Engineering a Bandwidth-Scalable Optical Layer for a 3D Multi-core Processor with Awareness of Layout Constraints

Luca Ramini[1], *Davide Bertozzi*[1] *and Luca P. Carloni*[2]

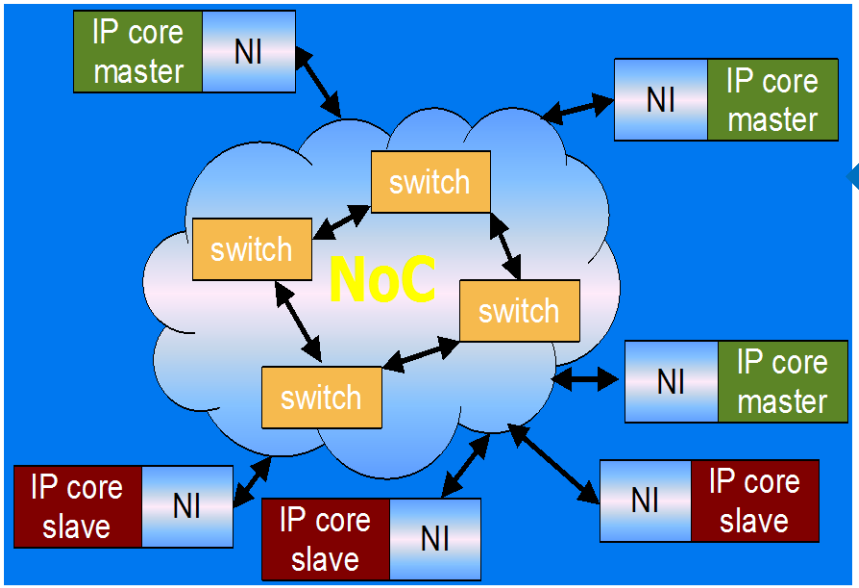
[1] *UNIVERSITY OF FERRARA*

[2] COLUMBIA UNIVERSITY

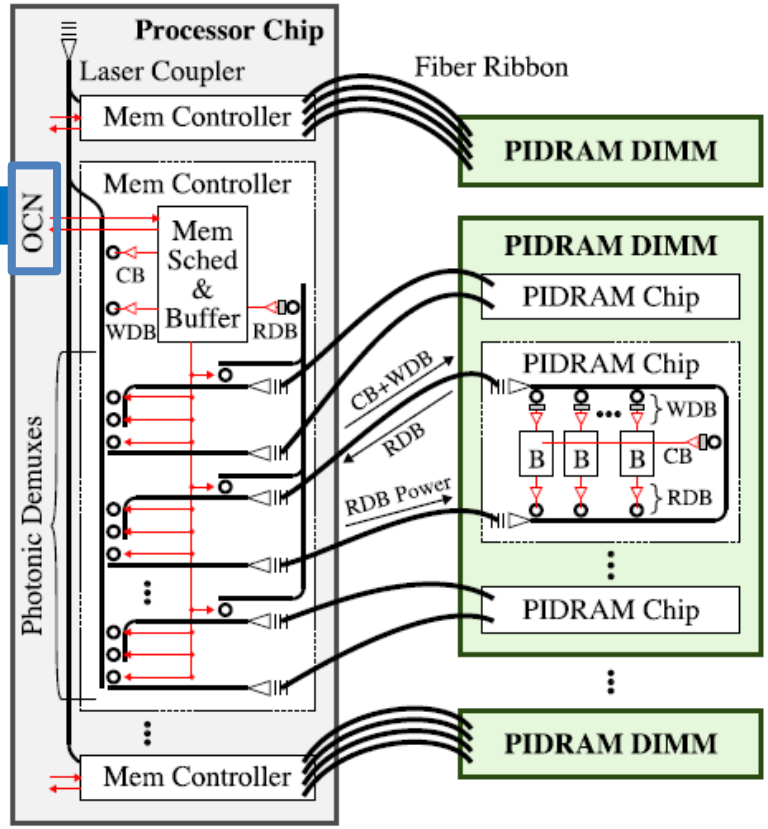
Trends and Challenges

- The **performance** of future multi-core processors **will only scale** with the number of integrated cores **if there is a corresponding increase in memory bandwidth.**
- **Silicon Photonic Technology** is being investigated as a way **to improve** pin **bandwidth density** and **power of DRAM memory devices.**

(S. Beamer et al., "Re-Architecting DRAM Memory Systems with Monolithically Integrated Silicon Photonics")



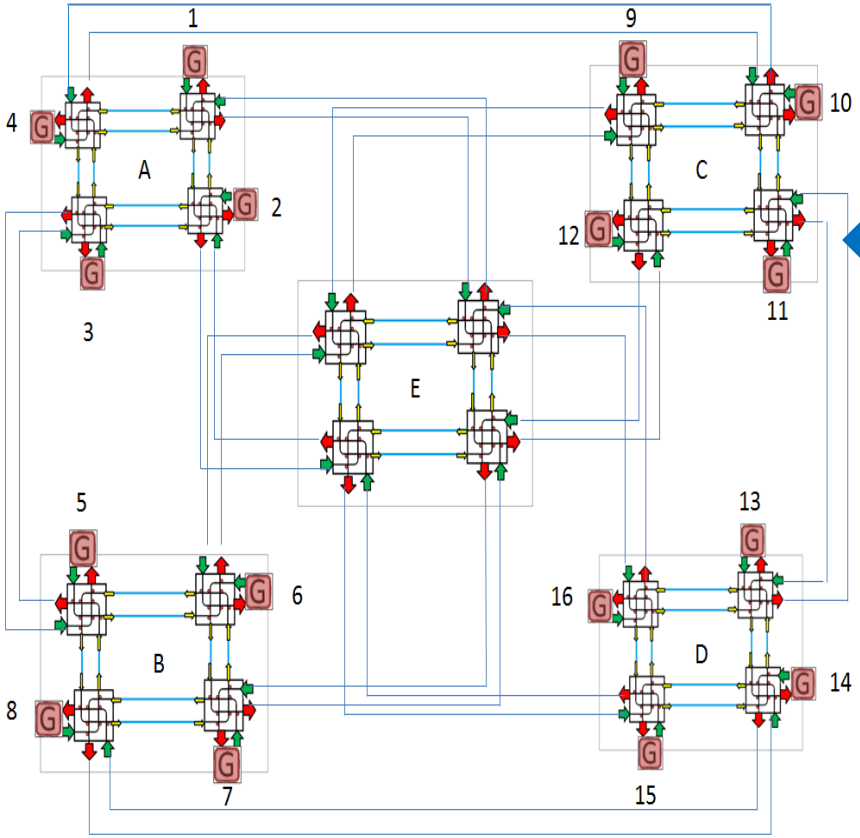
Processor-memory communication is typically accomplished by an **Electronic NoC:**

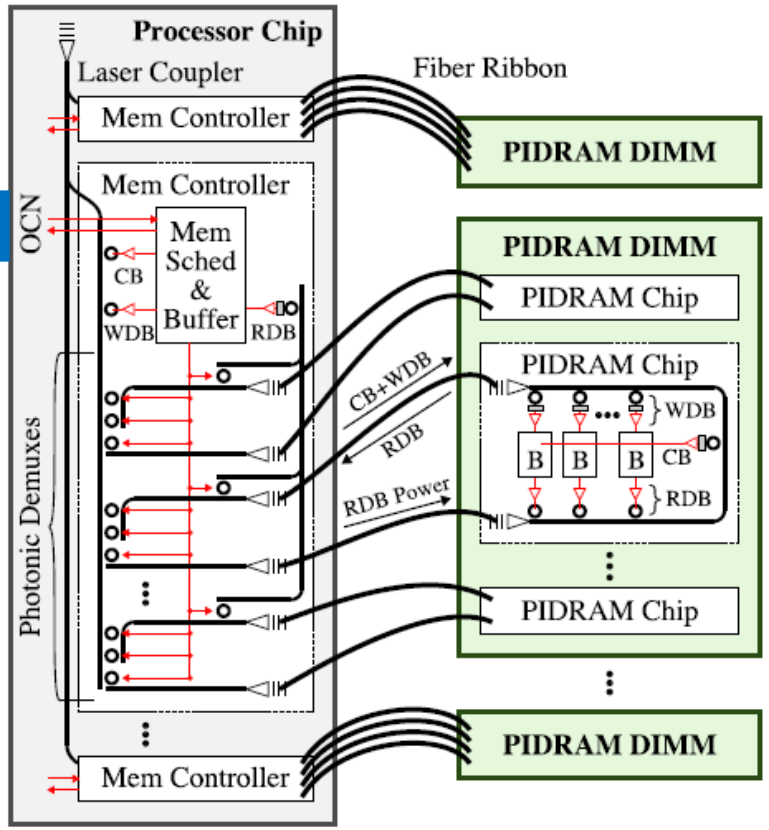


Trends and Challenges

Performance Gap between such Electronic NoCs and optical off-chip links **(high-bandwidth density, data-rate transparency , distance- independence)**

The only way to bridge this gap is to bring the **Photonic interconnect technology** deeper into the chip



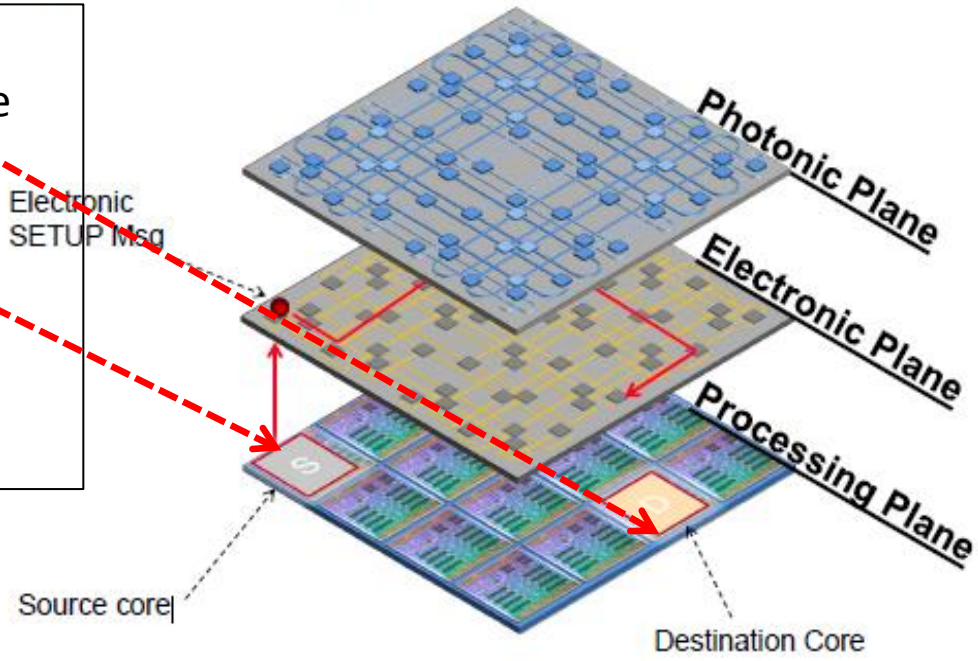


State-of-the-Art: Active ONoCs

3D STACKING APPROACH

In order to reserve a communication path between a couple **Source – Destination** the following steps must be accomplished :

1) **Path Setup Request**
2) **Path Ack**
3) **Transmission data**
4) **Teardown**



- Optical path control *(Shacham'07)* is expensive (hybrid NoC, path setup latency/contention)
- **Might not be the most appropriate mechanism** for **cost**- and/or **latency-constrained communications** (control applications where response time is the key metric**,** *Akesson2011*)
-**ALL-OPTICAL** approaches do exist, although require **frequent E/O and O/E conversions** *(Cianchetti'09)* or rely on **optimistic assumptions** on optical device properties *(Vantrease'08)*

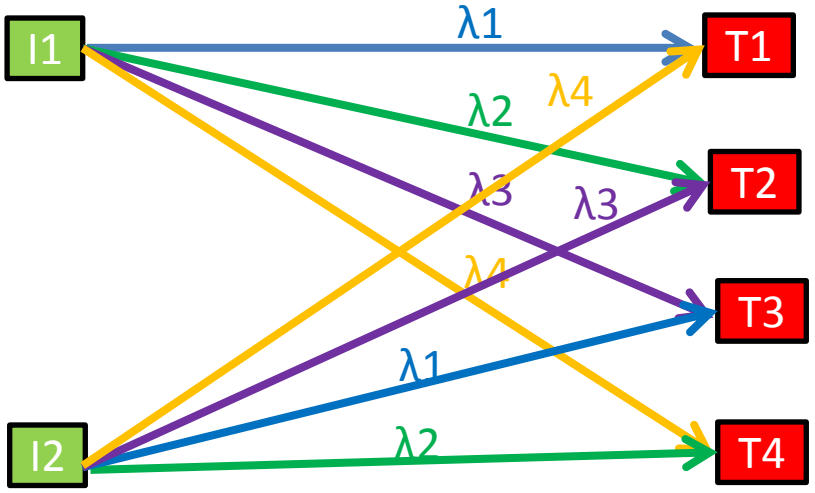
Our choice: ***Passive*** *Photonic NoCs (PPNoCs)*

WAVELENGTH-SELECTIVE -ROUTING

Packet routing depends solely on the wavelength of **its carrier signal.**
It **is configured at design time for a source-destination pair**

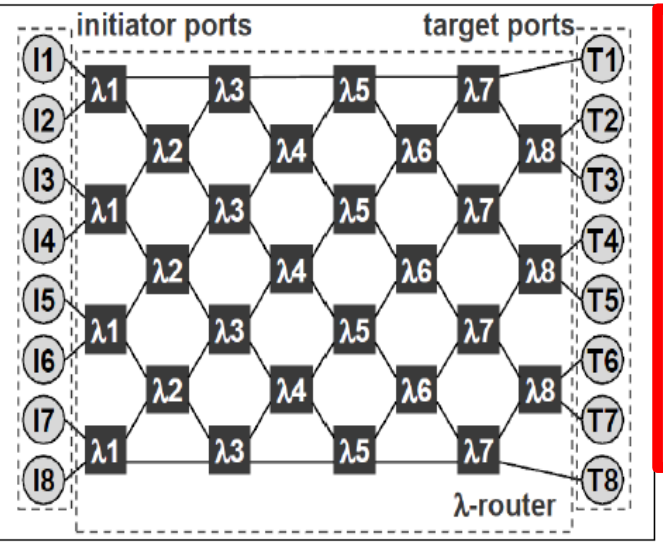
- ❑ **It does not depend on** ongoing transmissions by other nodes
- ❑ **No time is spent in Routing/ Arbitration**

Appealing property for a Processor-Memory network in mixed criticality systems



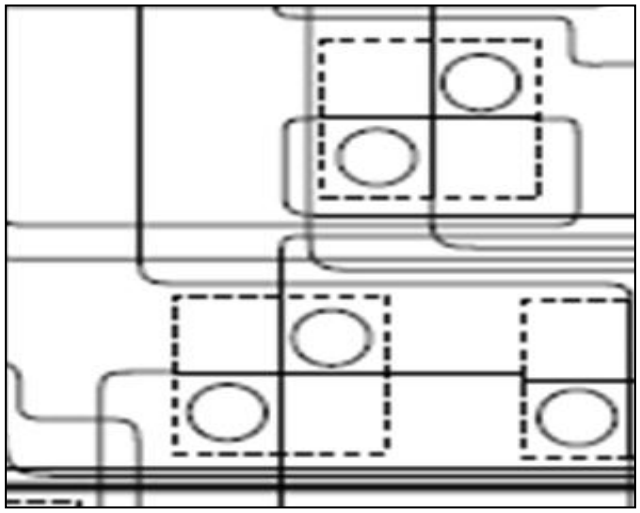
Although **PPNoCs** are well known in literature,
the implications of their actual layout constraints have been mostly overlooked so far, thus resulting in theoretical results with **poor practical relevance**

Key Contributions: Layout Constraints



Layout constraints **question** the practical feasibility of appealing **logic topologies**

the design of their associated **physical topologies** is mandatory for realistic assessments



Key effect this work is going to quantify:

The number of waveguide crossings on the actual **layout may be much larger than in the logic scheme** due to the mapping constraint on a 2D surface

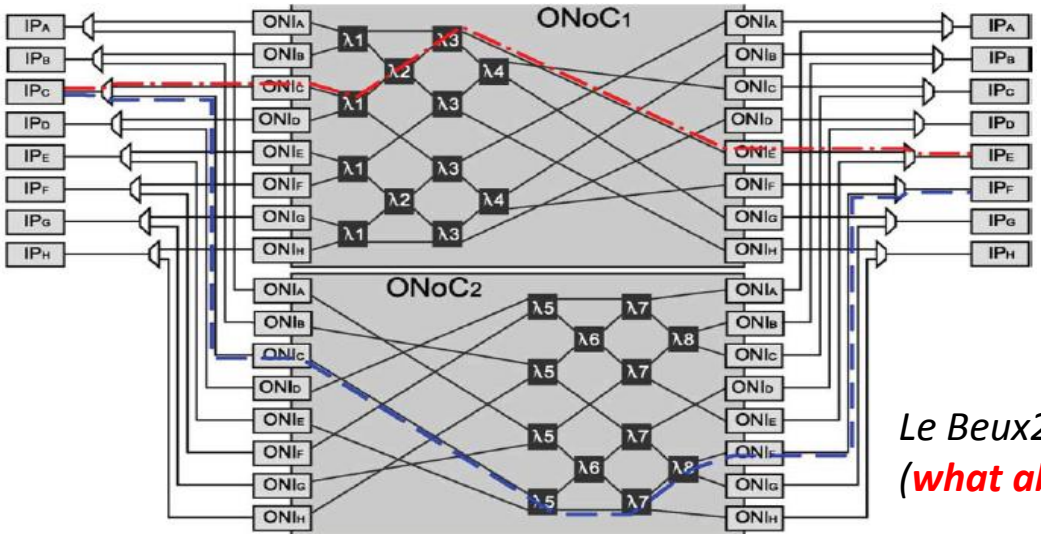
THE INSERTION LOSSES may DEGRADE to such an extent **that may render a topology unusable** or change relative topology comparison results

These effects are tightly design-specific, hence urging the choice for an experimental setting:
Processor-memory communication in a 3D stacked multi-core processor

Key Contributions: Network Partitioning

We question GLOBAL connectivity in PPNoCs and explore topology optimizations relying on the principle of network partitioning

Logic scheme



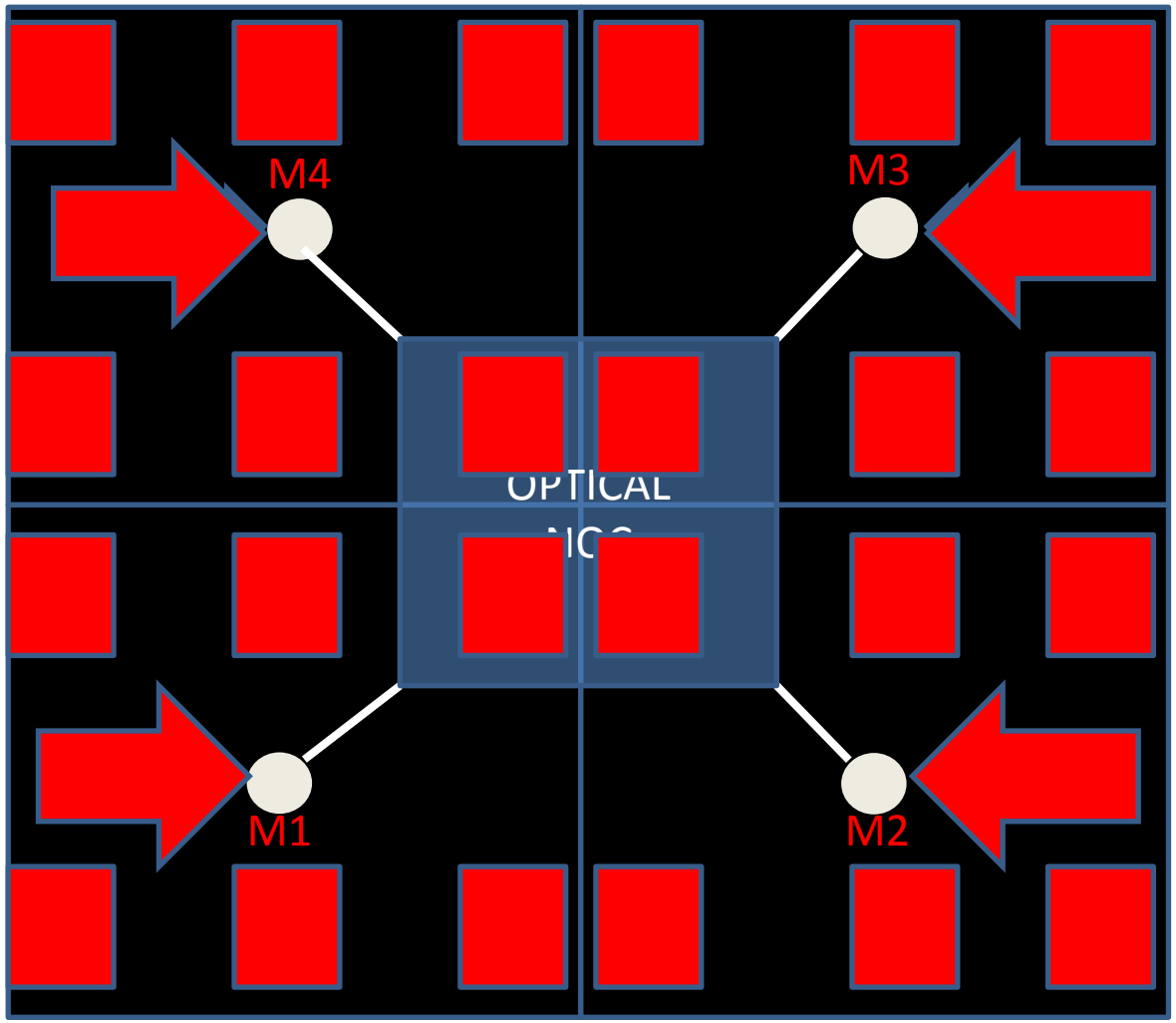
Le Beux2010
(what about the physical one?)

- Network partitioning as a way **of sharing wavelengths and laser sources**
- Network partitioning as a way **of simplifying connectivity patterns and improving physical design**
- Network partitioning as a way **of exploiting distinct traffic classes**

We aim at quantifying the insertion loss improvements that network partitioning can bring with respect to global connectivity

Key Contributions: Bandwith Scalability

We aim at exploring **Bandwidth Scalability Techniques** under a fixed number of network gateways and memory controllers, where just the number of cores of the electronic layer scales up.

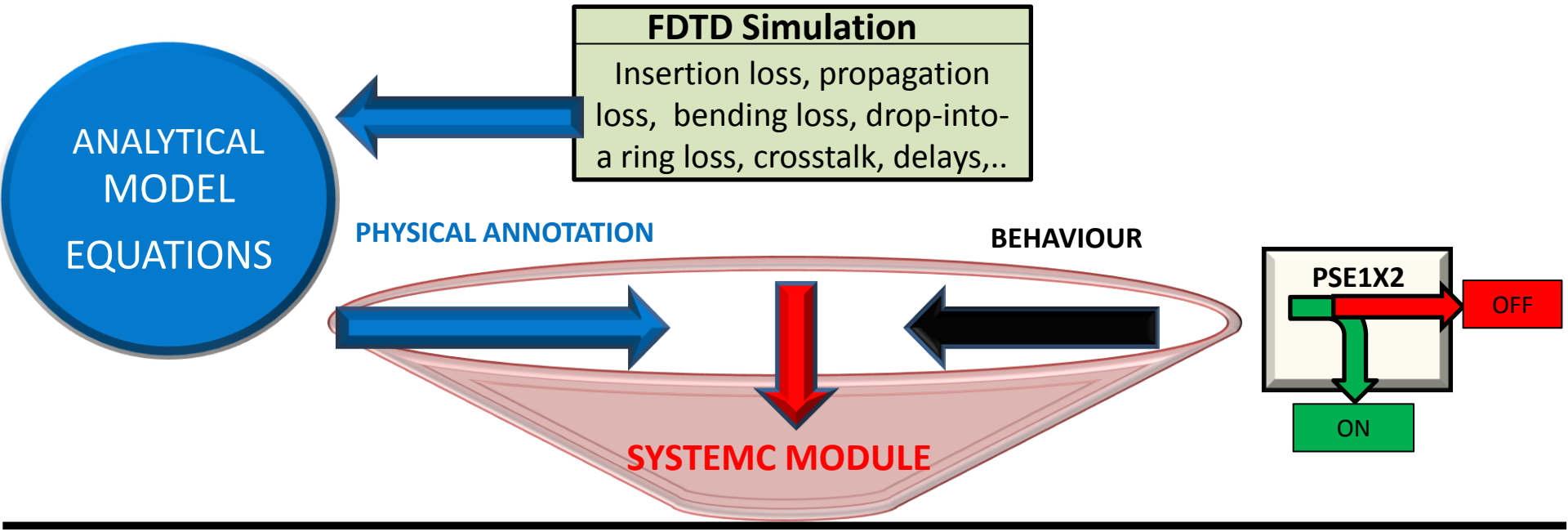


We present the first quantitative analysis of two relevant techniques**:**

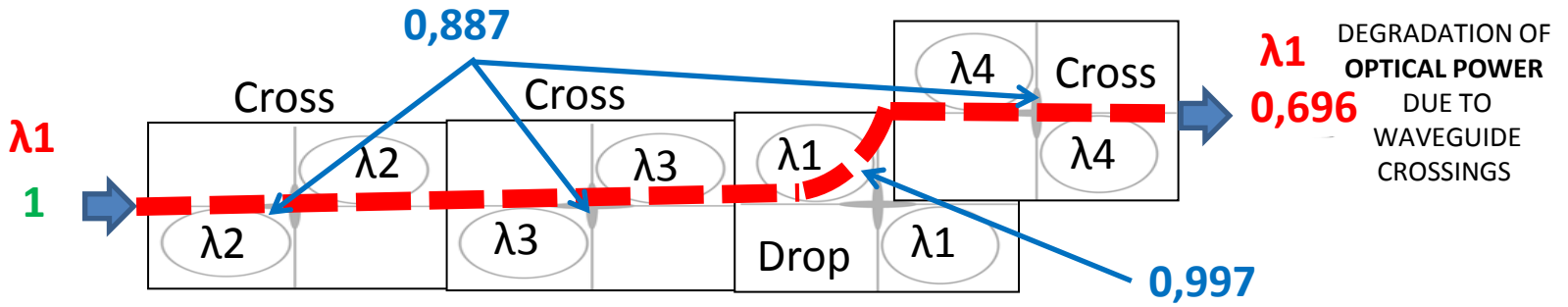
Spatial Parallelism (SPM) **and Broadband Passive Switching** (BPS).

Exploration Tool

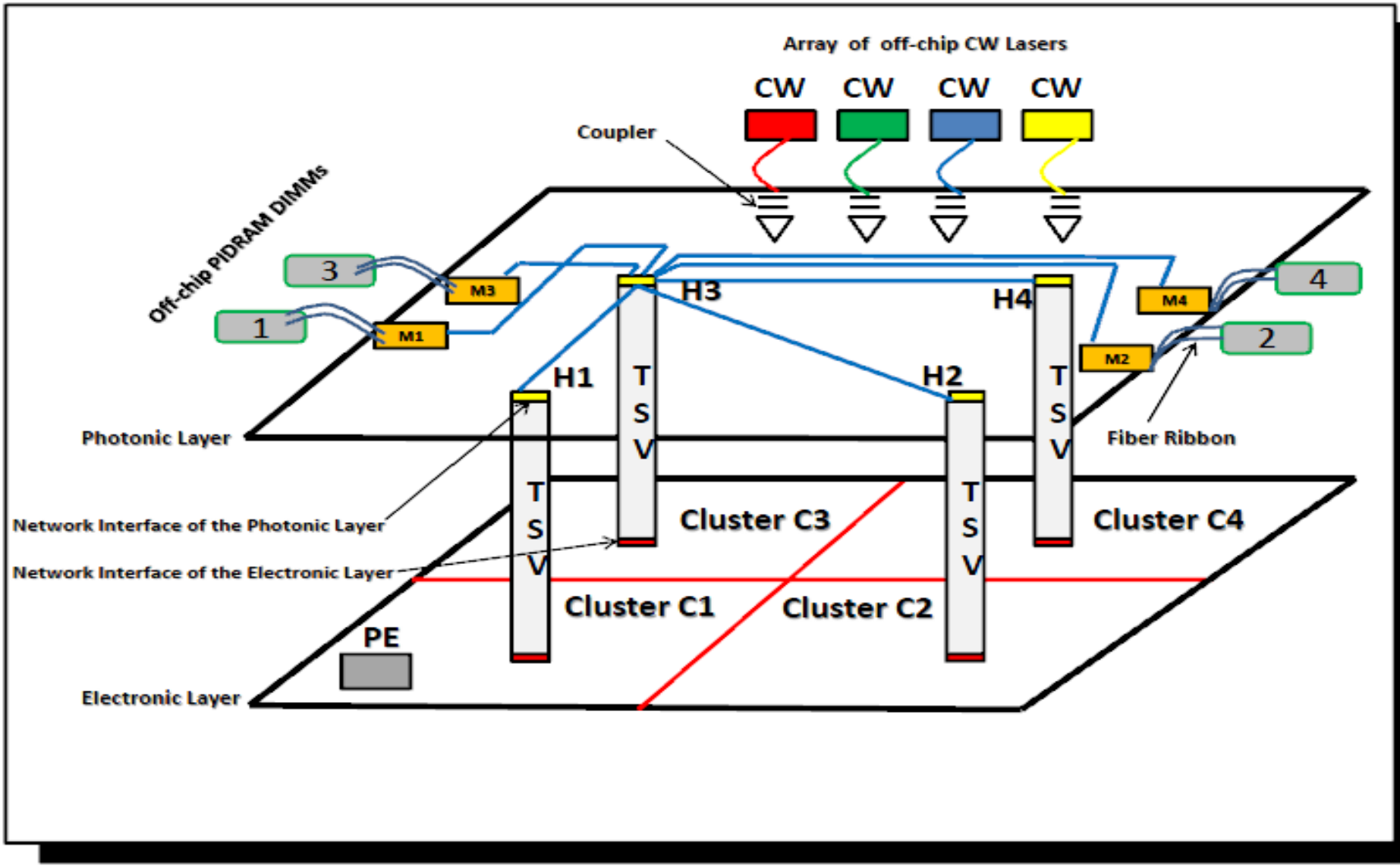
In order to preserve **technology-awareness** in the analysis, we rely on a **SystemC modeling and simulation environment** where routing functionality is merged with FDTD-derived technology annotations in the models of the optical devices.



Example



Target Architecture: 3D Stacked Multi-core Processor



Target Architecture: The Electronic Layer

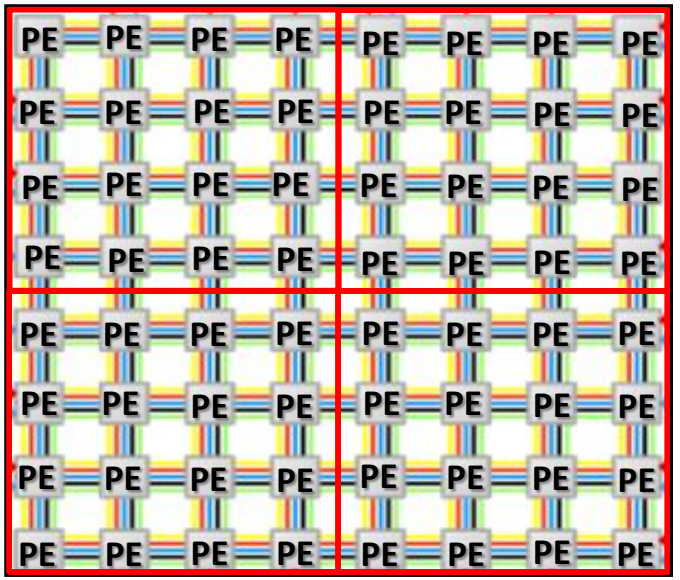
The Electronic Layer **consists of 64 homogeneous processor cores** connected by an Electronic NoC with a 2D Mesh Topology.

Assumptions:

Cores are grouped into **4 clusters *Ci* of 16 cores each**

Each cluster has **its own access to the optical layer** which is vertically stacked on top of the electronic layer.

E-NoC: 64 cores connected to a 2DMesh



Clusters and Aggregation Factor

The **number of cores** inside each cluster represents the ***Aggregation Factor(A.F.)***.

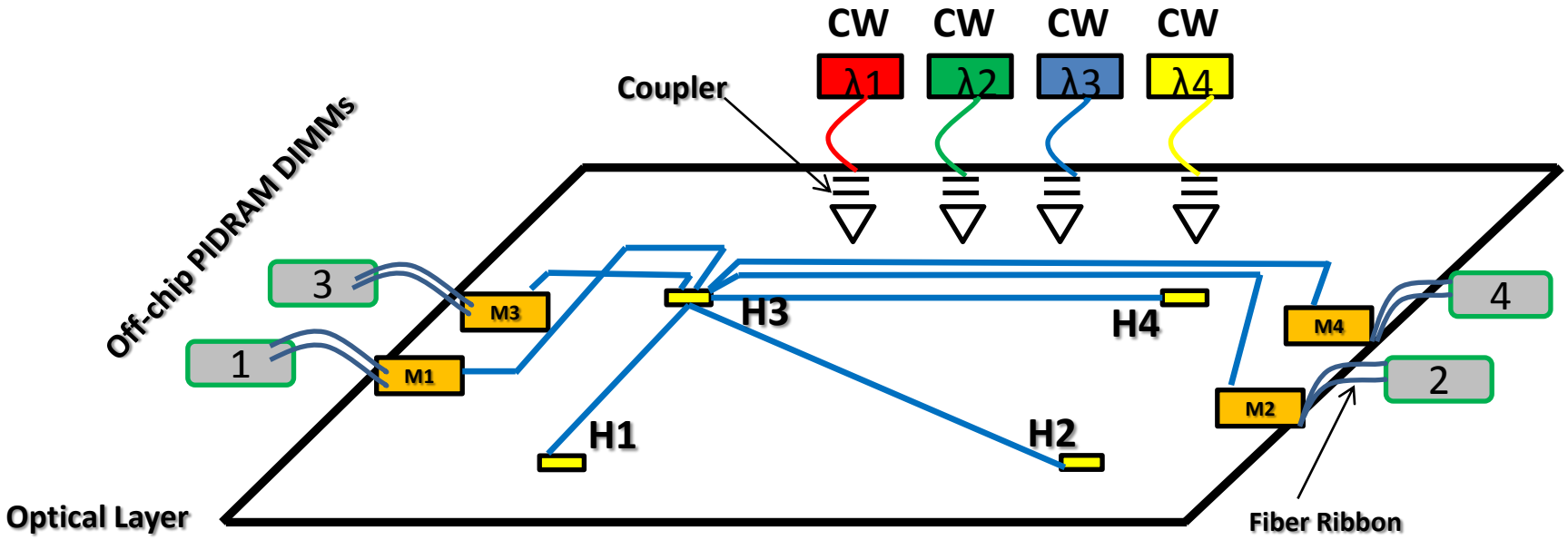
A.F. is **design- and technology- dependent**, since the cost (power and latency) for domain crossing dictates the most convenient boundary between the electronic and the optical N0C for cost-effective long range communication.

Target Architecture: The Optical Layer

The Cluster **Gateways** to the optical layer are defined as the ***Hubs (Hi)***

Optical Power: is provided by an array of off-chip Continuous Wave (CW) lasers.

Wavelength Sharing: the same wavelengths can be shared by all the Initiators.



The **Optical Layer** offers three kinds of communications:

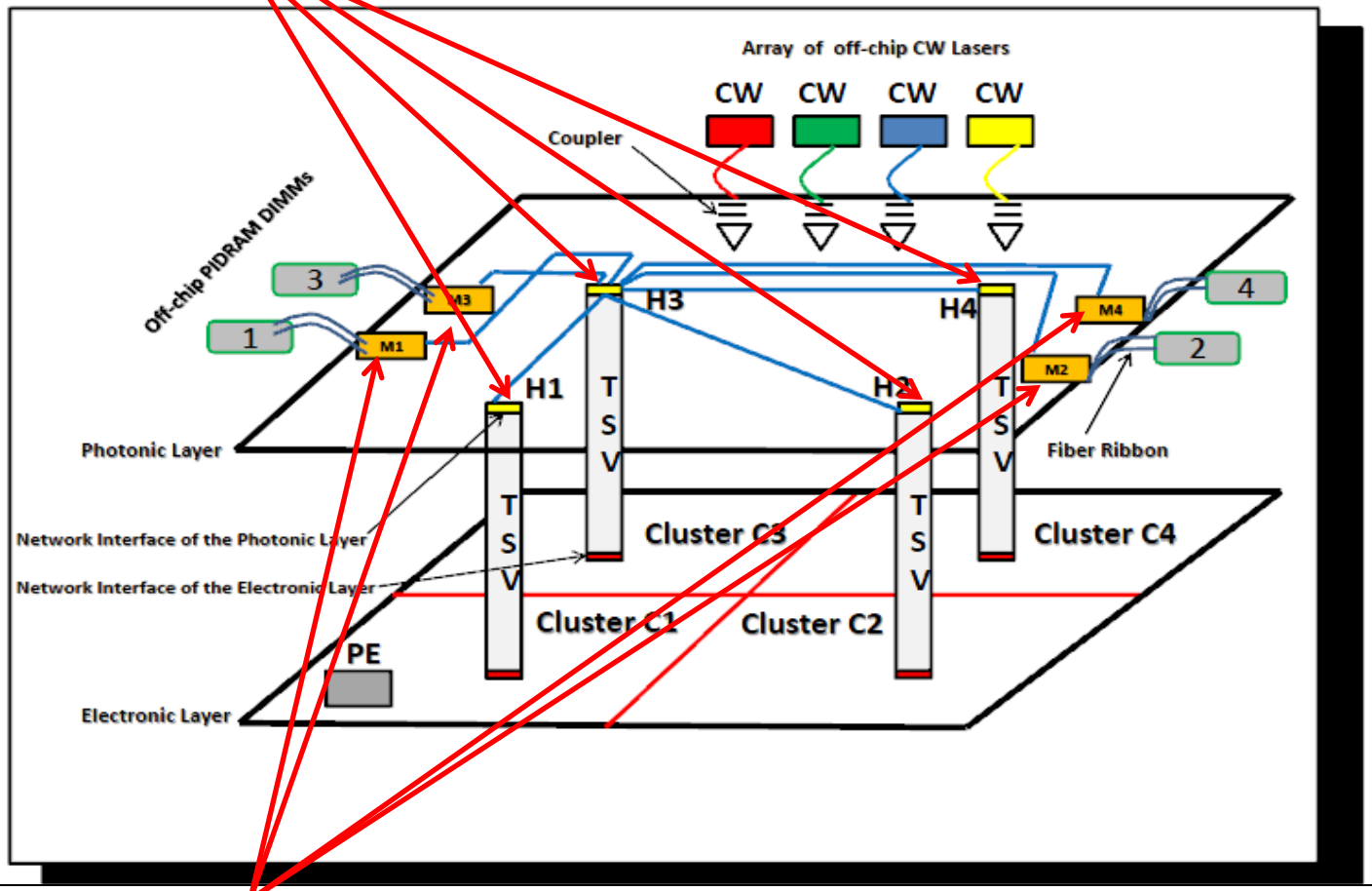
(a) **Among clusters**

(b) **From a cluster to a memory controller of an off-chip DRAM DIMM**

(c) **From a memory controller to a cluster**

LAYOUT CONSTRAINTS

Layout constraints : The Hubs are positioned in the middle of the clusters



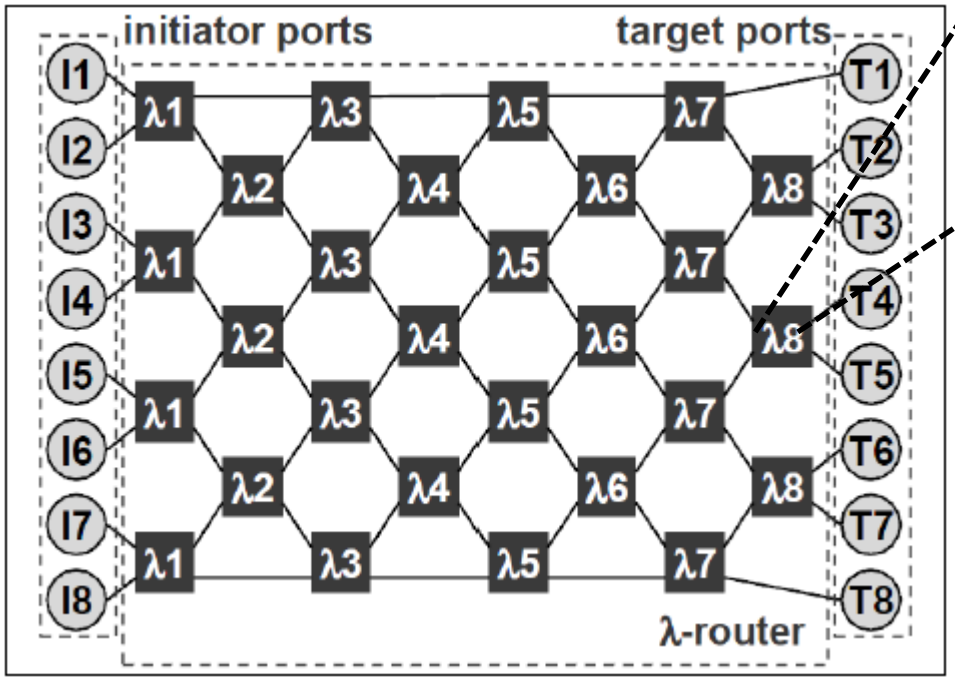
Layout constraints : The Memory Controllers are positioned pairwise at opposite positions of the chip thus reflecting a common industrial practice (e.g. Tilera TILE64)

Passive Optical NoC Design

The Passive optical layer consists of **8 initiators** that may communicate with **8 targets**

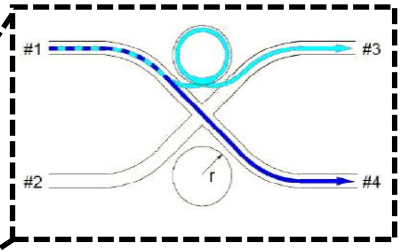
The **most straightforward solution** consists of an **8x8 Passive Optical-NoC (Global connectivity)**
We pick the **LAMBDA ROUTER** topology: **8 stages of 4 and 3 add-drop filters**

8x8 PPNoC



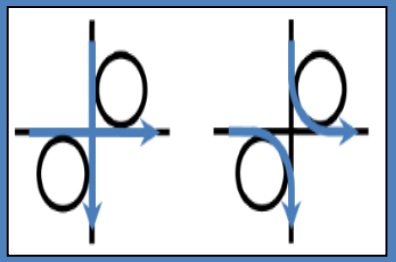
A. Scandurra and I.O'Connor, "Scalable CMOS-compatible photonic routing topologies for versatile networks on chip",NoC-Architeture,2008

2x2 Add-Drop Optical Filter



This solution needs **of 8 different Resonance Wavelegths**

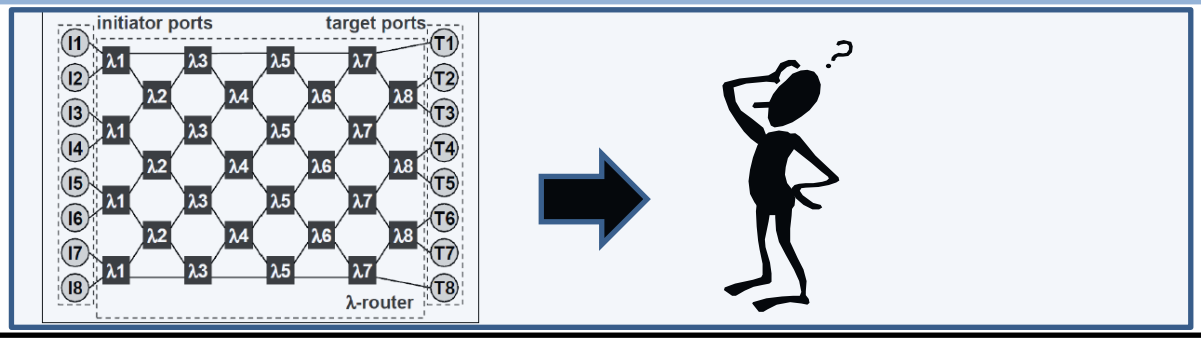
We replace their 2x2 ADF with a PSE 2x2



PSE 2x2

Easer layout design and same routing functionality

Passive Optical NoC Design



Since the Actual Floorplan **is subject to Specific Constraints** the Physical topology is **Radically different** from the Logic scheme

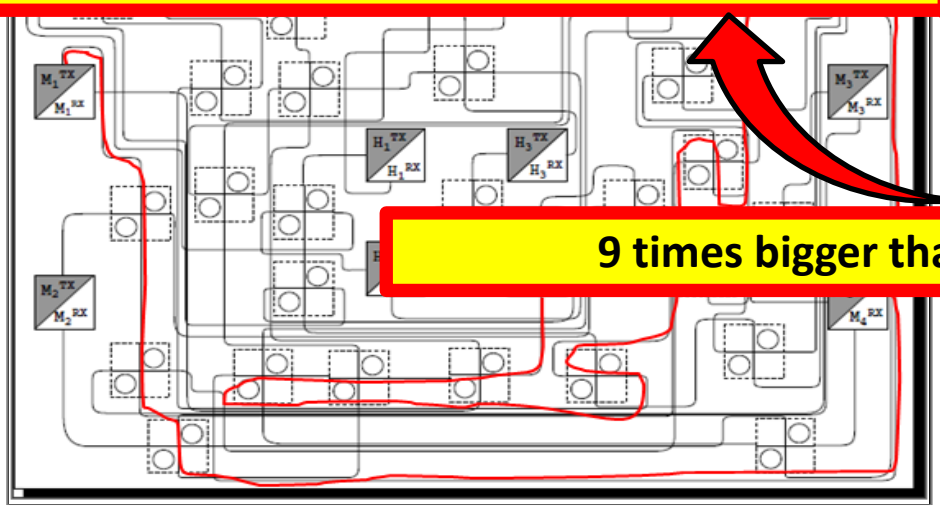
The logic scheme **does not fit** real-life placement constraints

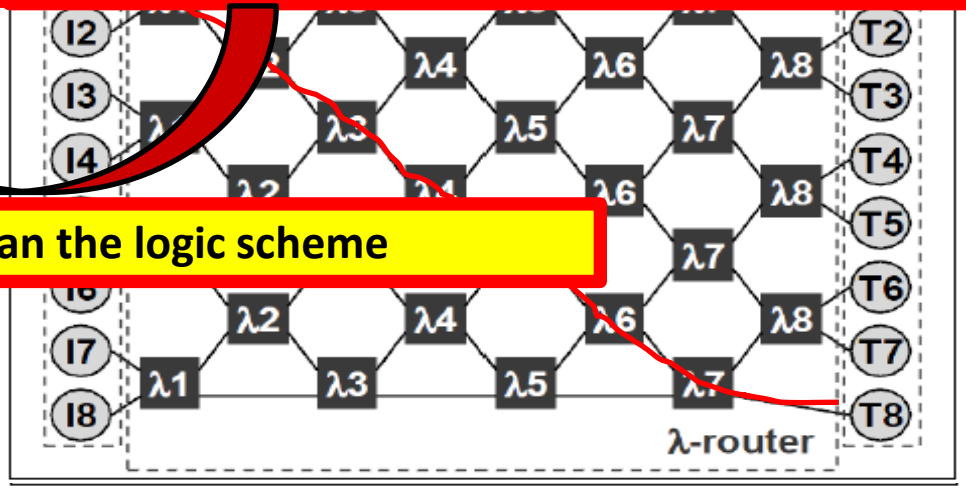
The logic scheme **imposes** that **all the initiators are placed on the left** of the Chip whereas **all the Targets on the right**.

Experimental Results (SystemC)

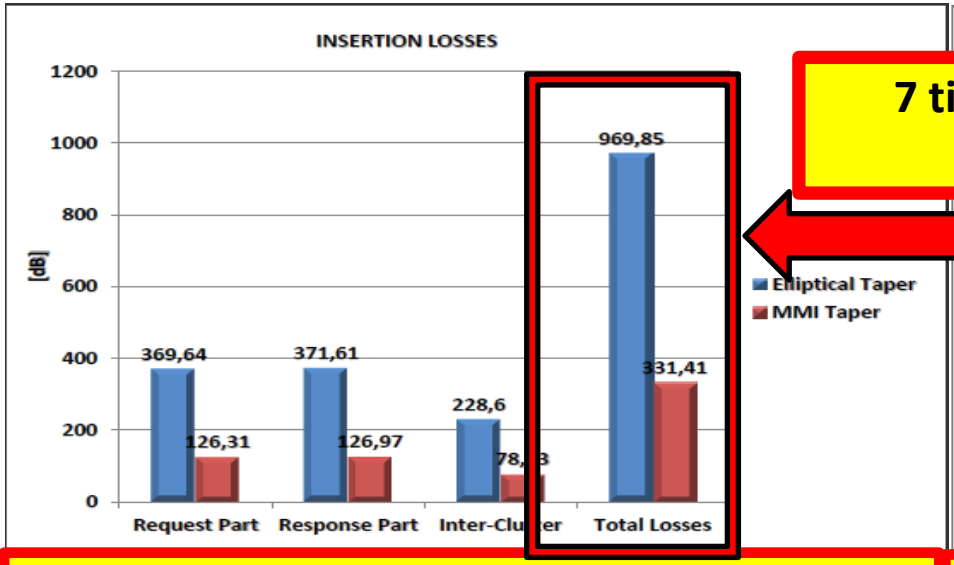
The critical path insertion-loss achieves 33.3 dB with E.T. and 11.4 dB with MMI Taper

The critical path insertion-loss achieves 3.6 dB with E.T. and 1.24 dB with MMI Taper

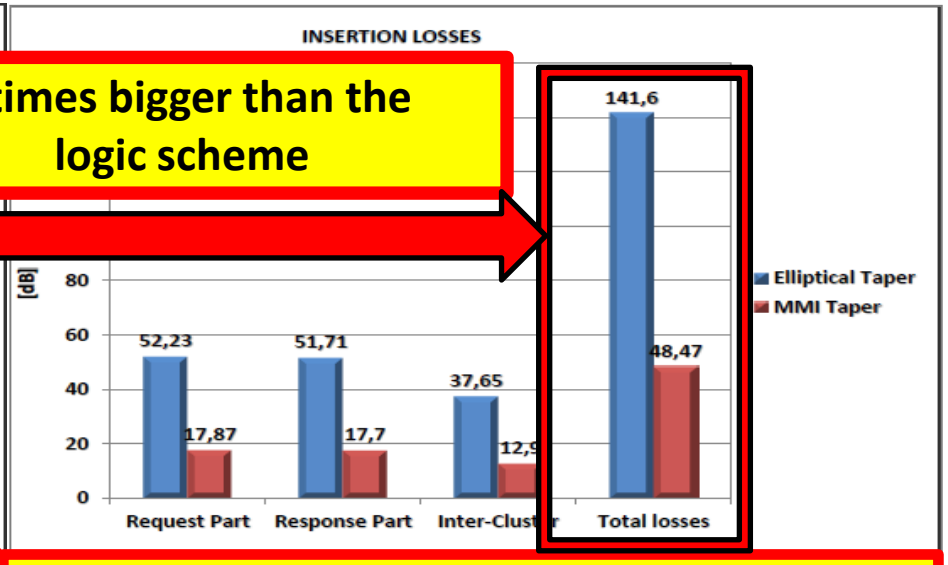




9 times bigger than the logic scheme



7 times bigger than the logic scheme



Total Losses are almost **7 times higher than ideal case, thus achieving 331 dB, with MMI Taper.**

Total Losses are **not capable** to stay below **48 dB, with MMI Taper at every intersection**

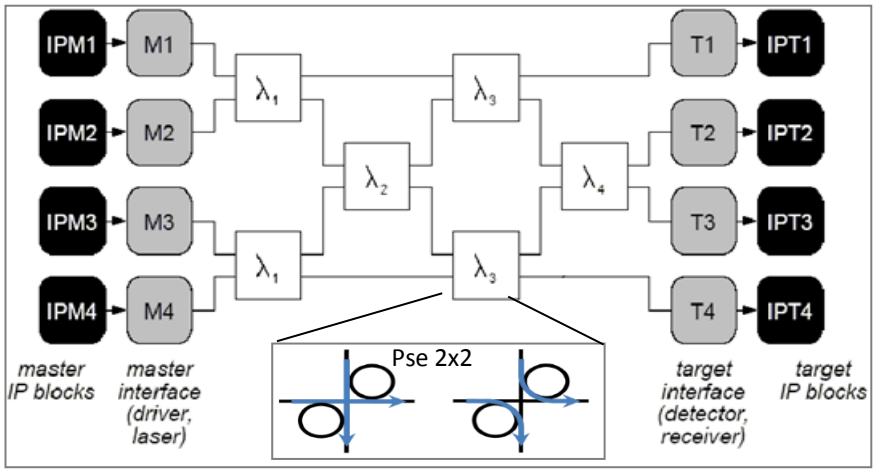
Partitioned Solution

The Global **PPNoC** is partitioned into 3 sub-networks, each dedicated to a different traffic class

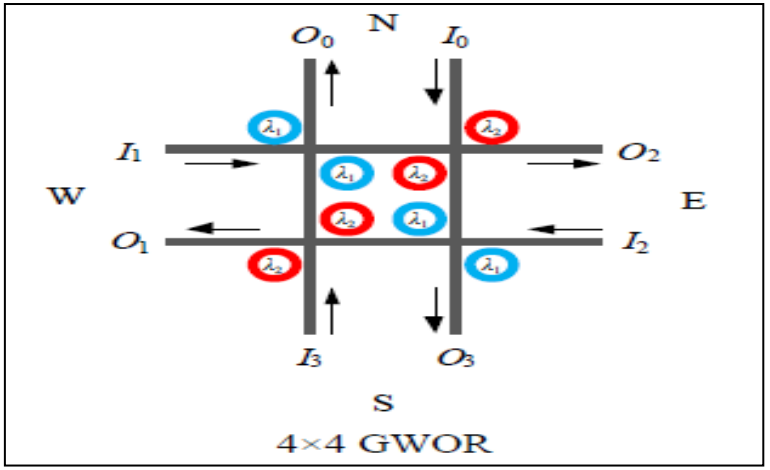
The **network for memory access requests** is obtained by scaling down the 8x8 PPNoC to 4 Initiators and 4 Targets (**4x4-λ Router)** .

In a similar way, we design the **network for memory responses** with the same features of **Request Network** .

We opt for a different topology for **Inter-Cluster Communications**: **4x4 GWOR, since** its scheme **has a good matching** with the placement of **HUBS** on the optical layer (along a square).

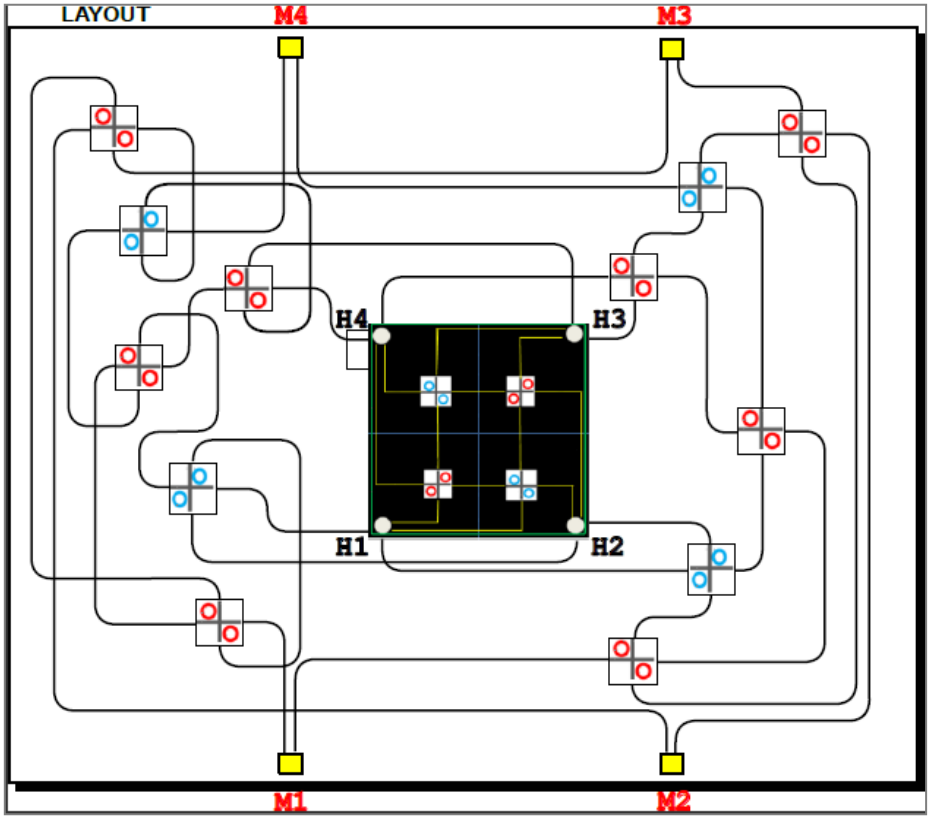


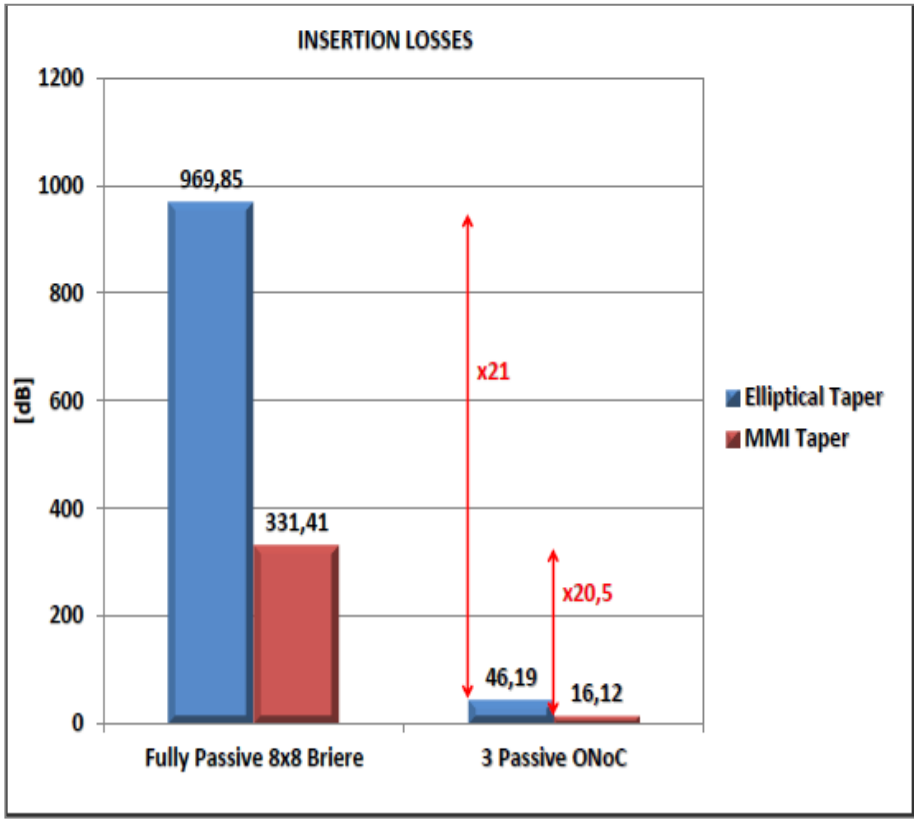
4x4-λ Router for Request as well as Response memory transactions



4x4-GWOR for Inter-Cluster Communications

Partitioned Solution





In the **8X8 PPNoC** every initiator modulates the same 8 wavelengths, thus requiring 8 different external Laser sources;
On the contrary, by adopting the **Partitioned Solution**, wavelengths can be reused across the multiple networks, **thus requiring only 4 Continuous Wave Lasers**.

Bandwidht Scalability in Passive Networks

Successive generations of our 3D- System <u>will integrate more cores</u>

A) Increasing the number of Hubs ***may not be a cost-effective choice for some time*** *in order to amortize the cost for electro-optical conversion and for the Optical NoC infrastructure support (e.g. Laser sources, distribution network of the Optical power).*

B) The same consideration ***holds for the number of Memory Controllers, which could stay the same for <u>a few device generations</u> (photonic integration may prevent DRAMS from being a performance bottleneck for some time)***

BANDWIDTH SCALABILITY TECHNIQUES are needed
TO INCREASE <u>THE PEAK OF INJECTION RATE from HUBS</u>

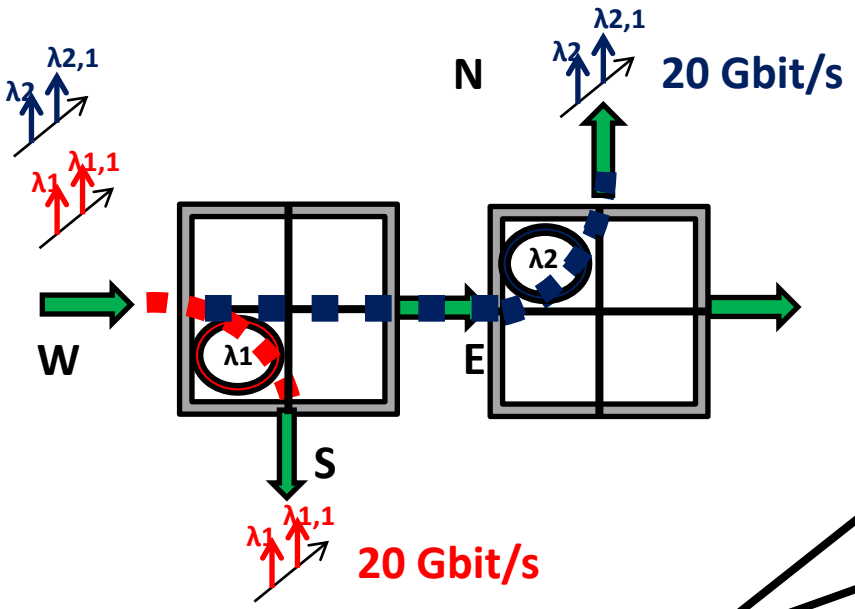
The peak Bandwidh can be Increased to accomodate the memory traffic that the hubs aggregate from a larger number of cores in the cluster
***(Assumed so far to be 40 Gbit/s for each Hub, i.e., 4 wavelenghts modulated at 10 Gbit/s*)**

BPS: Broadband Passive Switching Technique

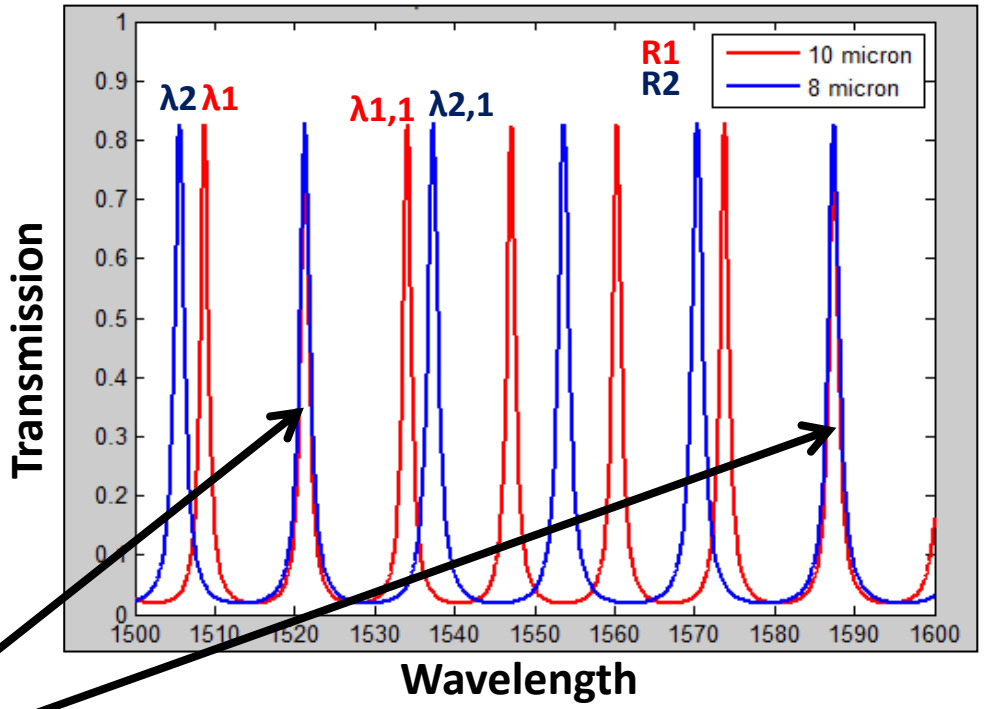
BPS consists **of embedding Multiple Virtual Networks into the same set of waveguides**, using spare wavelengths which may be available **depending on the maturity of the technology**

One possibility is to leverage as much as possible the wavelengths in the resonance band of a Micro Ring Resonator (**MRR**).

1x2PSEs cascading



Transmission Responses with different values of radius



The design of the radius should be carefully engineered

This overlapping provides Routing Fault

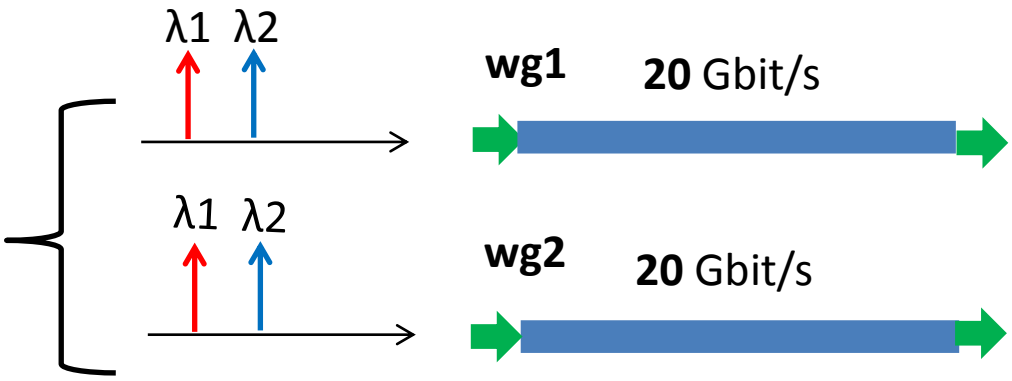
SPM: Spatial Parallelism Technique

Another way to achieve **higher network bandwidth** is simply to replicate the network.

All the replicated networks **must be laid out** in a way **to minimize waveguide crossings.**

λ1 , λ2: modulation at **10 Gbit/s**

Multiple physical networks can be used to forward more bits.



The optical power which is transmitted on a certain number of distinct wavelenghts (λ1, λ2), is physically and homogeneously coupled on different waveguides (wg1,wg2)

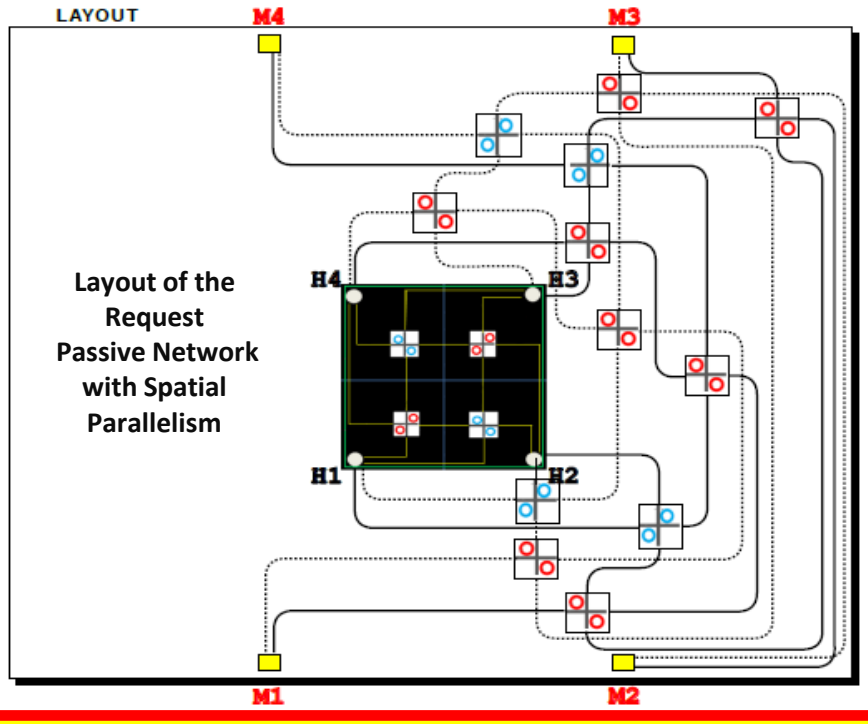
SPM uses **the same additional number of modulators and detectors** but on different waveguides.

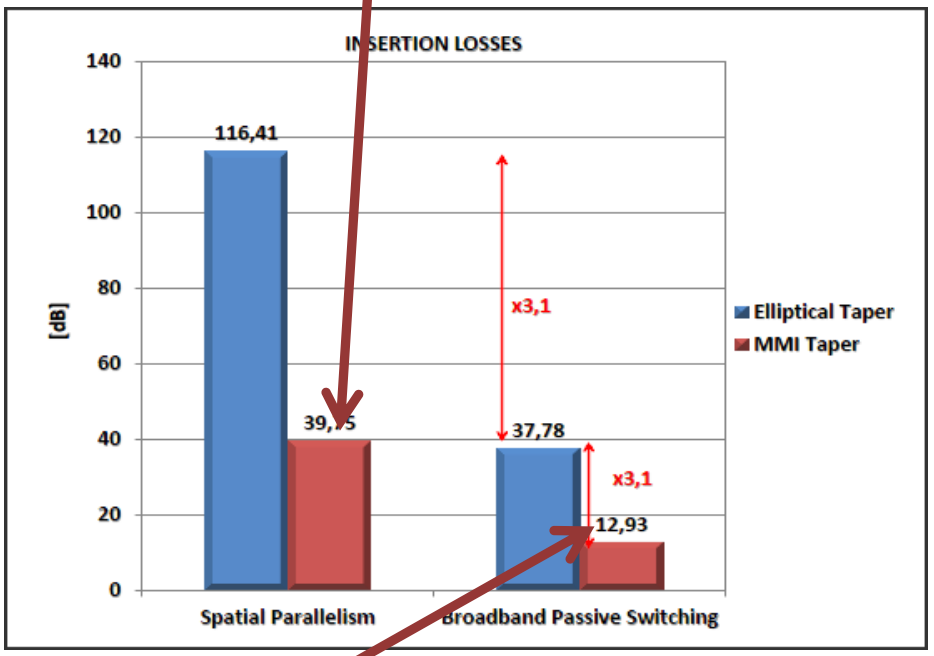
In the **BPS** and **SPM** techniques, the total power provided by the optical source sub-system should be more or less the same, **since in all cases the networks are replicated** (either **virtually** or **physically**).

Spatial Parallelism vs. Broadband Passive Switching

The insertion-loss comes **either** from new wavelengths on the same waveguides (**BPS)** or from the same wavelengths on additional waveguides (**SPM)**.

By using MMI taper optimization, **SPM** is not able to go below **39 dB** of total insertion-loss





BPS preserves the nominal insertion-loss of **around 12 dB**, whereas it grows **up to 3x** in **SPM** due to the waveguide crossings that the real layout constraints impose.

SPM has a critical path insertion-loss which is **4 times larger than BPS**

These losses are not comparable with those analyzed before ,since **the new plot refers to an injection rate** from each hub that has been doubled and now peaks **at 80 Gbit/s.**

Conclusions and Future Works

❑ In this paper we have quantified the deviation between quality metrics of logic topology as opposed to physical ones for passive optical NoCs.

❑ This discrepancy stems from the mapping of the **logic connectivity scheme** onto **the real layout** subject to **placement constraints of communication actors** and their **network interfaces.**

As a case study, we have considered a processor-memory network in a 3D-stacked multi-core processor, pointing out that:

➢ The **Insertion losses in the physical topology were one order of magnitude larger than expected** , due to the high number of waveguide crossings needed to lay it out

➢ **With respect to global connectivity**, **Optical** NoC partitioning materializes around **20x lower insertion losses** as well as an effective reuse of wavelengths and off-chip laser sources.

➢ **Real layout constraints heavily penalize SPM as bandwidth scalability technique**, since the additional waveguide crossings made **insertion-losses 3x larger than in the nominal case**. On the contrary, **BPS preserved such nominal values at the cost of more 2x optical sources**.

❑ **Future Works**: the **IL degradation associated with physical implementation of alternative topologies is being investigated**, in addition to their **node scalability properties**.

ACKNOWLEDGEMENTS

This work has been partially supported by the **PHOTONICA project** (under the "**FIRB-Futuro in Ricerca**" program, funded by the Italian Government and by **the National Science Foundation (under Award number: 5-25083**).

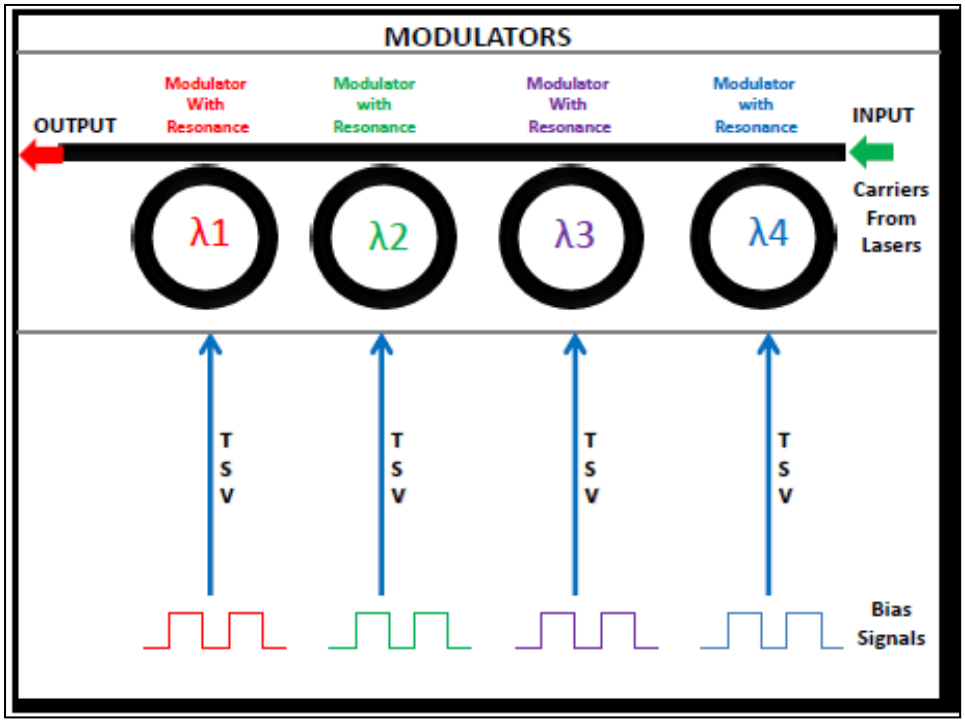
THANKS TO EVERYONE

Luca Ramini (**luca.ramini@unife.it**)

Backup

Notice that: There are no electronic devices in the optical layer **thus potentially resulting in Low-Cost Fabrication for this layer.**

Array of Modulators in the Optical Layer



1) The latest technological developments about **3D-integration enable TSVs** with a pitch of 5um x 5um and therefore a large TSV integration density **(up to 160K TSVs in a 10mmx10mm die).**

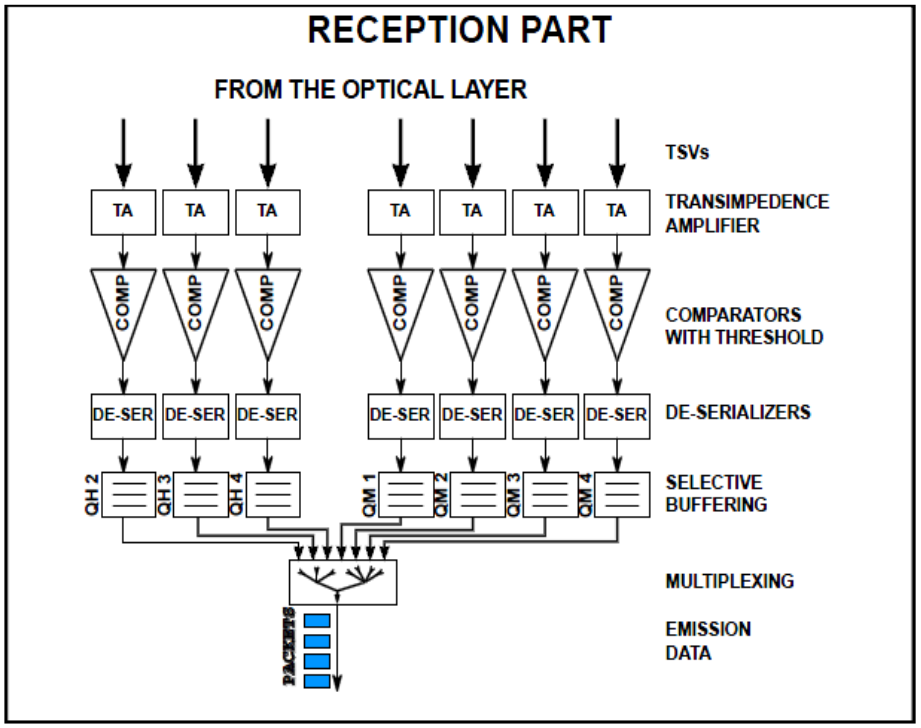
2) TSVs can deliver high-speed transmission from 1 Gbit/s to 10 Gbit/s.

The Modulation rates of each wavelength is **10 Gbit/sec**.
As a consequence, every hub offers a peak Bandwidth of **40 Gbit/sec.**

This **performance motivates us** to use **TSVs** to provide the biasing signal to the optical modulators in the optical plane

Opto-Electronic Network Interface

Electronic-Side



Digital Comparators and **De-Serializers** complete the domain conversion.

Buffers are associated with packet source and from here on the electronic network interface functions come into play.

For istance

1)**Association** of Memory Responses with memory requests.

2)**Packetization** for the E- NoC.

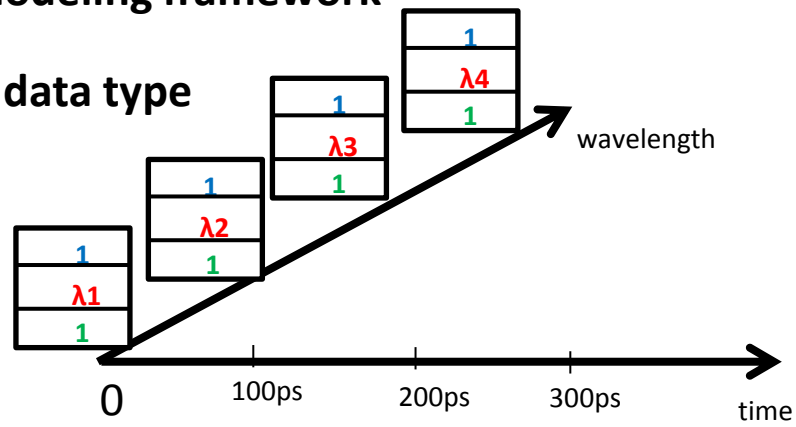
Optical link Modeling in SystemC

The Optical link model is at the core of our SystemC modeling framework

sc_signal channel is instantiated with a user definied **data type**

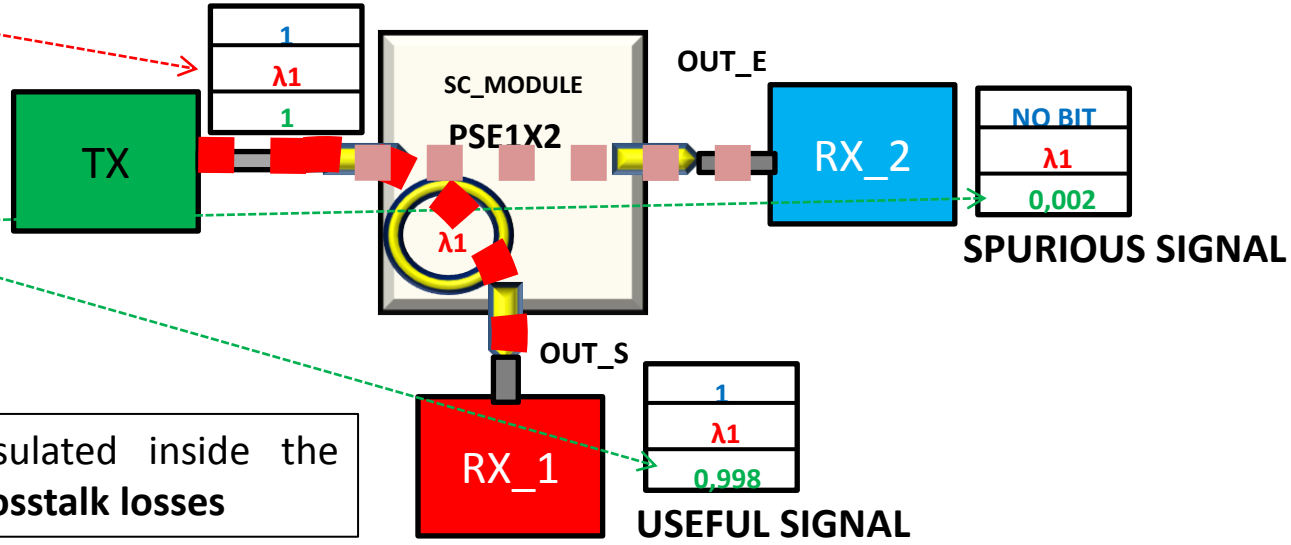
Our link can support WDM

By extending the user defined data type to represent multiple wavelengths (and associated logic values and signal amplitudes) which may be transmitted at the same time into the communication channel



The **wavelength** is used by the router for routing decisions

The **signal amplitude** **preserves technology awareness**

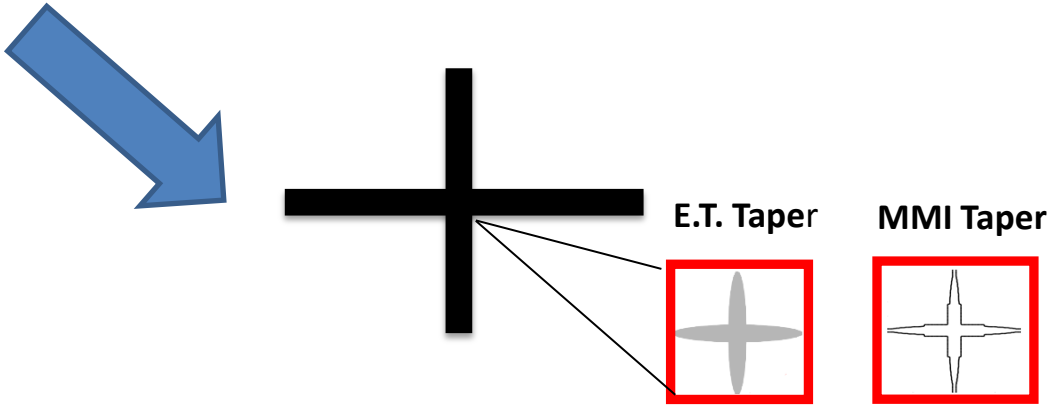


The **analytical model** encapsulated inside the router returns **Insertion and Crosstalk losses**

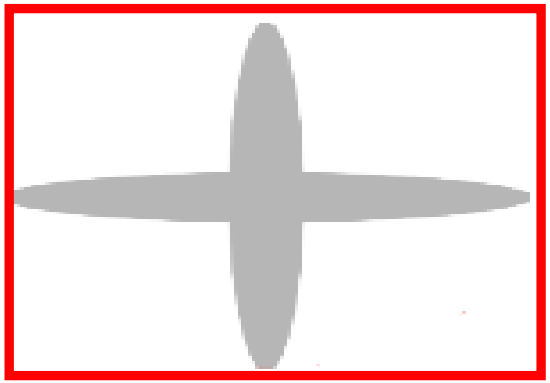
CROSSING WAVEGUIDES

Losses: 0.5 dB/single cross with **Elliptical Taper**
Losses: 0.18 dB/single cross with **MMI Taper**

Latency : 1 ps (through)



Elliptical Taper



Multi-Mode-Interference Taper