e-Photon ONe

Multiple TCP Flow Performance Study over Optical Burst Switched Networks

M. Casoni, A.M. Guidotti and C.Raffaelli

(casoni.maurizio@unimore.it)

Department of Information Engineering
University of Modena and Reggio Emilia
Italy

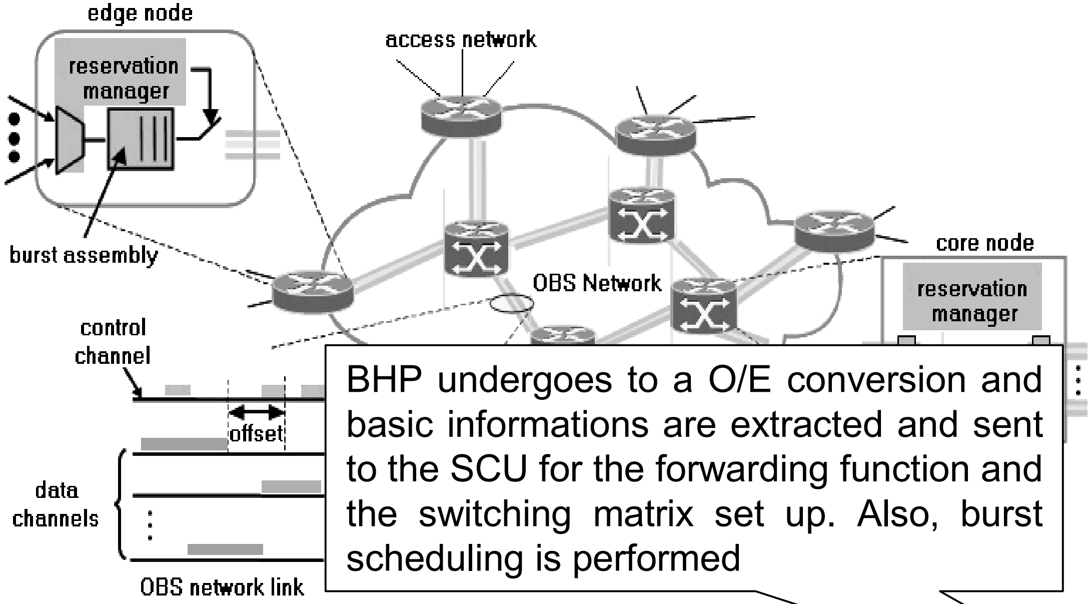
- Introduction: Optical Burst Switching scenario
- Assembly algorithms
  - Per-flow, mixed flow
- The TCP scenario
  - Single flow, multiple flows
- Numerical results
  - Analysis and Simulation (ns)
- Conclusions

- DWDM technique
  - **Transmissionrate in the range of Tbit/s**
- Architectural semplification
  - **From IP over ATM over SONET over WDM to IP over WDM**
- Need to exploit in an effective way the huge transmission bandwidth with IP traffic
  - **Wavelength Routing**
    - ✓ all-optical data network
    - ✓ Low flexibility for IP traffic
  - **Optical Packet Switching**
    - ✓ Ideal transfer mode for IP traffic
    - ✓ Severe technological constraints → not feasible in the short/middle term
      - Optical components immature

Goal: better sinergy between the mature electronic technologies and the new optical tecnologies (mid-term solutions)

- **Switching granularity between WR and OPS**
  - *Burst concept:* aggregation of IP packets with common features (e.g. destination and QoS), considered as the basic optical unit
- **Time and <u>space separation</u> of data and control (header) fields**
  - Control packet employs dedicated channel and precedes the relative data burst
    - ✓ All-optical network, buffer-less and data trasparent
    - ✓ Hybrid opto-electronic network for control signals (*out-of-band signaling*)
  - Simplification of the electronic processing of the control packets at intermediate nodes
  - Reduction of the opto-electronic functionalities required to router

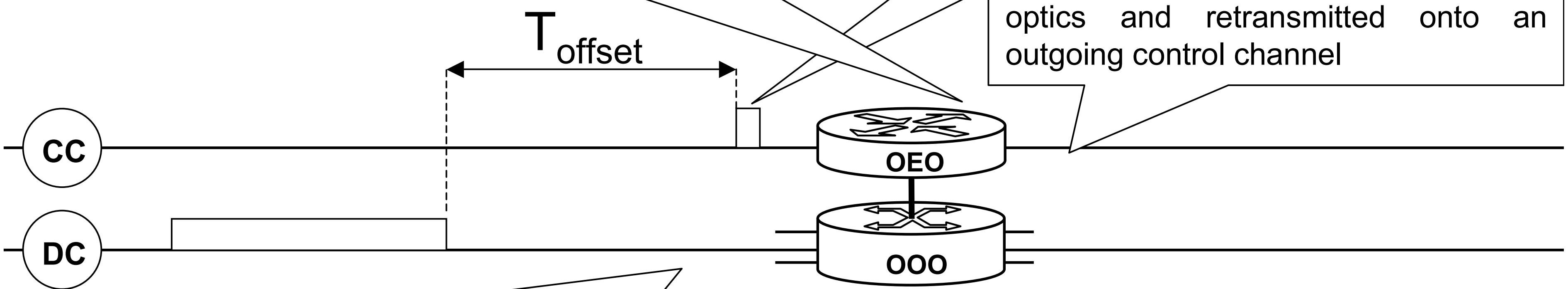
- **Two types of nodes:**
  - **Edge node**
    - **Interface between legacy network and optical domain**
    - **Assembly / Disassembly**
  - **Core node**
    - **BHP processing**



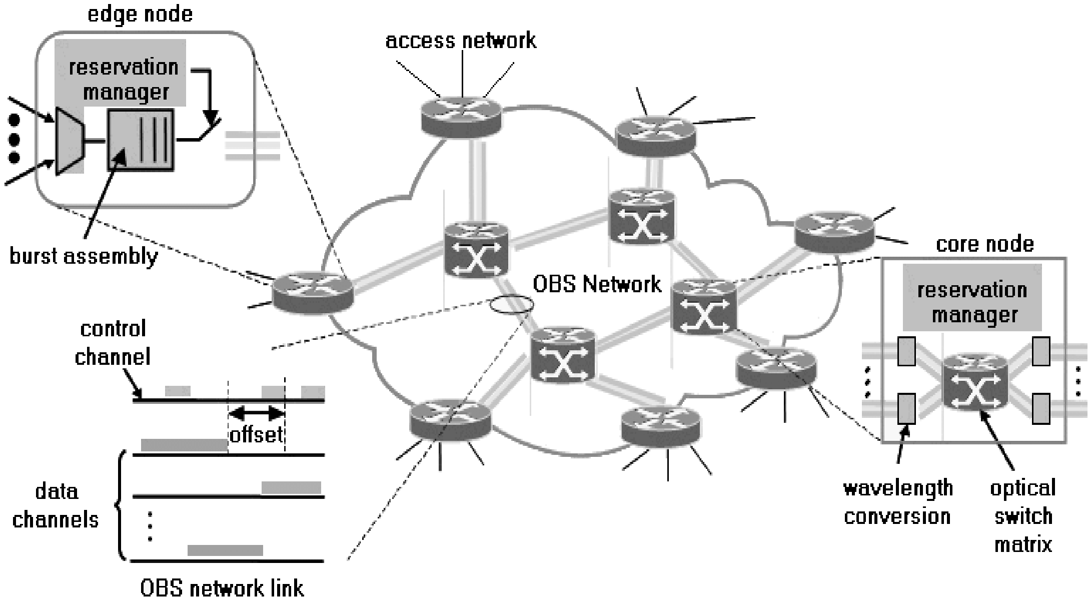
BHP undergoes to a O/E conversion and basic informations are extracted and sent to the SCU for the forwarding function and the switching matrix set up. Also, burst scheduling is performed

Control packet (BHP) arrival at the interface of a core node

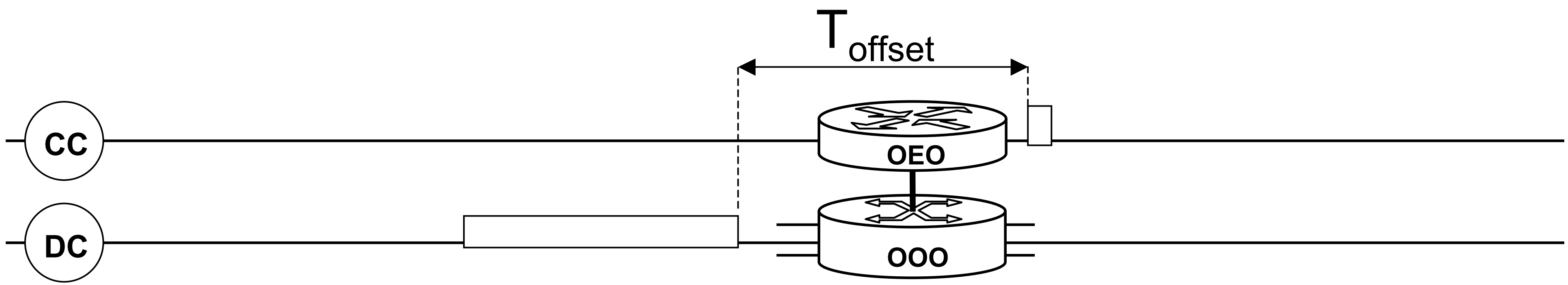
BHP is updated, reconverted into optics and retransmitted onto an outgoing control channel

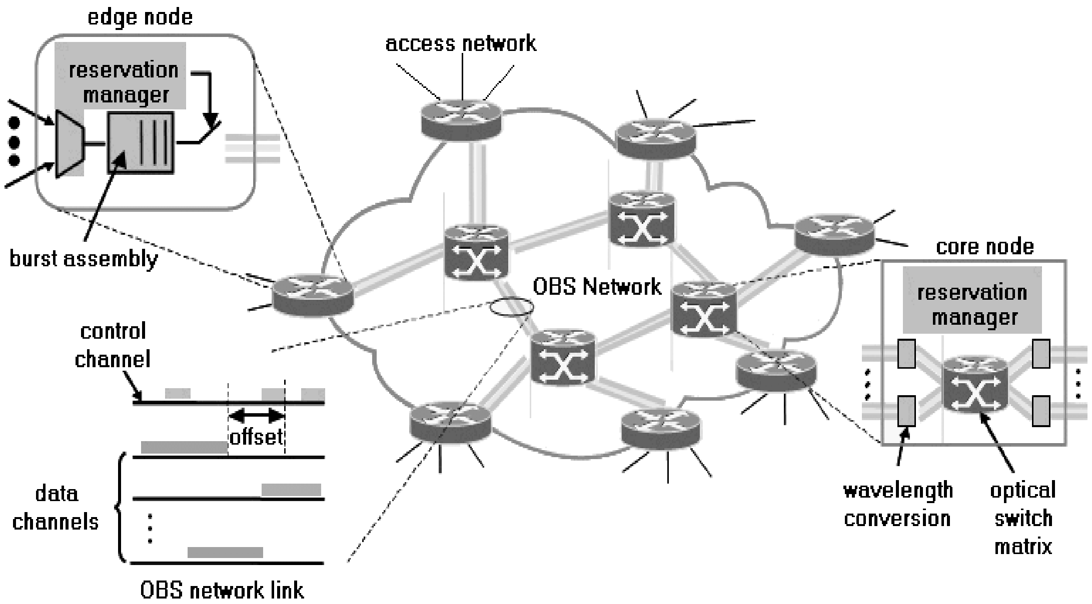


When the data burst arrives at the core node, the optical matrix is already set up and all internal devices configured. Data burst does not experience any delay and remain in the optical domain

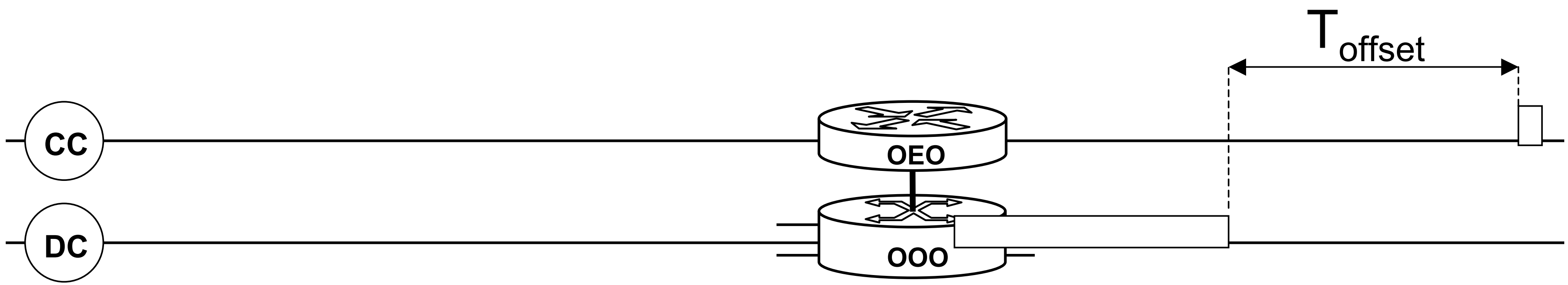


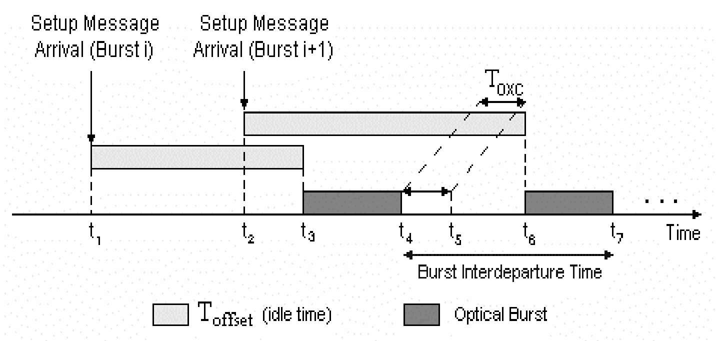
- **Two types of nodes:**
  - **Edge node**
    - **Interface between legacy network and optical domain**
    - **Assembly / Disassembly**
  - **Core node**
    - **BHP processing**
    - **Routing table look-up**
    - **Scheduling**

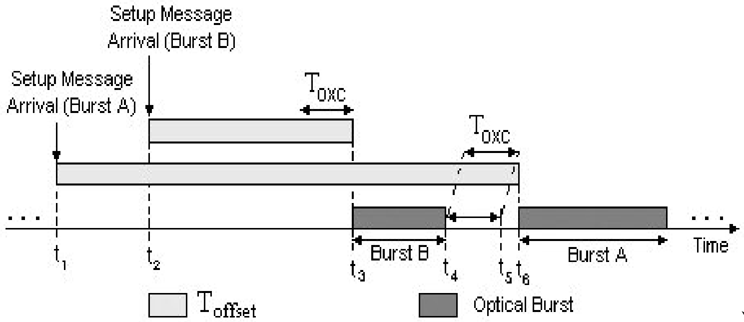


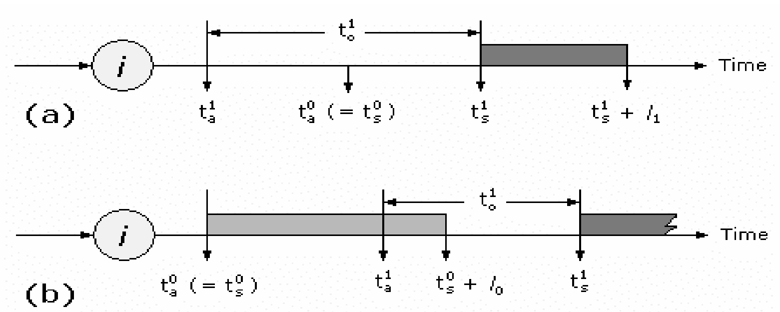


- **Two types of nodes:**
  - **Edge node**
    - **Interface between legacy network and optical domain**
    - **Assembly / Disassembly**
  - **Core node**
    - **BHP processing**
    - **Routing table look-up**
    - **Scheduling**





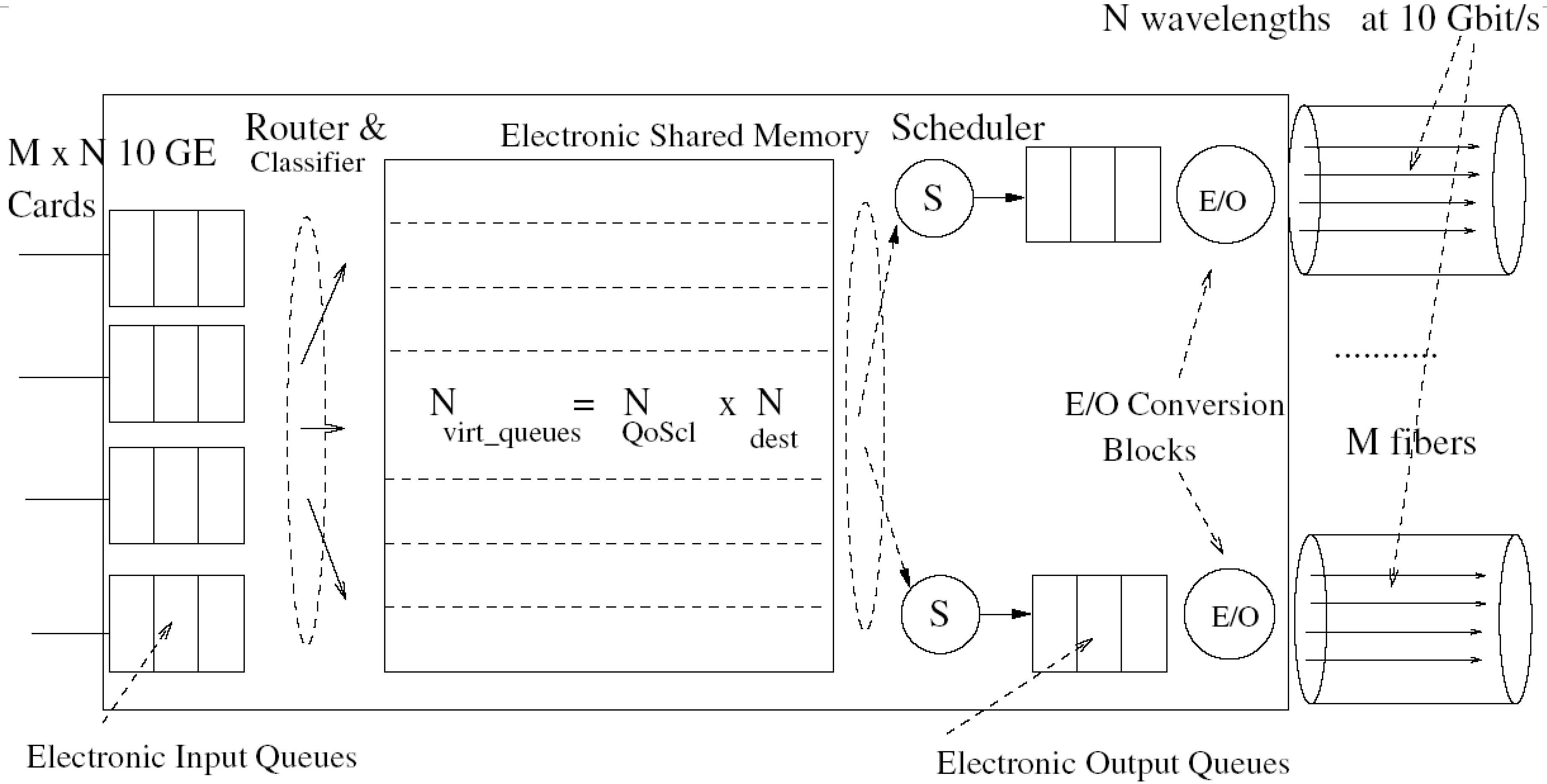




- Delayed reservation (Just Enough Time)
  - **BHP contains informations about the start and the duration of the correspondent data burst**
  - **Output wavelength is reserved only for the data burst length**
    - Efficient resource utilization
  - **Possibility to exploit empty spaces between 2 previously scheduled bursts (Void Filling)**

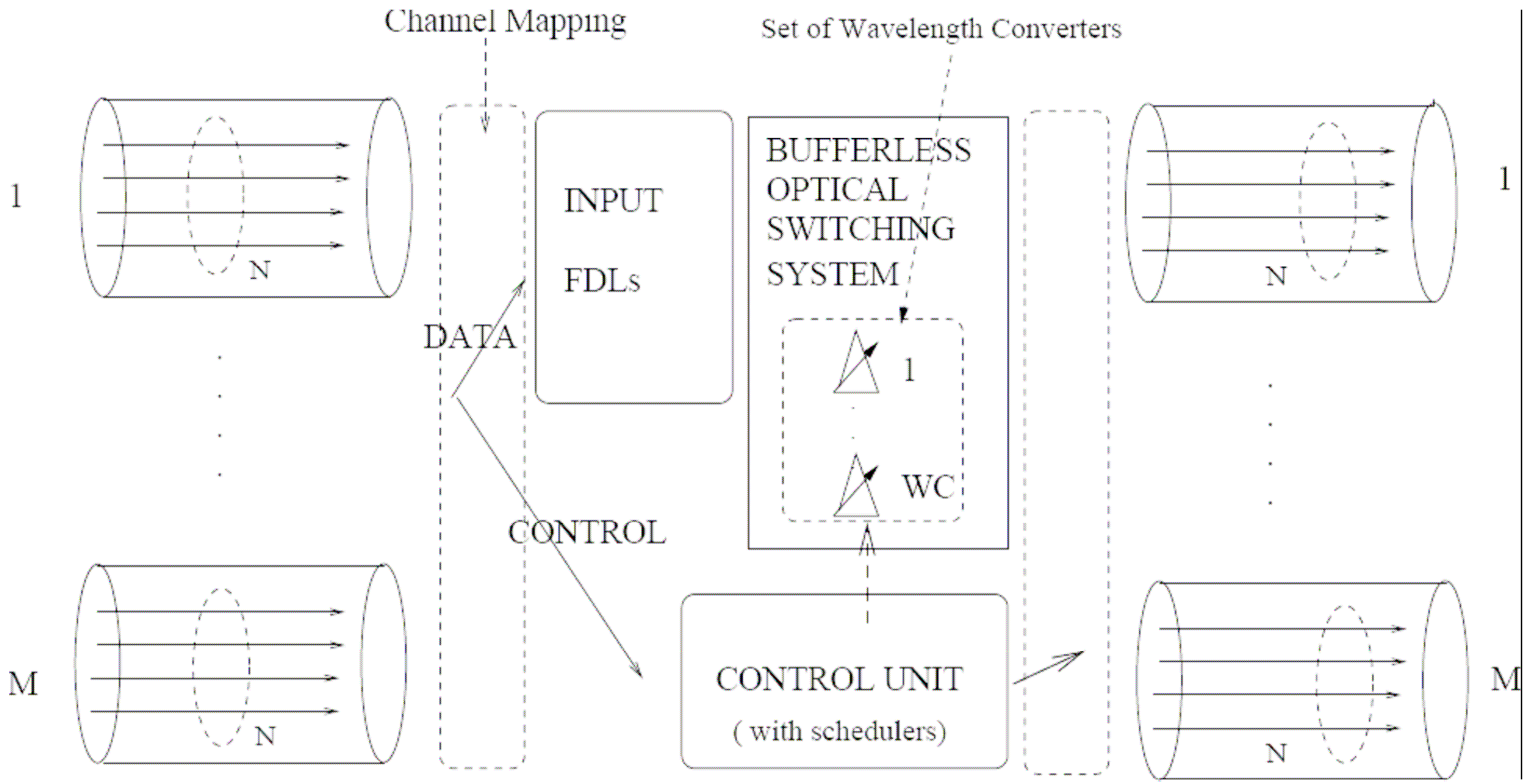
- QoS
  - **Additional extra offset allows to BHP to reserve the output channel in advance**
    - Lower blocking probability
    - Implementation of different QoS classes

Edge node



N wavelengths at 10 Gbit/s
M x N 10 GE
Cards
Router &
Classifier
Electronic Shared Memory
Scheduler
S
S
E/O
E/O
N virt_queues = N QoScl x N dest
E/O Conversion
Blocks
....
M fibers
Electronic Input Queues
Electronic Output Queues

Core Router Architecture

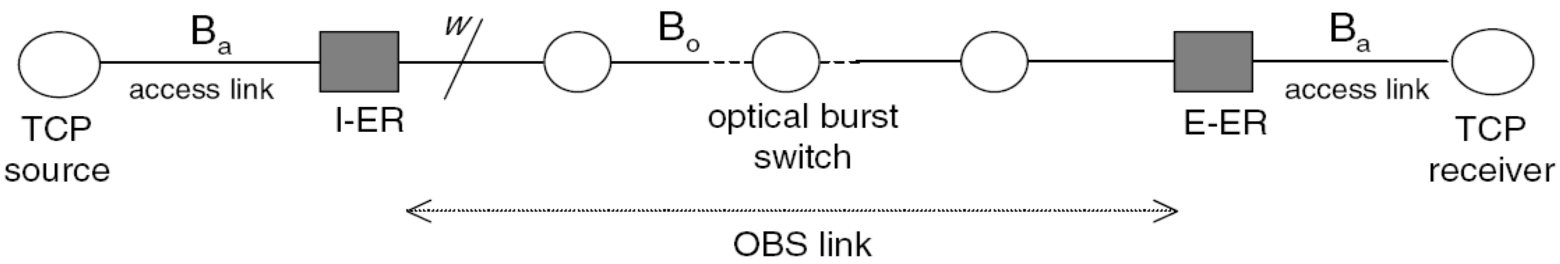


- Time based
- Threshold based
- Adaptive

- Per TCP flow
  - Optical burst contains data of the same class and same flow
  - Per flow queuing is needed at the edge node
- Mixed TCP flows
  - Optical burst may contain data from different flows of the same class
  - Per class queuing is required at the edge node

- **Time based assembly: T_b ; L_s: average TCP segment size**
- **W_m: advertised window; Number of collected segments:**

$$N_{\max} = \min\left(\left\lfloor \frac{T_b B_a}{L_s} \right\rfloor, W_m \right),$$

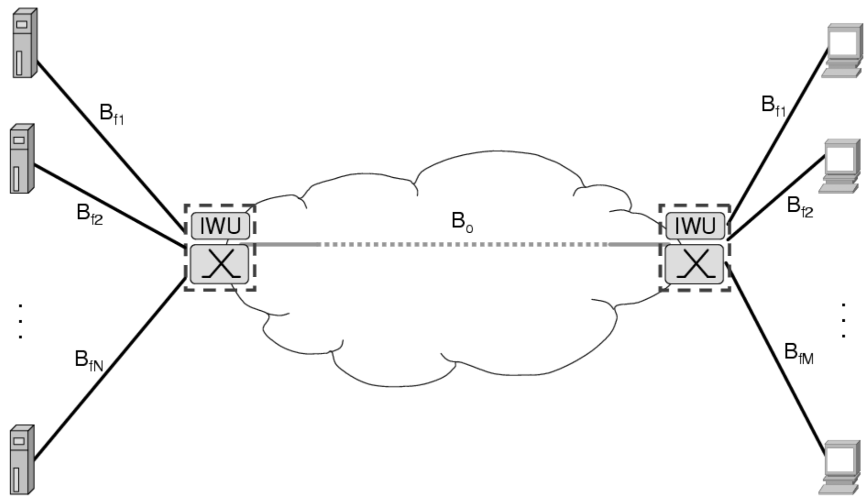


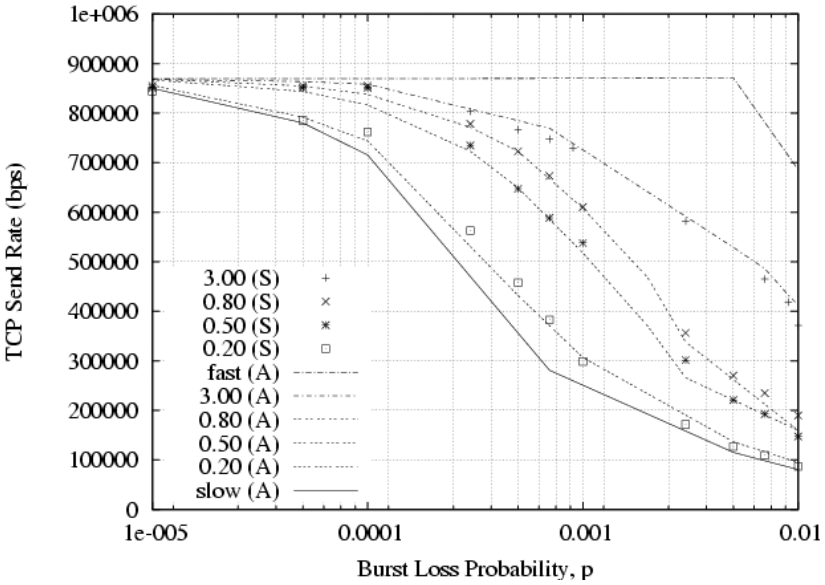
M.Casoni, C.Raffaelli, "Analytical Framework for End-to-End Design of Optical Burst-Switched Networks", Optical Switching and Networking, February 2007, pp.33-43, Elsevier Journal.

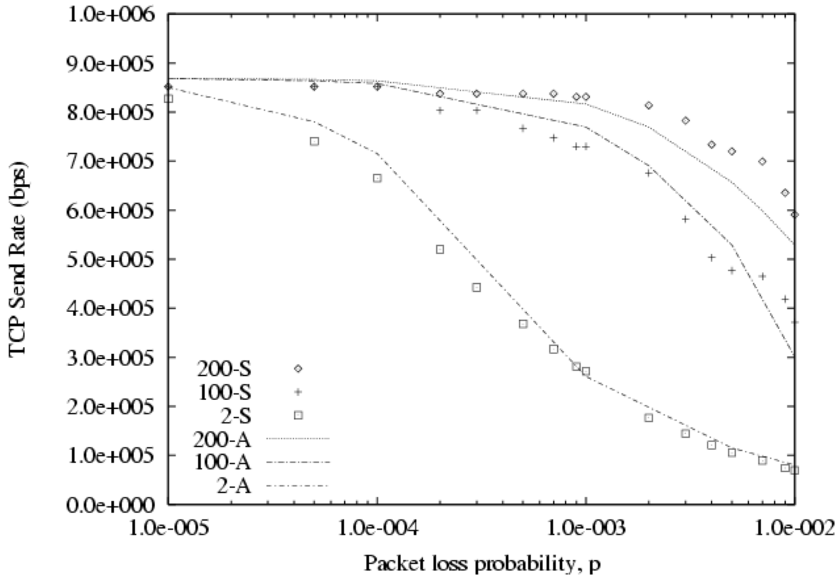
- Performance comparison between per-flow and mixed-flow;
- Two metrics evaluated: send rate and fairness
- Intra fairness ratio:

$$F_{intra} = \frac{Bw_{P\min}}{Bw_{P\max}}$$

$$\sum_{i=1}^{N} B_{fi} = B_a$$





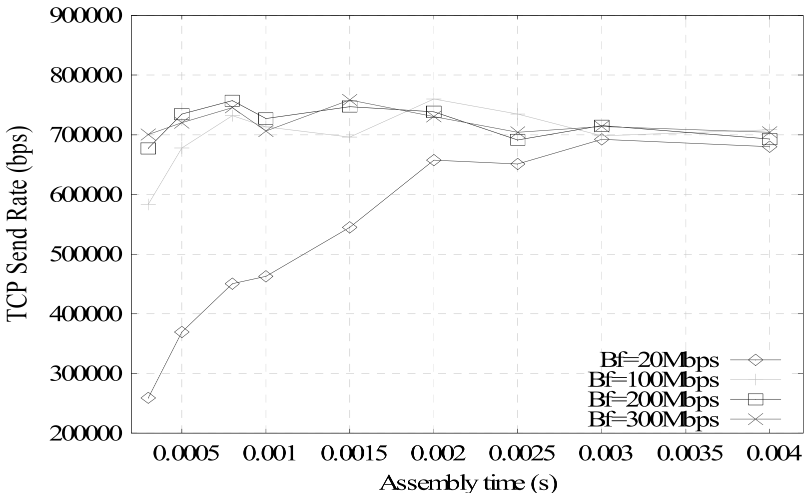


RTT_0= 600 ms; W_m= 128 segments; L_s = 512 bytes; B_o = 2.5 Gbit/s

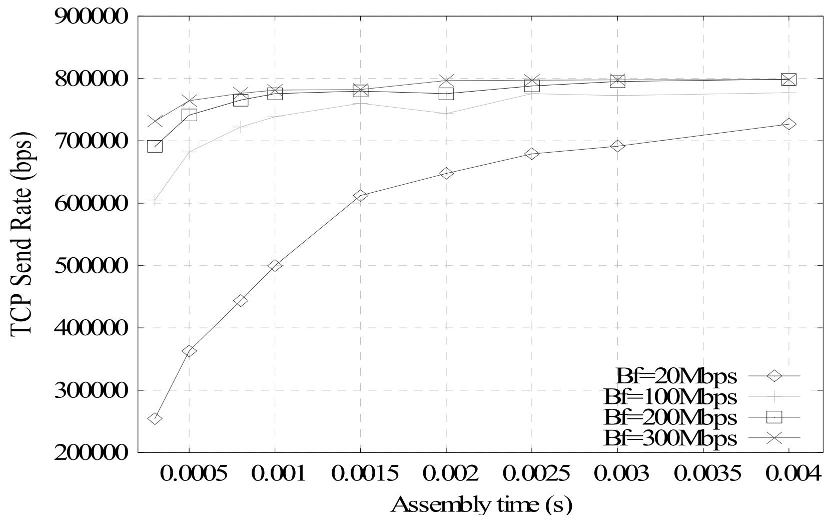
- Assembly time T_b from 0.2 to 3 ms
- B_a = 100 Mbit/s
- Send rate improves by increasing T_b due to the corrrelation benefit

- Access bandwidth B_a from 2 to 200 Mbit/s
- T_b = 3 ms
- Send rate vs. loss probability: design & evaluation
- Correlation benefit

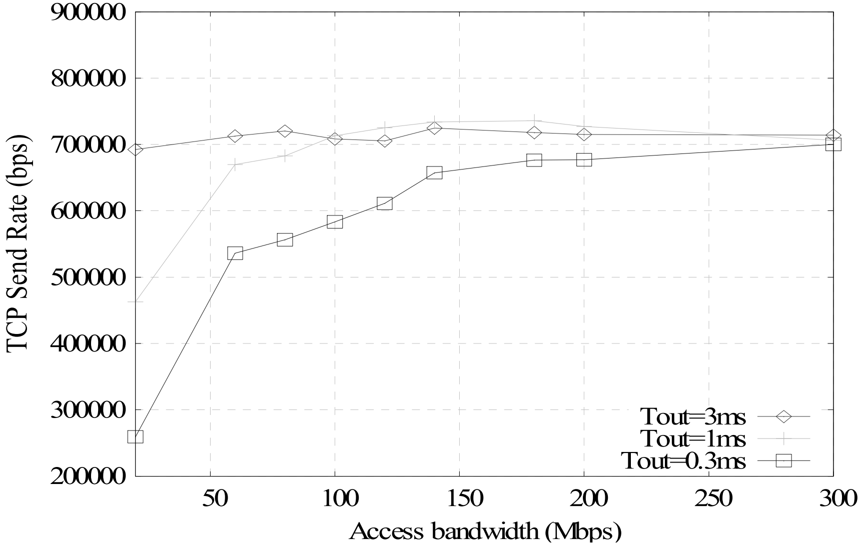
Mixed flow



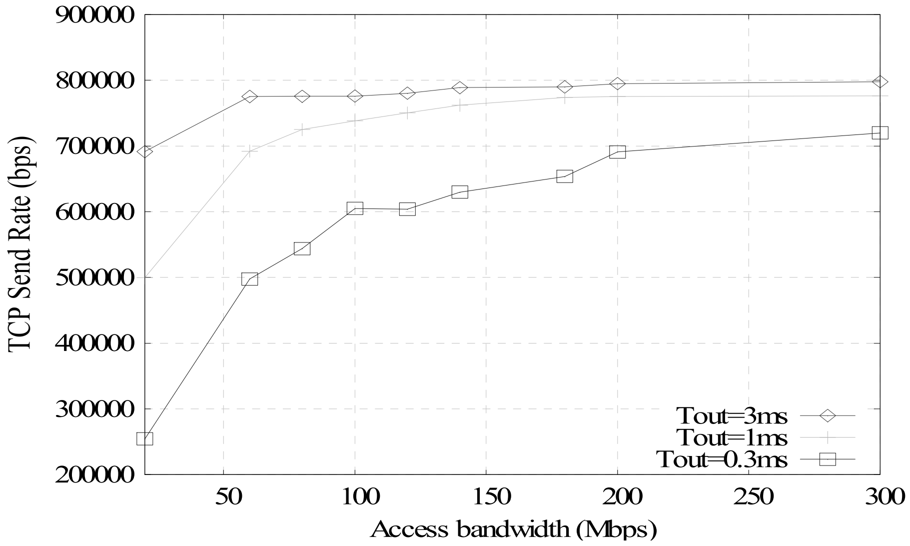
Per flow



- RTT_0= 600 ms; W_m= 128 segments; L_s = 512 bytes; B_o = 2.5 Gbit/s; N = 10; burst loss = 10^{-3}
- Access bandwidth B_f from 20 to 300 Mbit/s
- Asymptotic behaviour for increasing values of assembly time and access bandwidth
- Once all segments within the congestion window are put in the same burst, send rate gets to constant value
- Mixed flow: floor at roughly 700 kbit/s reached for T_b = 3 ms and no more gains for B_f > 100 Mbit/s
- Per flow: floor at roughly 800 kbit/s reached for longer T_b (5 ms) and no more gains for B_f > 200 Mbit/s
- Per flow performs slightly better because with mixed flow many flows get penalised by a loss, leading to an average reduction of the send rate



TCP Send Rate (bps)
Access bandwidth (Mbps)
Tout=3ms
Tout=1ms
Tout=0.3ms



TCP Send Rate (bps)
Access bandwidth (Mbps)
Tout=3ms
Tout=1ms
Tout=0.3ms

	mixed flow	per flow
Bf=20 Mbps	0.87	0.74332
Bf=100 Mbps	0.89944	0.85714
Bf=200 Mbps	0.93333	0.95455
Bf=300 Mbps	0.92697	0.93050

- Assembly time out T_b = 3 ms
- Fairness improves for both strategies when Bf increases
- The higher Bf the larger the burst size is and more segments are included in it
- Effects of a burst loss on an individual flow are lower and counterbalanced by the many more successfully transmitted segments

- OBS network has been investigated focusing on TCP performance
- Effects of burst assembly at edge nodes
  - Per flow and mixed flow with time based assembly
- Two main scenarios investigated
  - One TCP flow
  - Multiple TCP flows
- TCP send rate studied as a function of assembly time, loss, access bw

- Results:

1. Send rate not only decreases when burst loss increases (trivial) but with diverse slopes in relation to assembly time and access bandwidth
2. Both for per flow and for mixed flow send rate shows asymptotic behavior due to correlation benefit
3. Possible to determine the "optimal" assembly time and access bandwidth

THANK YOU FOR YOUR ATTENTION

casoni.maurizio@unimore.it
casoni@ieee.org

http://www.dii.unimo.it/casoni

… suggestions are very very welcome

Acknowledgements
This work has been partially funded by the Commission of EU through the *IST-FP6 Network of Excellence* <u>*e-Photon/ONe+*</u>