HIGH-SPEED NETWORKS: INTERNET A GBITS/S RATE

C. Pham
Université de Pau et des Pays de l'Adour
Département Informatique
http://www.univ-pau.fr/~cpham
Congduc.Pham@univ-pau.fr

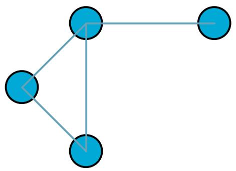
Copyright

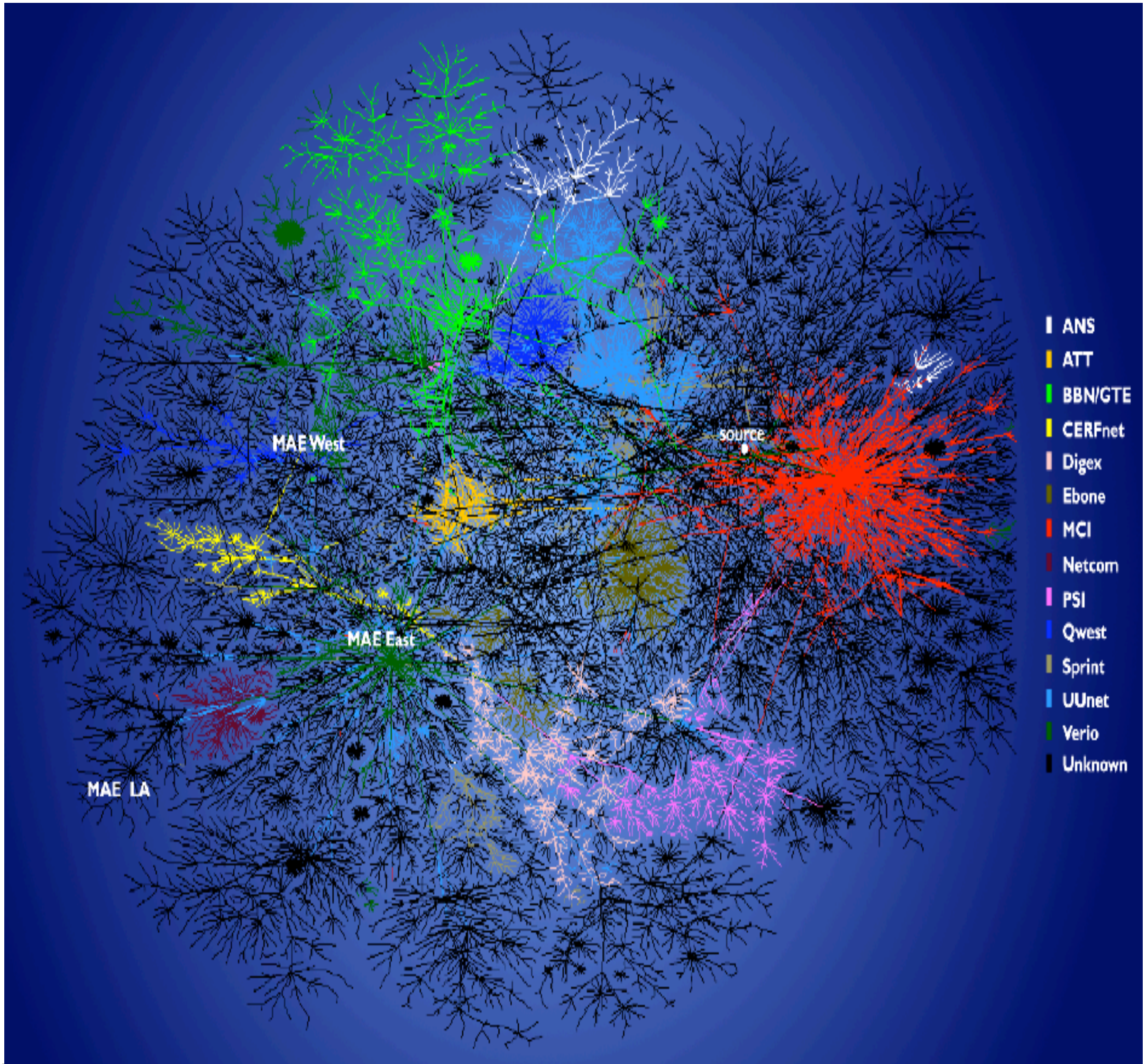
Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

THE NEW FACE OF THE INTERNET

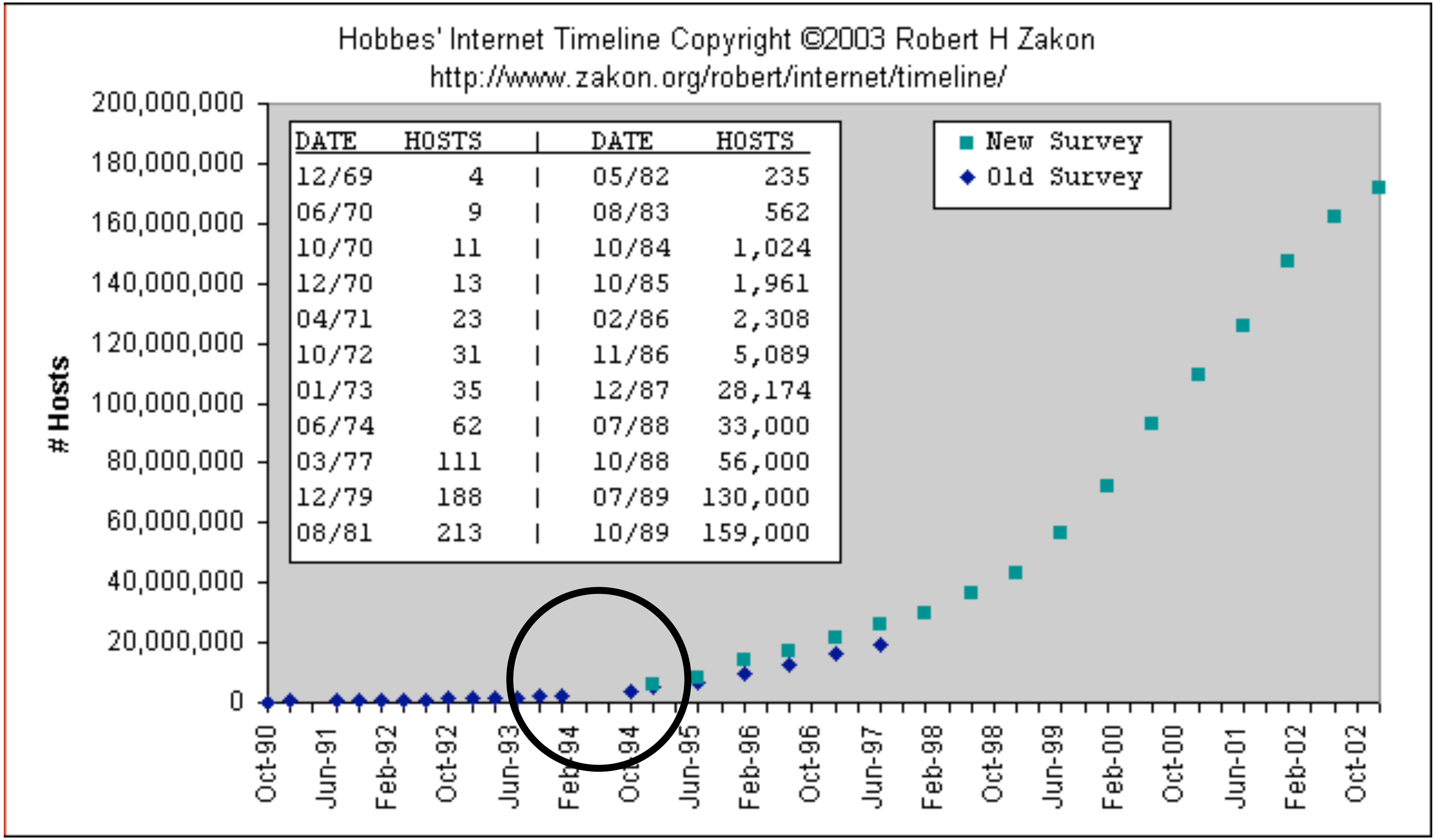
Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

The big-bang of the Internet

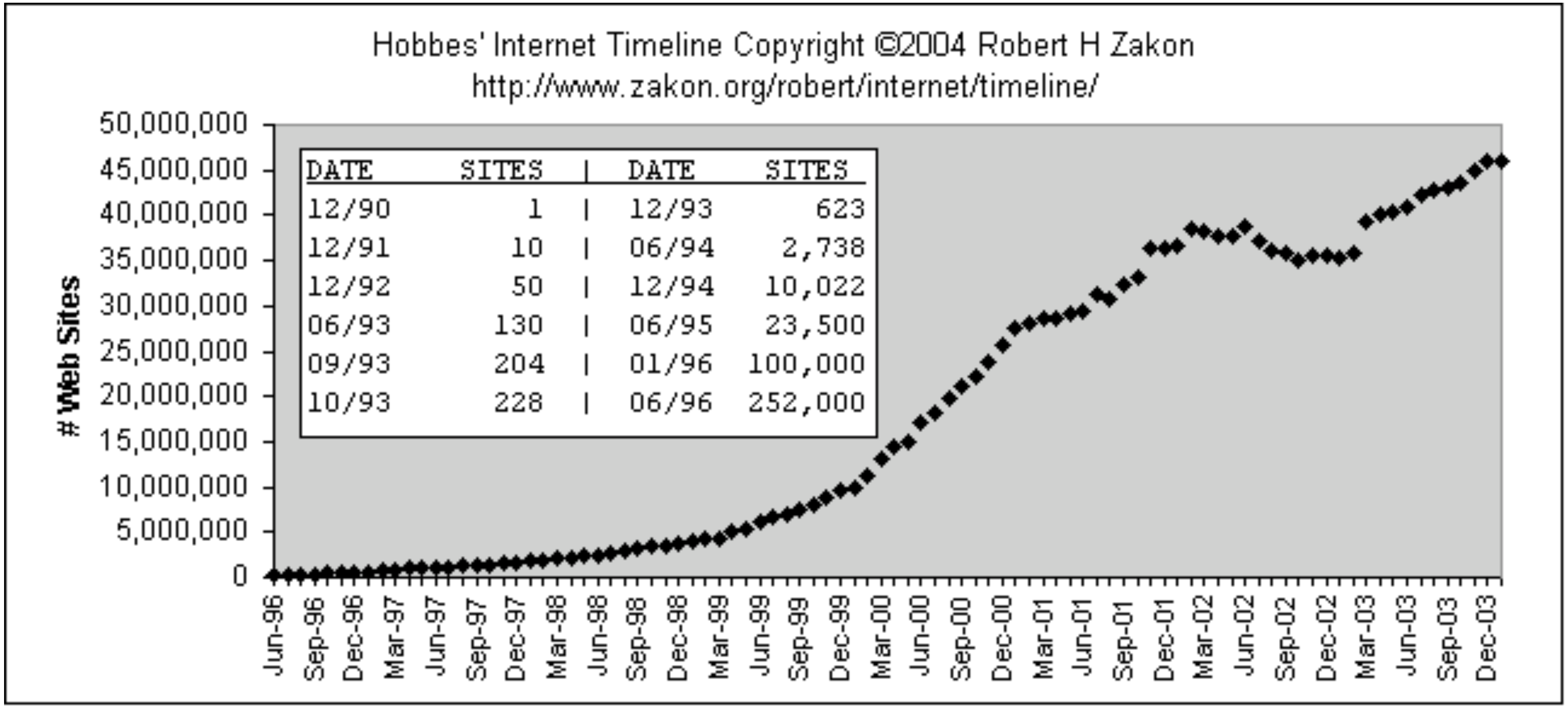




Internet host

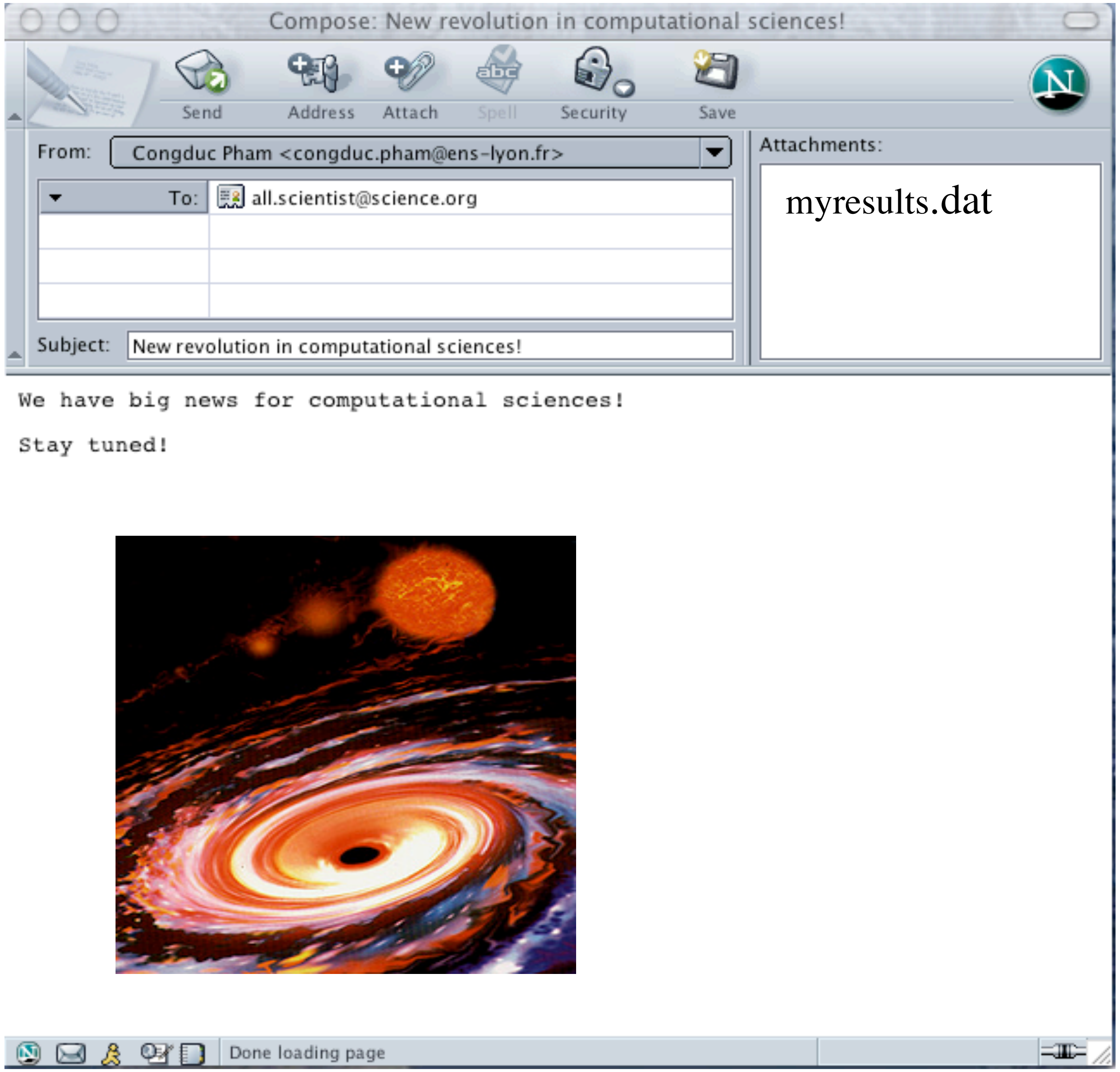


www.web-the-big-bang.org



Internet usage: e-mail…

- Convenient way to communicate in an informal manner
- Attachments as a easy way to exchange data files, images…



Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

…and surfing the web

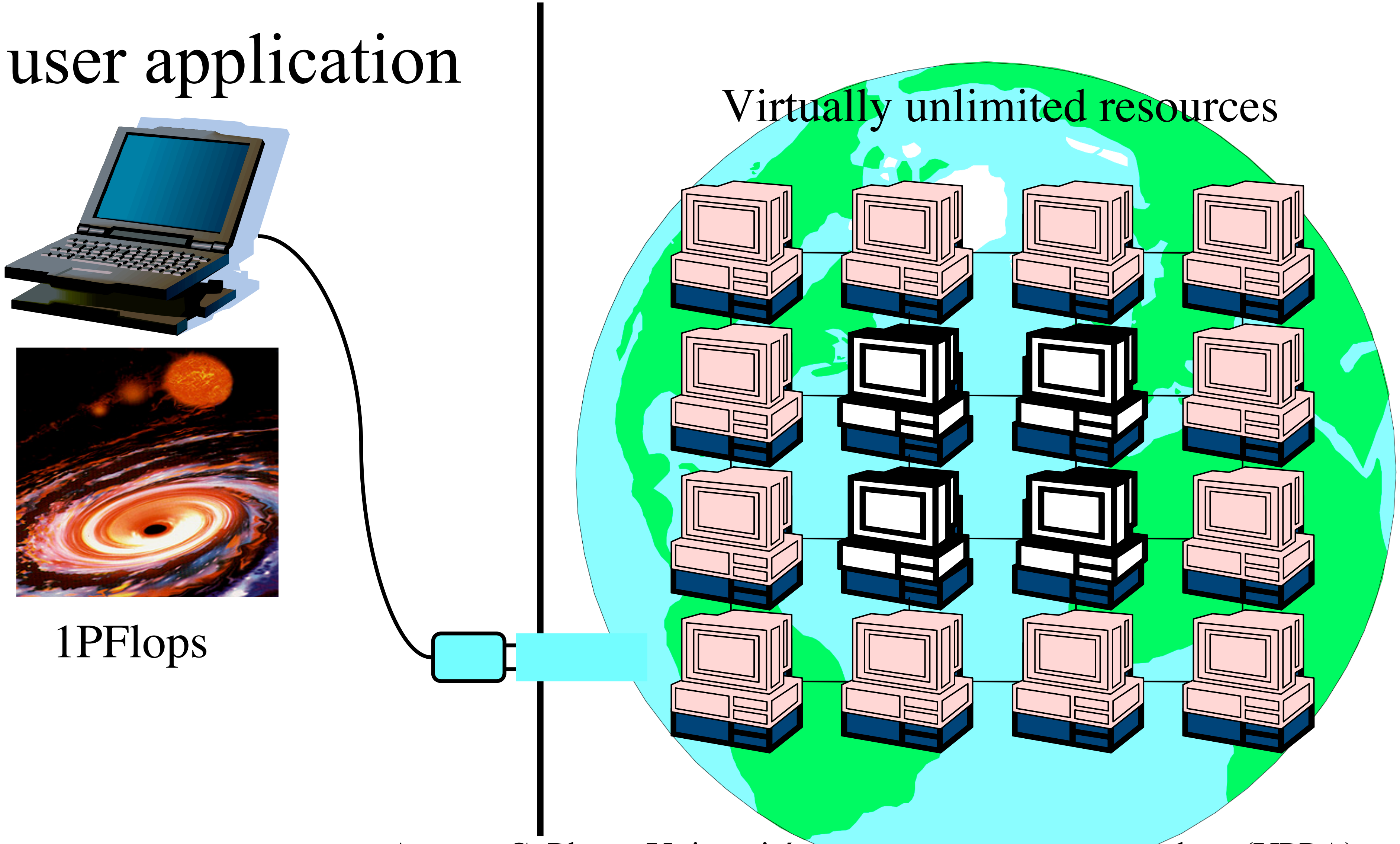
- **A true revolution for rapid access to information**
- **Increasing number of apps:**
  - e-science,
  - e-commerce, B2B, B2C,
  - e-training, e-learning,
  - e-tourism
  - …

New applications on the information highways

Think about...

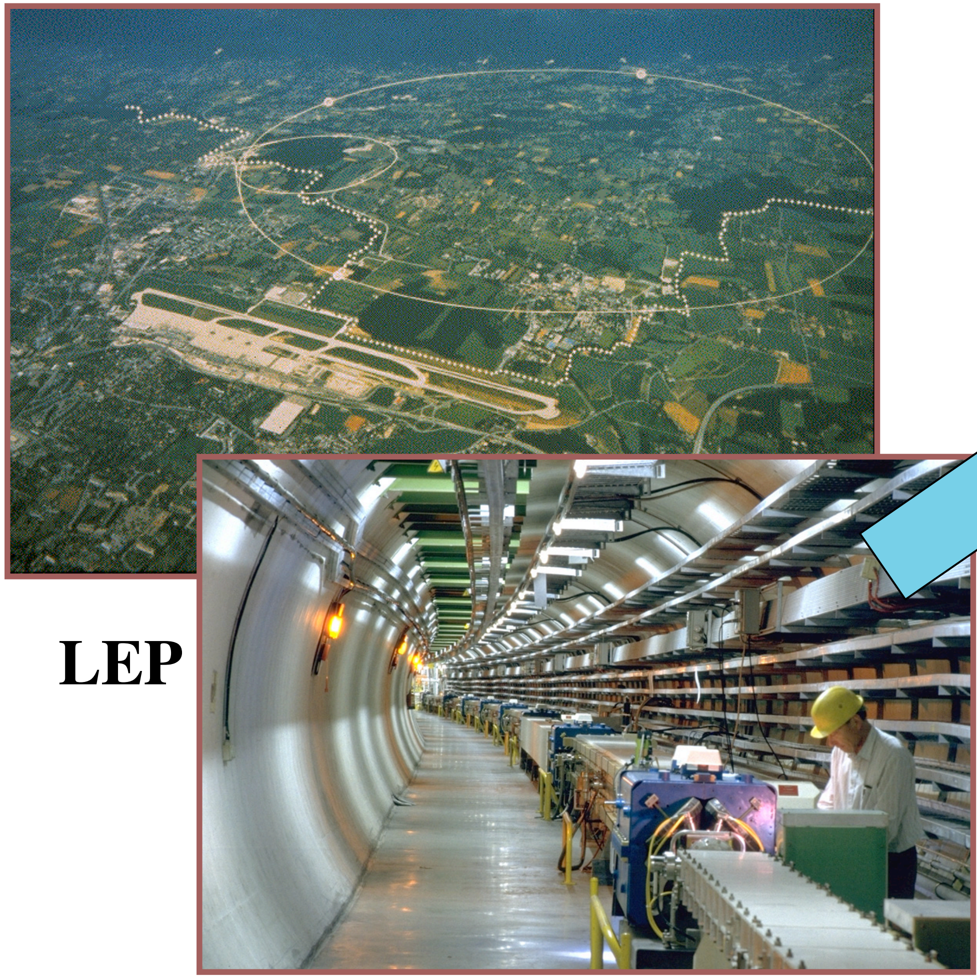
- video-conferencing
- video-on-demand
- interactive TV programs
- remote archival systems
- tele-medecine
- virtual reality, immersion systems
- high-performance computing, grids
- distributed interactive simulations

Computational grids

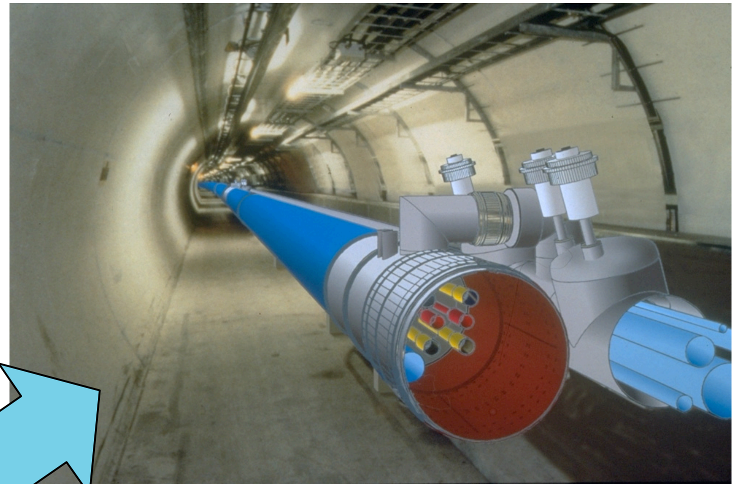


from Dorian Arnold: Netsolve Happenings

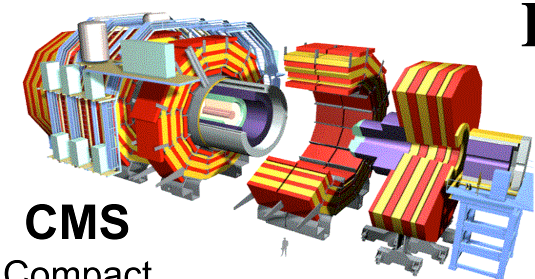
High Energy Physics at CERN



LEP

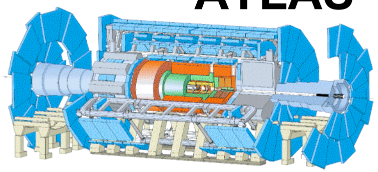


LHC



CMS
Compact Muon Solenoid

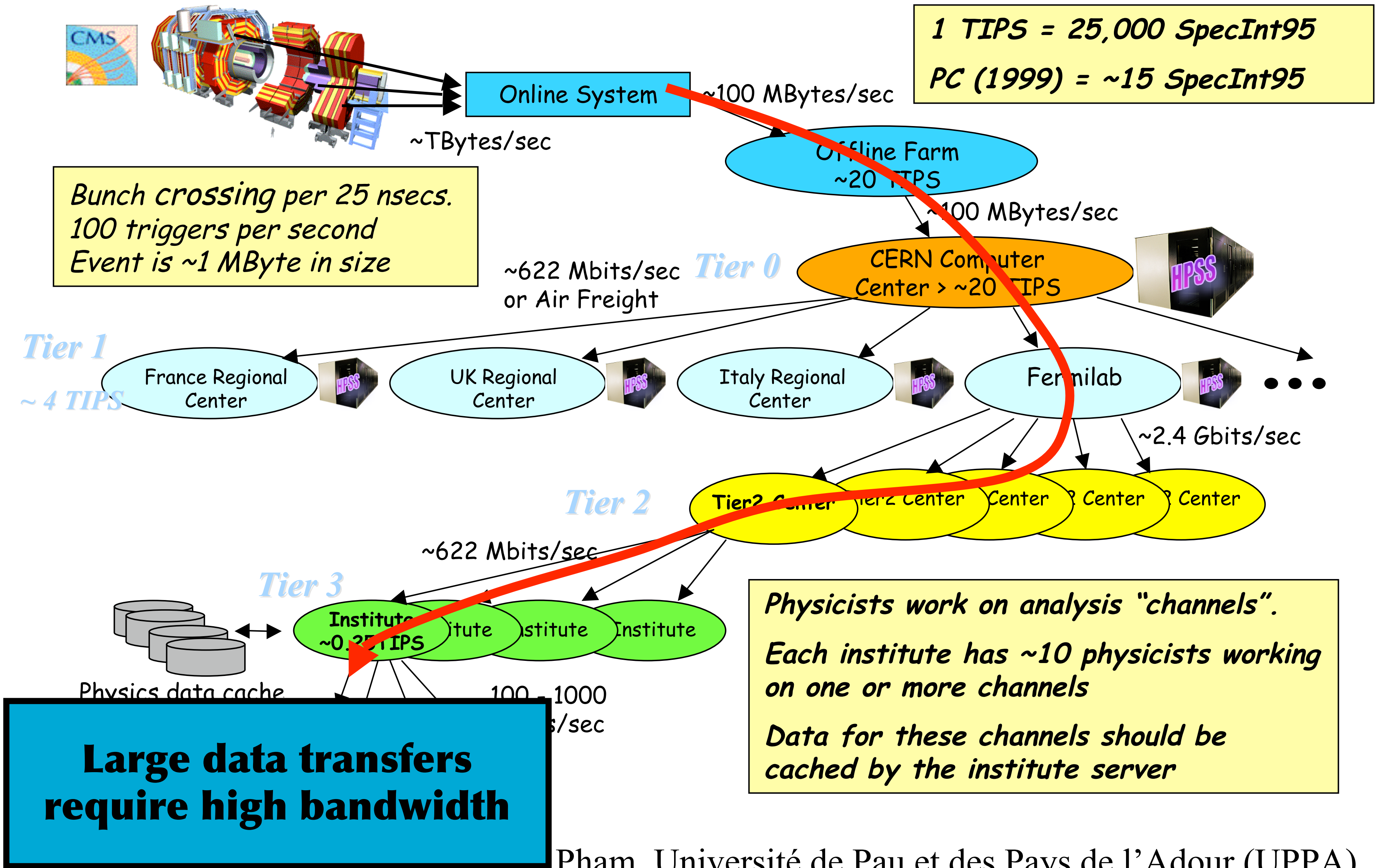
ATLAS



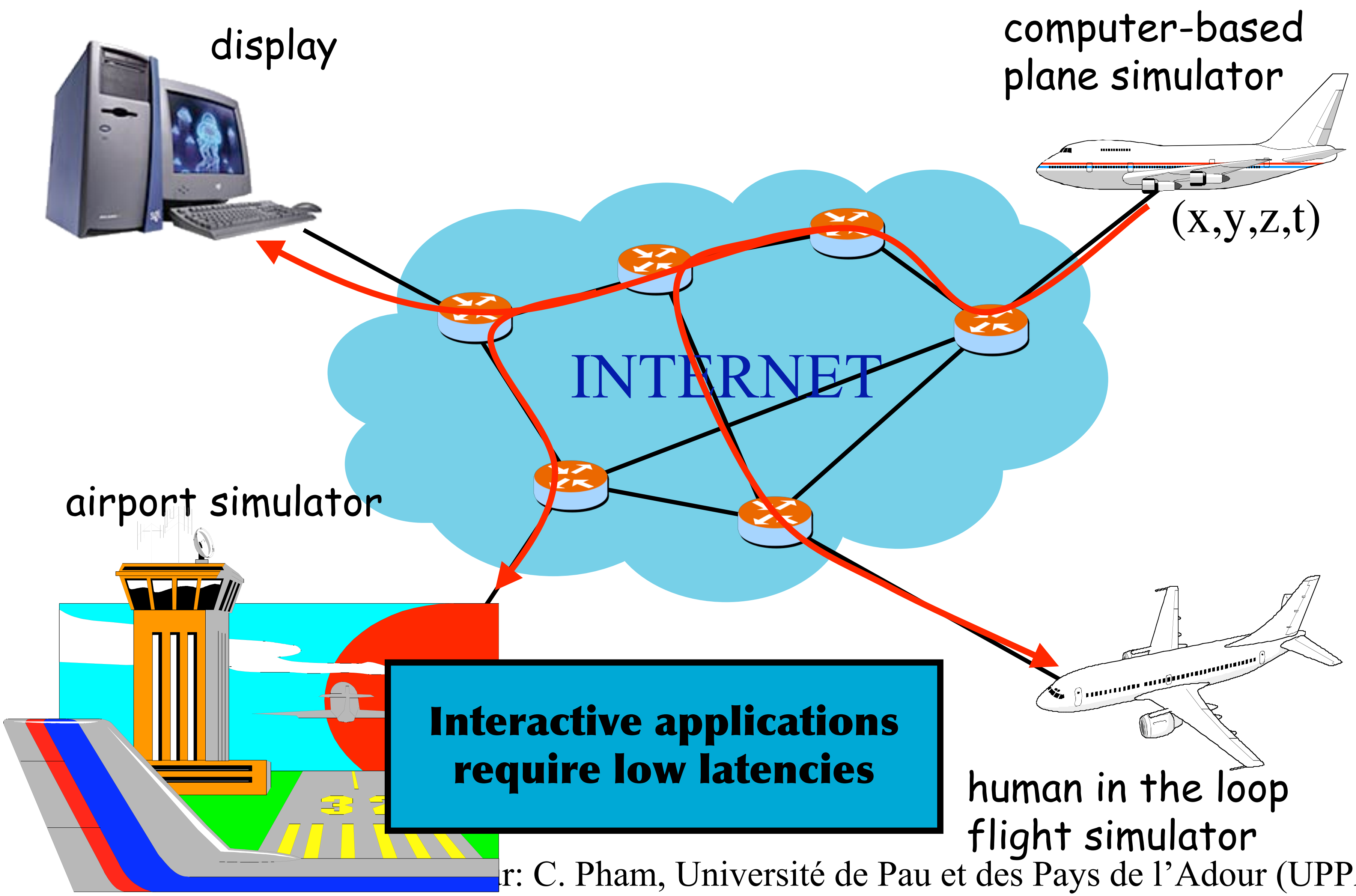
3.5 Petabytes/year ≈ 10^9 events/year

Images from EDG (DataGrid) project

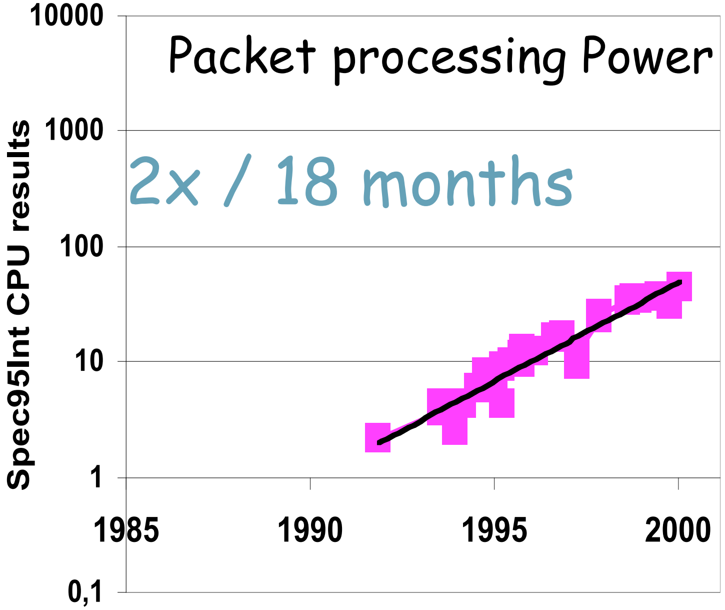
Distributed Databases



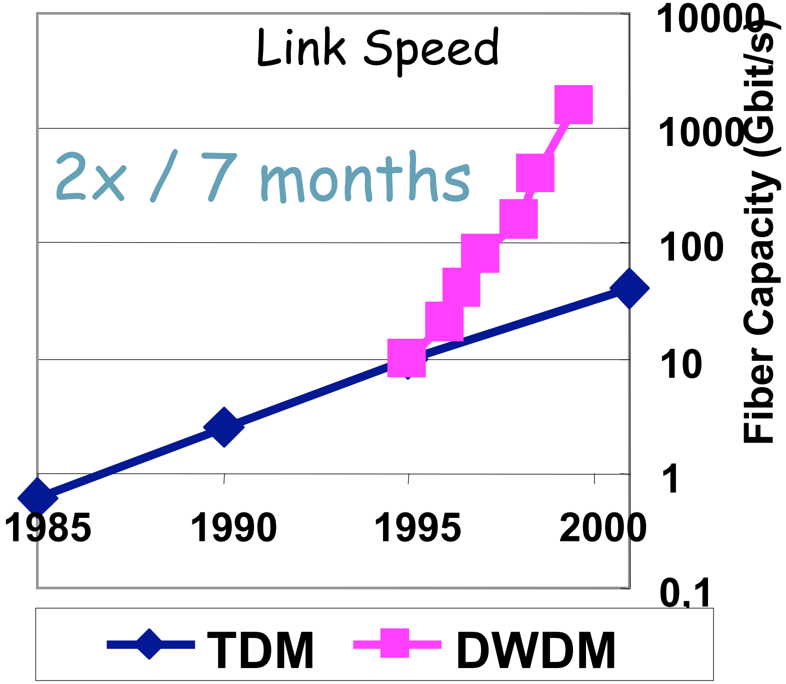
Wide-area interactive simulations



The optical revolution



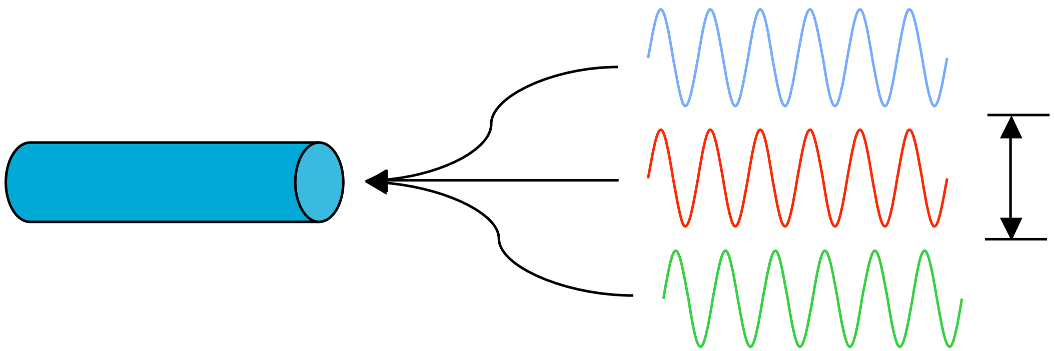
From McKeown

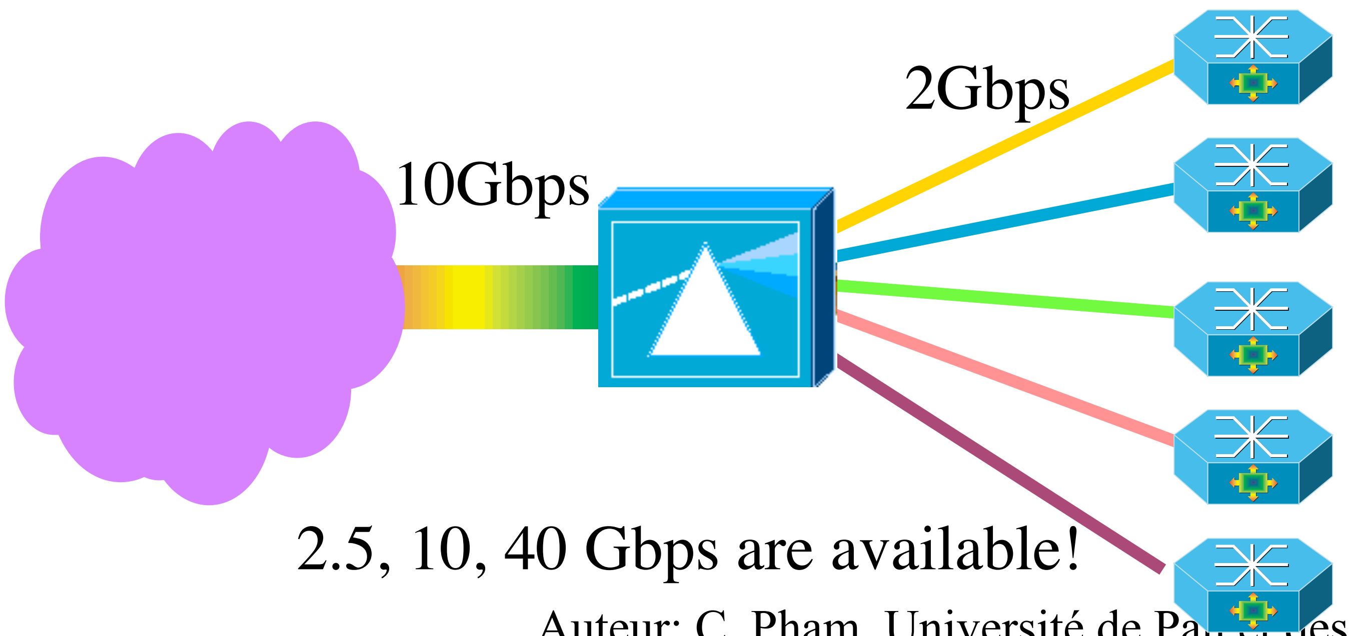


Demand: about 111 million km of cabled optical fiber / year

DWDM, bandwidth for free?

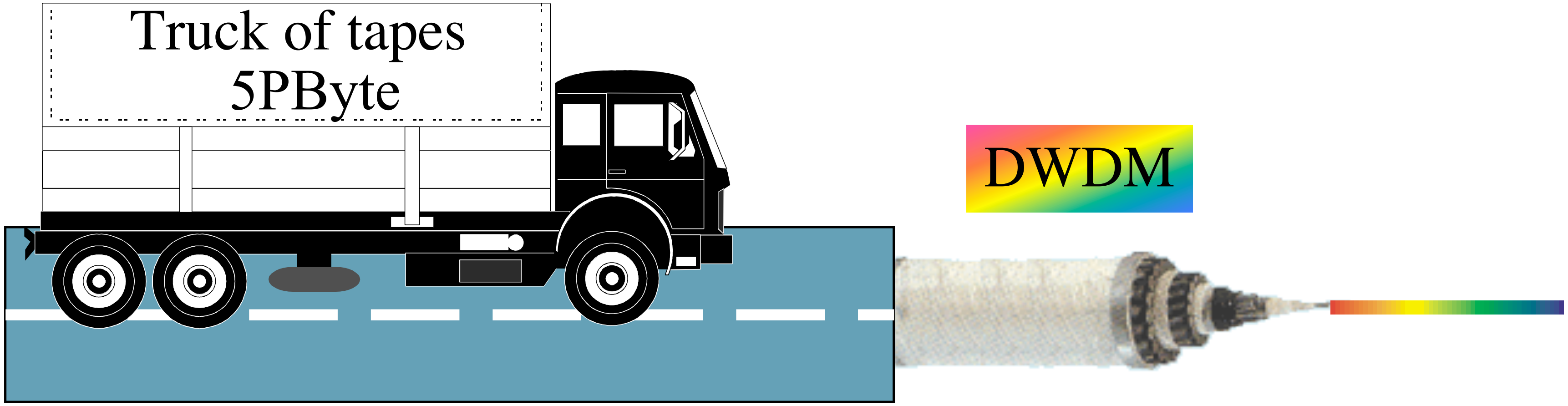
DWDM: Dense Wavelength Division Multiplexing





2.5, 10, 40 Gbps are available!

The information highways



NEWS of Dec 15th, 2004

A throughput of 1.28 Tbits/s has been achieved on a 430kms regular monomode fiber between France Telecom and Deutsch Telecom using 8 DWDM channels (EU project TOPRATE)

Revisiting the truck of tapes

(18 of 18)

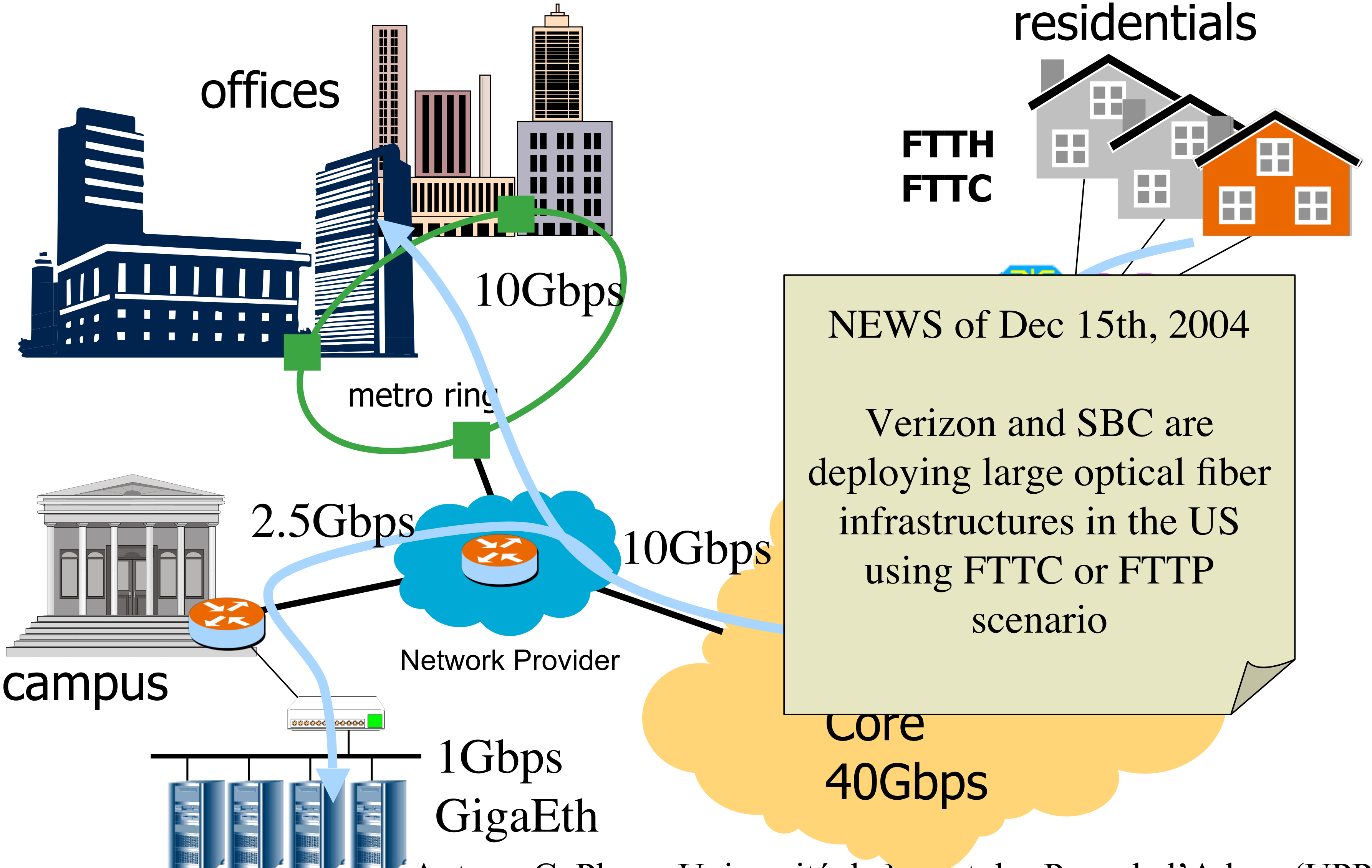
Consider one fiber

- Current technology allows for 320 λ in one of the frequency bands
- Each λ has a bandwidth of 40 Gbit/s
- Transport: $320 * 40*10^9 / 8 = 1600$ GByte/sec
- Take a 10 metric ton truck
- One tape contains 50 Gbyte, weights 100 gr
- Truck contains (10000 / 0.1) * 50 Gbyte = 5 PByte
- Truck / fiber = 5 PByte / 1600 GByte/sec = 3125 s ≈ one hour
- For distances further away than a truck drives in one hour (50 km) minus loading and handling 100000 tapes **the fiber wins!!!**

Example from A. Tanenbaum, slide from Cees De Laat

Fibers everywhere?

Where's the killer application?



NEWS of Dec 15th, 2004

Verizon and SBC are deploying large optical fiber infrastructures in the US using FTTC or FTTP scenario

High Performance Routers

©cisco

©Juniper

©Procket Networks

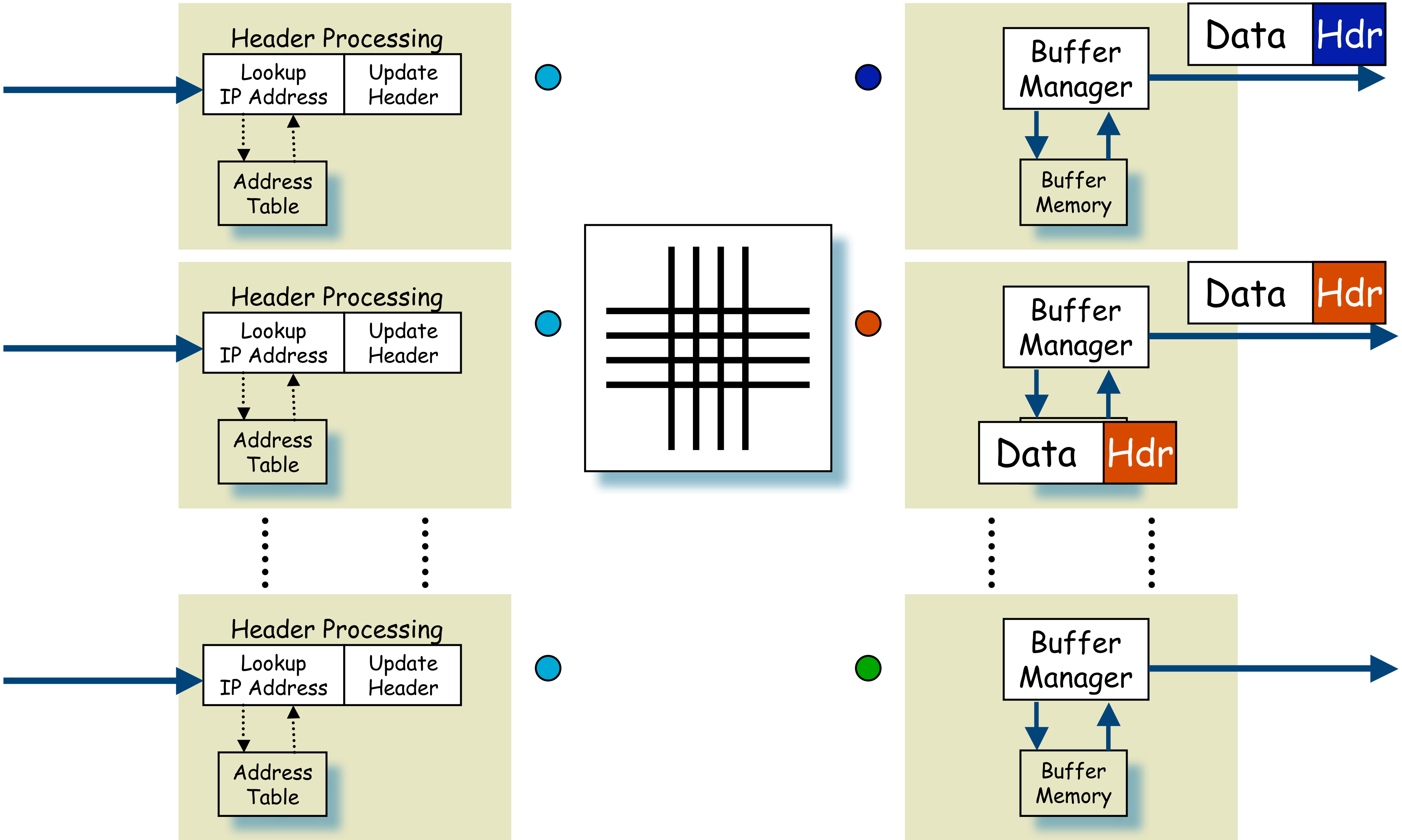
©Alcatel

©Lucent

©Nortel Networks

and more…

General router architecture



Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

Performance constraints

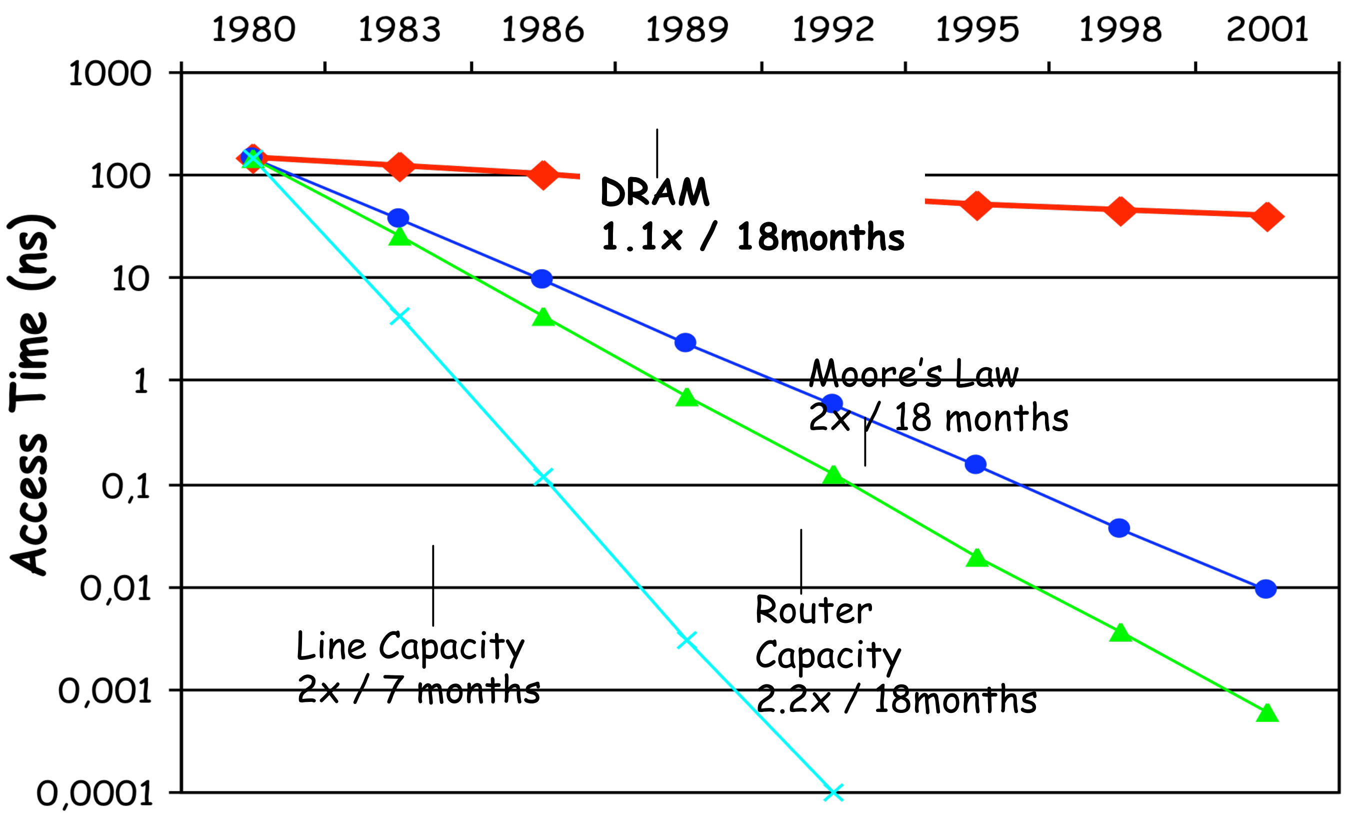
- At gigabit rate, millions of packets must be routed per seconds!

Year	Line	Linerate (Gbps)	40B (Mpps)	84B (Mpps)	354B (Mpps)
1997-98	OC3	0.155	0.48	0.23	0.054
1998-99	OC12	0.622	1.94	0.92	0.22
1999-00	OC48	2.5	7.81	3.72	0.88
2000-01	OC192	10.0	31.25	14.88	3.53
2002-03	OC768	40.0	125	59.52	14.12
	1GE	1.0	3.13	1.49	0.35

Memory Bandwidth

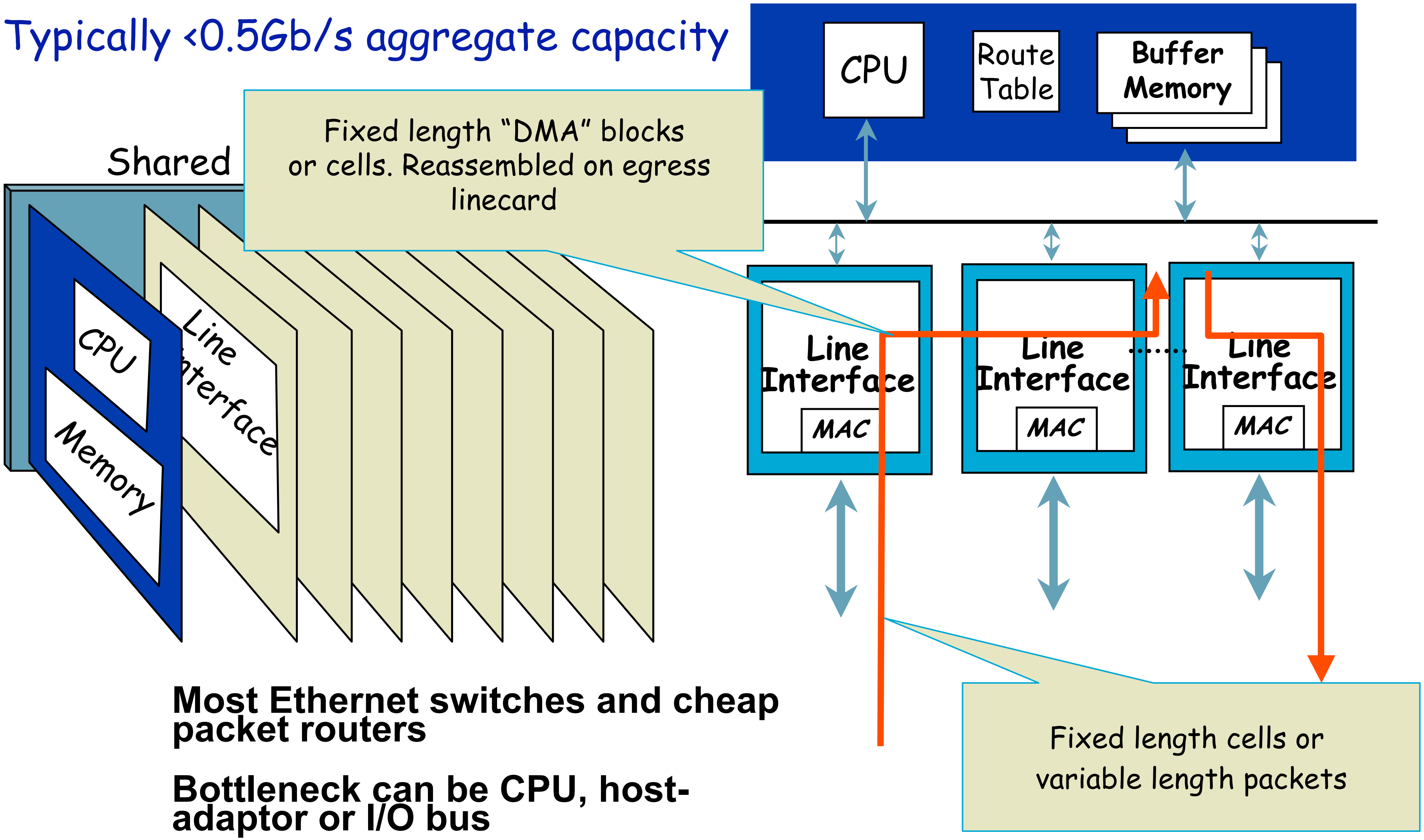
Commercial DRAM

- The bottleneck is memory speed.
- Memory speed is not keeping up with Moore's Law.



1st generation routers

Typically <0.5Gb/s aggregate capacity

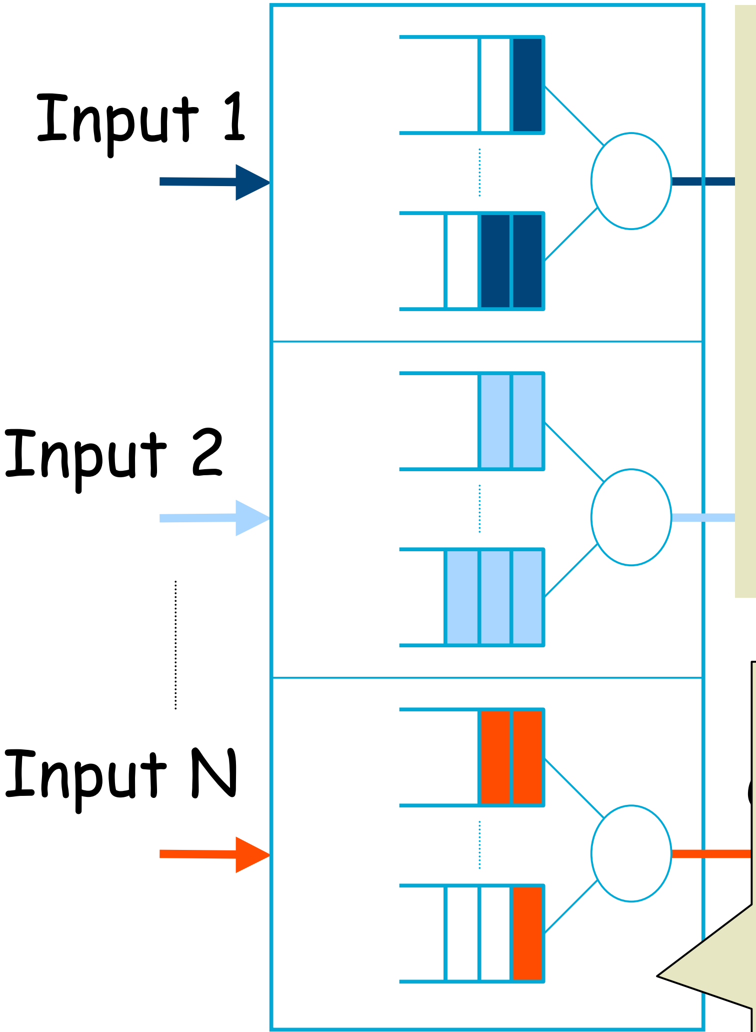


Most Ethernet switches and cheap packet routers

Bottleneck can be CPU, host-adaptor or I/O bus

First Generation Routers

Queueing Structure: Shared Memory



Numerous work has proven and made possible:

- Fairness
- Delay Guarantees
- Delay Variation Control
- Loss Guarantees
- Statistical Guarantees

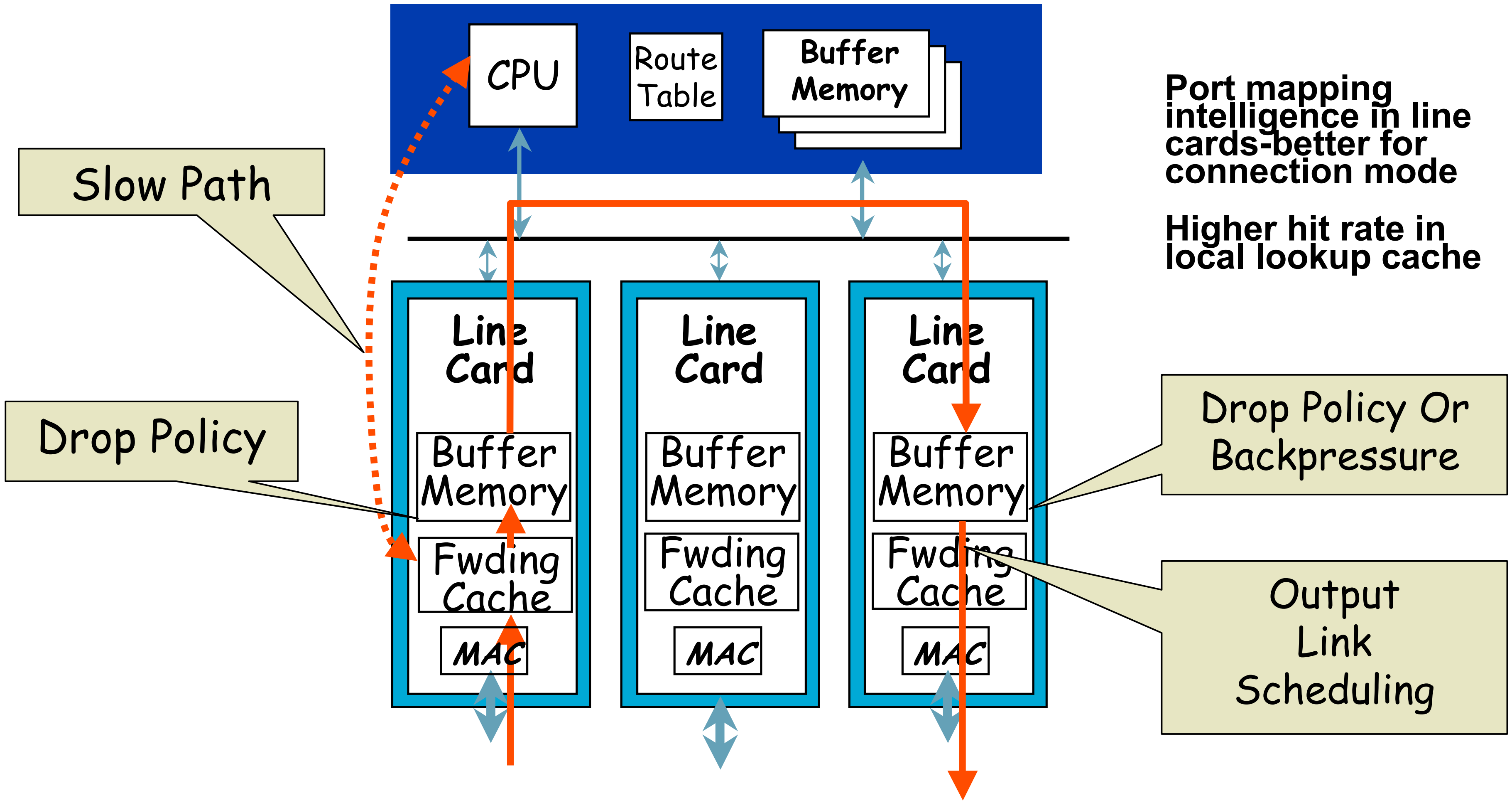
Large, single dynamically allocated memory buffer:
N writes per "cell" time
N reads per "cell" time.
Limited by memory bandwidth.

Limitations of shared memory

- **First generation router built with 133 MHz Pentium**
  - Mean packet size 500 bytes
  - Interrupt takes 10 microseconds, word access take 50 ns
  - Per-packet processing time is 200 instructions = 1.504 µs
- **Copy loop**
  - register <- memory[read_ptr]
  - memory [write_ptr] <- register
  - read_ptr <- read_ptr + 4
  - write_ptr <- write_ptr + 4
  - counter <- counter -1
  - if (counter not 0) branch to top of loop
- **4 instructions + 2 memory accesses = 130.08 ns**
- **Copying packet takes 500/4 *130.08=16.26µs; interrupt 10 µs**
- **Total time = 27.764 µs => speed is 144.1 Mbps**
- **Amortized interrupt cost balanced by routing protocol cost**

2nd generation routers

Typically <5Gb/s aggregate capacity

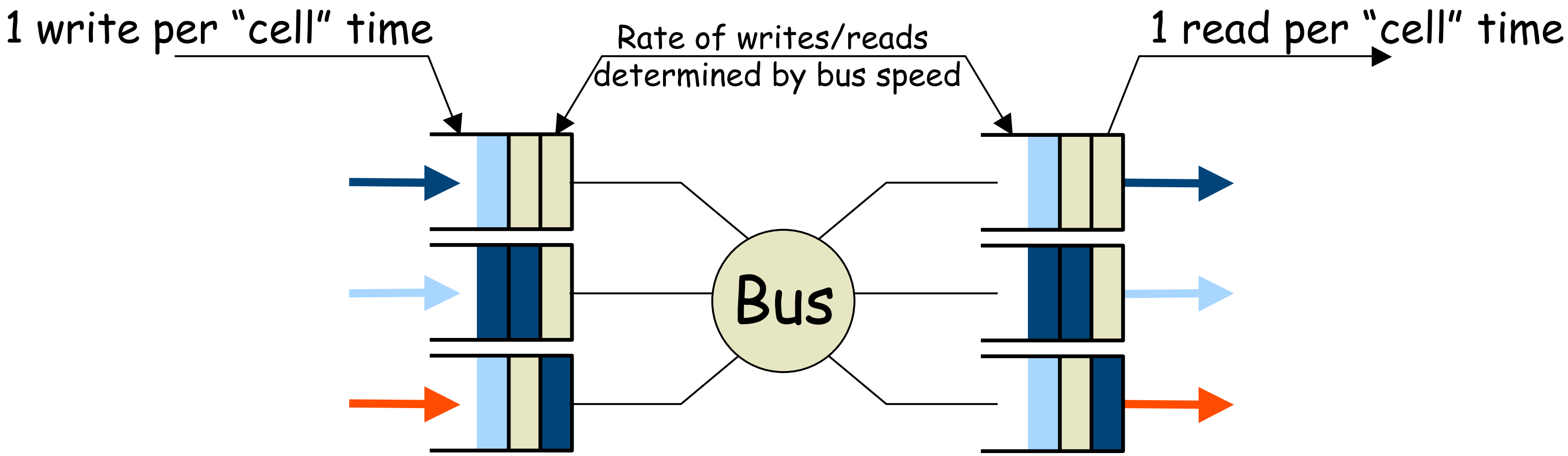


Port mapping intelligence in line cards-better for connection mode

Higher hit rate in local lookup cache

Second Generation Routers

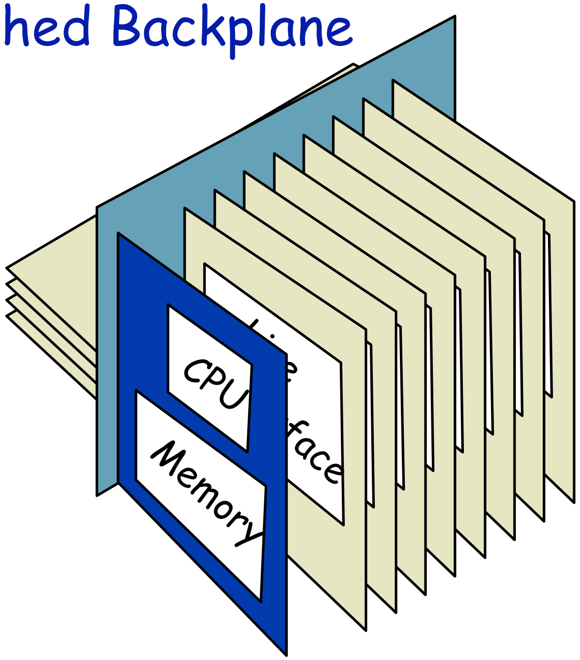
Queueing Structure: Combined Input and Output Queueing

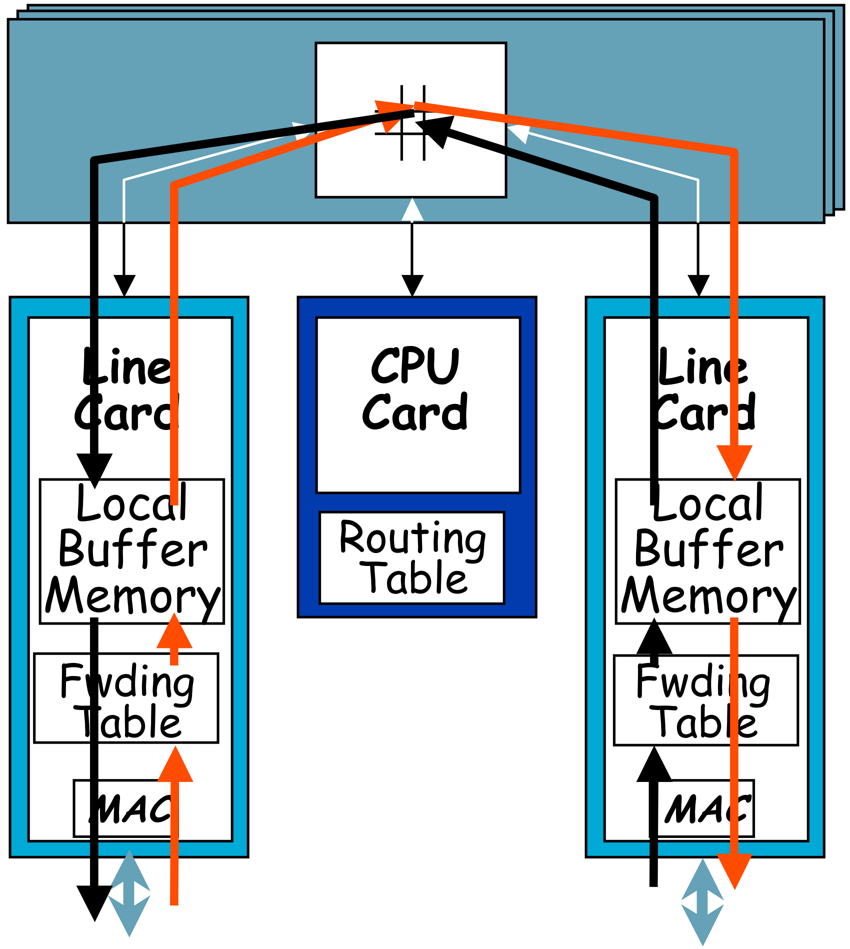


Third Generation Routers

- Third generation switch provides parallel paths (fabric)

Switched Backplane

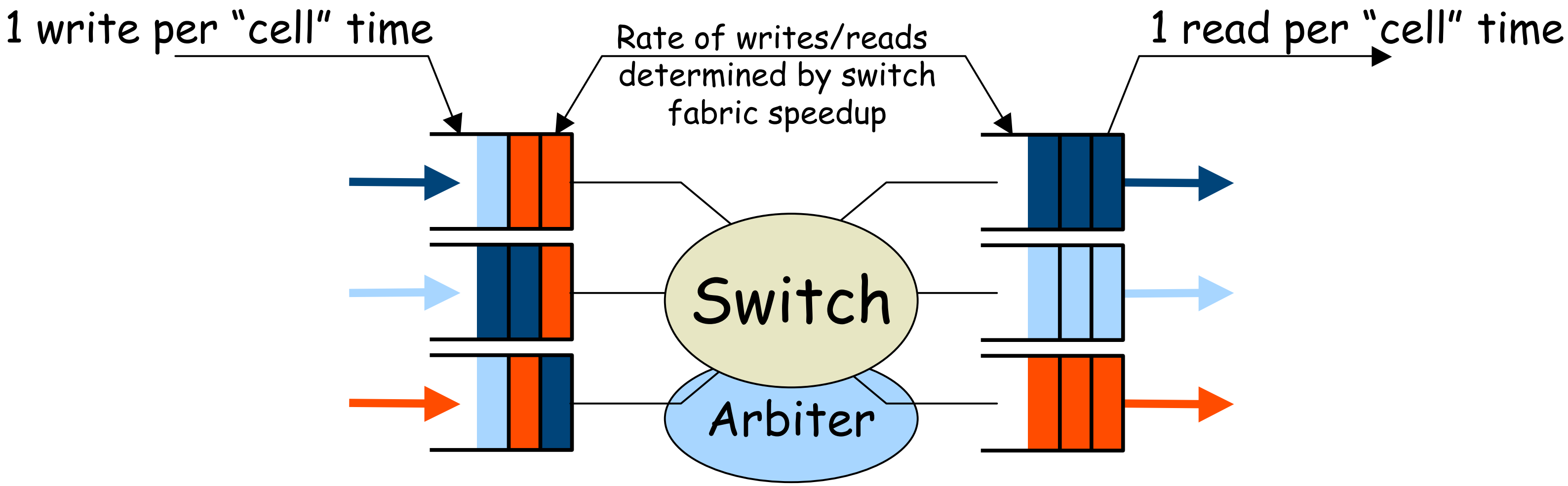




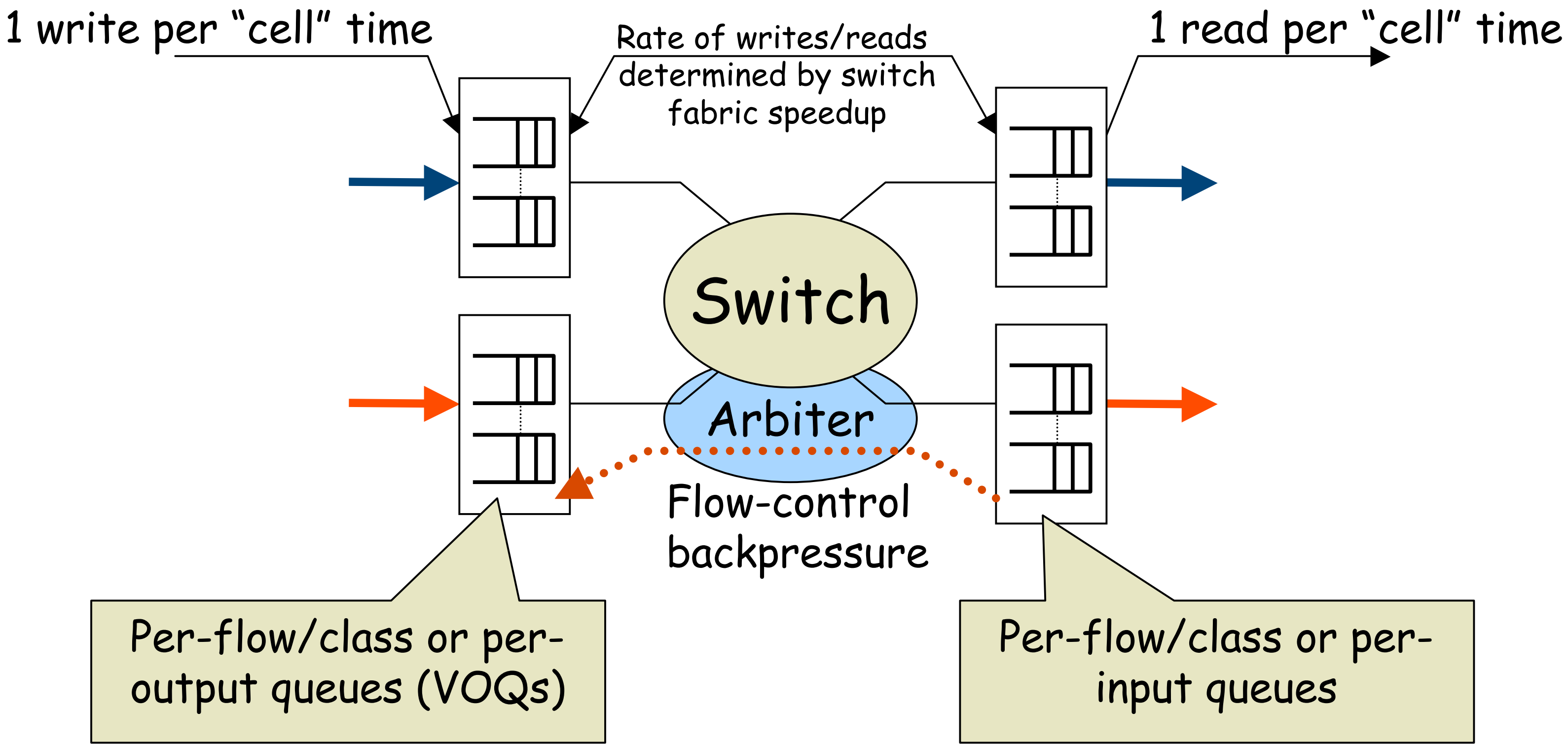
Typically <50Gb/s aggregate capacity

Third Generation Routers

Queueing Structure

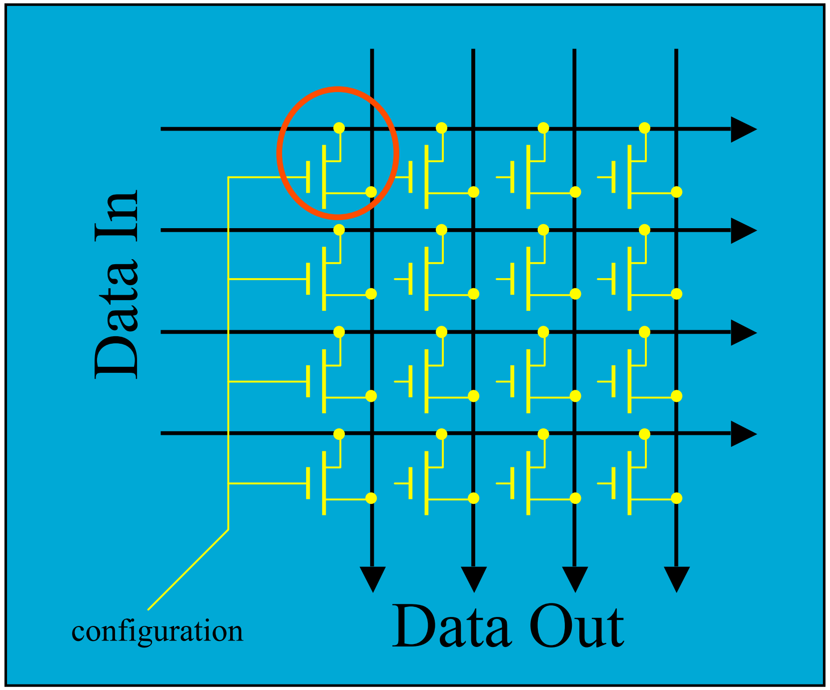


Third Generation Routers
Queueing Structure



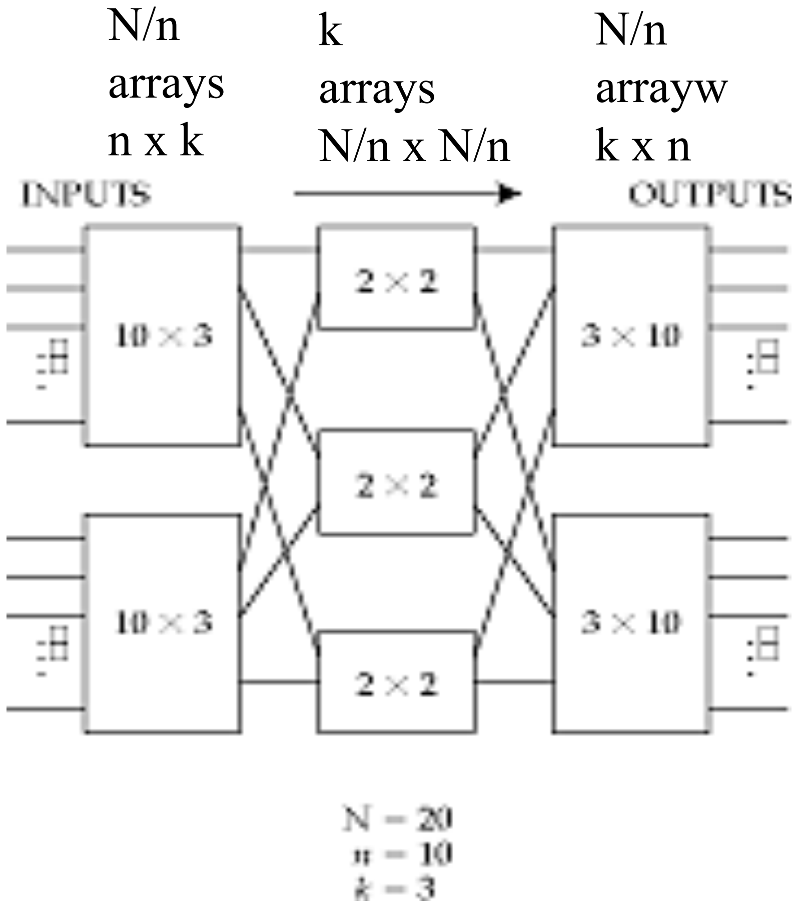
Crossbar

- **Simplest possible space-division switch**
- **Crosspoints can be turned on or off, long enough to transfer a packet from an input to an output**
- **Expensive**
- **Internally nonblocking**
  - but need N2 crosspoints
  - time to set each crosspoint grows quadratically

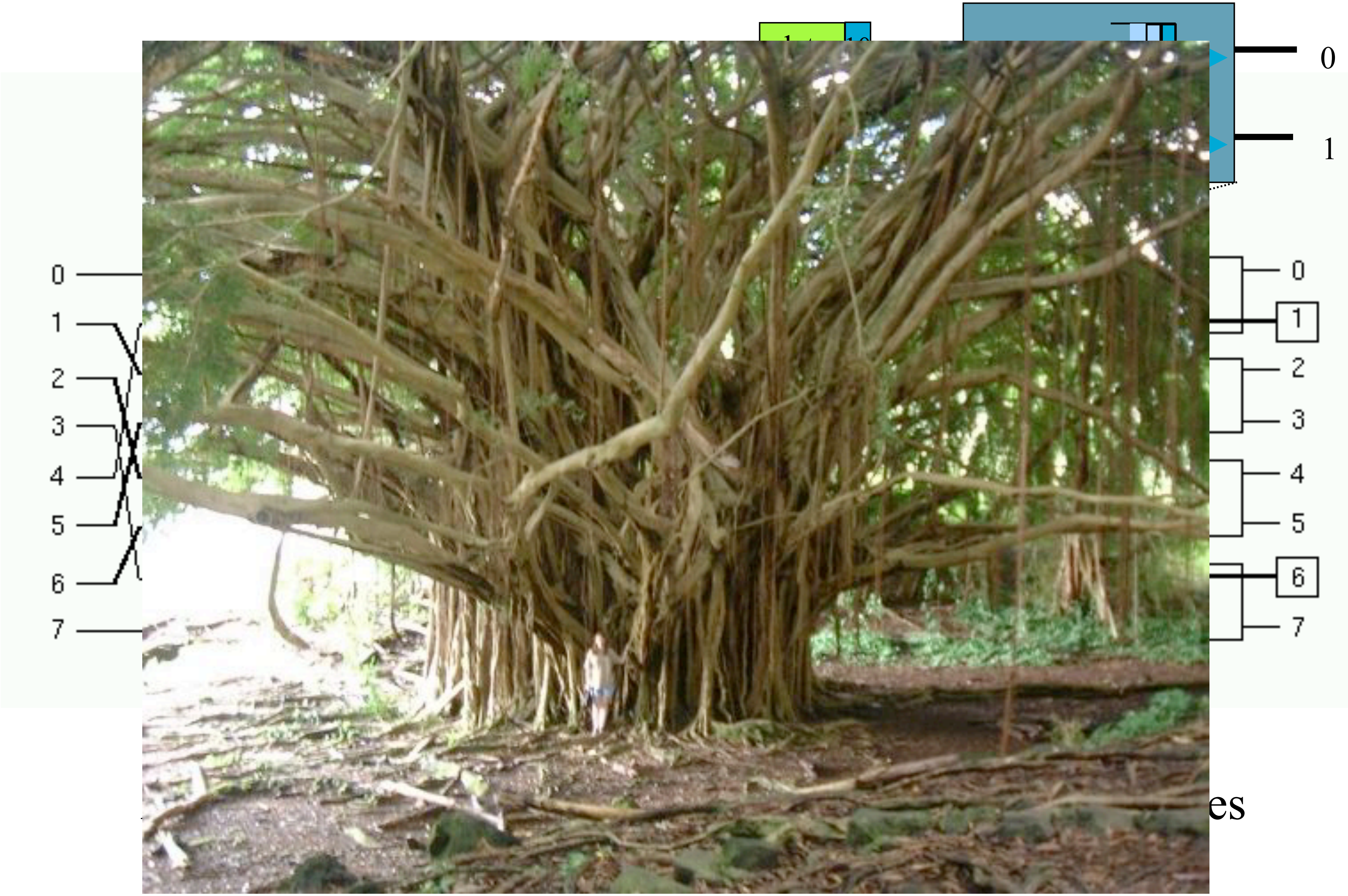


Multistage crossbar

- In a crossbar during each switching time only one cross-point per row or column is active
- Can save crosspoints if a cross-point can attach to more than one input line
- This is done in a multistage crossbar



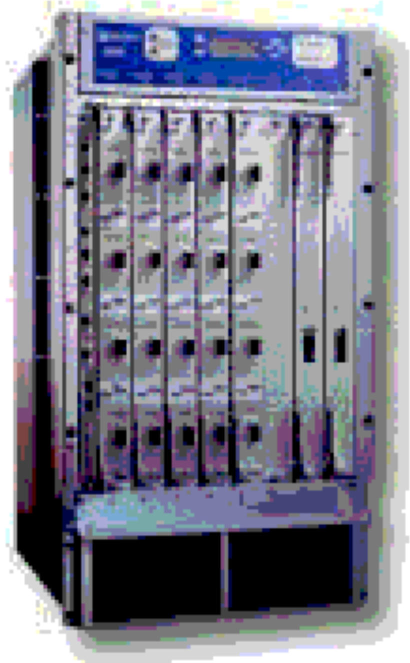
Banyan multi-stage networks



Example: Juniper high performance router

M160 Router Overview

The highest performance Internet backbone router available today, providing uncompromising performance with services enabled.



- **Flexible interface types**
  - Speeds from T1/E1 to OC-192c/STM-64
  - First shipping OC-192c/STM-64 interface on the market
- **Internet Processor II ASIC**
  - Performance *with* services
  - Aggregate throughput capacity exceeds 160 Gbps
- **Fully redundant system**
- **JUNOS Internet software**
- **8 slot chassis**
  - >24 Gbps full duplex throughput per slot
  - Space for 32 PICs per chassis
- **Two units per 7' rack**
  - 19"/48cm (w) x 35"/89cm (h) x 29"/74cm (d)

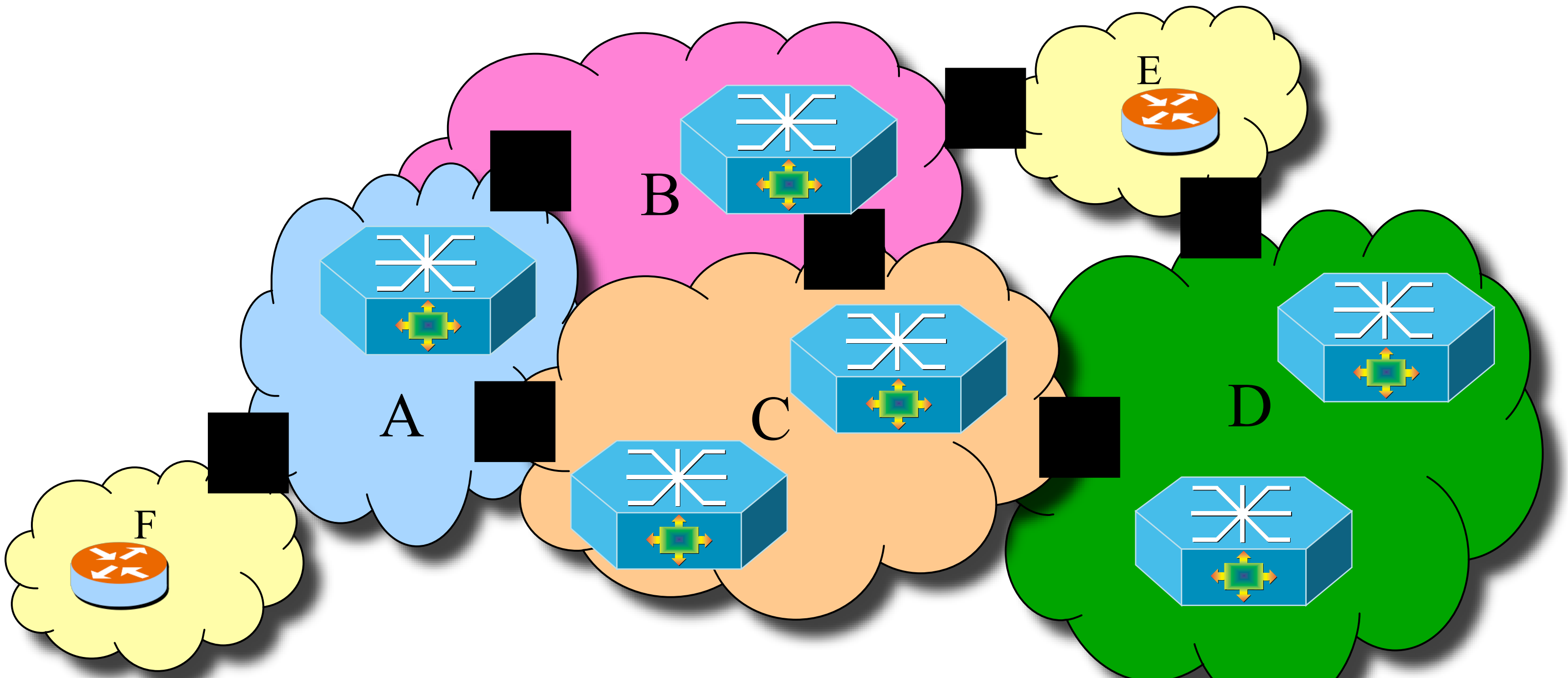
4/12/01

13

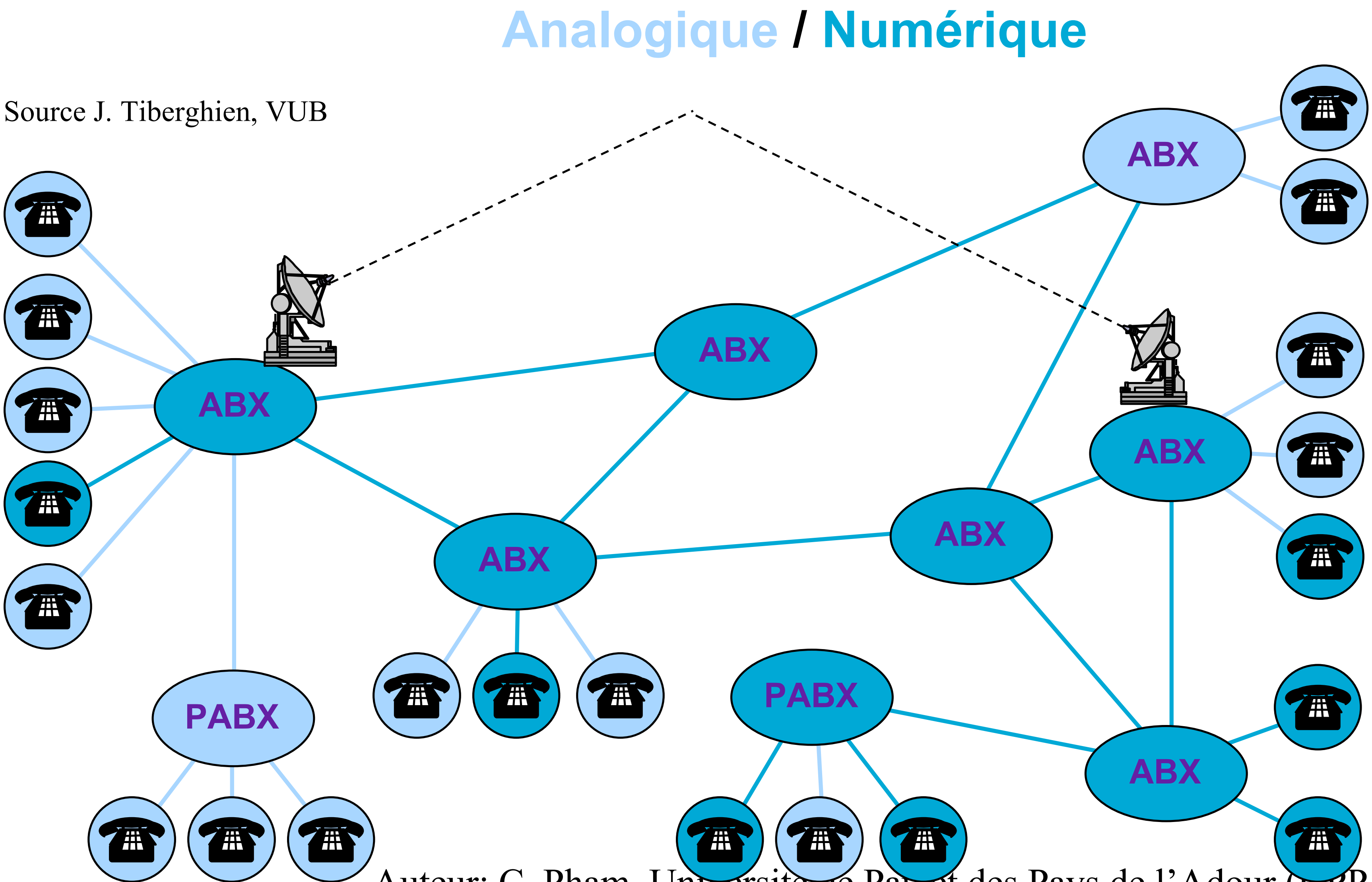
Juniper

Operator's infrastructure

- Backbones are optical: OC48 (2.5Gbps), OC192 (10Gbps), OC768 (40Gbps) soon, usually over-provisioned
- POPs available worldwide

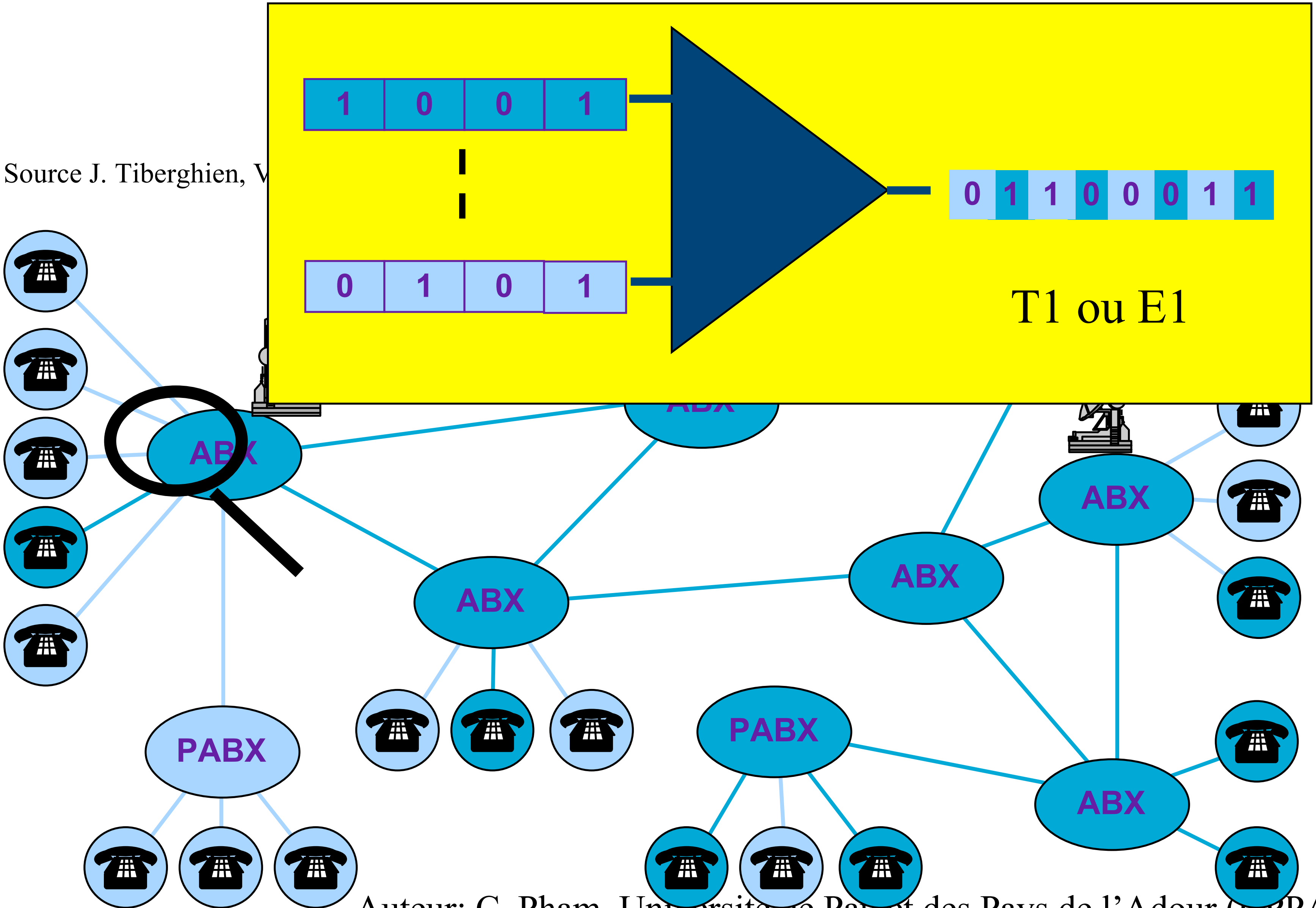


Back in time:
The telephone network, E.164 addressing



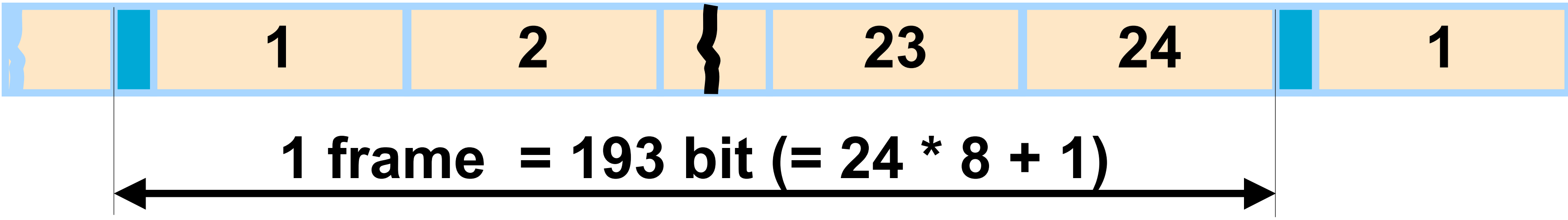
Close-up on accesses

Source J. Tiberghien, V



Bell D2 system (DS1/T1)

(CCITT G733)

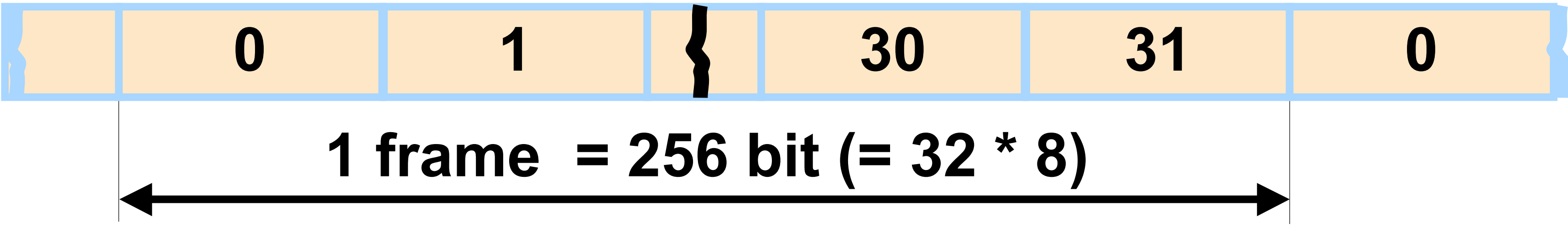


Frame duration :	**125 μS**	**= 1 / 8000**
Number of channels :	**24**	
Frame length :	**193 bit**	**= 8 * 24 + 1**
Bit frequency :	**1544 kHz**	**= 193 / 125.10^{-6}**
Signaling :	**least significant bit stolen once every 6 frames**	
Signaling rate :	**1.3 kb/s**	**= 8000 / 6**

Frame synchronization by bit 0

CEPT 30 (E1)

(CCITT G732)



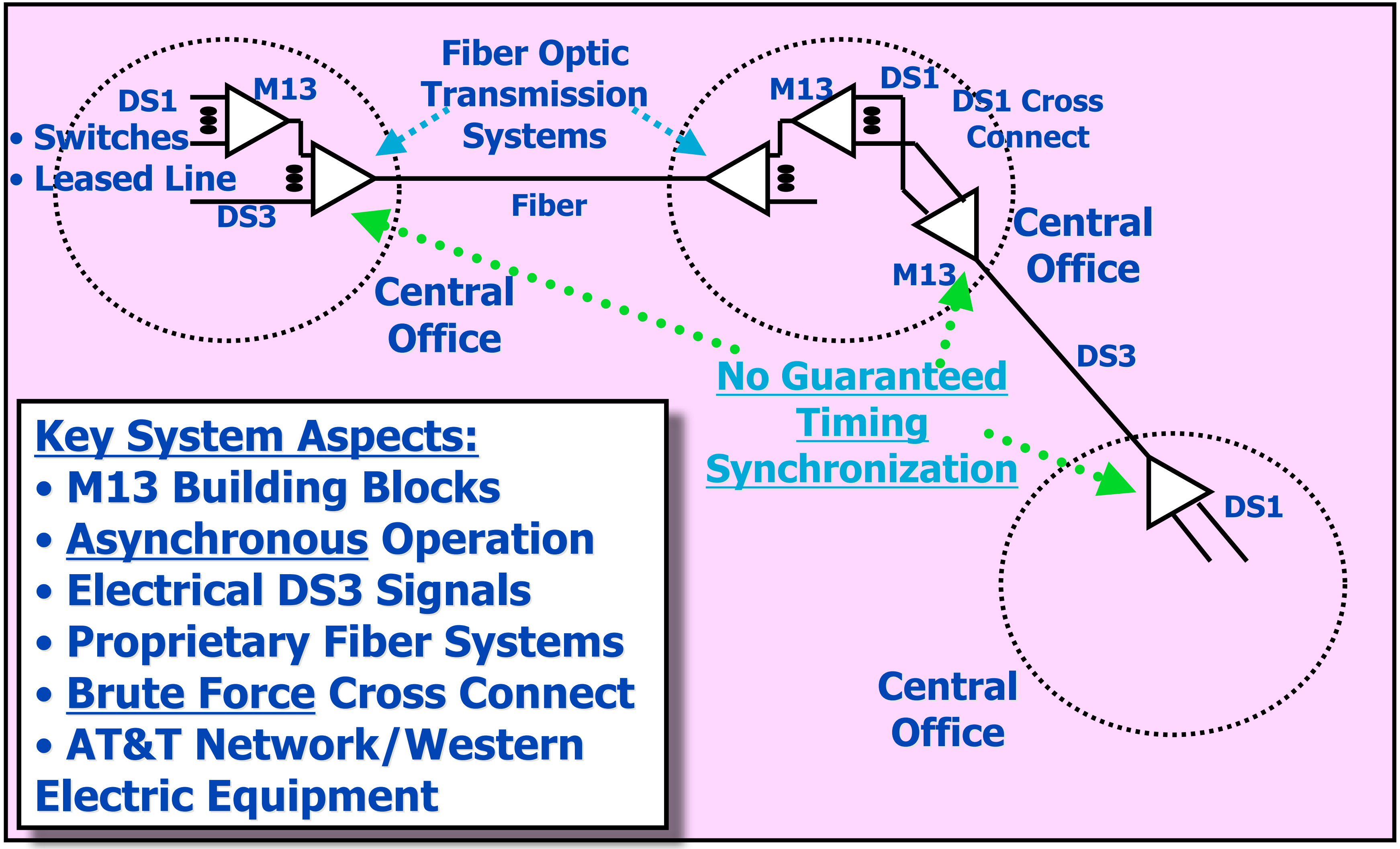
Frame duration :	**125 µS = 1 / 8000**
Number of channels :	**30**
Frame length :	**256 bit = 8 * 32**
Bit frequency :	**2048 kHz = 256 / 125.10^{-6}**
Signaling :	**Slot 16 reserved**
Channel Signaling : 2 kb/s	**= 64 / 32 kb/s**
Common Signaling :	**64 kb/s**

Frame sync. and link management by slot 0

T1/E1 represents the main underlying technologies for dedicated leased lines: Transfix in France for example

Bandwidth from 2400bps to 45Mbps (T3 in US)

Digital Telephony in 1984



Key System Aspects:

- M13 Building Blocks
- Asynchronous Operation
- Electrical DS3 Signals
- Proprietary Fiber Systems
- Brute Force Cross Connect
- AT&T Network/Western Electric Equipment

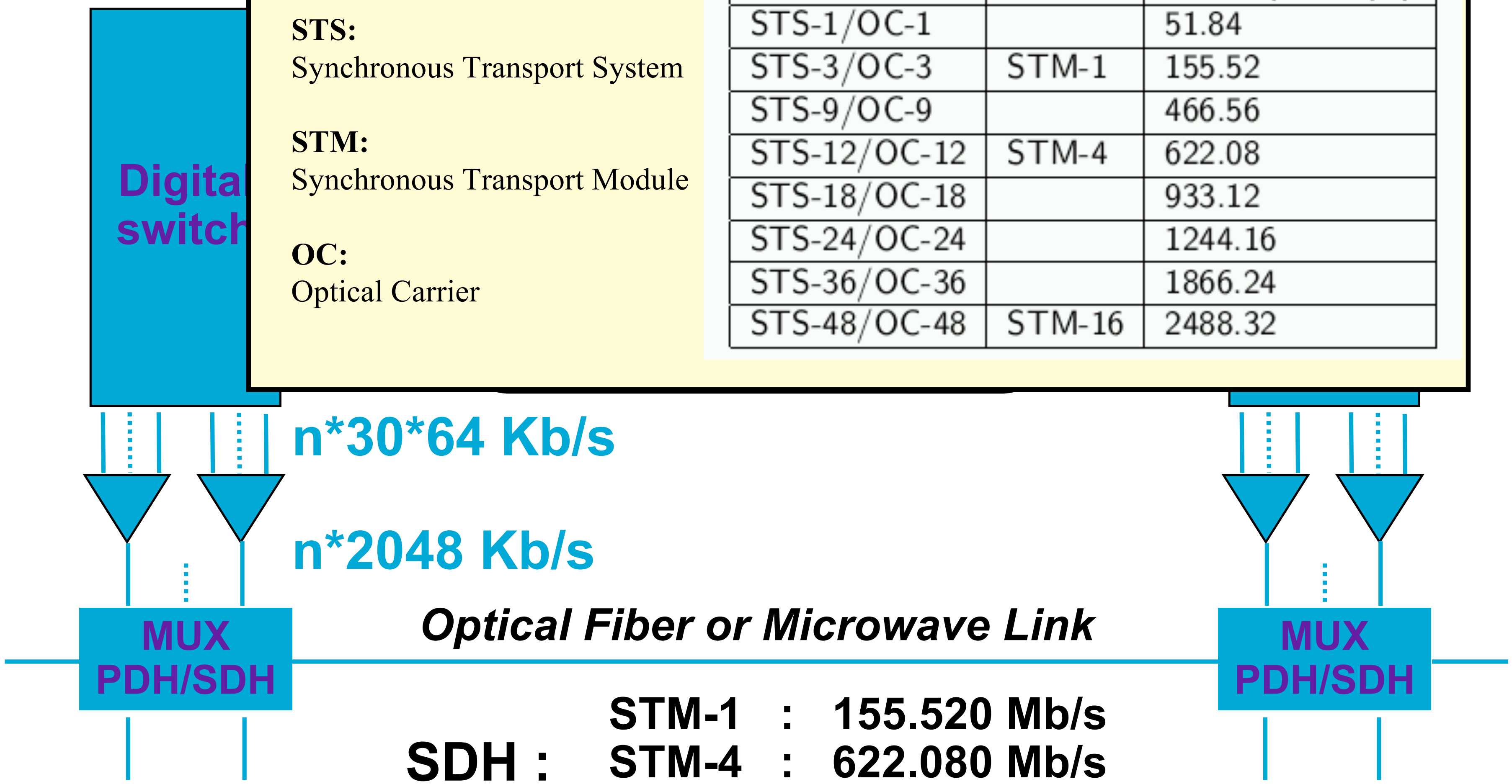
The core networks and SONET/SDH

STS:
Synchronous Transport System

STM:
Synchronous Transport Module

OC:
Optical Carrier

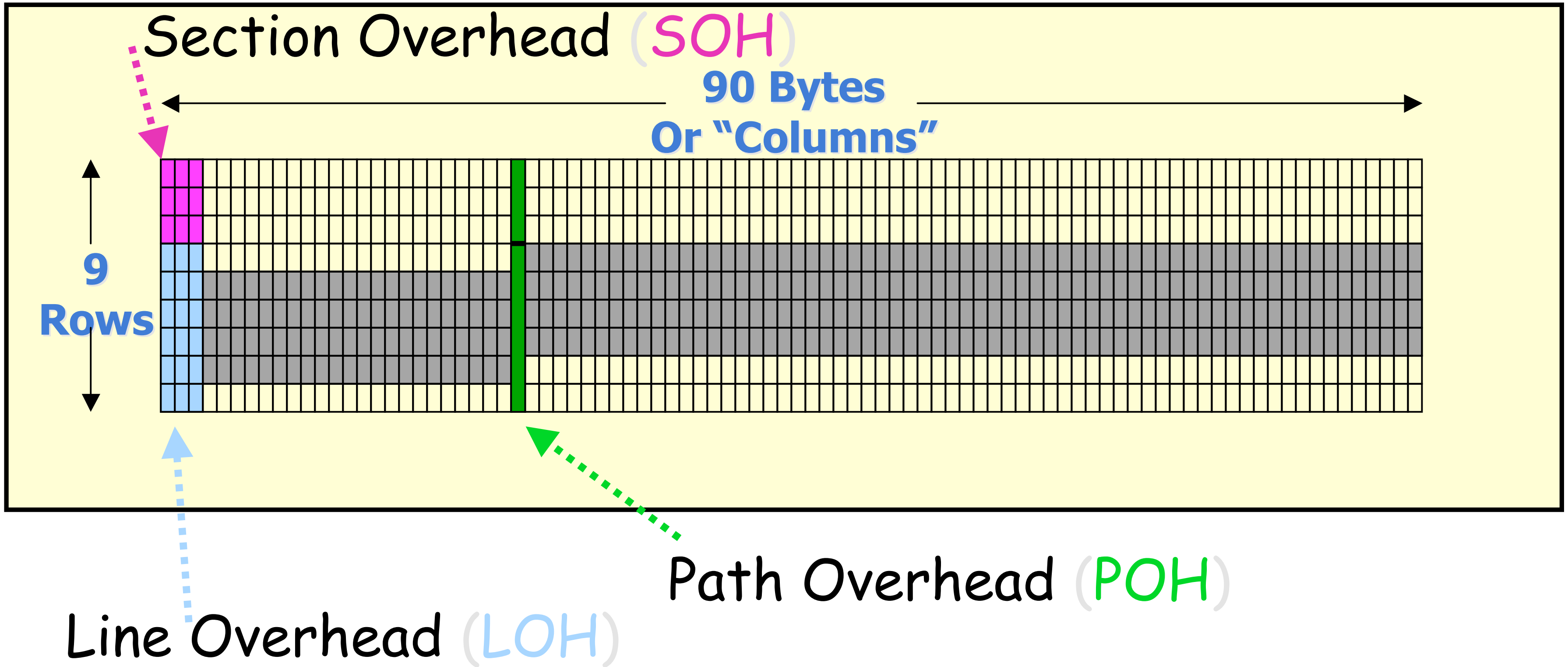
SONET	SDH	Débit (Mbits/s)
STS-1/OC-1		51.84
STS-3/OC-3	STM-1	155.52
STS-9/OC-9		466.56
STS-12/OC-12	STM-4	622.08
STS-18/OC-18		933.12
STS-24/OC-24		1244.16
STS-36/OC-36		1866.24
STS-48/OC-48	STM-16	2488.32



SDH :
STM-1 : 155.520 Mb/s
STM-4 : 622.080 Mb/s
STM-16 : 2488.320 Mb/s

The SONET frame

- **Basic frame length is 810 bytes**
  - Sent every 125us, raw throughput of 51.84 Mbits/s (STS-1)
  - Better seen as a block with 90 colomns and 9 lines
  - SDH has STM-1 which corresponds to an STS-3

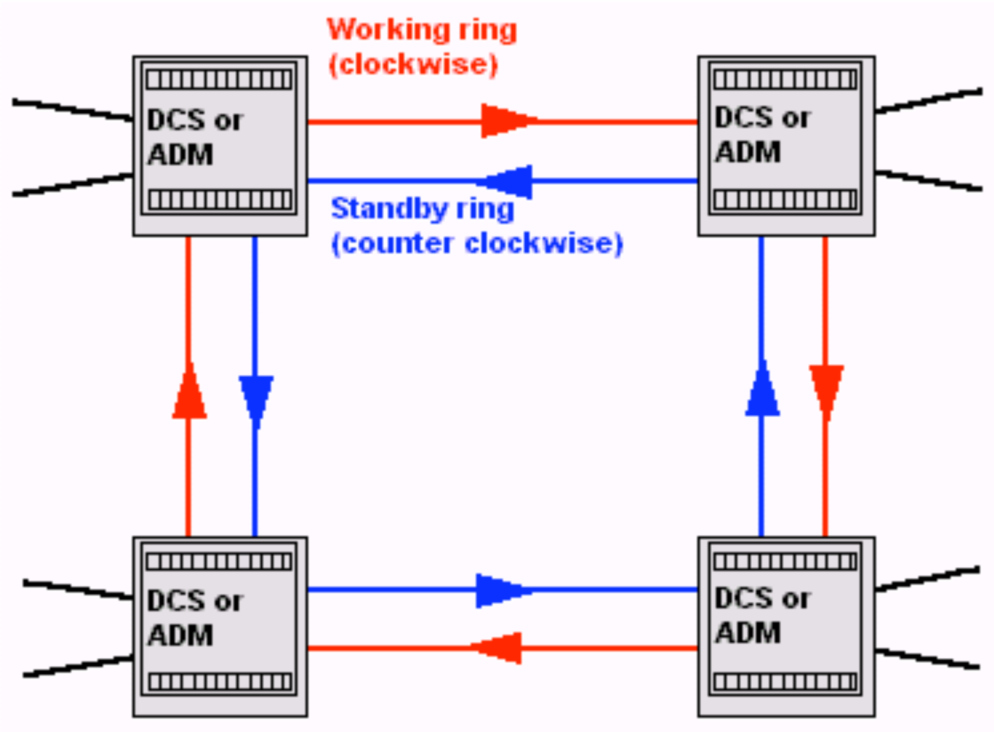


SONET/SDH and resiliancy

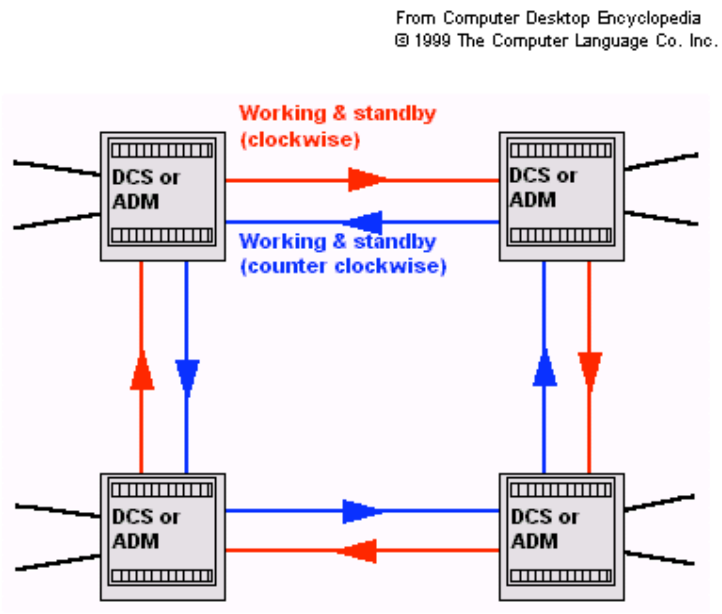
- SONET/SDH has built-in fault-tolerant features with multiple rings
- Ex: simple case

From Computer Desktop Encyclopedia

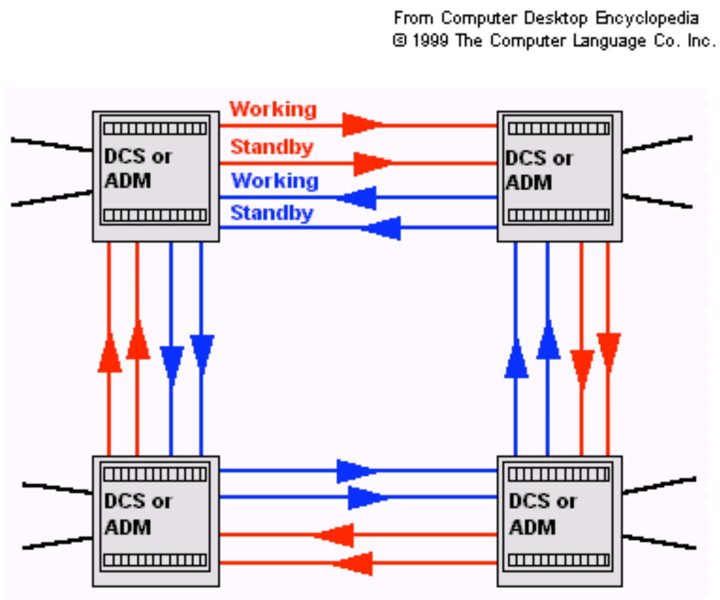
DCS
(Digital Cross-Connects)



SONET/SDH and resiliancy



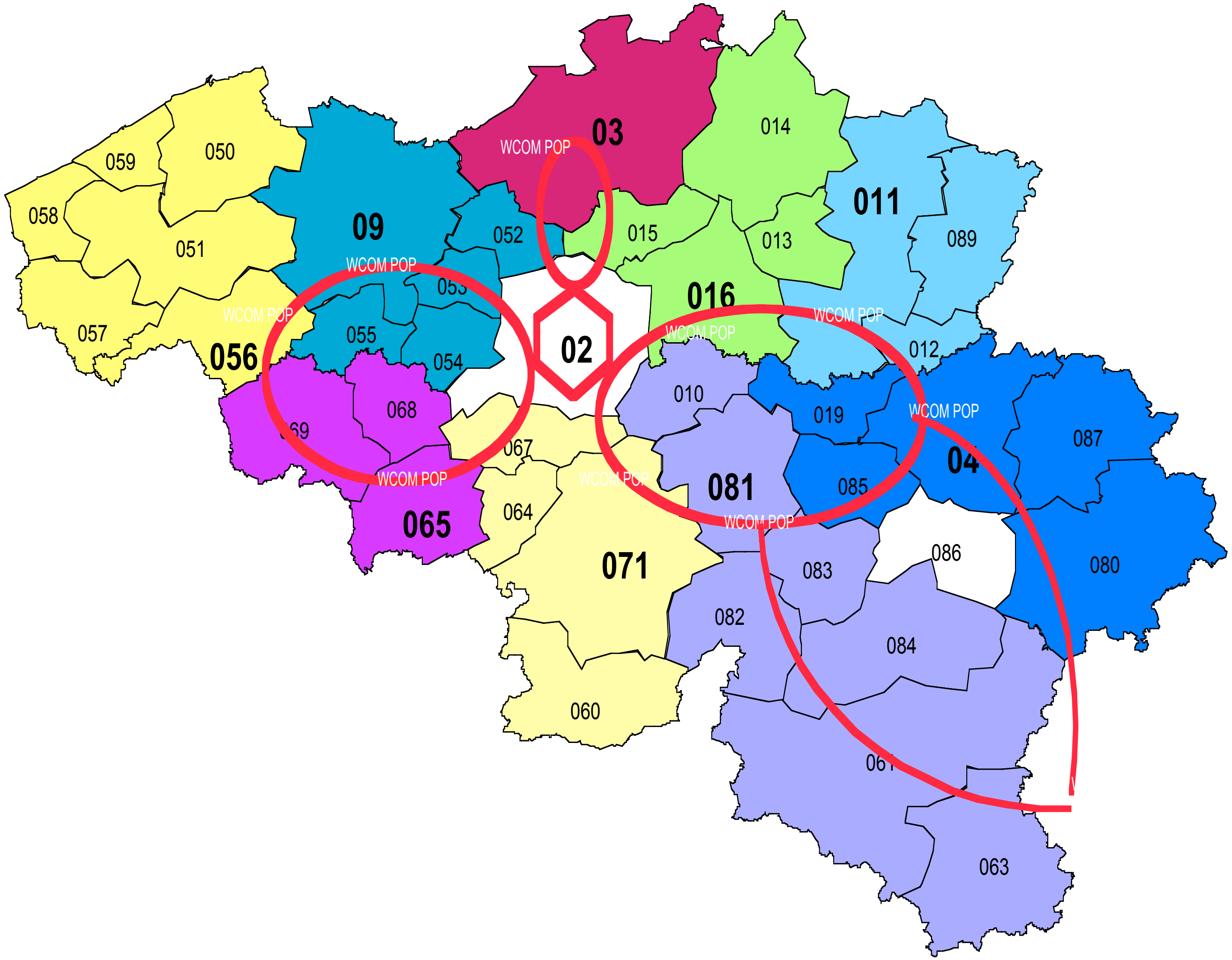
bi-directional



Found in most operators

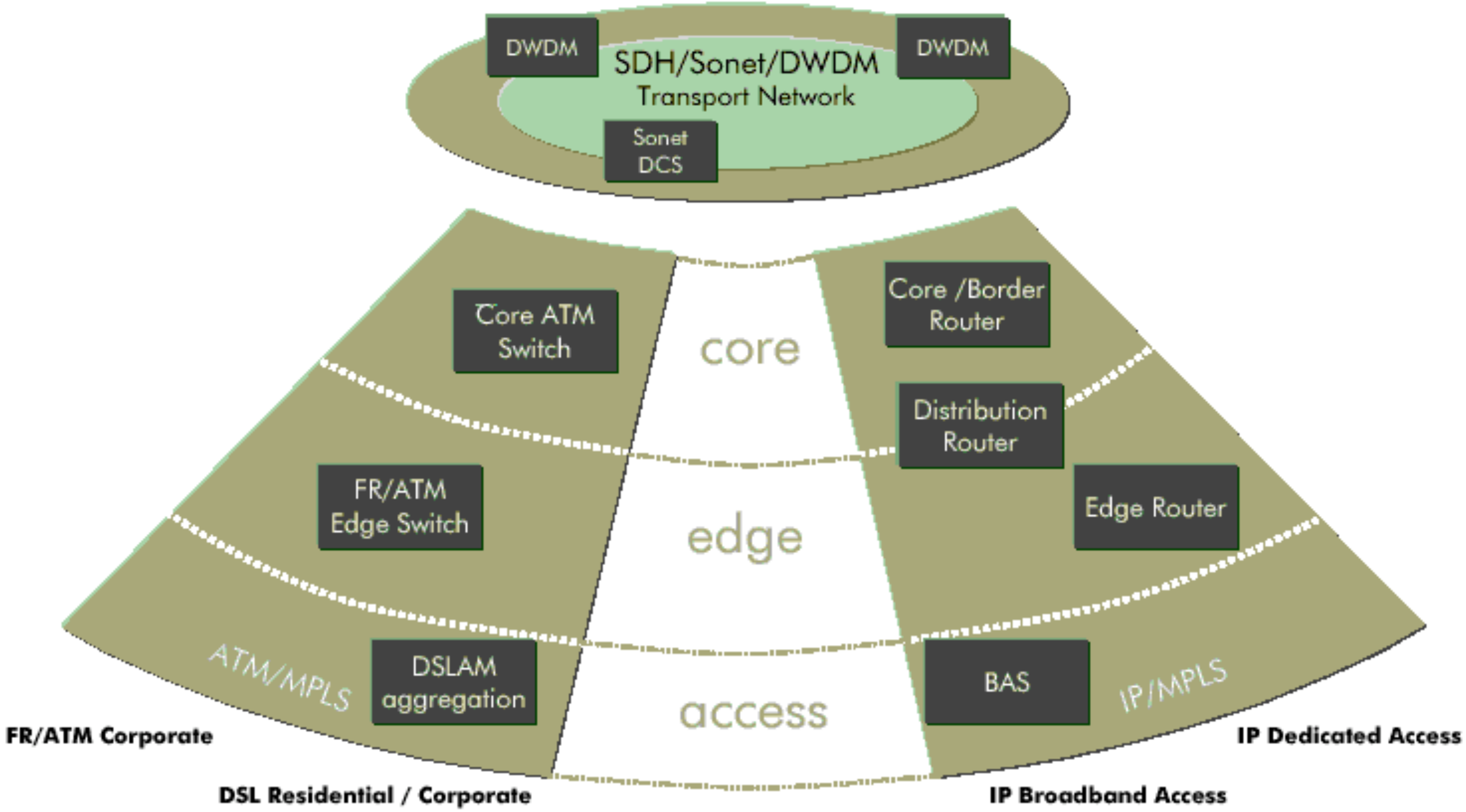
SDH Rings

The Worldcom Belgian Network



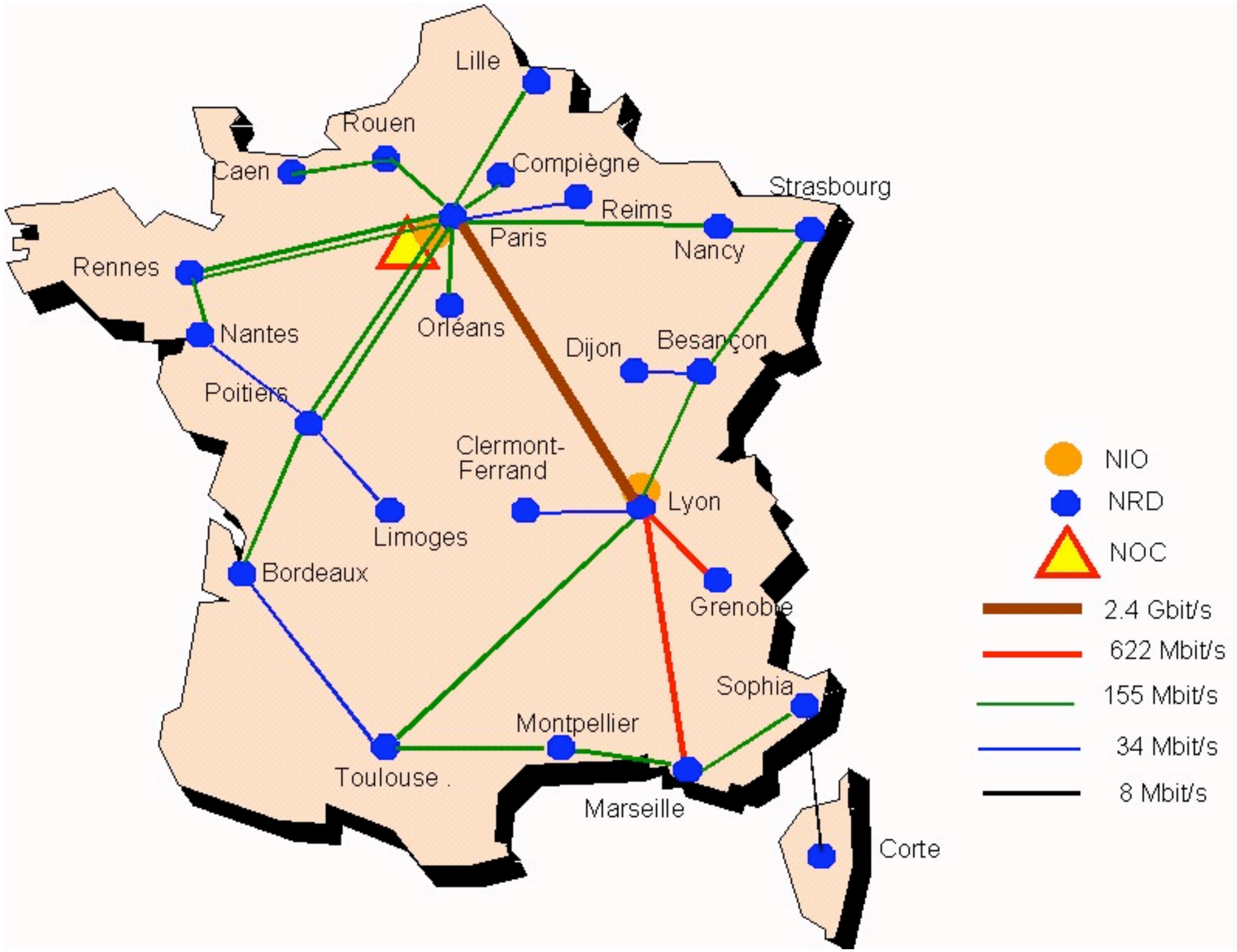
Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

Synthesis of networking technologies



source Alcatel

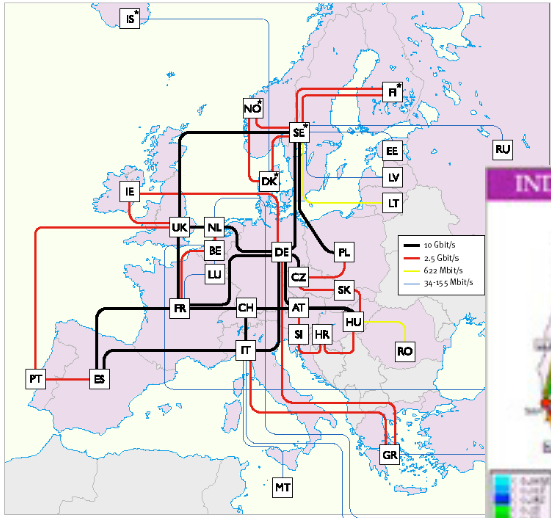
RENATER 2bis



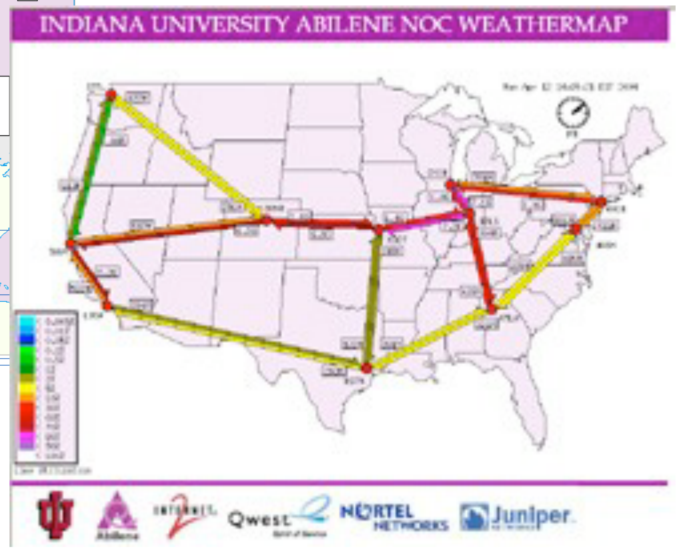
Auteur: C. Pham, Université de Pau et des Pays de l’Adour (UPPA)

The new networks

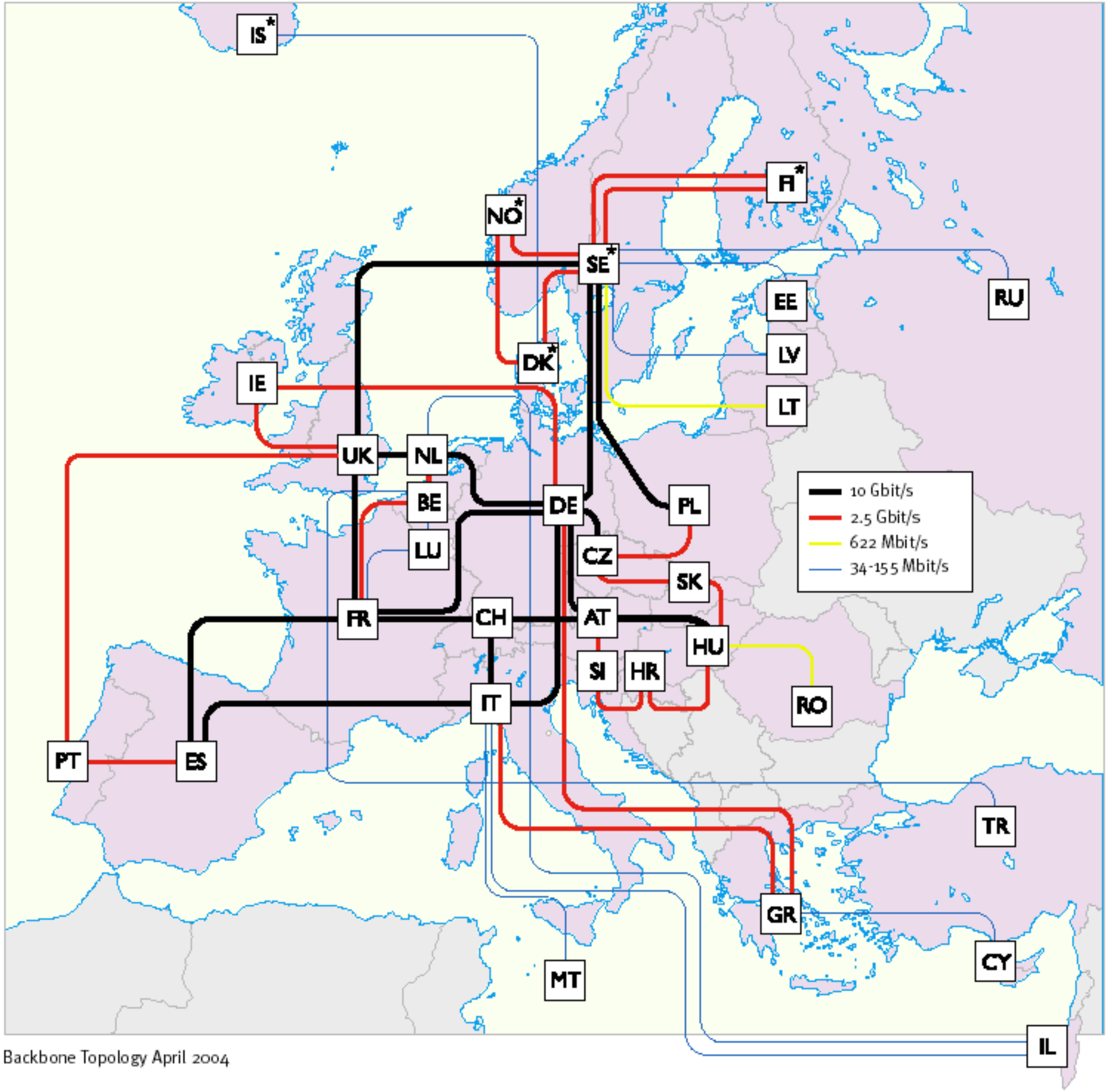
- vBNS
- Abilene
- SUPERNET
- DREN
- CA*NET
- GEANT
- DATATAG
- …much more to come!



Backbone Topology April 2004



GEANT



Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

NEXT GENERATION NETWORKS (NGN)

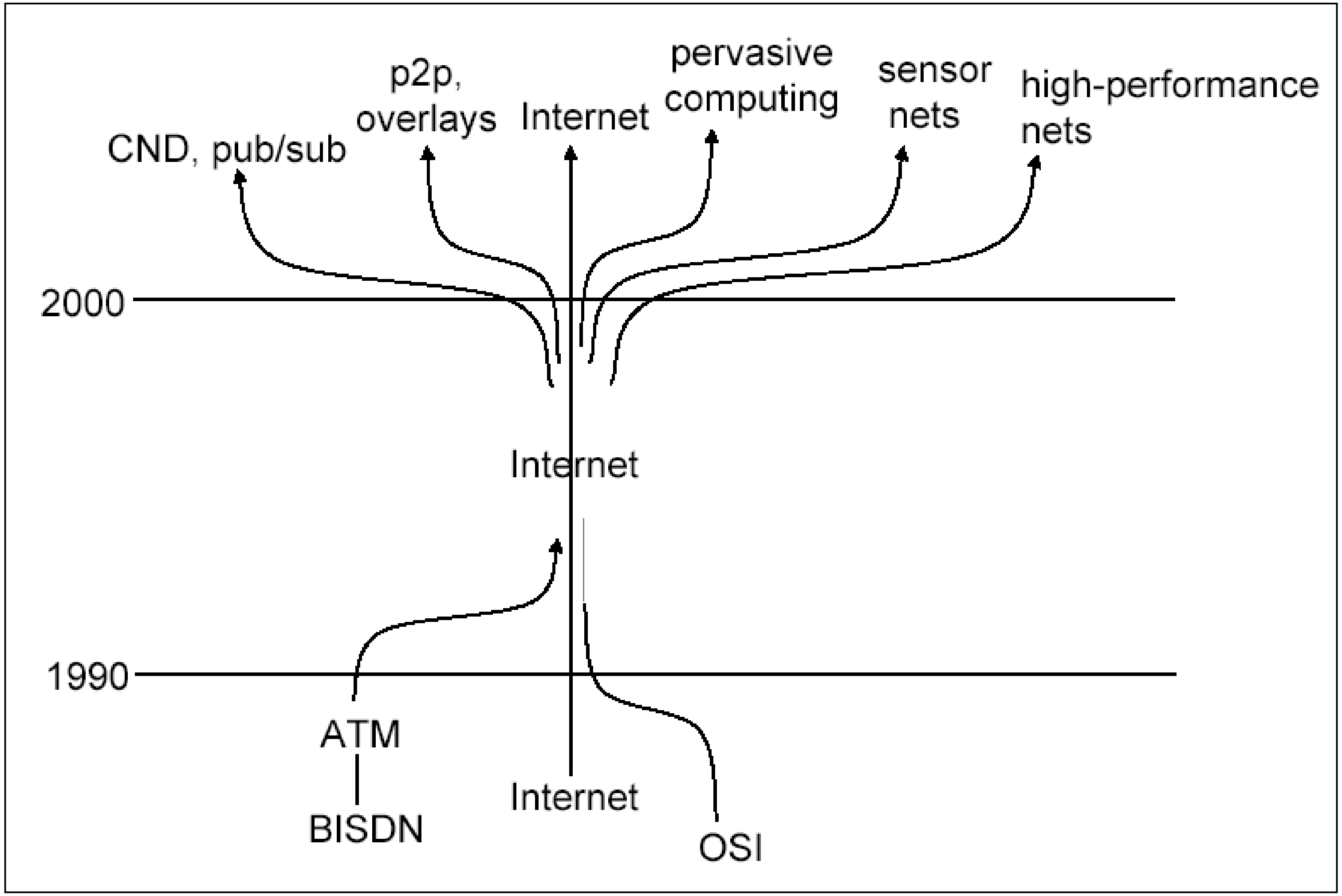
Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

Definition

- One definition for NGN could be
  - « ***The next-generation network seamlessly blends the public switched telephone network (PSTN) and the public switched data network (PSDN), creating a single multiservice network. Rather than large, centralized, proprietary switch infrastructures, this next-generation architecture pushes central-office (CO) functionality to the edge of the network. The result is a distributed network infrastructure that leverages new, open technologies to reduce the cost of market entry dramatically, increase flexibility, and accommodate both circuit-switched voice and packet-switched data.*** » **which is proposed by the IEC (International Engineering Consortium) at** http://www.iec.org/online/tutorials/next_gen/

- **PURPOSE: reduce OPEX!**
- **KEYWORDS: convergence, quality of service, control management, reconfiguration**

Towards all IP

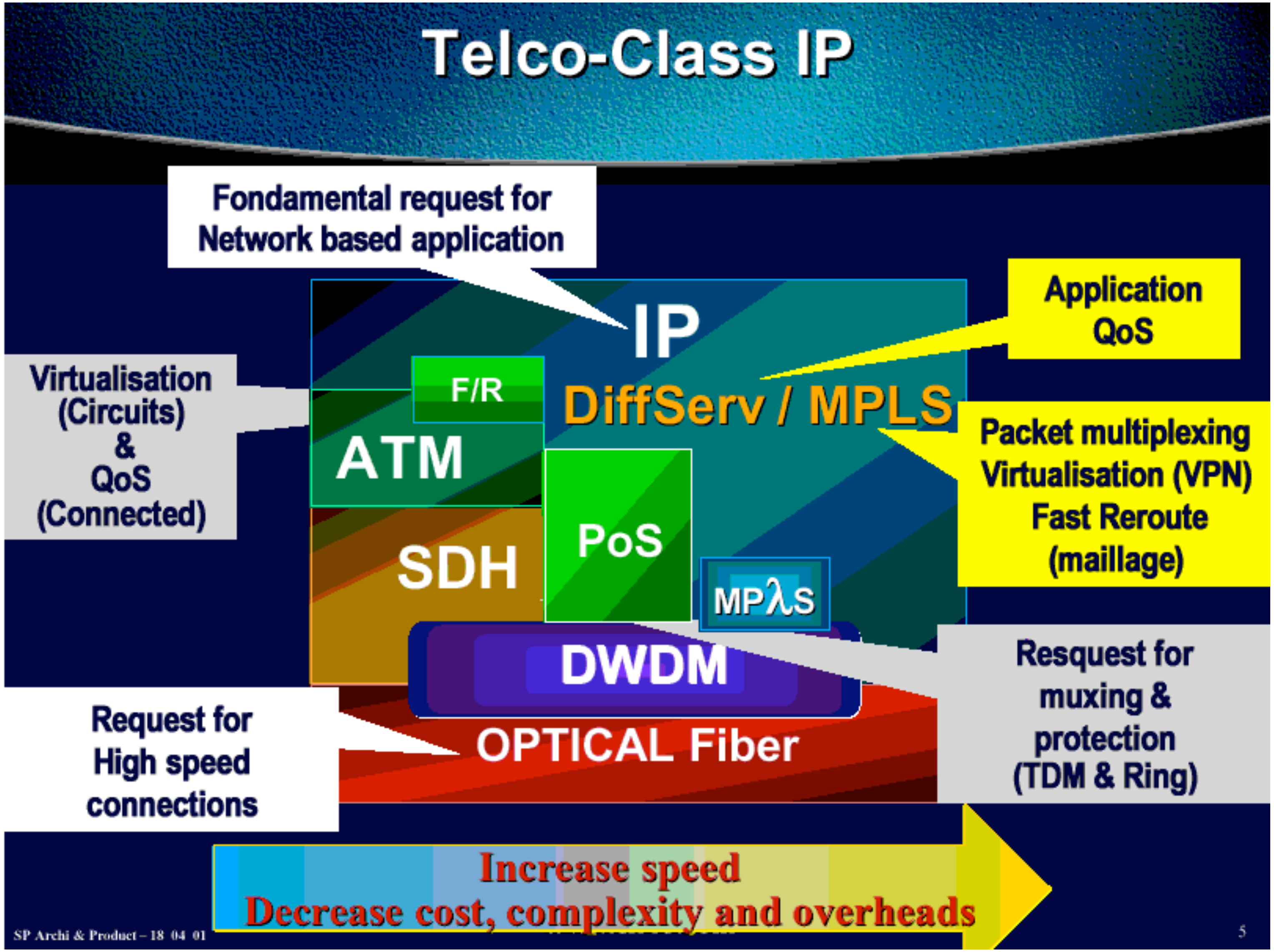


From Jim Kurose

A whole new world for IP

Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)

Towards all-IP next generation networks



Auteur: C. Pham, Université de Pau et des Pays de l'Adour (UPPA)