Energy Informatics
01 Introduction to Computer Networking

Christian Schindelhauer
Technical Faculty
Computer-Networks and Telematics
University of Freiburg

Overview

- Organization of the Course „Energy Informatics“
- Challenges of Computer Networks
  - Size, complexity, technology
- Fundamental concepts in Computer Networks
  - Layers
  - Modulation
  - Encoding
  - Protocols
  - Distributed Systems
- Network Layers
  - Physical, Data link, Network, Transport, Application

The Team

- Lecturers:
  - Prof. Dr. Christian Schindelhauer
    Computer Networks and Telematics
  - Prof. Peter Thiemann
    Programming Languages
  - Prof Georg Lausen
    Databases and Information Systems
- Tutor
  - Anas Alzoghbi

Organization

- Web-page:
  - http://cone.informatik.uni-freiburg.de/lehre/aktuell/einfo-ws15
  - with slides, exercises

First Week

Mon, 25.01.	9:00 – 13:00	**Introduction to Computer Networking**	CS	R1
Tue, 26.01.	9:00 – 13:00	**Internet Protocols**	CS	R1
		Exercises session		R1
Wed, 27.01.	9:00 – 13:00	**Network Algorithms**	CS	R1
		Exercises session		R1
Thur, 28.01.	9:00 – 13:00	**Network Security**	CS	R1
		Exercises session		R1
Fri, 29.01.	9:00 – 13:00	**IT-Structures for Smart Grids**	CS	R1
		Exercises session		R1

Second Week

Mon, 01.02.	9:00 – 13:00	**Introduction to Relational Databases**	GL	R1
		Exercises session	CS	R1
Tue, 02.02.	9:00 – 13:00	**Database Querying: SQL**	GL	R1
		Exercises session	AA	R1
Wed, 03.02.	9:00 – 13:00	**Database Behavior: SQL**	GL	R1
		Exercises session	AA	R1
Thur, 04.02.	9:00 – 13:00	**Database Modeling**	GL	R1
		Exercises session	AA	R1
Fri, 05.22.	9:00 – 13:00	**Interfaces: XML**	GL	R1
		Exercises session	AA	R1

Third Week

Mon, 08.02.	13:00-15:00	**Exercise Session for XML**	AA	R1
	15:00-17:00	**Requirements & Specification**	PT	R1
Tue, 09.02.	13:00 – 15:00	**Data Collection**	PT	R1
	15:00 – 17:00	**Exercise session**	AA	R1
Wed, 10.02.	13:00 – 15:00	**Analysis**	PT	R1
	15:00 – 17:00	**Exercise session**	AA	R1
Thur, 11.02.	13:00 – 15:00	**Modeling**	PT	R1
	15:00 – 17:00	**Exercise session**	AA	R1
Fri, 12.02.	13:00 – 17:00	**Project**	PT	R1

CoNe Freiburg

Types of Networks

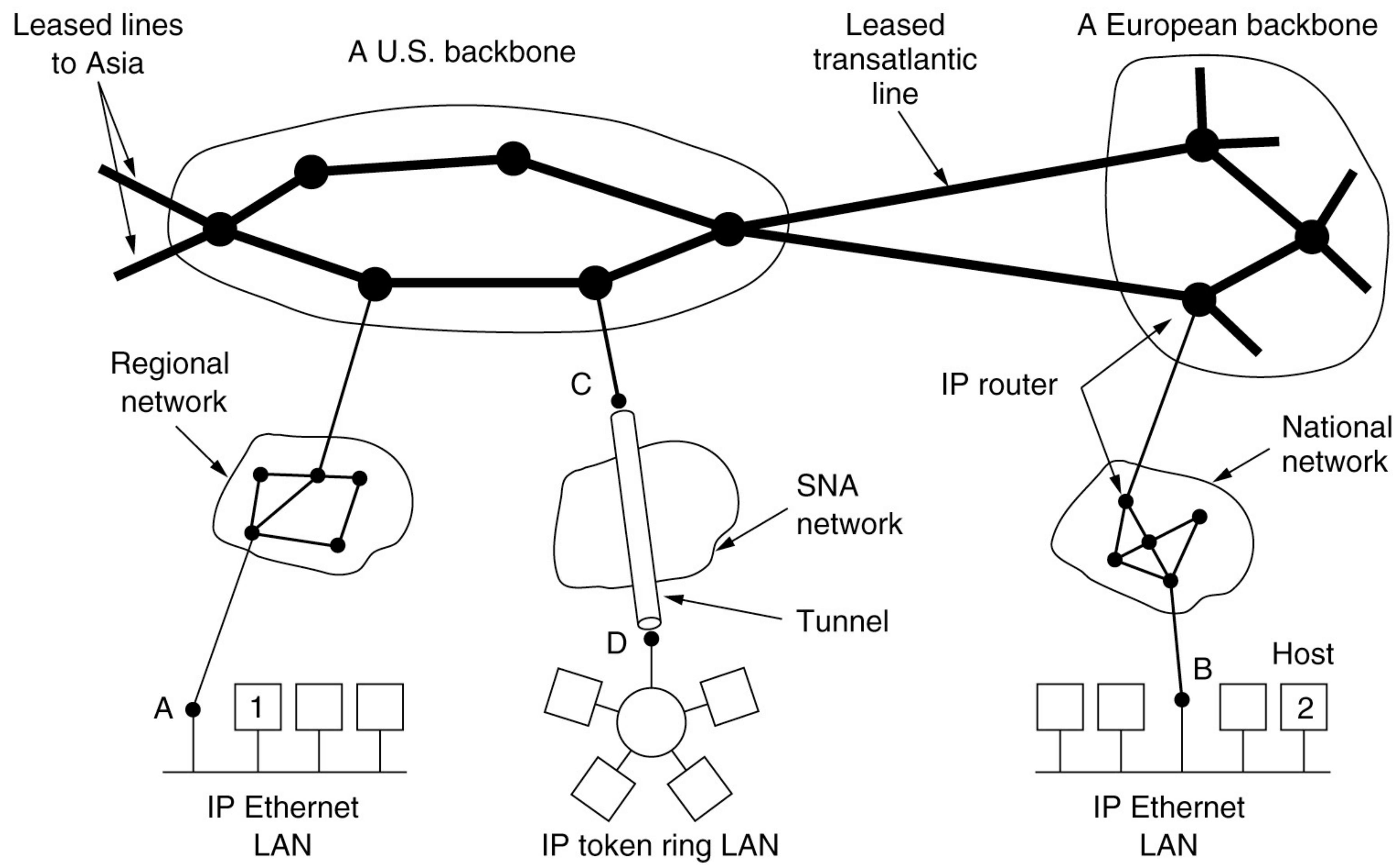
Interprocessor distance	Processors located in same	Example
1 m	Square meter	Personal area network
10 m	Room	Local area network
100 m	Building	Local area network
1 km	Campus	Local area network
10 km	City	Metropolitan area network
100 km	Country	Wide area network
1000 km	Continent	Wide area network
10,000 km	Planet	The Internet

(Tanenbaum)

UNI FREIBURG

The Internet

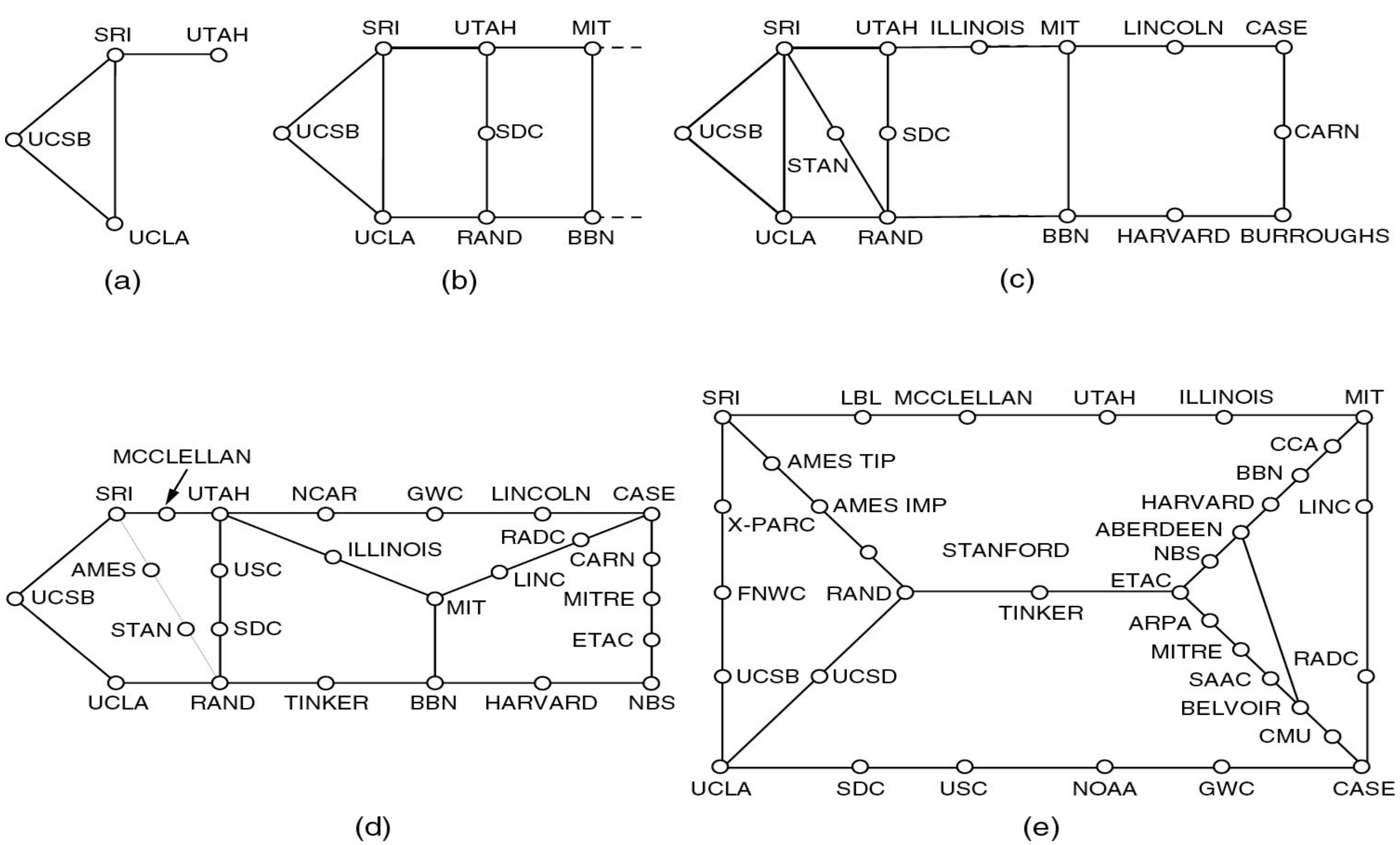
- global system of interconnected WANs and LANs
- open, system-independent, no global control



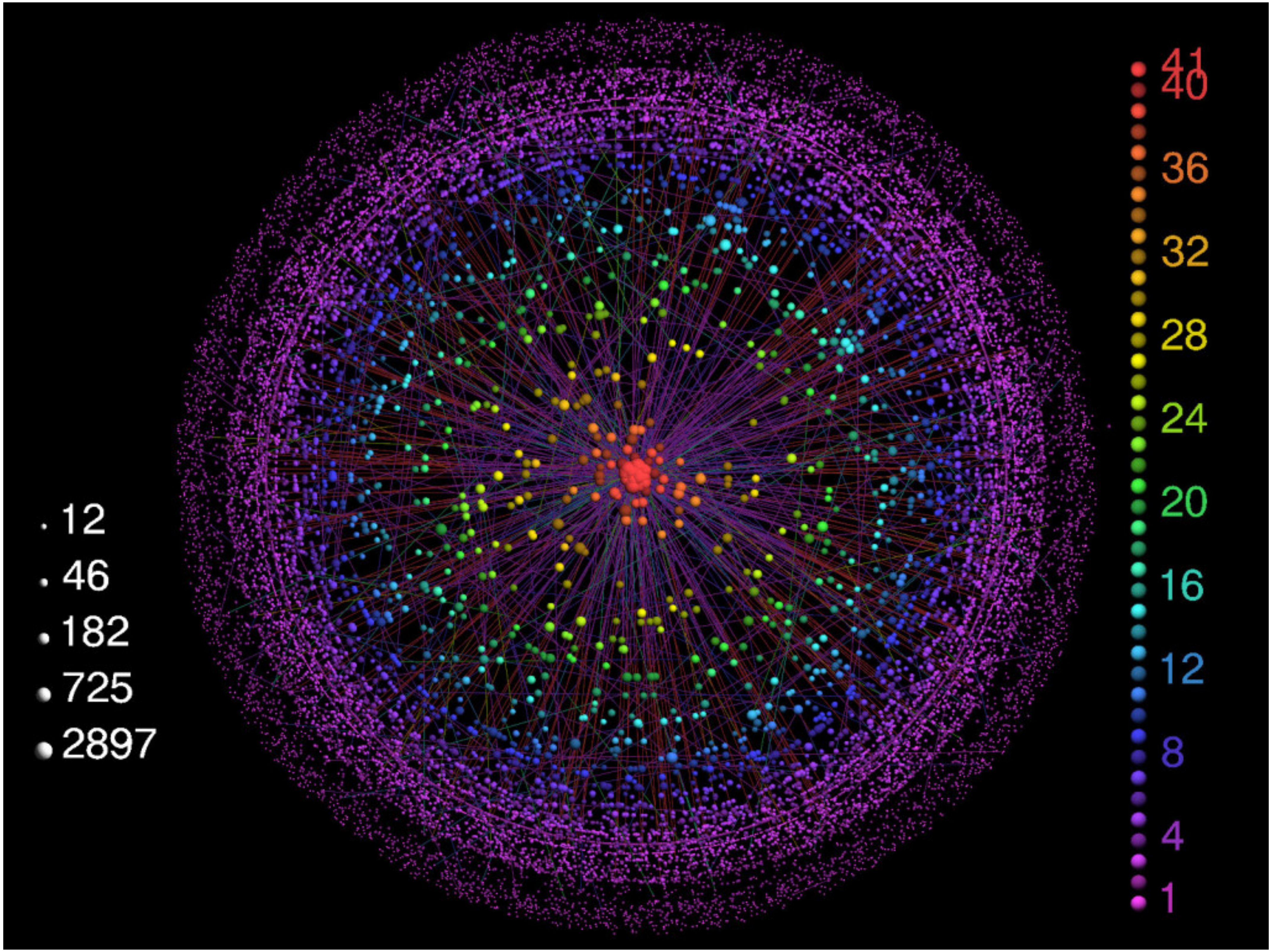
[Tanenbaum, Computer Networks]

ARPANET

ARPANET (a) December 1969 (b) July 1970 (c) March 1971 (d) April 1972 (e) September 1972



Internet ~2005



An Open Network Architecture

- Concept of Robert Kahn (DARPA 1972)
  - Local networks are autonomous
    - independent
    - no WAN configuration
  - packet-based communication
  - “best effort” communication
    - if a packet cannot reach the destination, it will be deleted
    - the application will re-transmit
  - black-box approach to connections
    - black boxes: gateways and routers
    - packet information is not stored
    - no flow control
  - no global control
- Basic principles of the Internet

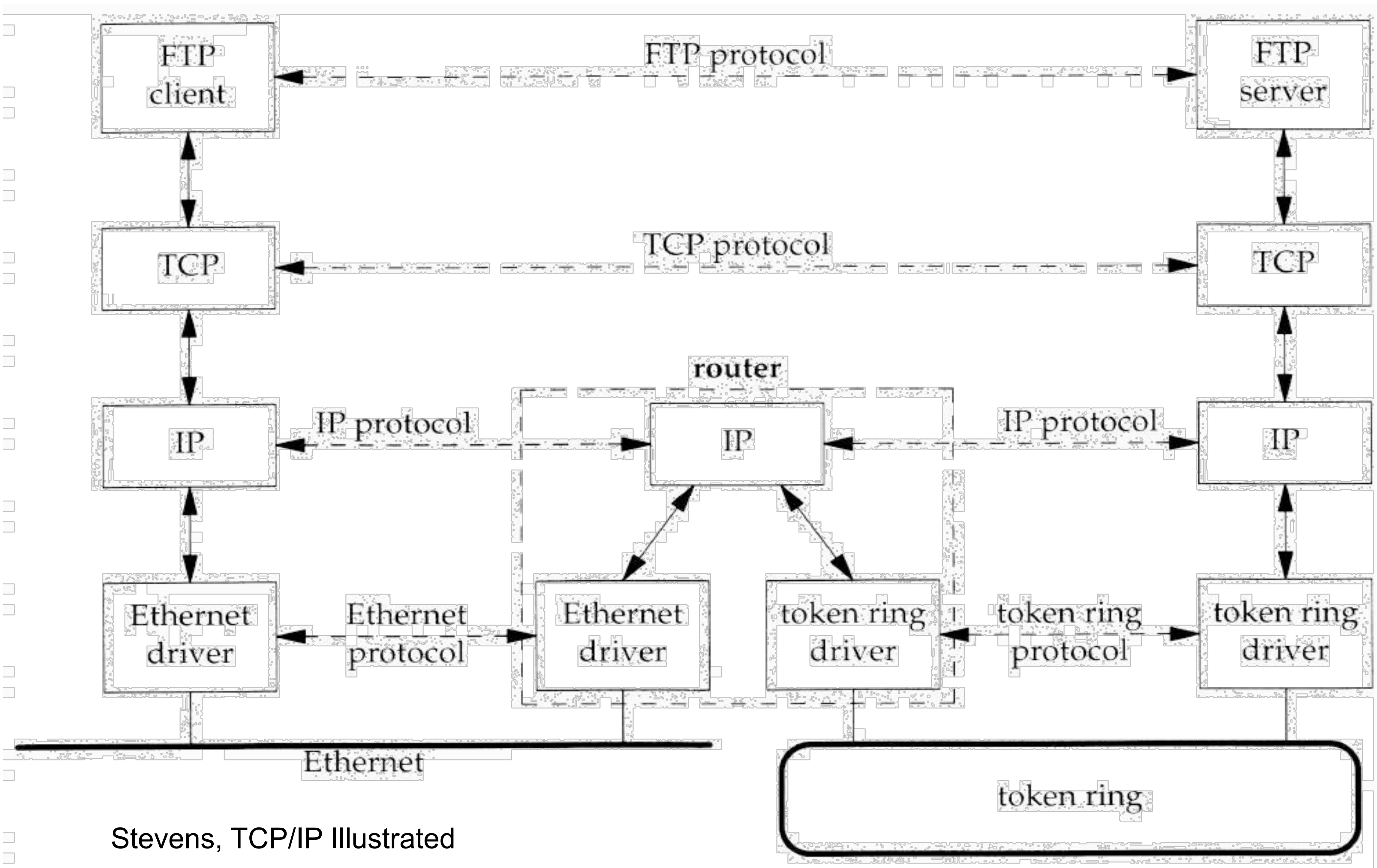
Protocols of the Internet

Application	Telnet, FTP, HTTP, SMTP (E-Mail), ...
Transport	TCP (Transmission Control Protocol) UDP (User Datagram Protocol)
Network	IP (Internet Protocol) IPv4 + IPv6 + ICMP (Internet Control Message Protocol) + IGMP (Internet Group Management Protoccol)
Host-to-Network	LAN (e.g. Ethernet, W-Lan etc.)

TCP/IP Layers

- 1. Host-to-Network
  - Not specified, depends on the local networ,k e.g. Ethernet, WLAN 802.11, PPP, DSL
- 2. Routing Layer/Network Layer (IP - Internet Protocol)
  - Defined packet format and protocol
  - Routing
  - Forwarding
- 3. Transport Layer
  - TCP (Transmission Control Protocol)
    - Reliable, connection-oriented transmission
    - Fragmentation, Flow Control, Multiplexing
  - UDP (User Datagram Protocol)
    - hands packets over to IP
    - unreliable, no flow control
- 4. Application Layer
  - Services such as TELNET, FTP, SMTP, HTTP, NNTP (for DNS), …
  - Peer-to-peer networks

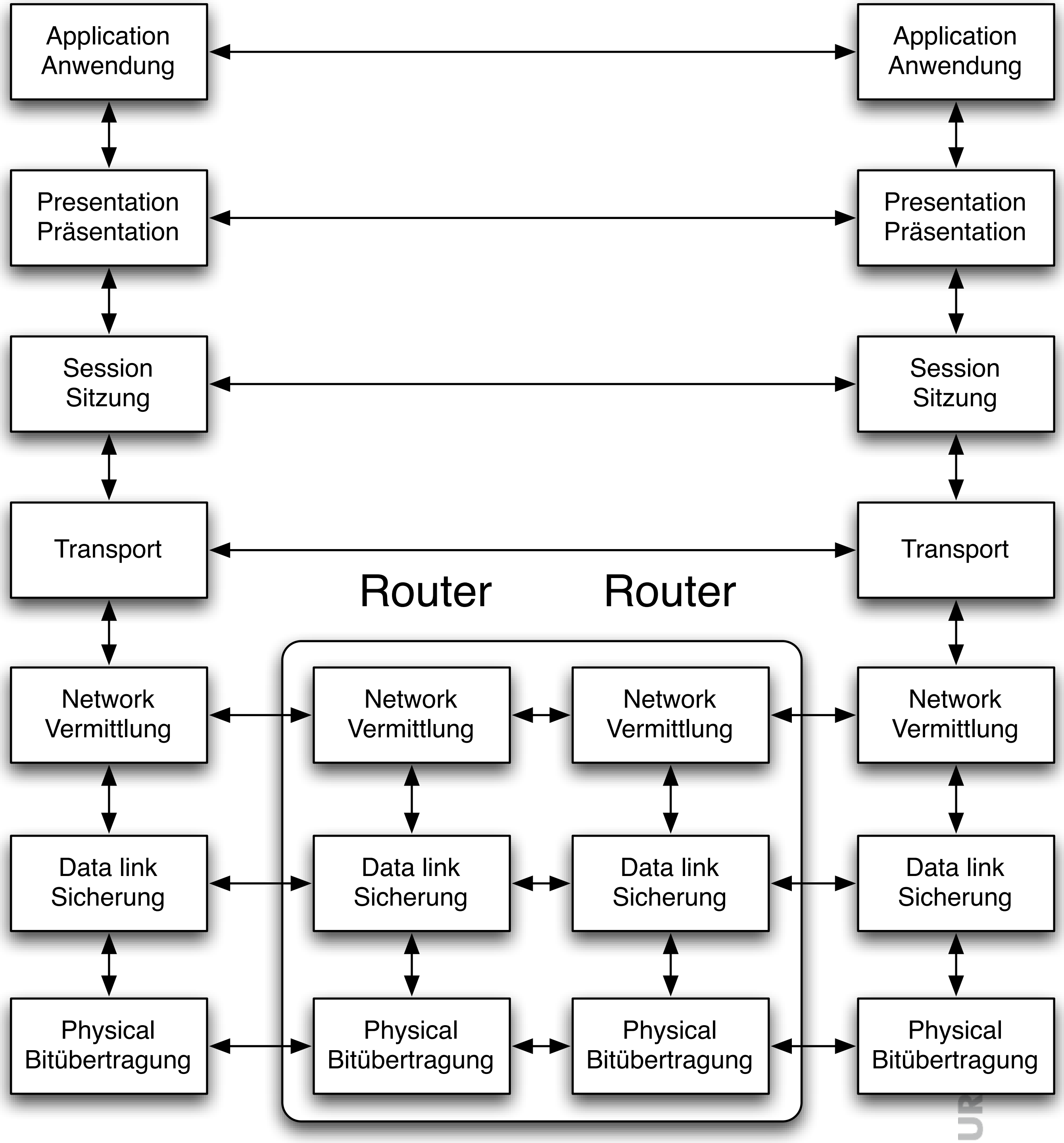
Example: Routing between LANs



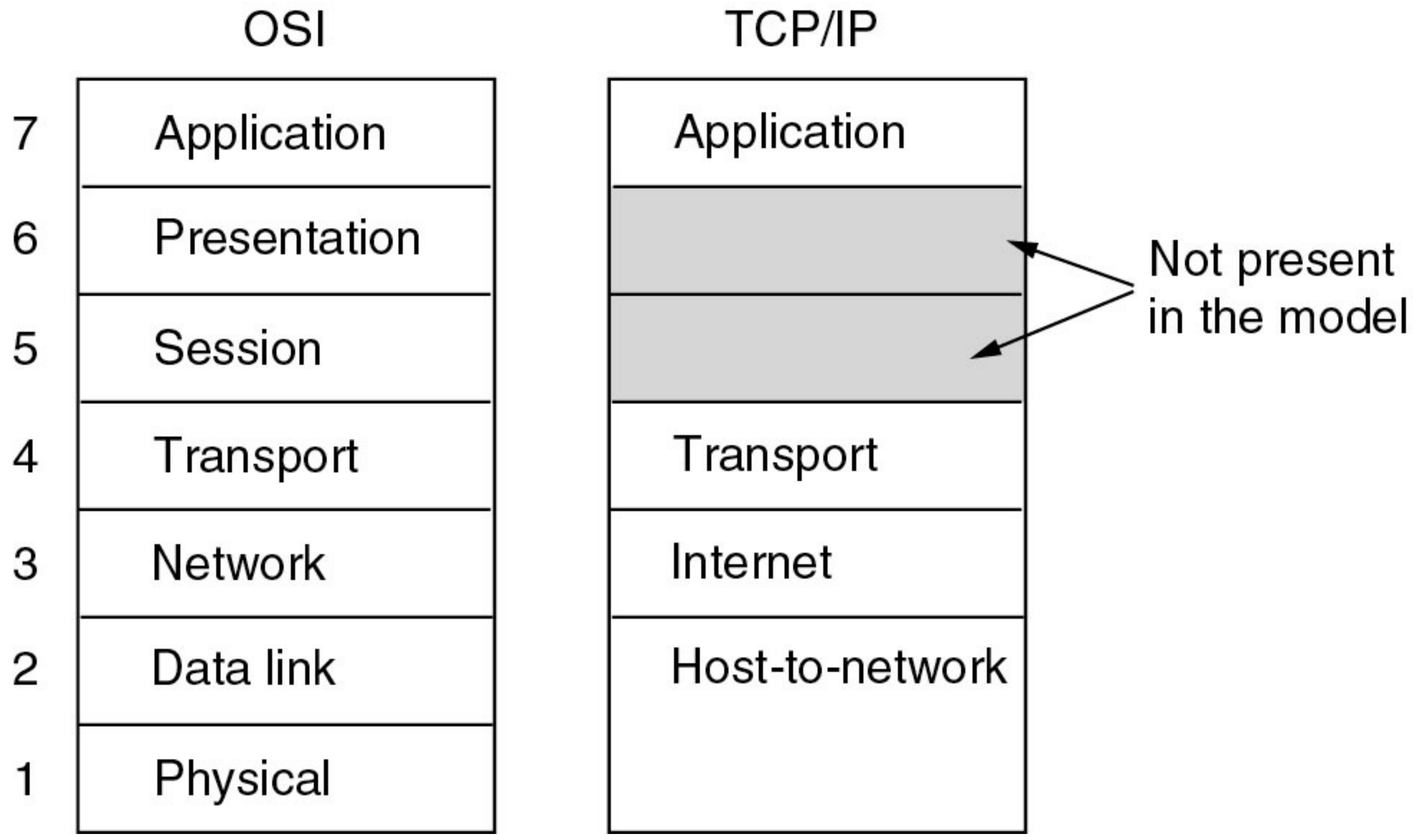
Stevens, TCP/IP Illustrated

ISO/OSI Reference model

- 7. Application
  - Data transmission, e-mail, terminal, remote login
- 6. Presentation
  - System-dependent presentation of the data (EBCDIC / ASCII)
- 5. Session
  - start, end, restart
- 4. Transport
  - Segmentation, congestion
- 3. Network
  - Routing
- 2. Data Link
  - Checksums, flow control
- 1. Physical
  - Mechanics, electrics

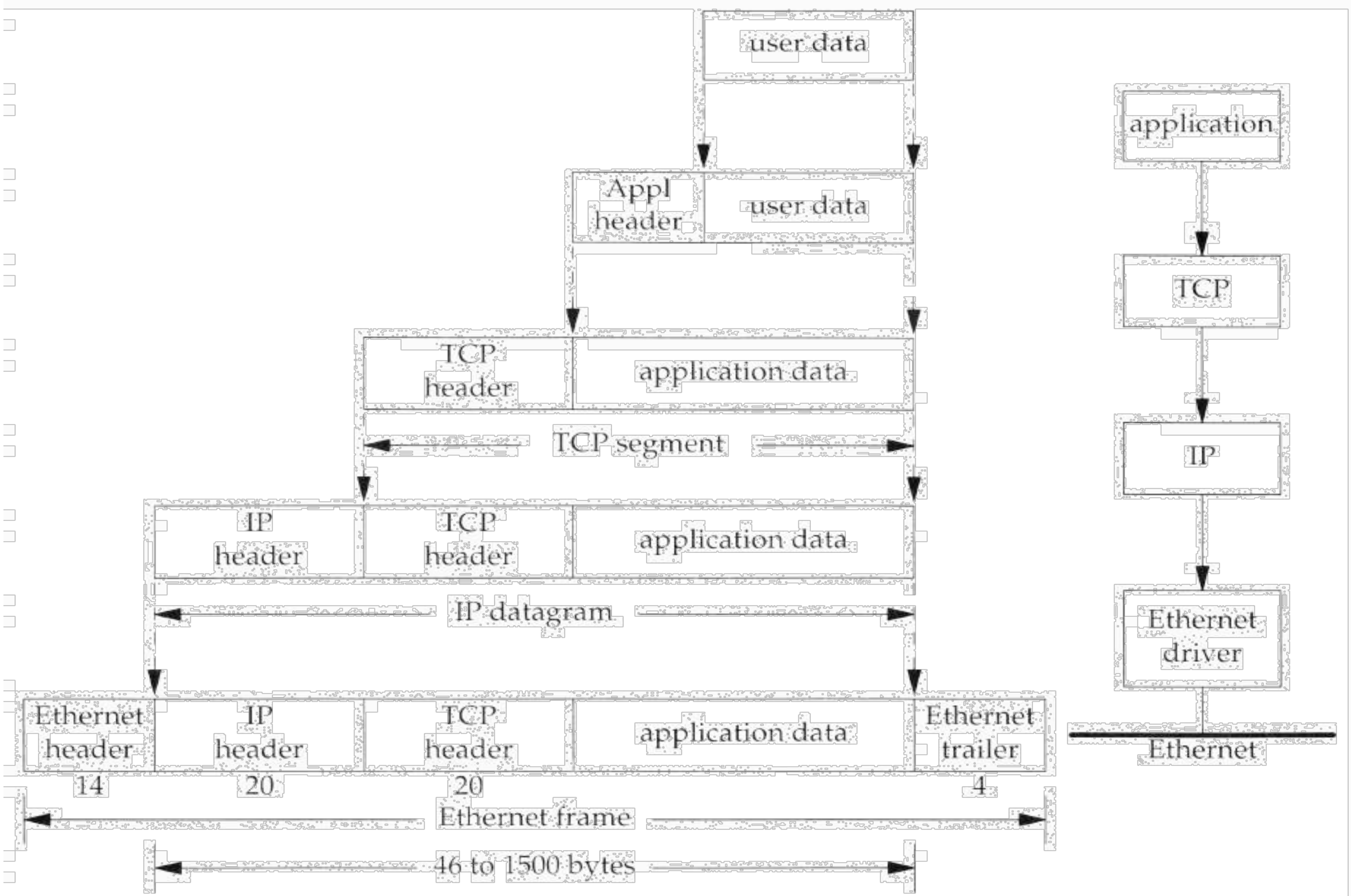


Reference Models: OSI versus TCP/IP



(Aus Tanenbaum)

Data/Packet Encapsulation



Stevens, TCP/IP Illustrated

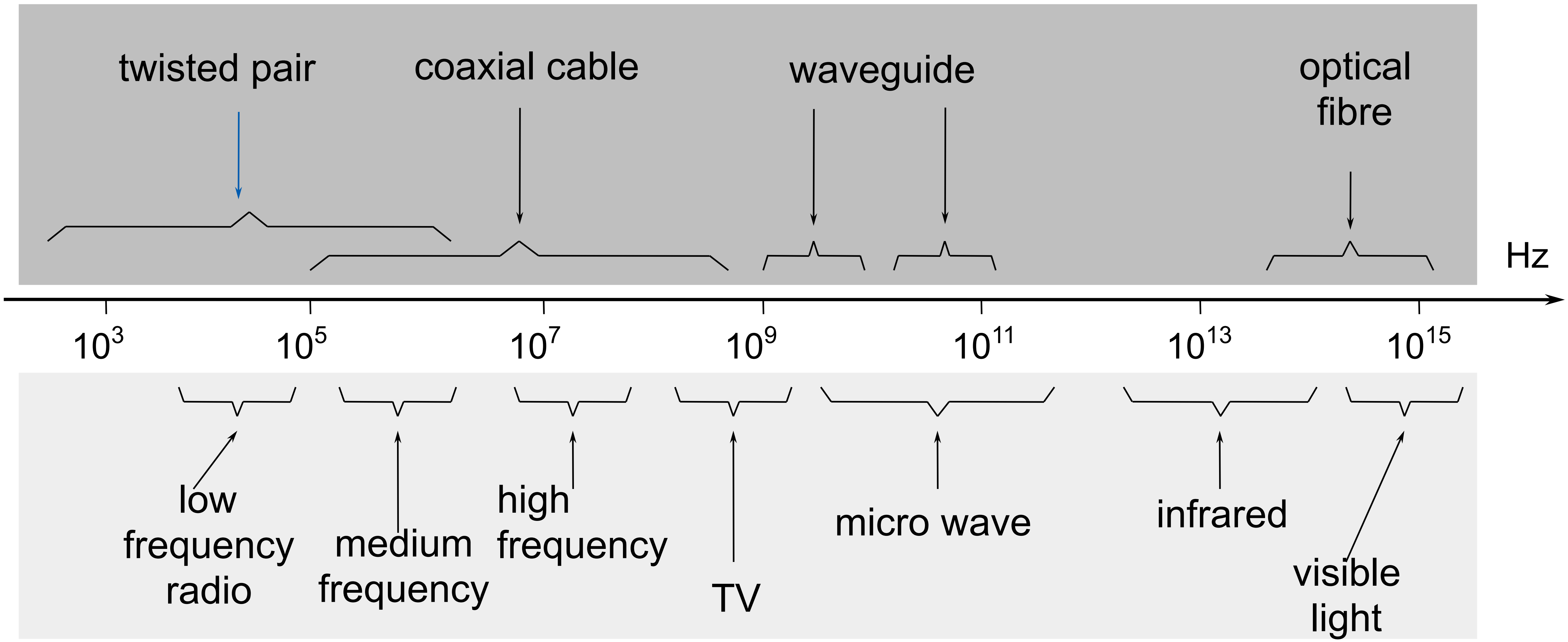
Physics – Background

- Moving particles with electric charge cause electromagnetic waves
  - frequency f : number of oscillations per second
    - unit: Hertz
  - wavelength λ: distance (in meters) between two wave maxima
  - antennas can create and receive electromagnetic waves
  - the transmission speed of electromagnetic waves in vacuum is constant
  - speed of light $c \approx 3 \cdot 10^8$ m/s
- Relation between wavelength, frequency and speed of light:

$$\lambda \cdot f = c$$

Electromagnetic Spectrum

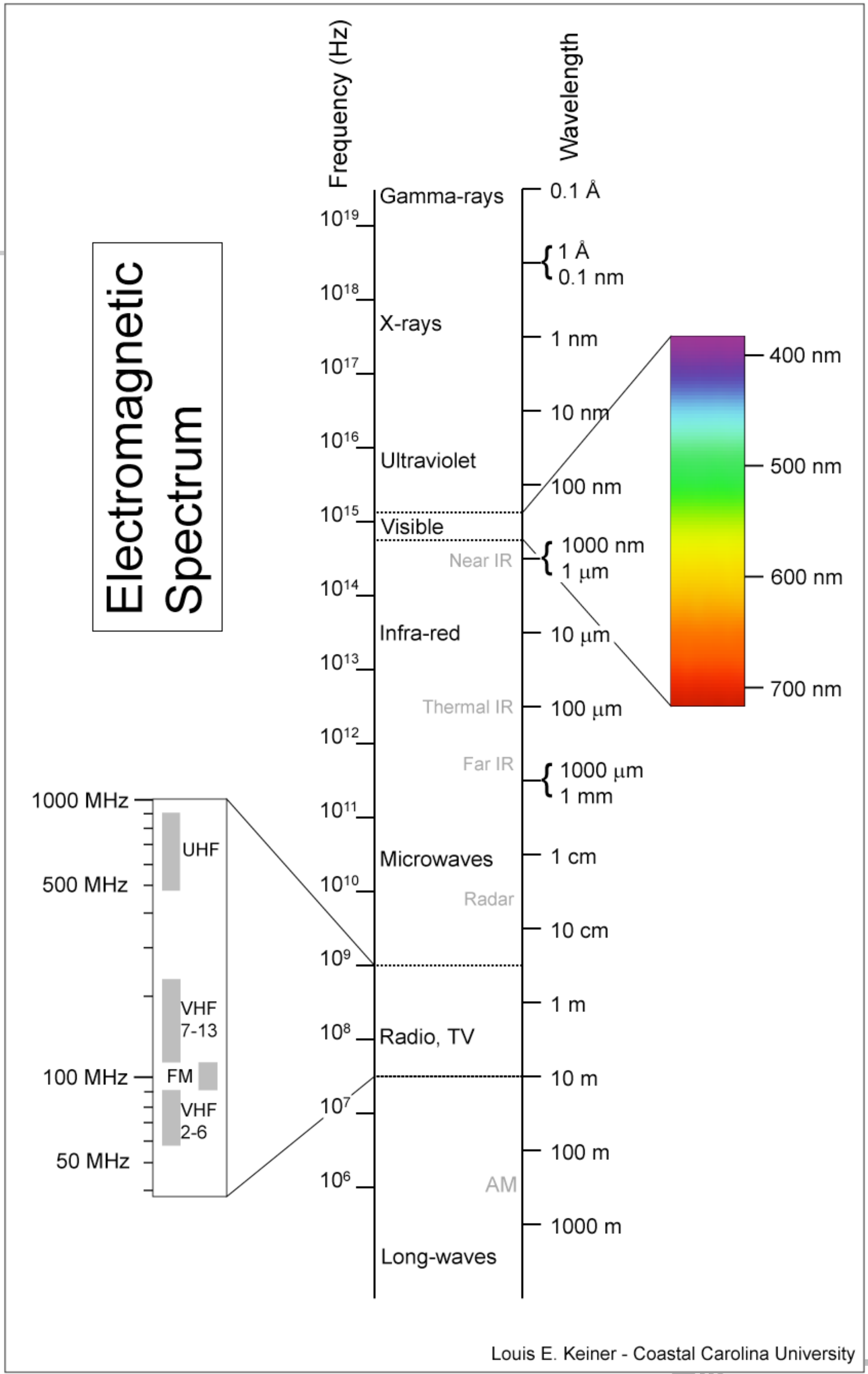
guided media



unguided media

Bands

- LF Low Frequency
- MF Medium Frequency
- HF High Frequency
- VHF Very High Frequency
- UHF Ultra High Frequency
- UV Ultra Violet light



Noise and Interference

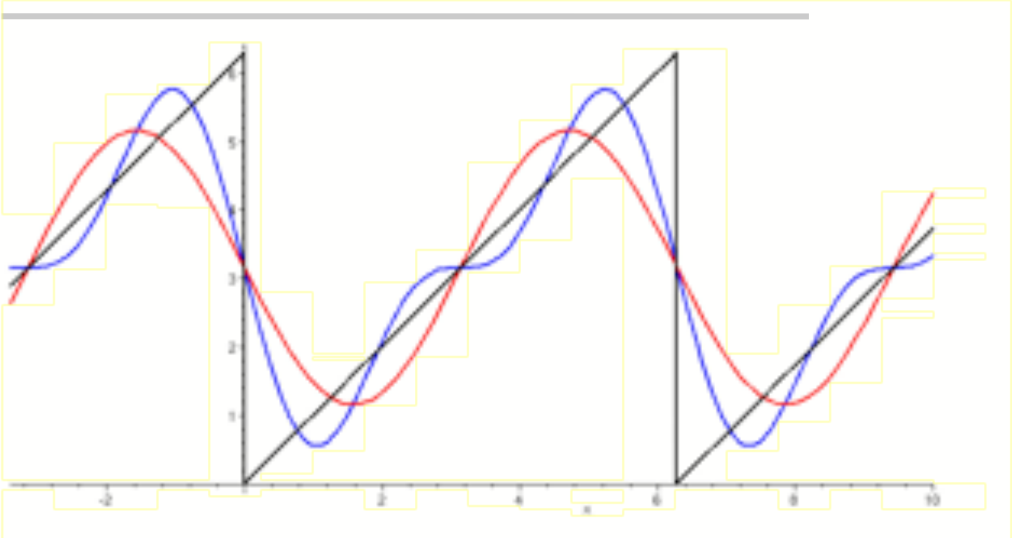
- Noise
  - inaccuracies and heat development in electrical components
  - modeled by normal distribution
- Interference from other transmitters
  - in the same spectrum
  - or in neighbored spectrum
    - e.g. because of bad filters
- Effect
  - Signal is disrupted

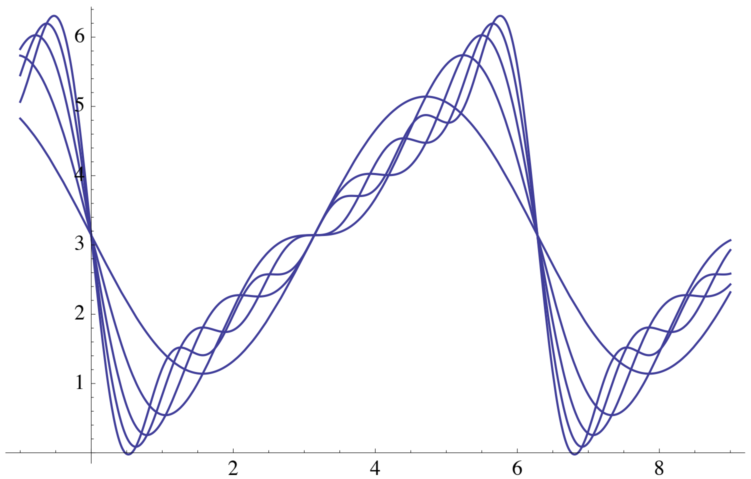
Signal Interference Noise Ratio

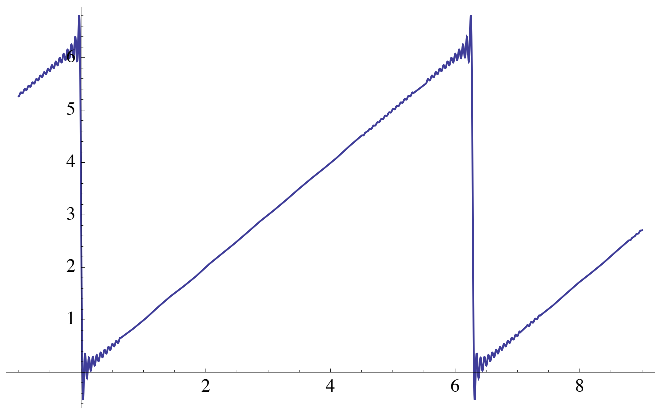
- reception energy = transmission energy · path loss
  - path loss ~ $1/d^{\gamma}$
    - $\gamma \in [2,5]$
- Signal to Interference and Noise Ratio = SINR
  - S = (desired) Signal energy
  - I = energy of Interfering signals
  - N = Noise
- Necessary condition for reception

$$\mathrm{SINR} = \frac{S}{I+N} \geq \mathrm{T}\mathit{hreshold}$$

Computation of Fourier Coefficients







$$f(x) = x \text{ , für } 0 < x < 2\pi$$

$$f(x) = \pi - 2\left(\frac{\sin x}{1} + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \ldots\right)$$

Fourier Analysis for General Period

- Theorem of Fourier for period T=1/f:
  - The coefficients c, a_n, b_n are then obtained as follows

$$g(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(2\pi k f t) + b_k \sin(2\pi k f t)$$

$$a_k = \frac{2}{T} \int_0^T g(t) \cos(2\pi n f t) dt$$

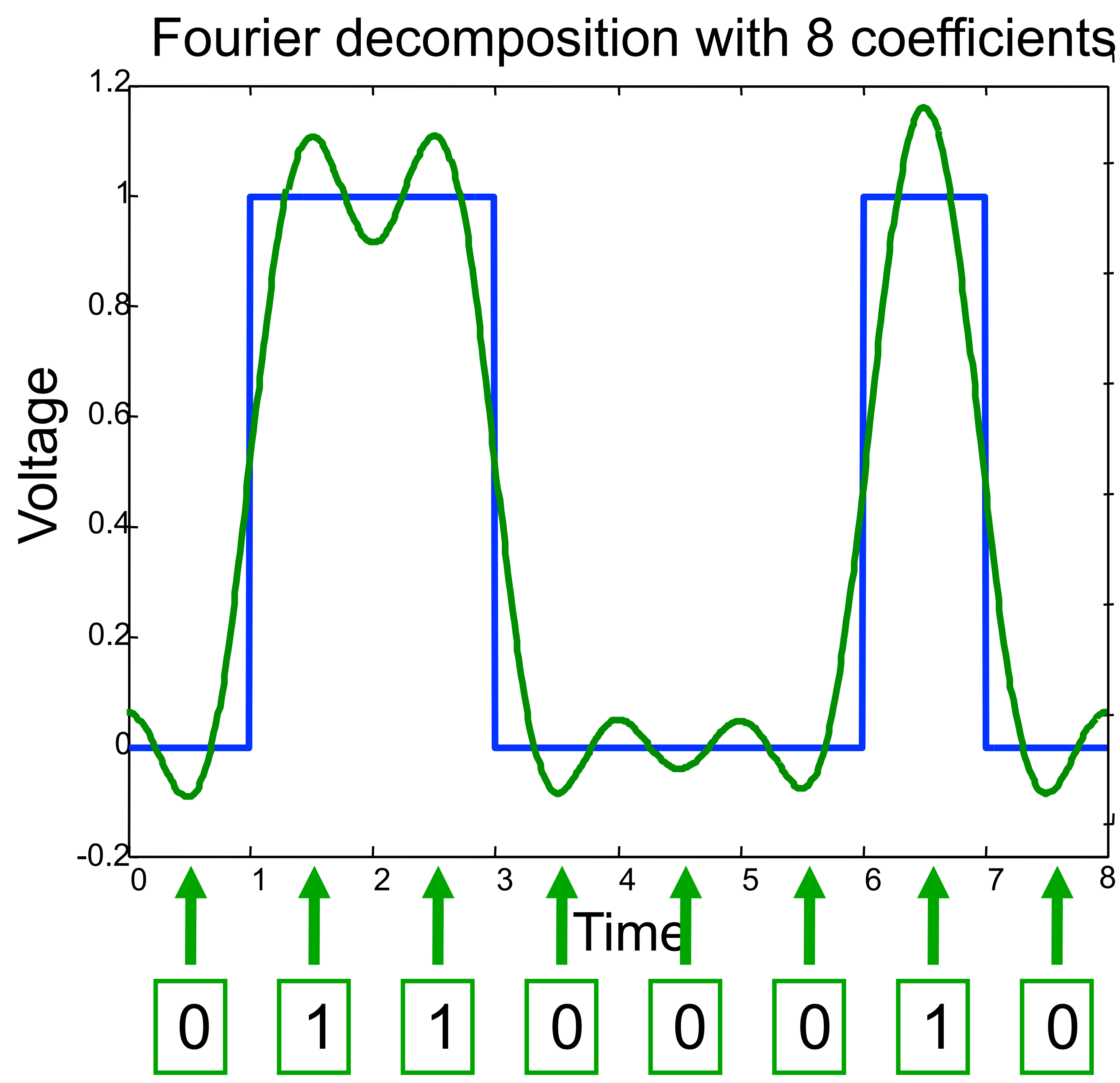
$$b_k = \frac{2}{T} \int_0^T g(t) \sin(2\pi n f t) dt$$

- The sum of squares of the k-th terms is proportional to the energy consumed in this frequency:

$$(a_k)^2 + (b_k)^2$$

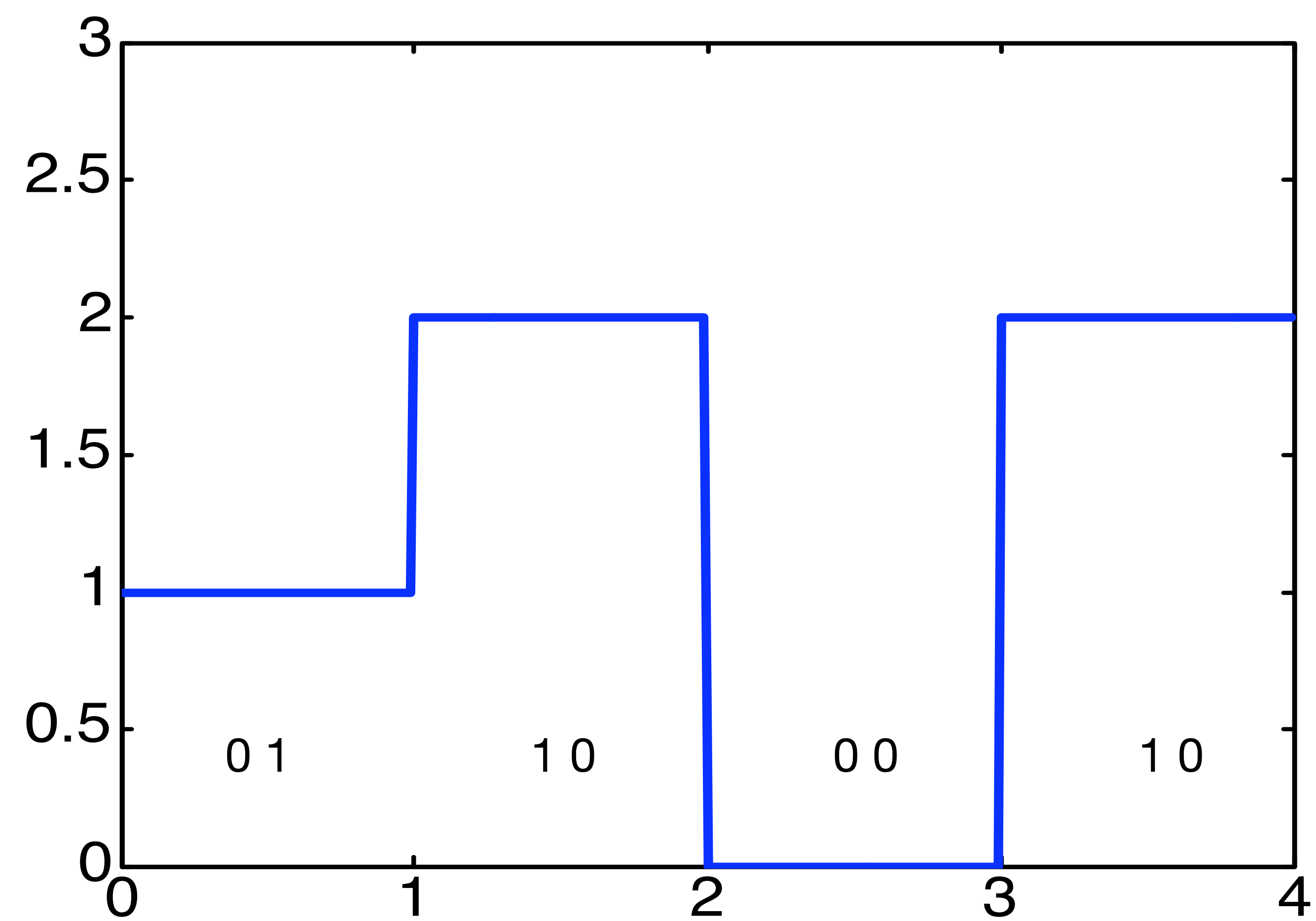
How often do you measure?

- How many measurements are necessary
  - to determine a Fourier transform to the k-th component, exactly?
- Nyquist-Shannon sampling theorem
  - To reconstruct a continuous band-limited signal with a maximum frequency f_{max} you need at least a sampling frequency of $2 f_{max}$.



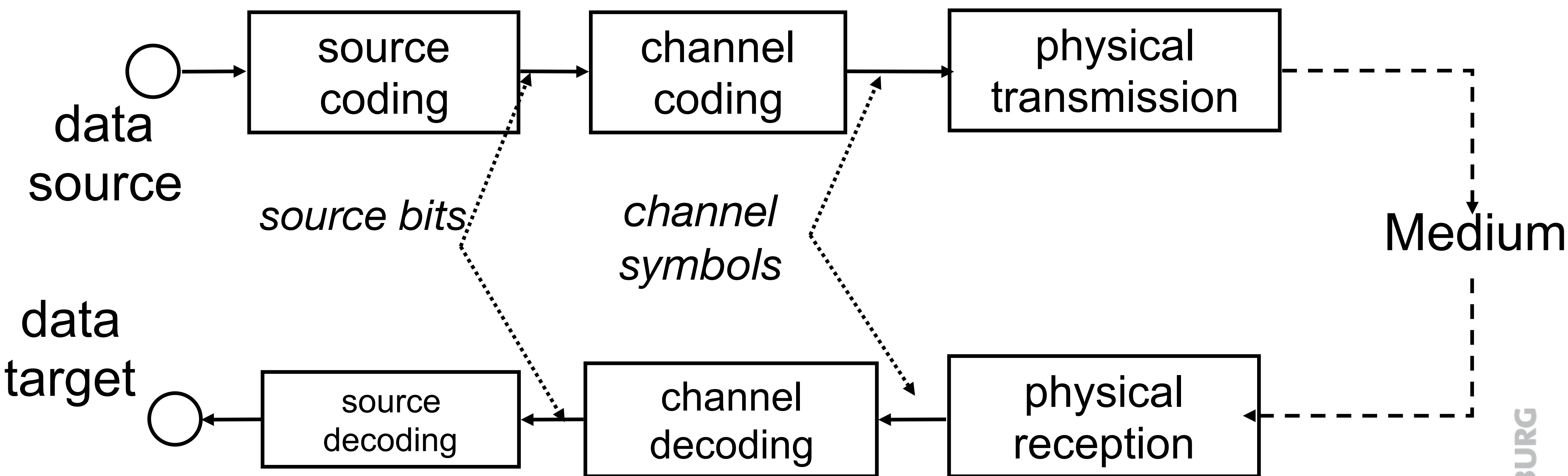
Symbols and Bits

- For data transmission instead of bits can also be used symbols
  - E.g. 4 Symbols: A, B, C, D
    - A = 00, B = 01, C = 10, D =
- Symbols
  - Measured in baud
  - Number of symbols per se
- Data rate
  - Measured in bits per secor
  - Number of bits per second
- Example
  - 2400 bit/s modem is 600 baud (uses 16 symbols)



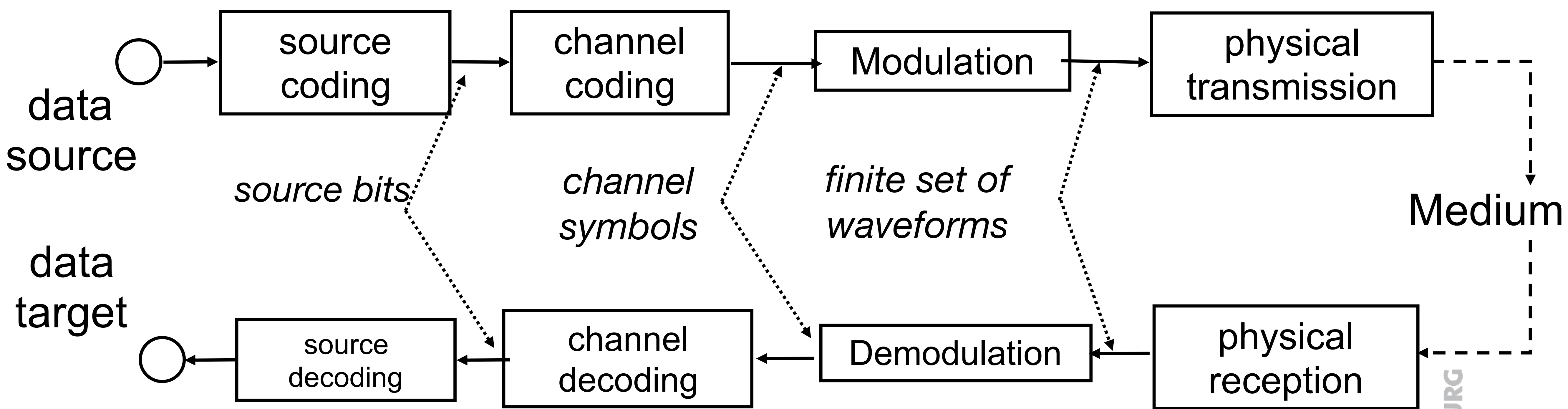
Structure of a *Baseband* Digital Transmission

- Source Coding
  - removing redundant or irrelevant information
  - e.g. with lossy compression (MP3, MPEG 4)
  - or with lossless compression (Huffman code)
- Channel Coding
  - Mapping of source bits to channel symbols
  - Possibly adding redundancy adapted to the channel characteristics
  - physical transmission
- Conversion into physical events



Structure of a *Broadband* Digital transmission

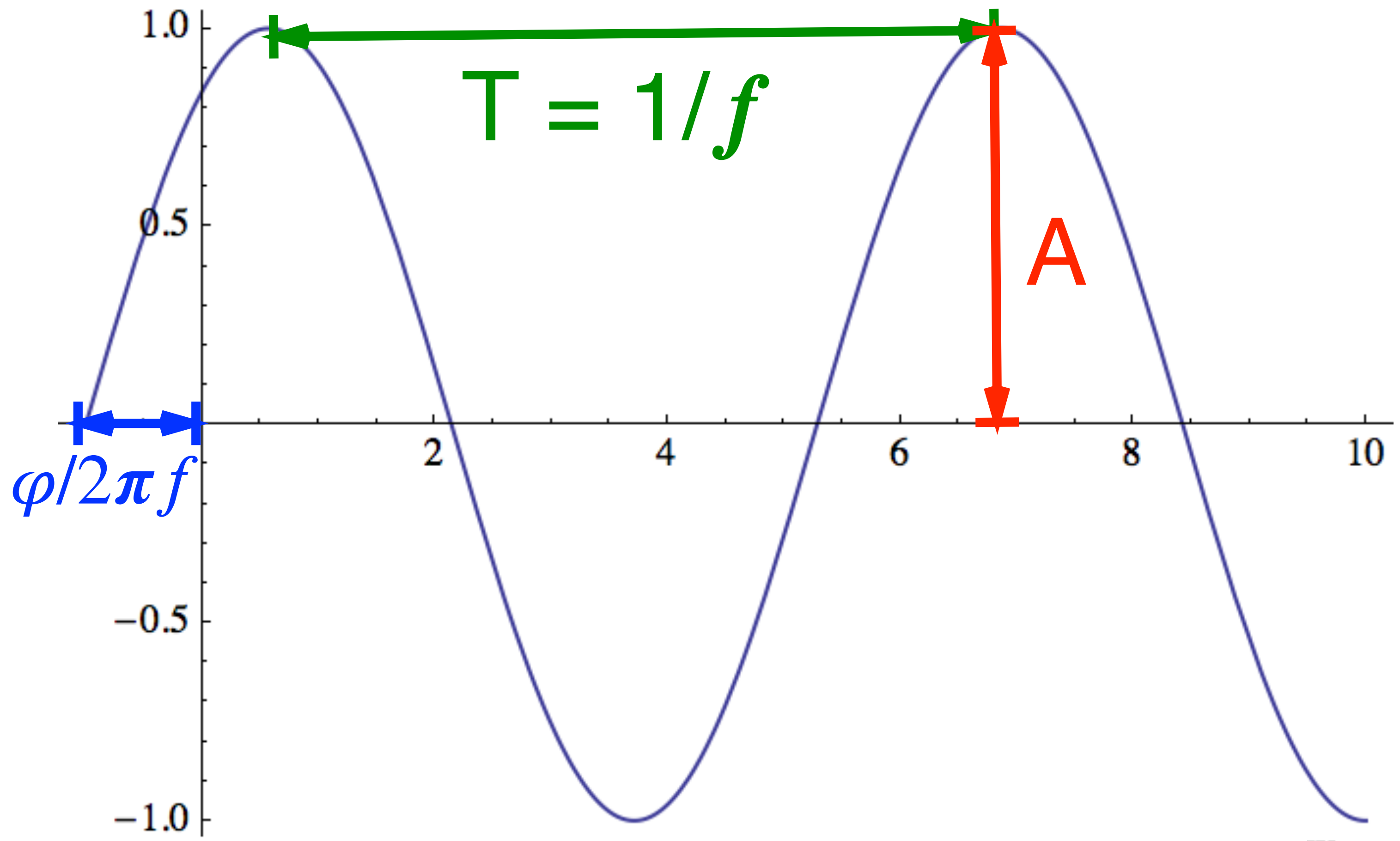
- MOdulation/DEModulation
  - Translation of the channel symbols by
    - amplitude modulation
    - phase modulation
    - frequency modulation
    - or a combination thereof



Broadband

- Idea
  - Focusing on the ideal frequency of the medium
  - Using a sine wave as the carrier wave signals
- A sine wave has no information
  - the sine curve continuously (modulated) changes for data transmission,
  - implies spectral widening (more frequencies in the Fourier analysis)
- The following parameters can be changed:
  - Amplitude A
  - Frequency f=1/T
  - Phase φ

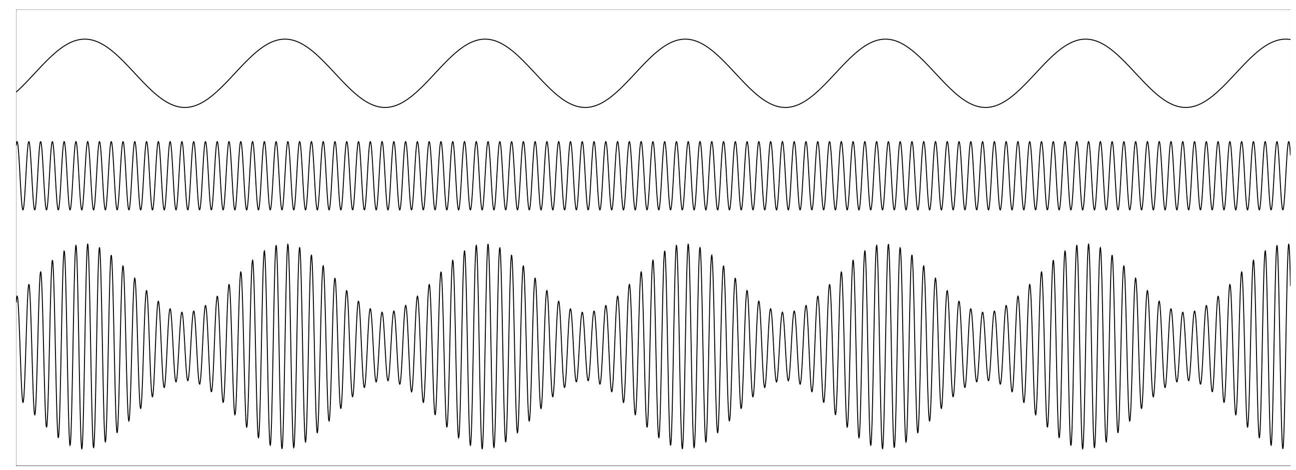
$$s(t) = A \sin(2\pi f t + \phi)$$



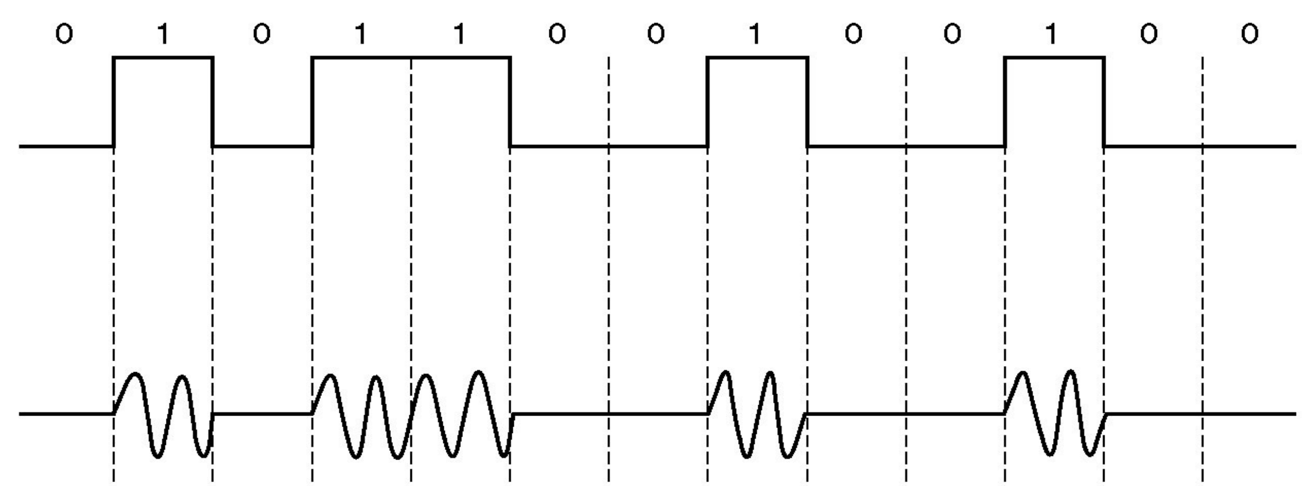
Amplitude Modulation

- The time-varying signal s (t) is encoded as the amplitude of a sine curve:

$$f_A(t) = s(t)\,\mathrm{sin}(2\pi f t + \phi)$$



- Analog Signal
- Digital signal
  - amplitude keying
  - special case: symbols 0 or 1
    - on / off keying

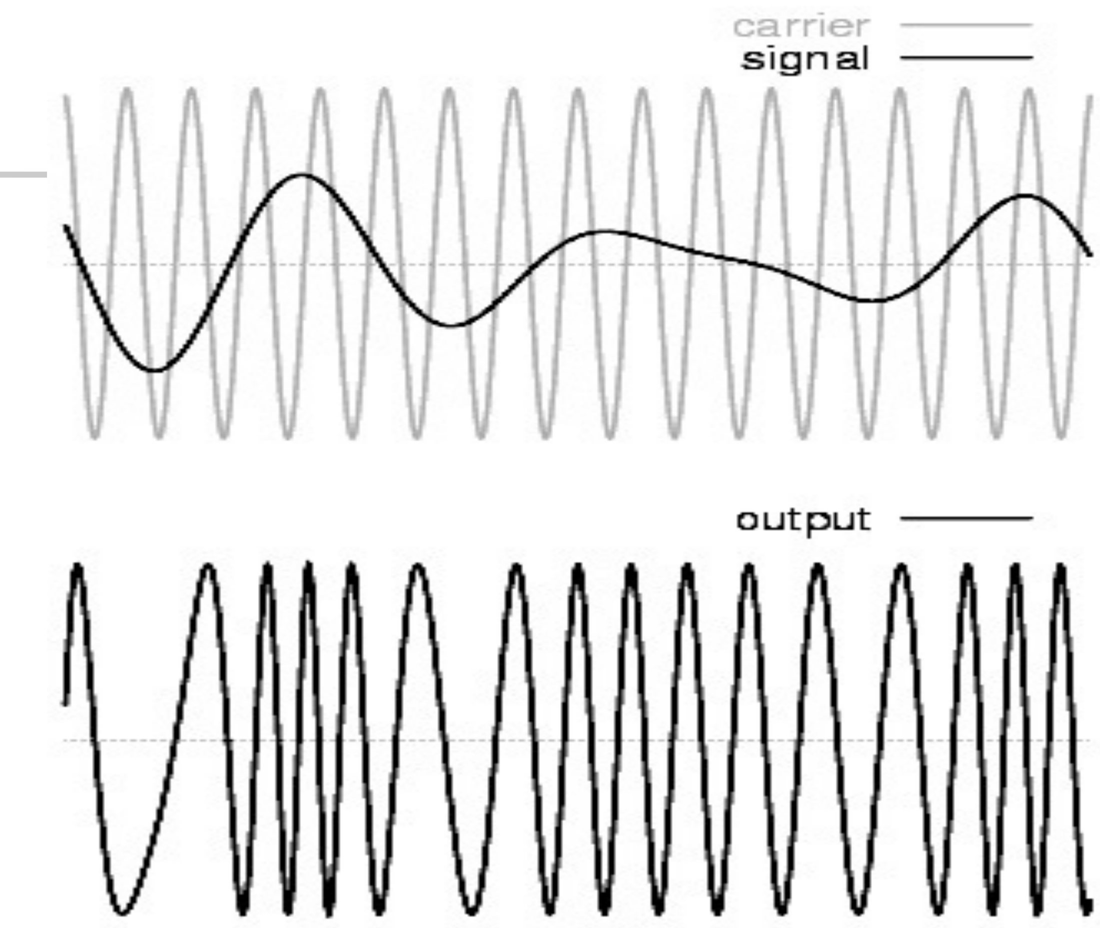


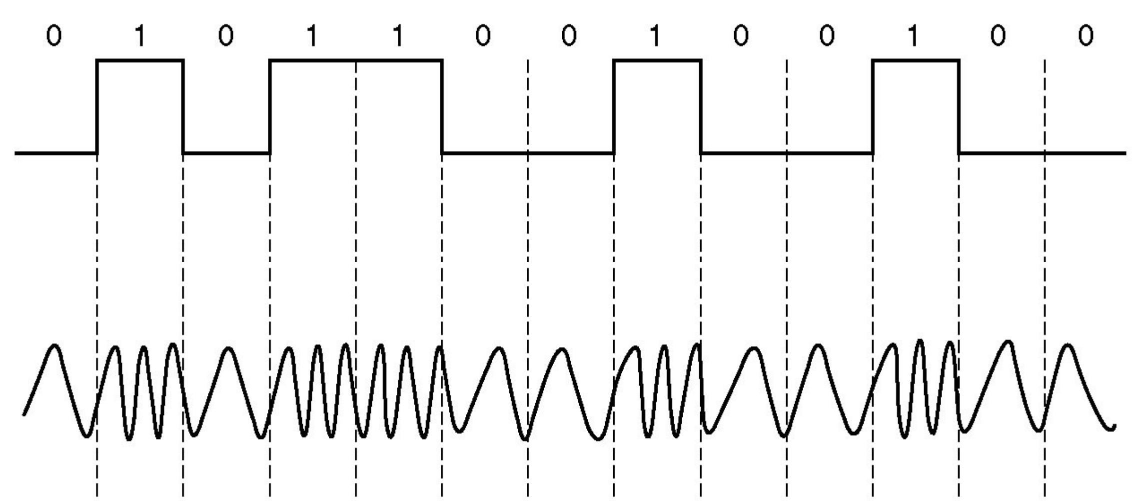
Frequency Modulation

- The time-varying signal s (t) is encoded in the frequency of the sine curve:

$$f_F(t) = a\,\mathsf{sin}(2\pi s(t)t + \phi)$$

- Analog signal
  - Frequency modulation (FM)
  - Continuous function in time
- Digital signal
  - Frequency Shift Keying (FSK)
  - E.g. frequencies as given by symbols



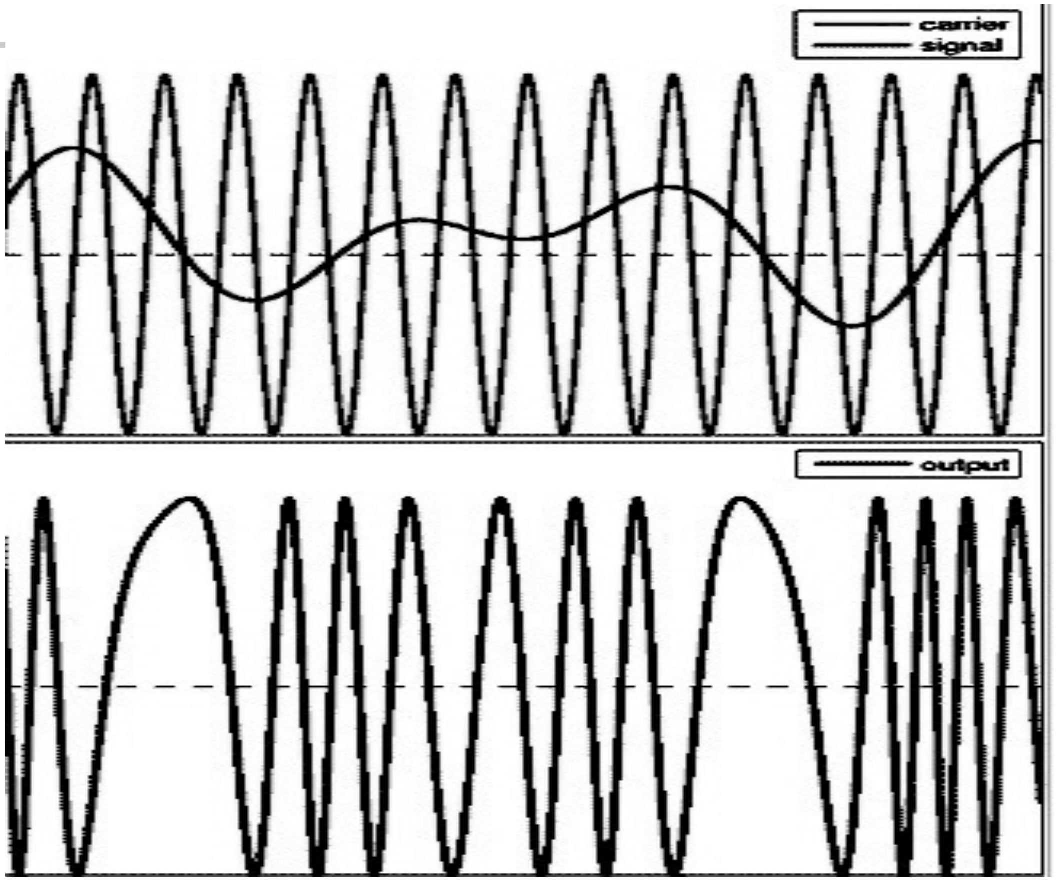


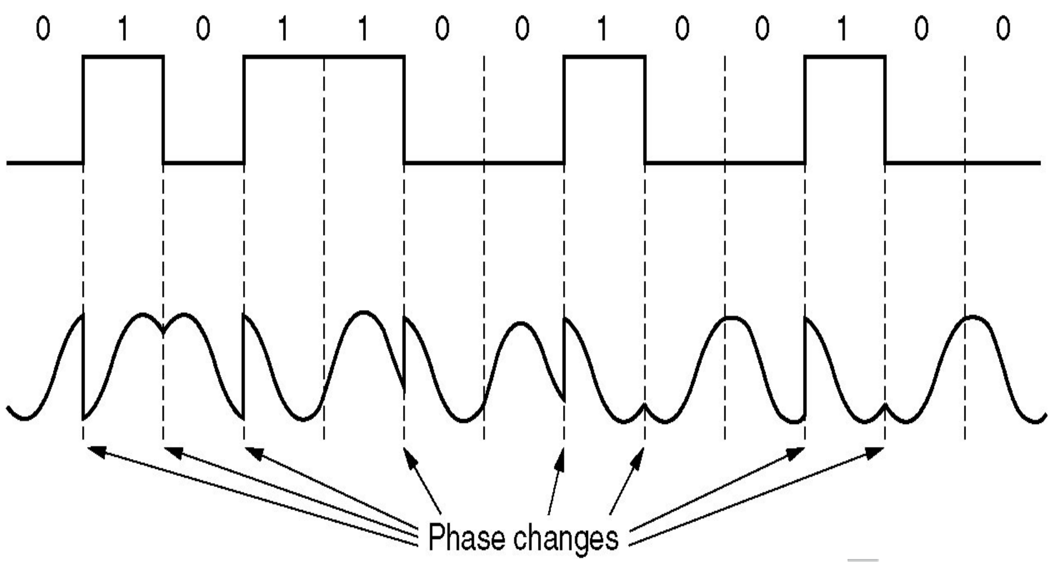
Phase Modulation

- The time-varying signal s (t) is encoded in the phase of the sine curve:

$$f_P(t) = a \sin(2\pi f t + s(t))$$

- Analog signal
  - phase modulation (PM)
  - very unfavorable properties
  - es not used
- Digital signal
  - phase-shift keying (PSK)
  - e.g. given by symbols as phases





Digital and Analog signals in Comparison

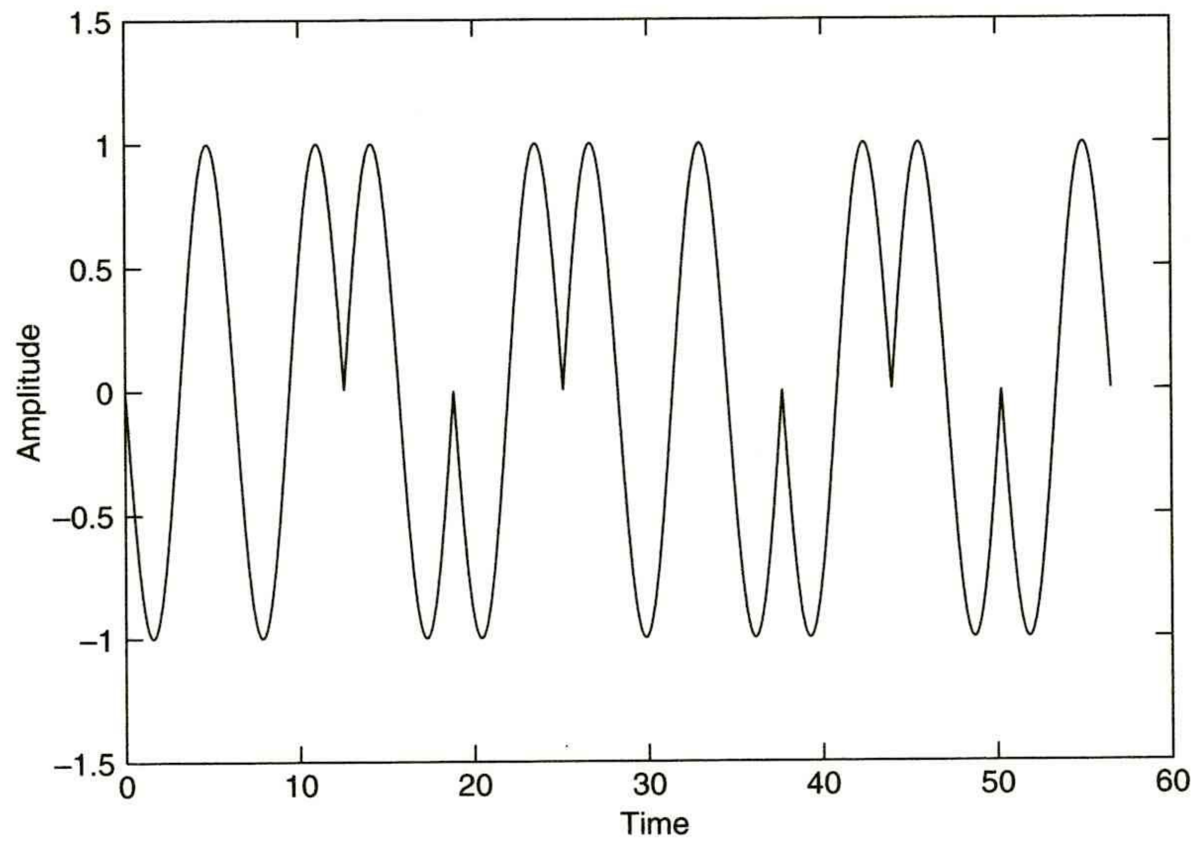
- For a station there are two options
  - digital transmission
    - finite set of discrete signals
    - e.g. finite amount of voltage sizes / voltages
  - analog transmission
    - Infinite (continuous) set of signals
    - E.g. Current or voltage signal corresponding to the wire
- Advantage of digital signals:
  - There is the possibility of receiving inaccuracies to repair and reconstruct the original signal
  - Any errors that occur in the analog transmission may increase further

Phase Shift Keying (PSK)

- For phase signals $\phi_i(t)$

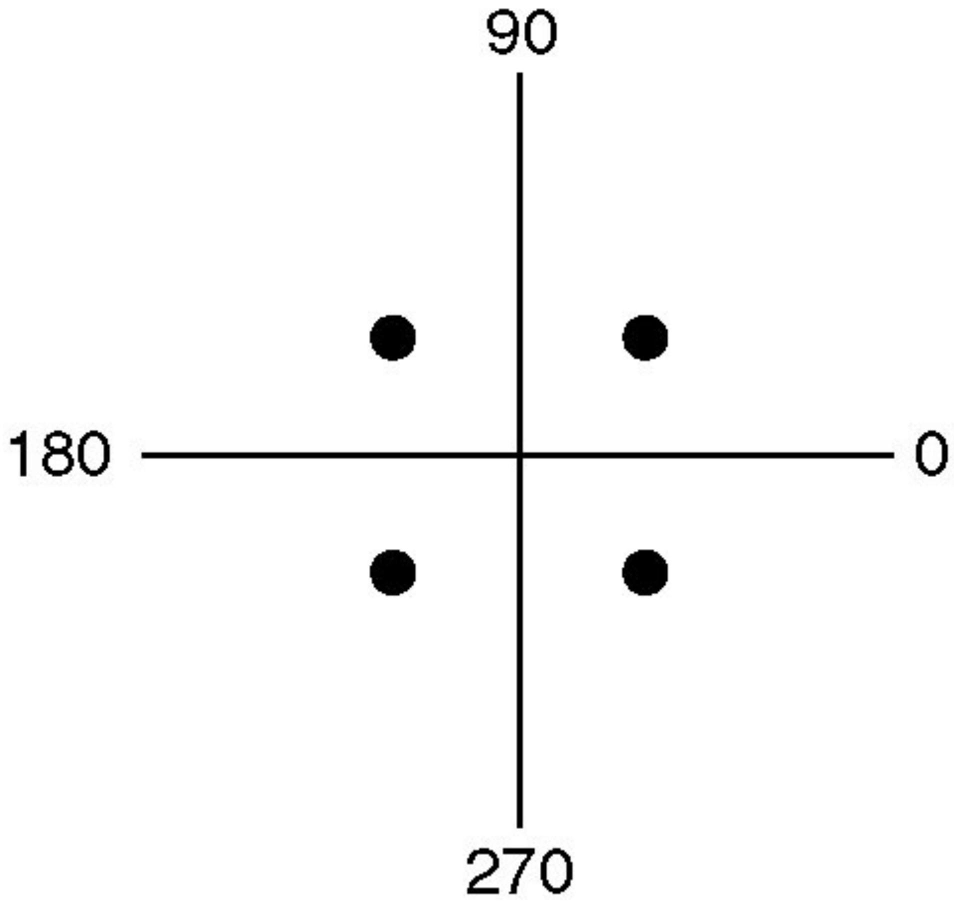
$$s_i(t) = \sqrt{\frac{2E}{T}} \cdot \sin(\omega_0 t + \phi_i(t))$$

- Example:



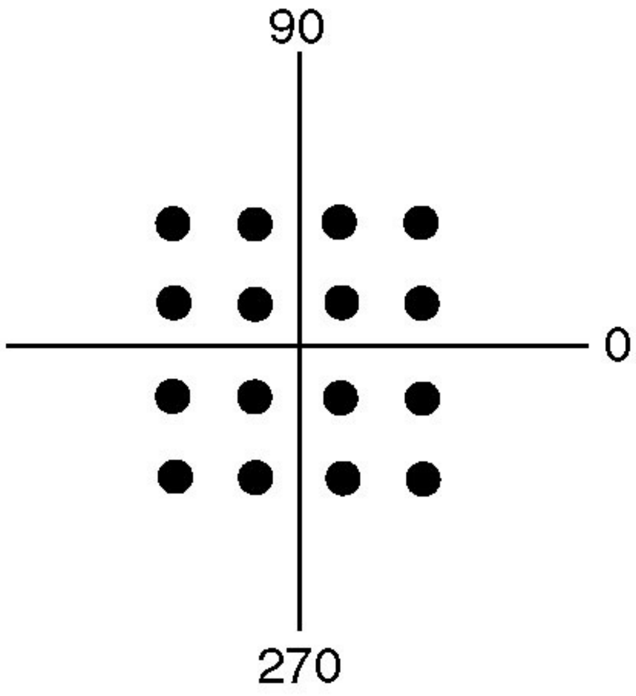
PSK with Different Symbols

- Phase shifts can be detected by the receiver very well
- Encoding various Symoble very simple
  - Using phase shift e.g. π / 4, 3/4π, 5/4π, 7/4π
    - rarely: phase shift 0 (because of synchronization)
  - For four symbols, the data rate is twice as large as the symbol rate
- This method is called Quadrature Phase Shift Keying (QPSK)



Amplitude and Phase Modulation

- Amplitude and phase modulation can be successfully combined
  - Example: 16-QAM (Quadrature Amplitude Modulation)
    - uses 16 different combinations of phases and amplitudes for each symbol
    - Each symbol encodes four bits ($2^4 = 16$)
  - The data rate is four times as large as the symbol rate



Nyquist's Theorem

- Definition
  - The band width H is the maximum frequency in the Fourier decomposition
- Assume
  - The maximum frequency of the received signal is f = H in the Fourier transform
    - (Complete absorption [infinite attenuation] all higher frequencies)
  - The number of different symbols used is V
  - No other interference, distortion or attenuation of
- Nyquist theorem
  - The maximum symbol rate is at most 2 H baud.
  - The maximum possible data rate is a bit more than $2 \log_2 H V$ / s.

Do more symbols help?

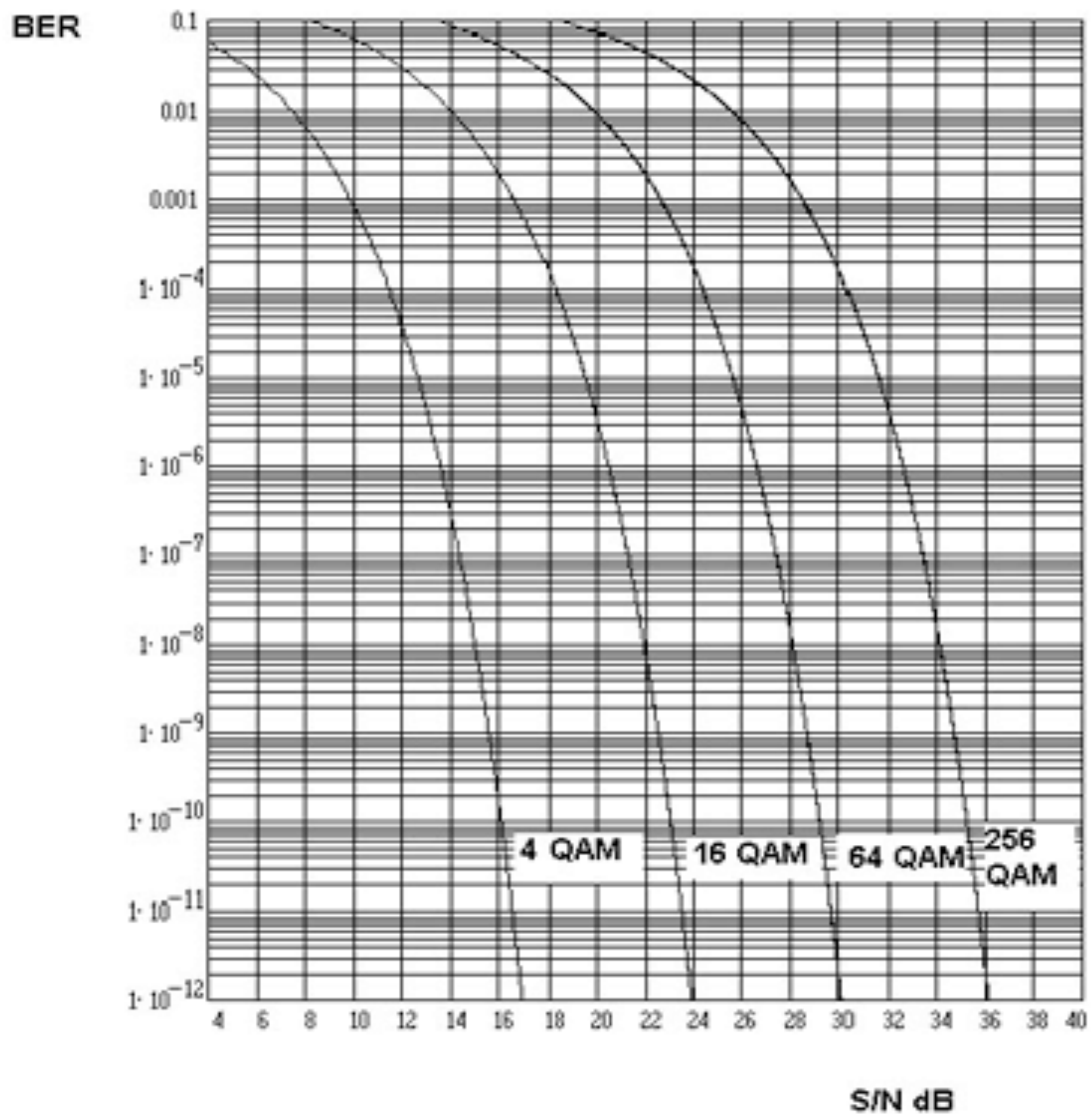
- Nyquist's theorem states that could theoretically be increased data rate with the number of symbols used
- Discussion:
  - Nyquist's theorem provides a theoretical upper bound and no method of transmission
  - In practice there are limitations in the accuracy
  - Nyquist's theorem does not consider the problem of noise

The Theorem of Shannon

- Indeed, the influence of the noise is fundamental
  - Consider the relationship between transmission intensity S to the strength of the noise N
  - The less noise the more signals can be better recognized
- Theorem of Shannon
  - The maximum possible data rate is $H \log_2(1 + S / N)$ bits/s
    - with bandwidth H
    - Signal strength S
- Attention
  - This is a theoretical upper bound
  - Existing codes do not reach this value

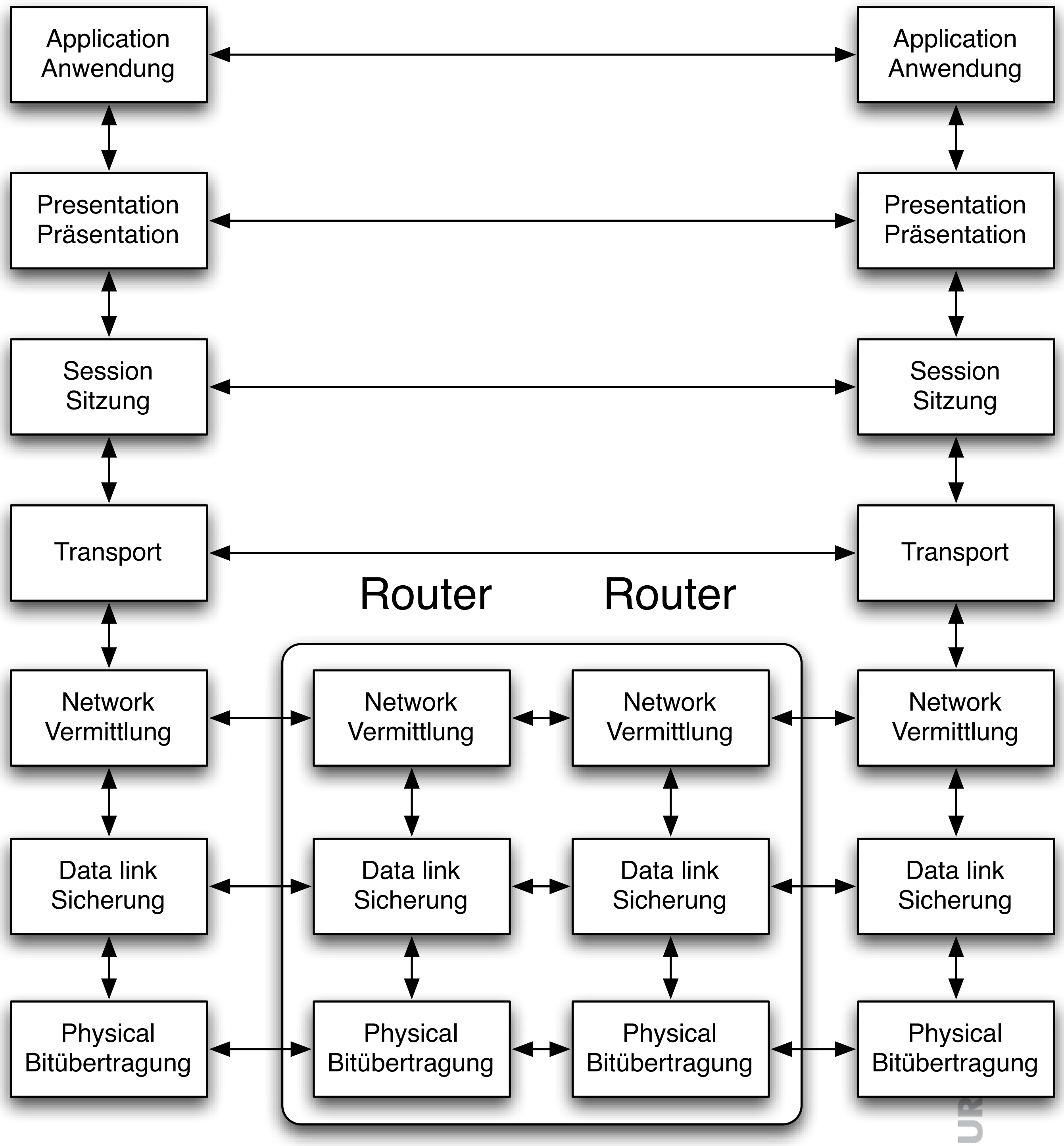
Bit Error Rate and SINR

- Higher SIR decreases Bit Error Rate (BER)
  - BER is the rate of faul received bits
- Depends from the
  - signal strength
  - noise
  - bandwidth
  - encoding
- Relationship of BER and SINR
  - Example: 4 QAM, 16 QAM, 64 QAM, 256 QAM



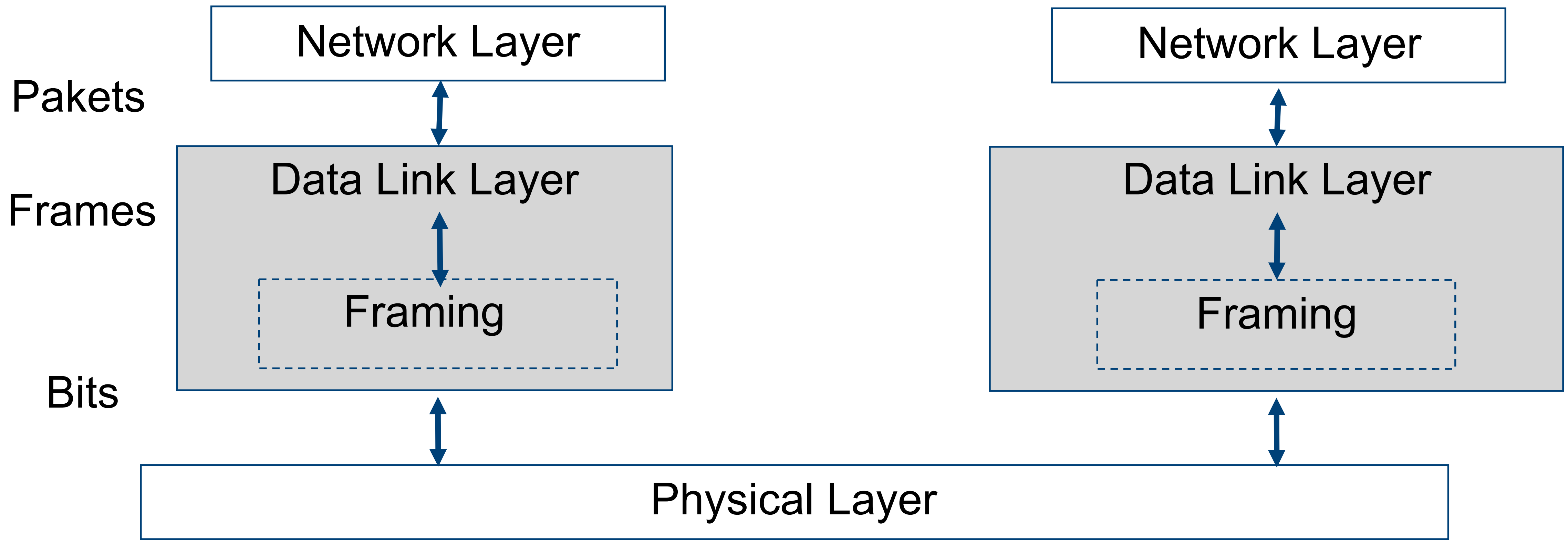
ISO/OSI Reference model

- 7. Application
  - Data transmission, e-mail, terminal, remote login
- 6. Presentation
  - System-dependent presentation of the data (EBCDIC / ASCII)
- 5. Session
  - start, end, restart
- 4. Transport
  - Segmentation, congestion
- 3. Network
  - Routing
- 2. Data Link
  - Checksums, flow control
- 1. Physical
  - Mechanics, electrics



Data Link Layer: Frames

- Framing for the physical layer into „frames“
  - for error control



Data Link Layer: Error Control

- Error detection
  - erroneous bits?
- Error correction
  - correction of bit errors
  - Forward Error Correction
    - Redundant coding without addition transmssions
  - Backward Error Correction
    - After detection resend frame

Sessions

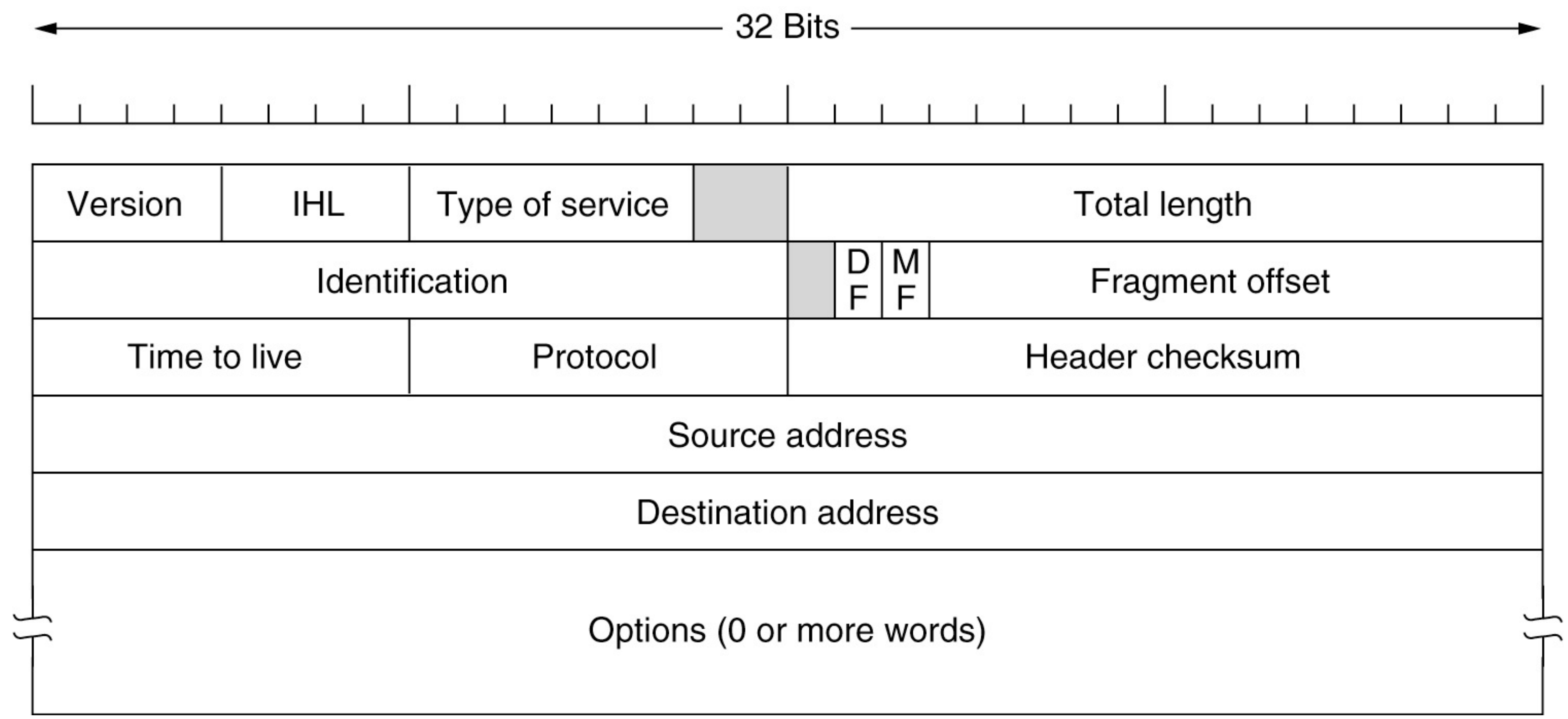
- Use of conections
  - control of the connection status
    - correctness of the protocol
  - error control
    - common context between sender and receiver

Flow control

- Problem: fast Sender ans slow receiver
- Adaption of the sending frame rate for the receivers

IPv4-Header (RFC 791)

- Version: 4 = IPv4
- IHL: IP header length
  - in 32 bit words (>5)
- Type of service
  - optimize delay, throughput, reliability, monetary cost
- Checksum (only IP-header)
- Source and destination IP-address
- Protocol identifies protocol
  - e.g. TCP, UDP, ICMP, IGMP
- Time to Live:
  - maximal number of hops



IP addresses and Domain Name System

- IP addresses
  - every interface in a network has a unique world wide IP address
  - separated in Net-ID and Host-ID
  - Net-ID assigned byInternet Network Information Center
  - Host-ID by local network administration
- Domain Name System (DNS)
  - replaces IP addresses like 132.230.167.230 by names, e.g. falcon.informatik.uni-freiburg.de and vice versa
  - Robust distributed database

Internet IP Addresses
Classfull Addresses until 1993

- Classes A, B, and C
- D for multicast; E: “reserved”

32 Bits

Class				Range of host addresses	
A	0	Network	Host	1.0.0.0 to 127.255.255.255	128 NWs; 16 M hosts
B	10	Network	Host	128.0.0.0 to 191.255.255.255	16K NWs; 64K hosts
C	110	Network	Host	192.0.0.0 to 223.255.255.255	2M NWs; 256 hosts
D	1110	Multicast address		224.0.0.0 to 239.255.255.255	
E	1111	Reserved for future use		240.0.0.0 to 255.255.255.255	

codes classes

Classless IPv4-Addresses

- Until 1993 (deprecated)
  - 5 classes marked by Präfix
  - Then sub-net-id prefix of fixed length and host-id
- Since 1993
  - Classless Inter-Domain-Routing (CIDR)
  - Net-ID and Host-ID are distributed flexibly
  - E.g.
    - Network mask /24 or 11111111.11111111.11111111.00000000
    - denotes, that IP-address
      - 10000100. 11100110. 10010110. 11110011
      - consists of network 10000100. 11100110. 10010110
      - and host 11110011
- Route aggregation
  - Routing protocols BGP, RIP v2 and OSPF can address multiple networks using one ID
    - Z.B. all Networks with ID 10010101010* can be reached over host X

Routing Tables and Packet Forwarding

- IP Routing Table
  - contains for each destination the address of the next gateway
  - destination: host computer or sub-network
  - default gateway
- Packet Forwarding
  - IP packet (datagram) contains start IP address and destination IP address
    - if destination = my address then hand over to higher layer
    - if destination in routing table then forward packet to corresponding gateway
    - if destination IP subnet in routing table then forward packet to corresponding gateway
    - otherwise, use the default gateway

IP Packet Forwarding

- IP -Packet (datagram) contains...
  - TTL (Time-to-Live): Hop count limit
  - Start IP Address
  - Destination IP Address
- Packet Handling
  - Reduce TTL (Time to Live) by 1
  - If TTL ≠ 0 then forward packet according to routing table
  - If TTL = 0 or forwarding error (buffer full etc.):
    - delete packet
    - if packet is not an ICMP Packet then
      - send ICMP Packet with
      - start = current IP Address
      - destination = original start IP Address

Introduction to Future IP

- IP version 6 (IP v6 – around July 1994)
- Why switch?
  - rapid, exponential growth of networked computers
  - shortage (limit) of the addresses
  - new requirements towards the Internet infrastructure (streaming, real-time services like VoIP, video on demand)
- evolutionary step from IPv4
- interoperable with IPv4

Capabilities of IP

- dramatic changes of IP
  - Basic principles still appropriate today
  - Many new types of hardware
  - Scale of Internet and interconnected computers in private LAN
- Scaling
  - Size - from a few tens to a few tens of millions of computers
  - Speed - from 9,6Kbps (GSM) to 10Gbps (Ethernet)
  - Increased frame size (MTU) in hardware

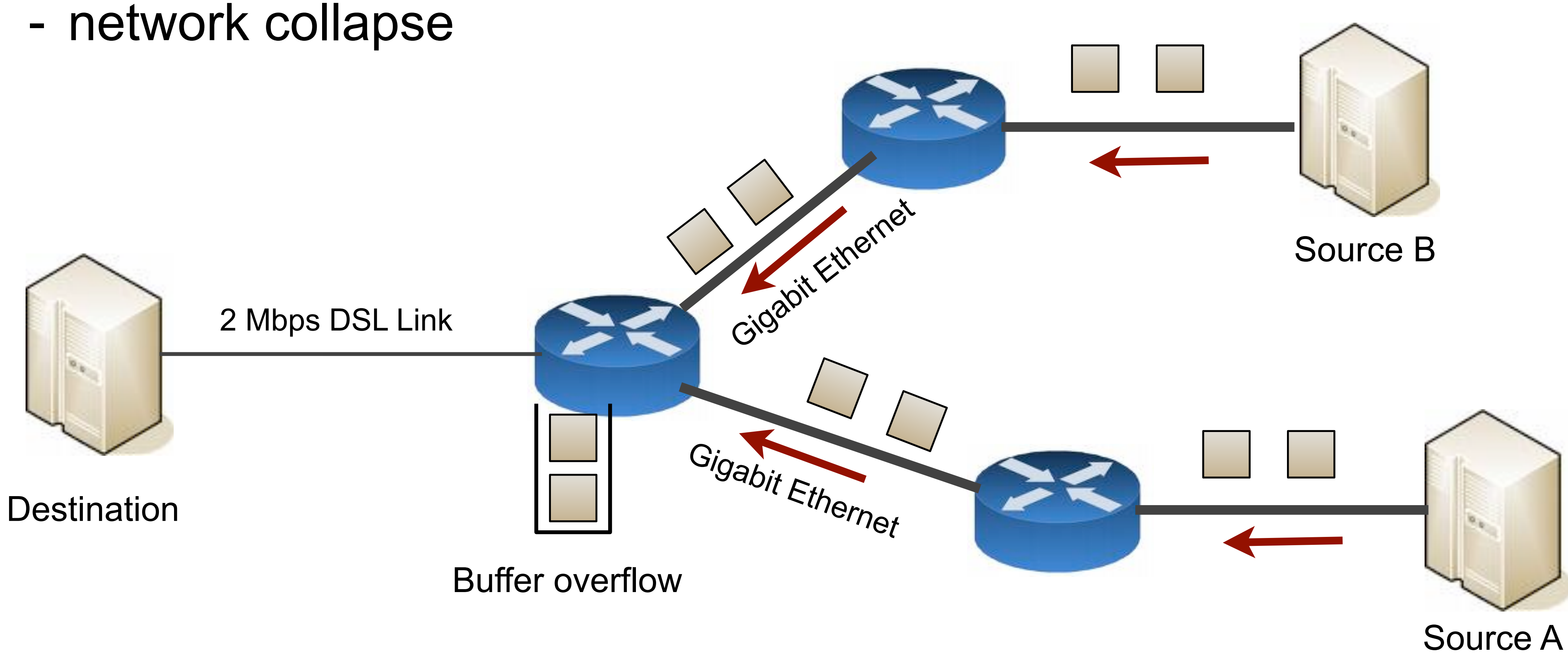
IPv6-Header (RFC 2460)

- Version: 6 = IPv6
- Traffic Class
  - for QoS (priority)
- Flow Label
  - QoS or real-time
- Payload Length
  - size of the rest of the IP packet
- Next Header (IPv4: protocol)
  - e..g. ICMP, IGMP, TCP, EGP, UDP, Multiplexing, ...
- Hop Limit (Time to Live)
  - maximum number of hops
- Source Address
- Destination Address
  - 128 bit IPv6 address

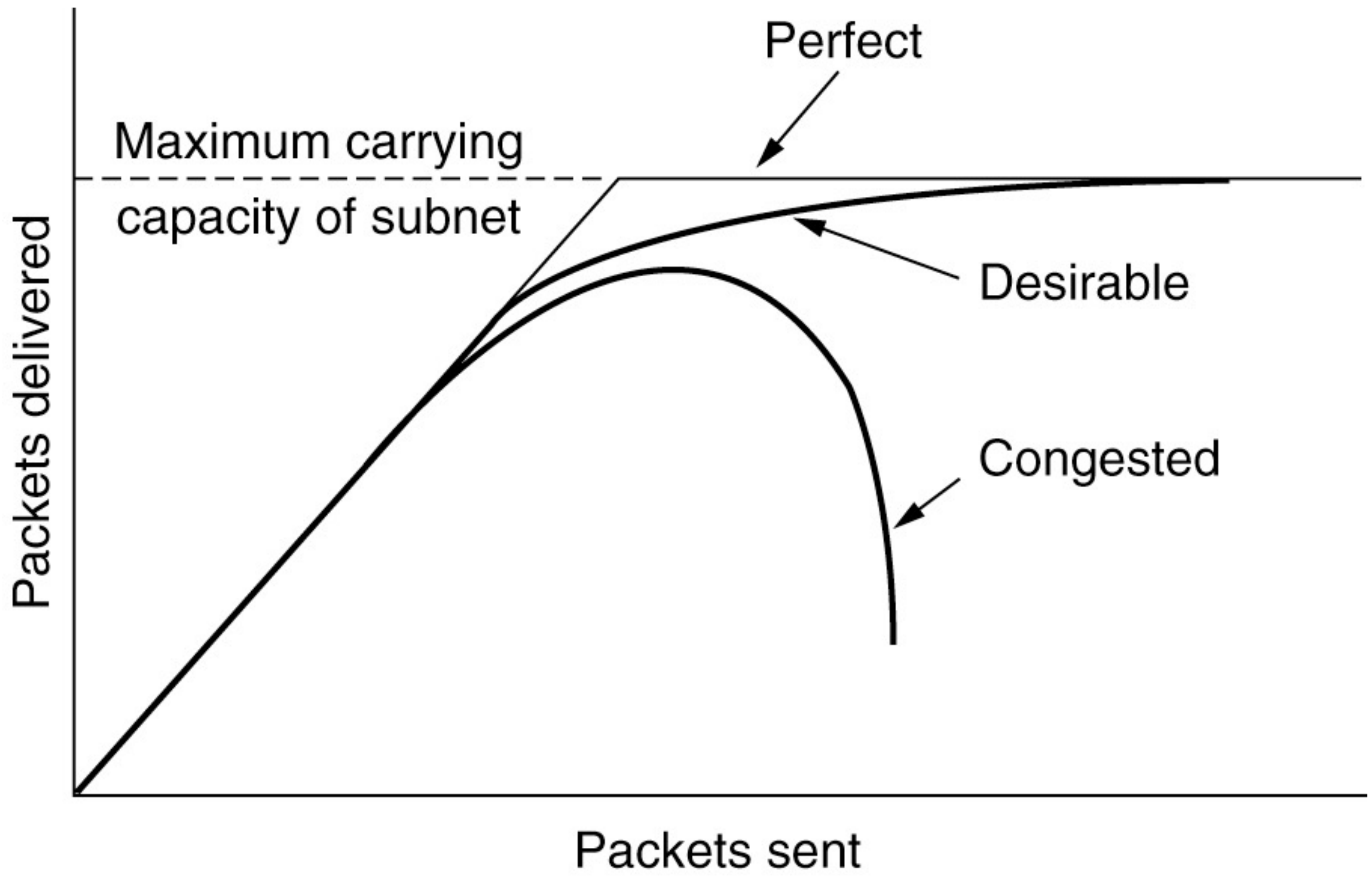
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version| Traffic Class |           Flow Label                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Payload Length        |  Next Header  |   Hop Limit   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+                                                               +
|                                                               |
+                         Source Address                        +
|                                                               |
+                                                               +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+                                                               +
|                                                               |
+                      Destination Address                      +
|                                                               |
+                                                               +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Network Congestion

- (Sub-)Networks have limited bandwidth
- Injecting too many packets leads to
  - network congestion
  - network collapse



Congestion and capacity

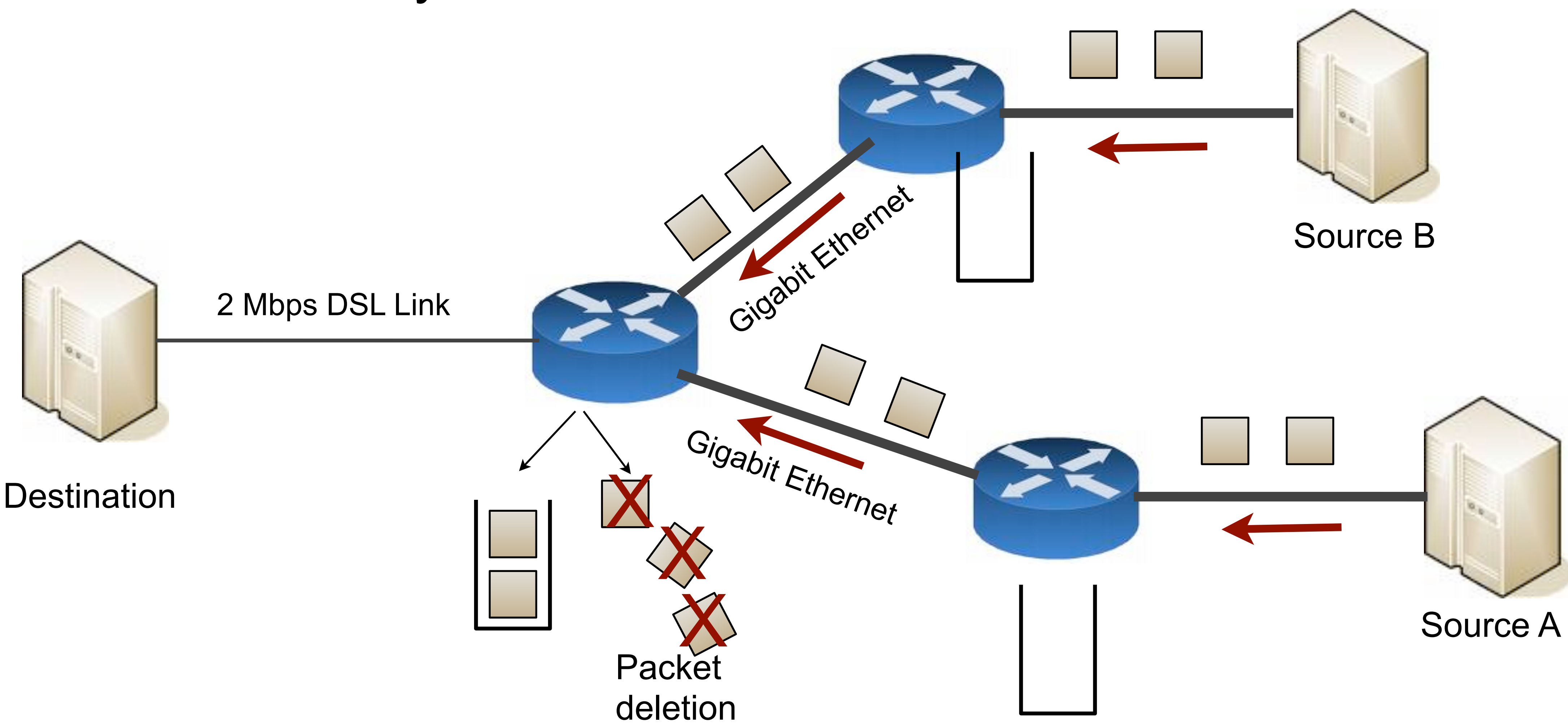


Congestion Prevention

Layer	Policies
Transport	• Retransmission policy • Out-of-order caching policy • Acknowledgement policy • Flow control policy • Timeout determination
Network	• Virtual circuits versus datagram inside the subnet • Packet queueing and service policy • Packet discard policy • Routing algorithm • Packet lifetime management
Data link	• Retransmission policy • Out-of-order caching policy • Acknowledgement policy • Flow control policy

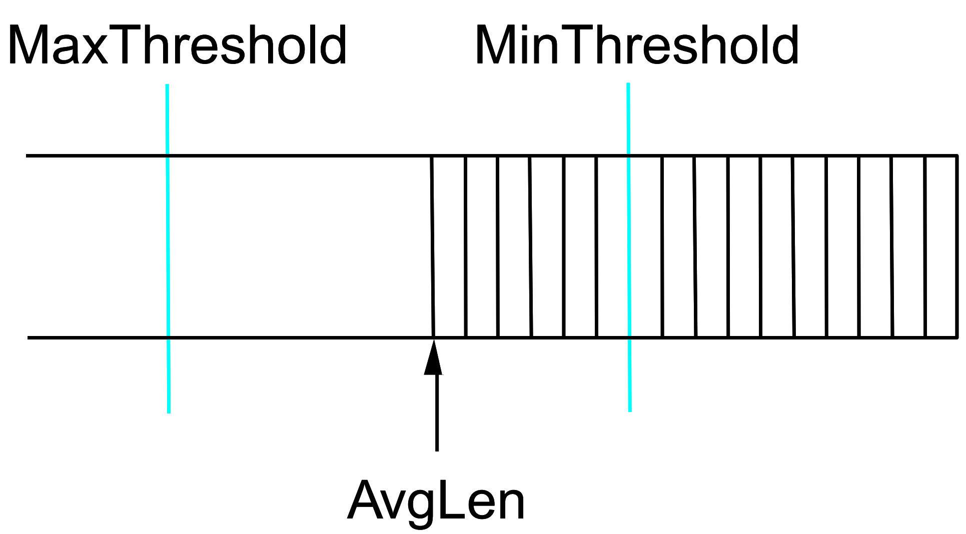
Congestion Prevention by Routers

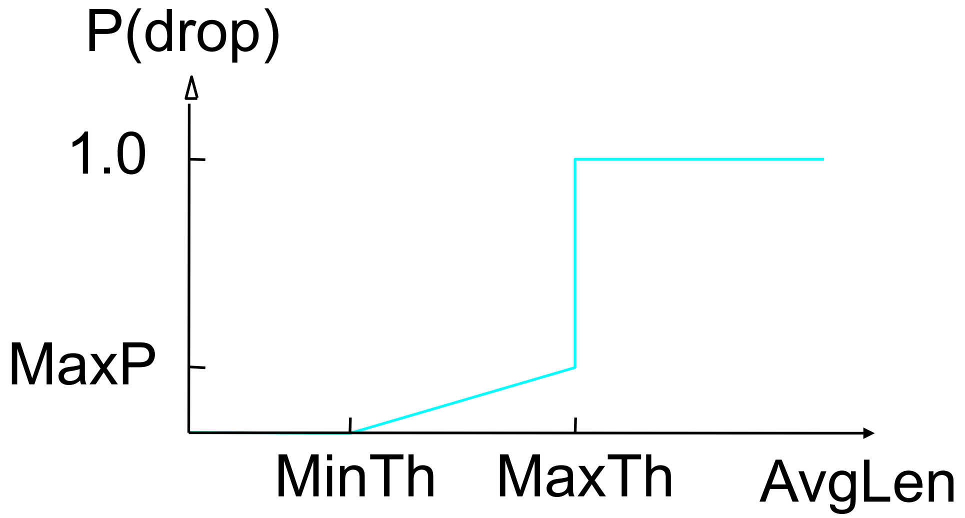
- IP Routers drop packets
  - Tail dropping
  - Random Early Detection



Random early detection (RED)

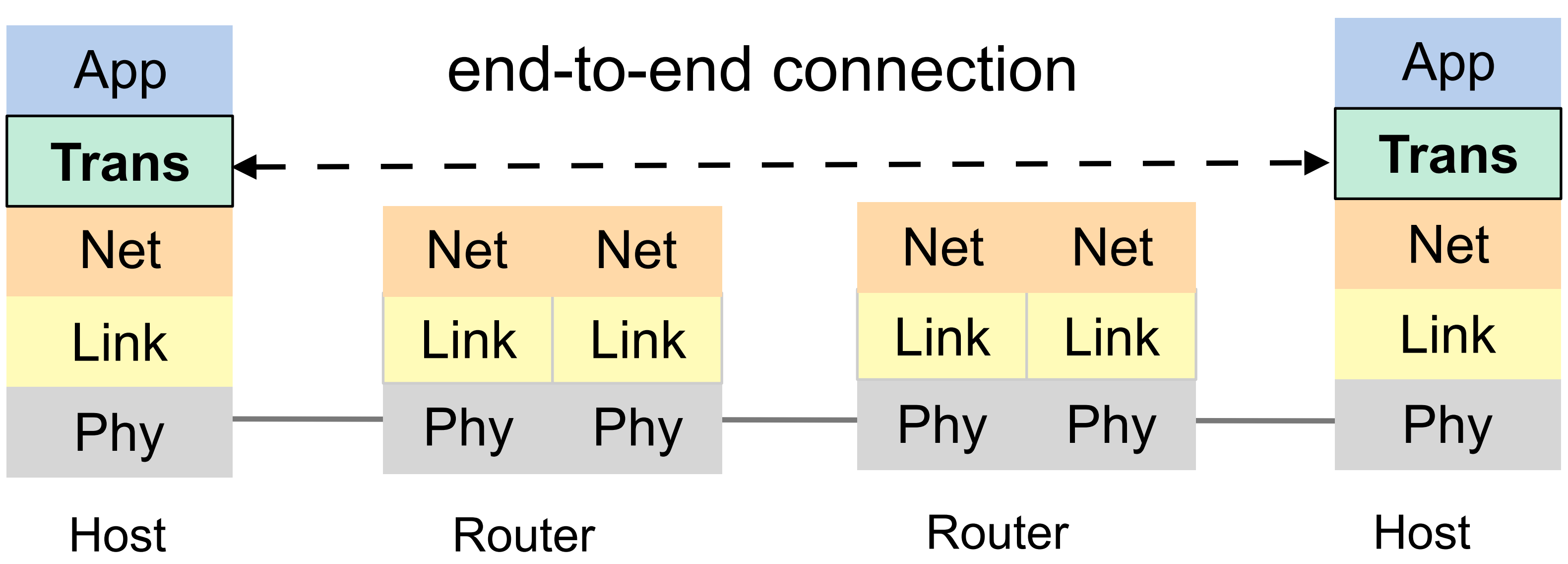
- Packet dropping probability grows with queue length
- Fairer than just “tail dropping”: the more a host transmits, the more likely it is that its packets are dropped





The Transport Layer

- TCP (Transmission Control Protocol
  - connection-oriented
  - delivers a stream of bytes
  - reliable and ordered
- UDP (User Datagram Protocol)
  - delivery of datagrams
  - connectionless, unreliable, unordered



UDP-Header

- Port addresses
  - for parallel UDP connections
- Length
  - data + header length
- Checksum
  - for header and data

```
0      7 8     15 16    23 24    31
 +--------+--------+--------+--------+
 |     Source      |   Destination   |
 |      Port       |      Port       |
 +--------+--------+--------+--------+
 |                 |                 |
 |     Length      |    Checksum     |
 +--------+--------+--------+--------+
 |
 |          data octets ...
 +---------------- ...
```

The Transmission Control Protocol (TCP)

- Connection-oriented
- Reliable delivery of a byte stream
  - fragmentation and reassembly (*TCP segments*)
  - acknowledgements and retransmission
- In-order delivery, duplicate detection
  - sequence numbers
- Flow control and congestion control
  - window-based (receiver window, congestion window)
- challenge: IP (network layer) packets can be dropped, delayed, delivered out-of-order ...

TCP-Header

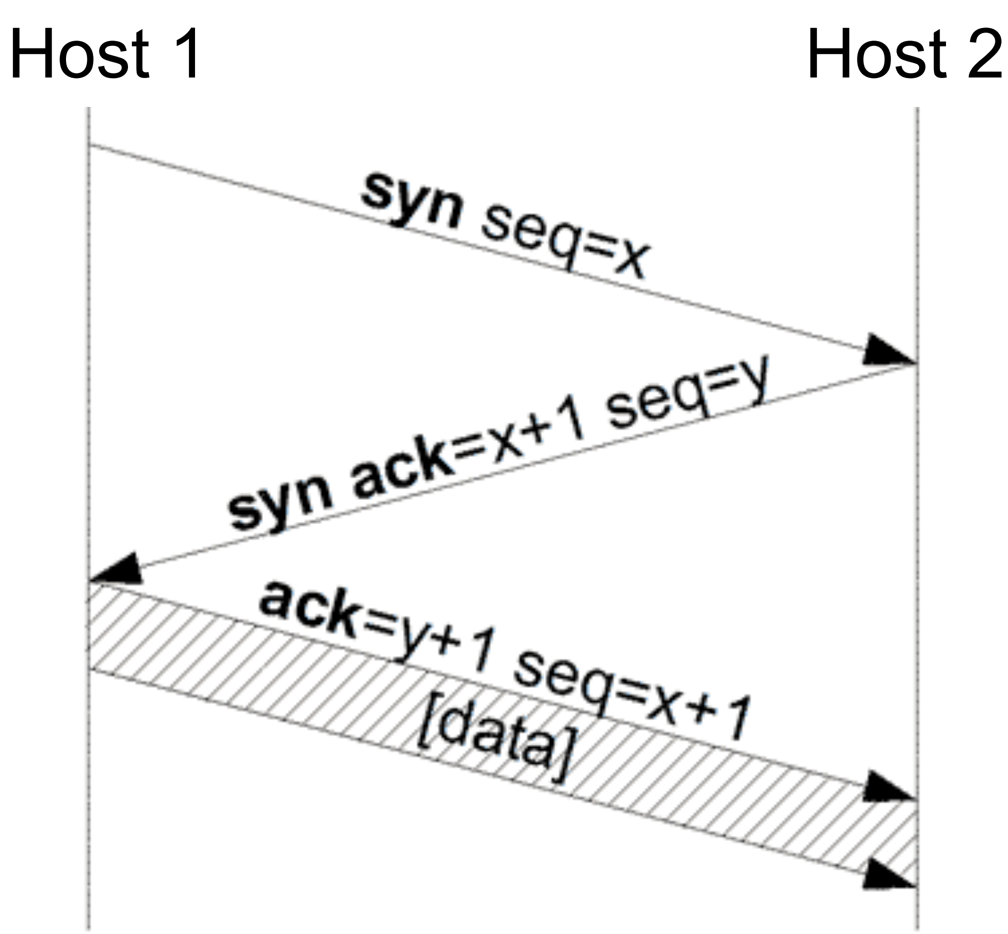
- Sequence number
  - number of the first byte in the segment
  - bytes are numbered modulo 2^{32}
- Acknowledge number
  - activated by ACK-Flag
  - number of the next data byte
    - = last sequence number + last amount of data
- Port addresses
  - for parallel TCP connections
- TCP Header length
  - data offset
- Check sum
  - for header and data

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Source Port          |       Destination Port        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Sequence Number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Acknowledgment Number                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Data |           |U|A|P|R|S|F|                               |
| Offset| Reserved  |R|C|S|S|Y|I|            Window             |
|       |           |G|K|H|T|N|N|                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Checksum            |         Urgent Pointer        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Options                    |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

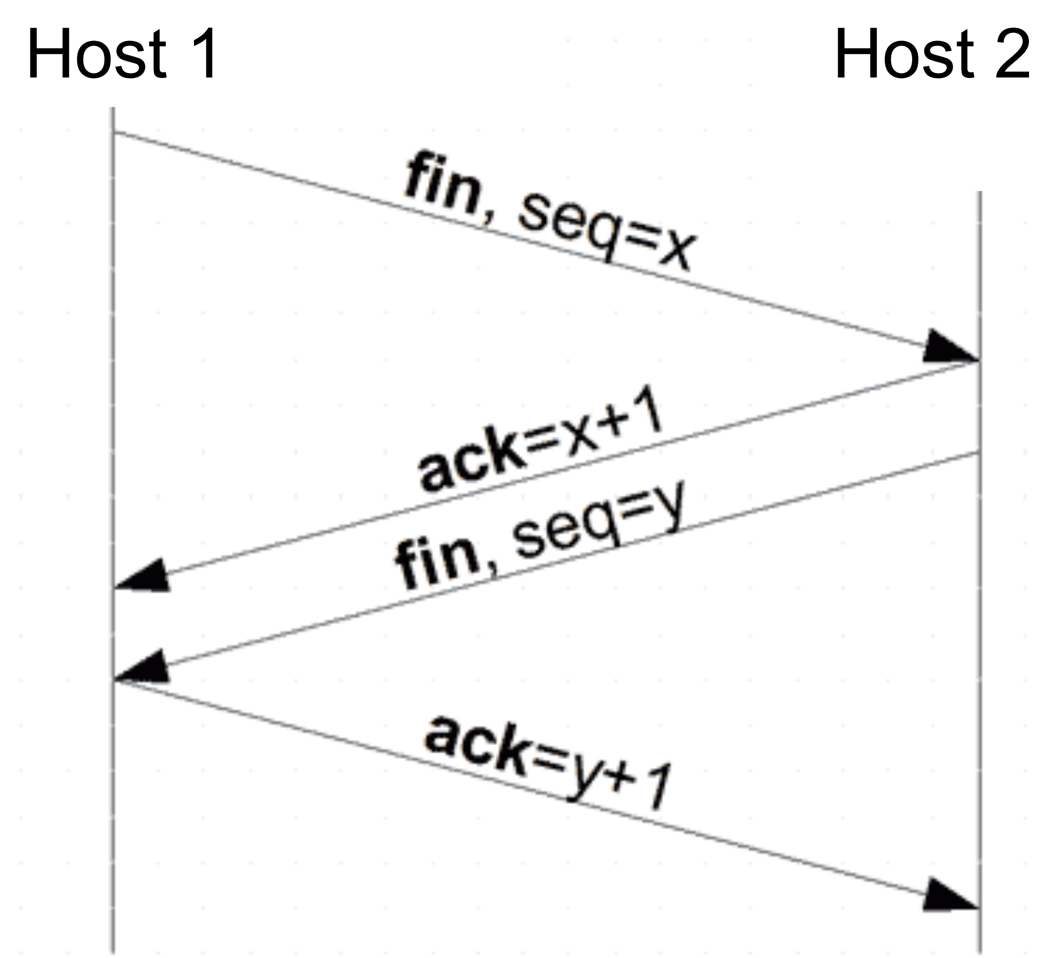
TCP Connections

- Connection establishment and teardown by 3-way handshake

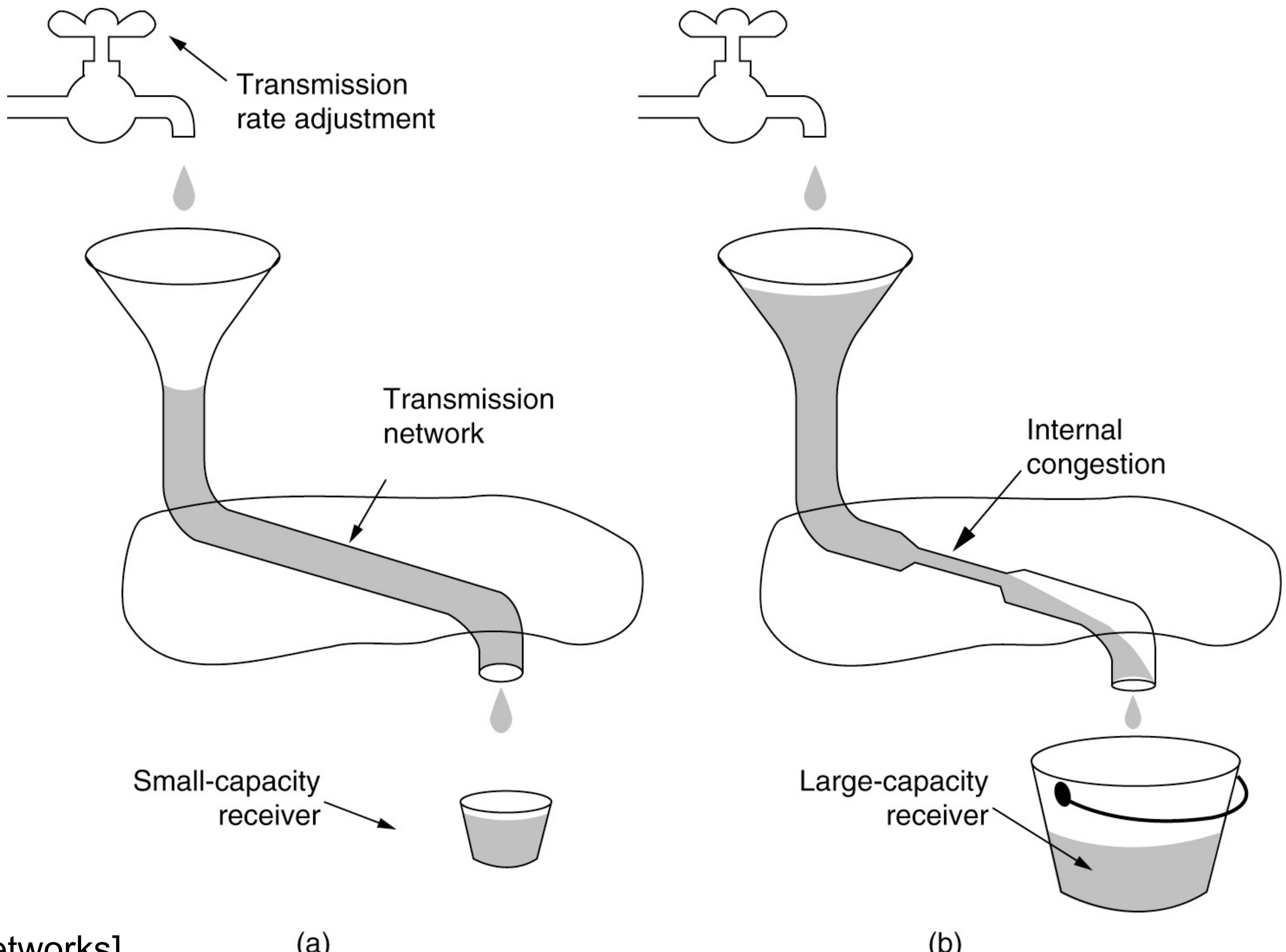
Connection establishment



Connection termination



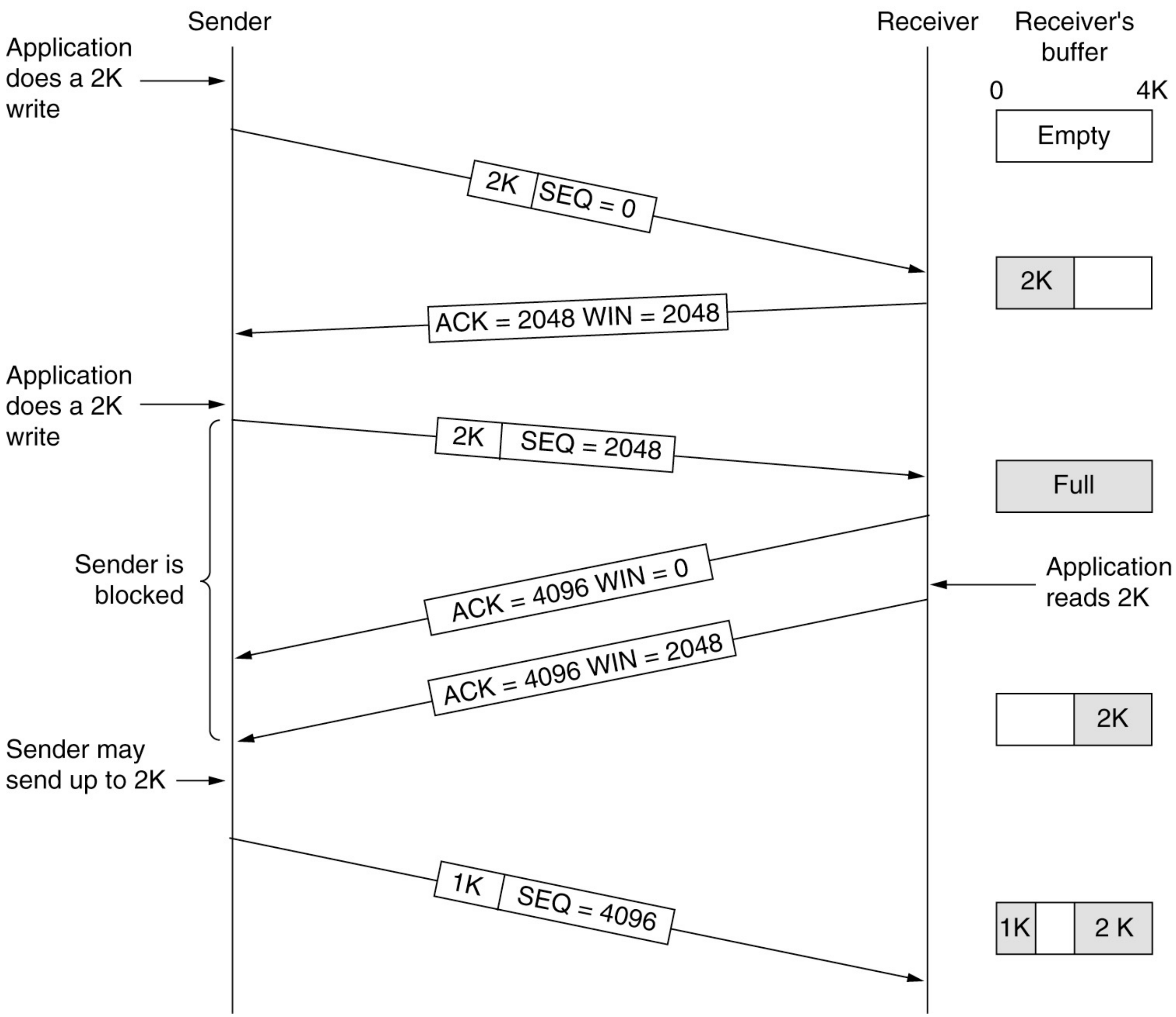
Flow control and congestion control



[Tanenbaum,
Computer Networks]

Flow Control

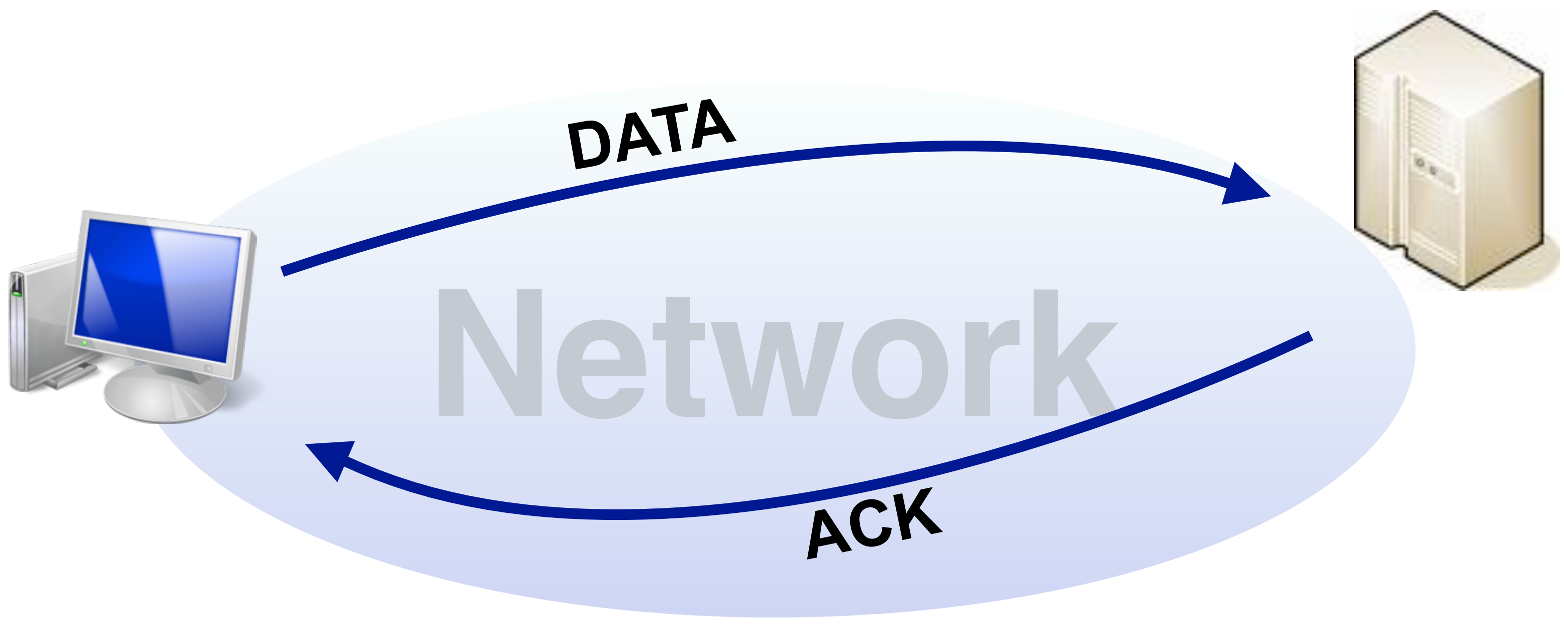
acknowledgements and window management



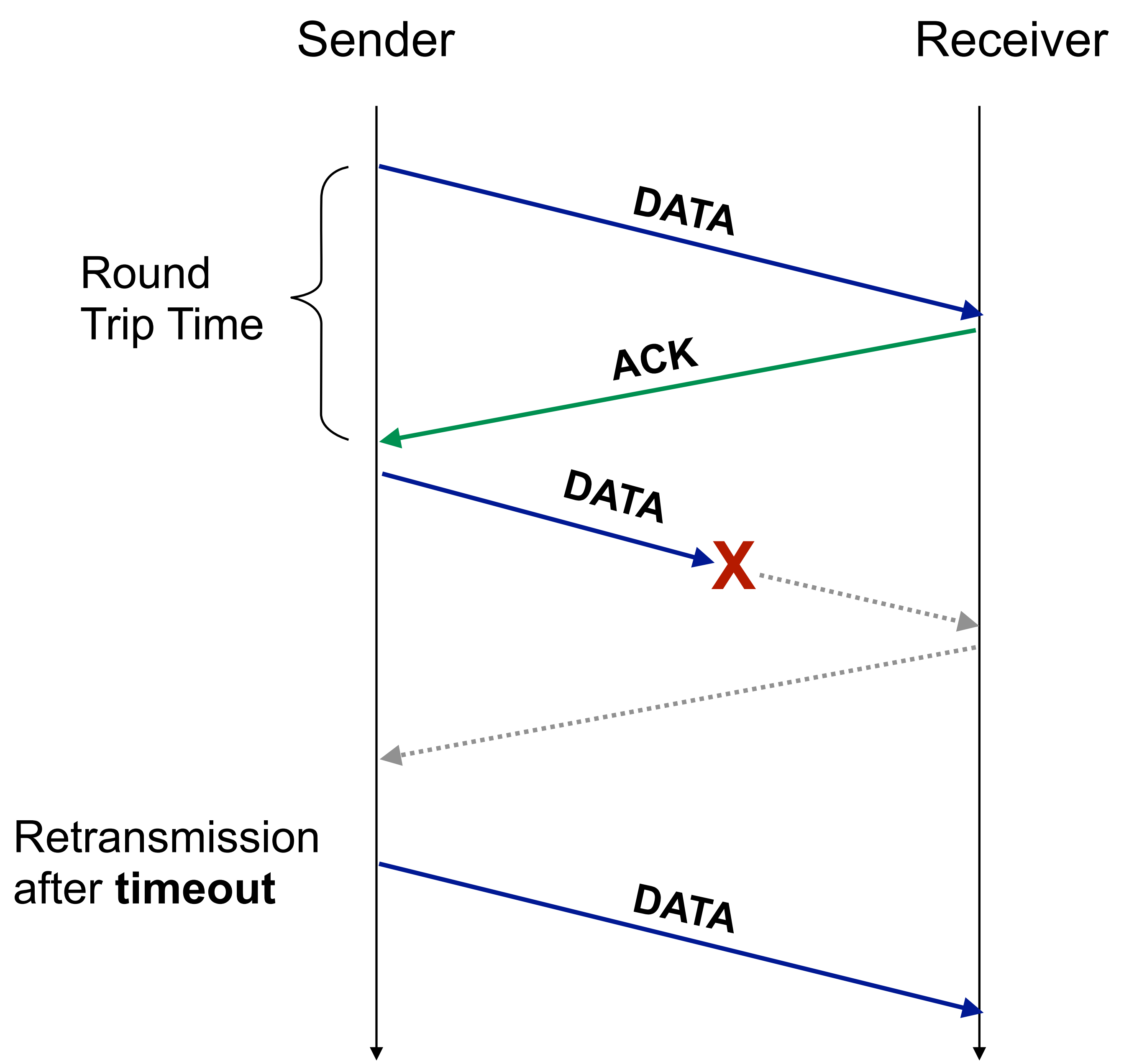
Retransmissions

- Retransmissions are triggered, if acknowledgements do not arrive
  ... but how to decide that?

- Measurement of the round trip time (RTT)



Retransmissions and RTT



Estimation of the Round Trip Time (RTT)

- If no acknowledgement arrives before expiry of the **Retransmission Timeout (RTO)**, the packet will be
retransmitted
  - RTT not predictable, fluctuating

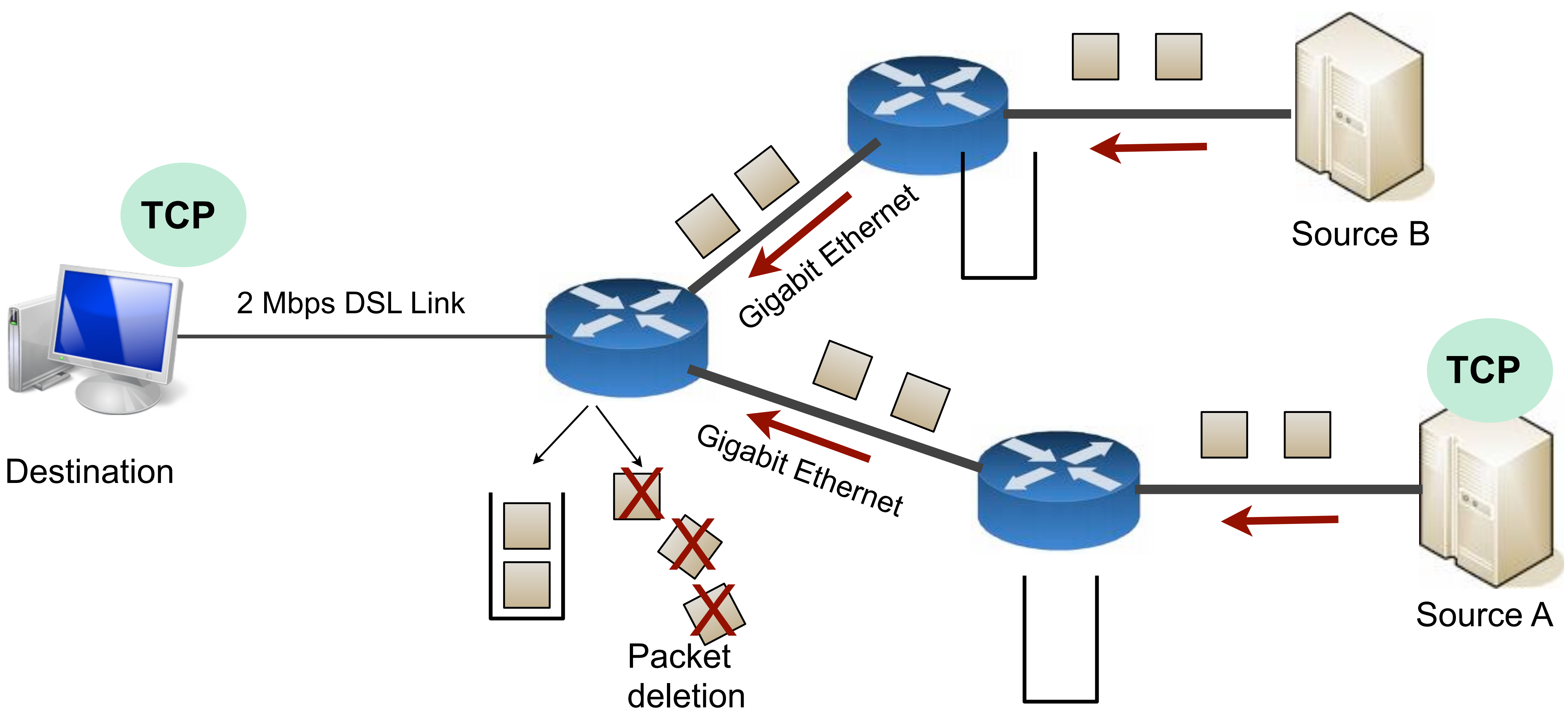
- **RTO derived from RTT estimation:**
  - RFC 793: (M := last RTT measurement)
    - RTT $\leftarrow \alpha$ RTT + (1-α) M, where $\alpha = 0{,}9$
    - RTO $\leftarrow \beta$ RTT, where $\beta = 2$
  - Alternative by Jacobson 88 (using the deviation D):
    - D $\leftarrow \alpha'$ D + (1-α') |RTT - M|
    - RTT $\leftarrow \alpha$ RTT + (1-α) M
    - RTO $\leftarrow$ RTT + 4D

TCP - Algorithm of Nagle

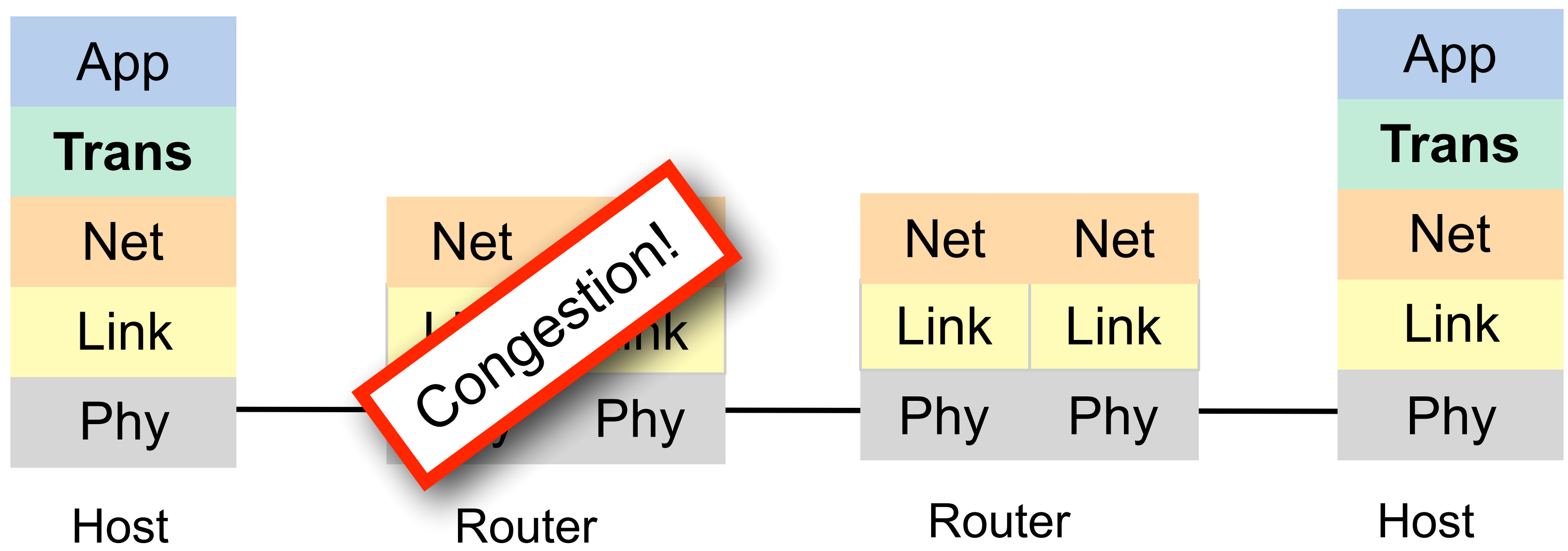
- How to ensure
  - small packages are shipped fast
  - yet, large packets are preferred
- Algorithm of Nagle
  - Small packets are not sent, as long as acks are still pending
    - Package is small, if data length <MSS
  - when the acknowledgment of the last packet arrives, the next one is sent
- Example:
  - terminal versus file transfer versus ftp
- Feature: self-clocking:
    - Quick link = many small packets
    - slow link = few large packets

Congestion revisited

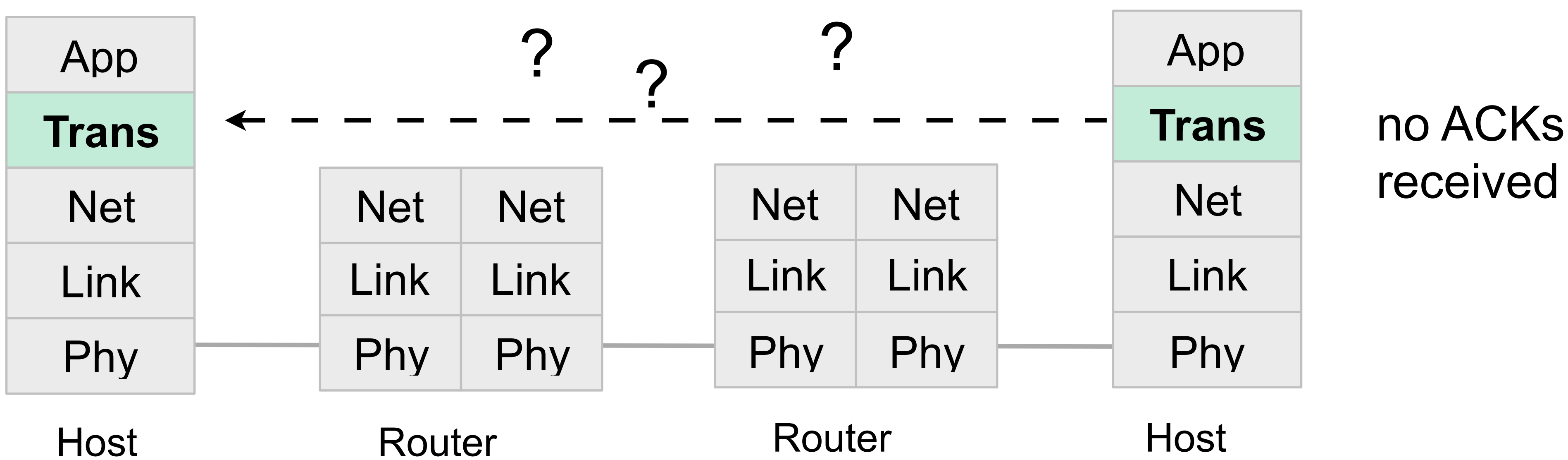
- IP Routers drop packets
- TCP has to react, e.g. lower the packet injection rate



Congestion revisited

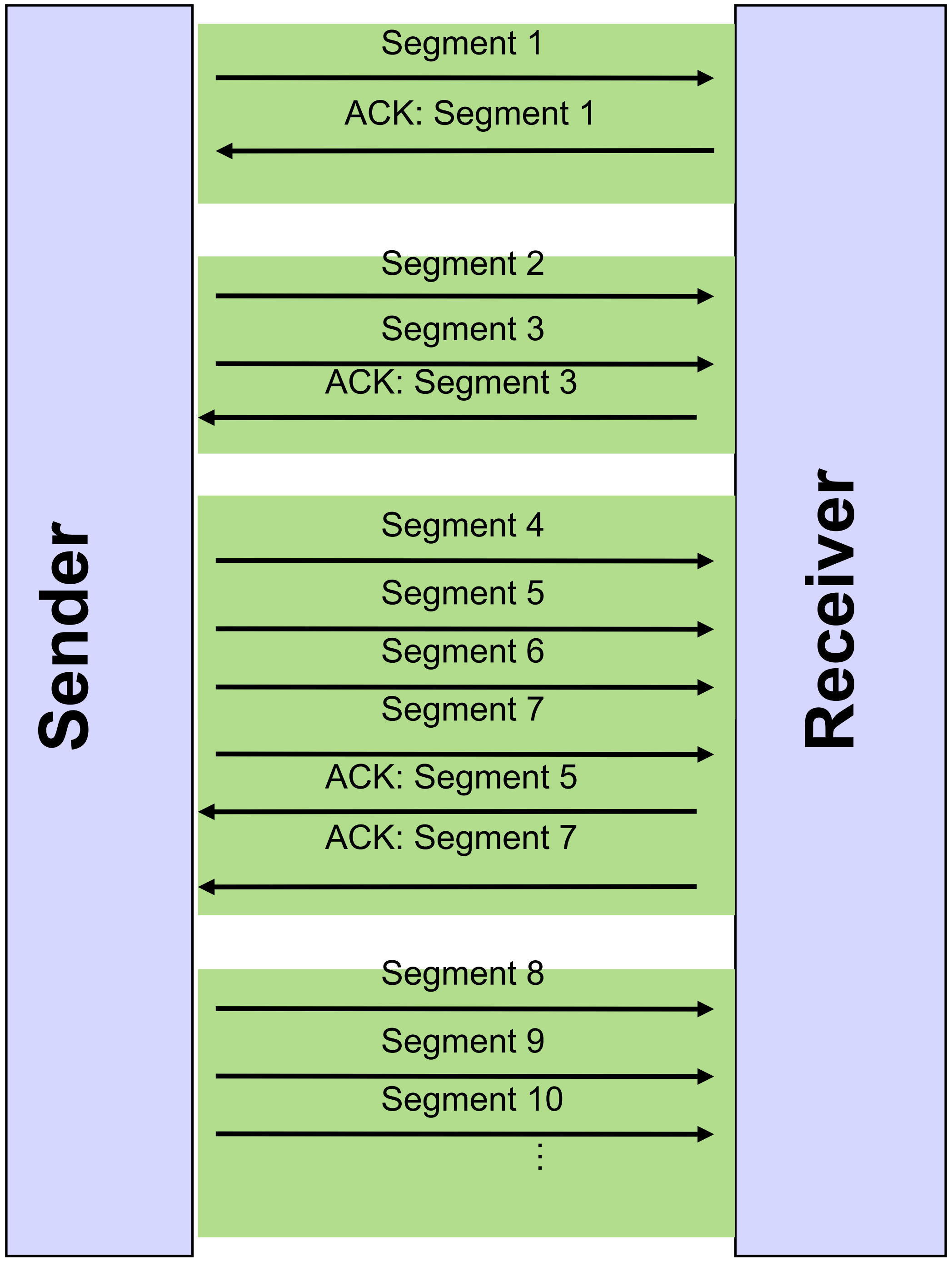


from a transport layer perspective:

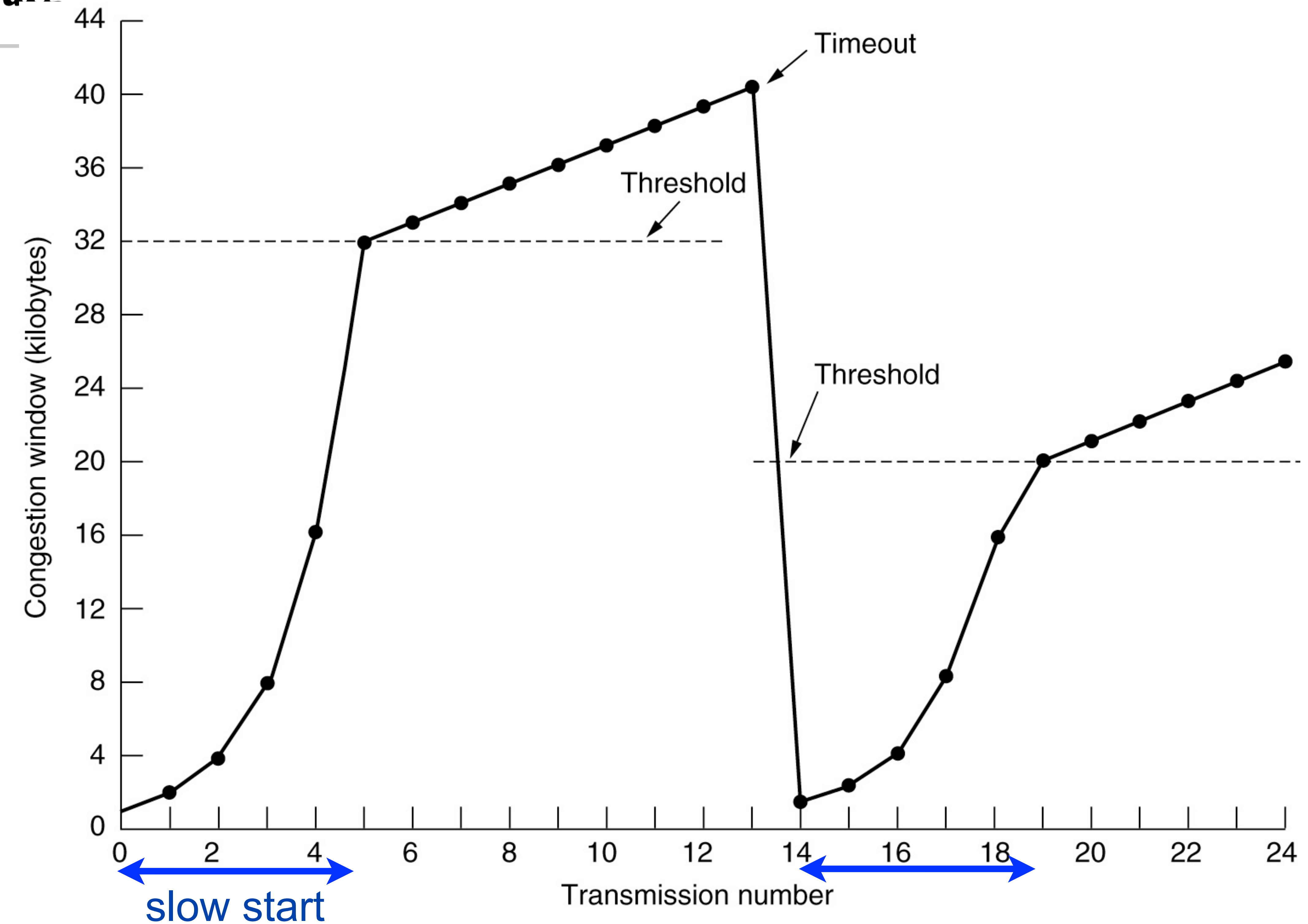


Data rate adaption and the congestion window

- Sender does not use the maximum segment size in the beginning
- Congestion window (cwnd)
  - used on the sender size
  - sending window: min {wnd,cwnd} (wnd = receiver window)
  - S: segment size
  - Initialization:
    - cwnd $\leftarrow$ S
  - For each received acknowledgement:
    - cwnd $\leftarrow$ cwnd + S
  - ...until a packet remains unacknowledged



Slow Start of TCP Tahoe



TCP Tahoe's slow start

- **TCP Tahoe, Jacobson 88:**
  - Congestion window (cwnd)
  - Slow Start Threshold (ssthresh)
  - S = maximum segment size
- **Initialization (after connection establishment):**
  - cwnd ← S ssthresh ← 65535
- **If a packet is lost (no acknowledgement within RTO):**
  - multiplicative decrease of ssthresh
    cwnd ← S ssthresh ← $\max\left\{2 \times S, \frac{\min\{cwnd, wnd\}}{2}\right\}$
- **If a segment is acknowledged and cwnd ≤ ssthresh then**
  - slow start: cwnd ← cwnd + S
- **If a segment is acknowledged and cwnd > ssthresh, then**

cwnd ← cwnd + S/cwnd

x: # Packets per RTT

$x \leftarrow 1$ | $y \leftarrow \max$

$x \leftarrow 1$ | $y \leftarrow x/2$

$x \leftarrow 2 \cdot x$, until $x = y$

$x \leftarrow x + 1$

Fast Retransmit and Fast Recovery

- TCP Tahoe [Jacobson 1988]:
  - If only one packet is lost
    - retransmit and use the rest of the window
    - Slow Start
  - Fast Retransmit
    - after three duplicate ACKs, retransmit Packet, start with Slow Start
- TCP Reno [Stevens 1994]
  - After Fast Retransmit:
    - ssthresh ← min(wnd,cwnd)/2
    - cwnd ← ssthresh + 3 S
  - Fast recovery after Fast retransmit
    - Increase window size by each single acknowledgement
    - cwnd ← cwnd + S
  - Congestion avoidance: if P+x is acknowledged:
    - cwnd ← ssthresh

y ← x/2
x ← y + 3

The AIMD principle

- TCP uses basically the following mechanism to adapt the data rate x (#packets sent per RTT):
  - Initialization:

    $\boxed{x \leftarrow 1}$

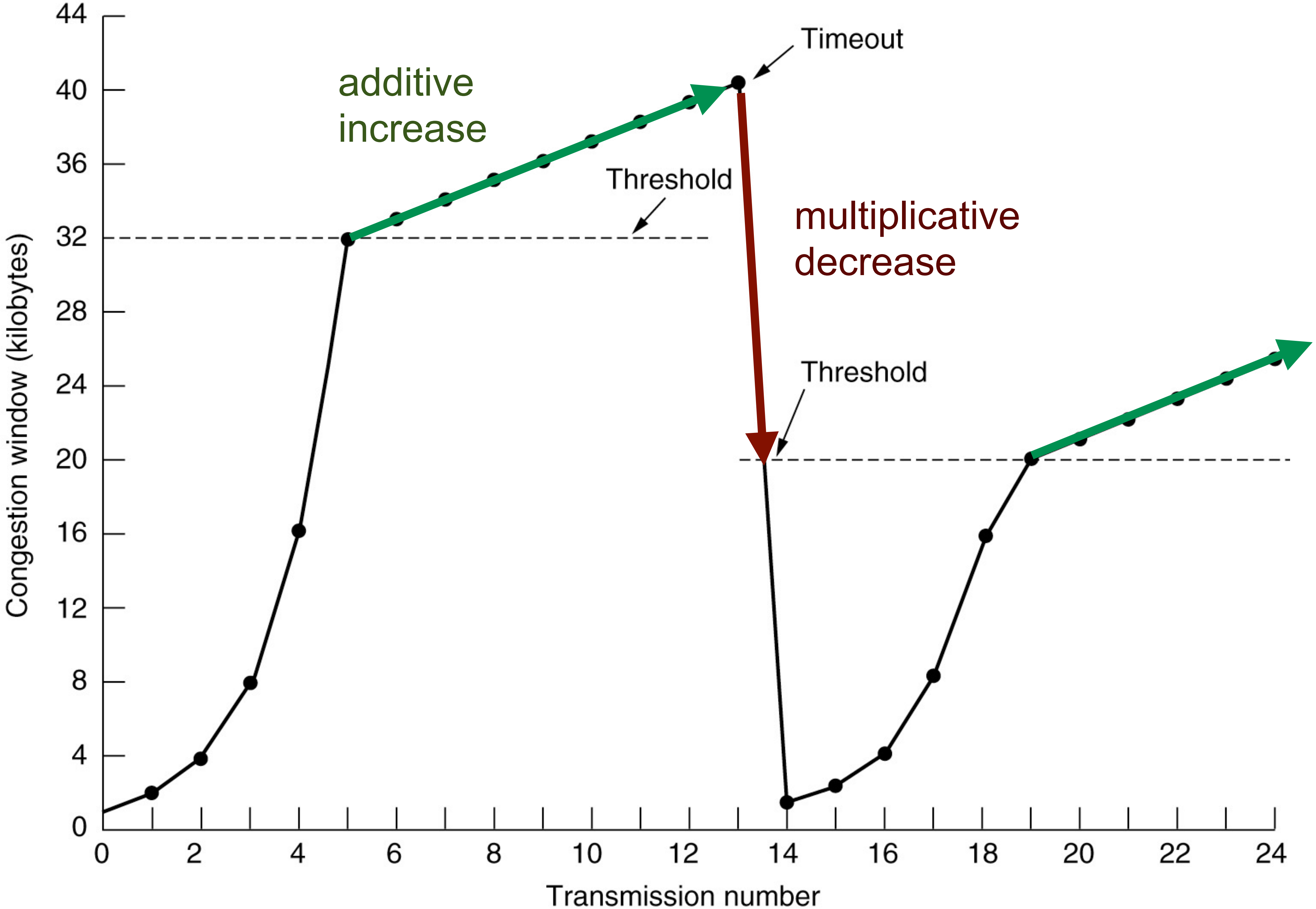
  - on packet loss: multiplicative decrease (MD)

    $\boxed{x \leftarrow x/2}$

  - if the acknowledgement for a segment arrives, perform additive increase (AI)

    $\boxed{x \leftarrow x+1}$

AIMD

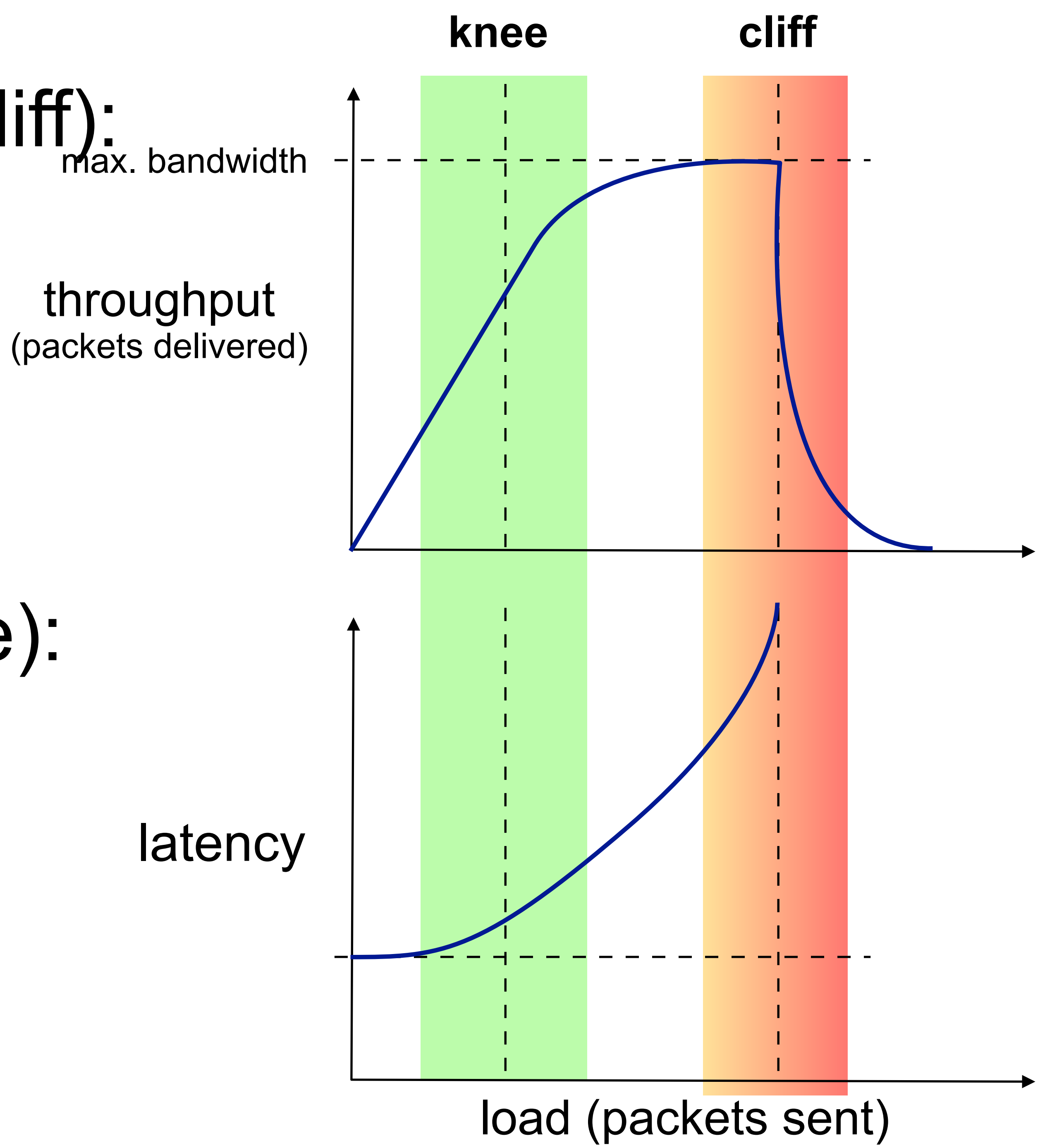


Timeout
additive
increase
Threshold
multiplicative
decrease
Threshold
Congestion window (kilobytes)
Transmission number
0
4
8
12
16
20
24
28
32
36
40
44
0
2
4
6
8
10
12
14
16
18
20
22
24

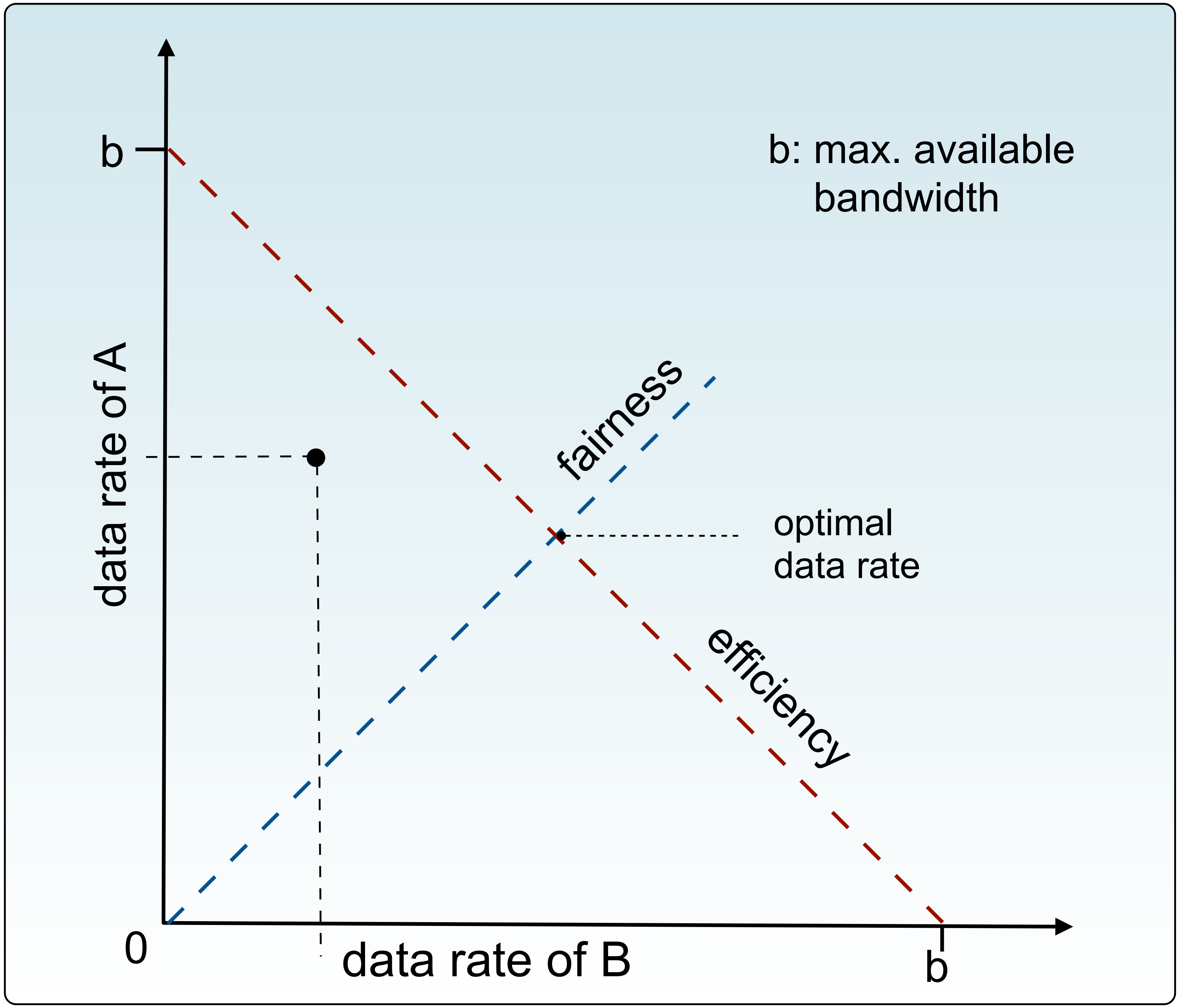
CoNe Freiburg

Throughput and Latency

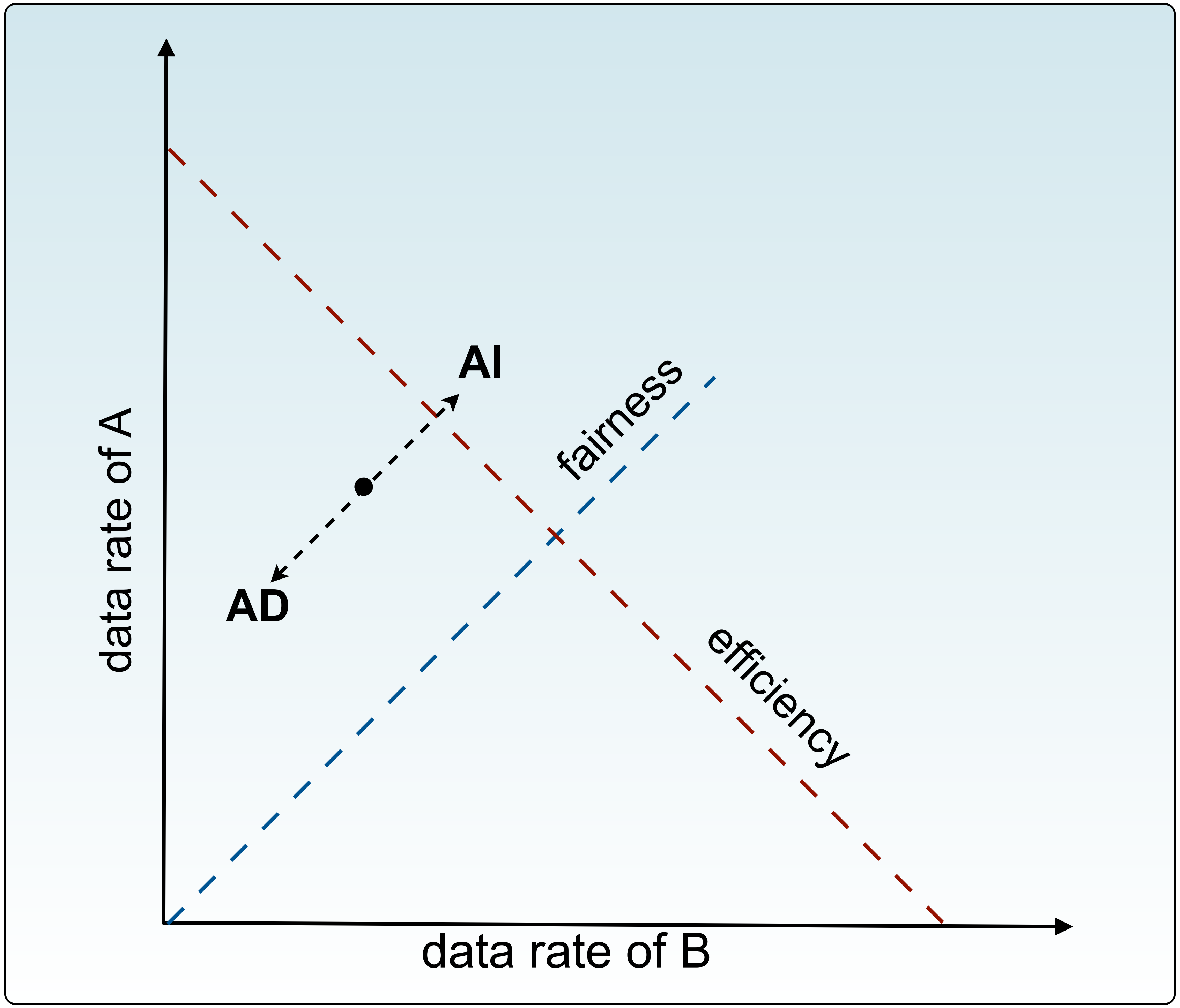
- Congested situation (cliff):
  - high load
  - low throughput
  - all data packets are lost
- Desired situation (knee):
  - high load
  - high throughput
  - few data packets get lost



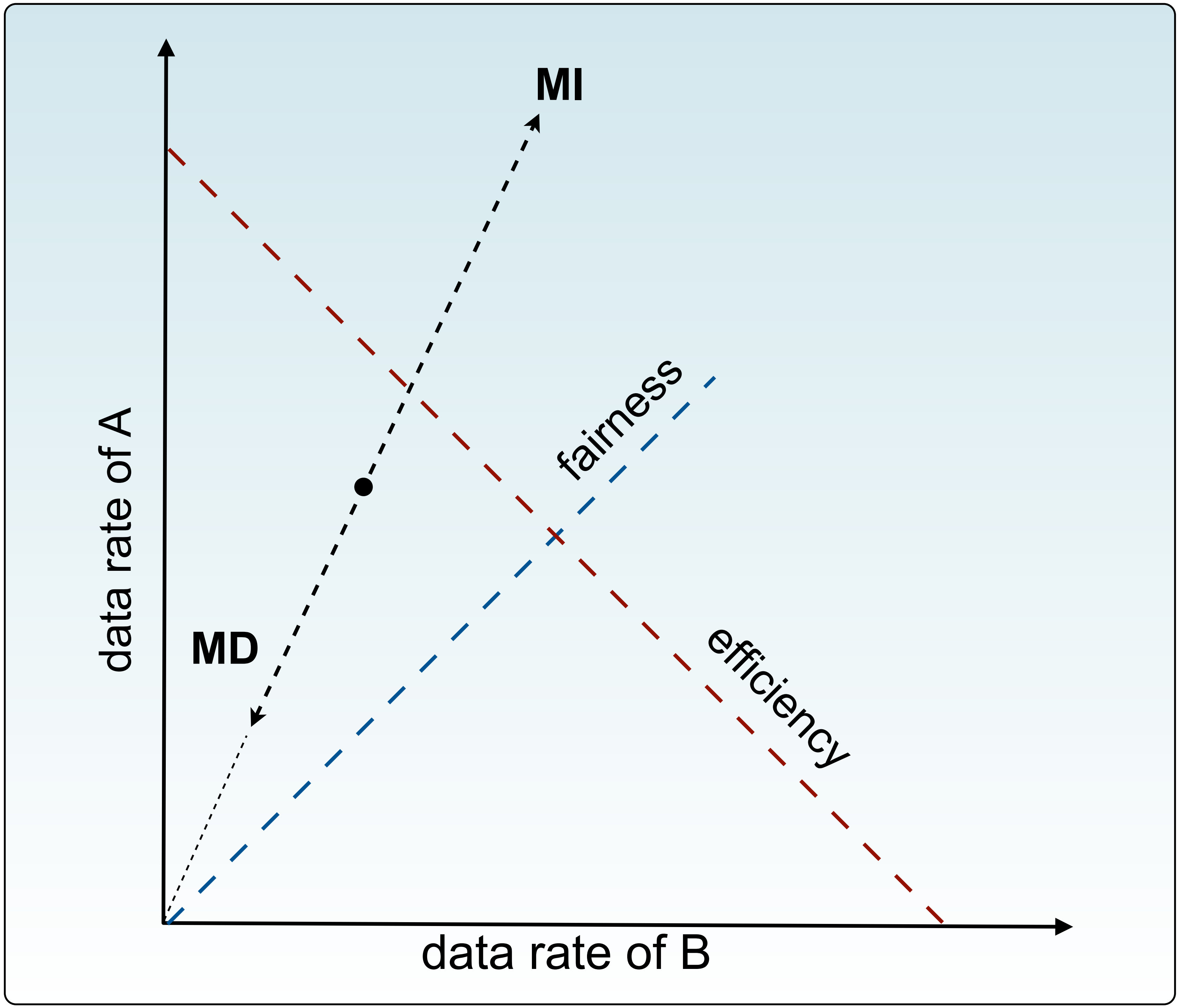
Vector diagram for 2 participants



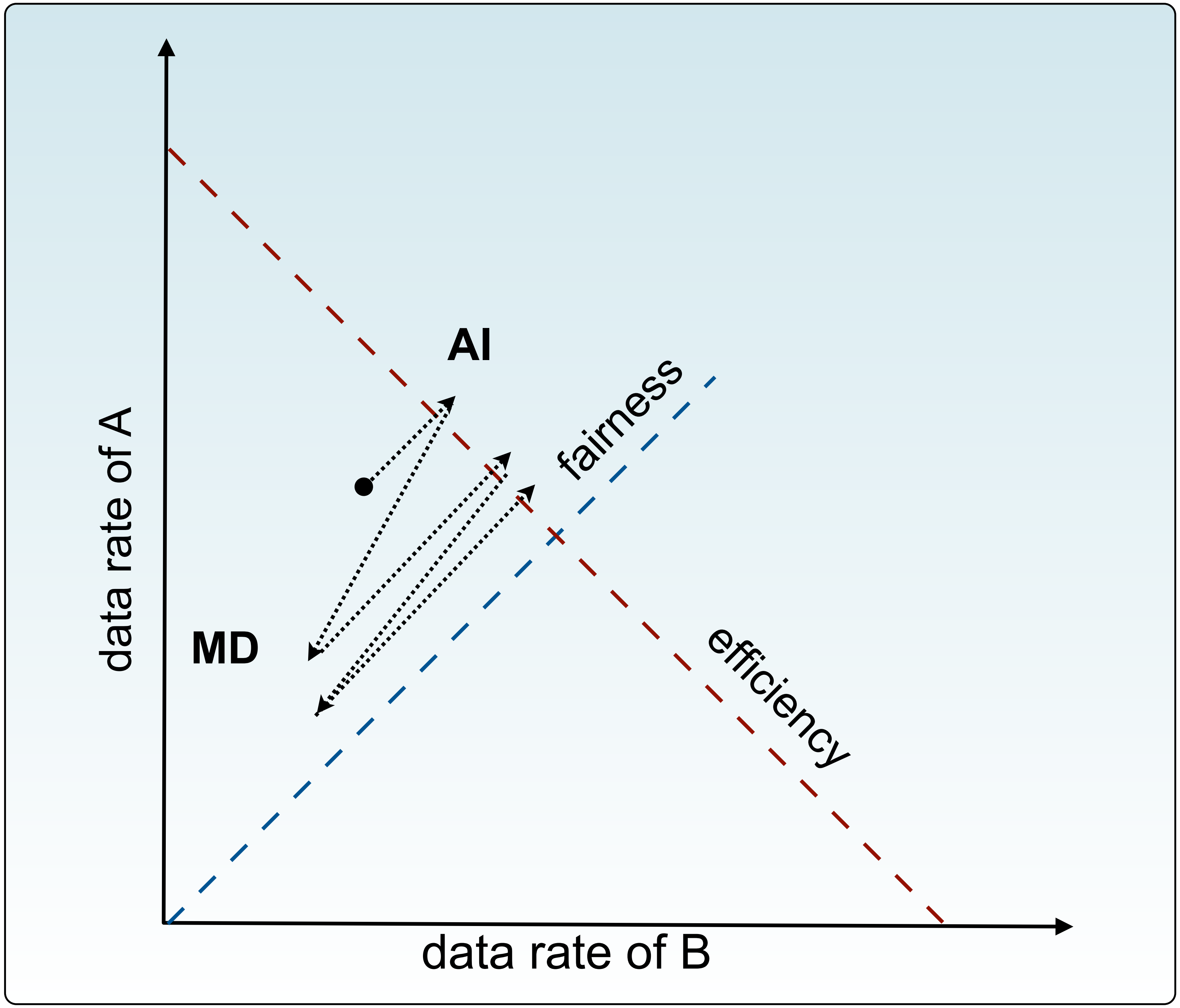
AIAD Additive Increase/ Additive Decrease



MIMD: Multiplicative Incr./ Multiplicative Decrease

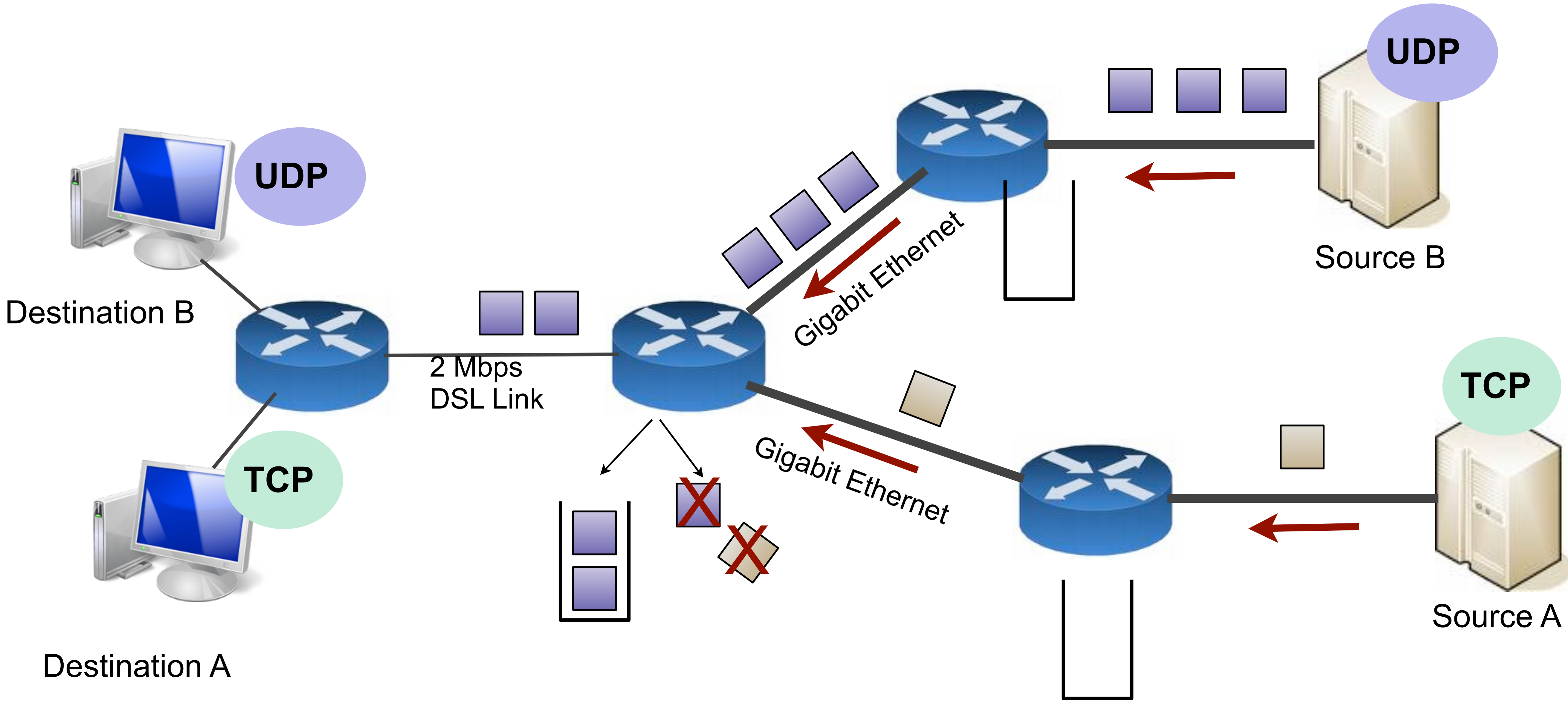


AIMD: Additively Increase/ Multiplicatively Decrease



TCP vs. UDP

- TCP reduces data rate
- UDP does not!



TCP - Conclusion

- Connection-oriented, reliable,
  in-order delivery of a byte stream
- Flow control and congestion control
  - Fairness among TCP streams
  - Unfair behavior of other protocols, e.g. UDP
  - Impact on latency
  - Tweaking the congestion avoidance mechanism has an impact on other applications

Energy Informatics
01 Internet Layers

Christian Schindelhauer
Technical Faculty
Computer-Networks and Telematics
University of Freiburg