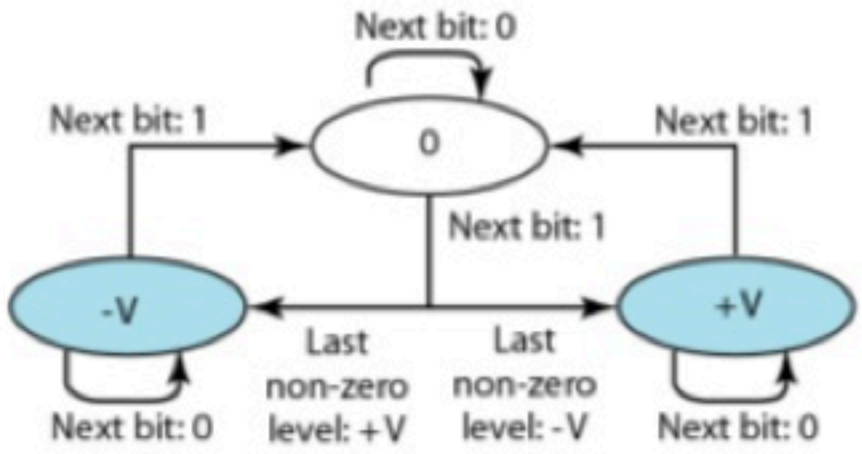
The baud rate is 31.25 Mbaud ~ 31.25 MHz bandwidth for 4b/5b+MLT-3 :-)

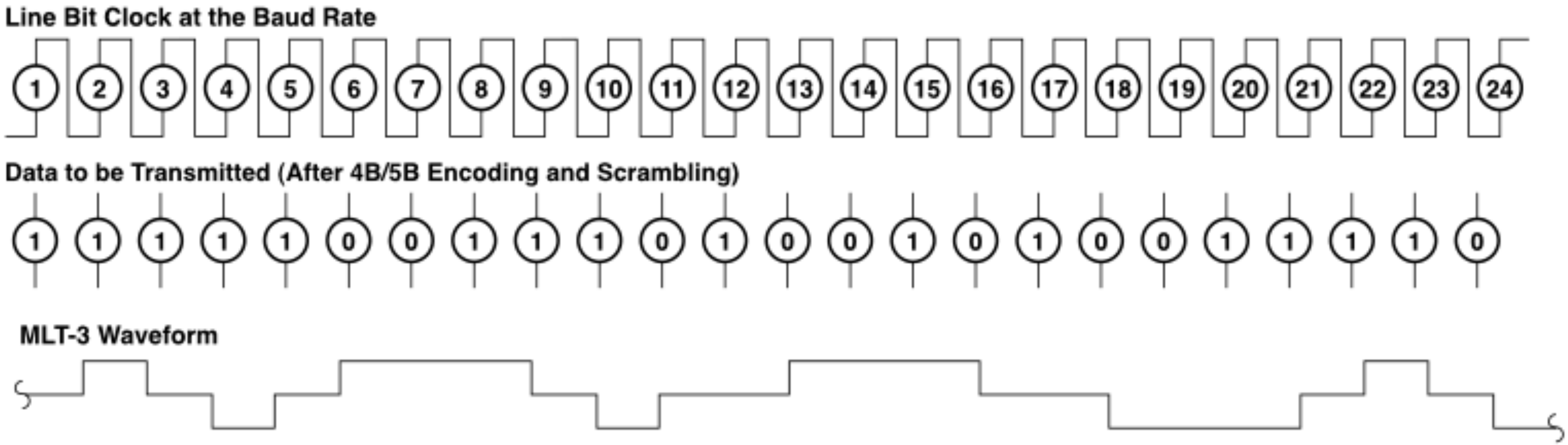
Example MLT-3 Encoding



MLT-3 Line Encoder

The MLT-encoder can be thought of as a finite state machine
This example labels the three output values: -V; 0; +V





Eye Diagram showing MLT-3 Encoding

Oscilloscope plot using an eye diagram

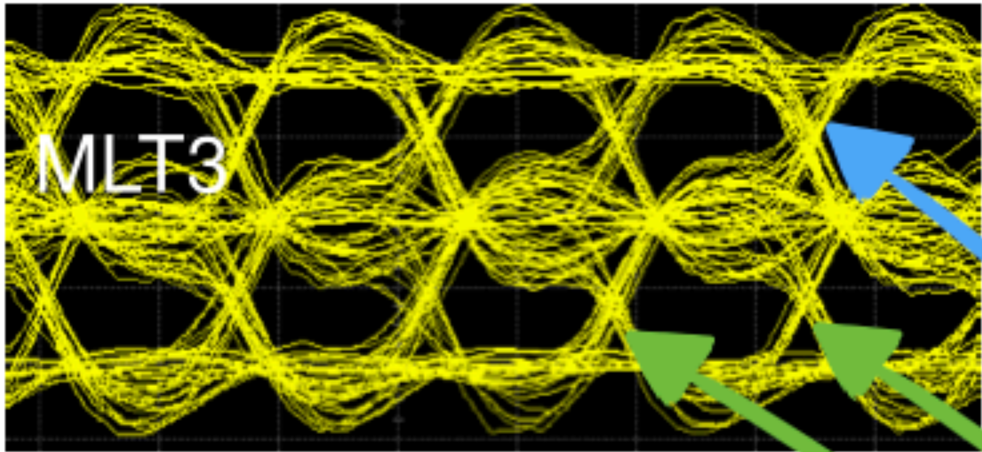
The eye diagram plots voltage v. time

With a timebase trigger for multiple scans through the waveform

+1

0

-1



Eye Diagram

2ns/Division

<1 bit >

Three distinct levels are clear

Noise is evident causing blurring in the vertical axis

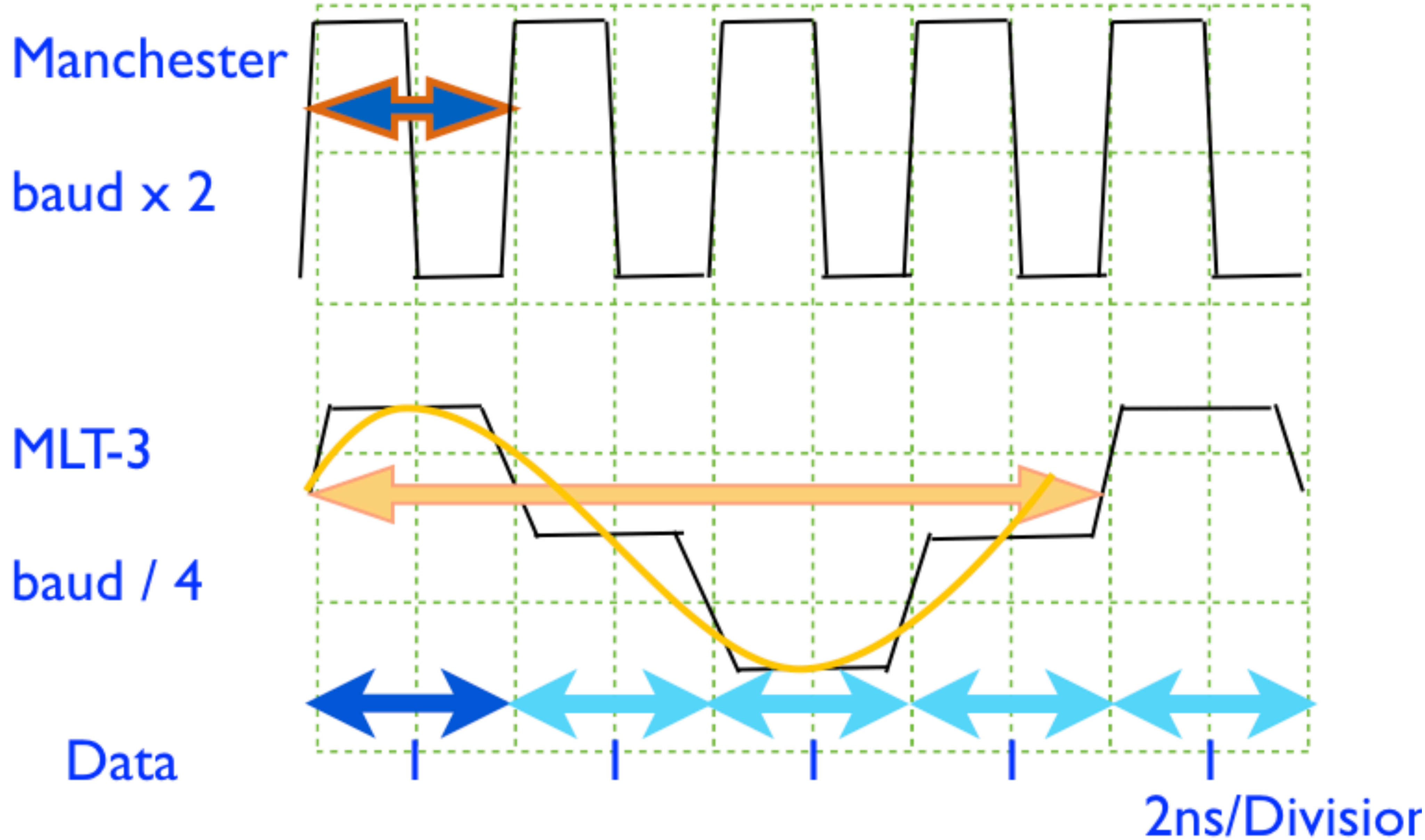
Transitions in level only occur at the edge of bits

Transitions in level only occur between adjacent levels

The slew rate is limited, i.e. the rise time for transitions

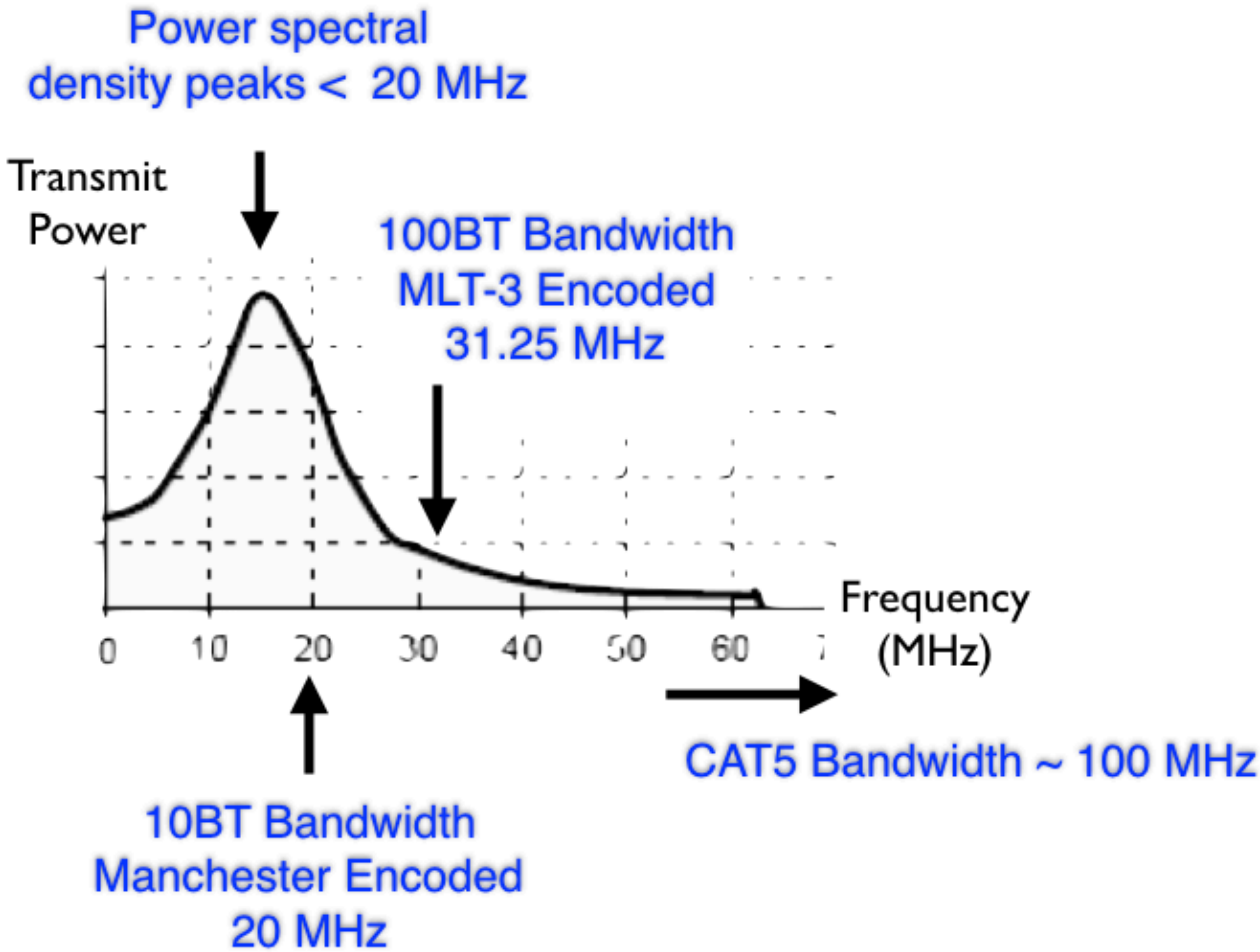
How does MLT-3 Encoding compress the Frequency?

Fastest change results when sending 1,1,1,1 etc



Max fundamental frequency = 100*5/4*1/4 = 31.25 Mbaud

MLT-3 Power Spectrum



Summary for Fast Ethernet

- **10BT uses UTP cable**

  You should understand the bandwidth limits of this cable preventing transmission of a 100 Mbps Manchester data

- **Fast Ethernet uses a New Waveform*:***

  4b5b clock encoding
  Special values are used to delimit the start/end of frames
  Scrambler included (reset at start of each frame).
  MLT-3 line encoding to compress bandwidth by 1/4

  You should should understand how to encode and decode a MLT-3 waveform

… the topic continues in next presentation

Fast Ethernet

100BT Line Encoding

Clock encoding (4b5b)

Scrambling

Level encoding (MLT-3)

Module 6.2

MLT-3 Encoding of Repetitive Signals

What happens if the Ethernet frame payload has repeating values

e.g., 0x000000000000 encodes as....

11110; 11110; 11110; 11110; in 4b5b encoding

Which is itself repetitive...

4b5b	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0
MLT-3	+	0	-	0	0	+	0	-	0	0	+	0	-	0	0

Looking again at this waveform we see it repeats….
In the frequency domain, there will be more power at specific frequencies

MLT-3 and Data Patterns

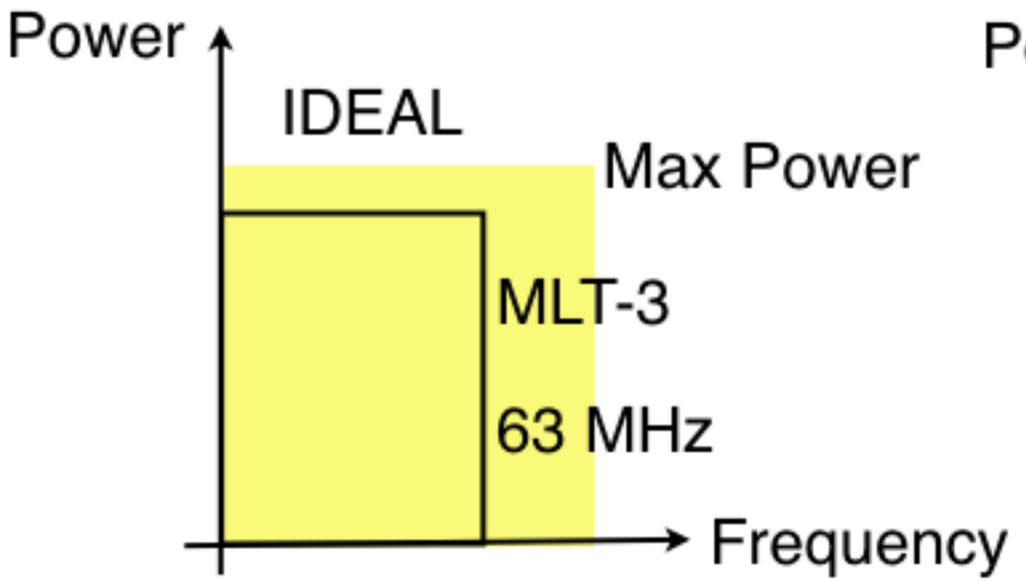
A problem occurs when same set of bytes are ***repeated*** over the cable

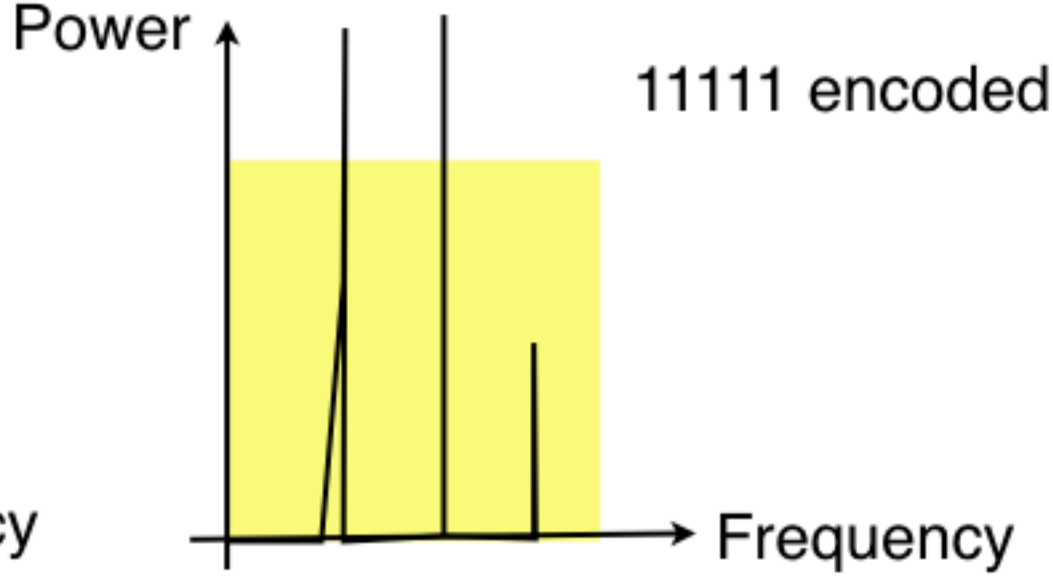
Results in a repetitive waveform with ***distinct frequency components*** (resulting in interference)

111111 = results in power concentrated at 31.25 MHz, 52.5 MHz, etc

10101 = results in power concentrated at 16.13 MHz, 31.25 MHz, etc

... clearly the spectrum is a function of the payload data!



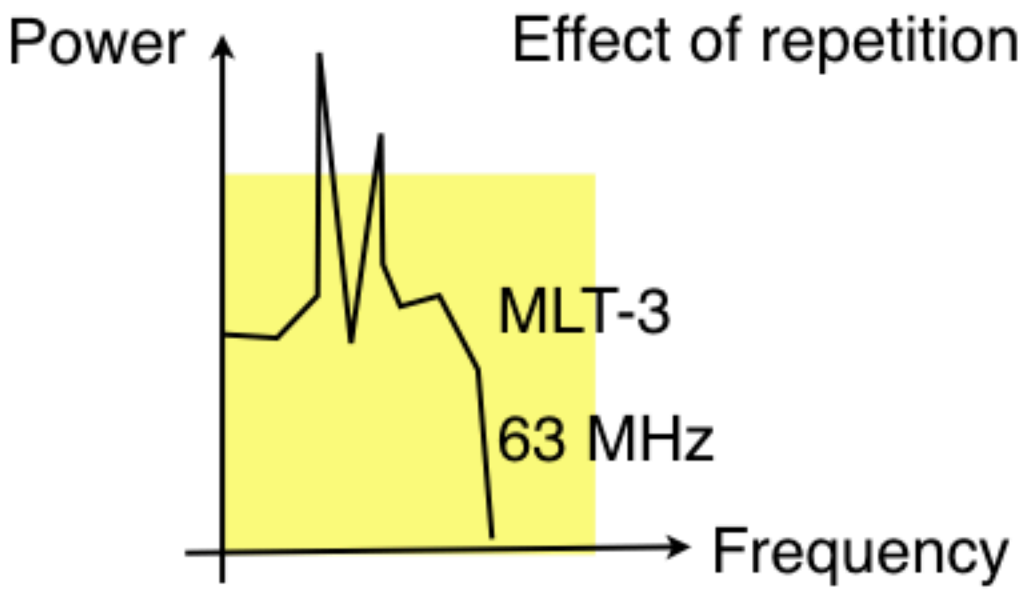


MLT-3 : Interference

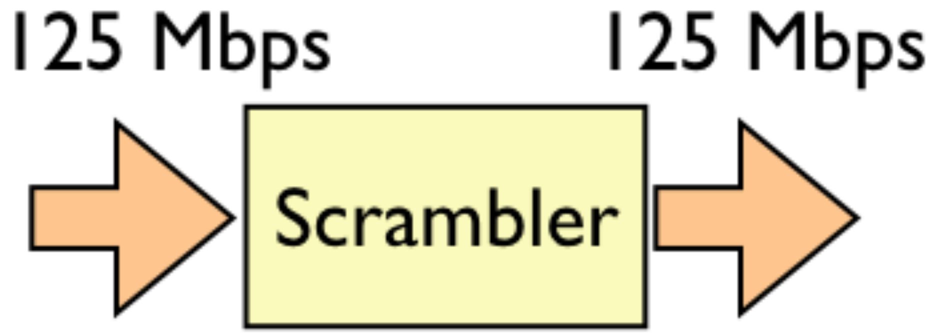
The peaks exceed the permitted power density allowed for the cable

Causes interference to other cables and equipment!

The actual spectrum ***must not*** be a function of the payload data!



Scrambler



A Scrambler "XOR"s the data with a known pseudo-random sequence

Original Data: 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0

Scrambler Sequence: 0 1 0 1 1 0 1 0 1 1 0 1 0 1 0 1 0 0 1

Sent Sequence: 1 1 1 0 1 0 1 0 1 1 0 1 0 1 0 0 0 0 1

The sent values of bits re randomised at the sender

The values restored at receiver using the same scrambler

Received Sequence: 1 1 1 0 1 0 1 0 1 1 0 1 0 1 0 0 0 0 1

Scrambler Sequence: 0 1 0 1 1 0 1 0 1 1 0 1 0 1 0 1 0 0 1

Original Data: 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0

Note 1: A random scrambler sequence has equal numbers of 1's and 0's

Note 2: Both sender and receiver reset sequence at the start of frame

Scramblers can be implemented using flip flops and XOR gates.

MLT transmission - Scrambler

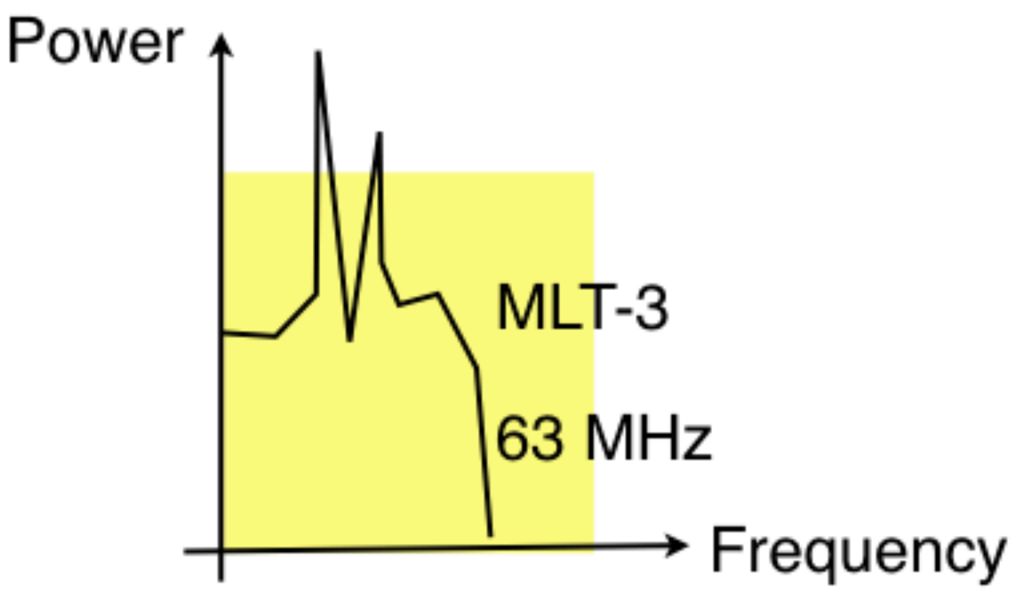
Scrambling is needed to ensure a ***smooth spectral response***

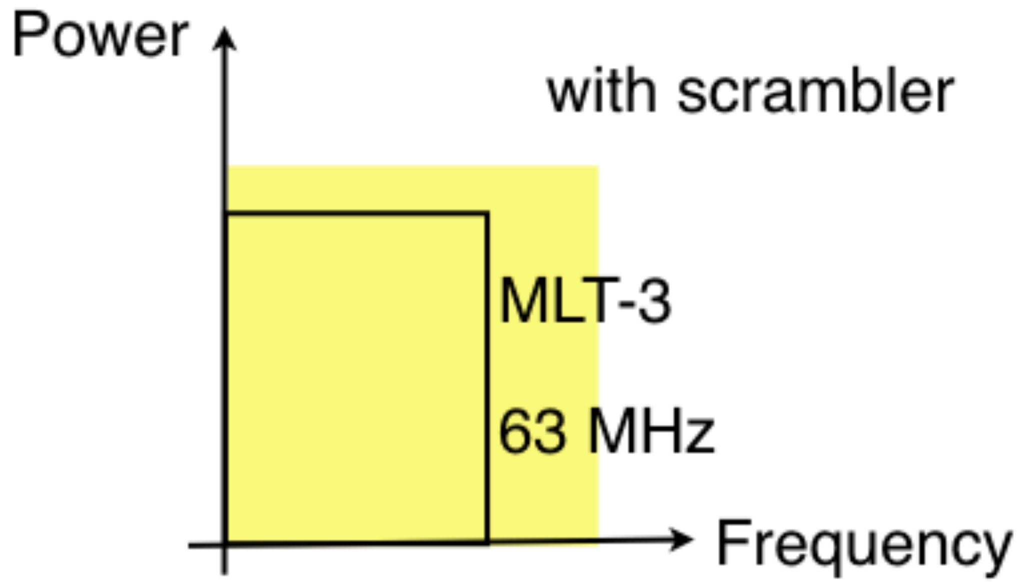
A scrambler changes the output of the 4b/5B encoder in some deterministic way, that may be restored at the receiver prior to decoding.

Scrambled data appears random to the MLT-3 encoder.

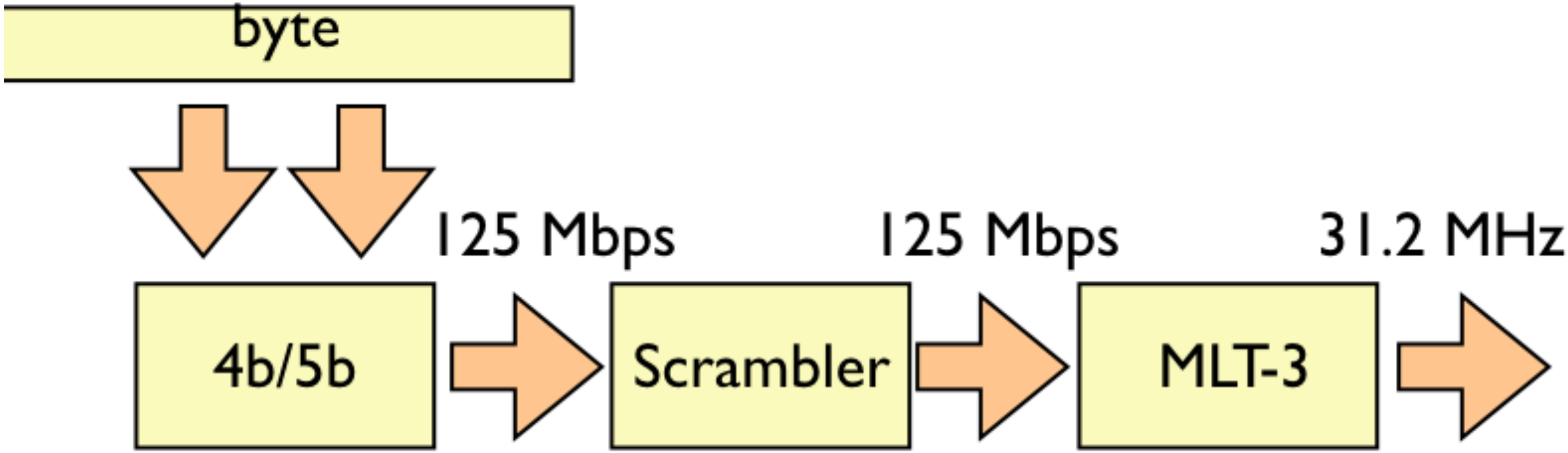
Power is ***spread*** rather than focussed at particular frequencies

Waveform matches the transmission properties of the cable





100BT Transmission over UTP



4 bits (1/2 byte) processed at a time

4 bits encoded to 5 bits (4b/5b encoding)

- ≤3 bits changed in 5 bits

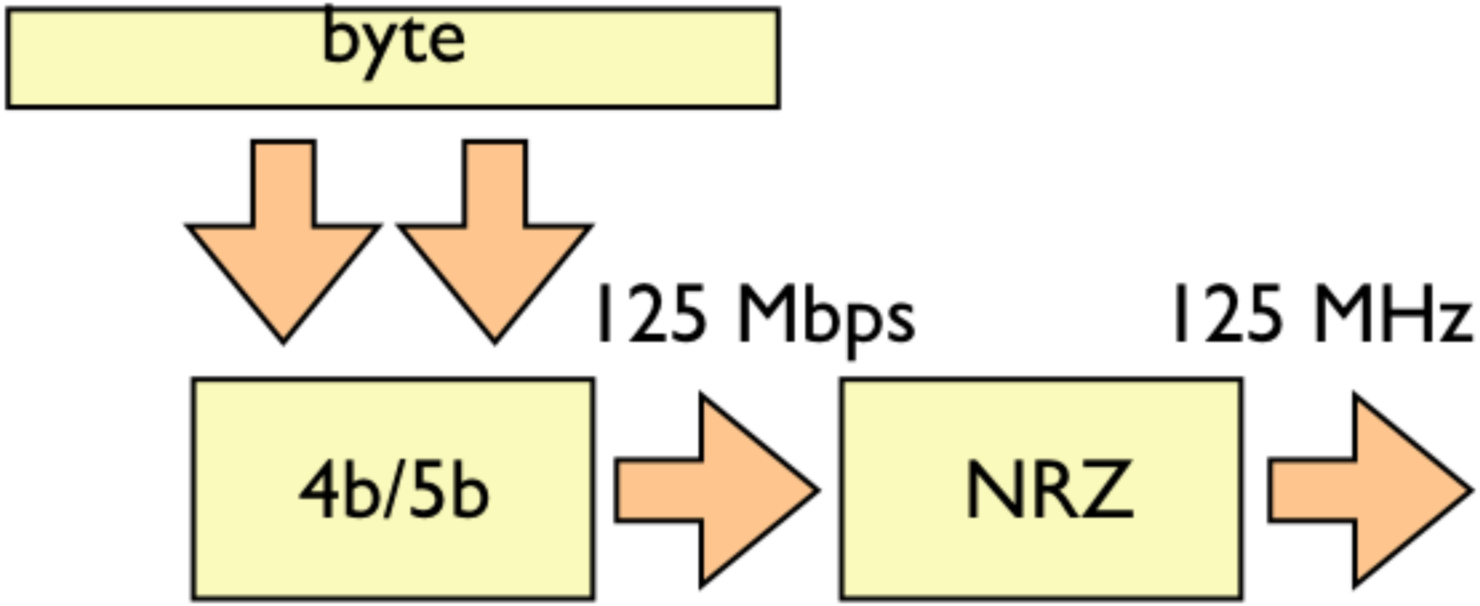
Scrambled

- Bit values randomised to disperse energy at sender

Line interface uses MLT-3 encoding (3 signal levels)

Chipset >> x2 the complexity need for 10BT

100BF Transmission over Fibre



4 bits (1/2 byte) processed at a time

4 bits encoded to 5 bits

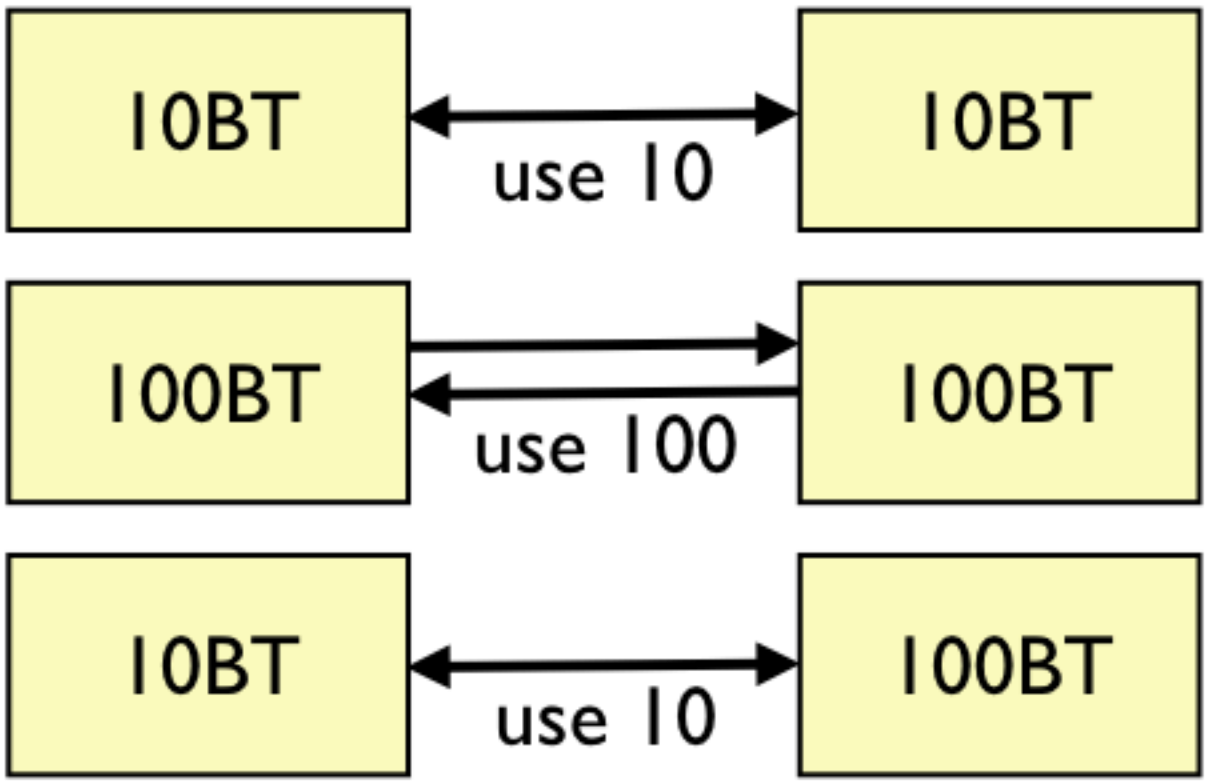
NRZ encoded (2 signal levels)

Fibre Bandwidth is very large

- The bandwidth of the fibre does not limit the bit rate
- There is also no need for scrambling (energy dispersal)

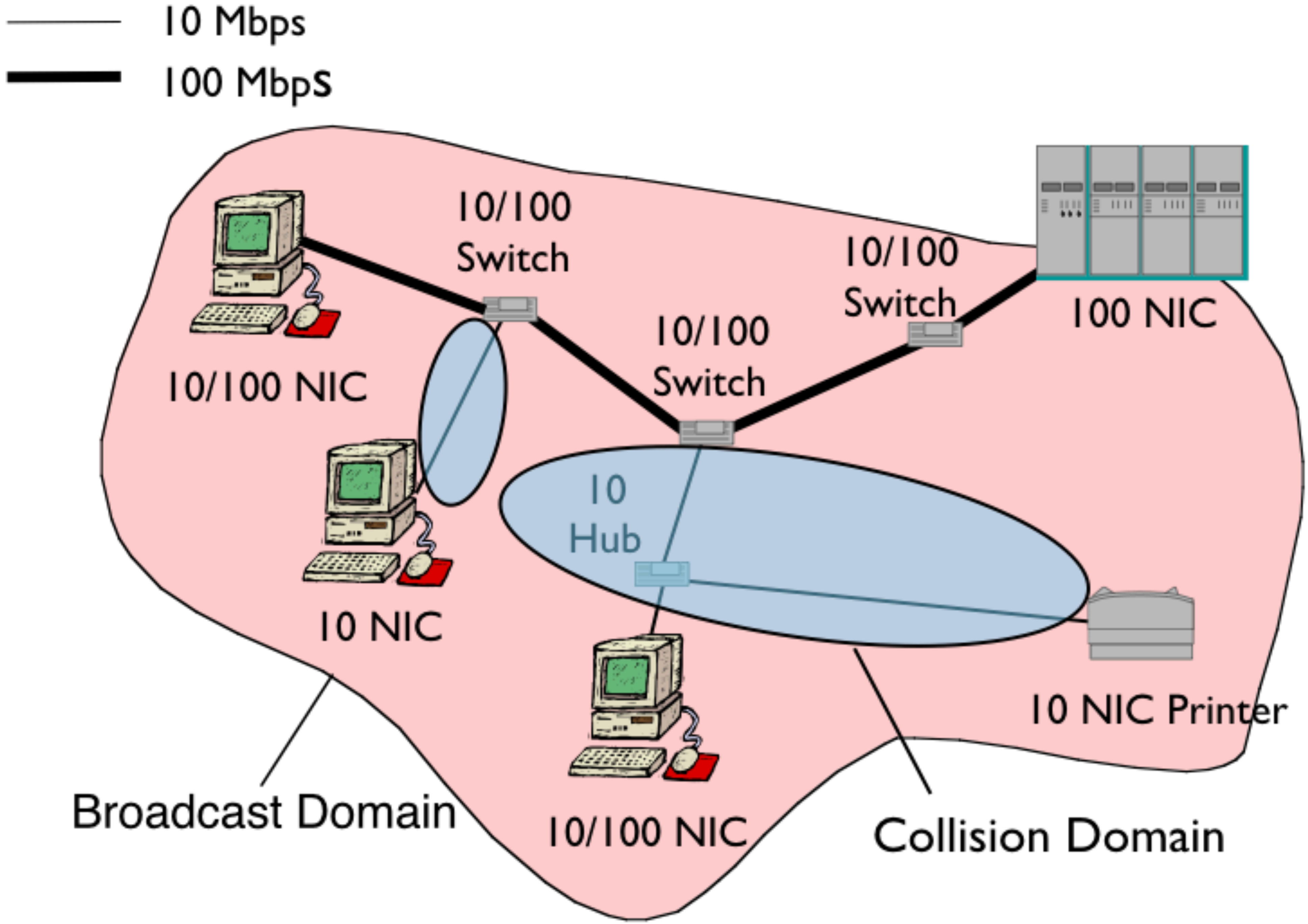
Full Duplex and Speed Auto-negotiation

100BT is full duplex:
A NIC can send at 100 Mbps while receiving at 100 Mbps



Most 100BT NICs also include an embedded 10BT NIC
Auto-negotiation allows systems to find the lowest inter-operable physical layer (including whether to use CSMA/CD)

Collision Domains



Summary Fast Ethernet

- **New 100BT Waveform:**

  4b5b clock encoding
  MLT-3 line encoding to compress bandwidth
  Effect of repetitive sequences on the MLT-3 spectrum
  Scrambling to perform spectral dispersion

- **100BF Fibre interface also specified**

  4b5b clock encoding clock encoding
  NRZ (non-return to zero) line encoding

- **Plug-and-play with 10BT technologies**

  Auto-negotiation to determine segment type
  Connection of cable segments using bridges/switches

- **100B technologies also used by other industries**

100Mbps Industrial Ethernet

Module 6.3

Industrial Ethernet

Industrial Ethernet

A ***range*** of industry standards exist built on Ethernet

Topology depends on application: star, ring or bus

e.g.10BASE-T1L, allows up to 10 in-line connectors to the bus

e.g. A ring topology increases robustness to link failure

Greater Component Protection

Harsher mechanical factors in connectors, and equipment

Greater resilience to electromagnetic interference

More robust switches (e.g. mounted on DIN-RAIL)

100Base-T1: Automotive Ethernet

100 Mbps

One electrical pair

Point-to-Point cables: 15m reach

4 inline connectors

3-level Pulse Amplitude Modulation (PAM)

Echo cancellation, DSP

Power Spectral Density shaping designed for automotive emissions

>10 Gbps Optical physical specification (802.3cy)

Chipset >> x3 the complexity need for 100BTx

Summary Industrial Ethernet

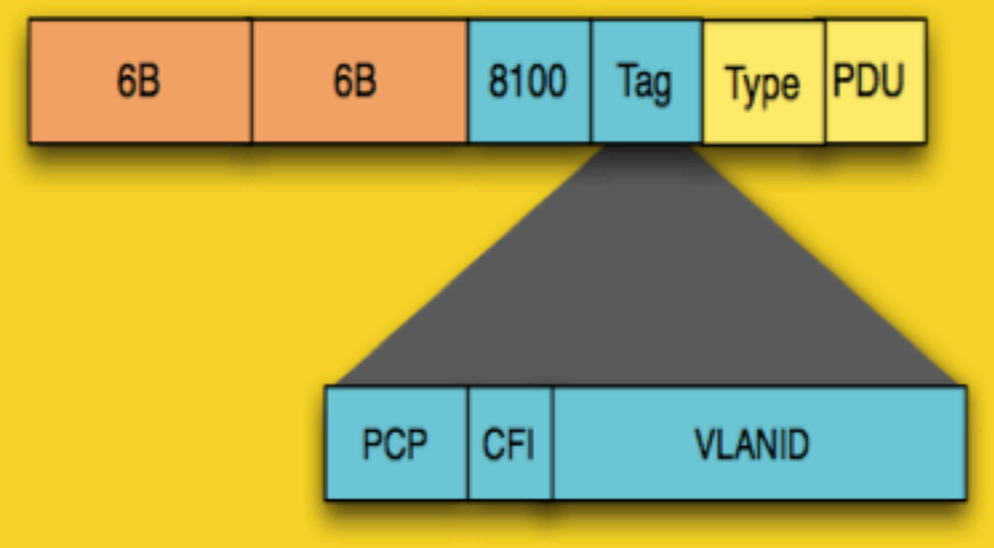
- **A variety of different industry requirements**
  Greater resilience to interference
  Higher cost acceptable in some cases

- **Various standards address these requirements**
  Different physical layer designs
  Different network topologies: Bus, Star, Tree, etc

Virtual LANs

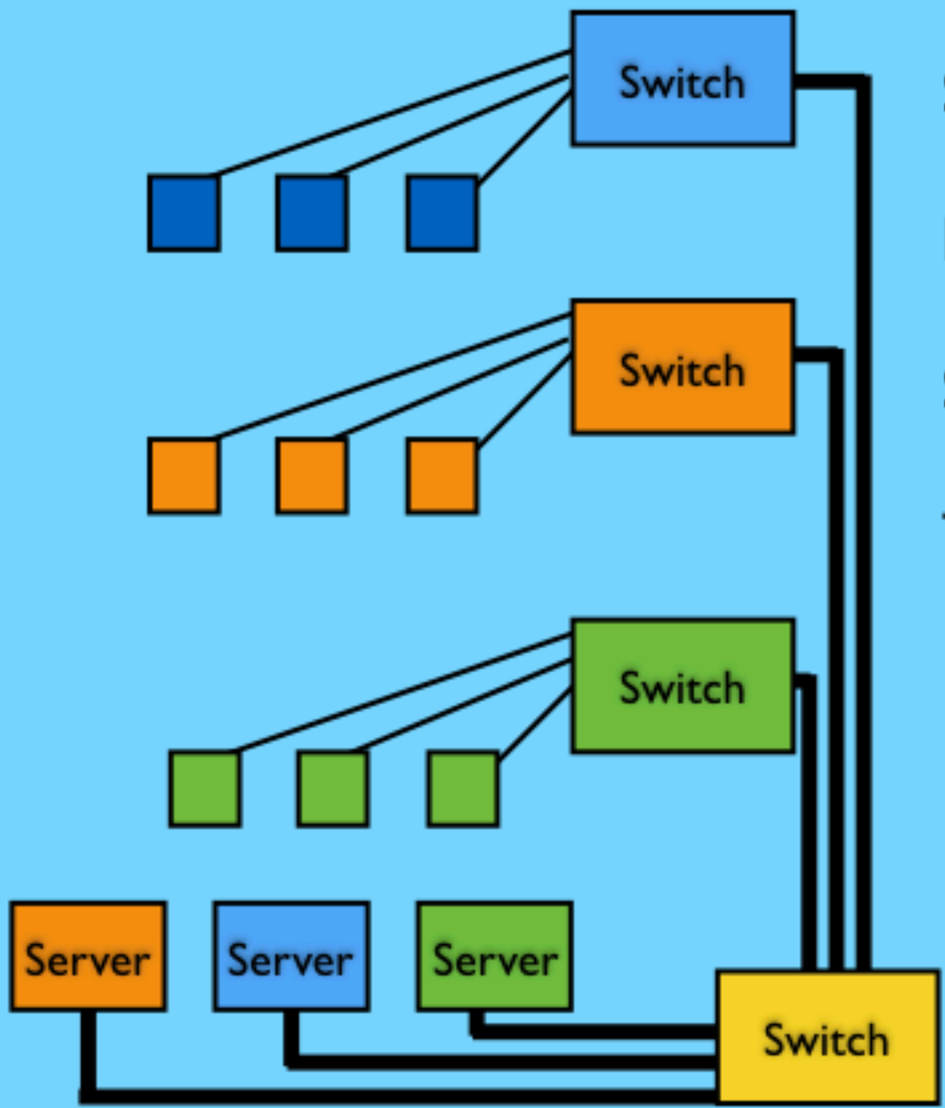
Multiple LANs

Virtual LANs



Module 6.4

Enterprise Stage 3



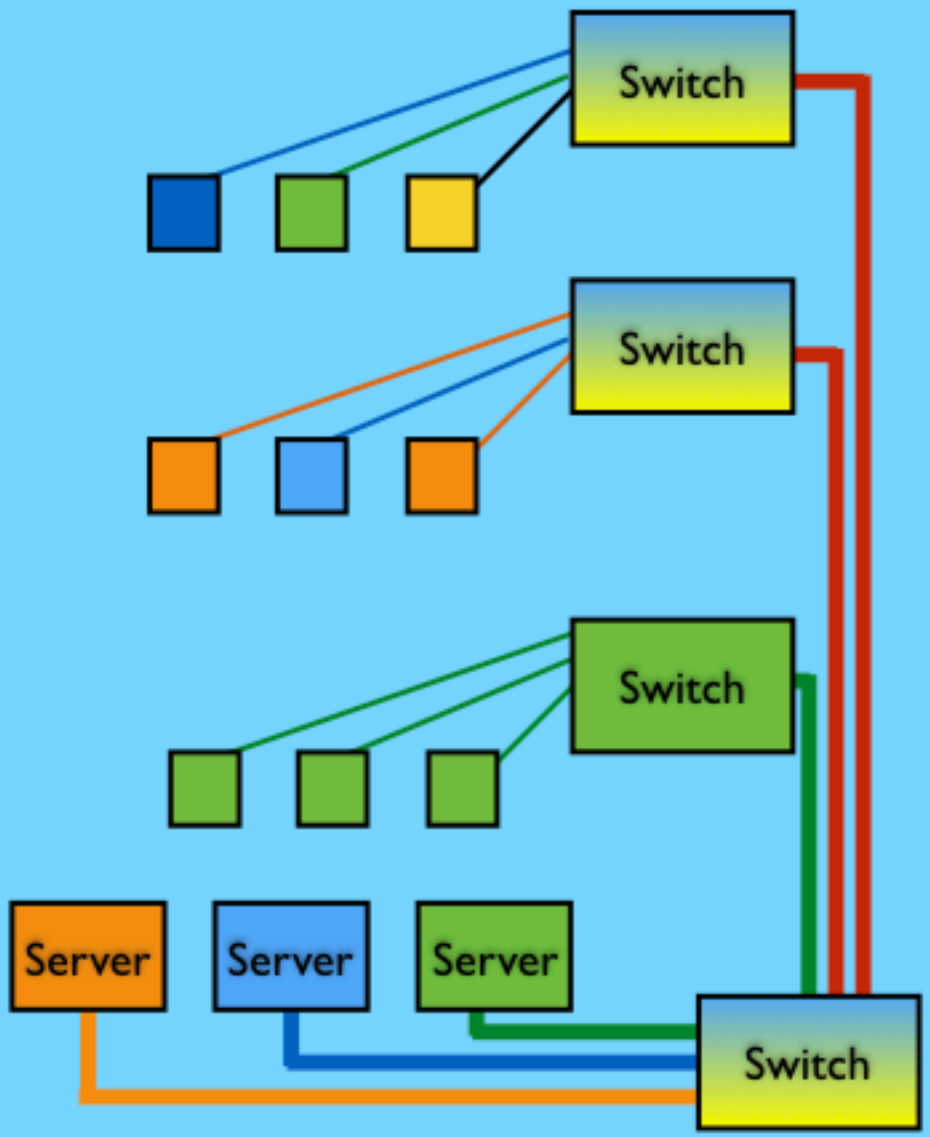
Switches connect workgroups

Higher speed links connect to a switch

Servers connect to the switch

There are multiple LANs

Enterprise Stage 4



Switches are VLAN enabled

There are separate virtual networks

Trunks connect switches
Could carry one VLAN (green trunk)

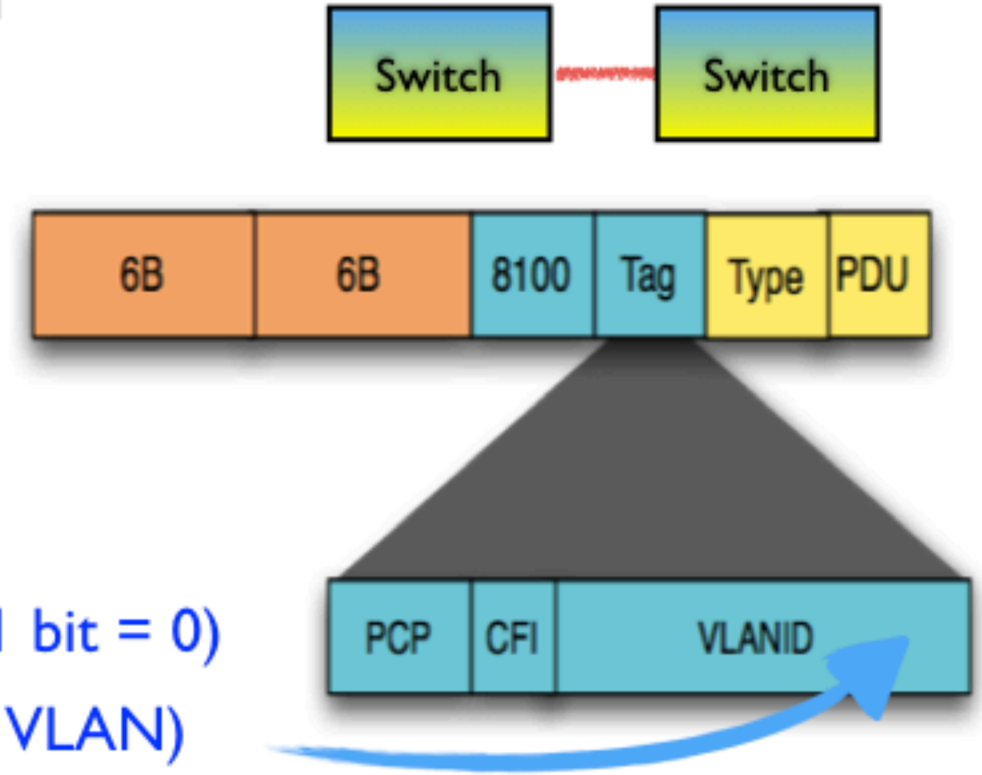
Or many VLANs (red trunk)
The Red trunk uses VLAN Tagging

Tagged Ethernet Frames

An Ethernet frames can include a Tag field

The EtherType for a Tag is set to 8100

A 6 byte tag follows the EtherType



Each IEEE 802.1pQ Tag comprises:

Priority Field (3-bit)

CR - Canonical Format Indicator (CFI) (1 bit = 0)

VLAN-ID (12 b number) (0 indicates no VLAN)

The last 2 bytes is the EtherType of the encapsulated Payload (PDU)

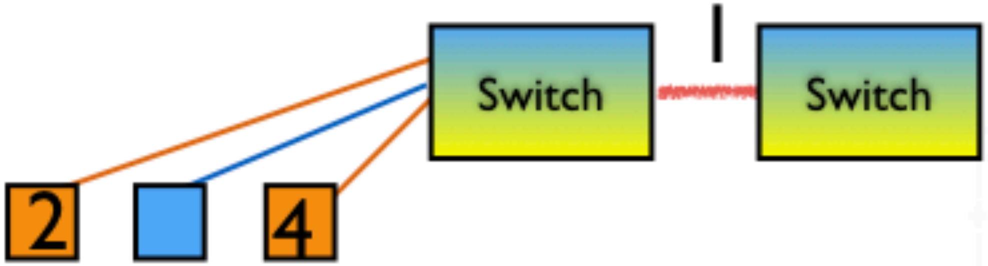
Tagged frames are often sent between VLAN switches

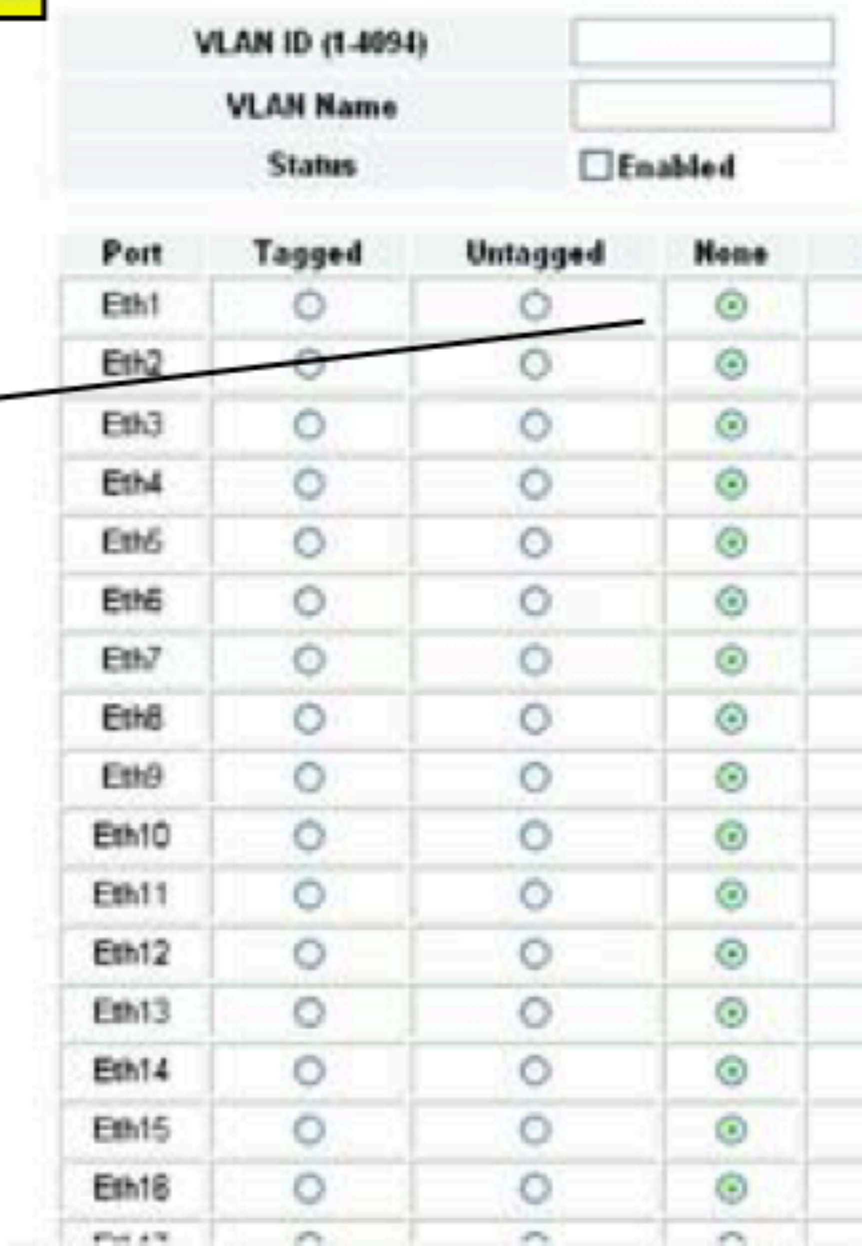
Interface port adds a tag to each sent frame indicating the VLANID

Interface port checks the tag on each received frame and removed

Indicated VLANID used in the address table to lookup how to forward frame

VLANs





VLAN ID (1-4094)	
VLAN Name	
Status	☐ Enabled

Port	Tagged	Untagged	None
Eth1	○	○	◉
Eth2	○	○	◉
Eth3	○	○	◉
Eth4	○	○	◉
Eth5	○	○	◉
Eth6	○	○	◉
Eth7	○	○	◉
Eth8	○	○	◉
Eth9	○	○	◉
Eth10	○	○	◉
Eth11	○	○	◉
Eth12	○	○	◉
Eth13	○	○	◉
Eth14	○	○	◉
Eth15	○	○	◉
Eth16	○	○	◉

Each port is in one of 3 modes:

None

- VLAN cannot use this port

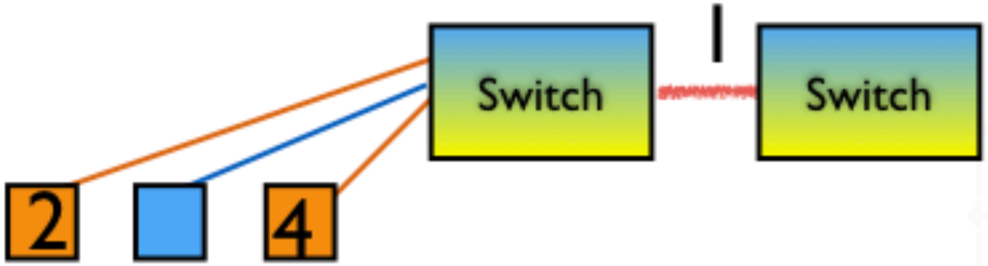
Tagged

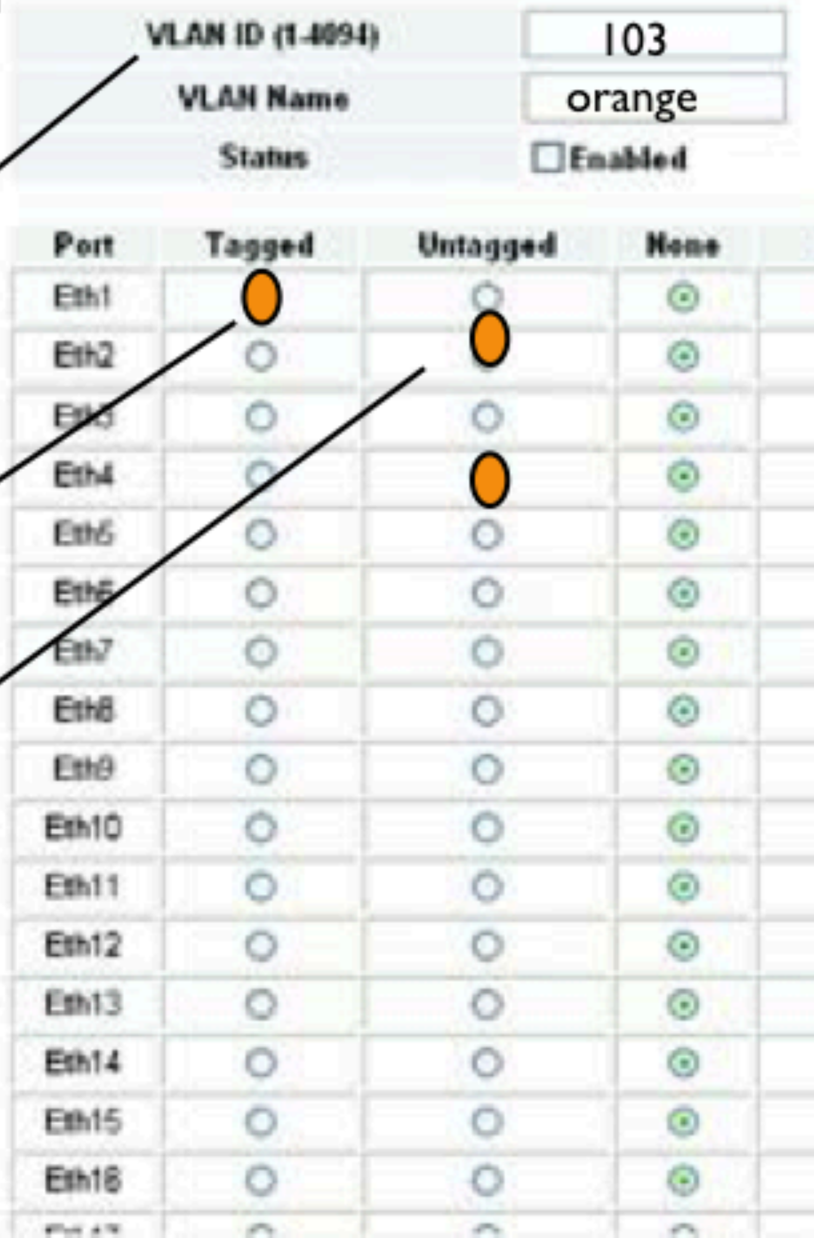
- All frames sent tagged

Untagged

- Frames sent untagged
- for a specific VLAN

Orange VLAN





Each VLAN is named

- And assigned a VLANID (tagged as 103)

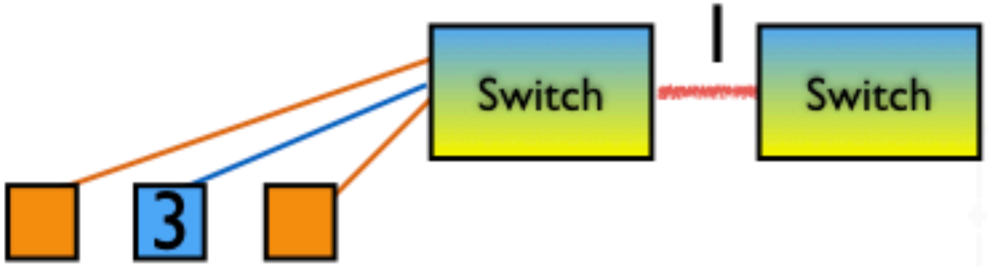
Tagged port interfaces

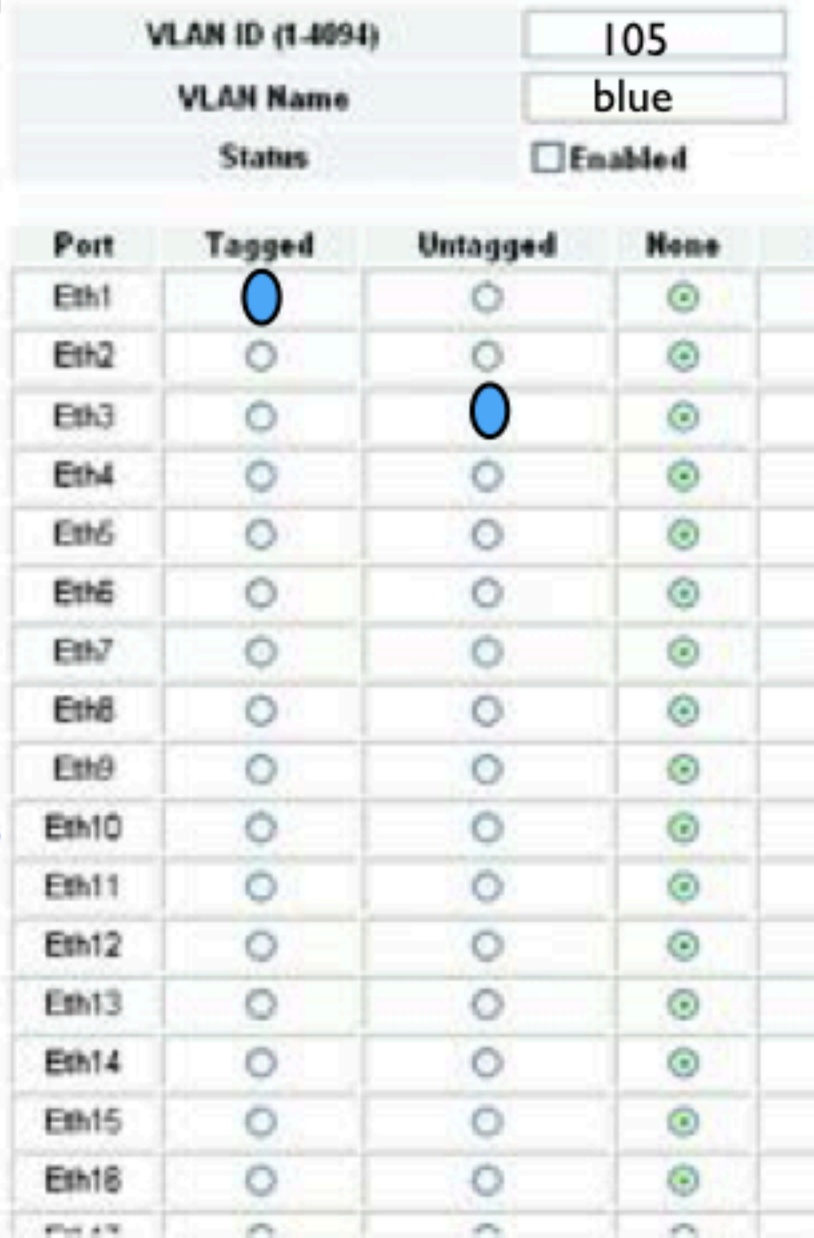
- All frames sent tagged (using VLAN103)

Untagged port interfaces

- Frames sent untagged (using VLAN103)

Blue VLAN





By default, interfaces assigned to None group

Tagged ports can belong to a VLAN

Eth1 carries VLAN105 tagged; Eth3 untagged.

Virtual LANS: VLANs

- **VLANs add information in the address table**
  - Each port can be configured to belong to a VLAN
  - The VLAN groups ports into separate broadcast domain
  - Broadcast packets flooded only within a VLAN

- **Trunk Ports**
  - A port can be configured to become a trunk port
  - A VLAN Tag header is added to each frame sent
  - The VLAN-ID to identify the VLAN associated with a frame
  - The same trunk port can be a member of multiple VLANs

- **Each VLAN is a separate IP network**
  - A server might connect a trunk port
    - It can receive packets from multiple VLANs
    - Each VLAN is treated as a separate interface
  - A router can forward packet between different VLANs

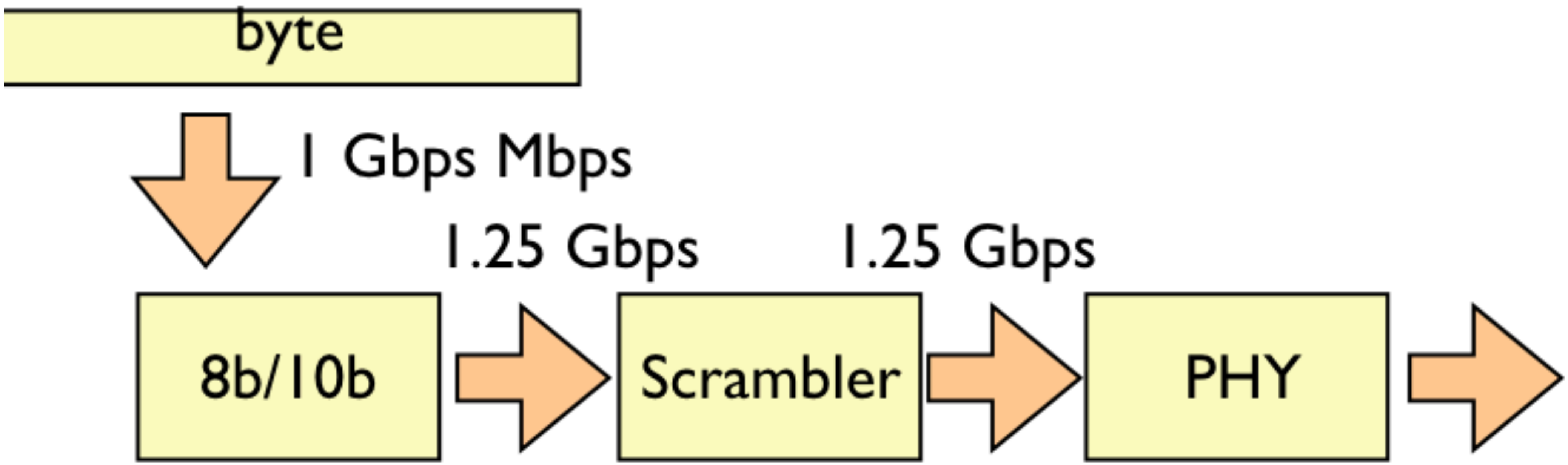
Gigabit Ethernet

1000 Mbps

1 Gigabit Ethernet

Module 6.5

GBE Transmission



8 bits (1 byte) processed at a time

8 bits encoded to 10 bits (constant disparity)

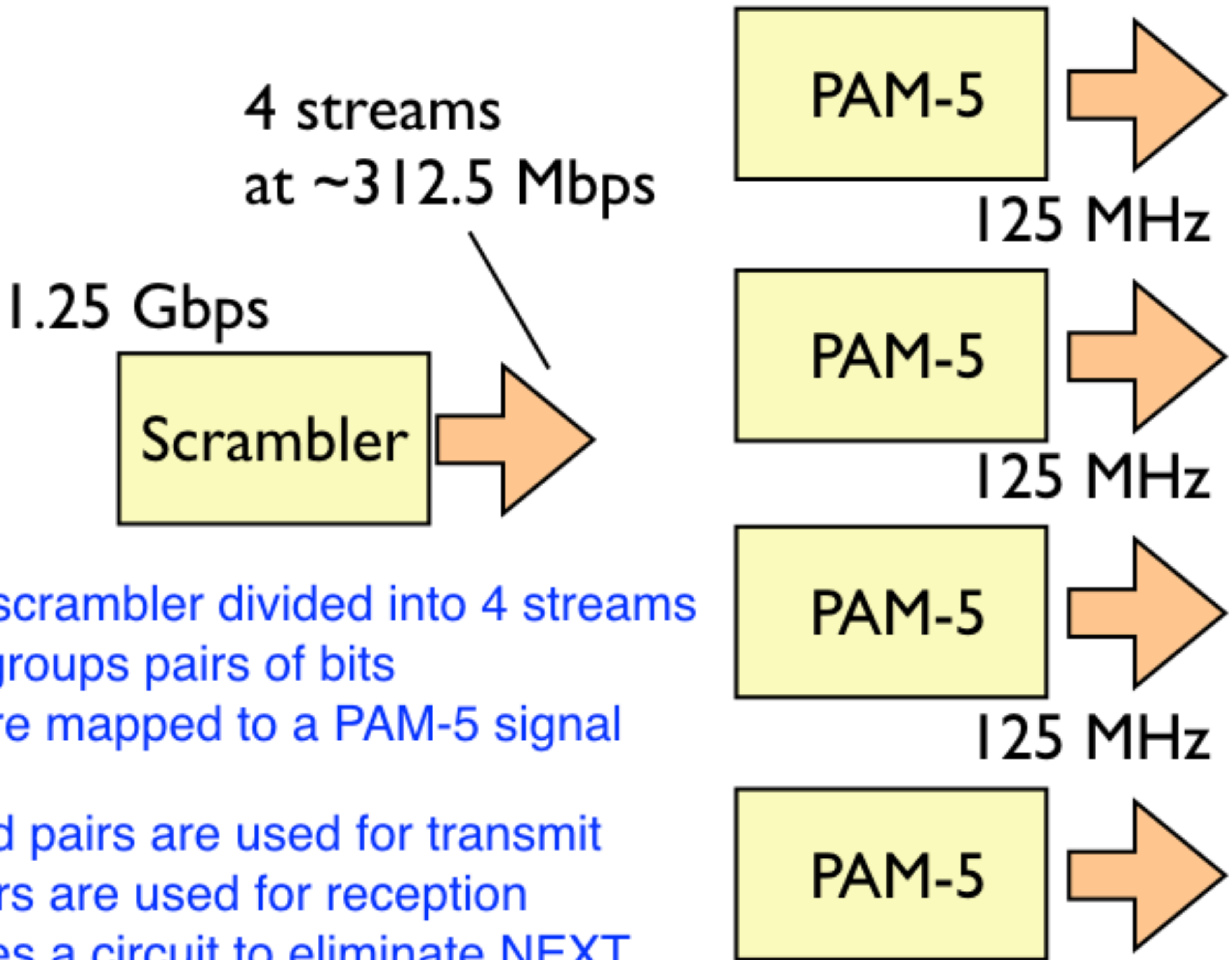
- Each value contains 5 ones or 5 zeros

Scrambled

- Bits randomised to disperse energy

1000BT PAM-5 Transmission over UTP

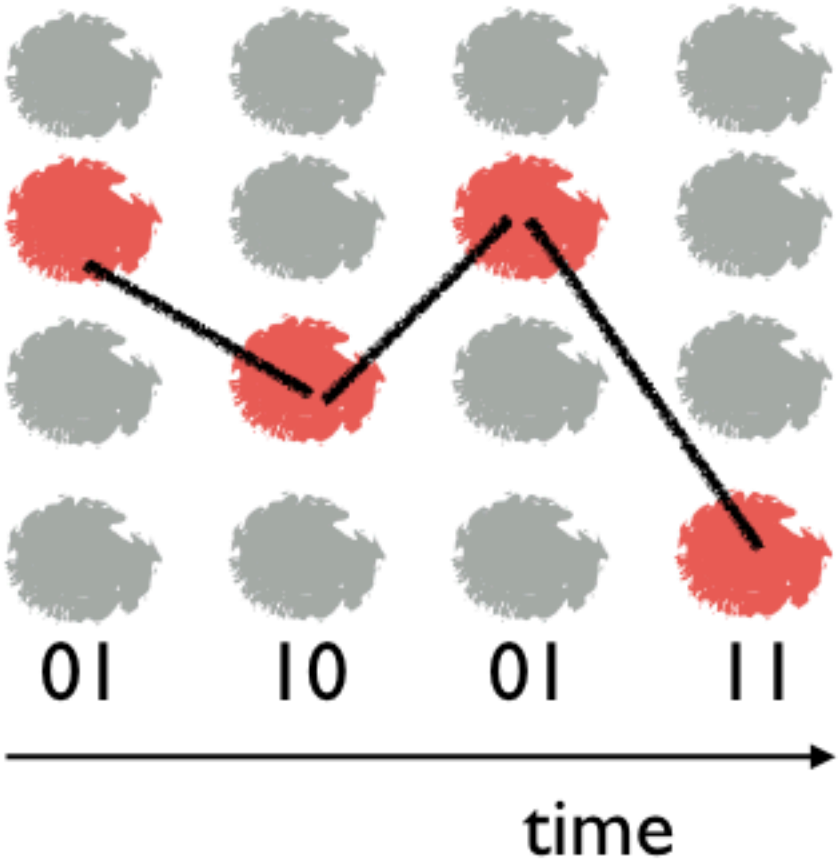
Signal needs to be with 125 MHz limit of CAT-5e



Output of the scrambler divided into 4 streams
Each stream groups pairs of bits
Pairs of bits are mapped to a PAM-5 signal

All four twisted pairs are used for transmit
The same pairs are used for reception
 This requires a circuit to eliminate NEXT

PAM-4 Transmission



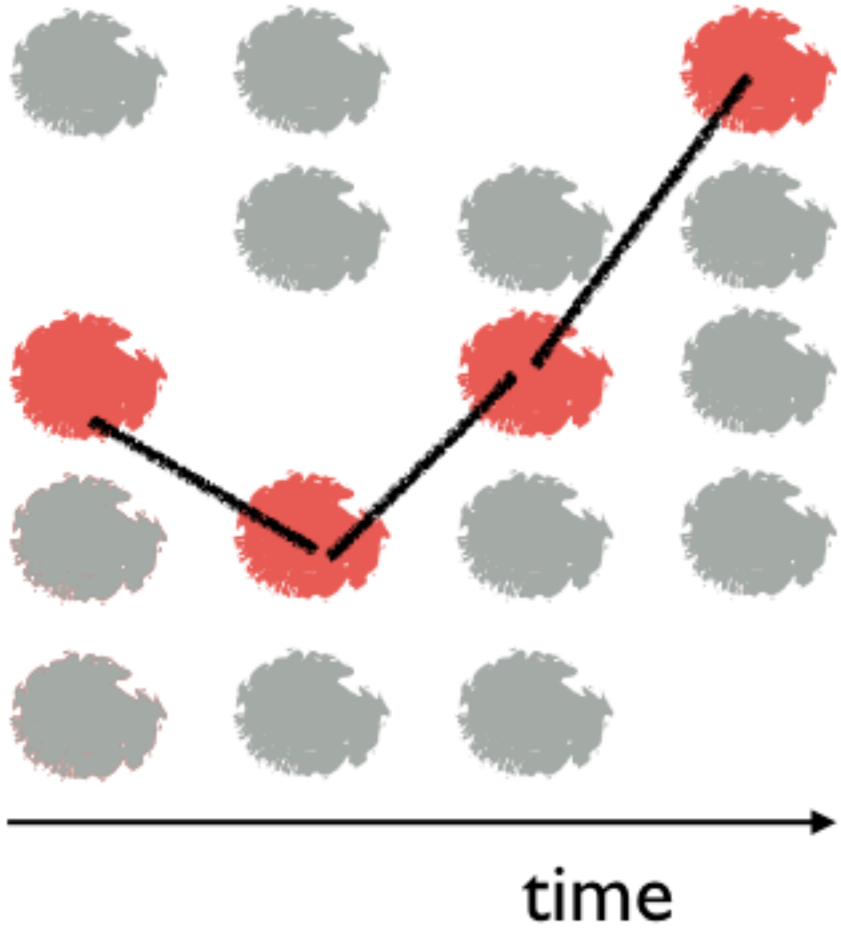
bin	Line signal
00b	+2
01b	+1
10b	-1
11b	-2

Groups two bits & maps into one PAM-4 baud (4 level)
Any transition is allowed between bauds (not just adjacent level as in MLT)

A receiver decodes one baud to 2 received bits

Not used in 1000BT

1000BT PAM-5 Transmission



Line signal
0
+1
+2
-1
-2

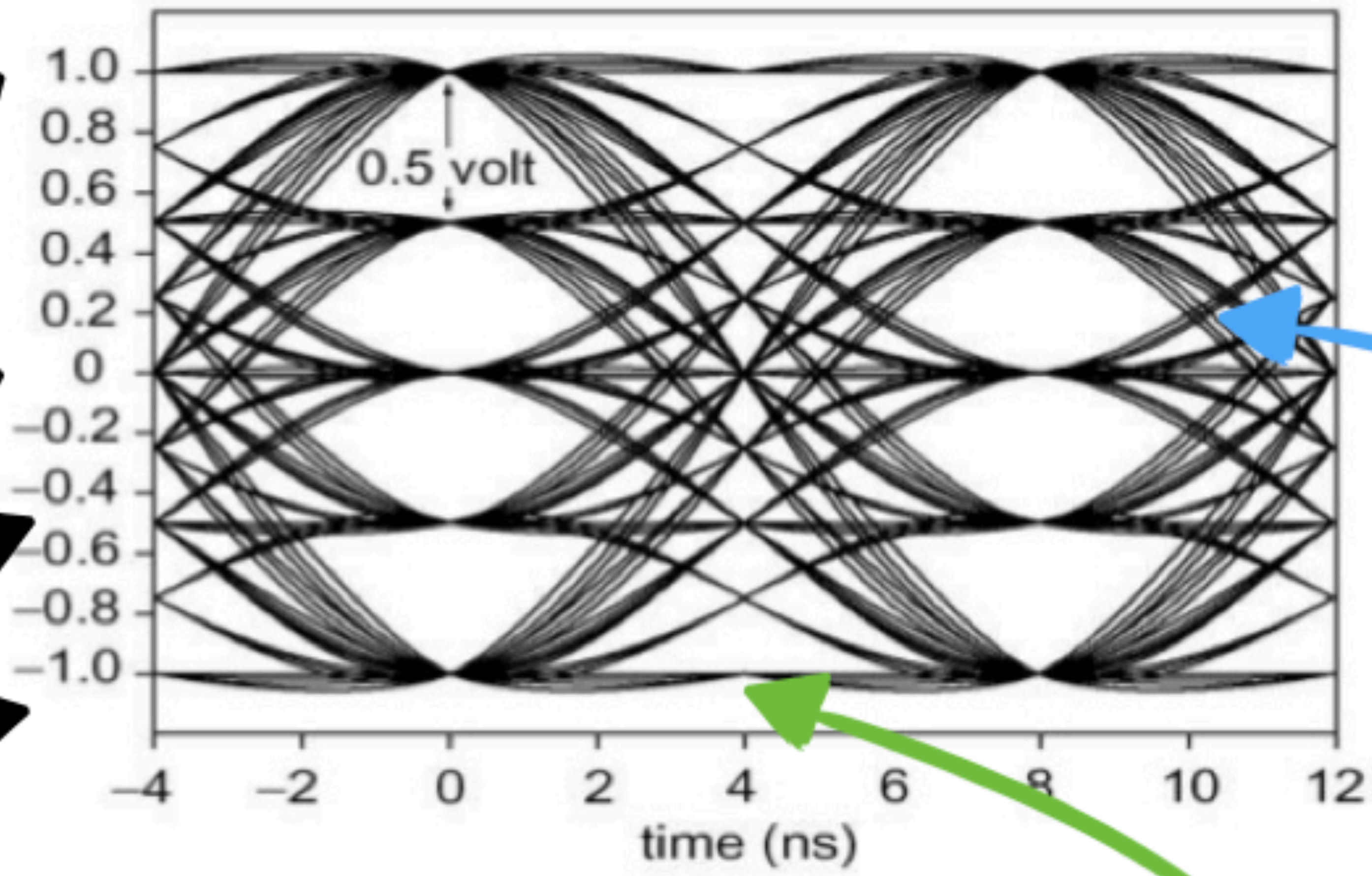
Method used in 1000BT

Groups two bits & maps to PAM-5 signal (5 levels)

4D mapping of the 2b to PAM-5levels is complex - that changes each baud and is designed to optimise immunity to noise across all 4 pairs in the cable

Overall FEC results in a 6 dB signal to noise improvement

Eye Diagrams showing PAM-5

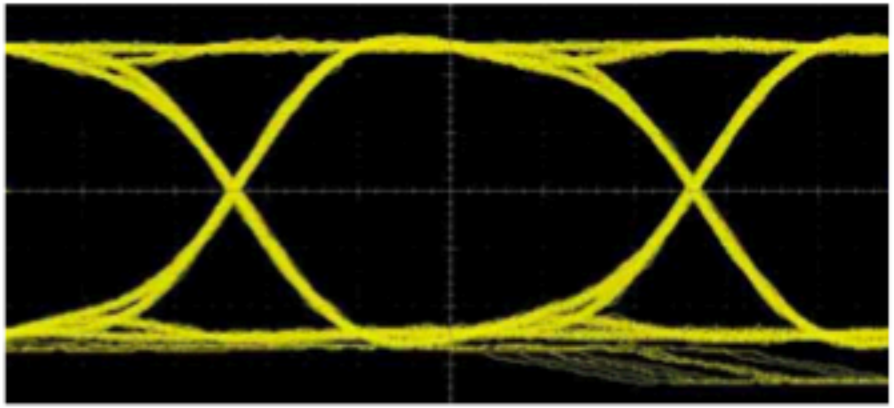


Five distinct levels are clear (no noise in these plots)

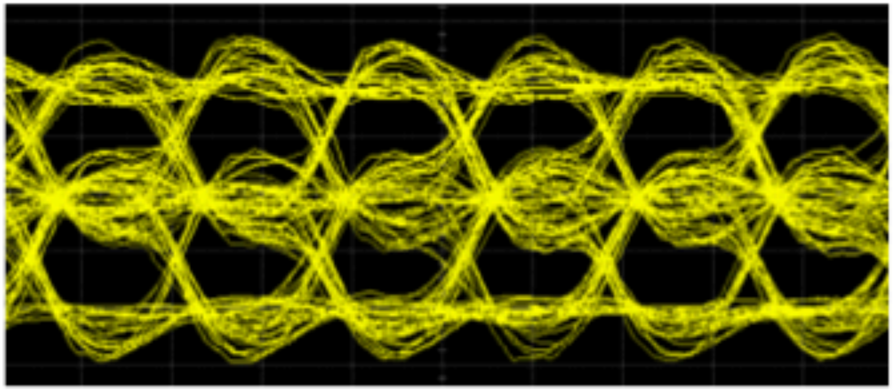
Transitions occur to **any** of the 5 next levels

The slew rate is limited, i.e. the rise time for transitions

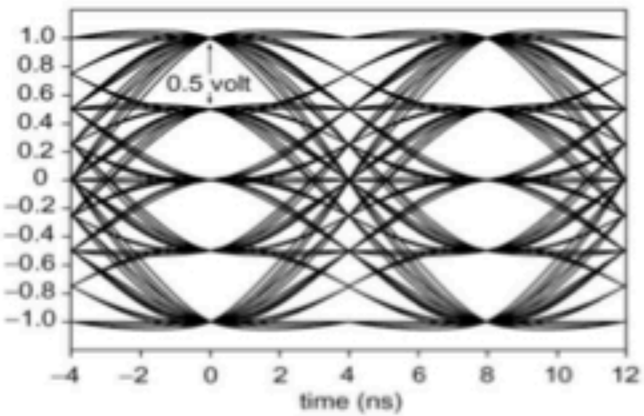
Eye Diagrams showing the various waveforms



Manchester signal

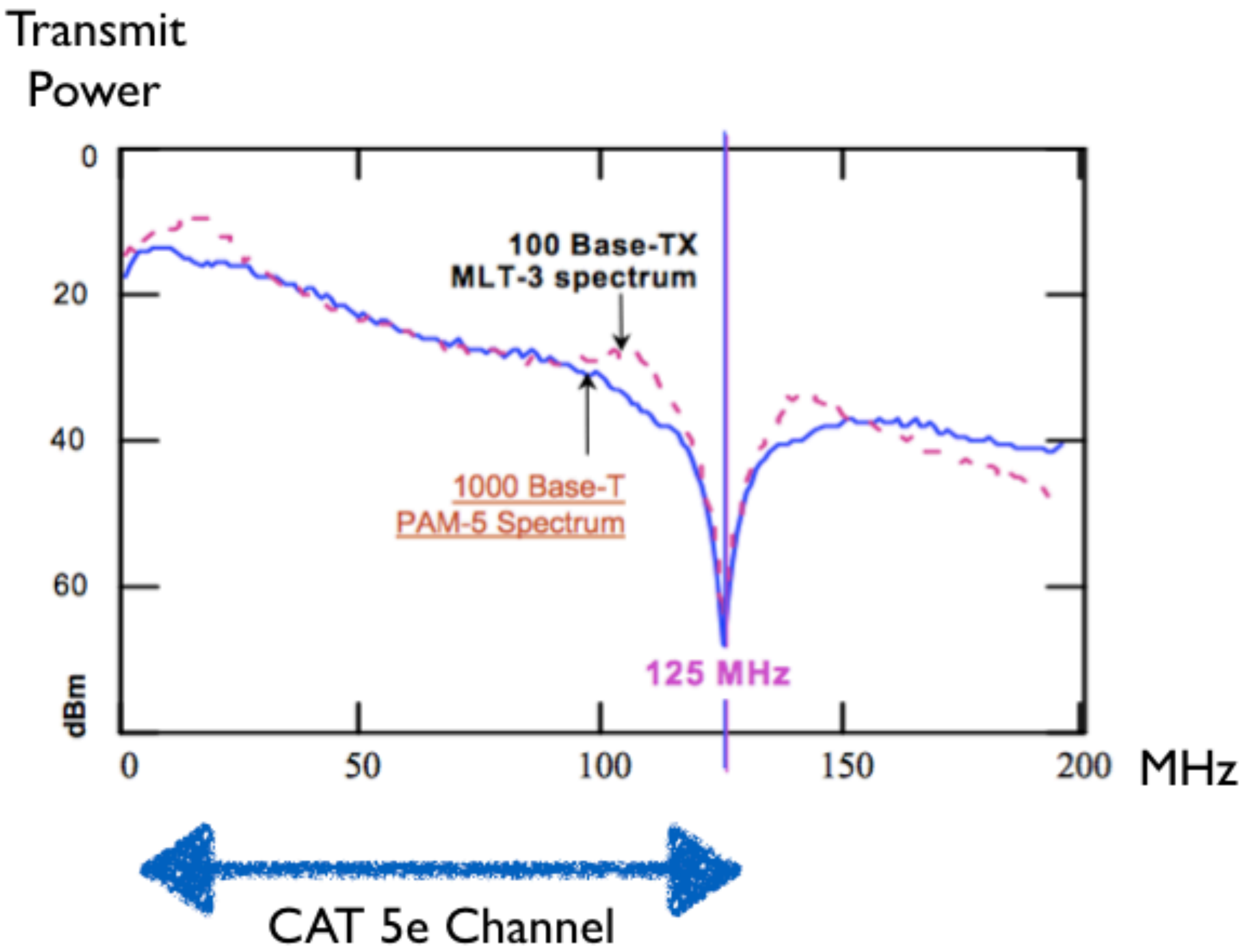


MLT3 signal



PAM5 signal

Gigabit Ethernet Spectrum

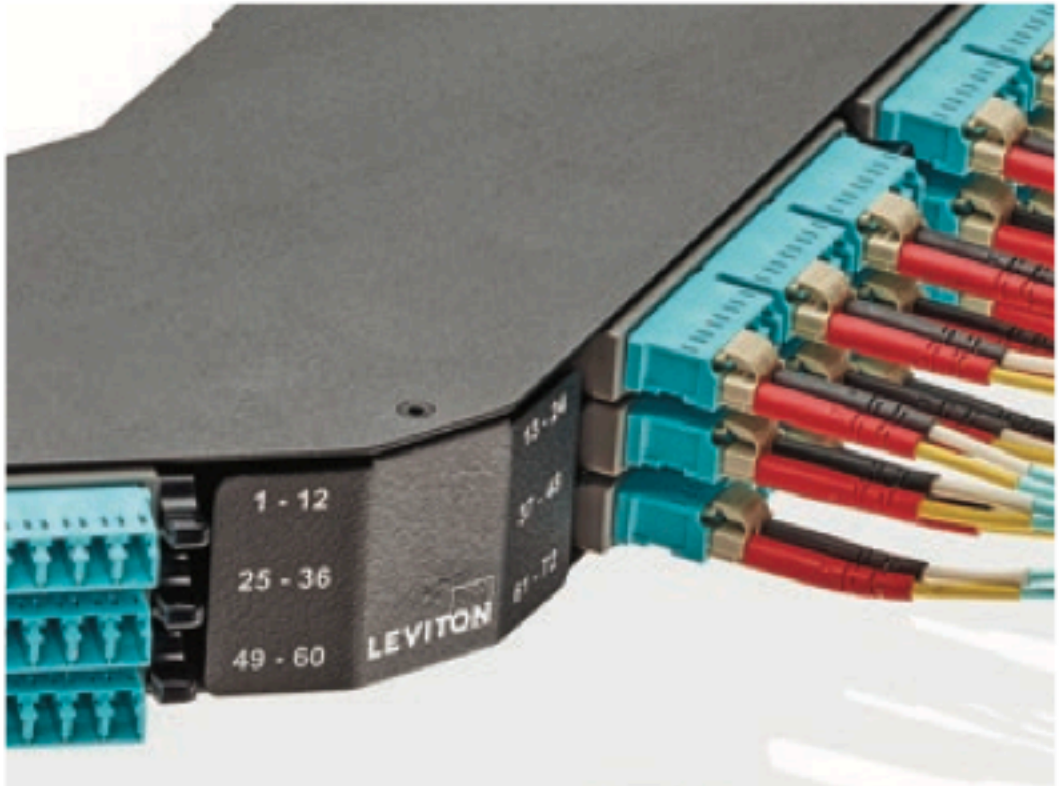


Gigabit Fibre Ethernet

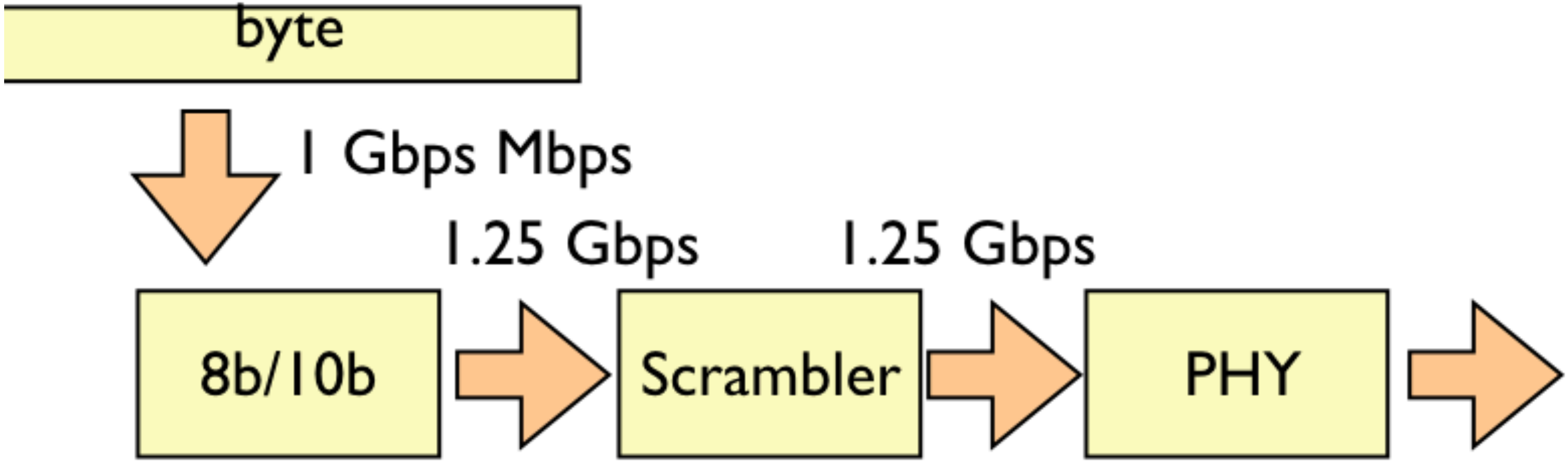
Standardised by IEEE 802 Committee

1 Gbps Fibre - Distance

5km	(short haul)
70km	(long haul)
300 km	(with optical repeaters)

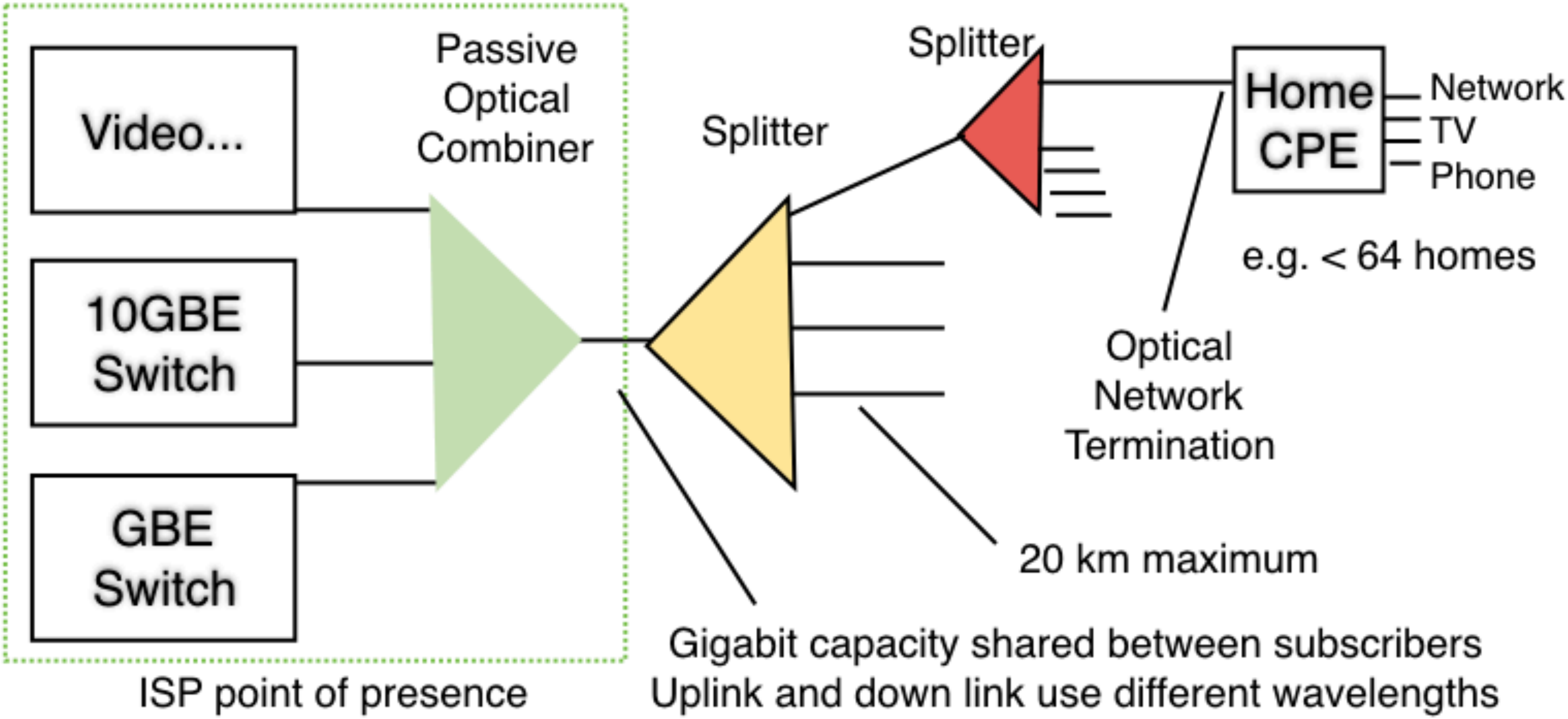


GBE Fibre Transmission



Transmitted directly using the PHY (i.e. using a Fibre transceiver)

E/G Passive Optical Network To The Home



GBE can also be used to connect homes to ISPs

This is shared technology for the "last mile" of an ISP

Unpowered passive optical splitters (fibre <20 km)

12-256 fibres linked to a splitter, typically <64

Specified 2004, by 2014, over 40 million installed EPON ports

Slot Time in GBE

10 Mbps Traditional Ethernet (Half Duplex)

- Bit time 0.1 μS
- Minimum frame size (512 bits), Slot time 51.2 μS for CSMA/CD

100 Mbps Fast Ethernet (Full Duplex)

- Bit time 0.01 μS
- Kept same minimum frame size (512 bits), slot-time 5.12 μS
- 5-4-3 rule had to be abandoned (Hubs are seldom used)

1000 Mbps GBE (Full Duplex)

- Bit time 0.001 μS (1 nS)
- IFG 12B (0.096 μS)
- For 1500B payload, n=1526 (incl overhead) = 12.304 μS
- Small frames VERY inefficient
- 64B frame => (64)/(512+12) = 12% efficiency

GBE allowed several frames to be sent as a burst

- Burst of small frames allowed with total size up to 8192 bytes

Summary GB Ethernet

- **1000BT Waveform:**

  Uses all four pairs of UTP (all eight wires)
  Data sent as four streams
  8b/10b line encoding
  PAM-5 signal compresses waveform (with complex mapping)
  Scrambling to perform spectral dispersion

- **Range of fibre interfaces also specified**

- **Plug-and-play with 10B / Fast Ethernet technologies**

  Auto-negotiation to determine segment type

  Connection using bridges/switches

-

10 Gigabit Ethernet ++

10 Gbps

10GT 802.3an (2006)

Variety of products had emerged by 2014; common in data centres

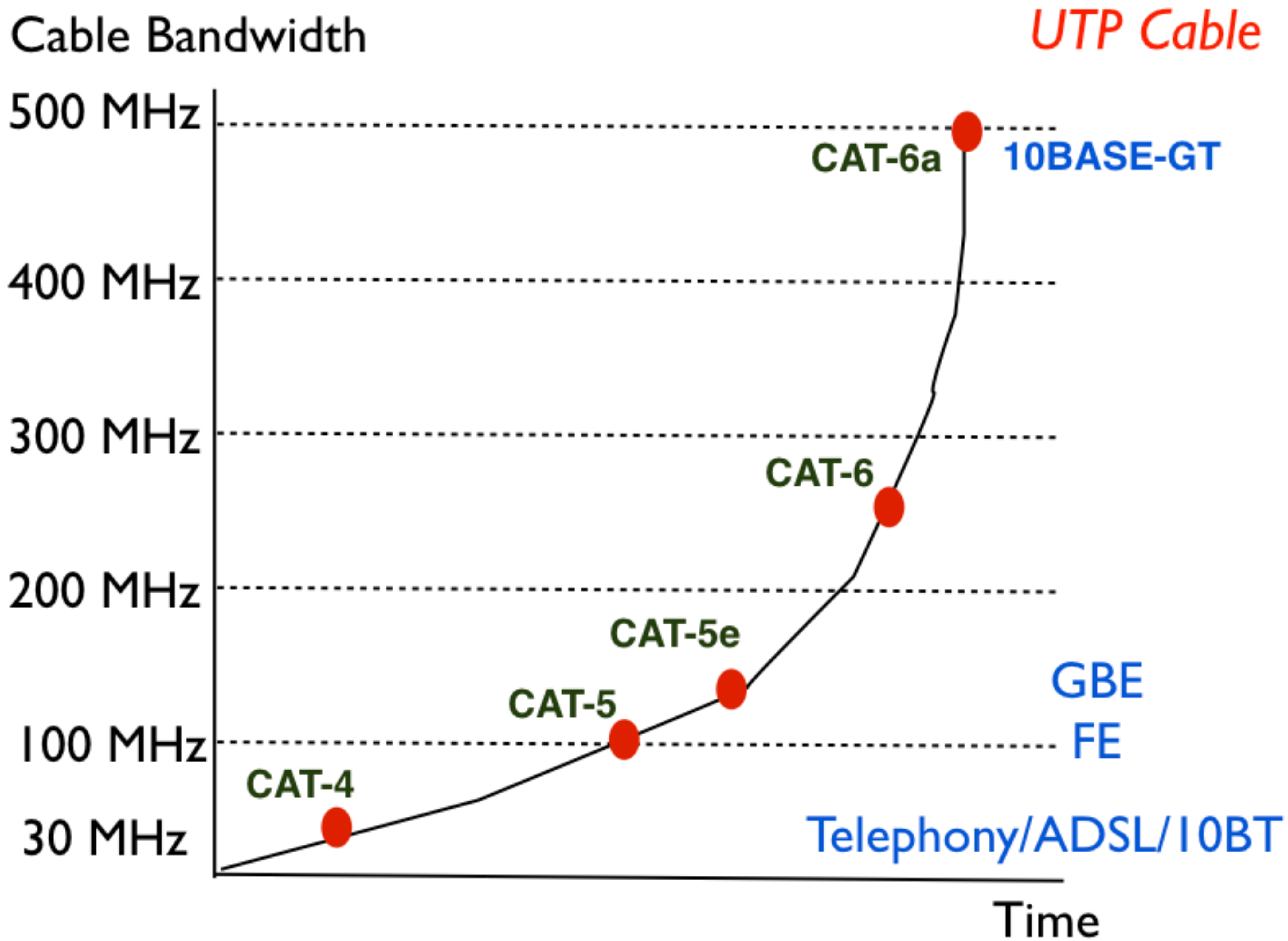
Waveform:

64B/66B encoding (3% overhead), effective rate of 10.3 Gbps
Encoded as 128 DSQ with LDPC*
Symbol rate 3,200 000 000 Symbols/sec

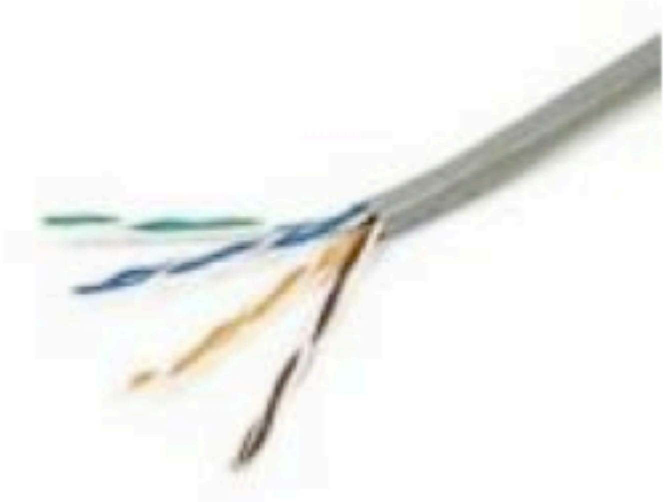
10G BASE-T

UTP over much shorter distances
CAT6 cable <~ 55 m**
Screened CAT6a cable <100m

** Transceiver significantly more complex - approx 10M gates, 10W*
*** This shorter distance or higher spec cable targets data centre applications, rather than home/enterprise applications.*


Cable Bandwidth
UTP Cable
500 MHz
400 MHz
300 MHz
200 MHz
100 MHz
30 MHz
CAT-6a
10BASE-GT
CAT-6
CAT-5e
GBE
CAT-5
FE
CAT-4
Telephony/ADSL/10BT
Time

Category 6/6a Cabling



Thicker wires that are much more tightly twisted
Cross-shaped former in centre
Better cable insulation

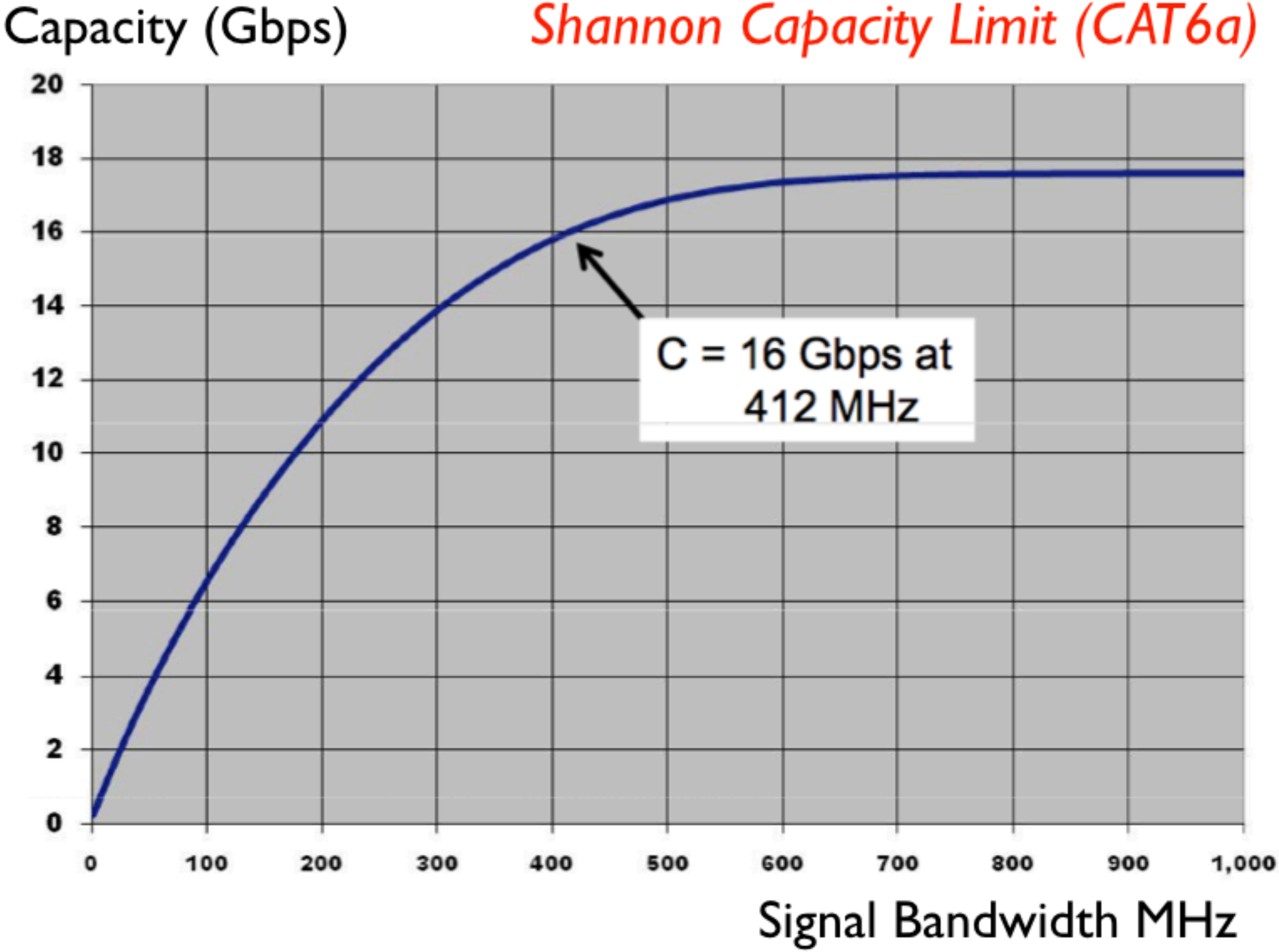
CAT-6
250 MHz bandwidth
Maximum length: 100m (Max 90m solid wire)
10BASE-GT limits length to 55m

CAT-6a
500 MHz bandwidth
Maximum length 100m (Max 90m solid wire)
10BASE-GT limits length to 100m

Results in a ***much thicker*** cable

Current cost x2 for CAT-5e

Shannon Capacity Limit (CAT6a)



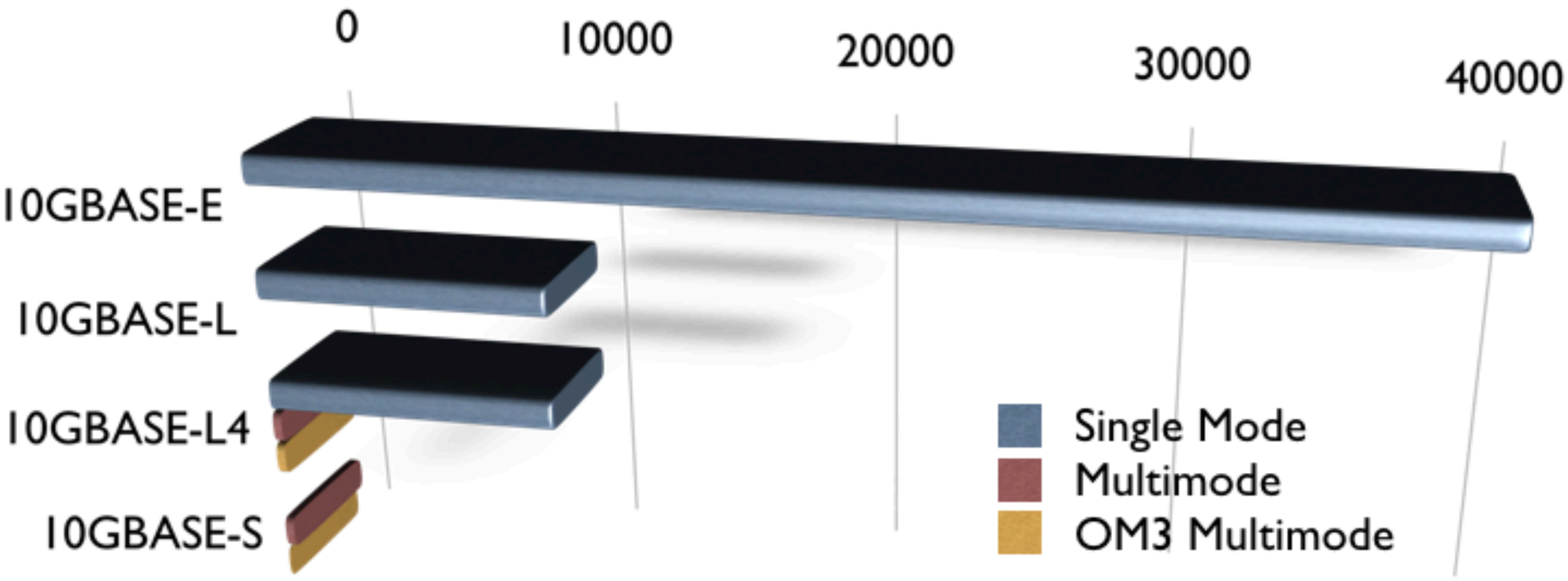
10 Gbps over fibre

Two sets of versions

LAN & WAN versions share a common “transceiver” format

Various fibre physical “sublayers” have been defined

64b/66b encoding (10GBASE-X uses 8b/10b)



Work in 2017 on 10km optical pays for 50, 200 and 400 Gbps

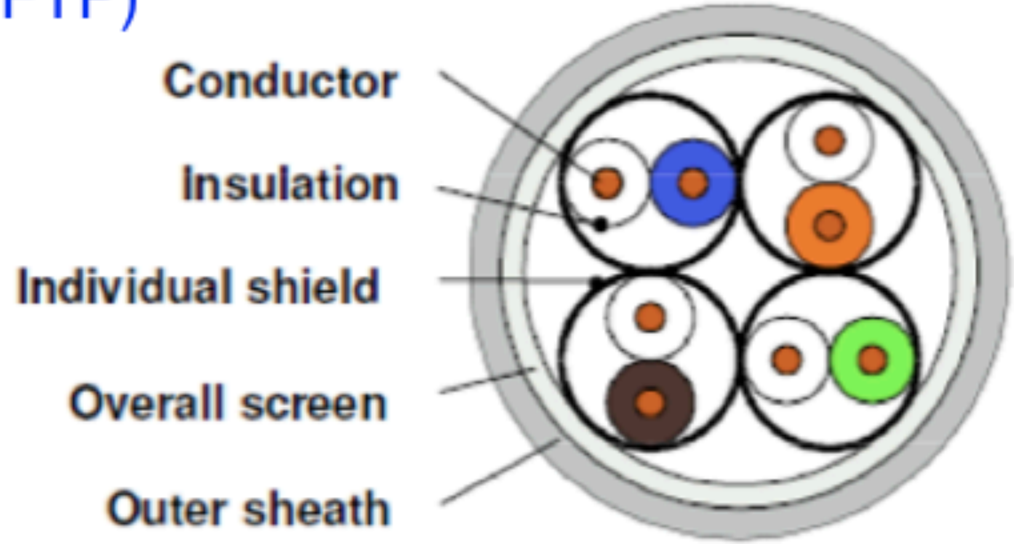
40 Gbps over copper

40G BASE-T

IEEE 802.3bq over copper UTP (From 2013)

- 4 Pairs up to 30 m*
- Each pair operating at 10 Gbps (128 DSQ coded with LDPC)
- Signal bandwidth 1600 MHz (SFTP) **

Cat 7   Cat7a



*** Shorter distance targets data centre applications, rather than home/enterprise applications.**

Category 8 Cabling

Shielded cable

Bandwidth of up to 2 GHz (2000 MHz)

Higher Cost! used for high rate digital video

100 Gbps over Copper

10 Mbps (1 lane)*

100 Mbps (1 lane)

1 Gbps (1 lane)

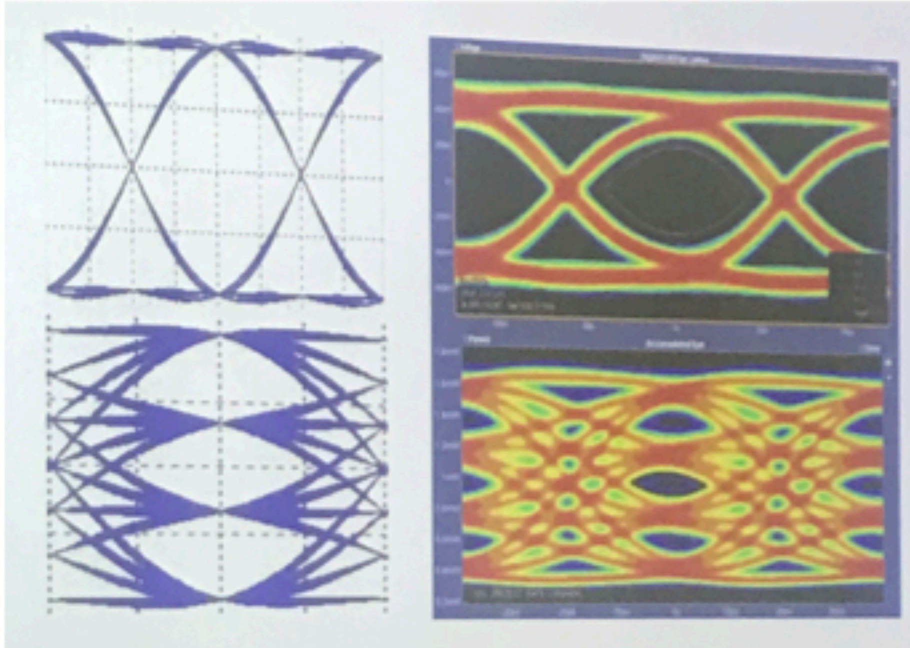
10 Gbps - Data Centre (1 x 10 Gbps lane)
40 Gbps - To high-end server (4 x 10 Gbps lanes)
25 Gbps - (1 x 25 Gbps lane)
50 Gbps - Data Centre (2 x 25 Gbps lanes)
50 Gbps - Wide area (2 x 50 Gbps lanes)

100 Gbps - Data Centre (4 x 25 Mbps lanes)

100 Gbps - Wide area (10x10 Gbps lanes) 2020
200 Gbps (2 x 100 Gbps lanes) standard planned for 2021
400 Gbps (2 x 100 Gbps lanes) standard planned for 2021

*A lane is a processing engine where a 64b data block is sent and reassembled at the remote end.

100 Gbps over Fibre Optic Cable



<=25 Gbps

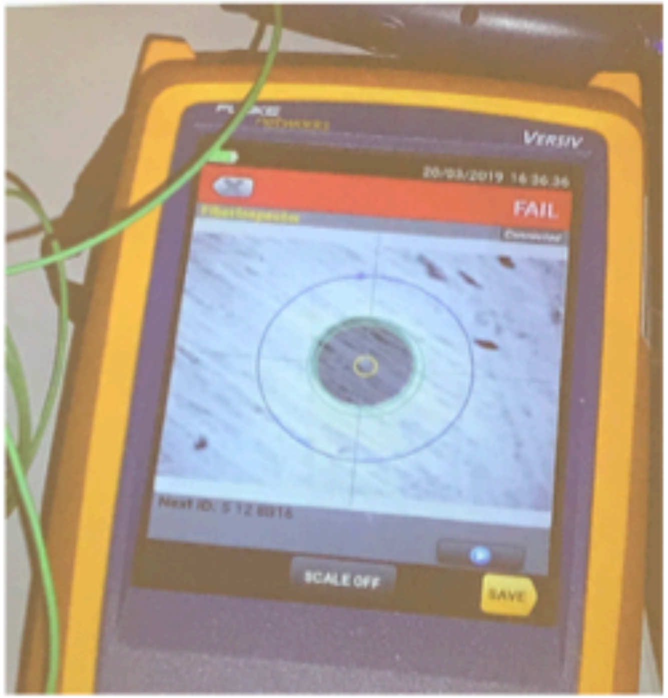
- On/Off (1 bit/pulse)
- 100 Gbps using 4 lanes of 25 Gbps

>= 50 Gbps < 100 Gbbps

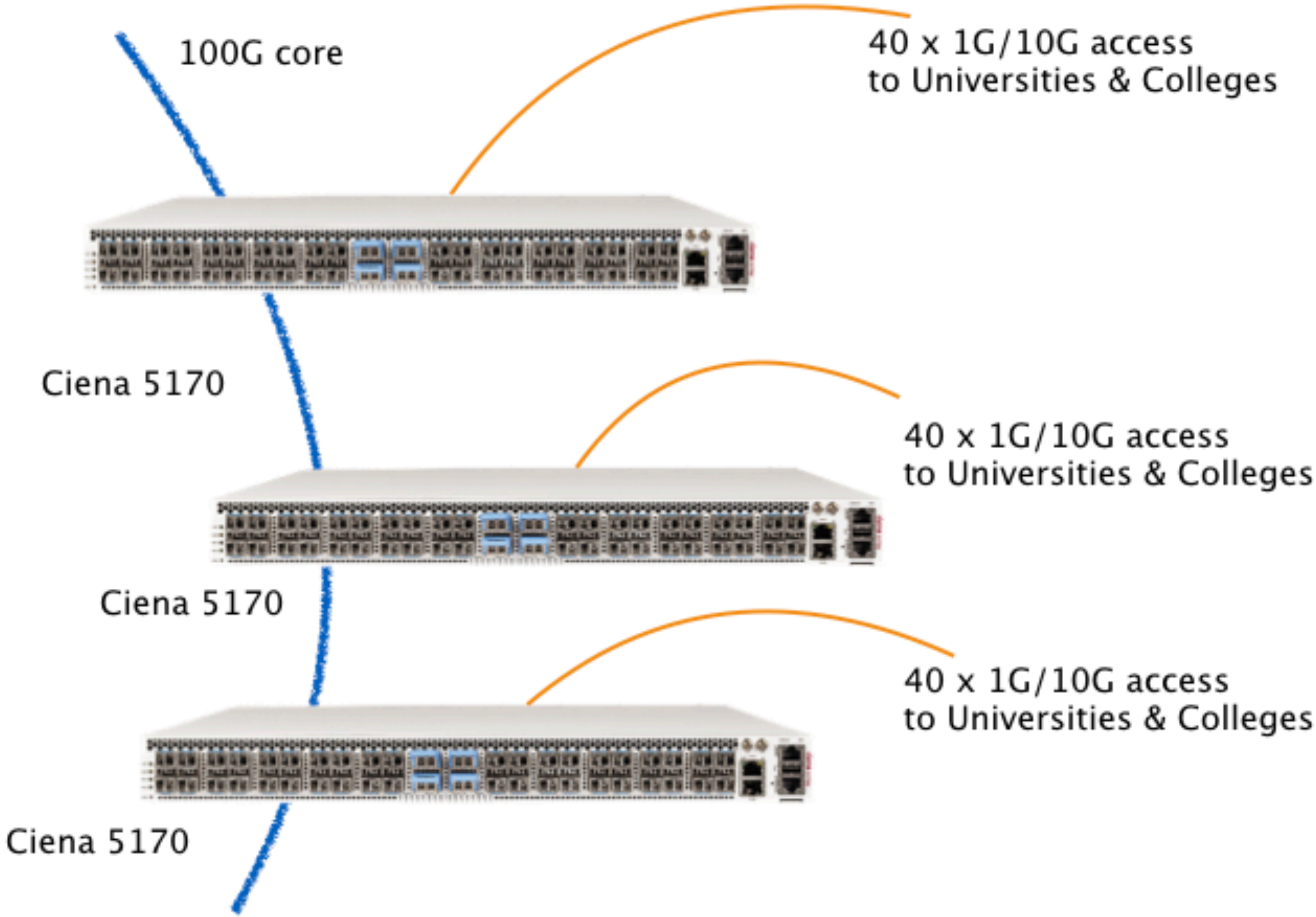
- PAM-4 (2 bits/pulse)
- Signal processing needed

100 Gbps

- QAM-16 (4 bits/pulse)
- i.e. 12.5 Gbaud
- Quality of lightpath becomes critical

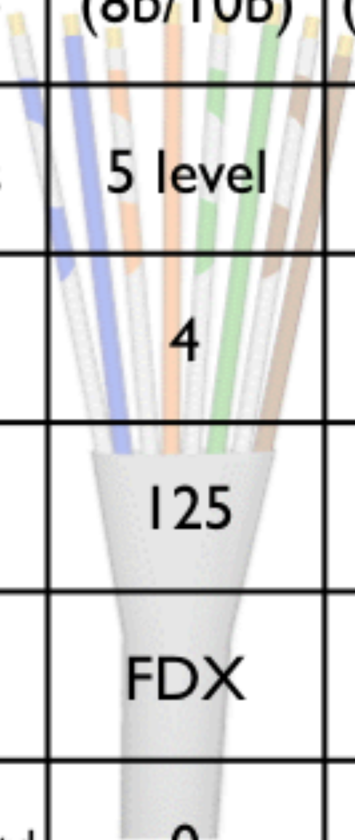


100 Gbps over Fibre



Evolution of the Ethernet Specification

	10Mbps	100Mbps	1 Gbps	10 Gbps	40 Gbps	100 Gbps
Cable	Fibre UTP Coax	Fibre UTP (CAT-5)	Fibre UTP (CAT5e)	Fibre UTP (CAT6)	Fibre UTP (CAT7)	Fibre
Encoding	Manc	(4b/5b)	(8b/10b)	(64b/66b)	(64b/66b)	(64b/66b)
Format	2 level	3 levels	5 level	16 levels	16 levels	16 levels
Pairs	2	2	4	4	4	
Bandwidth (MHz)	20	31	125	413	>1000	
Mode	HDX	HDX/ FDX	FDX	FDX	FDX	FDX
Hubs	4	(1 or 2 in std	0	0	0	0



Summary of GBE

- **New Physical Layer Technology**
  - Bandwidth limit for CAT-5 UTP
  - Fast Ethernet : 4b/5b+MLT-3 (over CAT-5/CAT-5e/Fibre)
  - GBE: 8b/10b+PAM-5 (over CAT-5 /CAT-5e/Fibre)
  - 10GBE: 64/66b (over CAT-6/Fibre/CX-4)
  - 1 GHz CAT-7 Screened cable

- **Switches**
  - Packet rate becomes major challenge for timing
  - Also places tight demands on design of address table

- **Next steps...**
  - 40 Gbps... was standardised 2010 (over Fibre)
  - First 100 Gbps Philadelphia, 2008.
  - 100 Gbps... variant of above now available (over Fibre)
  - Many variants being designed/built
  - 1 Tb/s in research labs

Ethernet continues to evolve

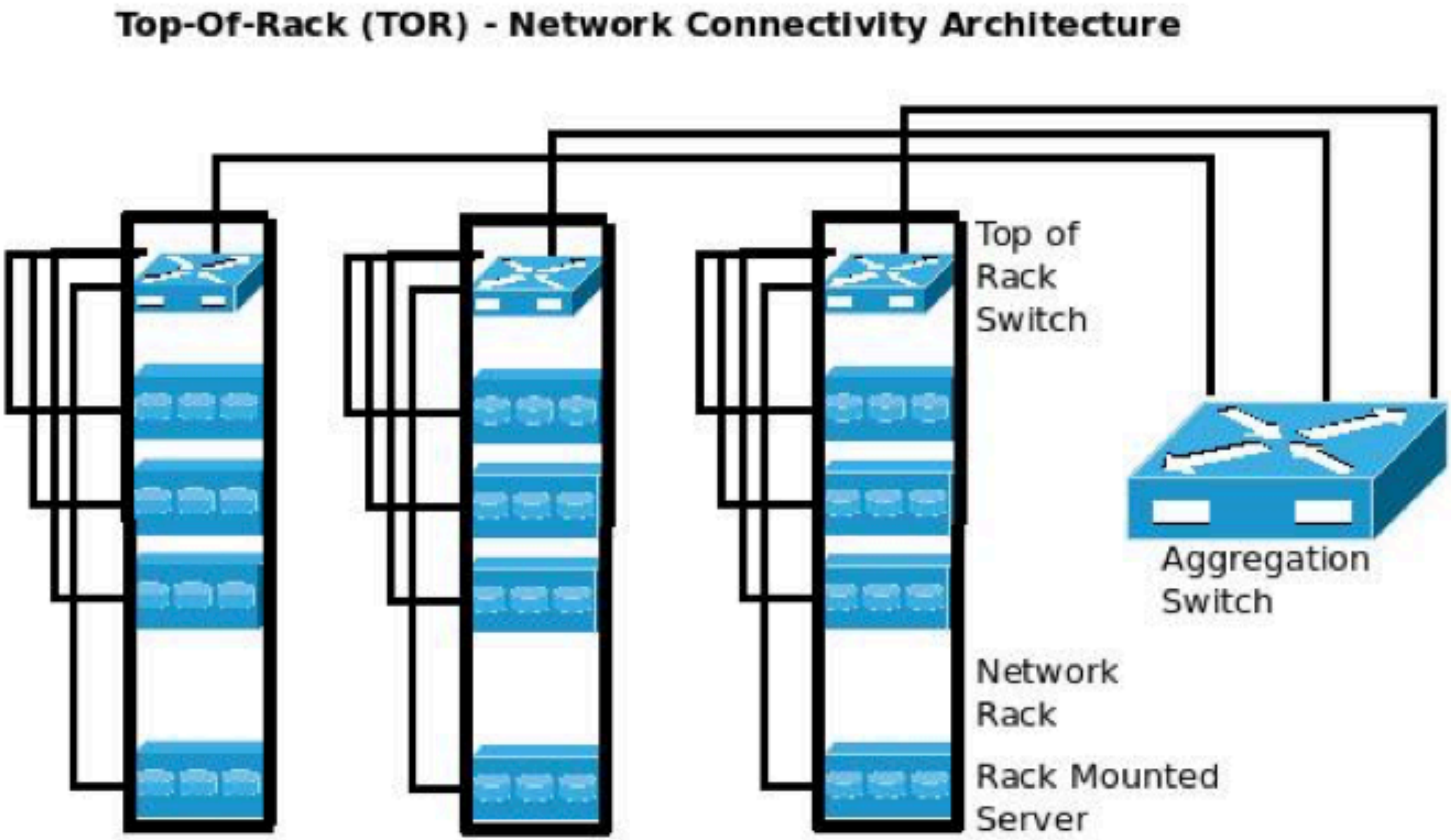
Data Centres

Module 6.7

One switch at the top of each rack

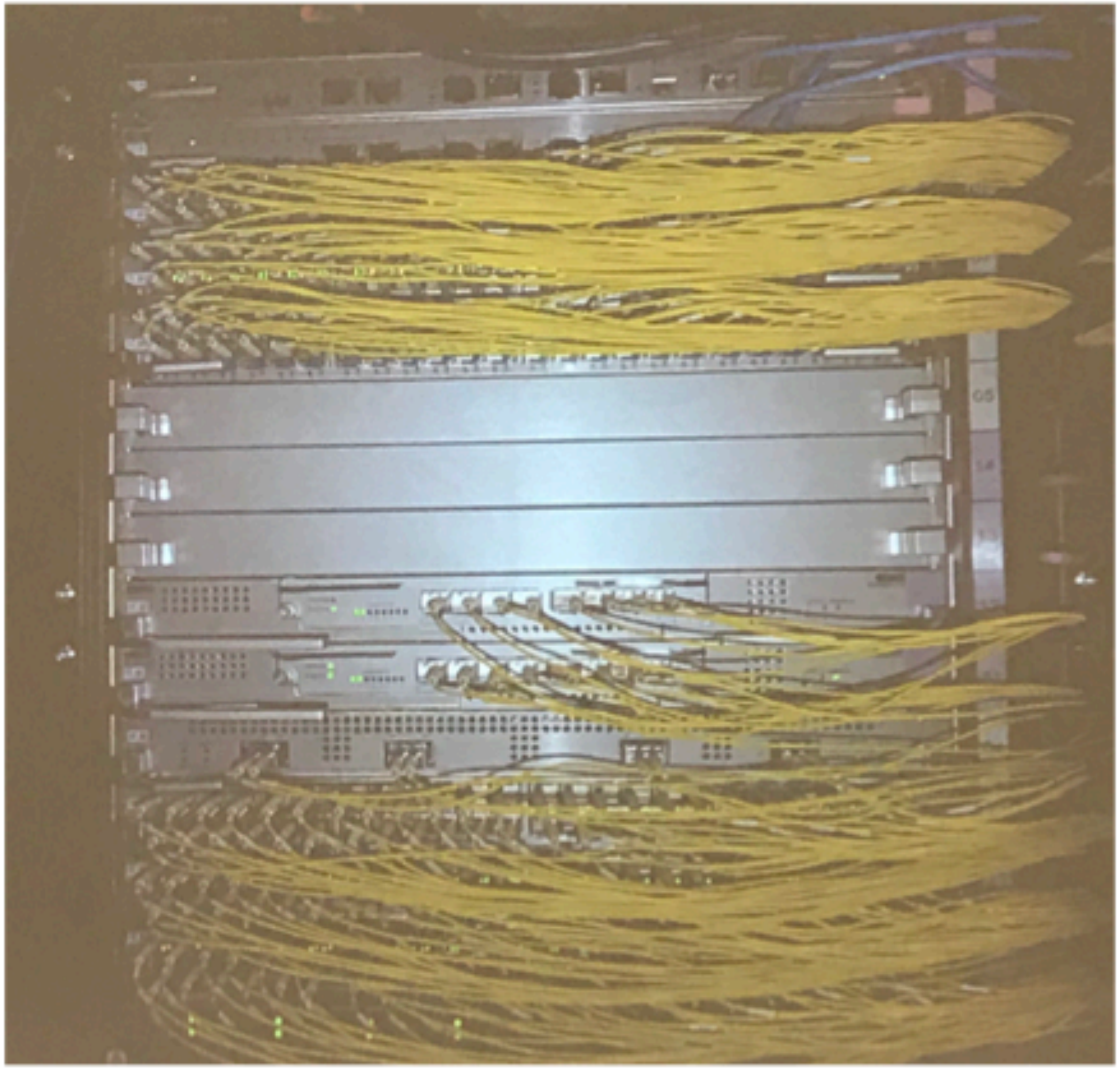
Connecting switches together poses a problem

standard switches have one or two "uplink ports"



Fine Patching in the Data Centre

Fibre Patch Panel

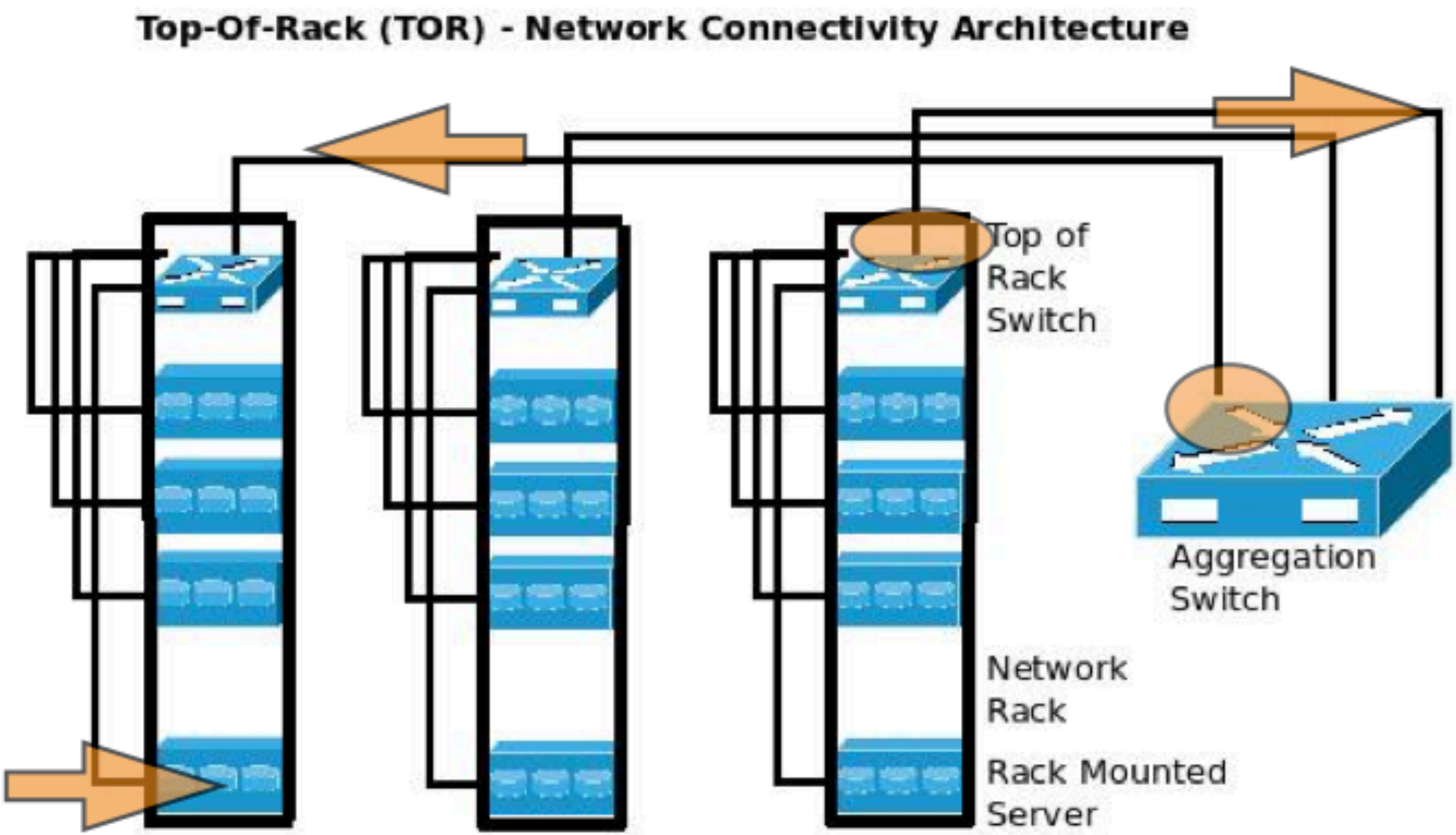


Priority-Based Flow Control Frames

A lot of packets can arrive in a very short time

Send a pause frame upstream when input buffer fills to a threshold

If next upstream switch congests, also send PAUSE on this upstream



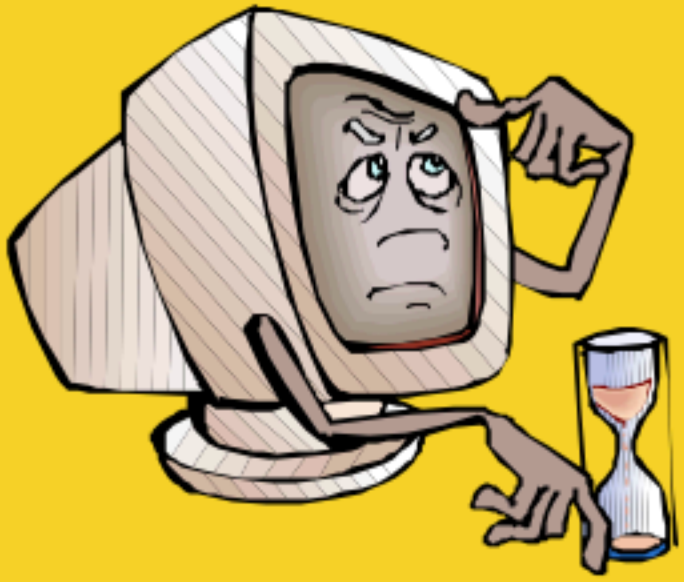
IEEE 802.1Qbb

Google Juniper Data Centre Switch

All switches meshed together
- effectively creates one 10,000 port switch

google jupiter DC switch

Performance



Two key performance measures:

Throughput

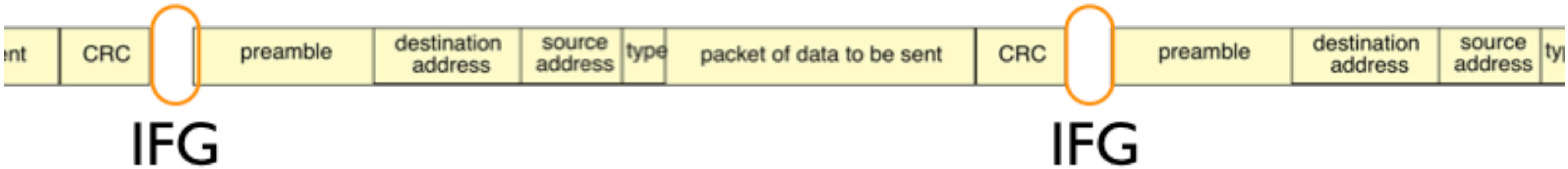
Utilisation

Ethernet Frames

Transmission

An 1000 byte frame takes 8000/(10 000 0000) at 10 Mbps
= 800 μS

An 1000 byte frame takes 8000/(1000 000 0000) at 1 Gbps
= 8 μS



Actually takes slightly longer because there must be an Interframe Gap between frames of 96 bit periods.

A 1000 B frame takes 809.6 μS at 10 Mbps

Example 1

Calculate the maximum frame rate of a node on a 10 Mbps Ethernet LAN.

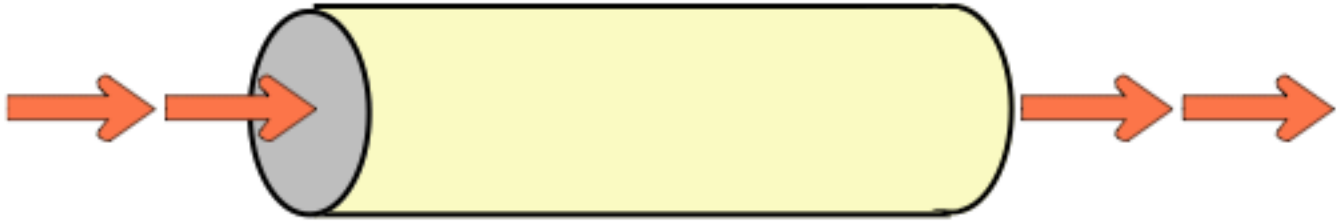
Frame Part	Minimum Size Frame
Inter Frame Gap (9.6μs)	
MAC Preamble (+ SFD)	
MAC Destination Address	
MAC Source Address	
MAC Type (or Length)	
Payload (Network PDU)	
Check Sequence (CRC)	
Total Frame Physical Size	

Example 1

Calculate the maximum frame rate of a node on an Ethernet LAN.

Frame Part	**Minimum Size Frame**
Inter Frame Gap (9.6μs)	
MAC Preamble (+ SFD)	
MAC Destination Address	
MAC Source Address	
MAC Type (or Length)	
Payload (Network PDU)	
Check Sequence (CRC)	
Total Frame Physical Size	

Throughput

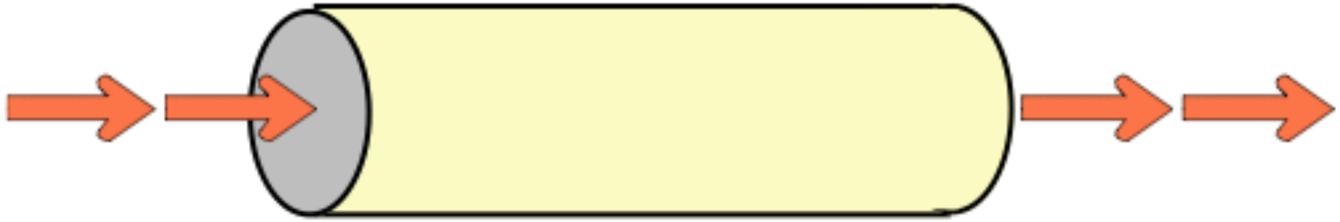


Defined as "the number of bits transferred per second from a given layer to the upper layer as a result of a conversation between two users of the layer"

Considers only data forwarded (i.e. not overhead)

Expressed in bits per second

Throughput



Defined as "*the number of bits transferred per second from a given layer to the layer above as a result of a conversation between two users of the layer*"

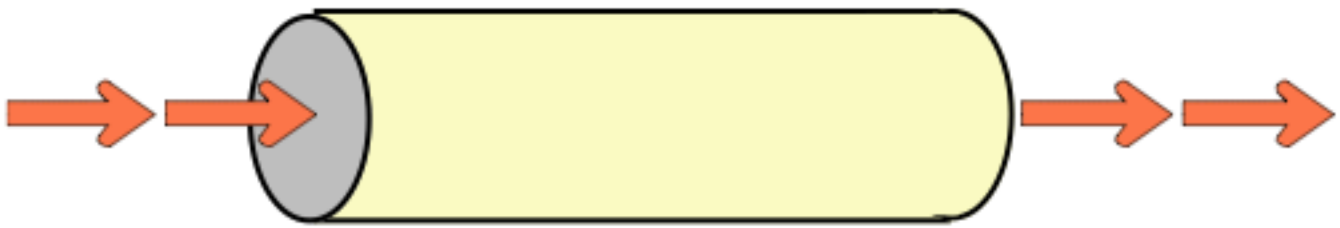
Considers only data forwarded (i.e. not overhead)

Expressed in bits per second

A source sends 1470 byte Ethernet frames at 10 frame/sec what is the throughput across the network?

Size of 1 frame = (1470-26)x8 bits
Throughput = 115.5 kbps

Calculation of Throughput



1) A source sends 1526 byte Ethernet frames at 50 frame/sec
What is the throughput across the network?

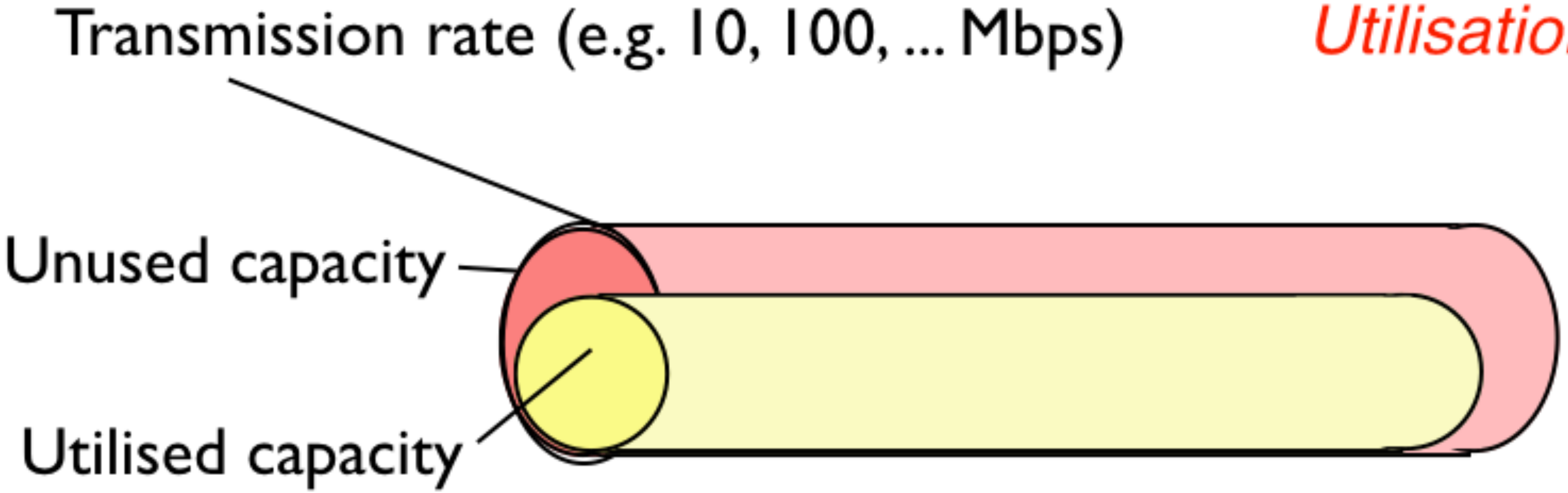
2) An application sends 25 PDUs per second with a size of 100 bytes
- what is the total network capacity consumed in bits per second?

3) Given that Ethernet also requires an Inter Frame Gap (IFG) of
9.6 μS before each frame, how long does it take at 10 Mbps to
transmit a frame that carries 46 bytes of PDU?

Example 2

Calculate maximum throughput of link service provided by 10 Mbps Ethernet

Frame Part	Maximum Size Frame
Inter Frame Gap (9.6μs)	
MAC Preamble (+ SFD)	
MAC Destination Address	
MAC Source Address	
MAC Type (or Length)	
Payload (Network PDU)	
Check Sequence (CRC)	
Total Frame Physical Size	



Utilisation

Defined as *"the total number of bits transferred at the physical layer to communicate a certain amount of data divided by the time taken to communicate the data."*

Includes all bits in all types of frame irrespective of whether they are corrupted or correctly received.

Expressed as a percentage of transmission rate.

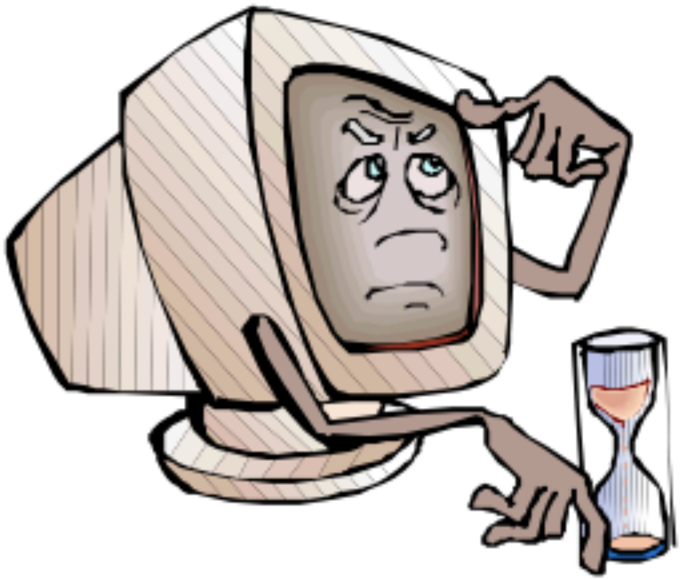
Measures link capacity used

Example 3

One node transmits 100 B frames at 10 frames per second, another transmits 1000 B frames at 2 frames per second, calculate the utilisation of a 10 Mbps Ethernet LAN.

Frame Part	Minimum Size Frame	
Inter Frame Gap (9.6μs)		
MAC Preamble (+ SFD)		
MAC Destination Address		
MAC Source Address		
MAC Type (or Length)		
Payload (Network PDU)		
Check Sequence (CRC)		
Total Frame Physical Size		

Over to you....



- Spend one session reviewing material on web.
- Answers to examples are at:
  - ./lan-pages/enet-calc.html
- Finally, do the revision questions....
  - ./questions/intro/index.html

National Cyber Security Centre

10 Steps to Cyber Security

Defining and communicating your Board's Information Risk Regime is central to your organisation's overall cyber security strategy. The National Cyber Security Centre recommends you review this regime – together with the nine associated security areas described below, in order to protect your business against the majority of cyber attacks.

Network Security

Protect your networks from attack. Defend the network perimeter, filter out unauthorised access and malicious content. Monitor and test security controls.

User education and awareness

Produce user security policies covering acceptable and secure use of your systems. Include in staff training. Maintain awareness of cyber risks.

Malware prevention

Produce relevant policies and establish anti-malware defences across your organisation.

Removable media controls

Produce a policy to control all access to removable media. Limit media types and use. Scan all media for malware before importing onto the corporate system.

Secure configuration

Apply security patches and ensure the secure configuration of all systems is maintained. Create a system inventory and define a baseline build for all devices.

Set up your Risk Management Regime

Assess the risks to your organisation's information and systems with the same vigour you would for legal, regulatory, financial or operational risks. To achieve this, embed a Risk Management Regime across your organisation, supported by the Board and senior managers.

Managing user privileges

Establish effective management processes and limit the number of privileged accounts. Limit user privileges and monitor user activity. Control access to activity and audit logs.

Incident management

Establish an incident response and disaster recovery capability. Test your incident management plans. Provide specialist training. Report criminal incidents to law enforcement.

Monitoring

Establish a monitoring strategy and produce supporting policies. Continuously monitor all systems and networks. Analyse logs for unusual activity that could indicate an attack.

Home and mobile working

Develop a mobile working policy and train staff to adhere to it. Apply the secure baseline and build to all devices. Protect data both in transit and at rest.

For more information go to www.ncsc.gov.uk @ncsc