# CS 152
# Computer Architecture and Engineering

# Lecture 24 – Networks

**2005-4-19**

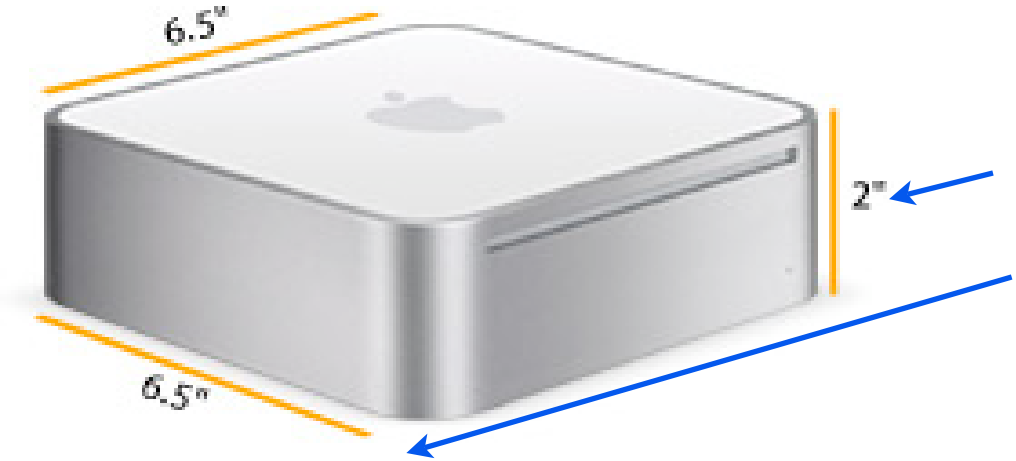
**John Lazzaro**

**(www.cs.berkeley.edu/~lazzaro)**

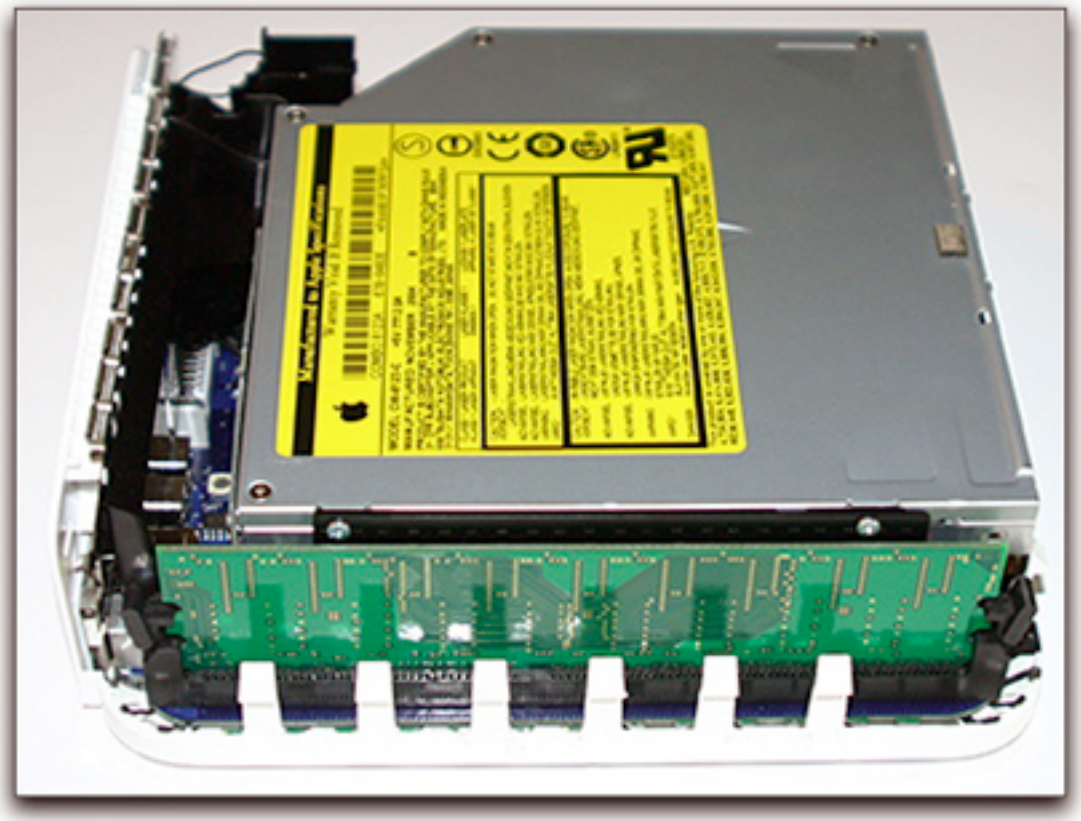
TAs: Ted Hong and David Marquardt

**www-inst.eecs.berkeley.edu/~cs152/**

# Last Time: Making Mac Mini

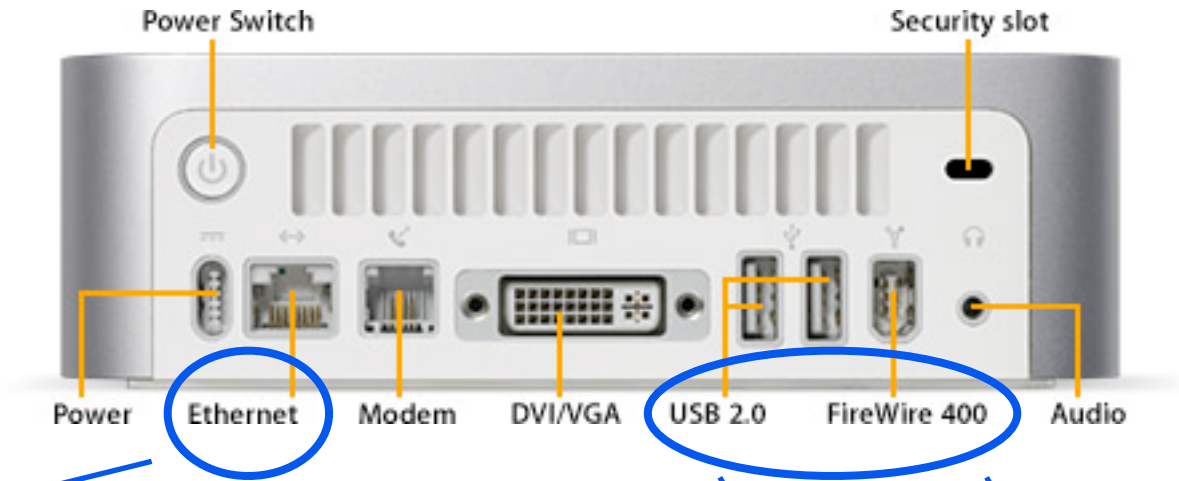



Size fixed by the "form factor" (physical size) of desktop DIMMS. Laptop DRAM is smaller, but too expensive for $499 price.



# Why are networks different from buses?

**Serial**: Data is sent "bit by bit" over one logical wire.




Network. Primary purpose is to connect computers to computers.

USB, FireWire. Primary purpose is to connect devices to a computer.

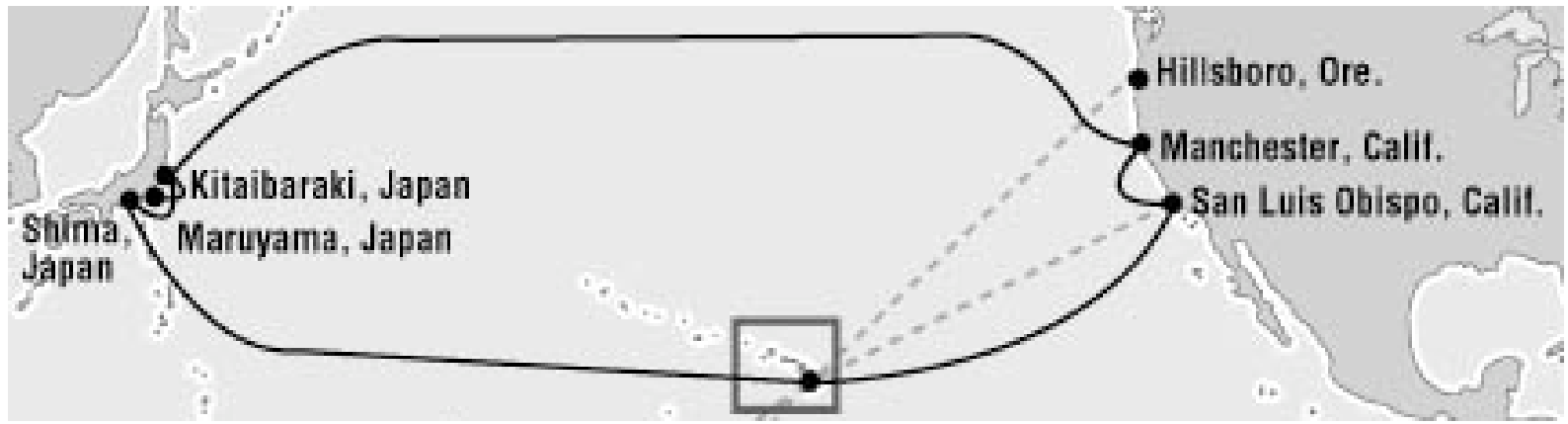
# Today: Networks

- **Link layers:** Using physics to send bits from place to place.
- **Internet:** A network of networks.
- **Routing:** Inside the cloud.

# Networking bottom-up: Link two endpoints

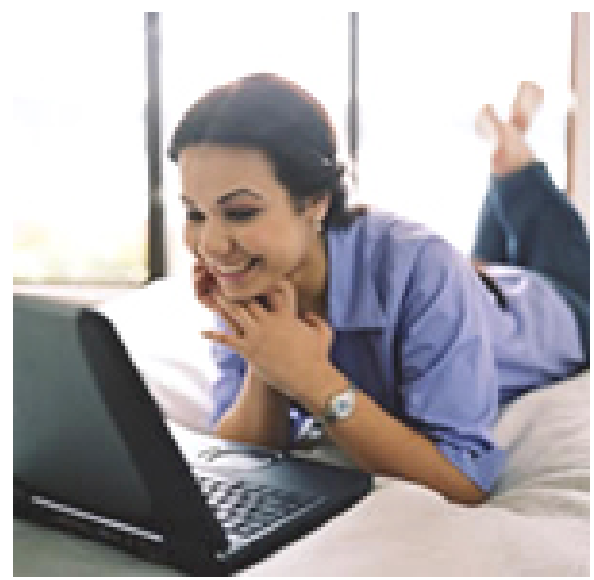
## Q1. How far away are the endpoints?

Japan-US undersea cable network




Physical media: optical fiber (photonics)

WiFi wireless from hotel bed to access point.



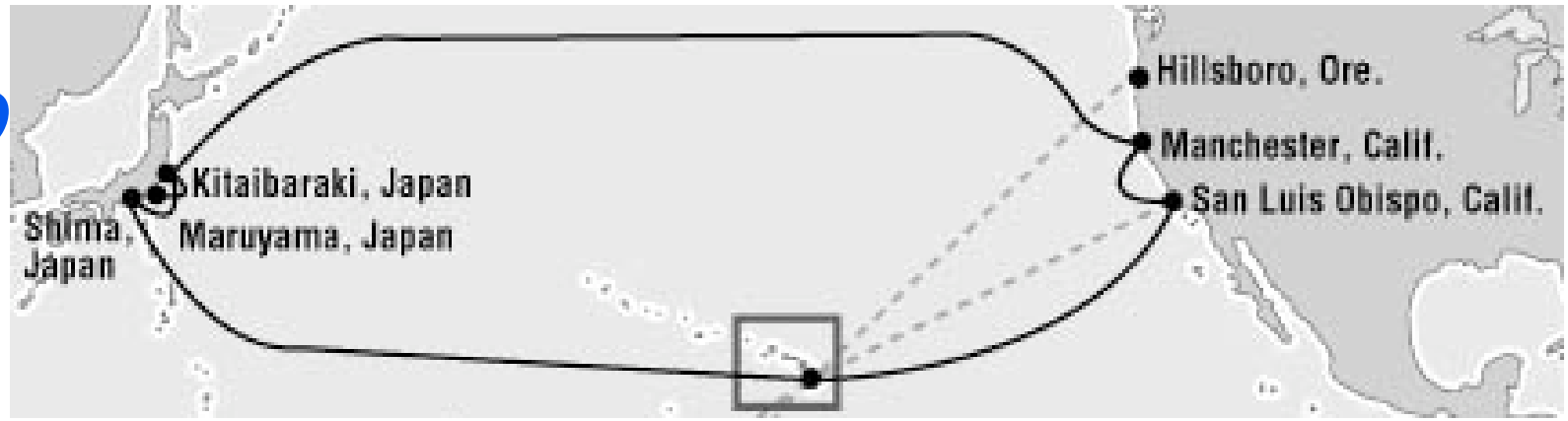
Distance + mobility + bandwidth influences choice of medium.

Physical media: unlicensed radio spectrum

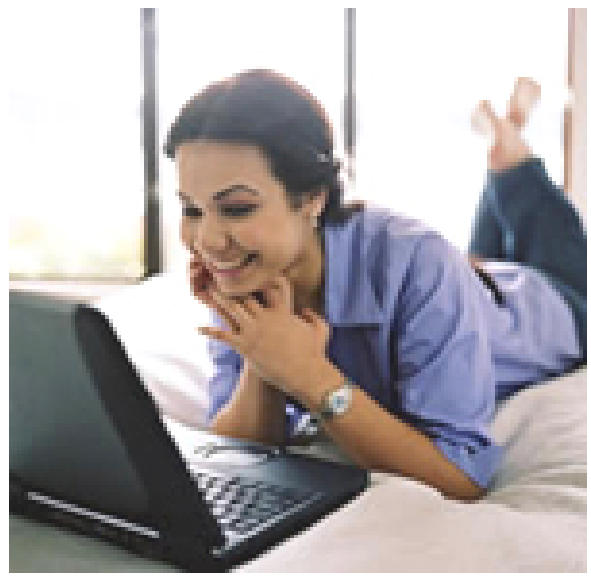
# Networking bottom-up: Link two endpoints

## Q2. Initial investment cost for the link.

$1B USD. A ship lays cable on ocean floor.




The price of the WiFi laptop card + the base station.



For expensive media, much of the "price" goes to pay off loans.

"Unlicensed radio" -- no fee to the FCC

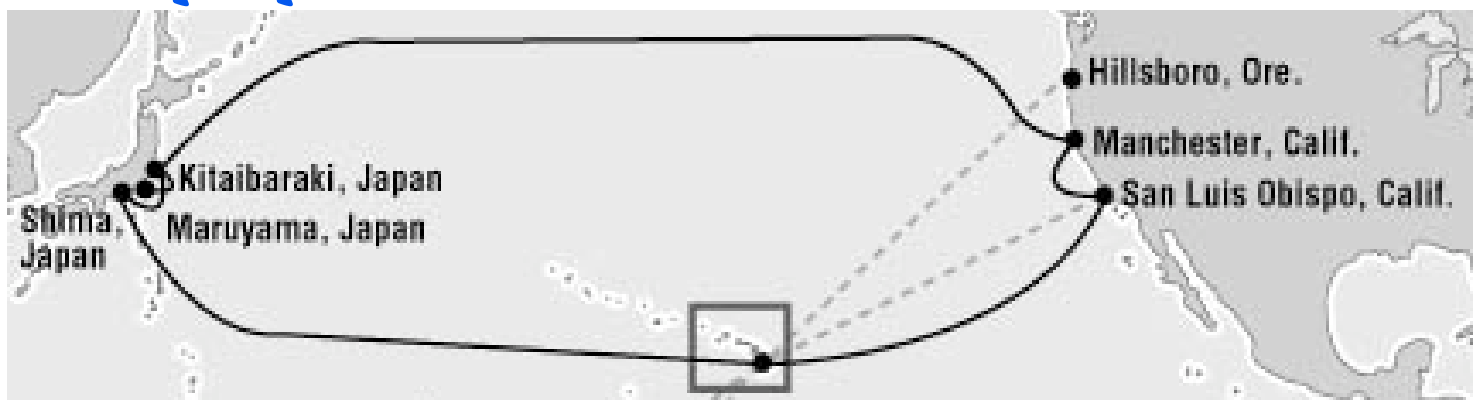
# Networking bottom-up: Link two endpoints

## Q3. How is the link imperfect?

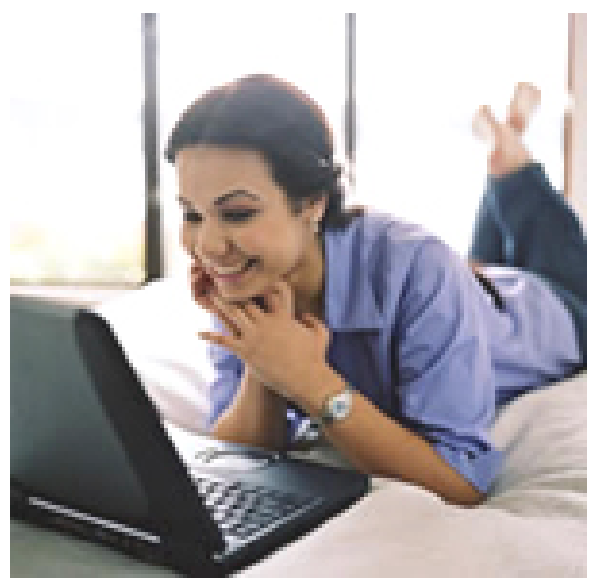
+++ A steady bitstream ("circuit"). No packets to lose.

+++ Only one bit flips per 10,000,000,000,000 sent.

--- Undersea failure is catastrophic




--- Someone walks by and the network stops working - "fading".



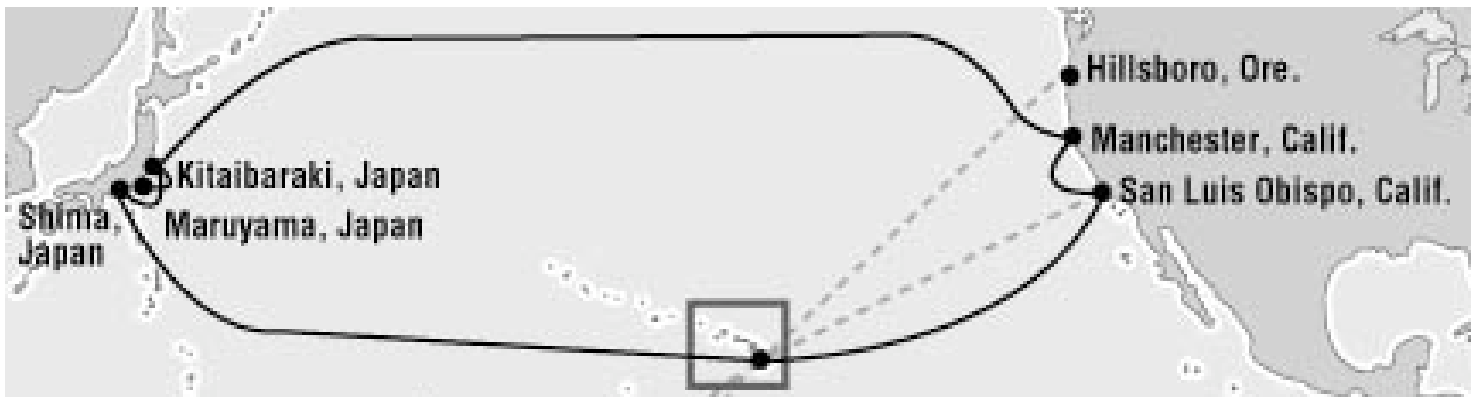
Solution: Short packets spaced in time to escape the fade. If lost, do retransmits.

# Networking bottom-up: Link two endpoints

## Q4. How does link perform?

BW: 640 Gb/s (CA-JP cable)

**Latency:** *% ping irt1-ge1-1.tdc.noc.sony.co.jp*
*PING irt1-ge1-1.tdc.noc.sony.co.jp (211.125.132.198): 56 data bytes*
*64 bytes from 211.125.132.198: icmp_seq=0 ttl=242 time=114.571 ms round-trip.*




Compare: Light speed in vacuum, SFO-Tokyo, 63ms RT.

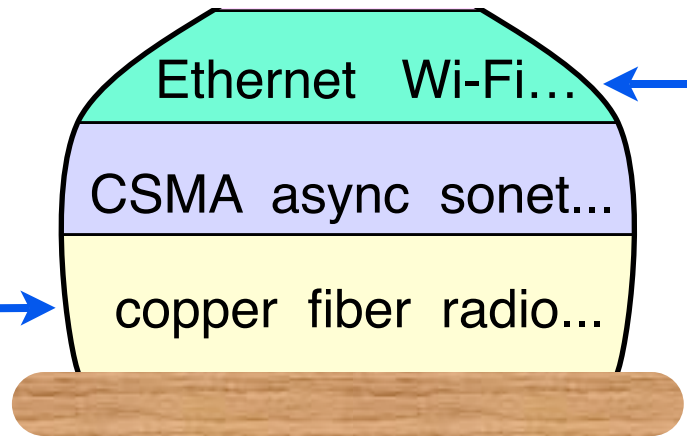
In general, risky to halve the round-trip time for one-way latency: paths are often different each direction.

BW: In theory, 802.11b offers 11 Mb/s. Users are lucky to see 3-5 Mb/s in practice.

Latency: If there is no fading, quite good. I've measured <2 ms RTT on a short hop.

# There are dozens of “link networks” ...

**Protocol Complexity**




Link networks

The undersea cable, the hotel WiFi, and many others ... DSL, Ethernet, ...

**Diagram Credit: Steve Deering**

# Web browsers do not know about link nets

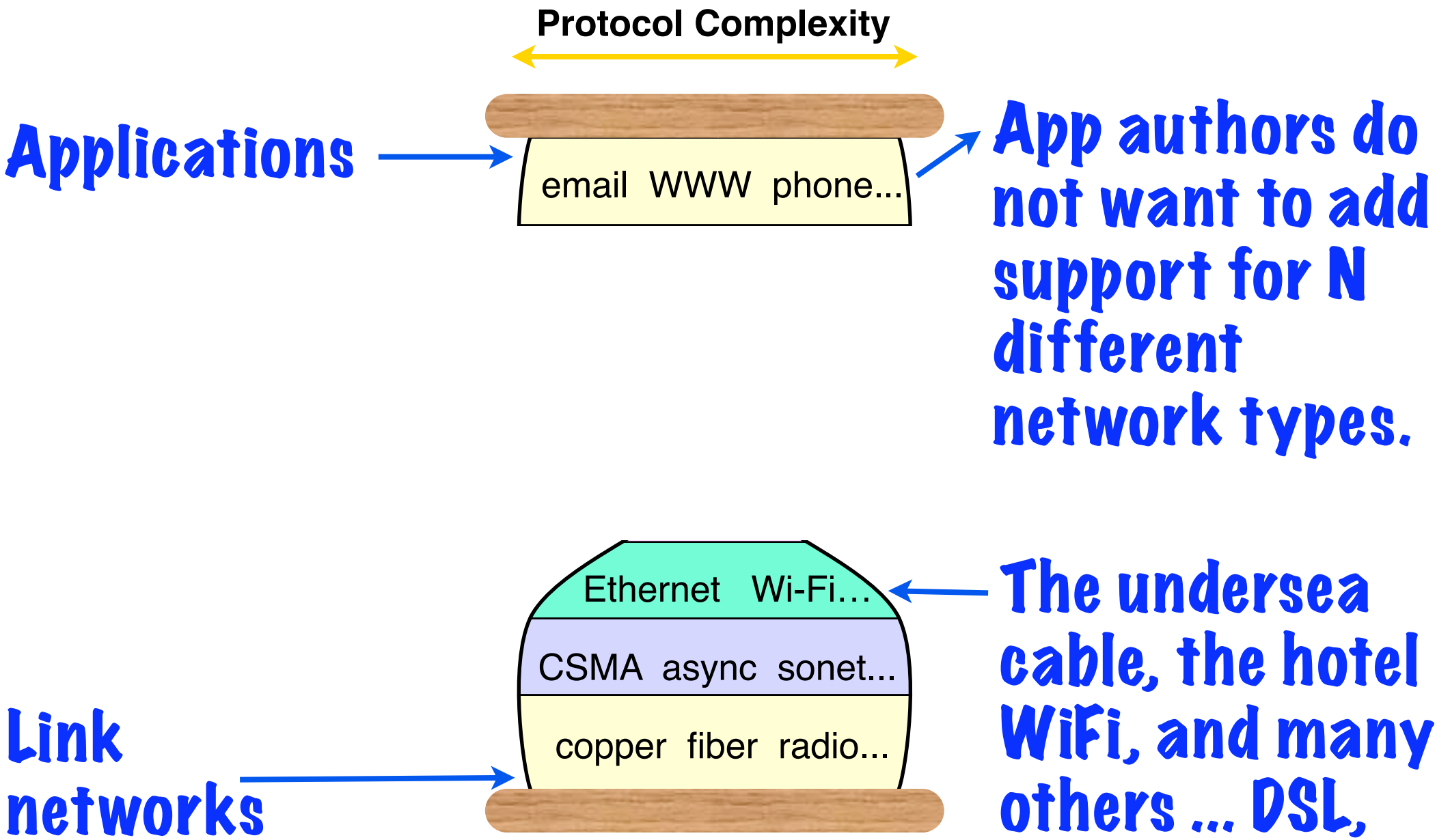
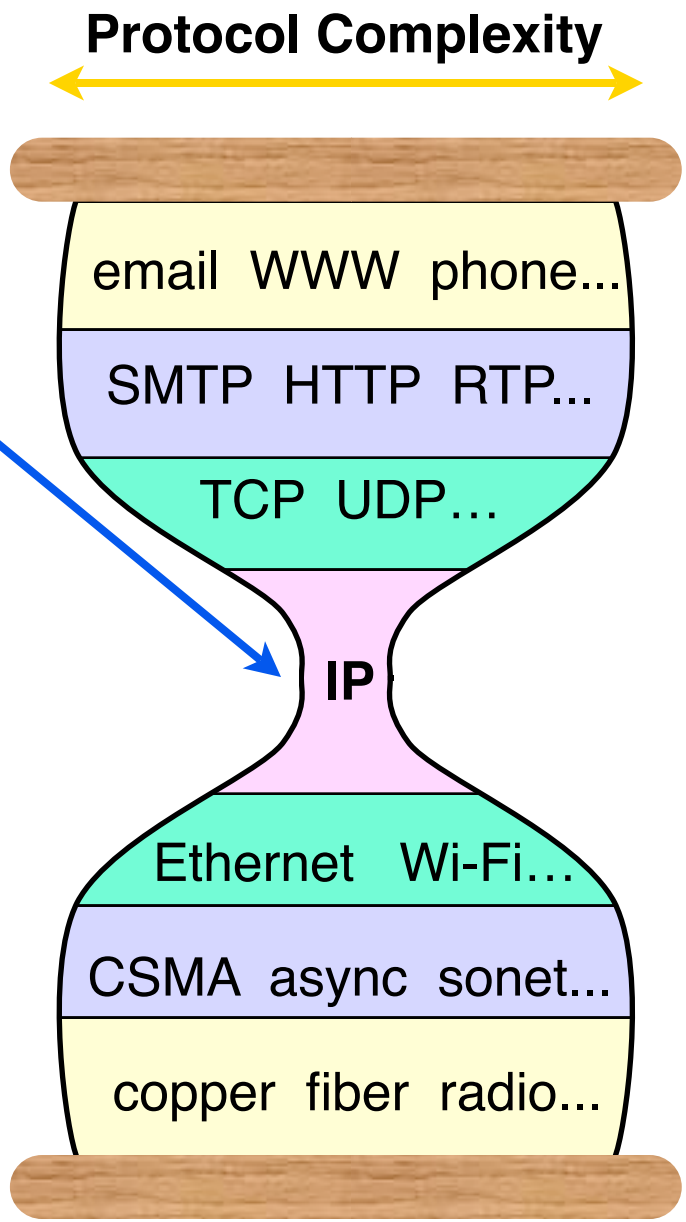



**Diagram Credit: Steve Deering**

# The Internet: A Network of Networks

**Internet Protocol (IP):** An abstraction for applications to target, and for link networks to support. Very **simple**, very successful.



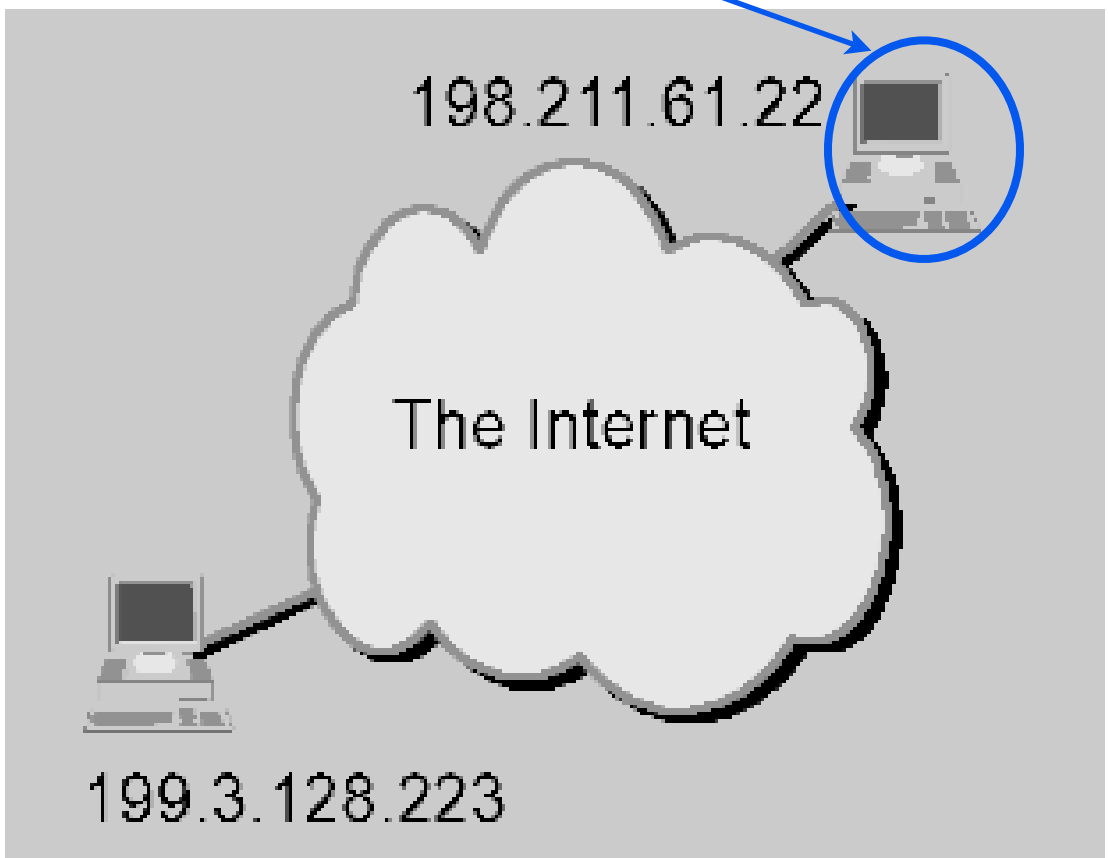

IP presents link network errors/losses in an abstract way (not a link specific way).

Link layer is not expected to be perfect.

**Diagram Credit: Steve Deering**

# The Internet interconnects "hosts" ...

**IP4 number for this computer:** 198.211.61.22




Every directly connected host has a unique IP number.

Upper limit of 2^32 IP4 numbers (some are reserved for other purposes).

Next-generation IP (IP6) limit: 2^128.

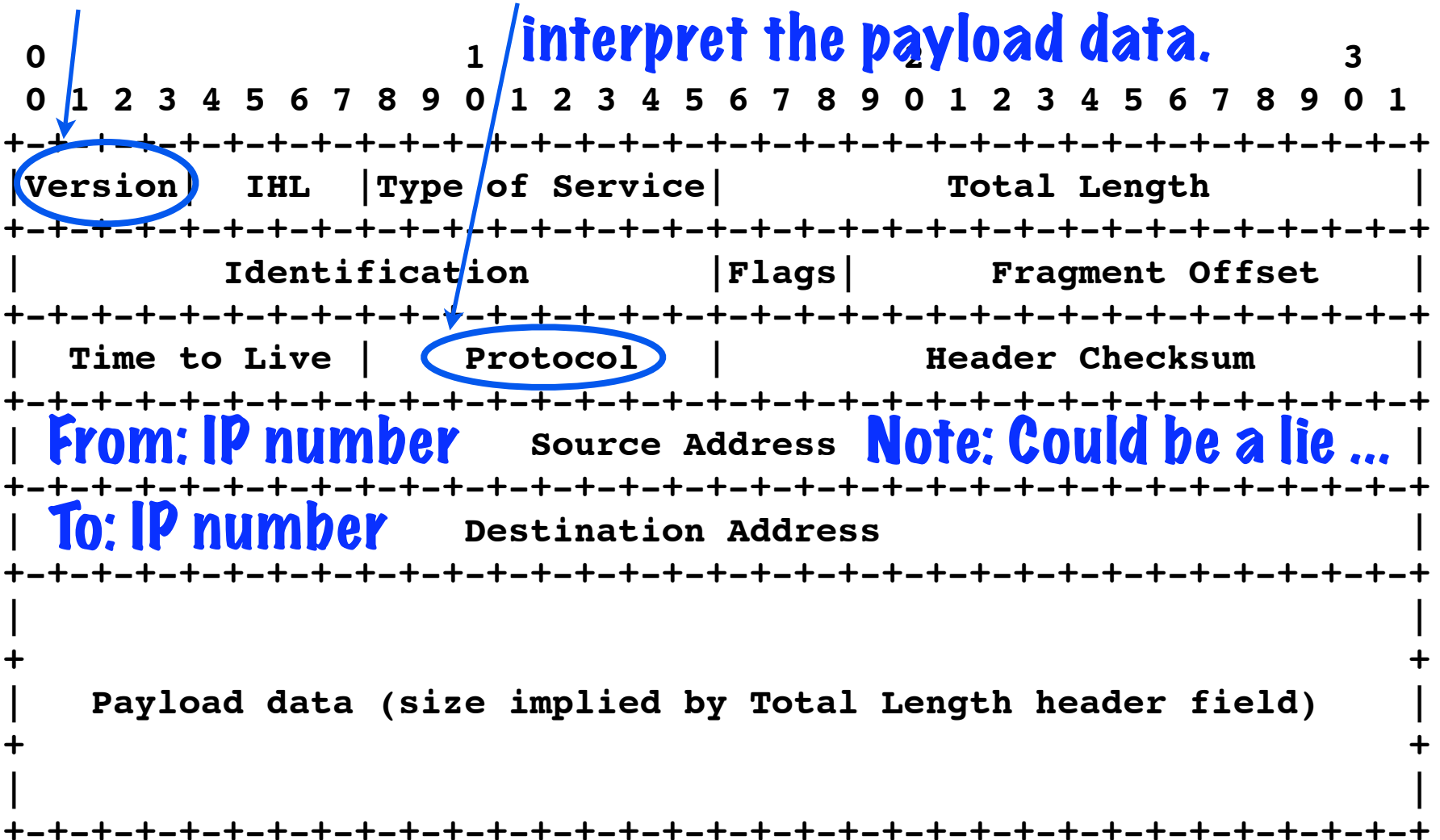
198.211.61.22 ??? A user-friendly form of the 32-bit unsigned value 3335732502, which is:

198*2^24 + 211*2^16 + 61*2^8 + 22

# Internet: Sends Packets Between Hosts

IP4, IP6, etc ...

How the destination should interpret the payload data.




From: IP number

Note: Could be a lie ...

To: IP number

Bitfield numbers

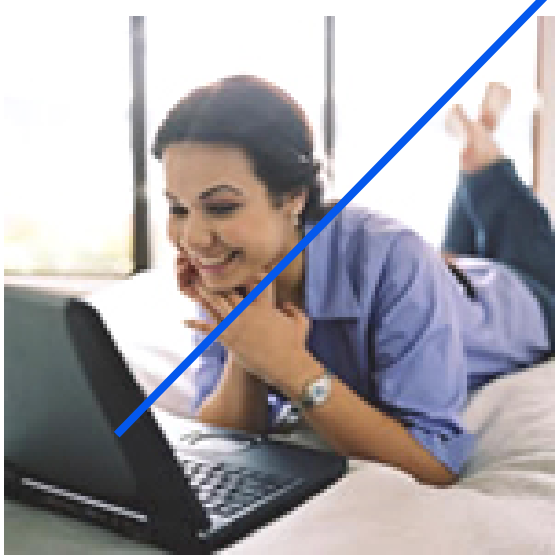
Header

Data

IHL field: # of words in header. The typical header (IHL = 5 words) is shown. Longer headers code add extra fields after the destination address.

# Link networks transport IP packets

ISO Layer Names:
- IP packet: "Layer 3"
- WiFi and Cable Modem packets: "Layer 2"
- Radio/cable waveforms: "Layer 1"



For this "hop", IP packet sent "inside" of a wireless 802.11b packet.

801.11b WiFi packet

IP Packet

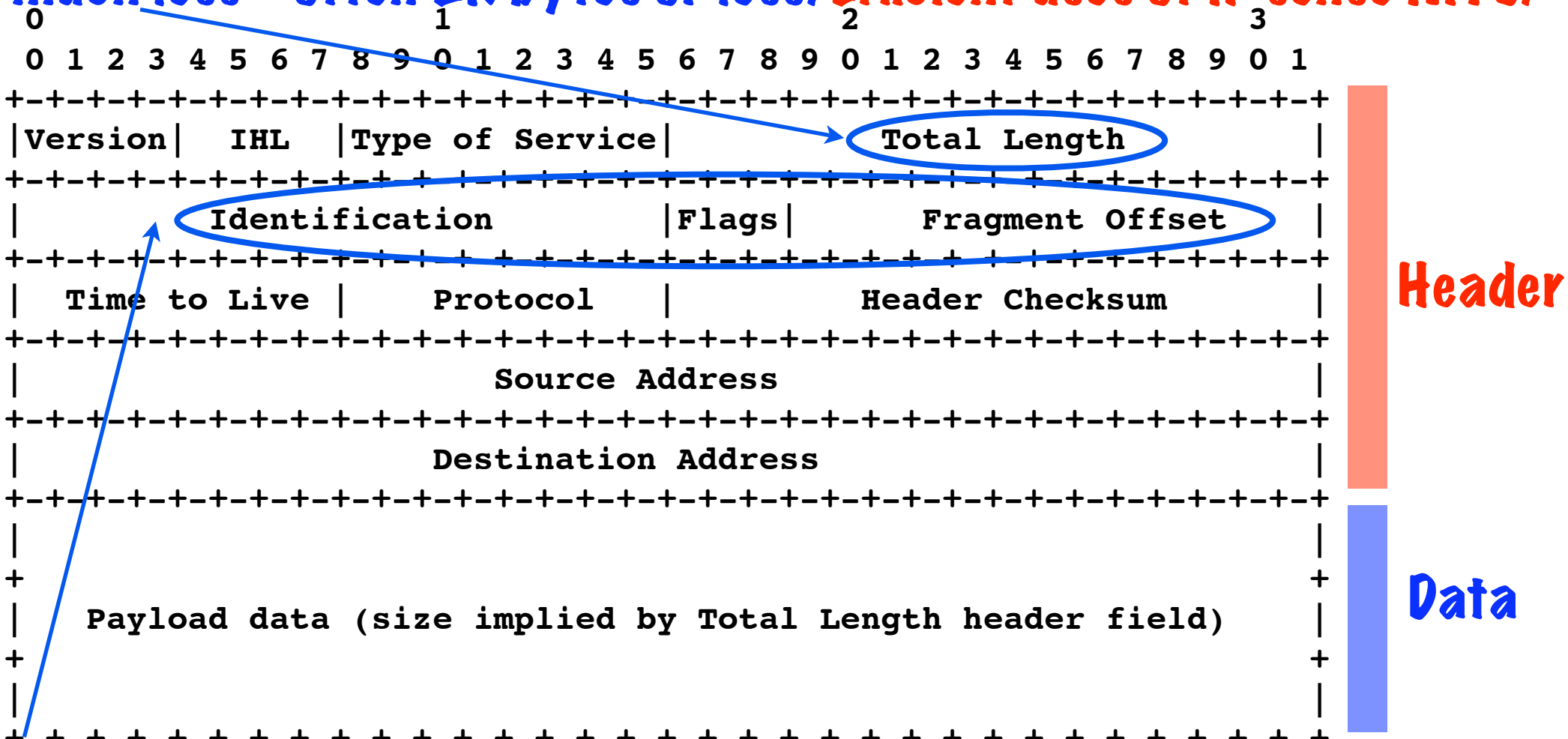

Cable modem packet

IP Packet

For this "hop", IP packet sent "inside" of a cable modem DOCSIS packet.

# Link layers "maximum packet size" vary.

Maximum IP packet size 64K bytes. Maximum Transmission Unit (MTU -- generalized "packet size") of link networks may be much less - often 2K bytes or less. Efficient uses of IP sense MTU.




Fragment fields: Link layer splits up big IP packets into many link-layer packets, reassembles IP packet on arrival.

# IP abstraction of non-ideal link networks:

- A sent packet may never arrive ("lost")
- If packets sent P1/P2/P3, they may arrive P2/P1/P3 ("out of order").

**Best Effort:** The link networks, and other parts of the "cloud", do their best to meet the ideal. But, no promises.

- Relative timing of packet stream not necessarily preserved ("late" packets).
- IP payload bits received may not match payload bits sent. IP header protected by checksum (almost always correct).

# How do apps deal with this abstraction?

"Computing" apps use the TCP (Transmission Control Protocol).

TCP lets host A send a reliable byte stream to host B. TCP works by retransmitting lost IP packets. Timing is uncertain.

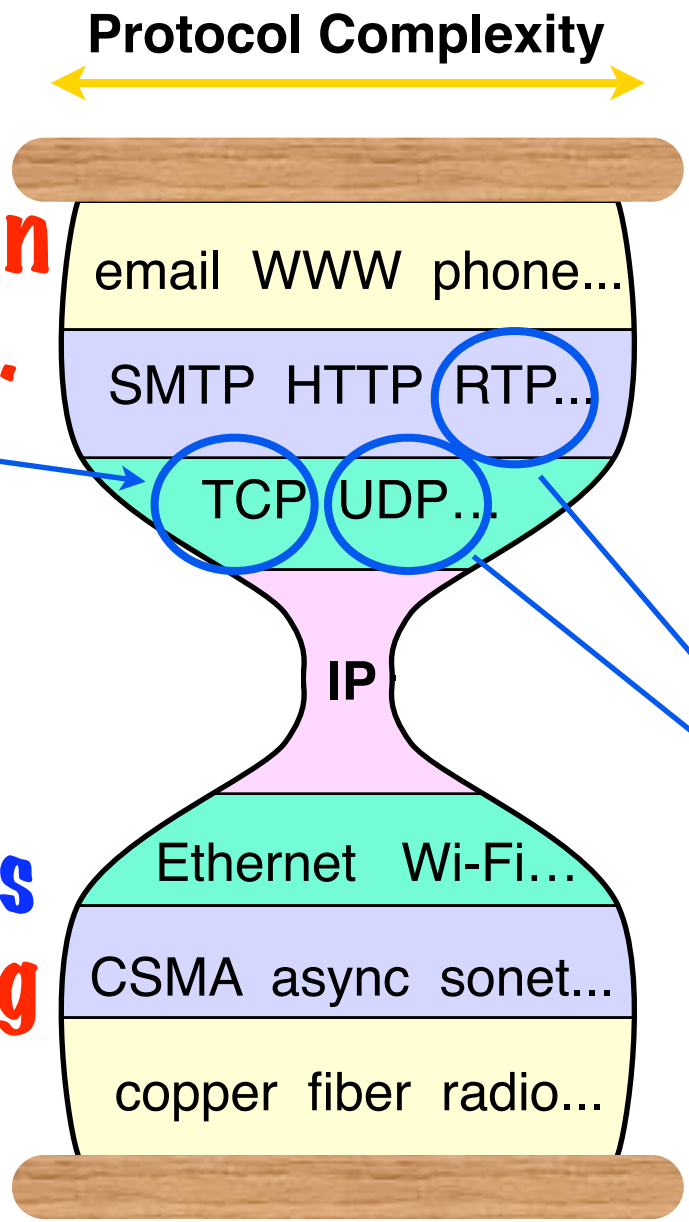


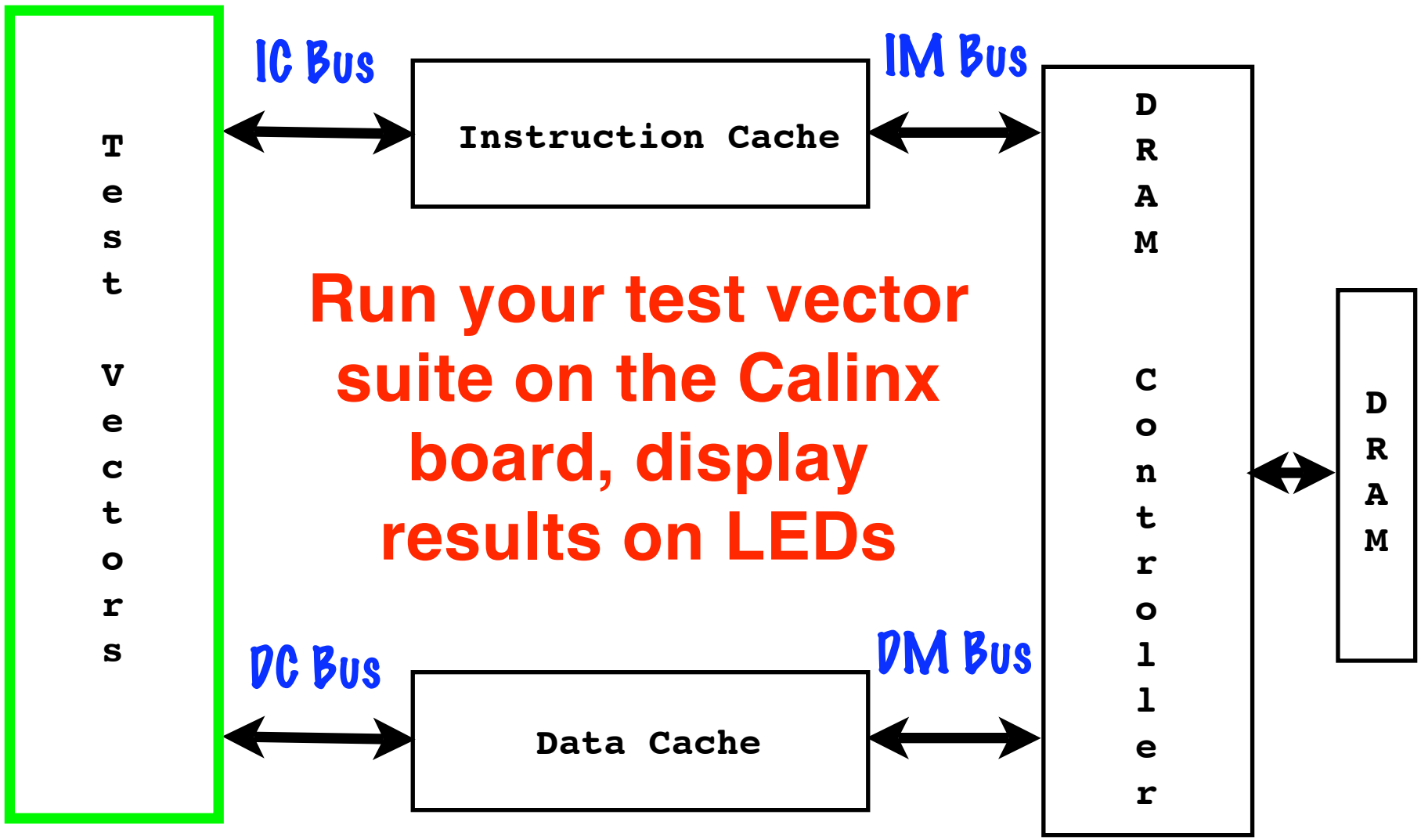

Diagram Credit: Steve Deering

Retransmission is bad for IP telephony: resent packets arrive too late.

IP telephony uses packets, not TCP. Parity codes, audio tricks used for lost packets.

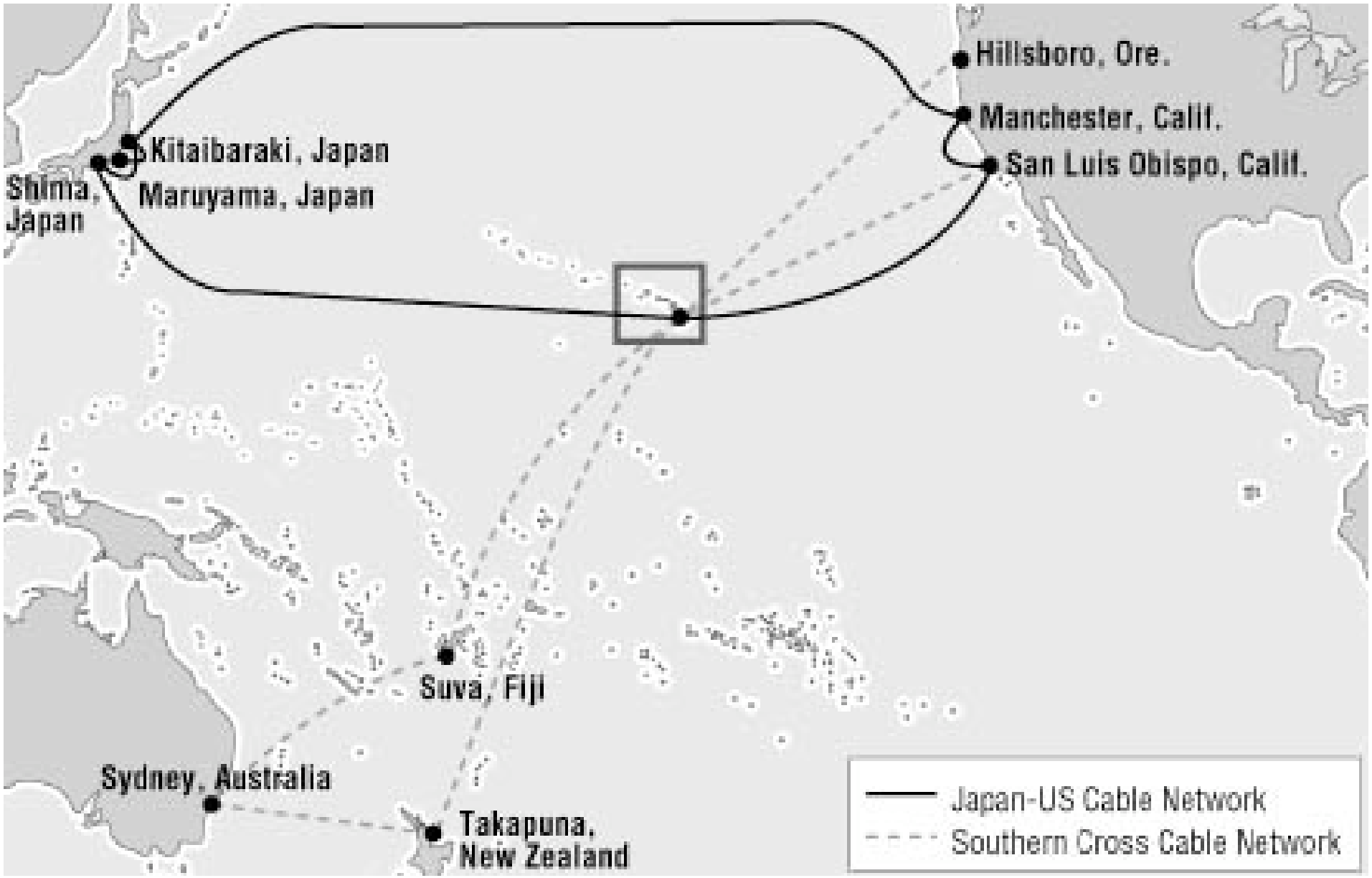
# This Friday: Memory System Checkoff

| | | | | |
|---|---|---|---|---|
| F<br>4/22 | | | | Final Project: Memory System Xilinx Checkoff<br>Midterm Review Homework Due in Section |

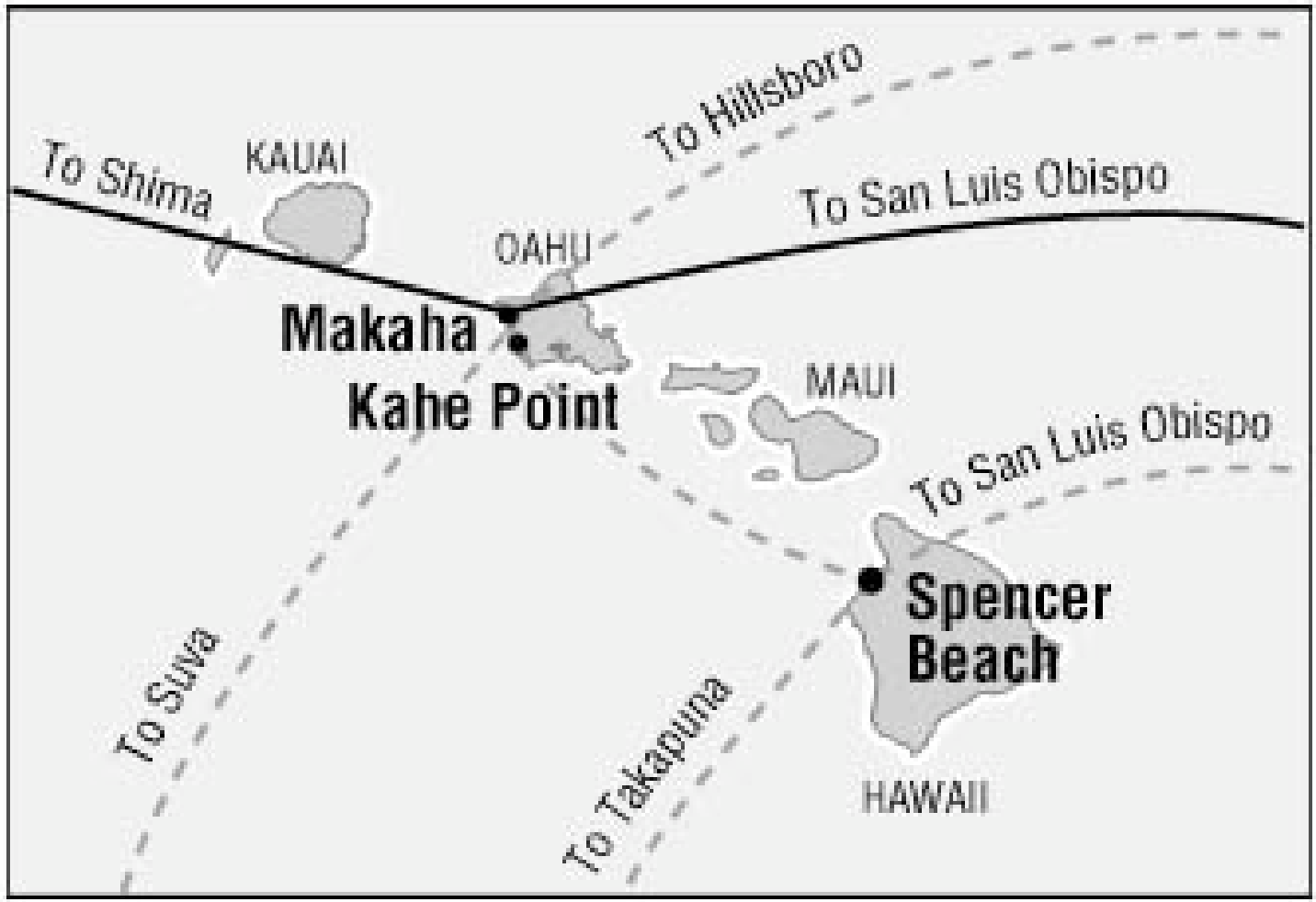


# Routing


# Undersea cables meet in Hawaii ...



# Routers: Like a hub airport




In Makaha, a router takes each Layer 2 packet off the San Luis Obispo (CA) cable, examines the IP packet destination field, and forwards to Japan cable, Fiji cable, or to Kahe Point (and onto big island cables).

# Example: berkeley.edu to sony.co.jp

## Passes through 21 routers ...

Leaving Cal ...

Getting to LA ...

Cross Pacific

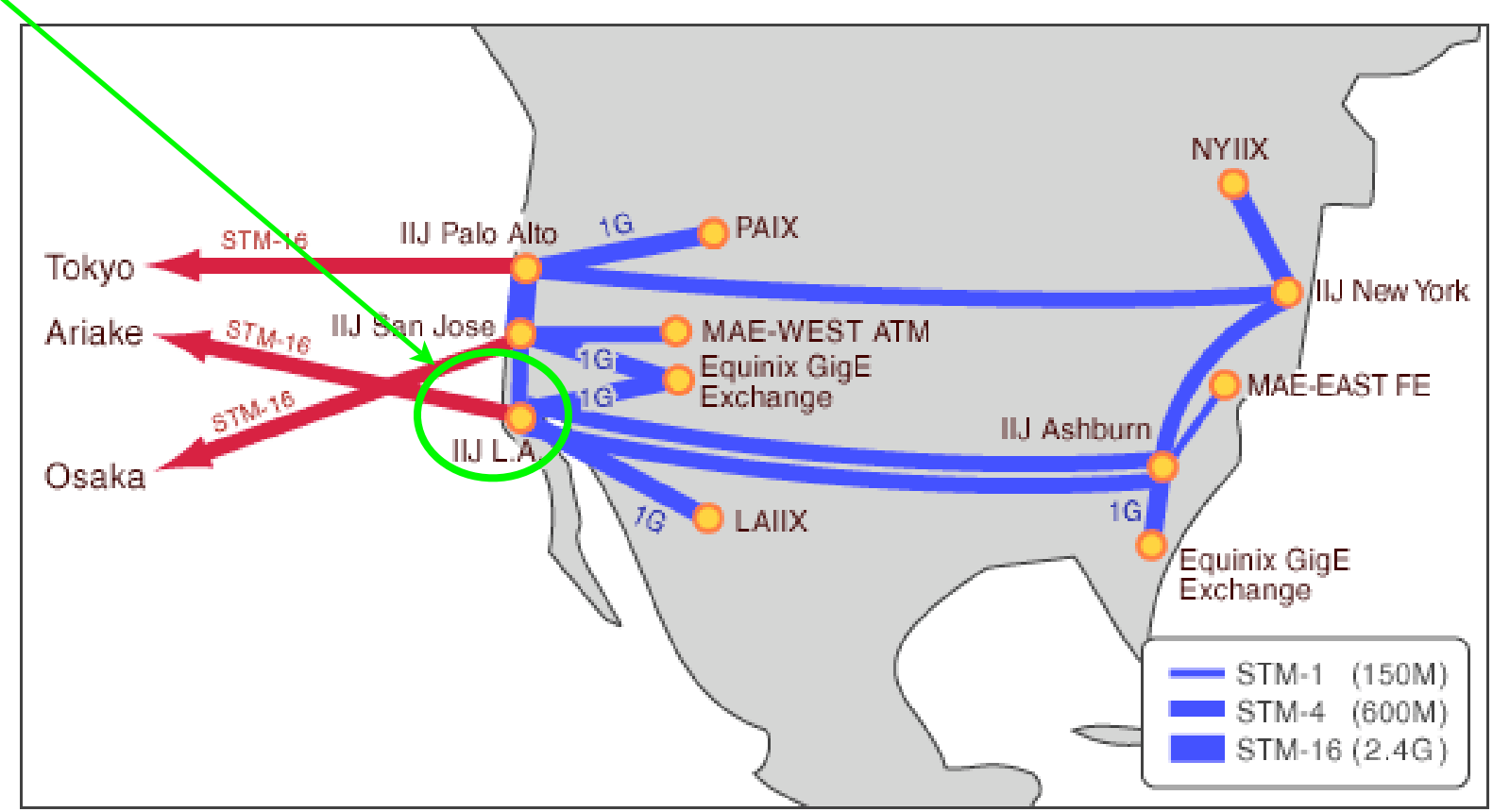
Getting to Sony

```
% traceroute irt1-ge1-1.tdc.noc.sony.co.jp
traceroute to irt1-ge1-1.tdc.noc.sony.co.jp (211.125.132.198), 30 hops max, 40
 1  soda3a-gw.eecs.berkeley.edu (128.32.34.1)  20.581 ms  0.875 ms  1.381 ms
 2  soda-cr-1-1-soda-br-6-2.eecs.berkeley.edu (169.229.59.225)  1.354 ms  3.097
 3  vlan242.inr-202-doecev.berkeley.edu (128.32.255.169)  1.753 ms  1.454 ms  1
 4  ge-1-3-0.inr-001-eva.berkeley.edu (128.32.0.34)  1.746 ms  1.174 ms  2.22 m
 5  svl-dc1--ucb-egm.cenic.net (137.164.23.65)  2.653 ms  2.72 ms  12.031 ms
 6  dc-svl-dc2--svl-dc1-df-iconn-2.cenic.net (137.164.22.209)  2.478 ms  2.451
 7  dc-sol-dc1--svl-dc1-pos.cenic.net (137.164.22.28)  4.509 ms  95.013 ms  7.7
 8  dc-sol-dc2--sol-dc1-df-iconn-1.cenic.net (137.164.22.211)  18.319 ms  4.324
 9  dc-slo-dc1--sol-dc2-pos.cenic.net (137.164.22.26)  19.403 ms  10.077 ms  13
10  dc-slo-dc2--dc1-df-iconn-1.cenic.net (137.164.22.123)  8.049 ms  20.653 ms
11  dc-lax-dc1--slo-dc2-pos.cenic.net (137.164.22.24)  94.579 ms  14.52 ms  21.
12  rtrisi.ultradns.net (198.32.146.38)  25.48 ms  12.432 ms  17.837 ms
13  lax001bb00.iij.net (216.98.96.176)  11.623 ms  25.698 ms  11.382 ms
14  tky002bb01.iij.net (216.98.96.178)  168.082 ms  196.26 ms  121.914 ms
15  tky002bb00.iij.net (202.232.0.149)  144.592 ms  208.622 ms  121.801 ms
16  tky001bb01.iij.net (202.232.0.70)  153.757 ms  110.29 ms  184.985 ms
17  tky001ip30.iij.net (210.130.130.100)  114.234 ms  110.095 ms  169.692 ms
18  210.138.131.198 (210.138.131.198)  113.893 ms  113.665 ms  114.22 ms
19  ert1-ge000.tdc.noc.ssd.ad.jp (211.125.132.69)  114.758 ms  138.327 ms  113.
20  211.125.133.86 (211.125.133.86)  113.956 ms  113.73 ms  113.965 ms
21  irt1-ge1-1.tdc.noc.sony.co.jp (211.125.132.198)  145.247 ms *  136.884 ms
```

## Cross ocean in 1 hop - link about 175 ms round-trip

# Left on Internet Initiative Japan (IIJ) in LA

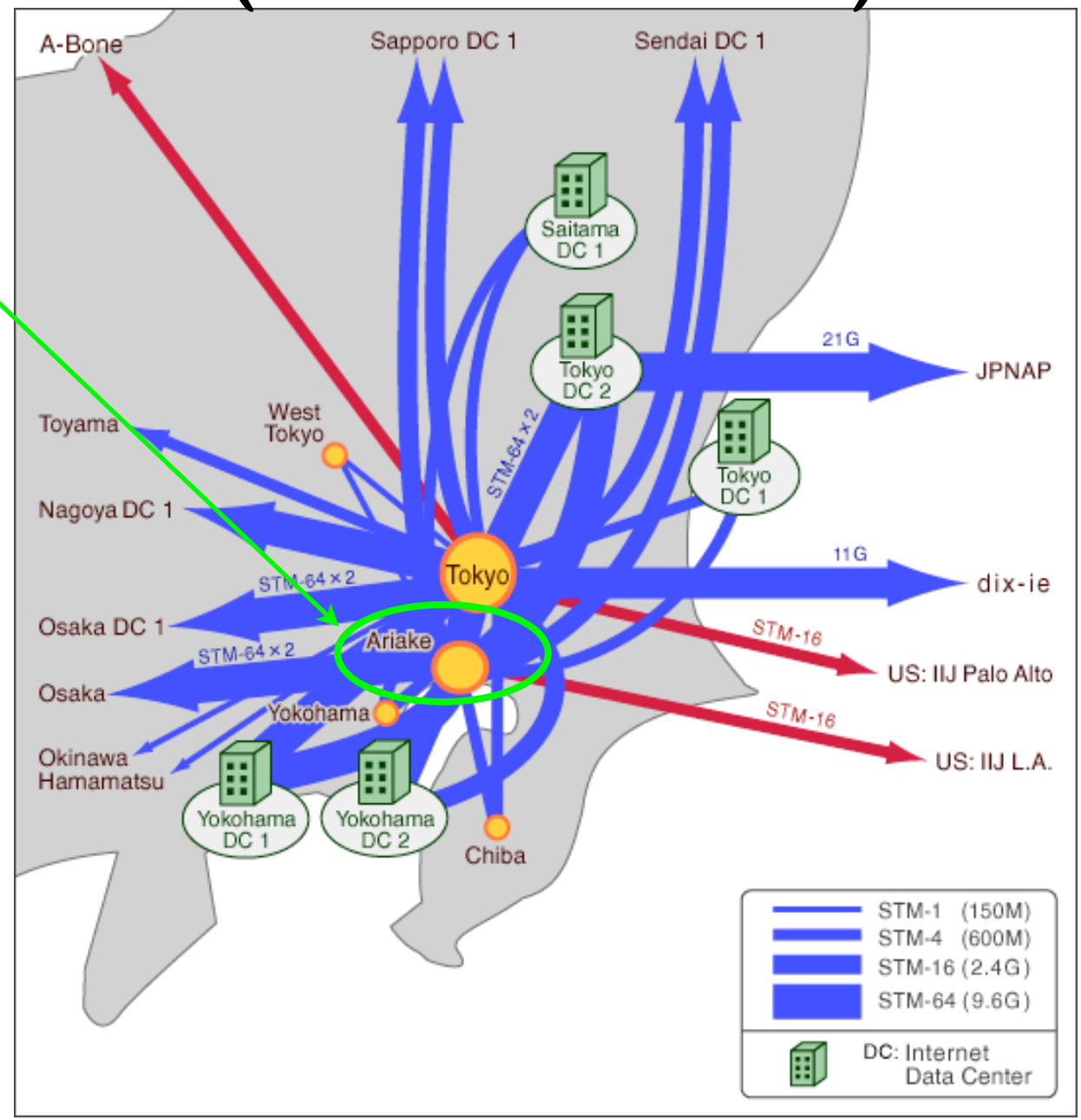
```
lax001bb00.iij.net (216.98.96.176)
```



# Arrived IIJ in Ariake (perhaps ...)

```
tky002bb01.iij.net (216.98.96.178)
```

Either map is out of date, DNS name above is misleading, or `traceroute` is incorrect!



# A↦B packet path may differ from B↦A

## Different paths: Different network properties (latency, bandwidth, etc)

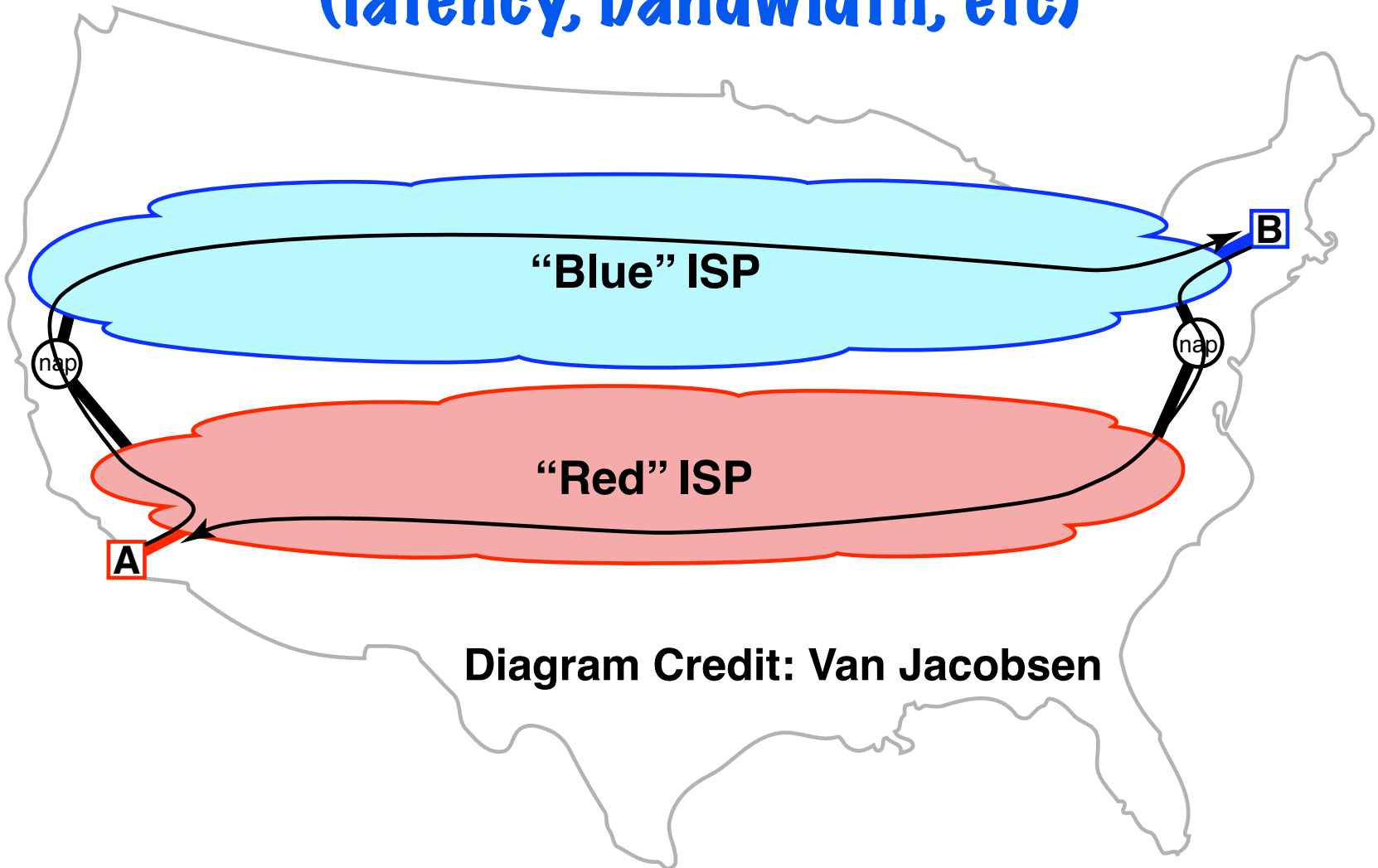


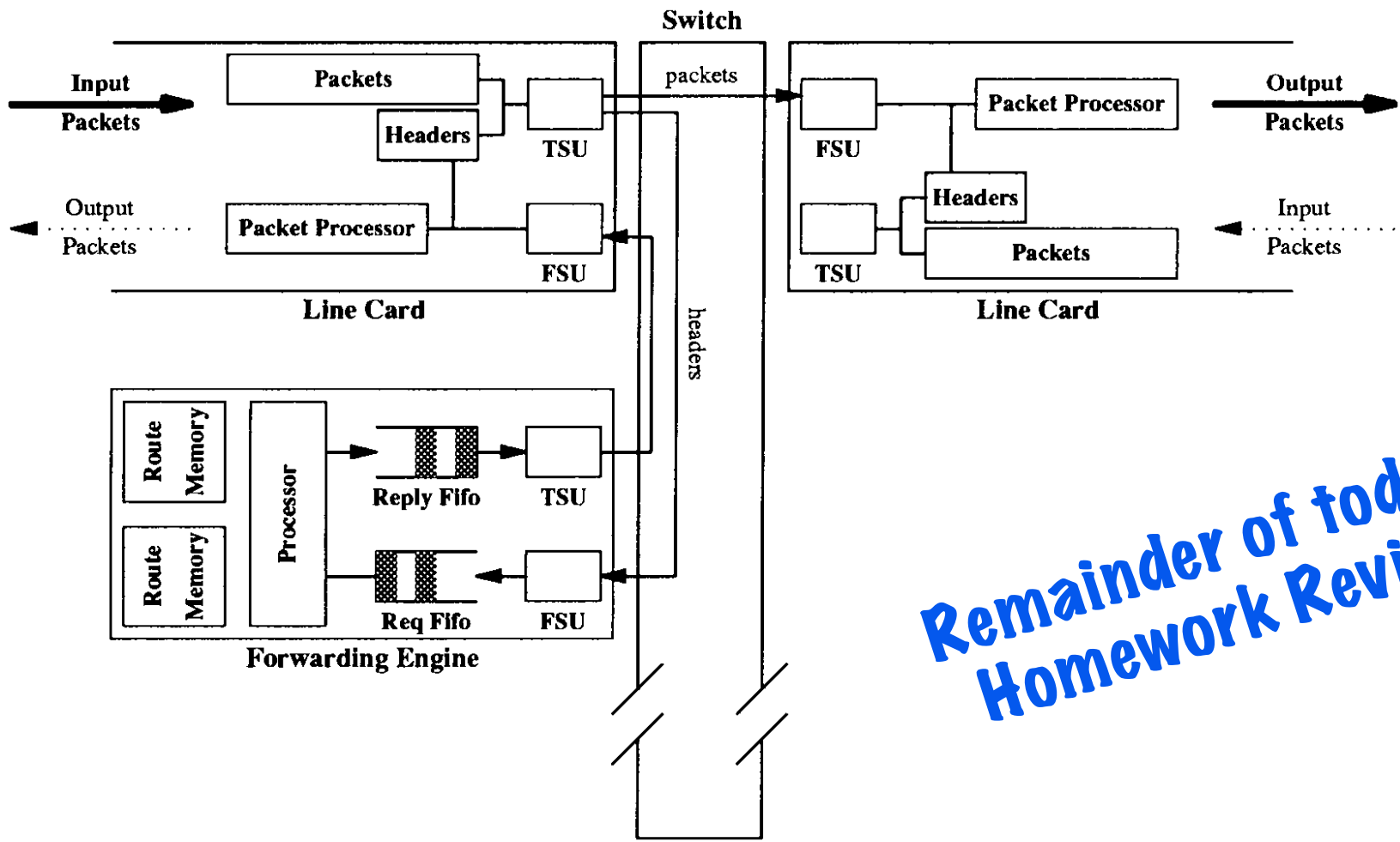

Diagram Credit: Van Jacobsen

**Economics: A and B use different network carriers … carriers route data onto their networks ASAP.**

# Next Time: How to design a router

## A 50-Gb/s IP Router

Craig Partridge, *Senior Member, IEEE*, Philip P. Carvey, *Member, IEEE*, Ed Burgess, Isidro Castineyra, Tom Clarke, Lise Graham, Michael Hathaway, Phil Herman, Allen King, Steve Kohalmi, Tracy Ma, John Mcallen, Trevor Mendez, Walter C. Milliken, *Member, IEEE*, Ronald Pettyjohn, *Member, IEEE*, John Rokosz, *Member, IEEE*, Joshua Seeger, Michael Sollins, Steve Storch, Benjamin Tober, Gregory D. Troxel, David Waitzman, and Scott Winterble




**Remainder of today: Homework Review**