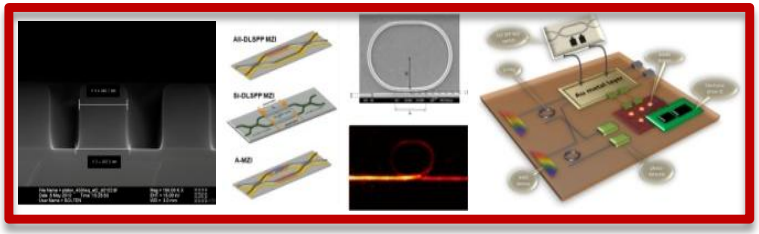
High-Speed Routing and Switching in Optical Communications & High-Performance Computing Systems



Dr. Nikos Pleros

Photonics Systems & Networks (PhosNET) Research Group
Dpt. of Informatics, Aristotle University of Thessaloniki

Outline

- ☑ *HPCS systems today: status and challenges*
- ☑ *Routing in HPC systems*
- ☑ *Optics for Routing in HPC*
- ☑ *Tb/s Si-Plasmonic Routers*
- ☑ *Optical RAM*

HPC examples..and metrics

No.1: Jaguar (USA)

No.2: Nebulae (China)

Has light anything to do with computing ?

- *1.75 Petaflop/s*

  (1 PF = 10^{15} calculations per sec)

- *410 m² floor space*
- *7 MW power consumption !!*

- *1.271 Petaflop/s*
- *Total 120640 cores*
- *2.25 MW power cons.*
- *relies on BladeSystem*

…a look inside

IBM's Roadrunner architecture

- *18x Connected Units*
- *270x Racks*

- ✓ *actually a small-range network*
- ✓ *…with 1.04 Pflop/sec and 384 Gb/s intra-CU traffic*
- ✗ *…and 2.5 MW power consumption !*

…and here comes light in

use optical fiber for the interconnection

✓ *…and enable Tb/s transmission speeds*

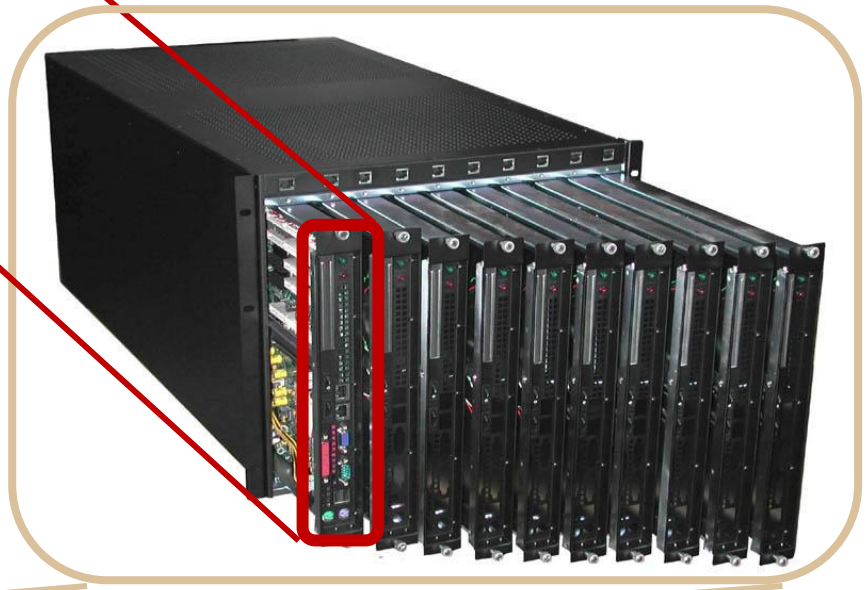
…is there any other problem ?

- *size and cable length*
- *ultra-small latency required*
  - *for fast and low-complexity parallelization*
- *power consumption…in MWs !!*
  - *consumes what a small plant can produce !!*

BladeCenters: a solution ?

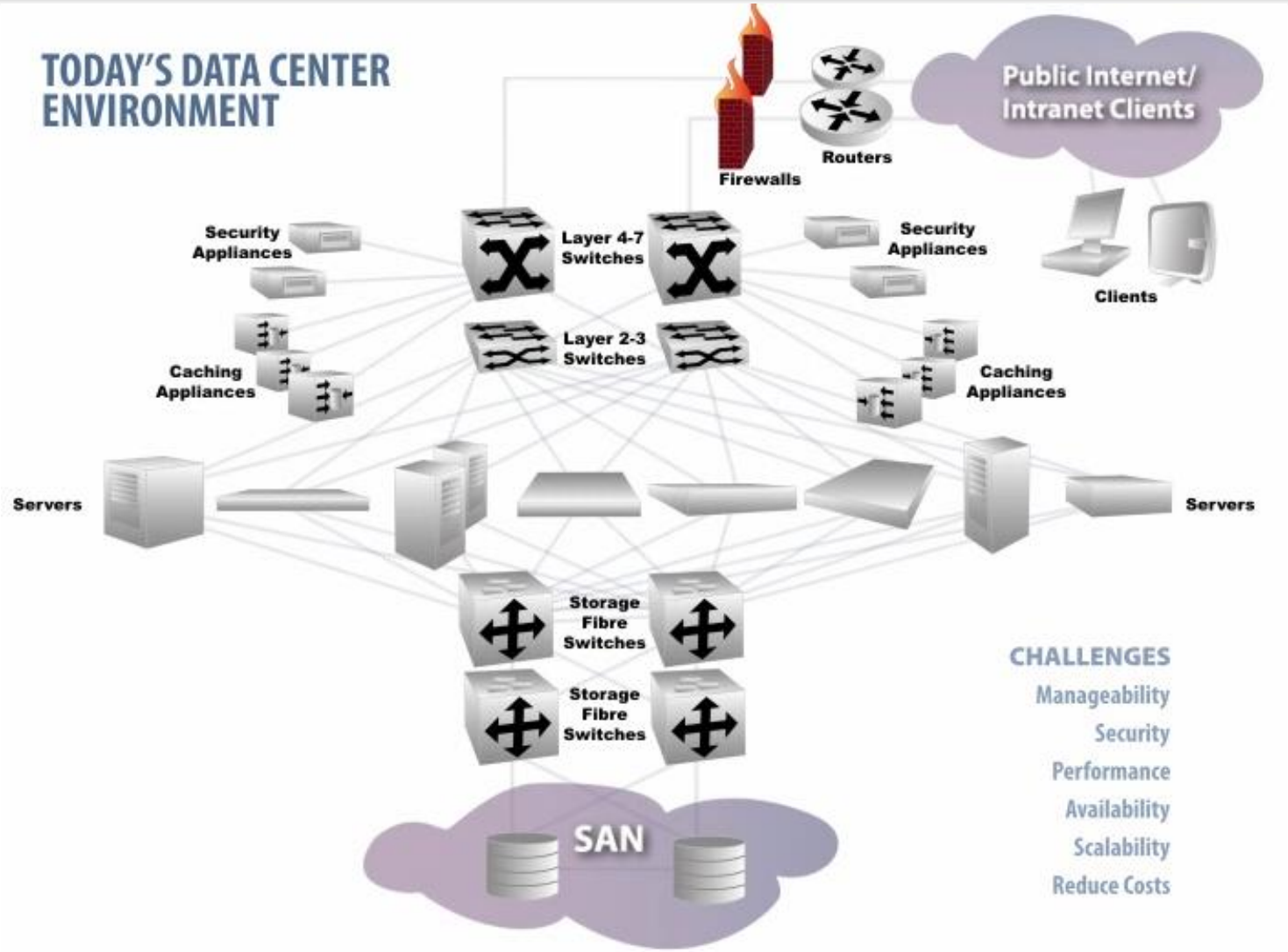
HPC architecture supported by IBM

Blade server *is a stripped down server computer, for minimizing physical space and energy requirements*

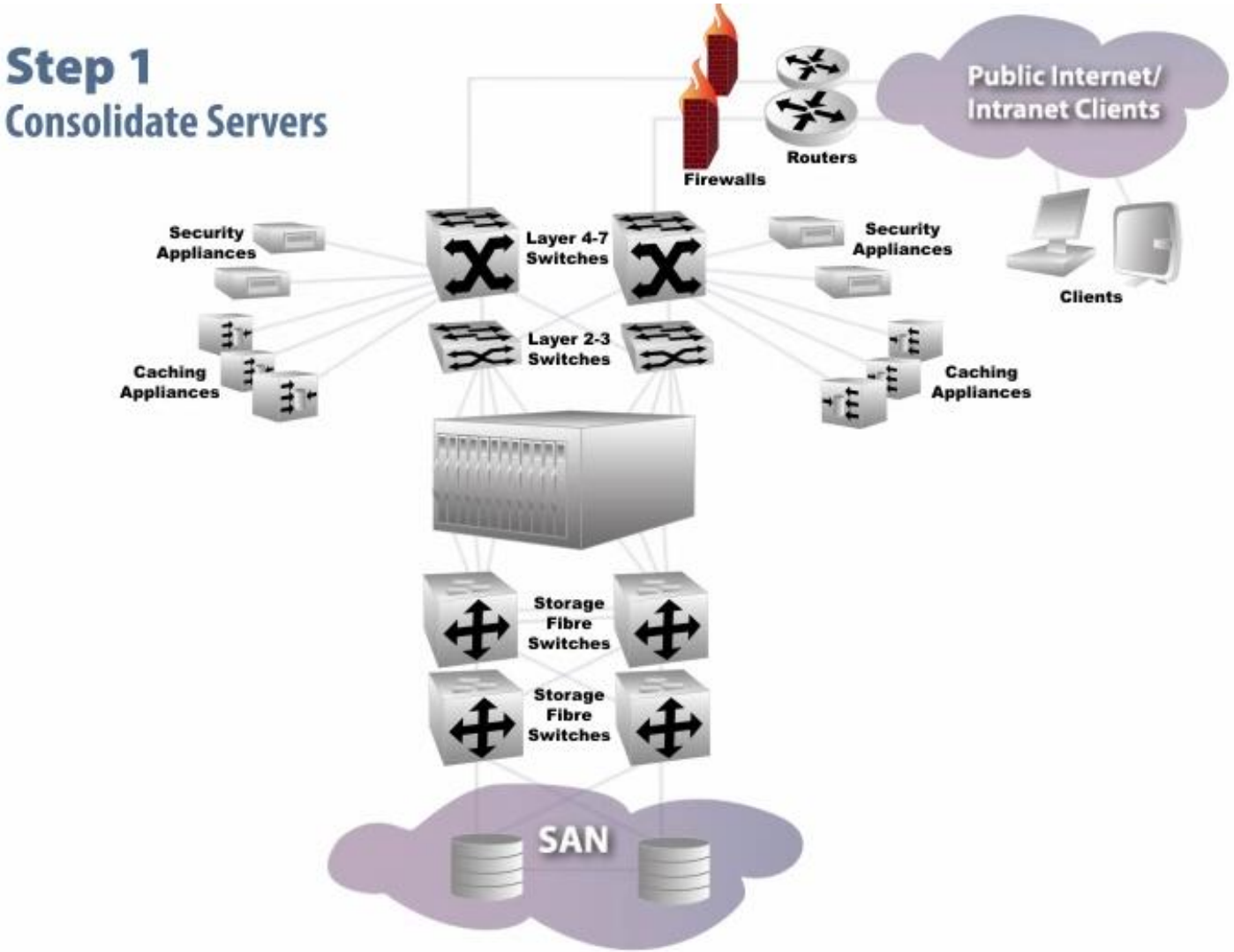


Blade enclosure*, hosts multiple blade servers, provides power, cooling, networking, interconnects & management*

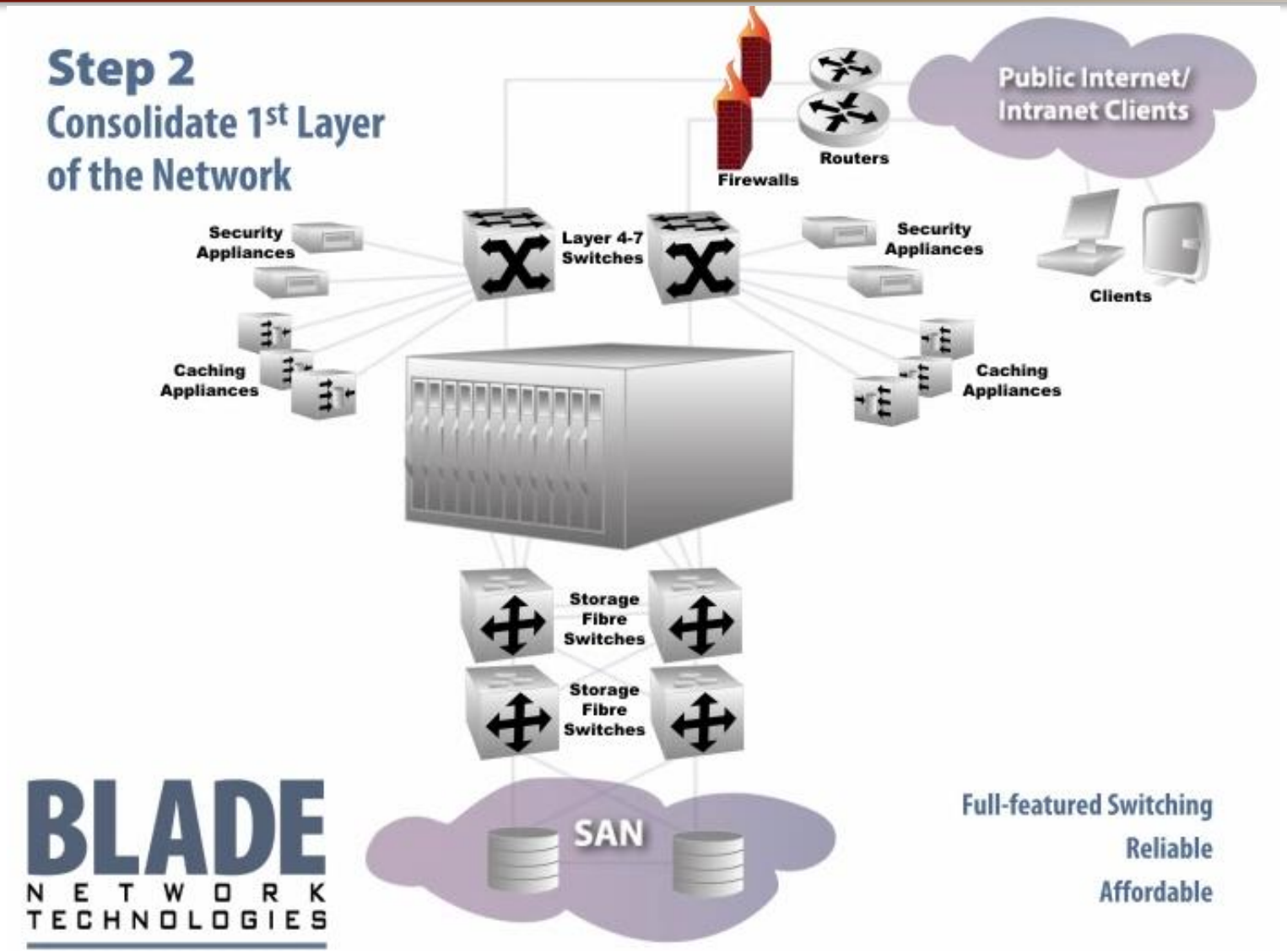
BladeCenters: The vision



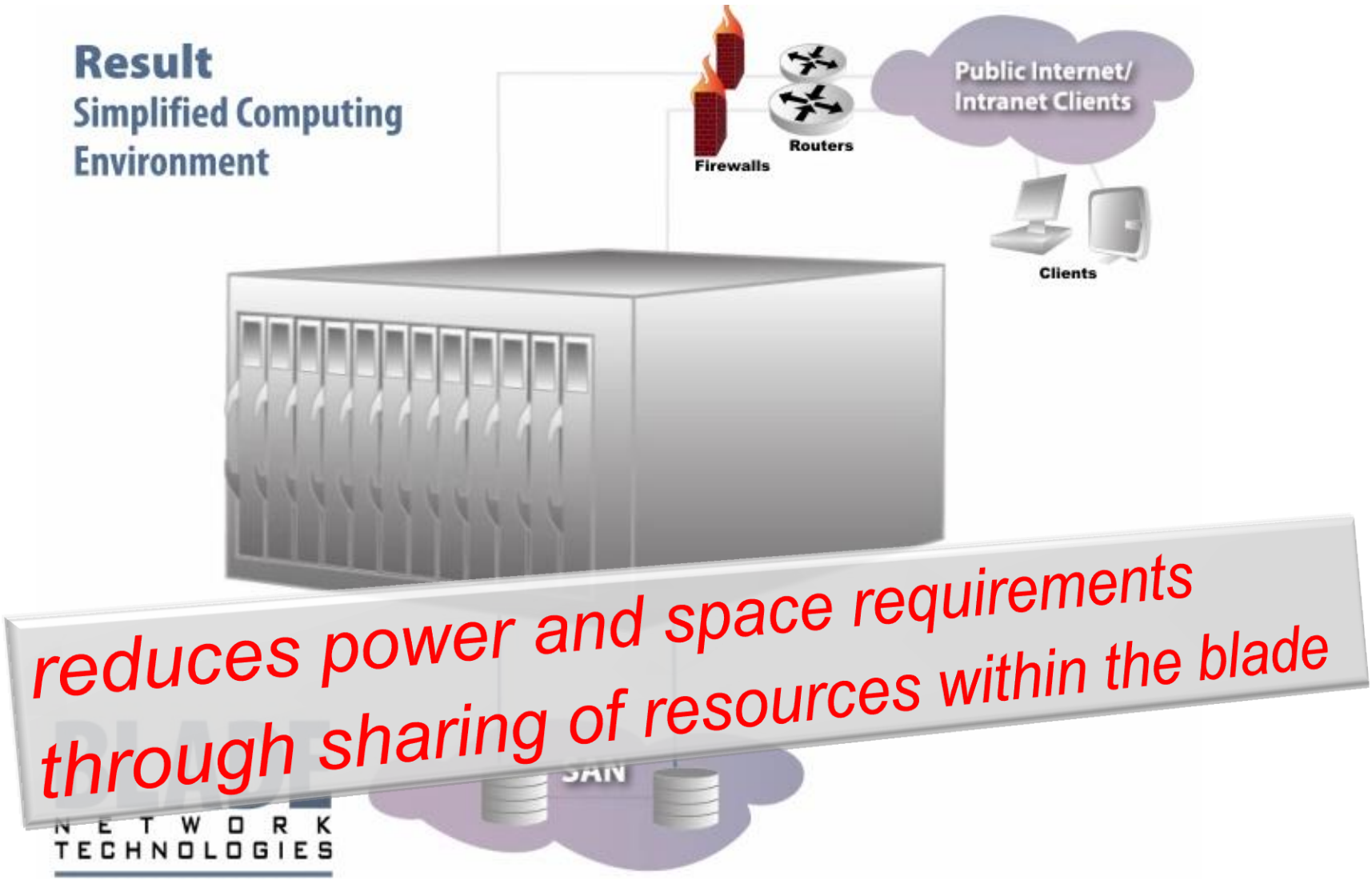
BladeCenters: The vision



BladeCenters: The vision



BladeCenters: The vision



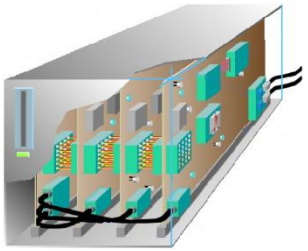
reduces power and space requirements through sharing of resources within the blade

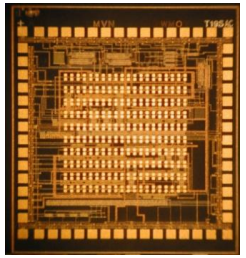
BladeCenters: a solution ?

- *creates large aggregate traffic…*
- *100's of Gb/s in miniature networks !*

- ***The Question:*** ***How to route this ?***
  - *…in consolidated network environment*
  - *…at inter-, intra-blade, backplane level*
  - *without consuming most of the blade power*

A new framework for photonics





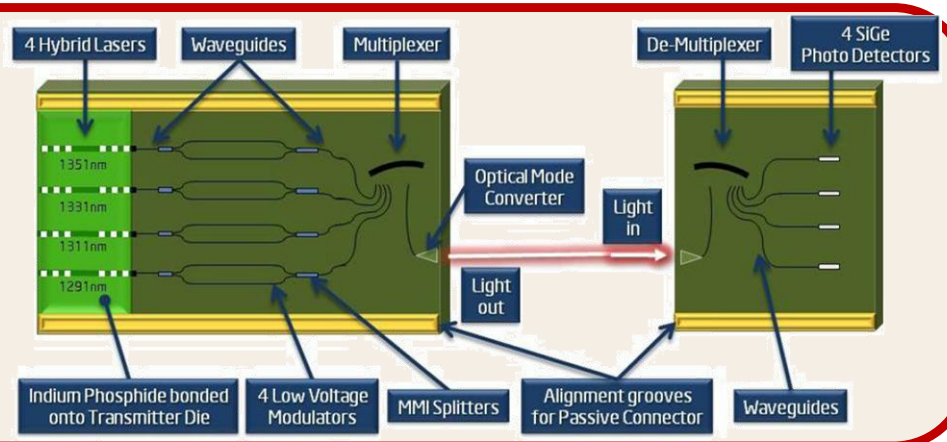
Wide Area Networks ***LANs*** ***rack-to-rack*** ***Backplane & chip-to-chip*** ***On-chip***

End 80's – early 90's *…early 2000* *…now*

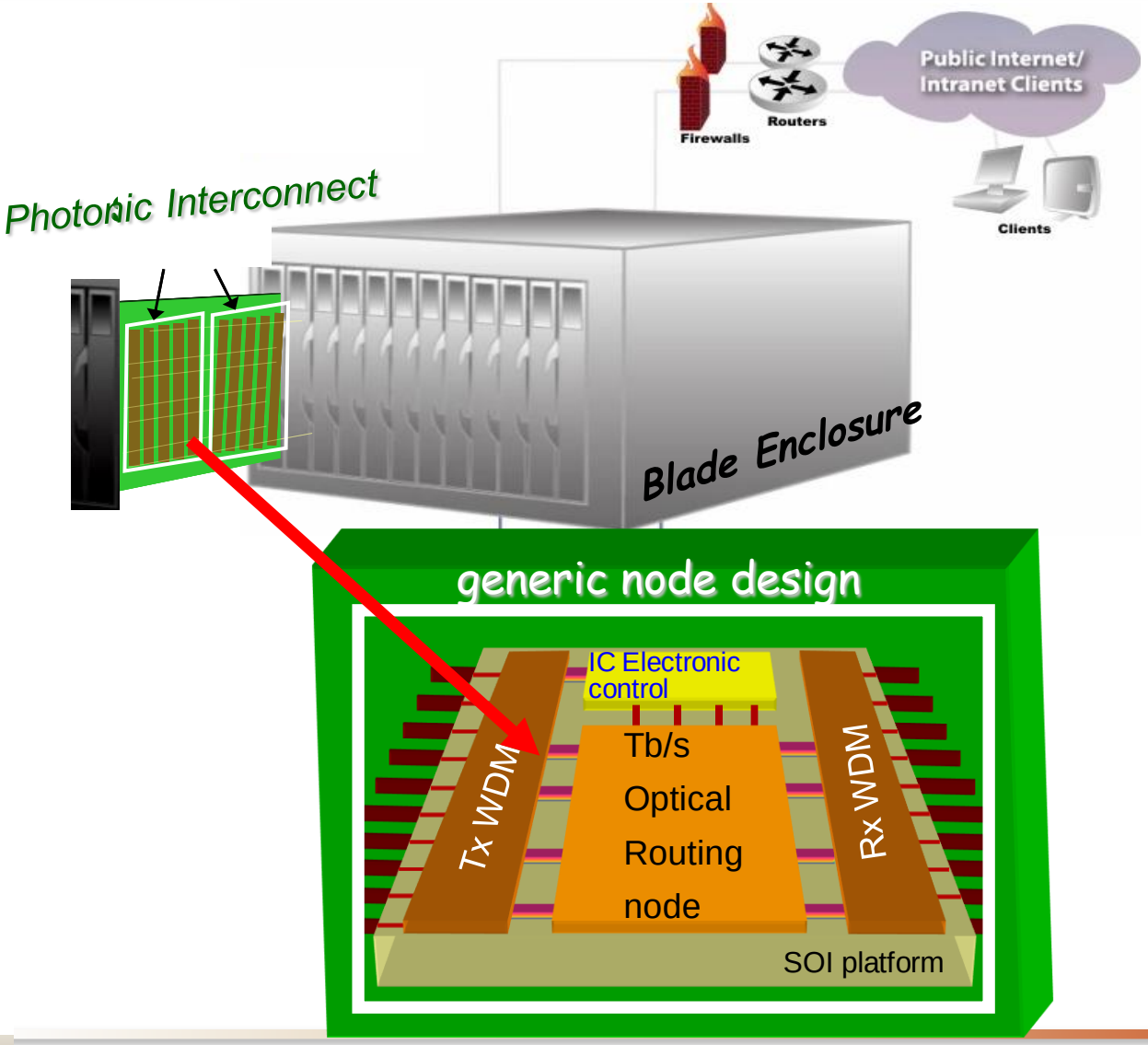
1000 km ***Network dimensions*** **1 m** **1 cm** **1 mm**

…and a new roadmap

- *Silicon Photonics integration platform*
- *Recent example: 50Gb/s optical bus (Intel USA, 2010)*



Need for chip-scale routers

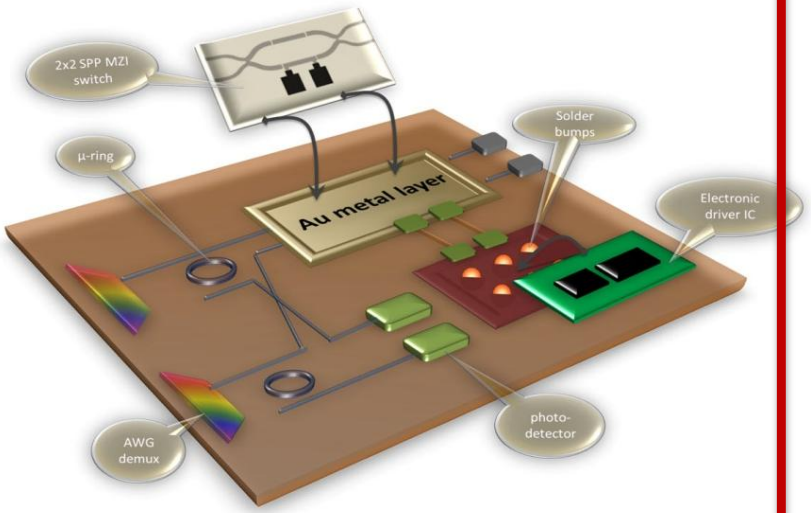


Tb/s optical routers on-chip

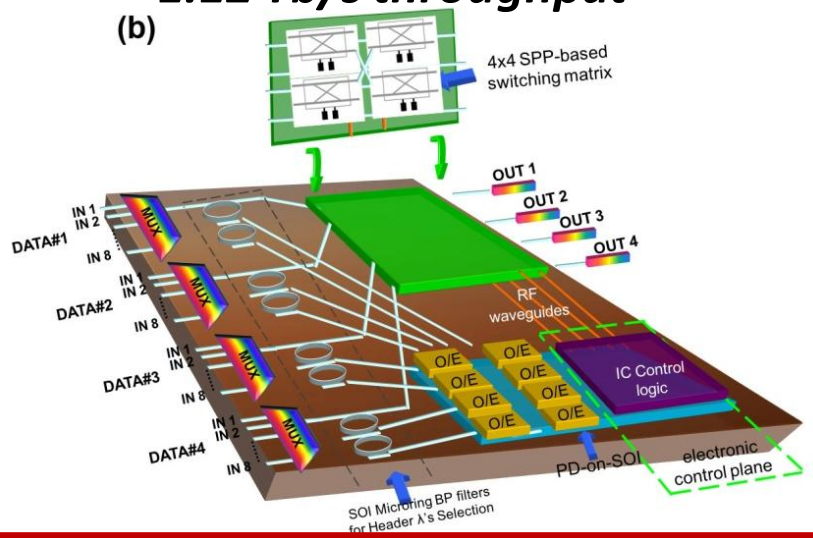
- *integrate plasmonics and silicon photonics platforms*
- *demonstrate integrated Tb/s routers:*
  - ✓ *mm² footprint*
  - ✓ *a few Watts power consumption*

Platon FP7 ...plasmonics "in-the-box"

2x2 Router
560 Gb/s throughput



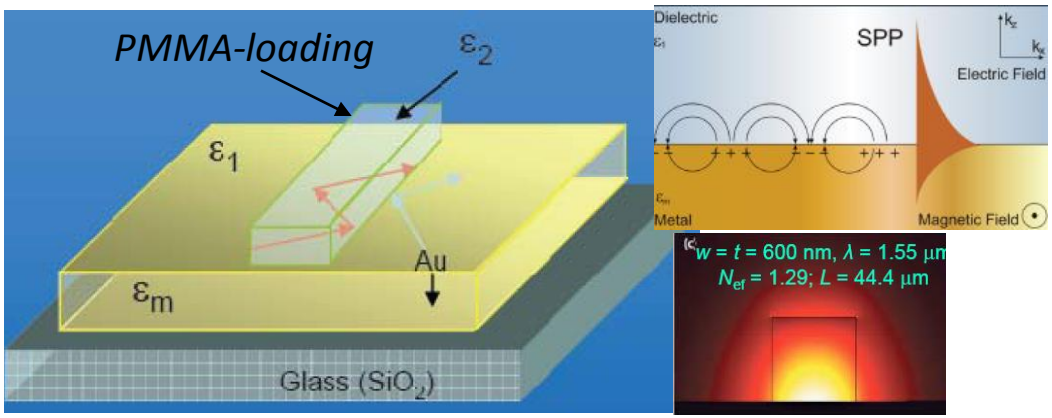
4x4 Router
1.12 Tb/s throughput



Plasmonics for switching

Dielectric-Loaded Surface Plasmon Polaritons

polymer strip (PMMA) on top of Au film



- ☑ *EM waves guided at the dielectric-gold interface*
- ☑ *small footprint (500x600nm waveguide dimensions)*

- ✓ *appropriate for interfacing photonics and electronics*
- ✓ *allows for thermooptic-induced switching phenomena*
- ✓ *low switching power consumption (few mWs)*
- ✘ *...but high propagation losses L_{prop}~45µm (while L_{π}~90µm)*

4x4 Si-Plasmonic Router

Technology & Architecture

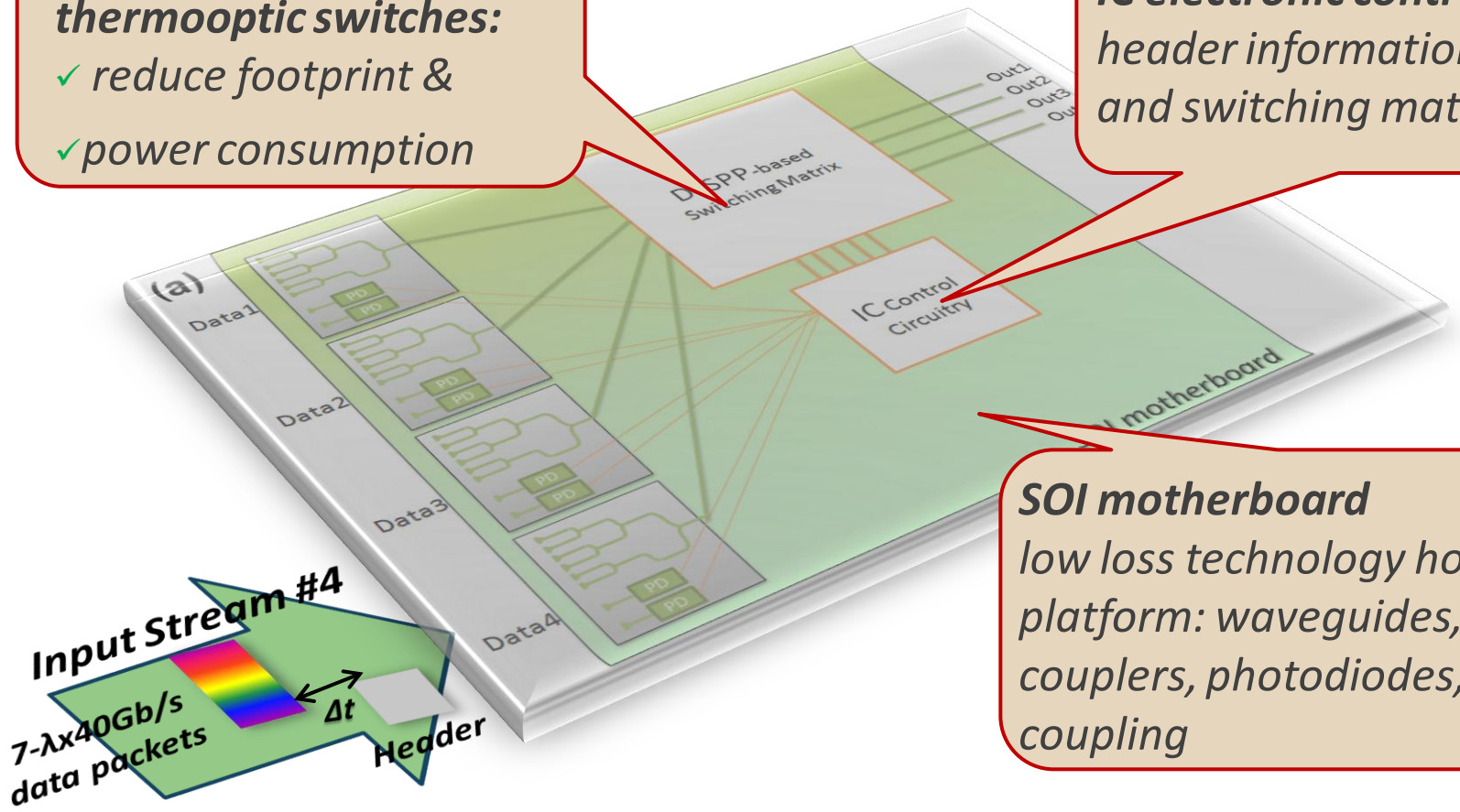
2x2 / 4x4 plasmonic thermooptic switches:

✓ *reduce footprint &*

✓ *power consumption*

IC electronic control circuit

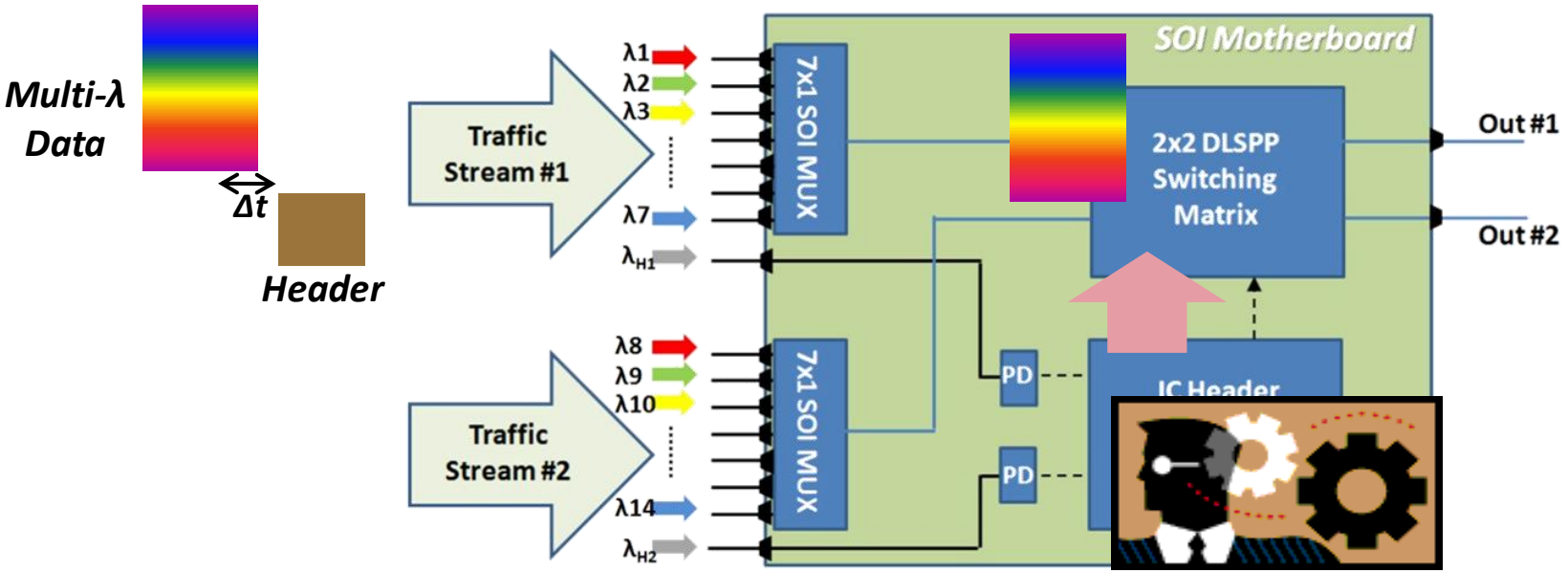
header information processing and switching matrix control



SOI motherboard

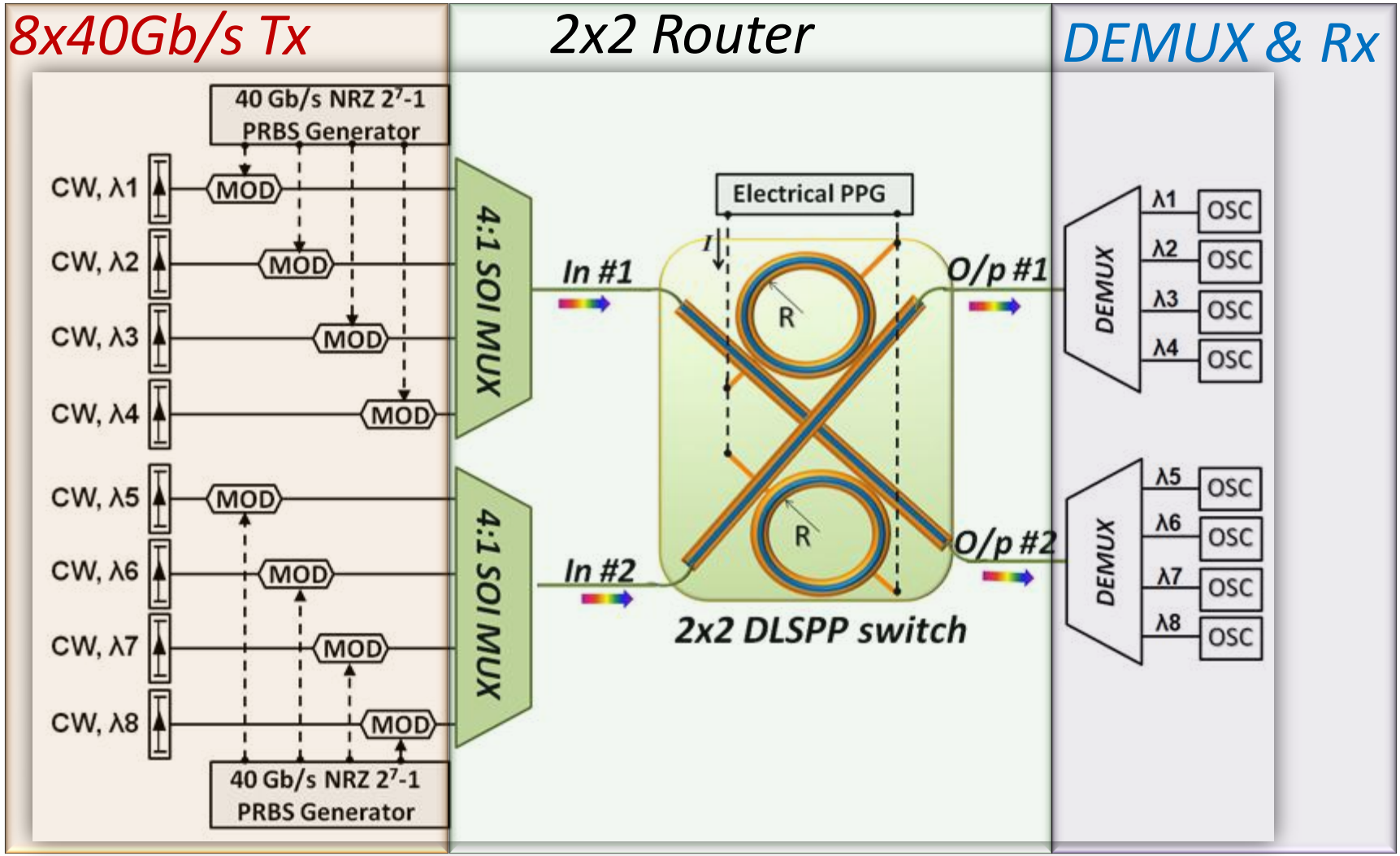
low loss technology hosting platform: waveguides, MUX, couplers, photodiodes, fiber coupling

Si-Plasmonic Router



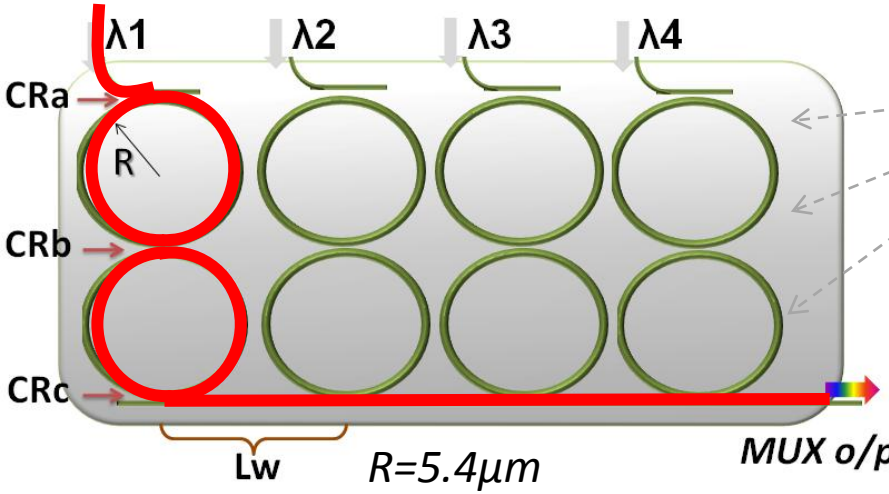
- *7xλ data packets at 40Gb/s : 280 Gb/s per input port*
- *1 extra wavelength for header (MHz data pulses)*
- *Time-offset between Header and Payload information for ensuring header processing in the IC (burst-mode network concept)*

A 320Gb/s 2x2 architecture

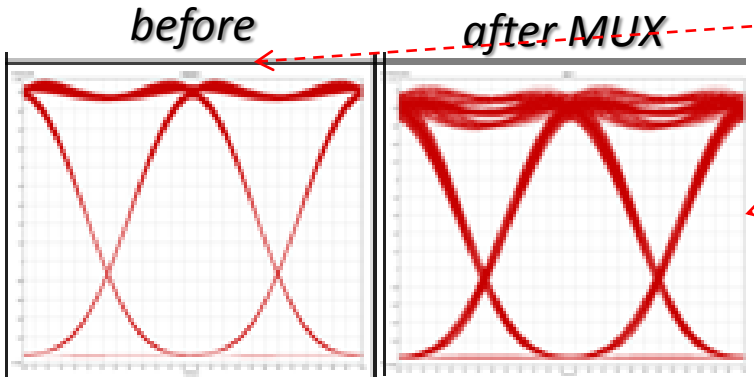


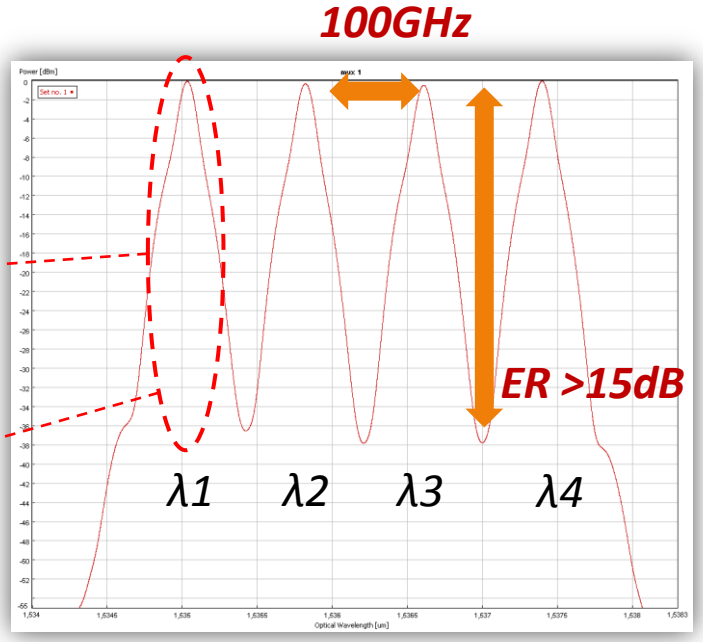
40Gb/s NRZ 4:1 SOI MUX

- ✓ *4 cascaded 2nd order silicon rings*



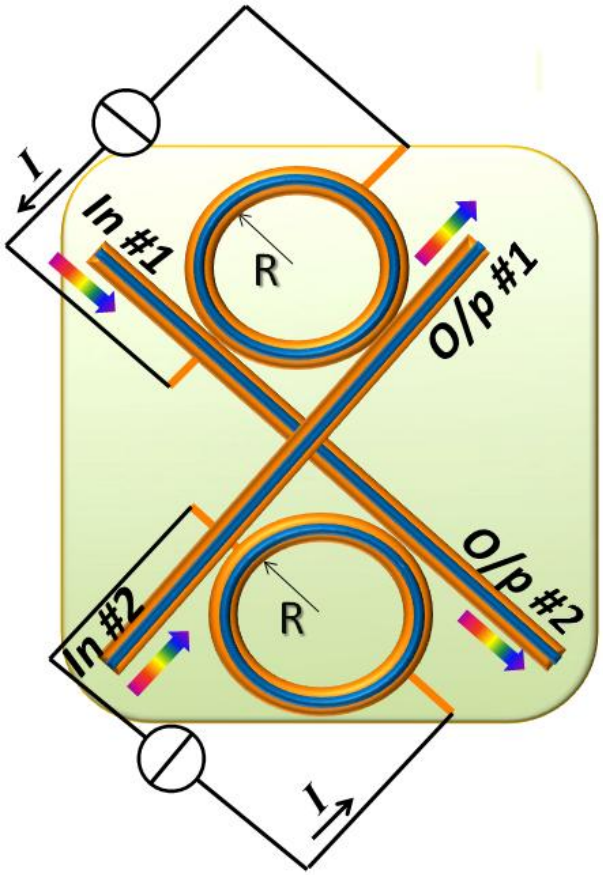
Gaps:
g1=200nm (power coupling of 0.06)
g2=460nm (power coupling of 0.0007)
g3= 200nm (power coupling of 0.06)

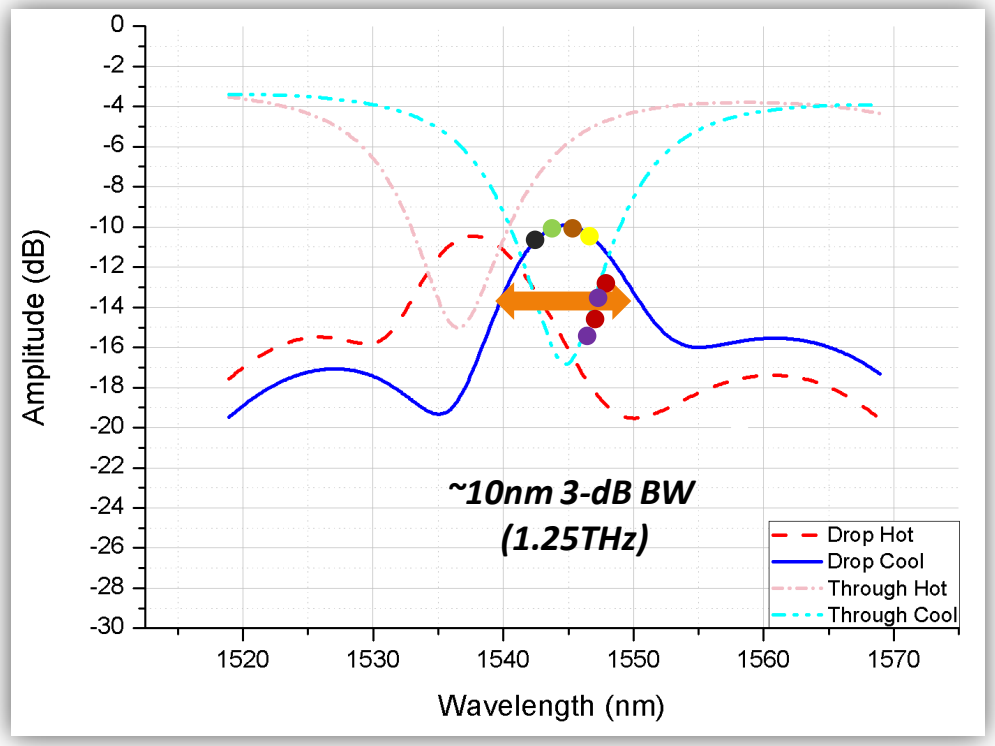




Broadband 2x2 Plasmonic Switch

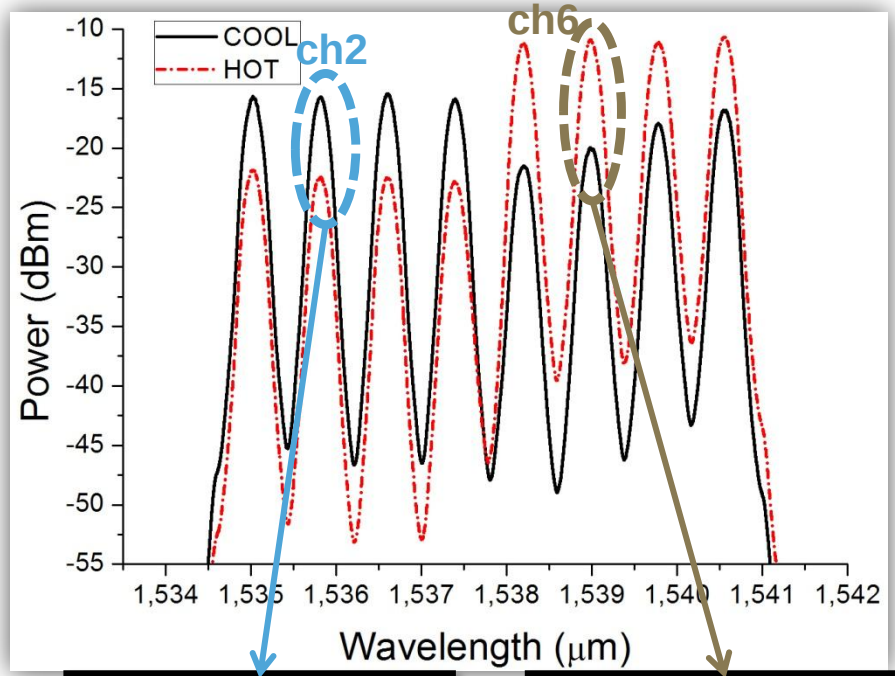
- ✓ *dual plasmonic ring resonator*
  - *R=5μm*

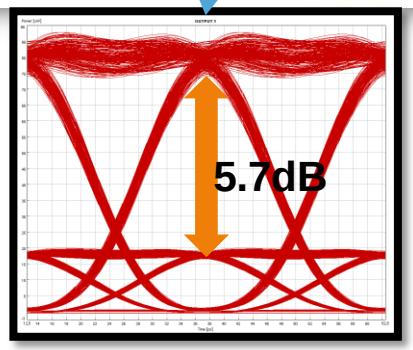


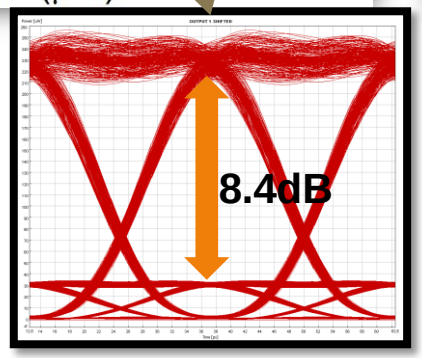


320Gb/s throughput routing

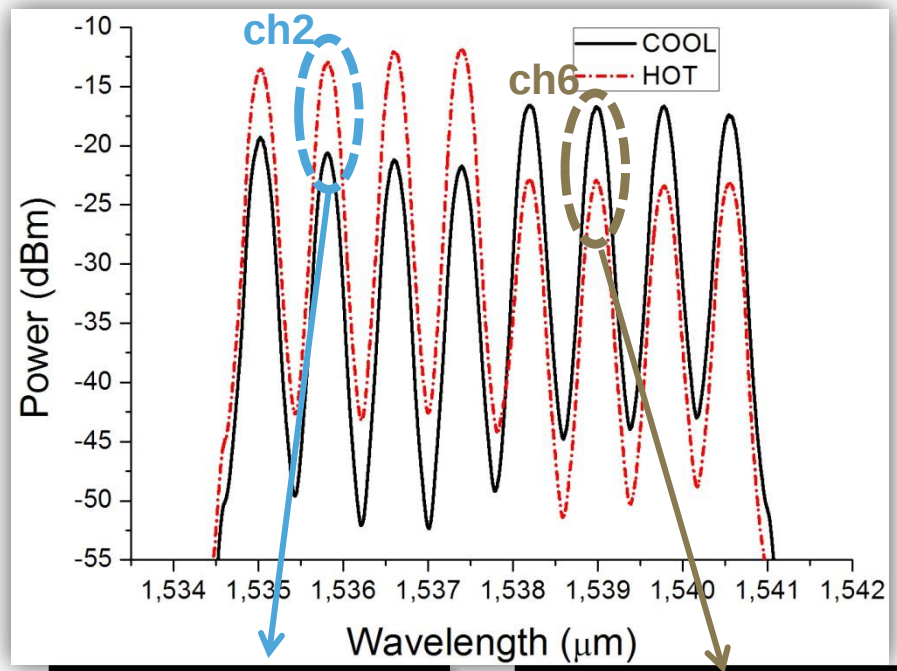
Output 1

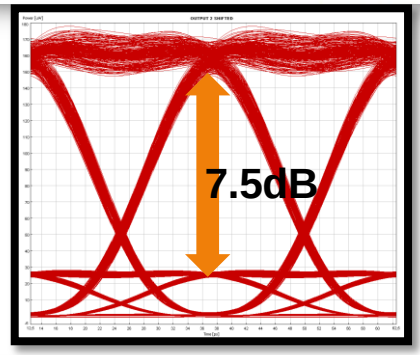


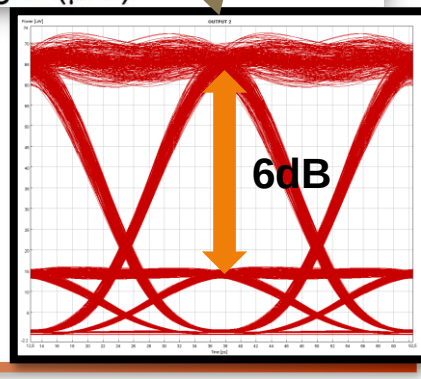




Output 2

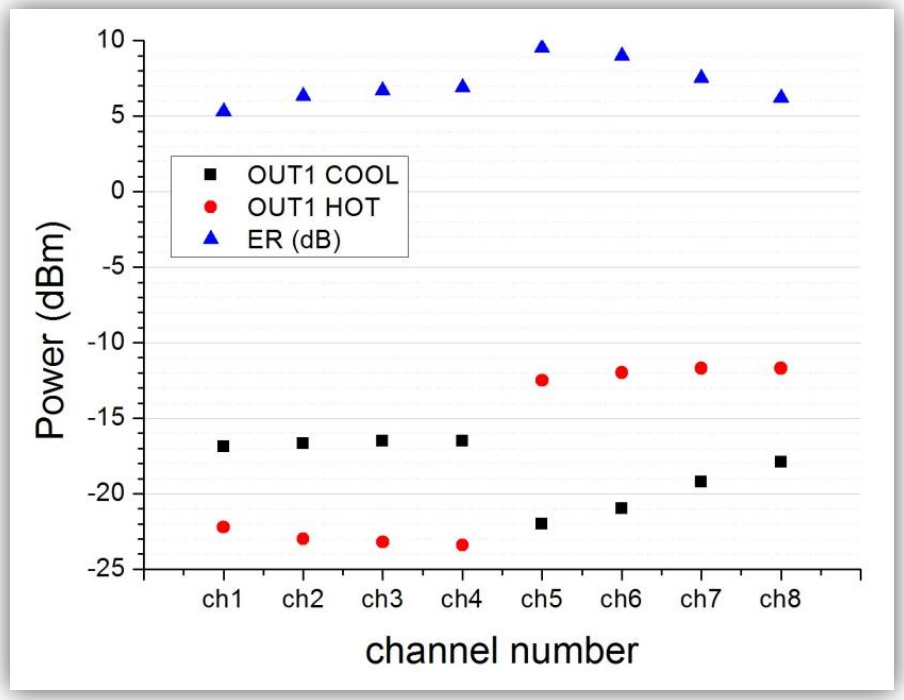




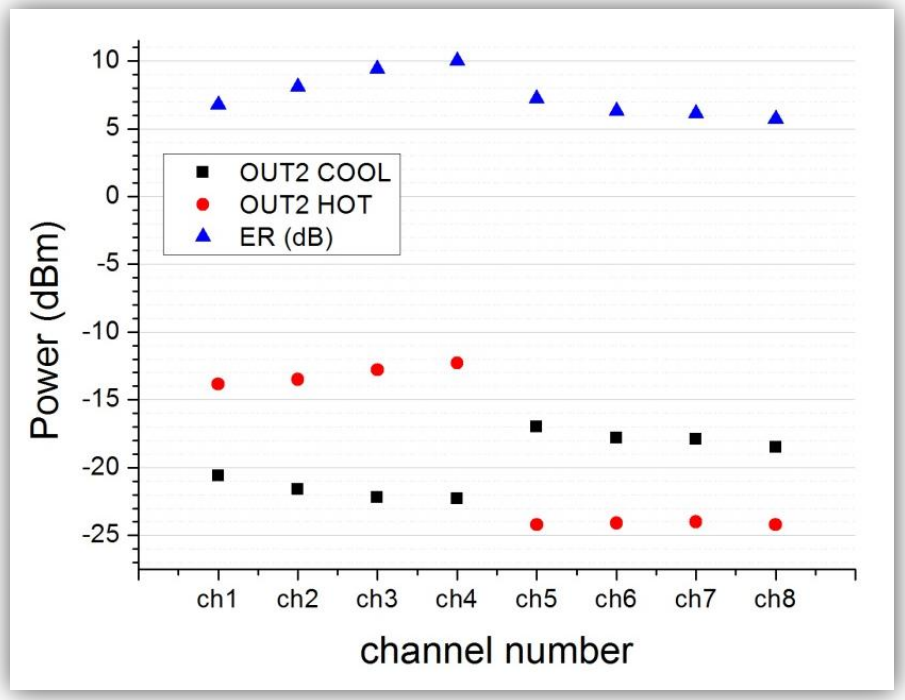


320Gb/s throughput routing

Output 1



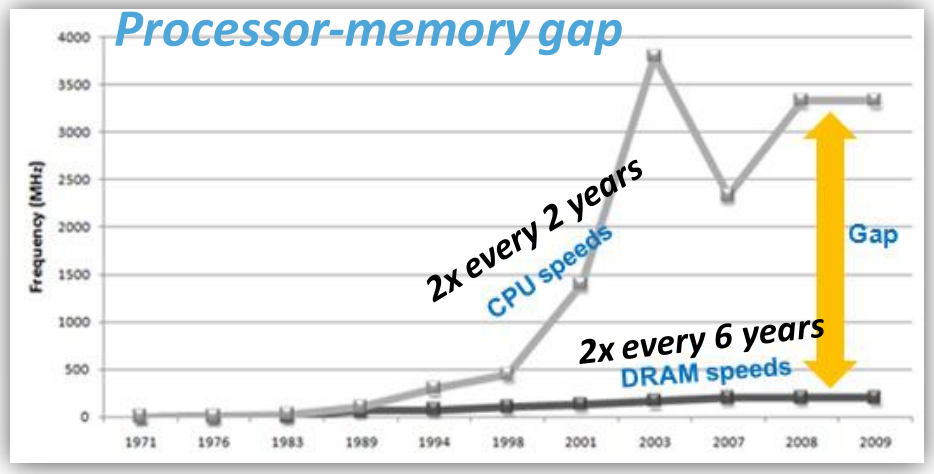
Output 2

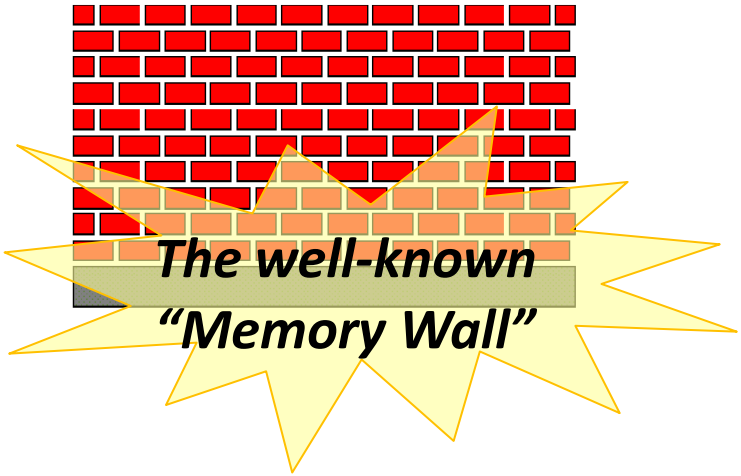


All channels having ER between 5.5 and 10 dB

What about buffering in HPC?

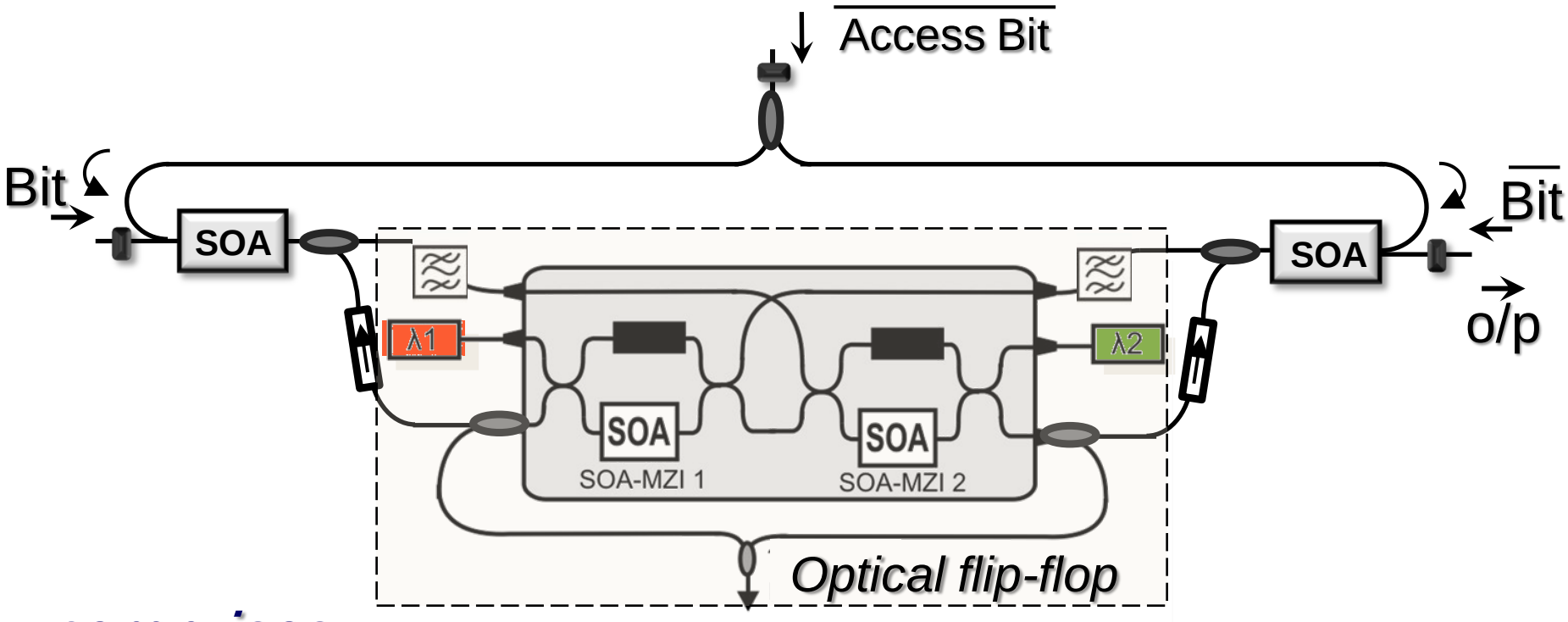
- *Latency of the entire HPC is limited by the nsec access time of electronic RAM*





- *…but electronic RAM is the only available solution for the HPC Storage Area*

Optical RAM

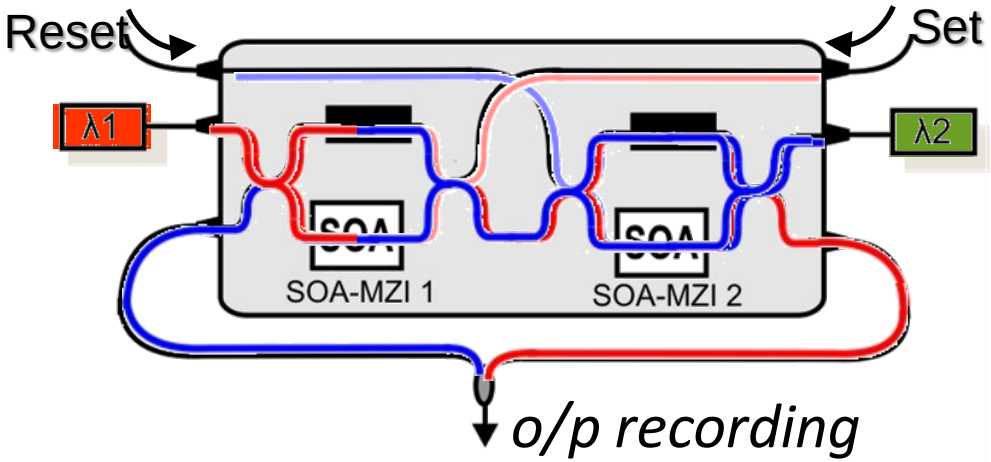


comprises:

- *integrated optical flip-flop as memory unit*
- *2 'ON-OFF' SOA switches controlled by* $\overline{\text{Access Bit}}$

Optical RAM

Memory unit:



Optical flip-flop using 2 coupled optical switches

- *Memory content = logical '1' when λ1 dominant*
- *Memory content = logical '0' when λ2 dominant*

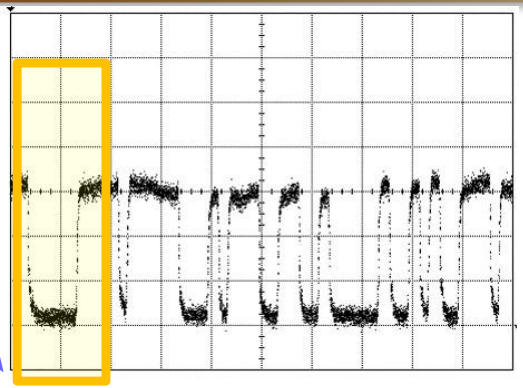
5GHz Optical Random Access Read

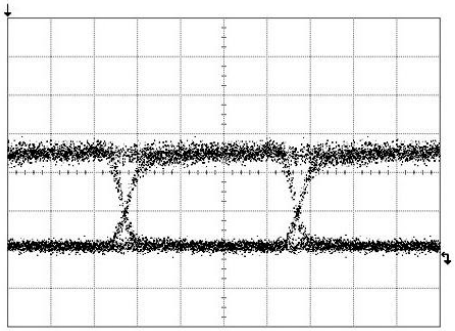
Inverted Access Bit

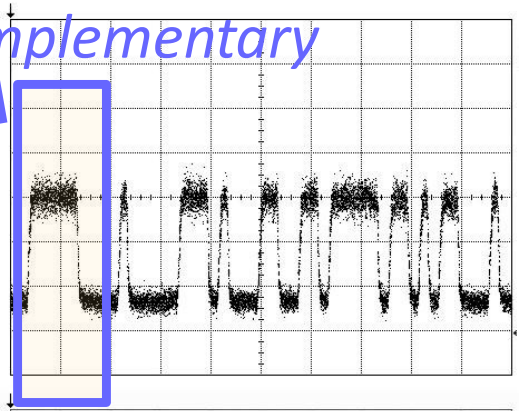
complementary

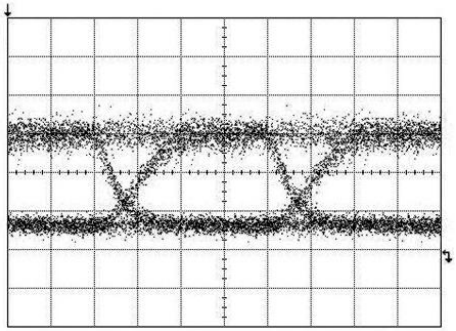
Read o/p @ $\lambda_{FF\#1}$ 1556nm

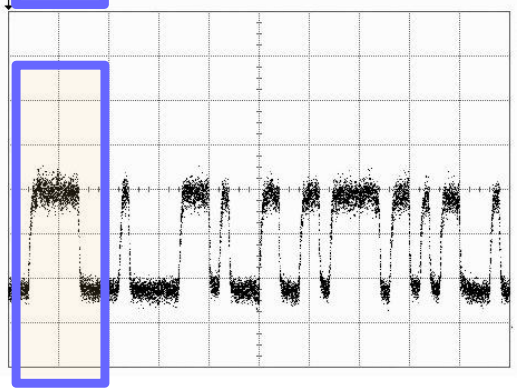
Read o/p @ $\lambda_{FF\#2}$ 1559nm

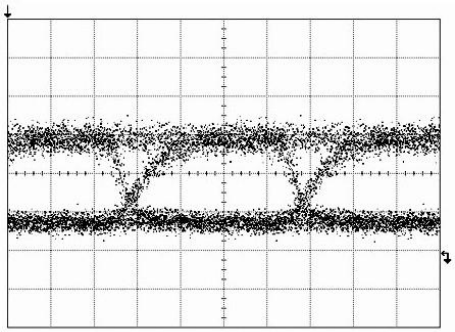












5GHz Optical Random Access Write

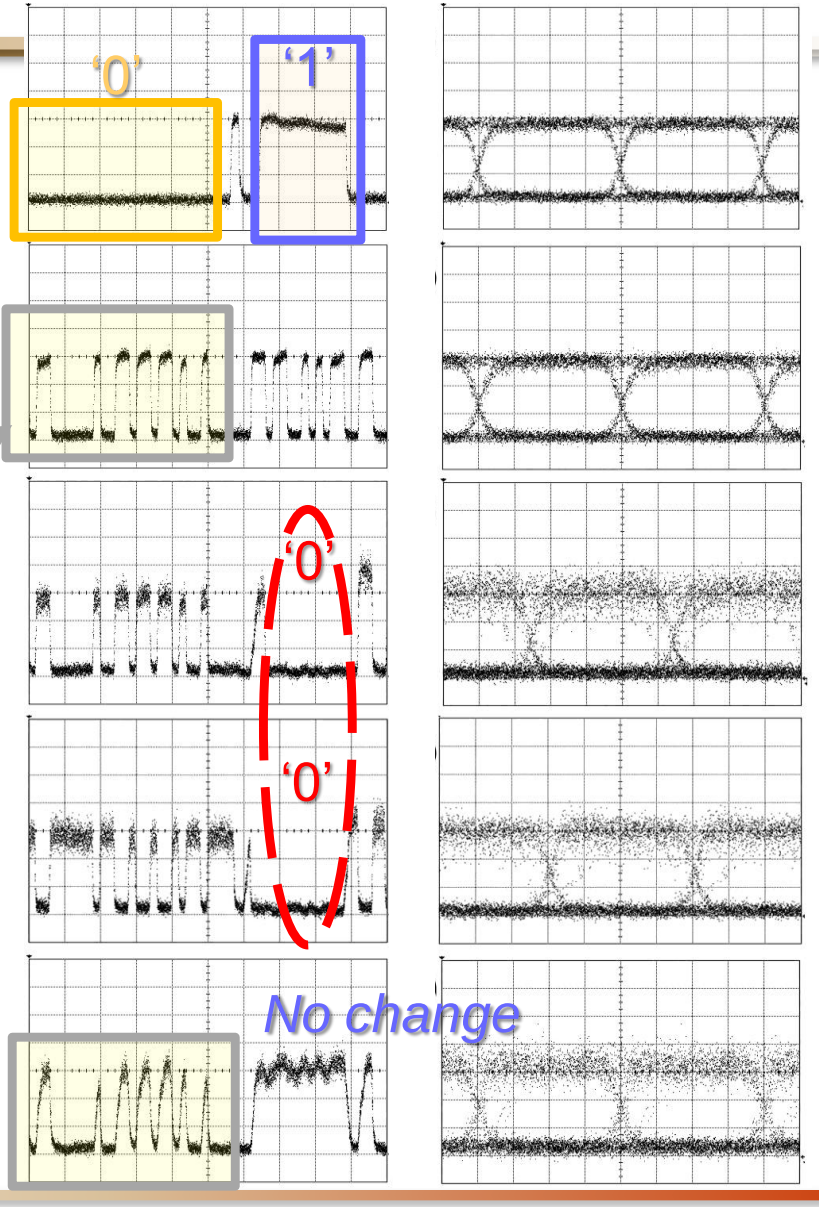
Inverted Access Bit

Incoming Bit signal

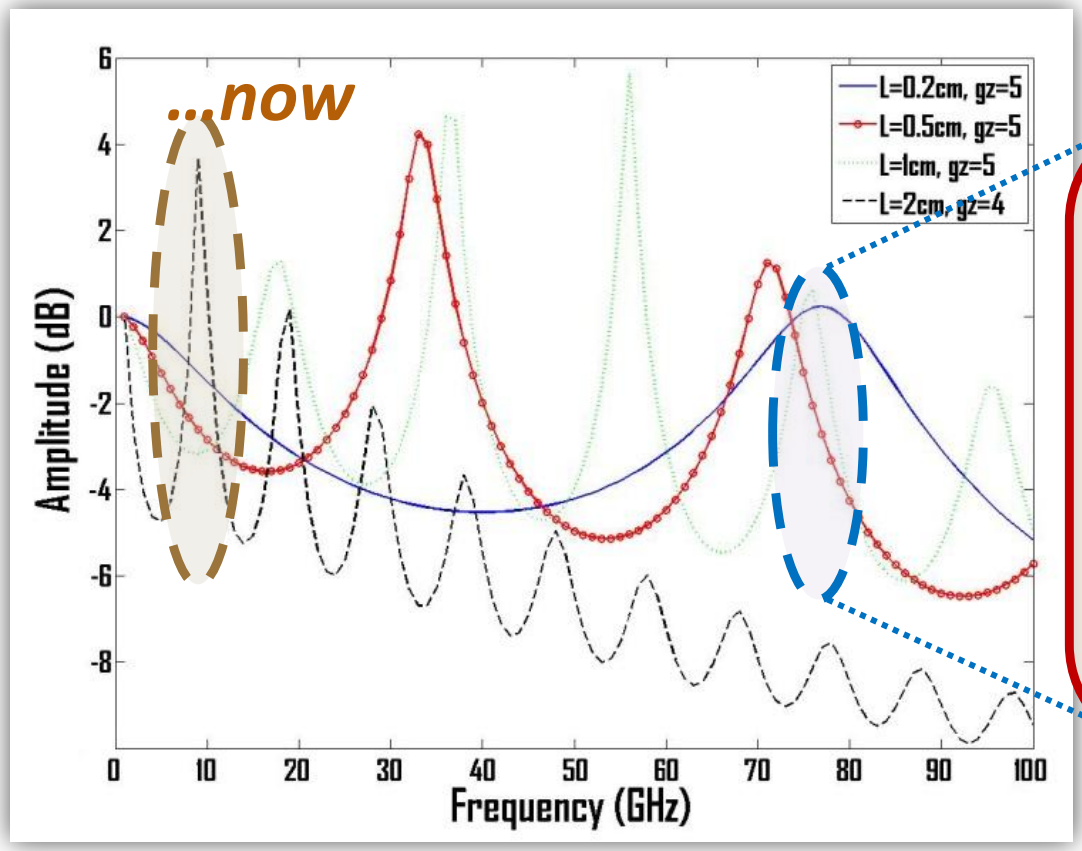
'Reset' signal

'Set' signal

Memory content



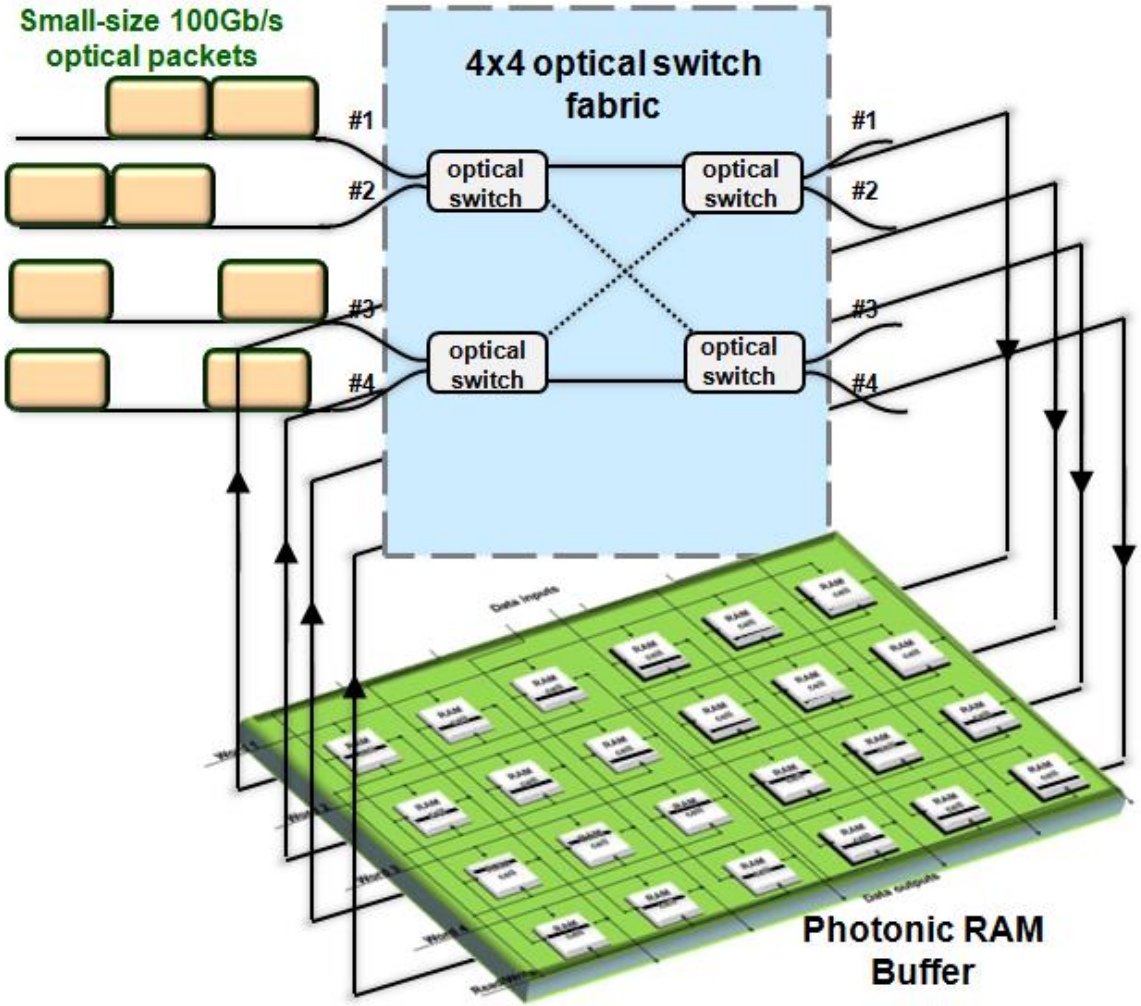
Towards 100GHz Optical RAM

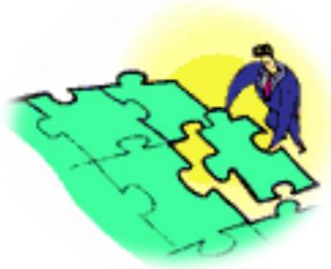


Optimized circuit design and silicon-integration can lead to 100GHz Read/Write

RAM Speed ~ c/nL

Towards true all-optical routers





THANK YOU !

The PhosNET team…

- *T. Alexoudi, D. Fitsios, G. Kalfas, G.T. Kanellos, A. Miliou, S. Papaioannou, D. Tsiokos , K. Vyrsokinos*

PhoS-net Research Laboratory