POPSTAR: a Robust Modular Optical NoC Architecture for Chiplet-based 3D Integrated Systems

Yvain Thonnart*, Stéphane Bernabé*, Jean Charbonnier*, Christian Bernard*, David Coriat*, César Fuguet*, Pierre Tissier*, Benoît Charbonnier*, Stéphane Malhouitre*, Damien Saint-Patrice*, Myriam Assous*, Aditya Narayan†, Ayse Coskun†, D. Dutoit*, P. Vivet*
*Université Grenoble Alpes, CEA-Leti, Grenoble, France; E-mail: {first.last}@cea.fr
†Boston University, Boston, MA 02215, USA; E-mail: {adityan, acoskun}@bu.edu

***Abstract*—Silicon photonics technology is now gaining maturity with increasing levels of design complexity from devices to large photonic integrated circuits. Close integration of control electronics with 3D assembly of photonics and CMOS open the way to high-performance computing architectures partitioned in chiplets connected by optical NoC on silicon photonic interposers. In this paper, we give an overview of our works on optical links and NoC for manycore systems, from low-level control of photonic devices to high-level system optimization of the optical communications. We detail the POPSTAR optical NoC topology and architecture (Processors On Photonic Silicon interposer Terascale ARchitecture) with electro-optical interface chiplets, the corresponding nested spiral topology for single-writer multiple-reader links and the associated control electronics, in charge of high-speed drivers, thermal stabilization and handling of the protocol stack, from data integrity to flow-control, routing and arbitration of the optical communications. The strengths and opportunities for this architecture will be discussed, with a shift in system & implementation constraints with respect to previous optical NoC proposals, and new challenges to be addressed.**

I. Introduction

High-performance computing and big-data applications have been craving for more integrated modular & scalable many-core systems. With increasingly complex CMOS technologies, single large-scale monolithic chips become more difficult and costly to develop, with lower fabrication yield on a large silicon area leading to increased production costs. Partitioning the design in smaller modular chiplets densely integrated with 3D-stacking on an interposer allows for re-use, increased yield on smaller area, and potentially simpler system development [1]. Nevertheless, this chiplet assembly requires efficient communication across the interposer, providing low end-to-end latency with high bandwidth density between chiplets with limited power consumption. While passive interposers have been developed and used in production since a few years, demonstrations of active CMOS interposers are just emerging [2]. As silicon photonics technology is now reaching maturity with a broad palette of optimized devices [3], and with several groups that have pushed the Network-on-Chip paradigm to new optical flavors, leading to optical NoC (ONoC) architectures and systems, we had investigated the breakeven point of passive, active and optical interposer technologies for large-scale manycore systems [4]. We will present the motivations (Section II), technology constraints (Section III) and design choices (Section IV) that led us to the definition of the POPSTAR ONoC topology and E/O chiplet for a robust modular architecture, which will be detailed in Section V.

II. Architecture Opportunities

A. Optical Networks on Chip & 3D-Systems on photonic interposers

Optical Neworks on Chip have been initially proposed a decade ago as an architecture proposal for manycore systems on chip [5][6]. Different approaches were taken to evolve from electrical NoC, with ring topologies, Clos networks and wavelength-routed optical lattices and 2D-meshes. In all cases, the major claim has always been a huge available bandwidth density thanks to the use of multiple laser wavelengths, followed by low latency and low power brought by silicon photonics by removing the wiring capacitance of electrical interconnects on long distances, i.e. 2-4cm scale for a full reticle chip. High bandwidth and low latency would indeed drastically improve data-limited high-performance applications.

While most early ONoC proposals were either purely focused on optical topologies or describing monolithic circuits integrating CMOS and photonics, an ambitious proposal was considering a tileable multi-chip ONoC system [7] assembling bridges and island chips with optical proximity coupling.

With a strong technological background in 3D integration, micro-bumping and thru-silicon vias, CEA-Leti has been advocating for dense packaging of chiplets on smart interposers integrating advanced communication architectures. In this context, it became possible to consider the design of 3D integrated systems on silicon photonic interposers. For modularity and integration of different technologies, dedicated electro-optical (E/O) chiplets can be used to drive the ONoC.

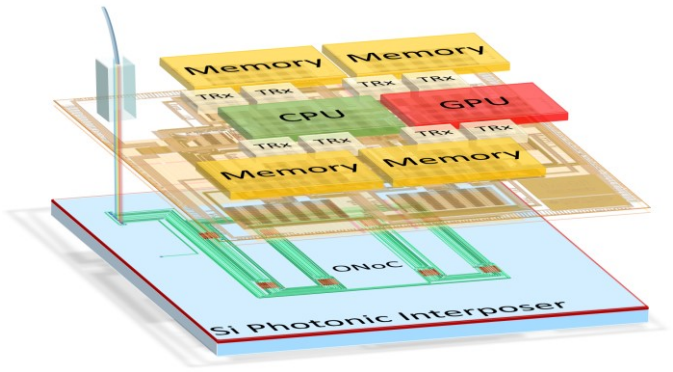

Fig. 1. Chiplet stacking on a silicon photonic interposer, showing heterogeneous compute dies, and dedicated E/O interface chiplets for optical communication in the ONoC on the photonic interposer.

B. *Microring-based Silicon photonic links*

Silicon photonic platforms such as [3] have been providing and optimizing photonic devices, light sources, waveguides, modulators, filters and detectors, allowing to build various optical subsystems. The objective of tight integration of an optical network-on-chip within a complex compute-intensive architecture imposes a low area overhead for the optical communication, at least on the electro-optical interface between CMOS and photonics. This clearly favors microring resonators (MRR) for modulators and filters [8] over longer devices such as Mach-Zehnder interferometers and electro-absorption modulators. Besides, the need for stable and energy-efficient laser sources pushes for external sources. Fig. 2 presents the basic architecture of a microring-based on-chip photonic link. An external laser source is coupled to the silicon photonics by a grating coupler to an optical waveguide. A MRR modulator is tuned to resonate at the laser wavelength, and electrically modulated at high speed to encode the data to be transmitted using on-off keying (OOK) or pulsed-amplitude modulation (PAM). The optical waveguide runs to the receiver, where a MRR is tuned to filter the modulated wavelength to a photodiode, which generates a photocurrent fed to a TIA for transmission to the receiver.

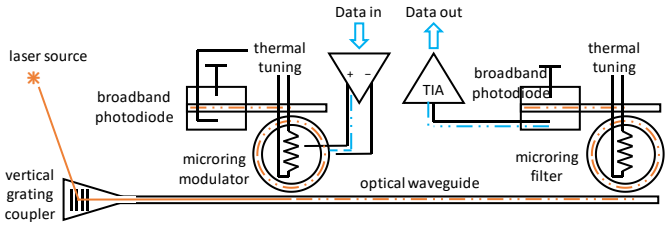

Fig. 2. Constituents of a microring-based on-chip photonic link

Thanks to high quality factors (Q>10,000) of the MRR leading to very narrow resonance bandwidths, it is possible to have several wavelengths combined in a single waveguide, allowing for wavelength-division multiplexing (WDM), as illustrated in Fig. 3. The different wavelengths can be used independently to transmit different data streams. Various ONoC architectures have proposed to add additional swichable MRR filters to route streams to different destinations.

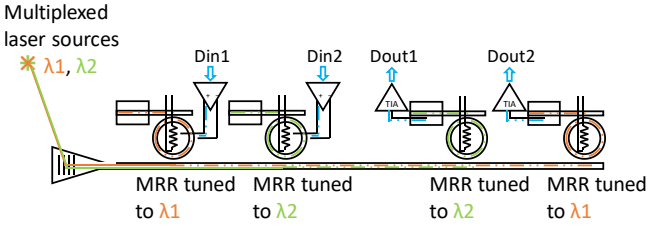

Fig. 3. WDM on-chip photonic link enabled by narrow resonance bandwidth of the MRR.

Nevertheless, the narrow resonance bandwidths require accurate control of the resonance, which cannot be fully settled at design time, and requires dynamic tuning based on monitoring of the optical signal on the drop-port of the MRR. Closed-loop locking on a wavelength is presented in [9].

Constraints and feasibility of large-scale WDM integration in an ONoC in view of this will be presented in the next section.

III. Technology Constraints

Microring resonance has a huge dependence on silicon waveguide manufacturing process, especially Si thickness, with resonance variations in the order of 1nm of wavelength shift per nm of average thickness over the MRR [7]. We characterized these variations at die and wafer level to identify the amount of drift to be expected for WDM links for neighboring rings (part of the same Tx or Rx) and for distant rings.

Fig. 4 shows wafer-level variability measurements for 8.7µm radius MRRs around 1310nm. The free-spectral range (FSR) in this band is around 7.2nm between successive resonance orders. The closest resonance near 1310nm has been selected in all cases. This show an average mismatch at 5cm distance well below half an FSR (which would be observed in the random case), which demonstrates that the same resonance order is indeed selected on all dies. Hence, the measurements show a worst-case geometrical variability of the resonant wavelength around 75pm/mm.

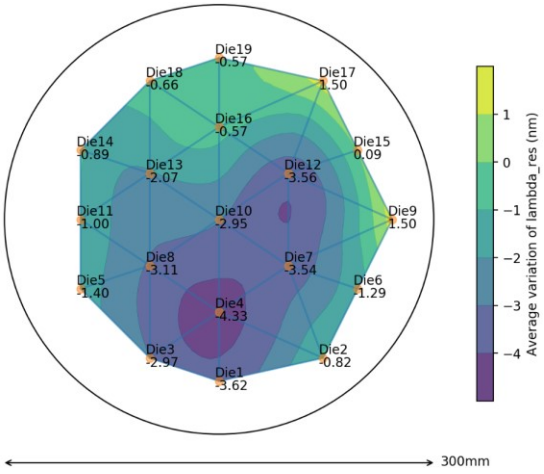

Fig. 4. Wafer-level variability measurements showing worst-case geometrical variability of MRR resonant wavelength below 75pm/mm

Figure 5 shows the resonant wavelength statistics for adjacent rings with 80µm pitch, which can be approximated by a Gaussian distribution. The resonance difference is extracted, with a standard deviation of 86pm. As the std. dev of a sum of Gaussian IID random variables is the square root of the sum of variances, dividing by $\sqrt{2}$ gives a local random variability of the resonance wavelength following a standard deviation of 60pm.

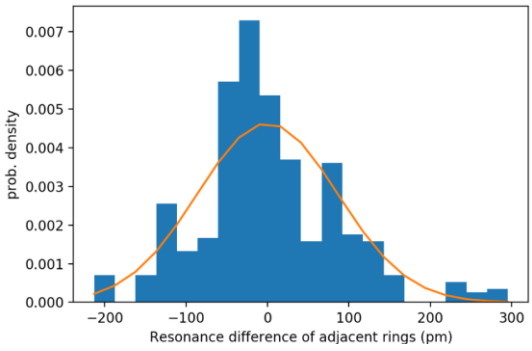

Fig. 5. Local random variability measurements of MRR resonant wavelength approximated by a gaussian distribution with 60pm standard deviation.

Based on these local and global variability measurements, we can consider the consequences at large-scale on the WDM photonic links.

Regarding the number of wavelengths, Q-factors above 13,000 at 1310nm give a 3dB bandwidth below 100pm. For rings in the 10µm radius range, this is about 1% of the FSR. With 7× margins to limit crosstalk to 0.1dB, this would allow for about 12 to 16 wavelengths evenly spaced in the FSR. Nevertheless, using an 80µm pitch for the rings, previous variability results show about 250pm variability at 3σ and over 1mm between rings, i.e. either crosstalk up to 0.4dB between 16 wavelengths or crosstalk below 0.1dB for up to 10 wavelengths.

This computation shows that it is possible to consider MRR groups working as a whole for WDM transmission within a 1mm×1mm region. However, geometric variability becomes dominant in the cm range with resonance shifts above 750pm, and no assumption should be done on the correspondence of resonances of different MRR groups. Furthermore, MRRs are very sensitive to thermal variations: 78pm/K of resonance shift was measured in [9]. As compute load can lead to temperatures differences in the 10K range, different MRR groups can incur both thermal shifts and process shifts in the 1nm range. Locally, however, the temperature hotspots created in the chiplets and observed at interposer level are below 1K/mm. Heat is indeed relatively well spread by the silicon bulk and metallization of the chiplets and interposer, as shown in [10].

In other words, it is possible to design a WDM Tx or Rx site with increasing MRR perimeters, but distant Tx and Rx sites may have very different resonances for identical MRR perimeters. The closest resonances can be observed for rings of different diameters and potentially with the previous or next order of resonance thanks to folding in the FSR.

To lock the MRR resonance to the closest laser wavelength, a shift of up to FSR/n may be needed for n wavelengths. This is done by leveraging the thermal sensitivity of the MRR with integrated resistive heaters, which are controlled to by a closed loop feedback between the optical power and the injected current for Joule heating [9]. The optical response depends on local ring temperature and laser wavelengths, as illustrated in Fig. 6. A discrete set of ring temperatures matches the setpoint, dynamically mapping the ring to a given WDM wavelength.

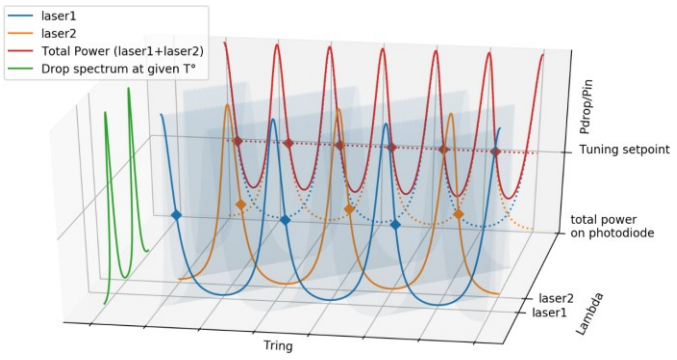

Fig. 6. Microring drop port transmission as a function of input wavelength and ring temperature shaped as a wave surface, showing resonances (here with low Q-factor) and free-spectral range periodicity (at a given temperature in green), with WDM multiplexing dividing the effective period measured on photodiode by the number of wavelengths (photocurrent depending on temperature in red).

The feedback loop allows to continuously maintain the ring at resonance within than 1pm of accuracy [9], preventing any thermally induced degradation of the SNR or data crosstalk between wavelengths. Regarding thermal coupling between rings, most of the heat flux coming from the heaters is sunk by the silicon substrate, and less than 0.02K/mW coupling is observed on adjacent rings, even when using a backside cavity opening below the ring to increase the local heating efficiency of the thermal tuning, as shown in Fig. 7. This shows the feasibility of individual operation of MRRs in an MRR group.

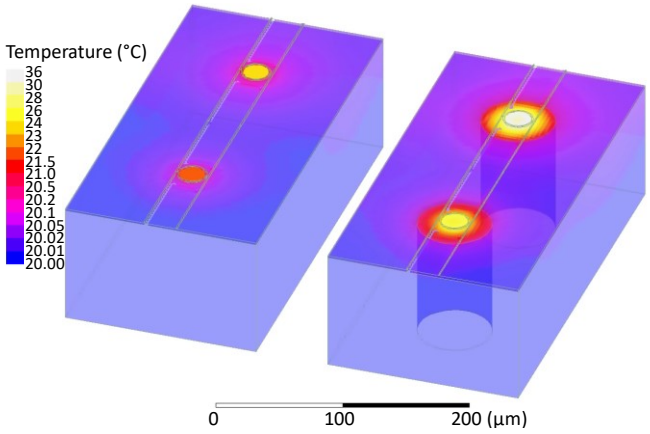

Fig. 7. Thermal simulation of individual ring tuning & coupling coefficients without and with 30µm radius back-side opening all the way to SOI buried oxide (simplified layout). Simulation shows heating efficiency around 250µW/K without opening and 60µW/K with opening, for mutual coupling below 0.02K/mW at 80µm distance in both cases.

Complete operation of a WDM MRR group is described in Fig. 8. Depending on application load, the chiplet temperature profile evolves and triggers remapping of the MRR resonances to the best laser wavelengths by discrete ring temperature jumps. Using WDM link as a single parallel data bus reduces remapping overhead and relieves constraints on global communication sequencing, with a single other endpoint affected by remapping.

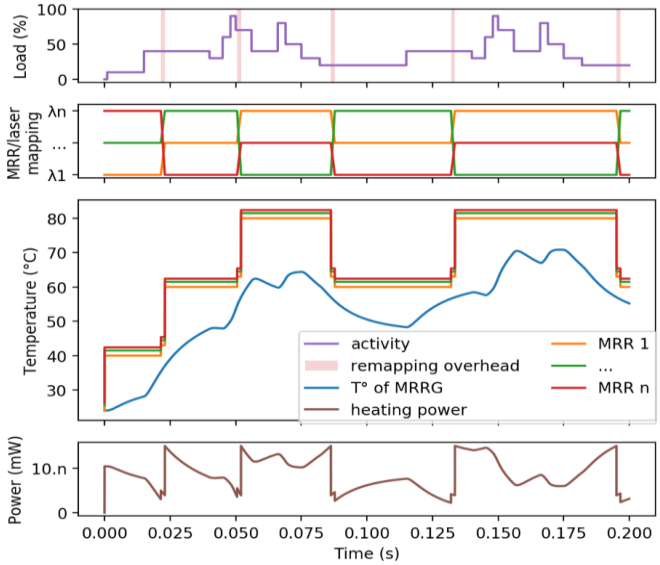

Fig. 8. Joint temperature remapping of all wavelengths of a WDM link depending on neighboring activity. A shared lock signal indicates link availability to the application level [9].

IV. WDM Extension beyond point-to-point link

WDM modulation and routing have different specifications for switching speed. High modulation rates are needed to amortize static laser and thermal power costs, but dense integration in ONoC prevents from targeting complex power-hungry pre-distortion and equalization schemes.

MRR modulation in the 10-20Gbit/s range can be achieved using depletion-mode PN-type MRR, where a diode junction in the optical path allows for fast voltage modulation by carrier depletion. However, voltage modulation efficiency in pm/V and Q-factor are correlated via dopant density, and resonance shifts are typically in the order of 20% of FWHM for CMOS voltages. Fig. 9 shows thru and drop transmissions for a PN ring near critical coupling. With limited shifts, a tradeoff exists between insertion losses (IL) and extinction ration (ER). For Q=20,000, 6dB of ER is achievable under 2.4V at the cost of 3dB of IL. As thermal tuning has >100µs lock time, the MRR needs to be constantly ready for transmission and locked to the drop port level corresponding to the 3dB IL point on the thru port.

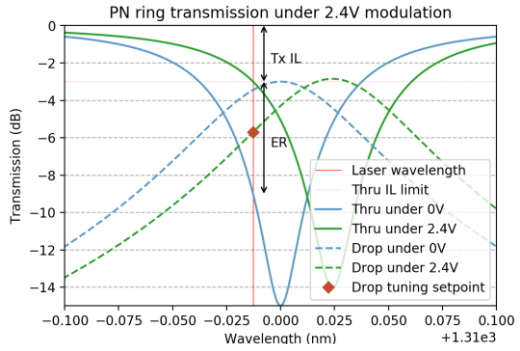

Fig. 9. Transmissions on thru and drop ports for a PN ring.

Conversely, filtering for routing is switched at packet rate in the 100MHz-1GHz range, and can be achieved using injection-mode PIN-type MRR, where carriers are injected in the intrinsic Si waveguide of the ring by forward biasing a PIN junction. This non-linear effect allows for resonance shifts above the FWHM, and much lower IL ot the thru port than PN-type MRR. The design tradeoff for PIN-type MRR is rather between thru IL out of resonance, and drop IL at resonance, which defines the losses of the deflected wavelength after a "turn" in the ONoC, in view of sufficient tuning power on the drop port. Fig. 10 shows the transmissions near resonance for a PIN ring designed for 0.7dB thru IL and 2dB drop IL for -10dB tuning setpoint.

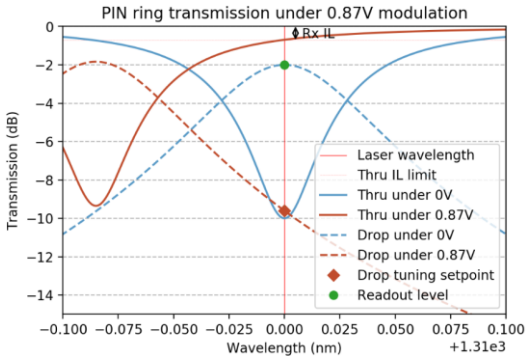

Fig. 10. Transmissions on thru and drop ports for a PIN ring.

Given these specificities of PN and PIN MRR and constraints on permanent tuning close to resonance for low-latency path setup, cascading several PN rings tuned to the same wavelength for multiple-writer topologies with several Tx sites sharing a waveguide is not scalable in terms of optical power budget (accumulation of 3dB PN thru IL). Neither is considering many jumps from thru port to drop port of several PIN rings for routing (accumulation of 2dB PIN drop IL). This is why we favor Single-Writer Multiple-Reader (SWMR) types of topologies, where light is kept within a single waveguide from Tx to any Rx. When scaling up the number of Rxes, only PIN thru IL and waveguide losses accumulate, as seen in Fig. 11. An input power level of 0dBm (1mW) has been chosen here after the grating coupler, with a 3dB margin wrt. maximum acceptable power level in MRRs. Continuous lines indicate the optical power level in the waveguide, while markers indicate the optical power levels measured on the photodiode after the drop port of each MRR. Robust tuning of the MRR is possible as long as the diamond markers stay above -18dBm sensitivity. Data readout is done on drop levels for "0" and "1" (round markers) with an average value above -15dBm. A summary of consistent device parameters from our 200mm silicon photonics platform is given in Table I.

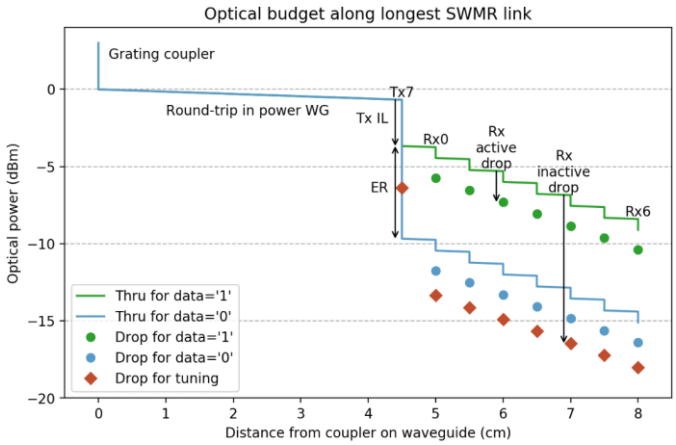

Fig. 11. Cascading Rxes on a SWMR link is better than MWSR/MWMR for optical budget based on PN rings for Tx and PIN rings for Rx. Overall scalability is determined by receiver sensitivity for low-frequency tuning on last Rx of the longest link.

TABLE I. Optical Device Parameters (w. data from [3][11])

Device	Parameter	Value
PN MRR modulator	Thru IL off-res. ("1")	-3dB
	Thru IL on-res ("0")	-9dB (ER 6dB)
	Drop IL tuning	-6dB
PIN MRR filter	Thru IL off-res. (deselected)	-0.7dB
	Drop IL on-res (selected)	-2dB
	Drop IL tuning	-10dB
Waveguide	Straight losses	-0.11dB/cm
	Critical radius (lossless)	20µm
	Crossings (1x1 MMI)	-0.25dB
Grating coupler	IL	-3dB
Laser power	Max power in MRR	3dBm
O/E sensitivity	Demod. sensitivity (10Gbps)	-15dBm
	Tuning sensitivity	-18dBm

V. POPSTAR ARCHITECTURE

The POPSTAR (Processors On Photonic Silicon interposer Terascale ARchitecture) topology combines SWMR links dedicated to each Tx into a ring-shaped bundle of nested spirals, as shown in Fig. 12. It is designed to present a replicable interface for 3D-stacking of identical electro-optical chiplets on a silicon photonic interposer made of 1 Tx row and N-1 Rx rows for N chiplets. One of the spirals has been highlighted in dashed blue for illustration: it corresponds to the longest optical path analyzed in the previous section. Optical power is injected from external sources on the exterior of the spirals, and brought to the Tx E/O chiplet by an optical power delivery waveguide. From the Tx, the modulated waveguide passes across Rxes of all other E/O chiplets, with an inwards shift on the spirals to the next Rx row at every chiplet. This avoids any waveguide crossings between SWMR channels.

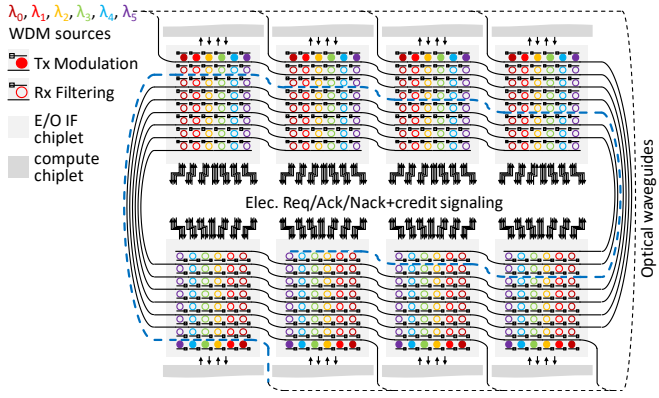

Fig. 12. POPSTAR uses a bundle of nested spiral waveguides, which allows identical chiplets to drive SWMR links connected to all the other chiplets.

In terms of complexity, the POPSTAR topology contains the same number of MRR as a complete non-blocking crossbar. It is actually a fully distributed optical crossbar controlled at Rx sites of E/O chiplets. Because of the SWMR choice, arbitration is fully contained in the Rx stages of E/O chiplets, and no centralized arbiter is required.

Signaling between E/O chiplets is done with electrical transmission lines to avoid any information loss in the control flow (e.g. due to a pending wavelength remapping). Fig. 13 presents the E/O communication protocol between chiplets.

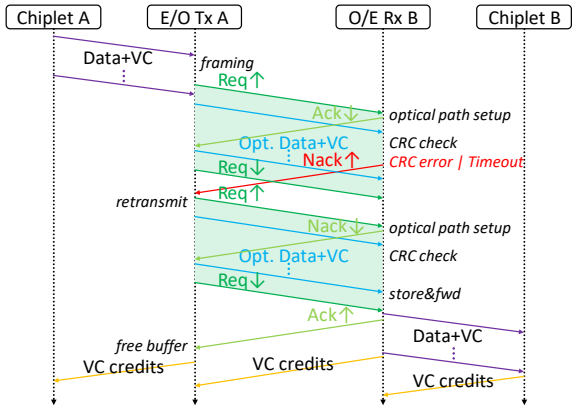

Fig. 13. Req/Ack/Nack+Credit protocol between E/O chiplets

An electrical "Request" (Req) signal is used to set-up the optical path between Tx and Rx by biasing the PIN MRR to lock on the drop port resonance peak for all the rings in the Rx WDM row. Backwards "Acknowledge/No-acknowledge" (Ack/Nack) signals notify the transmission status to the Tx for completion or retransmission, in a 4-phase protocol with "Req". Nevertheless, data transmission starts immediately after rising "Req", without waiting for Ack/Nack falling edge. Finally, depending on the target compute platform, credits for the different virtual channels (VC) can be transmitted electrically or piggy-backed in the backwards optical flow.

Throughput on these electrical lines is low wrt. optical data communication since at packet level, and simple low-power drivers and metal traces on interposer can be used. The main constraint is that "Req" is sent early enough to set-up the Rx MRR filters before optical data arrives. Electrical RF simulation showed that a latency of 2ns/cm is easily achieved in 2µm wide copper lines.

The architecture of the E/O chiplet is presented in Fig. 14. Tx and Rx operate independently from each other. Besides, Rx rows receiving data from other E/O chiplets are also independent from optical demodulation to storage buffers per direction and virtual channel. Indeed, for arbitration to be contained in Rx E/O chiplet, writing in Rx data buffers must be non-blocking between Txes.

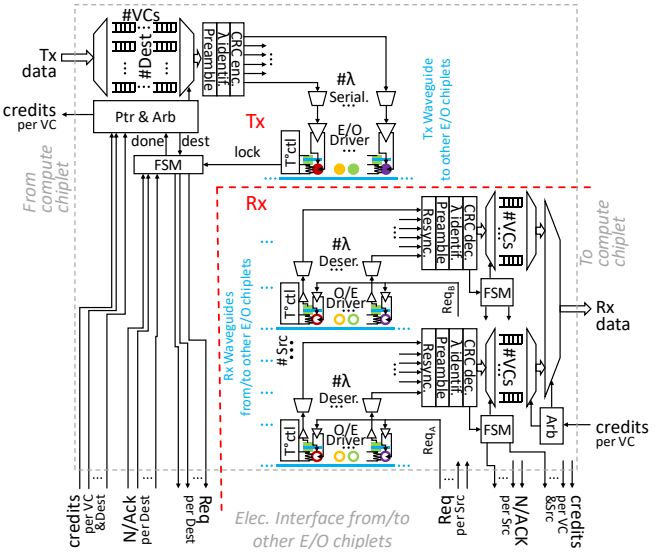

Fig. 14. E/O chiplet architecture with local interface to compute chiplet and interfaces to other E/O chiplets using optical waveguides for data, and electrical control lines for path setup and flow control.

Operation of the E/O chiplet is as follows: On the Tx side, data from the compute chiplet is demultiplexed into FIFO buffers dedicated to each virtual channel and destination. As soon as enough data is stored, a frame is formed, which can contain a variable number of flits. An arbiter between destinations and VCs selects the next frame to transfer provided enough credits have been received. "Req" is raised for this destination. A preamble is inserted at the beginning of the frame, along with an encoding of the logical wavelengths and a cyclic redundancy check per parallel word of the frame. Words are serialized and modulated and finally "Req" is deasserted.

On the Rx side, at the row selected by "Req", MRRs are activated to route data to the photodetectors, while "Ack" and "Nack" are lowered, and data is demodulated and deserialized, then synchronized to the Rx clock domain. The preamble is detected to identify the permutation of bits to apply on the deserialized data, and logical wavelength identification is done to recover the order of sub-words to reconstruct the parallel word. CRC check is computed, and "Ack" or "Nack" is raised, while data is stored in buffers per VC and source. An arbiter (round robin per source, priority per VC) selects the data to be sent first to the compute chiplet. Once the buffer is read, credits are sent to the Tx chiplet.

End-to-end latency for E/O/E communication from compute to compute chiplet is variable due to store & forward of the data frame until correct transmission is confirmed by CRC, and potential contention in the final arbiter. Minimum 0-load latency is 12 clock cycles for any pair of chiplets on the interposer.

VI. Discussion

The POPSTAR architecture has been designed to be robust and modular based on technological constraints, in a pragmatic approach toward implementation, leveraging various concepts introduced in previous works. Measured MRR models favor single waveguide paths to multi-hop architectures such as [5][11][13][20][23][26]. The serpentine ring crossbar of [6] presented a similar decentralized approach. SWMR, first introduced in [14], is used here as a single parallel WDM bus, and offers lower losses than MWxR [16][17][22]. Non-blocking operation was investigated in [15][18]. For robustness, we favored electrical lines for signaling and arbitration rather than optical arbiters [6] [19]. Thanks to [9], chip-level temperature homogenization has become unnecessary. While we have been investigating the idea of chiplets on optical interposers since [3], the disaggregation of an ONoC to different optical chiplets has been described in [21].

Our design approach was constructed from the bottom-up. System-level investigation of the architecture on different data-intensive applications has started with [10], showing potential for dynamic adaptation of the number of active wavelengths. More studies like [24][25] combined with an architectural reflection on processors and memory hierarchies would clearly open new perspectives for optimization and system integration.

VII. Conclusion

In this paper, we have presented the process, thermal, device and packaging constraints, which have led us to the definition of the POPSTAR architecture for electro-optical communication between chiplets stacked on a photonic interposer. A standard replicable E/O interface chiplet is in charge of buffering, routing, arbitration, serialization, driving and thermal tuning of MRR modulators and filters. The use of independent SWMR links each as a single parallel WDM bus allows for robust operation with coordinated tuning and switching of all MRR of a Tx or Rx WDM channel. This topology implements a low-latency distributed non-blocking crossbar, and fully contains arbitration on the Rx side for access to the compute chiplet. It provides low-latency communication to large-scale chiplet-based 3D integrated systems, and offers new architecture opportunities for data-intensive high-performance applications.

References

[1] J. Yin *et al.*, "Modular Routing Design for Chiplet-Based Systems," in *ISCA*, 2018.

[2] P. Vivet *et al.*, "A 220 Gops 96-core Processor with 6 chiplets 3D-stacked on an active interposer offering 0.6ns/mm latency 3 Tbit/s/mm2 inter-chiplet interconnects and 156mW/mm2@82% DC-DC converters", in *ISSCC*, 2020.

[3] B. Szelag *et al.*, "Multiple wavelength silicon photonic 200 mm R+D platform for 25Gb/s and above applications," in *Silicon Photonics and Photonic Integrated Circuits V*, vol. 9891, 2016.

[4] Y. Thonnart and M. Zid, "Technology assessment of silicon interposers for manycore SoCs: Active, passive, or optical?," in *NOCS*, 2014.

[5] M. Briere *et al.*, "System Level Assessment of an Optical NoC in an MPSoC Platform," in *DATE*, 2007.

[6] D. Vantrease *et al.*, "Corona: System Implications of Emerging Nanophotonic Technology," in *ISCA*, 2008.

[7] A. V. Krishnamoorthy et al., "Computer Systems Based on Silicon Photonic Interconnects," *Proceedings of the IEEE*, vol. 97, no. 7, 2009.

[8] W. Bogaerts et al., "Silicon microring resonators," *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 47–73, 2012.

[9] Y. Thonnart *et al.*, "A 10Gb/s Si-photonic transceiver with 150µW 120µs-lock-time digitally supervised analog microring wavelength stabilization for 1Tb/s/mm2 die-to-die optical networks", in *ISSCC*, 2018.

[10] A. Narayan, Y. Thonnart, P. Vivet, C. F. Tortolero, and A. K. Coskun, "WAVES: Wavelength Selection for Power-Efficient 2.5D-Integrated Photonic NoCs," in *DATE*, 2019.

[11] Q. Wilmart *et al.*, "Ultra low-loss silicon waveguides for 200 mm photonics platform", in *Group IV Photonics*, 2019.

[12] A. Shacham, K. Bergman, and L. P. Carloni, "Photonic Networks-on-Chip for Future Generations of Chip Multiprocessors," *IEEE Transactions on Computers*, vol. 57, no. 9. 2008.

[13] M. J. Cianchetti, J. C. Kerekes, and D. H. Albonesi, "Phastlane: A Rapid Transit Optical Routing Network," in *ISCA*, 2009.

[14] Y. Pan *et al.*, "Firefly: Illuminating Future Network-on-chip with Nanophotonics," in *ISCA*, 2009.

[15] A. Joshi *et al.*, "Silicon-photonic clos networks for global on-chip communication," in *NOCS*, 2009.

[16] Y. Pan, J. Kim, and G. Memik, "FlexiShare: Channel sharing for an energy-efficient nanophotonic crossbar," in *HPCA*, 2010.

[17] S. Le Beux, J. Trajkovic, I. O'Connor, G. Nicolescu, G. Bois, and P. Paulin, "Optical Ring Network-on-Chip (ORNoC): Architecture and design methodology," in *DATE*, 2011.

[18] Y.-H. Kao and H. J. Chao, "BLOCON: A Bufferless Photonic Clos network-on-chip architecture," in *NOCS*, 2011.

[19] Y. Pan, J. Kim, and G. Memik, "FeatherWeight: Low-cost optical arbitration with QoS support," in *MICRO*, 2011.

[20] Cheng Li, M. Browning, P. V. Gratz, and S. Palermo, "LumiNOC: A power-efficient, high-performance, photonic network-on-chip for future parallel architectures," in *PACT*, 2012.

[21] Y. Demir, Y. Pan, S. Song, N. Hardavellas, J. Kim, and G. Memik, "Galaxy: A High-performance Energy-efficient Multi-chip Architecture Using Photonic Interconnects," in *SC*, 2014.

[22] S. Bahirat and S. Pasricha, "METEOR: Hybrid Photonic Ring-mesh Network-on-chip for Multicore Architectures," *ACM Trans. Embed. Comput. Syst.*, vol. 13, no. 3s, p. 116:1–116:33, Mar. 2014.

[23] P. K. Hamedani, N. E. Jerger, and S. Hessabi, "QuT: A low-power optical Network-on-Chip," in *NOCS*, 2014.

[24] J. L. Abellán, C. Chen, and A. Joshi, "Electro-Photonic NoC Designs for Kilocore Systems," *J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 2, p. 24:1–24:25, Nov. 2016.

[25] J. Bashir and S. R. Sarangi, "NUPLet: A Photonic Based Multi-Chip NUCA Architecture," in *ICCD*, 2017.

[26] Z. Wang *et al.*, "CAMON: Low-Cost Silicon Photonic Chiplet for Manycore Processors," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2019*.