Current Status and Future Directions of Photonic Networks

Masayuki Murata
Cybermedia Center
Osaka University
e-mail: murata@cmc.osaka-u.ac.jp
http://www.anarg.jp/

Advanced Network Architecture Research

Advanced Network Architecture Research

Contents

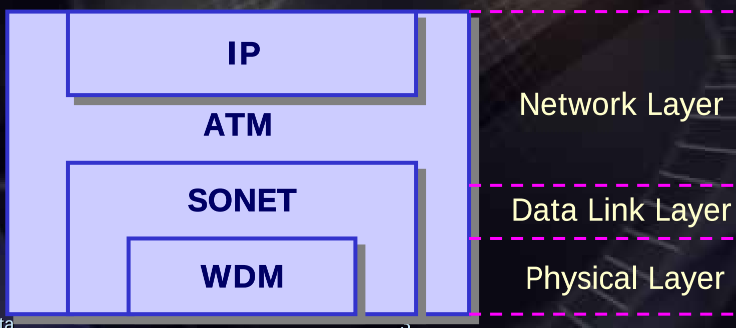
- Why We Need IP over WDM?
- Four Architectures of IP over WDM
  - WDM link
  - Lightpath network (so called, MPλS)
  - Photonic burst switching
  - Photonic packet switching
- Roadmap to Photonic Network
- Possibility of New Photonic Network Architecture

Osaka University

Advanced Network Architecture Research

Several Views for Photonic Internet

- IP over ATM over SONET over WDM
- IP over SONET over WDM
- IP over (PPP or HDLC over) WDM



Advanced Network Architecture Research

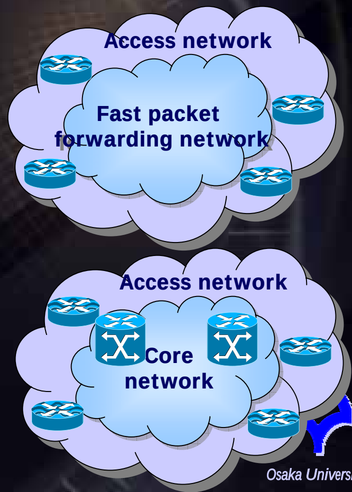
Multi-Layer Protocol Stack is Harmful

- Duplicated Functionalities
  - Same functions are possibly spread in two or more layers
  - Not easy to optimize the function distributed over multiple layers
  - We still have a possibility of functional partitioning
    - Routing Protocol in IP and WDM Layers
    - Reliability Control in IP and WDM Layers
- Inefficiency
  - IP over ATM over SONET over WDM network
    For 40 Byte IP packet, two ATM cells (106 Byte) are required

Advanced Network Architecture Research

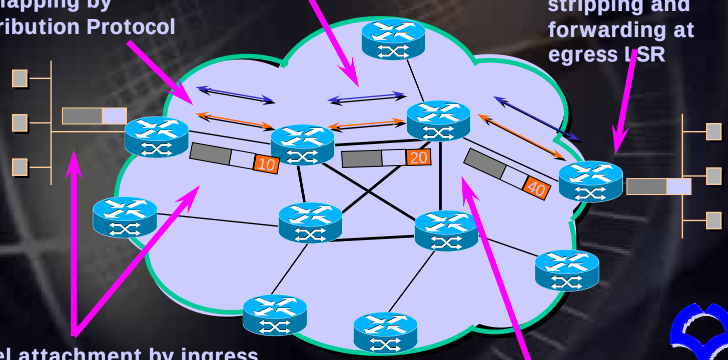
Partitioning Control Plane for Packet Switching and Forwarding

- GMPLS: Generalized MultiProtocol Label Switching
  - Packet Switching over Circuit Switching
  - Separates packet forwarding and switching
  - Provides a circuit between an edge node pair
- Cf. MPLS
  - Traffic engineering, QoS routing
- Cf. Legacy Model
  - Fast lookup of routing tables (longest prefix matching)
  - Parallel/pipeline processing



Advanced Network Architecture Research

Functions of MPLS

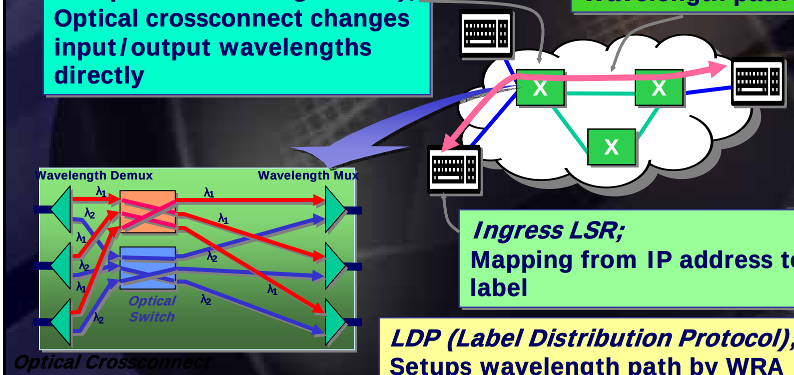


Advanced Network Architecture Research

Mapping from MPLS to GMPLS

LSR (Label Switching Router);
Optical crossconnect changes input/output wavelengths directly

LSP (Label-Switched Path);
Wavelength path (lightpath)



Ingress LSR;
Mapping from IP address to label

LDP (Label Distribution Protocol);
Setups wavelength path by WRA

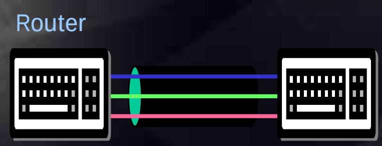
Advanced Network Architecture Research

Choice of Photonic Network Architecture

- Four alternatives
  1. WDM Link Network
     - Connects adjacent routers by a WDM link
  2. WDM Lightpath Network
     - Connects an edge node pair by a wavelength path (lightpath)
  3. Photonic Burst Switching Network
     - Establishes an on-demand path for burst
  4. Photonic Packet Switching Network
     - Directly handles IP packets for switching/forwarding

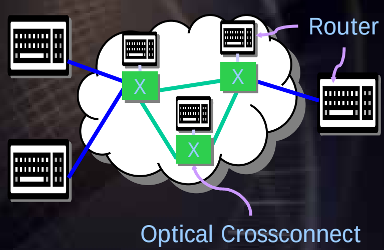
Photonic Network Architecture (1) WDM Link Network

- Connects adjacent routers by multiple wavelengths on the fiber
- Capacity is increased by the # of wavelengths
- × Simply moves bottleneck to electronic routers



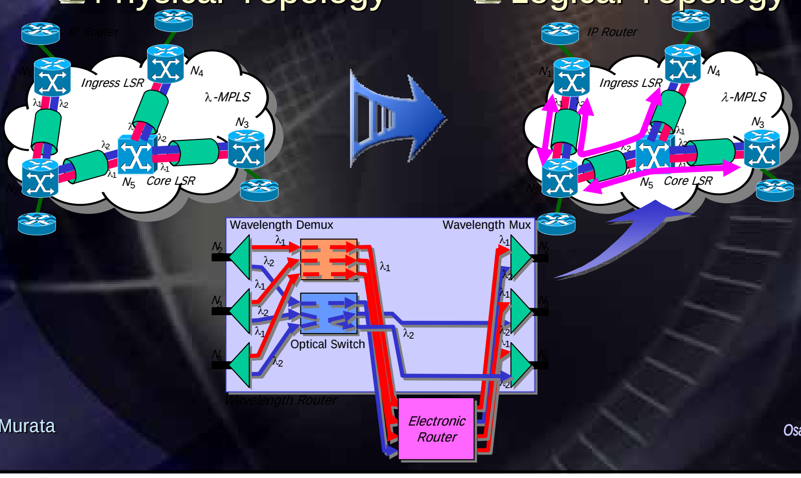
Photonic Network Architecture (2) Lightpath-based Network

- Logical topology based on wavelength paths
  - Cuts through nodes within the network by wavelength paths to resolve router bottleneck
  - Option: routing within the network by IP router attached to the crossconnect
  - RWA (Routing and Wavelength Assignment) Problem
- × Bottleneck at Ingress LSR
  - Can alleviate it by WDM ring to distribute the processing load
- × Granularity of Labels: wavelength
  - Difficult to realize label merging/splitting and four-layer switching



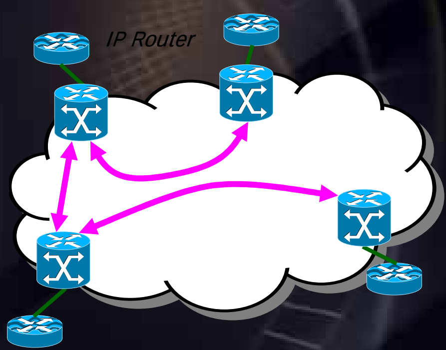
Logical Topology by Wavelength Routing

- Physical Topology
- Logical Topology



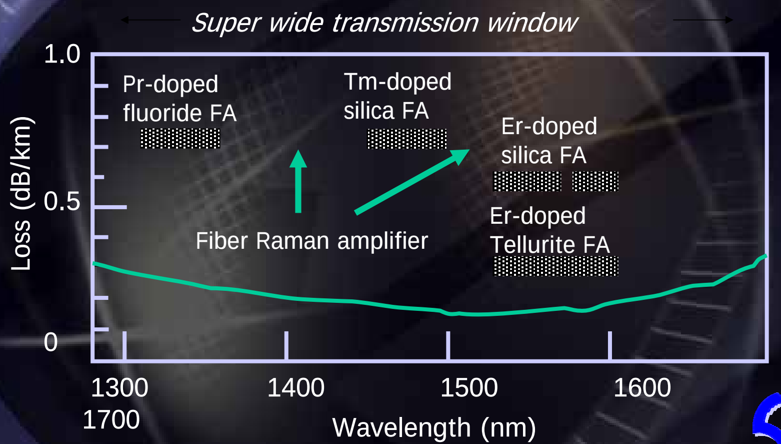
Logical View Provided to IP

- Redundant Network with Large Degrees
1. A smaller number of hop-counts between end-nodes
2. Decreases load for packet forwarding at the router
3. Relieves bottleneck at the core routers



Advanced Network Architecture Research

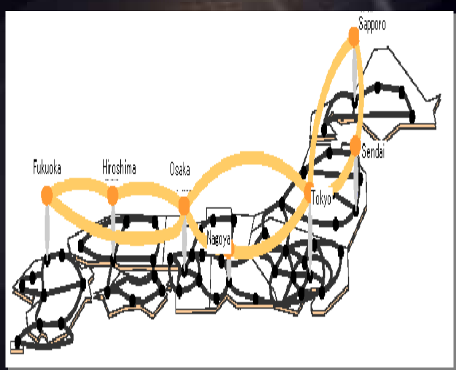
Possibility of 1000-Wavelength Division Multiplexing?



Advanced Network Architecture Research

Relation between Bandwidth and # of Wavelengths/Distance

- 50THz in Total
  - 20 GHz
    - 2.5 GBps,
      2500 Wavelengths
      6,400Km
  - 50 GHz
    - 10 Gbps
      1,000 Wavelengths
      400Km
  - 90 GHz
    - 40 Gbps
      556 Wavelengths
      25Km



http://www.ntt.co.jp/databook/setubi/

Advanced Network Architecture Research

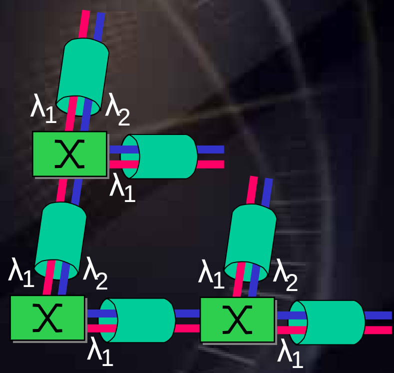
Effects of WDM with 1000 Wavelengths

- Given
  - Physical topology
  - Traffic demands, Traffic matrix
- Derives bounds on logical topology from given physical topology
  - The required # of wavelengths to satisfy the degree of logical topology
  - The traffic volume at the router for a given degree of logical topology
- Required # of wavelengths, Packet processing capability at the router
- Numerical Examples
  - NTT's Backbone
    - Min degree: 2, Max degree: 9, Max hop count; 9
  - Traffic Matrix from NTT's Telephone Network
    - 30 Gbps in Total (Increased by scale factor α)
    - http://www.ntt-east.co.jp/info-st/network/traffic/index.html

Advanced Network Architecture Research

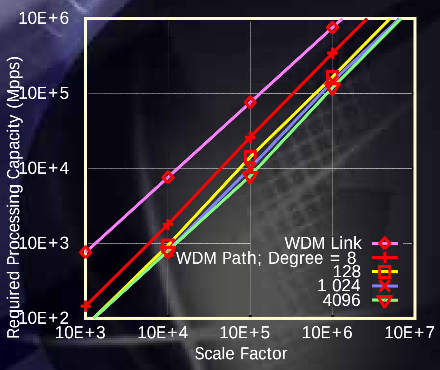
Derivations of Bounds

- Required # of wavelengths Δ:
  - Degree of physical topology: Δ_p
  - Degree of logical topology: Δ_l
  - $\Delta > \Delta_l / \Delta_p$
- Determine the traffic volume at the router from the degree of the logical topology
  - The # of one-hop paths: Δ_l
    The # of two-hop paths: Δ_l^2
    The # of three-hop paths: Δ_l^3
  - Assuming the shortest-path routing, determine the transit traffic
  - Get the sum of traffic on the path terminating at the router [bps] and translate it to [pps]

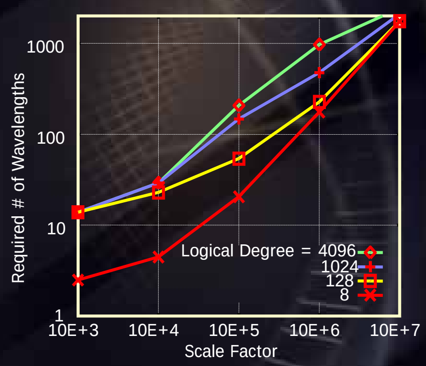


Advanced Network Architecture Research

The Case of NTT Backbone Network



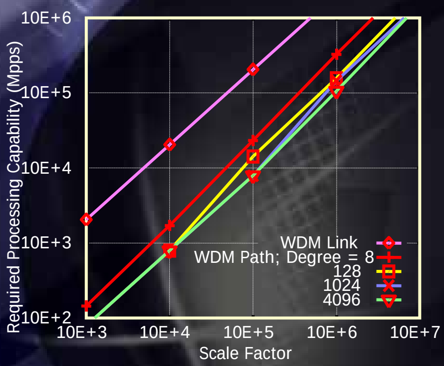
Traffic Volume at the Router



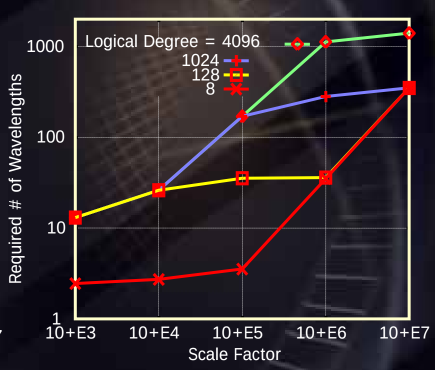
The Required # of Wavelengths

Advanced Network Architecture Research

Homogeneous Traffic Case



Traffic Volume at the Router

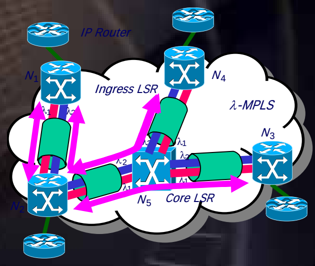


The Required # of Wavelengths

Advanced Network Architecture Research

Problems for Realizing Lightpath Network

- Functional Partitioning
  - IP routing and WDM routing
  - Quality of Protection
- Design of Lightpath Network
  - Same difficulty as an existing traffic engineering approach
    - Topology optimization for given traffic demand
  - We need incremental lighpath establishment meeting unpredictable traffic growth
- Path establishment based on GMPLS
  - Centralized vs. distributed
  - It should be chosen by on-demand or not



Advanced Network Architecture Research

Do We Need More “Intelligent” WDM Network?

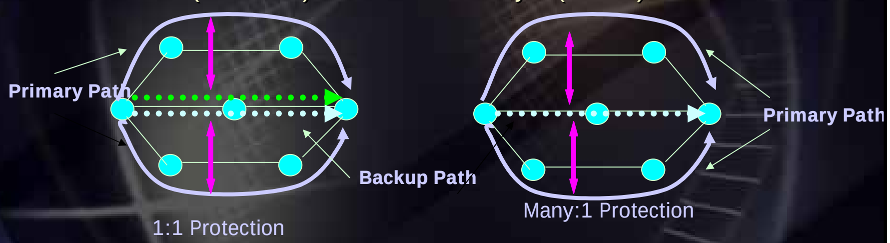
- WDM network itself has network control capabilities
  - Routing function
    - IP also has it!
  - Congestion control function
    - TCP also has it!
    - TCP over ATM (ABR service class) is difficult to work well Parameter tuning of control parameters in ABR is not easy
  - Connection establishment
    - IP is connectionless
    - Multimedia application does not require 10Gbps channel
    - Functional Partitioning vs. Multi-layered Functionalities?
- Important is reliability

Functional Partitioning between IP and WDM?

- Reliability functionalities offered by two layers
  - IP Layer: Routing
  - WDM Layer: Path Protection and Restoration
- WDM should provide its high-reliability mechanism to IP
  - Protection mechanisms
    - link protection
    - dedicated-path protection
    - shared-path protection

WDM Protection

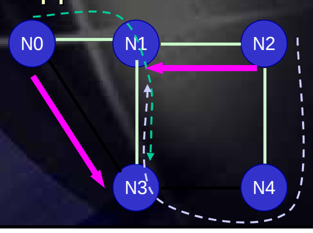
- Immediately switch to backup path on failure of nodes/links
  - In the order of 10ms
- 1:1 (dedicated) Protection vs. Many:1 (shared) Protection

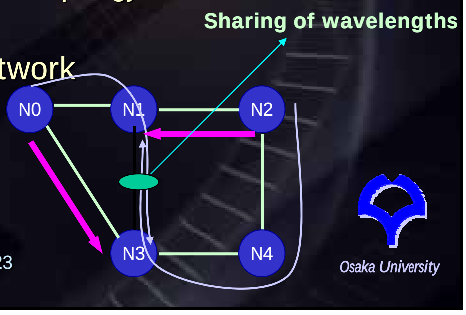


- Protection technique suitable to IP over WDM network?
  - IP has its own protection mechanism (i.e., routing) while it is slow
  - We want an effective usage of wavelengths
  - Many:1 protection is reasonable

Reliability Design Problem for Many-to-1 Path Protection

- Objective
  - Minimize the utilized wavelengths in total
- Given Conditions
  - The number of wavelengths on the fiber
  - Physical topology and logical topology
  - Primary routes
- Application to 5-node network





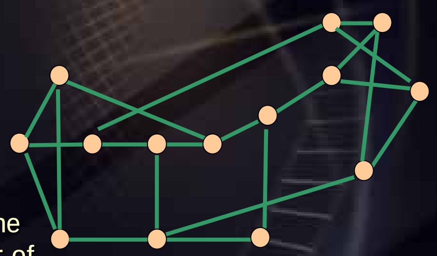
Application to Large-Scaled Network

- Previous formulation is MILP (Mixed Integer Linear Problem); difficult to be applied to the large-scaled network
- *Min-hop-first approach*
  - It is rare that lightpaths with a small number of hop counts contend the backup lightpaths
  - First protect the lightpaths with small hop-counts, and then protect lightpaths with large hop-counts using the remaining wavelengths
- *Largest-traffic-first approach*
  - Assign the paths in a descending order of the traffic loads
- *Random approach*
  - For only reference purpose

Advanced Network Architecture Research

Network Model

- NSFNET;
  - 14 node
  - 20 physical links
- Logical topology is first determined by MLDA
  - First set up the lightpaths between adjacent nodes
  - Set up the lightpaths for the path in a descending order of traffic volume



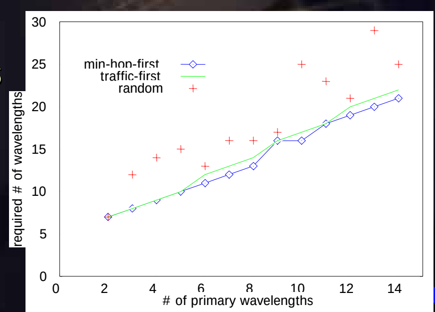
NSFNET

Osaka University

Advanced Network Architecture Research

Effect of Min-Hop-First Approach

- The required number of wavelengths to protect all lighpaths with
  - Random-selection method?
  - Traffic-large-first method?
  - Min-hop-first method?

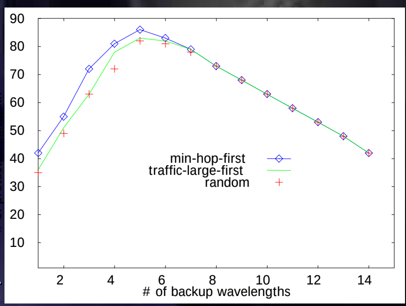


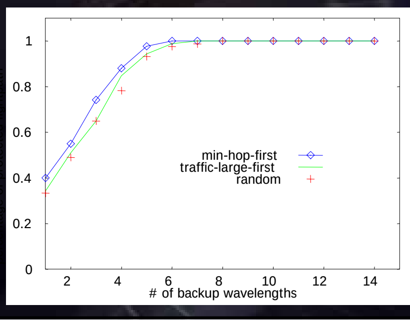
Osaka University

Advanced Network Architecture Research

Effect of Min-Hop-First Approach (Cont'd)

- Fix the wavelength on the fiber to be 16
- Change the number of wavelengths dedicated to primary lightpaths to compare the number of protected lightpaths

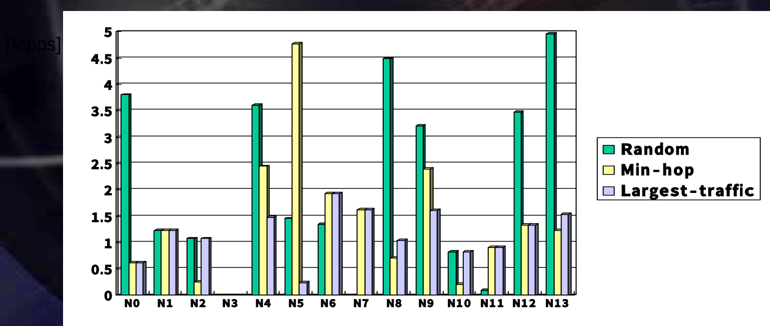




Advanced Network Architecture Research

Effect on IP Routers

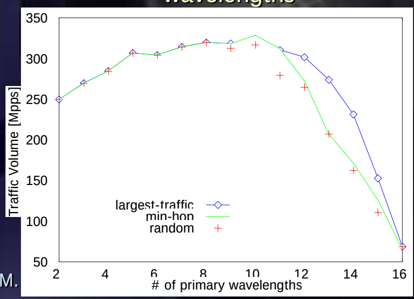
- A increased amount of traffic load at the IP router after the failure?
  - We choose the largest traffic increase among all possible cases for failures
- Primary wavelengths; 12, Backup wavelengths; 4
- Largest-traffic-first approach can decrease the traffic handled by the IP router

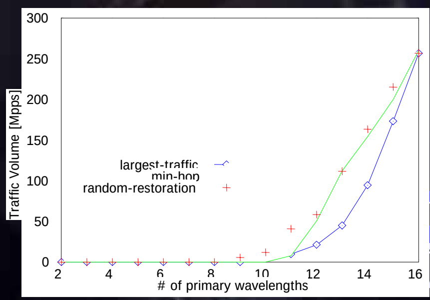


Advanced Network Architecture Research

Functional Partitioning in IP over WDM

- Total volume of the traffic **protected / non-protected** by backup lightpaths **before** routing table upgrade
- Sum of total traffic volume for each fiber failure scenario
- Largest-traffic-first approach provides better performance
  - Lots of traffic is protected by a lower number of backup wavelengths

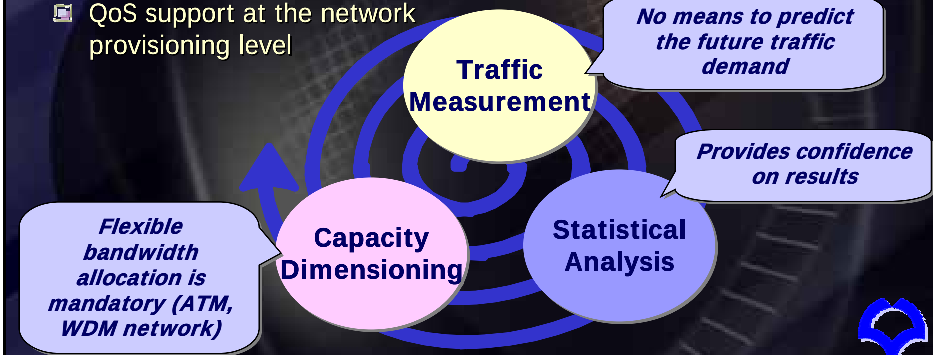




Advanced Network Architecture Research

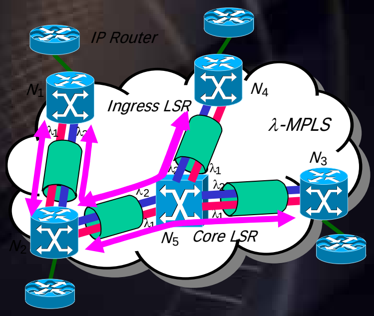
Spiral Approach for Network Dimensioning

- Incremental network dimensioning by feedback loop is an only way to meet QoS requirements of users
- QoS support at the network provisioning level



Incremental Network Dimensioning

- Initial Step
  - Topology design for given traffic demand
- Adjustment Step
  - Adds a wavelength path based on measurements of traffic by passive approach, and end-users' quality by active approach
  - Only adds/deletes wavelength paths; not re-design an entire topology
  - Backup paths can be used for assigning the primary paths
- Entirely Re-design Step
  - Topology re-design by considering an effective usage of entire wavelengths
  - Only a single path is changed at a time for traffic continuity



Routing and Wavelength Assignment (RWA) Problem

- WA: route is predefined and a "best" wavelength is selected on that route
  - Random, First-Fit
- RWA: Multi-path Routing
  - A "best" route is chosen in addition to the wavelength
  - Most-Used: a smallest number of wavelength available in the network is chosen, implying the centralized control

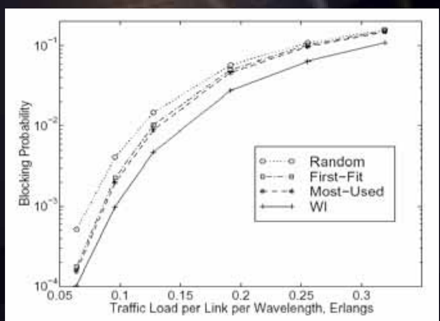


Figure 4 in E. Karasan , E. Ayanoglu. *Effects of Wavelength Routing and Selection Algorithms on Wavelength Conversion Gain in WM Optical Network,* ACM / IEEE Transactions on Networking, April 1998.

Distributing Lightpath Setup Control

- Which information should be distributed?
  1. Connectivity?
  2. Hop count?
  3. Traffic load status? (How many number of wavelengths are currently used?
  4. Which wavelengths are used?
- Example: information (1~3) is ditrbuted by routing protocol (OSPF), and wavelength assignment (4) is performed on demand basis
- Processing of control plane is performed in an optical domain
  - Wavelength assignment by photonic processing is realizable

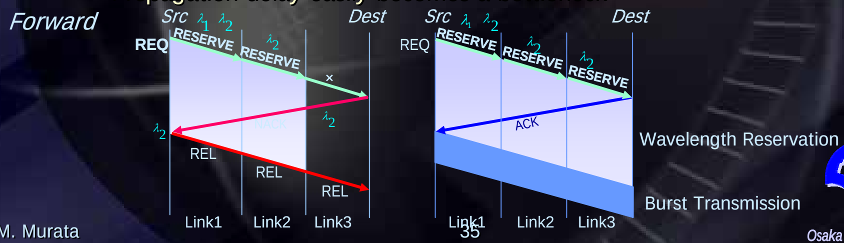
Photonic Internet Architecture (3)

- We have several burst switching in the past
  - Telephone network: TASI
    - Target service is voice; packet loss can be allowed
  - Packet Switching Network: Virtual Cut Through
    - If the output channel is not available, packet is stored in an electronic buffer
  - ATM : FRP (Fast Reservation Protocol)
    - Very large data transfer
    - ATM has a capability of flexible bandwidth usage
      When bandwidth is free, send the burst at 150Mbps. As the bandwidth is short, decrease the transmission speed as 75Mbps -> 37.5Mbps...
- WDM: OBS (Optical Burst Switching)
  - 10Gbps of channel capacity -> Granularity is very large
    - We can transmit DVD (4.7GB) within several seconds
    - The target is never real-time media
  - Bufferless
    - In other words, if the buffer is equipped with, we don't call it burst switching
- What is burst?
  - A large volume of data
  - If we support IP packet of short length, buffering at the send buffer is necessary, but TCP does not presume it

Advanced Network Architecture Research

OBS Protocol (1): Tell-and-Wait

- Two protocols based on ACK/NACK signals
  - Forward protocol; the wavelength is determined on the forward path
  - Backward protocol; the wavelength is determined on the backward path
- Wavelgnth is determined on the predefined path when the burst arrives
- ○ Buffering within the network is not necessary
  - An inherent difference from packet switching
- × Propagation delay easily becomes a bottleneck

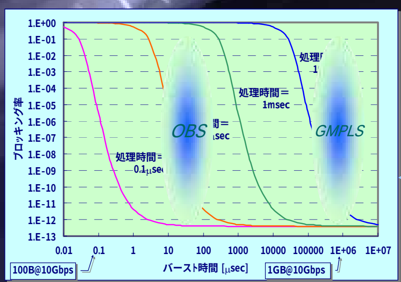


M. Murata

Osaka University

Advanced Network Architecture Research

Effect of Node Processing Time TAW vs. Circuit Switching

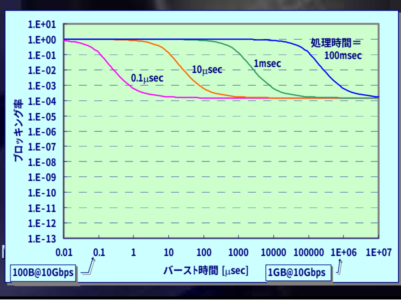


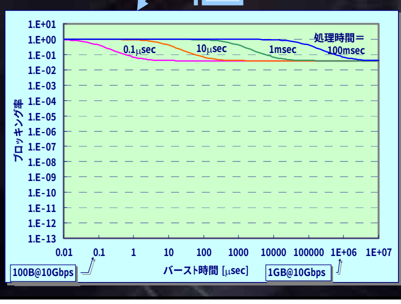
- TAW
  - The # of wavelengths: 32
  - Prop. Delay: 0
- The proc. Time should be smaller than burst length in one order of magnitude

Load on wavelength = 0.2

0.5

0.8

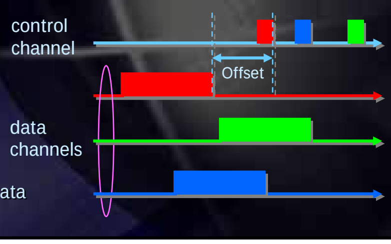


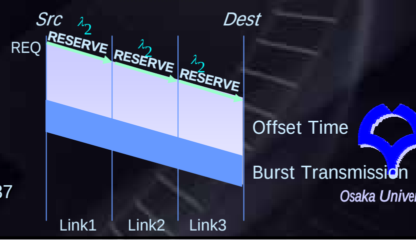


Osaka University

OBS Protocol (2): Tell-and-Go

- Eliminates the overhead due to prop. delay
- Payload should be delayed for header processing
- Header processing (path setup by wavelength reservation) is a key to high-speed burst tranfer
- × Burst may be dropped within the network
  - It is M/G/1/1 queueing system if wavelength change is not allowed





Burst Blocking Probability in OBS

- *b*: Burst length (Connection holding Time)
  s: Connection processing time
  p: Propagation delay
  W: The number of wavelengths
  λ: Burst arrival rate
- Total Traffic Load
  - TAW: $\rho = (b + s + p)\lambda$
  - TAG: $\rho = (b + s)\lambda$
- Traffic Load per Wavelength: $\rho_0 = \rho / W$
- Blocking Probability (M/G/W/W)

$$B = \frac{\rho^W / W!}{\sum_{n=0}^{W} \rho^n / n!}$$

Cause of Performance Degradation

- Increase of Hop Counts
  - Linearly increased
- Wavelength Change is not Allowed
  - Wavelength Continuity Problem
- Path is predetermined
  - WA (Wavelength Assignment) vs. RWA (Routing and Wavelength Assignment)
  - WA: Random, First-Fit
    - Distributed processing is possible
  - RWA : Multi-path Routing
    - Most-Used: Centralized control
  - Only wavelength assignment is possible in an optical domain
    - WA is promising

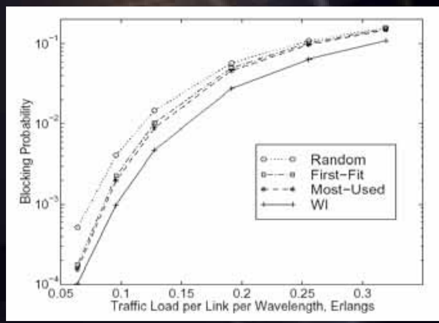
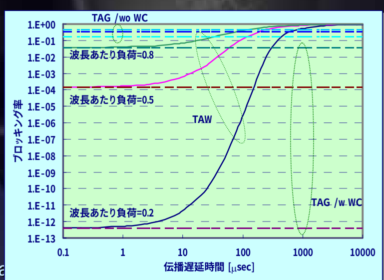
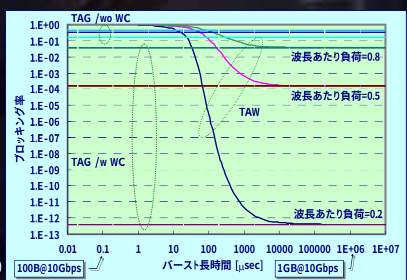


Figure 4 in E. Karasan , E. Ayanoglu. *Effects of Wavelength Routing and Selection Algorithms on Wavelength Conversion Gain in WM Optical Network*, ACM / IEEE Transactions on Networking, April 1998.

Tell-and-Wait vs. Tell-and-Go

- TAG (JET, JIT, ...)
  - Effects of propagation delay is eliminated
  - Instead, processing time at the node becomes large
  - If wavelength change is not allows, it is modeled as M/G/1/1 !
- The # of wavelengths = 32
- Burst length = 100 usec
  - Much larger than 10 usec of processing time





Advanced Network Architecture Research

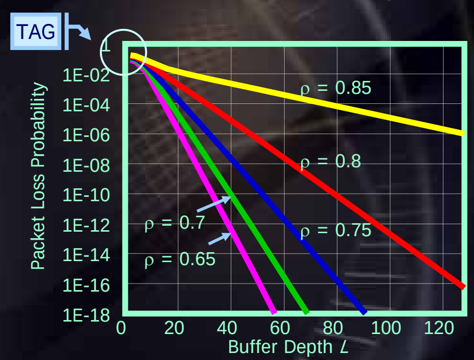
Problems of TAG

- If wavelength change is not allowed, the performance is considerably degraded
- It is only applicable to data transfer
- The burst may be lost during its transfer
- It must affect the upper-layer protocol
  - In general, lower-layer PDUs should be equal to or smaller than upper-layer PDUs

Advanced Network Architecture Research

Effect of Packet Buffering TAG vs. Packet Switching

- Packet Switching Presumes Packet Buffering
  - No wavelength changes; M/G/1/1+L
  - Wavelengh Changes; M/G/W/W+L
  - If FDL is equipped, some overhead is introduced to handle variable size packet
- Conditions
  - # of Wavelengths *W*=8
  - Wavelength changeable
- In principle, OBS can have “burst buffer”, but it requires very long FDL
  - 1Mbit burst = 100 υsec burst (@10Gbps) -> 20Km x *L*

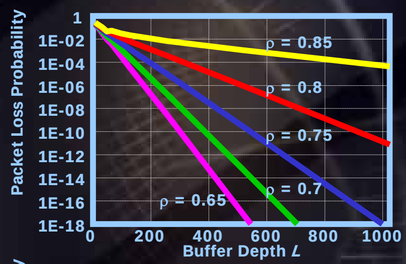


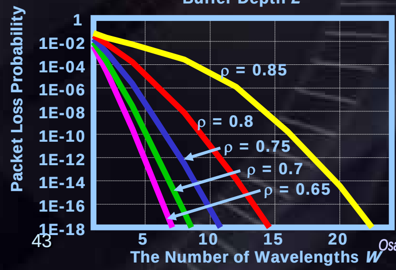
Masayuki Murata and Ken-ichi Kitayama, “Ultrafast photonic label switch for asynchronous packets of variable length,” *IEEE INFOCOM 2002*, June 2002.

Advanced Network Architecture Research

Advantage of Packet Switching

- No wavelngth changes
  - Even without wavelength changes, much performance improvement is achieved at the expense of large FDL buffers
  - 1Kbit packet = 0.1 μsec packet (@10Gbps) -> 20m x *L*
- The increase of # of wavelengths can much increase the performance
  - Wavelength interchangeable
  - Buffer size = 64





Advanced Network Architecture Research

Burst Switching is Actually Necessary?

- We should establish “Fast Circuit Switching” not “Burst Switching”
  - The technologies of TAW are applicable
- Remaining Issues
  - High-speed processing in control plane
    - Photonic processing
  - A Large Scale Network
    - Distributed processing of IP routing protocol is applicable

Photonic Internet Architecture (4)

- Photonic Packet Switching
  - Direct support of IP packets
    - Packet buffering by FDL
    - Asynchronous -> synchronous
    - Variable packets -> slotted time
  - Routing protocol based on GMPLS
  - Switching control in an electronic domain
  - Multiplexing effect is vary large

Problems of MPλS

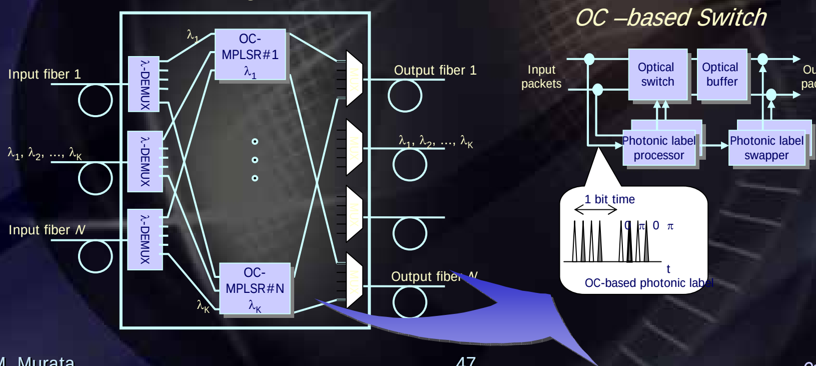
- Large granularity; wavelength
  - Label merging/splitting is not easy
  - Layer-four switching is difficult
- Generation method of logical topology
  - Traffic demand should be known a priori for lightpath optimization
  - Entire topology should be determined as an optimization problem
  - Our incremental network dimensioning can solve the problem
- Bottleneck at the Ingress LSR
  - WDM ring at the edge node is possible to distribute the packet forwarding

MPLS based on Photonic Label

- Packet forwarding using photonic label

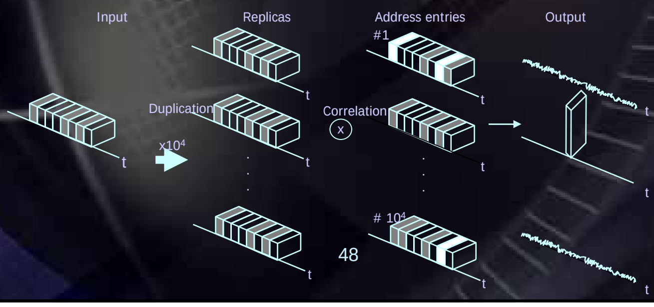
OC-based Switching Node

OC –based Switch



Packet Switching based on Photonic Label

- Optical code by BPSK
- Photonic label processing in optical domain
  - photonic label is tapped from packet header
  - optically dupilcated by optical amplification
  - power-splitted as many copies as the count of label entries in the table
  - optical correlations between the copies and the label entries is performed in parallel



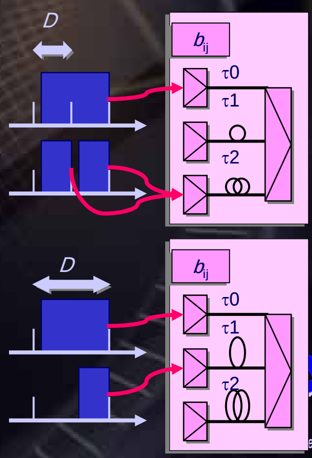
Photonic Packet Switch for Variable-Length Packet

- The length of every packet is counted by the delay line unit *D*
  - Small delay line unit *D* means small buffer capacity
  - If *D* is large, dummy byte insertion degrades the performance
- Buffer status counter b_{ij} for packets destined for wavelength *i* on output line *j*
  - For every arrival of packet with length *x*

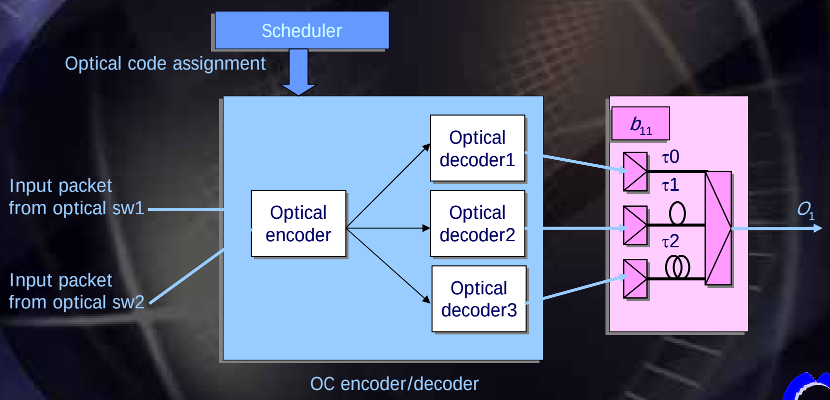
$$b_{ij} \leftarrow b_{ij} + \lceil x/D \rceil$$

For every delay unit *D*

$$b_{ij} \leftarrow b_{ij} - 1$$

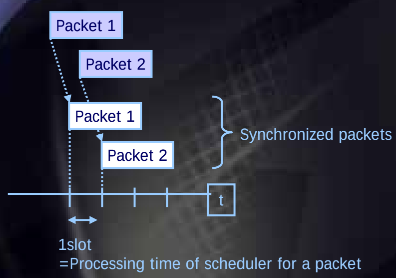


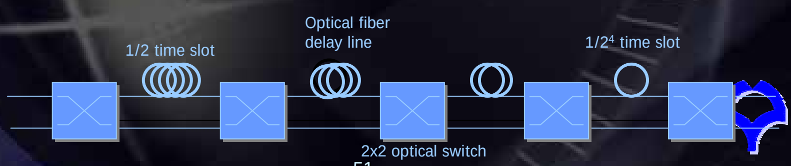
Optical Buffer using FDL



Advanced Network Architecture Research

Packet Synchronization using Packet Sequencer



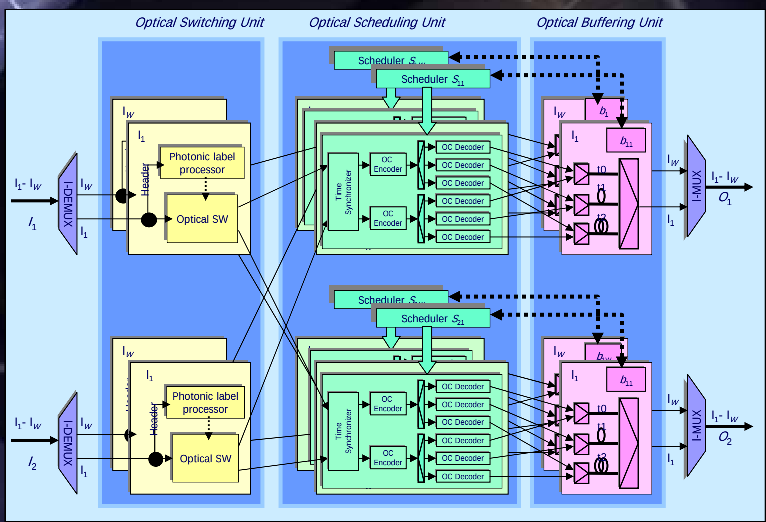


M. Murata

Osaka University

Advanced Network Architecture Research

Packet Switch Architecture without Wavelength Conversion

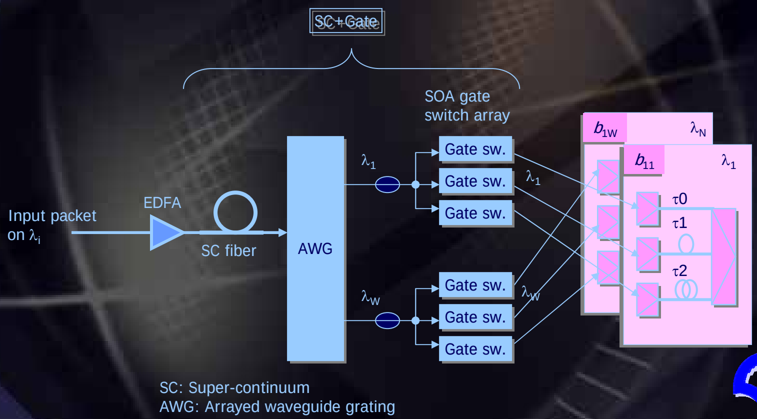


M. Murata

University

Advanced Network Architecture Research

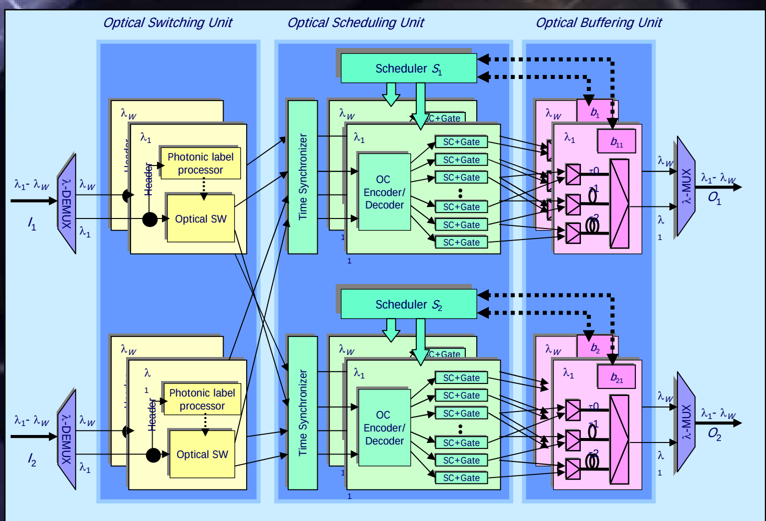
Effective Buffer Usage by Wavelength Conversion



SC: Super-continuum
AWG: Arrayed waveguide grating

Advanced Network Architecture Research

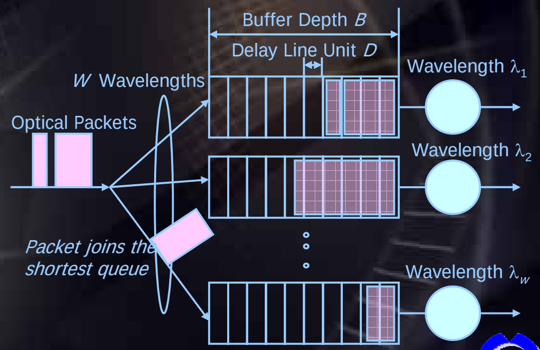
Packet Switch Architecture with Wavelength Conversion



Advanced Network Architecture Research

Packet Scheduling for WDM Buffer

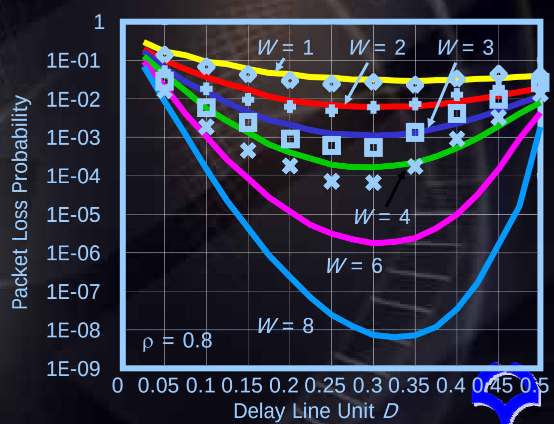
- Multi-server and Multi-queue with “Join-the-Shortest-Queue” scheduling policy
- Approximate analysis by exponentially-distributed packets and Poisson arrivals

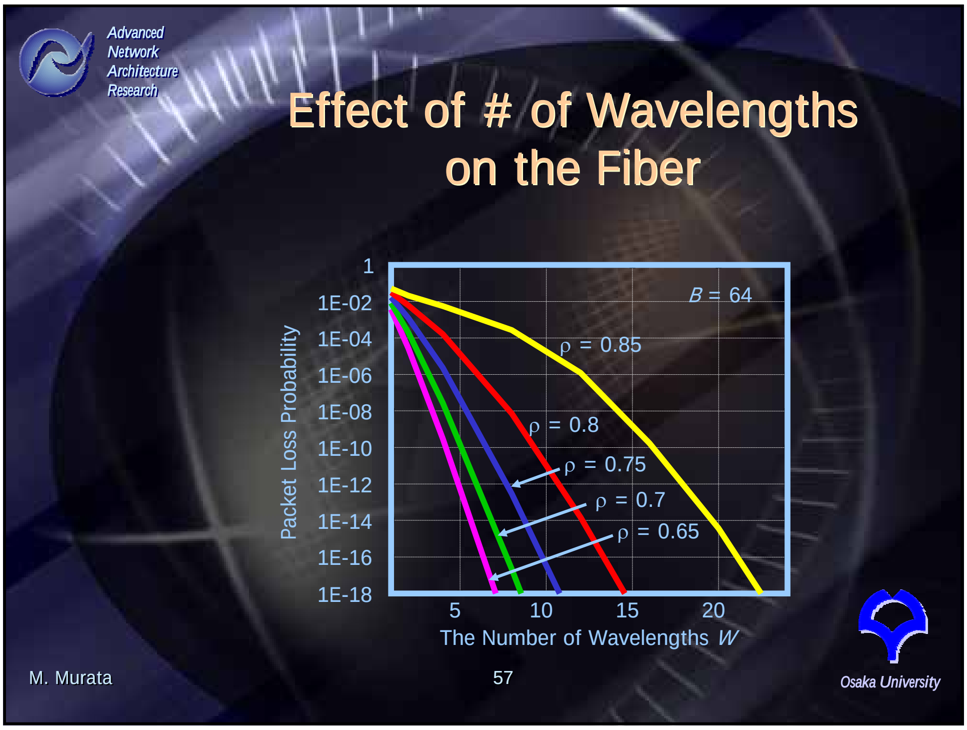


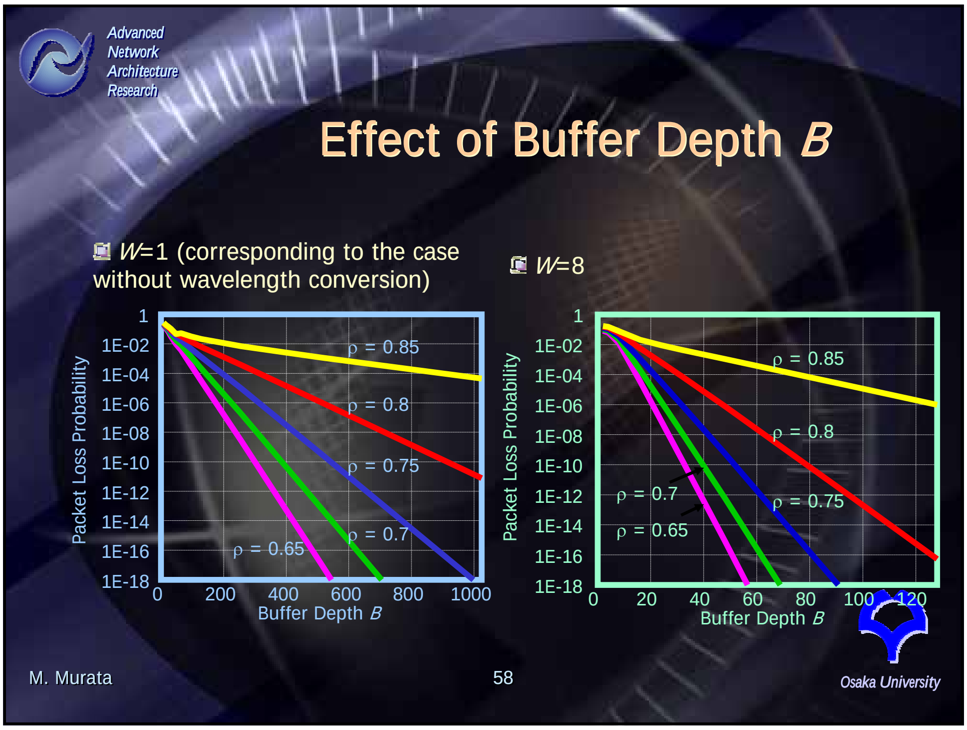
Advanced Network Architecture Research

Effect of Delay Line Unit *D*

- Buffer depth is fixed at *B* = 64
- Small delay line unit *D* means small buffer capacity
- If *D* is large, dummy byte insertion degrades the performance

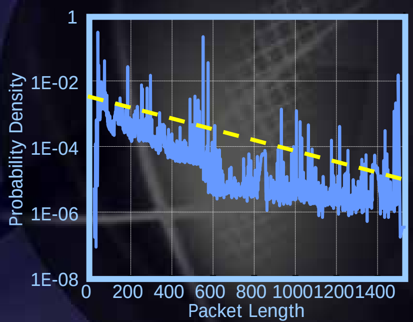



Advanced
Network
Architecture
Research
Effect of # of Wavelengths on the Fiber
Packet Loss Probability
1
1E-02
1E-04
1E-06
1E-08
1E-10
1E-12
1E-14
1E-16
1E-18
B = 64
ρ = 0.85
ρ = 0.8
ρ = 0.75
ρ = 0.7
ρ = 0.65
5
10
15
20
The Number of Wavelengths W
M. Murata
57
Osaka University

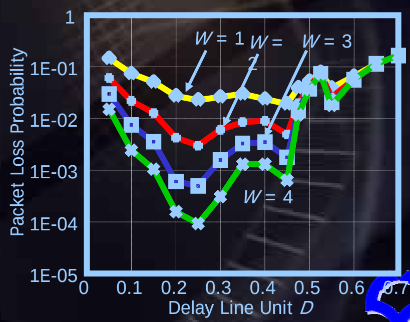

Advanced
Network
Architecture
Research
Effect of Buffer Depth B
W=1 (corresponding to the case without wavelength conversion)
W=8
Packet Loss Probability
1
1E-02
1E-04
1E-06
1E-08
1E-10
1E-12
1E-14
1E-16
1E-18
ρ = 0.85
ρ = 0.8
ρ = 0.75
ρ = 0.7
ρ = 0.65
0
200
400
600
800
1000
Buffer Depth B
Packet Loss Probability
1
1E-02
1E-04
1E-06
1E-08
1E-10
1E-12
1E-14
1E-16
1E-18
ρ = 0.85
ρ = 0.8
ρ = 0.7
ρ = 0.75
ρ = 0.65
0
20
40
60
80
100
120
Buffer Depth B
M. Murata
58
Osaka University

Effect of Distribution of IP Packet Length

- IP Packet Length Distribution



- Effect of delay line unit *D*

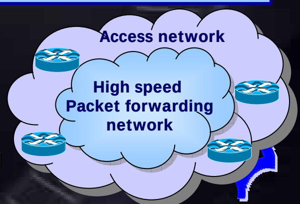


"WAN Packet Size Distribution," http://www.nlanr.net/NA/Learn/packetsizes.html

Packet Switching vs. Circuit Switching

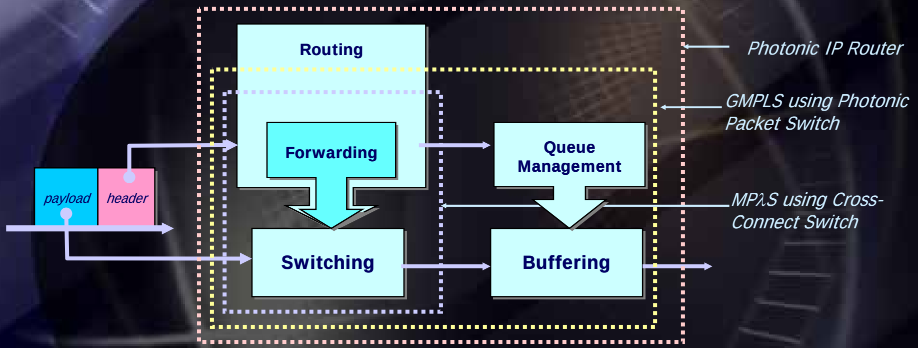
Function	Circuit Switching (Photonic XC)	Packet Switching (Electronic Router)
Utilization	Not bad	Generally believed to be good, but overprovisioning decreases utilization to attain low delays
E-to-E Path Availability	High by much cost	Maintained by routing protocol
Node Availability	High due to poor functionality	Low
Node Cost	Low due to poor functionality (half to 1/10)	High if we provide high functionality
Diversity of Service	Low	High

- Fusion of packet and circuit switching technologies
  - Access network: packet switching
  - Backbone network: WDM based circuit switching by vorture of GMPLS
  - To increase scalability, adaptability, flexibility, the # of wavelengths per fiber should be increased instead of increasing the wavelength capacity



- In future, photonic packet switching + GMPLS ?
  - Deployment?

Roadmap of IP over Photonic: Cross-Connect, Switch or Router?



- Burst switching is just same as MPλS using Cross-Connect Switch, implying that it does not add any functionality

Open the Wavelength to Users!

- Carrying IP packets is a future role of WDM?
- TCP inherently
  - Incurs packet loss
  - Takes a role of fair-share of bandwidth
- Open the wavelength to users (not to edge nodes)
  - On-demand lightpath setup
  - The # of wavelengths should be enough large (more than 1,000)
  - Application area
    - User-oriented VPN
    - Data grid to carry Tbyte data
    - Wide area SAN
  - Proprietary protocols can be adopted