Modeling a Photonic Network for Exascale Computing

José Duro, Salvador Petit, Julio Sahuquillo and María E. Gómez
Departamento de Ingeniería de Sistemas y Computadores
Universitat Politècnica de València, Spain
Email: jodugo1@gap.upv.es

***Abstract*—Photonics technology has become a promising and viable alternative for both on-chip and off-chip computer networks of future Exascale systems. Nevertheless, this technology is not mature enough yet in this context, so research efforts focusing on photonic networks are still required to achieve realistic suitable network implementations. In this context, system-level photonic network simulators can help to guide designers to assess the multiple design choices.**

Most current research is done on electrical network simulators, whose components work widely different from photonics components. Moreover, photonics technology adds new components that are not present in electrical networks. This paper discusses how a photonics simulation tool can be built by extending an electrical simulation framework. We summarize and compare the working behavior of both technologies –electrical and photonics–, and discuss the rationale behind the proposed extensions. Among others, the devised extensions model optical routers, wavelength-division multiplexing, circuit switching, and specific routing algorithms.

This work is aimed to provide support to investigate off-chip optical networks in the context of the European Exascale System Interconnect and Storage project (ExaNeSt) project. The experiments presented in this paper study multiple realistic photonic networks configurations and have been performed with excerpts of real traces. Experimental results show that, compared to electrical networks, optical networks can reduce the execution time of the workload by several orders of magnitude. Our study reveals that future optical technologies presenting a 3.2 Tbps aggregate link bandwidth will not provide additional performance benefits over state-of-the-art 1.6 Tbps optical links across the studied workloads, but 1.6 Tbps network links are enough to achieve the highest optical performance on computer networks. Regarding the link configuration, the bandwidth per optical channel is the parameter with highest impact on the network delay and so on the execution time, while for a given optical bandwidth per channel the better strategy is to reduce the phit size.

I. Introduction

The most powerful supercomputers in the world [1] are ranked by their computational power in terms of floating-point operations executed per second (FLOPS). The Sunway TaihuLight, the recent supercomputer leading the list at November 2016, realizes by 93 PetaFlops (10^{15}) with 10,5 million cores. The Top500 list tracks the computational power of supercomputers since 1971, and according to the current growing computational trend, it is expected that supercomputers will break the ExaFlop (10^{18}) barrier by 2020. Reaching this target, however, will be challenging and requires from multiple simultaneous solutions addressing, among others, computation at chip level (nodes of the system), data movement across the system, distributed storage, energy management, etc.

From the aforementioned challenges, the data movement challenge is probably the most critical to be achieved, mainly due to the increasing number of computing nodes and, therefore, the increasing communication requirements. *Exascale* networks will count with thousands of computing nodes, so data transmission among them becomes a major design concern, and new requirements rise not only in terms of throughput but also in energy consumption demands. In such systems, the underlying network technology [2], [3] is a critical design choice.

In this regard, photonics interconnects –both on-chip and off-chip– have emerged as a worth technology alternative addressing the key constraints of traditional electrical networks. This technology provides much more bandwidth than electrical technology with much less energy consumption [4], [5]. Optical Networks on-Chip (ONoCs) [6], [7] will become a viable option for the growing demand of high performance computing (HPC) applications that electric networks cannot efficiently deal with. On the other hand, off-chip photonics technology can provide what is required to cover the rising requirements in Exascale computing, contributing additional advantages over electrical technology such as the volume of the interconnection links (inter- and intra-cabinet) or the power savings. Depending on the technology node (from 90 nm to 22 nm), photonics technology is from 7 to 4 orders of magnitude smaller than electrical technology [8].

The Exascale challenge has become a major concern for supercomputers industry and countries around the world. In this sense, some research projects have spread in Europe to address the involved challenges. This work has been developed under the umbrella of the European Exascale System Interconnect and Storage project (ExaNeSt) [9] whose main aim is to build a system able to scale up to tens of millions of interconnected low-power consumption ARM cores to solve large-scale scientific and big data problems [10].

The development of such a system requires from multiple performance evaluation studies in order to guide the system construction. Regarding interconnects, simulation frameworks are being used to assess network bandwidth and network latency. Original ExaNest simulation frameworks model electrical interconnection networks and they are not prepared to simulate networks with photonic technology.

Modeling photonics networks on an electrical simulation framework is not a straightforward process but it requires from a sound knowledge on the basics of both photonics and electrical networks. Many aspects are widely different since photonics offers new possibilities and prevents from some others. For instance, photonics provides the capability of sending multiple messages concurrently on a given link, while it does not support flit storage at the intermediate network routers.

This work discusses the critical steps that have been carried out to model photonics interconnects by extending the ExaNeSt INSEE [11] simulator. Distinct design choices are discussed and evaluated. The original INSEE code has been widely modified in order to i) model the main photonics network components such as the network links or the routers; ii) model the photonic communications particularities since circuit switching and wavelength-division multiplexing are used; iii) model the routing; and iv) adapt network topologies to properly mimic the photonics working behavior.

The remainder of this paper is organized as follows. Section II presents some photonics background. Section III describes the baseline simulation framework. Section IV introduces the key differences between electric and photonics networks, and discusses the proposed models. Section V evaluates the proposal. Section VI summarizes the related work. Finally, Section VII presents some concluding remarks.

II. Background on Photonics Technology

As mentioned above, the requirements for Exascale computation over the current decade are expected to scale the network performance. In this section, we summarize recent advances in silicon photonics technology and its current state on both off-chip and on-chip interconnects.

A. Off-Chip Silicon Photonics

Silicon photonics-based interconnects are being widely deployed in data communication (datacom) systems due to their potential to achieve large scale and low cost integration together with low power operation. This potential relies on advantages like higher bandwidth capability, better distance/speed tradeoff and easier cable management. Regarding bandwidth, achieving more than 10 Gbps with conventional copper wires remains a challenge, while a single optic fiber can offer bandwidths in the Terahertz range. With respect to the distance/speed tradeoff, optic fibers are able to transmit data along several kilometers without bandwidth penalties. Finally, due to their inherent lightness and thinness, using optic fiber cables instead of copper ones, highly reduces cable density so considerably eases cable management.

Current state-of-the-art photonics technologies, however, require from pluggable transceivers to transform electrical signals to optical signals and vice versa, which limits the potential of a full silicon photonics system. To deal with this shortcoming, current research aims to integrate optical devices with logic chips. This goal has not been reached yet and it would make a significant impact on datacom network topology.

The bandwidth limit of current Quad Small Form-factor Pluggable (QSFP) transceivers based on Vertical-Cavity Surface-Emitting Laser (VCSEL) technology is by 40 Gbps. Nevertheless, silicon photonic interconnects are expected to reach the 100 Gbps mark and beyond in the near future. For instance, Intel Corporation and other big companies such as IBM or Cisco Systems have moved their silicon photonics efforts beyond research and development, and have produced engineering samples that run at speeds of up to 100 Gbps. Moreover, Intel and Corning are currently developing the MXC connector, which supports up to 64 fibers communicating at 25 Gbps, reaching an unprecedented data transmission capacity by 1.6 Tbps over a 300 meters distance.

B. On-Chip Silicon Photonics

The need of low latency and high bandwidth multi-core data transmissions has led CMOS-compatible photonic interconnects as an alternative technology to address these design issues in on-chip networks. Moreover, silicon photonics-based on-chip networks enable the implementation of silicon photonic routers, which are a key development for inter-rack and intra-rack *full*-optical networks based only on optical components –i.e. all-optical networks– for Exascale systems. In this regard, current efforts have concentrated on the realization of reliable hybrid silicon lasers, electro-optic modulators, ring resonators and receivers; the most critical building components of photonic circuits.

Laser sources inject light into the chip's waveguides. Laser sources are probably the most difficult devices to be integrated on silicon. *Duan et al.* have developed hybrid silicon/III-V lasers with less power consumption than previous works [12], [13], although not yet achieving ultra-low power consumption, which will significantly reduce packaging costs.

Electro-optical modulators establish the switching capacity, that is, the operation bandwidth of any photonic integrated circuit (PIC). High bandwidth modulation can be realized in silicon with free-carrier induced index change [14], using biased pn structures (carrier depletion) achieving up to 30-50 Gbps data rates [15], [16].

Optical ring resonators are the key component to leverage wavelength division multiplexing (WDM) [17] technology. WDM allows splitting up the optical signal into multiple independent wavelengths. A ring resonator captures specific optical wavelenghts; thus, it can redirect these wavelengths to other waveguides and receivers, so enabling the implementation of complex optical on-chip networks and photonic routers.

Finally, optical coherent receivers (also known as photodetectors), which convert the amplitude, phase, and polarization of an optical signal into the electrical domain have already been integrated, and provide very high data conversion rates (up to 224 Gb/s with PDM-16-QAM signals) [18], [19].

III. The INSEE Simulation Framework

This section summarizes the main characteristics of the Interconnection Network Simulation and Evaluation Envi-

ronment (INSEE) [11], which has been widely extended to model photonic networks. The INSEE simulator was originally developed with the aim of modeling electrical networks. This framework, developed at the University of Manchester, is publicly available and is one of main ExaNeSt project [10] simulation platforms.

INSEE implements multiple topologies (e.g. cubes and tree-like) and allows multiple traffic generation methods (e.g. synthetic, traces, and architectural simulators). Such a flexibility is not provided in other existing simulators [20]. INSEE achieves this flexibility with a frugal use of resources in terms of memory and CPU computing power, allowing the simulation of large systems (e.g. tens of thousands of nodes) in a few days at most.

The simulation environment consists of a functional simulator (FSIN) and a traffic generator (TrGen). Both components feature a modular design that can be augmented with new modules or extending the capabilities of the existing ones. We leveraged this modular design to implement multiple extensions, which allow INSEE to support all-optical networks.

As a simulator originally designed to study electrical networks, INSEE implements major electrical network components. Most of these components can be reused to simulate photonics based networks. Nevertheless, key components need to be highly modified to mimic the behavior either of photonics technology or photonics network components. This is the case of the routers, the switching technique, and the links. In addition, other techniques only apply to photonics technology. For instance, current photonics technology allows to populate a single fiber link with multiple channels.

IV. Proposed Photonics Extensions

This section presents the INSEE extensions developed to model photonics networks. For this purpose, we briefly discuss and compare the working behavior of the major components of a high-performance network. This analysis is carried out from the underlying technology (i.e. electrical vs photonics) perspective.

A. Optical Routers versus Electrical Routers

Electrical routers implement internal buffers that provide local temporal storage for in-transit packets (or a smaller data unit, depending on the switching technique). Packets are kept in these buffers in case they cannot advance due to traffic or network constraints.

As opposite, all-optical networks do not provide storage capacity at network routers, this means, that once the data is injected in the network it must travel without being blocked its paths. An interesting attempt to deal with this drawback could be the use of hybrid electro-optical routers. This approach, however, requires from electro-optical and opto-electrical conversions to write and read data into and from electric buffers, respectively, limiting the achievable bandwidth and latency improvements. As discussed below, the fact that all-optical networks do not provide buffering support, has important consequences for the switching technique used in all-optical networks.

B. Circuit Switching versus Packet Switching

Two main switching techniques, circuit switching and packet switching, can be used in electrical networks. In the former, a *circuit* path is established a priori that is then used to transmit the message; thus, routing, arbitration and switching are performed once for each message. In the latter, routing, arbitration and switching are performed on a per-packet basis.

Notice that packet switching cannot be supported by design by all-optical networks since, as mentioned before, buffering is not supported by optical routers. In addition, circuit switching is not implemented in modern electrical networks since it can be highly wasteful of scarce network bandwidth. This means that circuit switching is not usually modeled in current network simulators, so INSEE was enhanced with the purpose of using this technique in photonics networks.

C. Wavelength-Division Multiplexing

Although circuit switching can be considered a rather old switching method, it can bring important advantages in optical networks combined with wavelength-division multiplexing (WDM). WDM is a technique used in optical communications that consists in multiplexing in frequency a number of wavelengths onto the same optical cable. The amount of multiplexed wavelengths depends on the *separation* between them (e.g. as standard, with 100, 50 or 25 GHz there may be up to 40, 80 or 160 wavelengths respectively). When using WDM, the total link bandwidth (i.e. considering all the wavelengths), known as *aggregated bandwidth*, is given by the sum of the bandwidths provided by each individual wavelength.

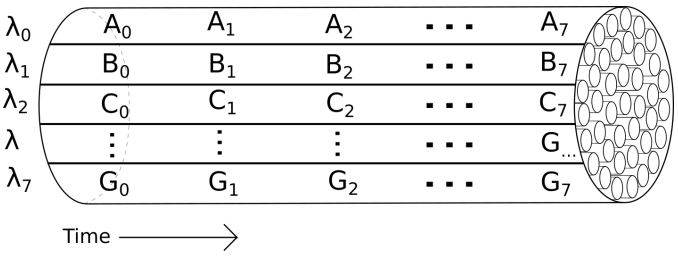
To model both WDM and circuit switching together in the baseline simulator, several design issues have been considered. First, since there are multiple wavelengths in the same link, circuit switching needs to be adapted since more than one path per link can be defined at the same time; that is, each channel (i.e. set of wavelengths) can be part of an eligible path. Therefore, when a message is ready to be injected into the network, the number of possible paths that it can reserve is much higher than in electric networks.

D. Photonic Links versus Electrical Links

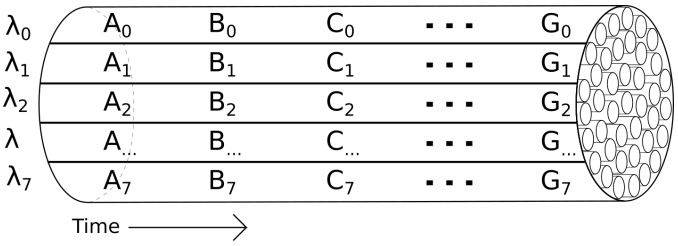
As explained above, electrical links only allow to send information of a single message or packet on a network cycle. In contrast, optical links are split in channels [1] each one using a different set of wavelengths.

To address this issue, we provide INSEE the support to configure the amount of wavelengths per optical link and to partition these wavelengths in independent channels, each one handling the transmission of a different message. In short, the first configuration option allows specifying the link bandwidth, and the second the channel bandwidth.

[1]Note: The term channel has been used in the literature also to refer to a single wavelength in optical technology. This paper uses this term from a *computer perspective* to refer to set a of wavelengths used to transfer the same message.



(a) Coarse-grain cycle approach.



(b) Packed wavelengths approach.

Fig. 1: Example of using eight wavelengths to transmit eight phits, referred to from A to G, with the studied transmision approaches.

E. Transmission units: Phits versus Bits

Virtual Cut-Through and Wormhole are the most widely used switching techniques in electrical networks. These techniques split the packet in small *flits*, which are the units of flow control. Flits are in turn divided in *phits* (physical units). A phit is the amount of bits that can be transferred in a single network cycle. In contrast, optical networks only can transfer one single bit per wavelength and per network cycle (note that optical cycles are much smaller than electrical cycles).

In general, electrical network simulators, and INSEE in particular, define the phit size as an integer amount of bytes (8 bits). Therefore, when adapting such a kind of simulator to work as a photonics simulator, it makes sense to keep the byte as the minimal transference unit per cycle. However, as mentioned above, optical links transfer one bit per wavelength in a given network cycle. Therefore, a new approach is required to fulfill this mismatch.

Two main approaches, *coarse-grain cycles* and *packed wavelengths*, have been devised to model the transmission of bits instead of phits in INSEE. The former approach defines coarse-grain cycles consisting of 8 *small* simulation cycles as the working cycle unit, which allows submitting 8 bits (i.e. 1 byte, the minimum phit size in the baseline simulator) per cycle using the same wavelength. The latter groups 8 wavelengths, which act as a single transmission unit; that is, 8 wavelengths are used to transmit 8 bits of the same message in a single photonic cycle, which implies that the minimum channel size is 8 wavelengths. Figure 1 presents an example where 8 phits (labeled from A to G) are transmitted in both approaches. In the coarse-grain cycles approach, each phit is transmitted in a different wavelength while in the packed wavelengths approach several wavelengths cooperate to transmit the same phit in parallel.

	# Channels	Channel Bandwidth
Coarse-grain cycles	40	40 Gbps
Hybrid	20	80 Gbps
Hybrid	10	160 Gbps
Packed wavelengths	5	320 Gbps

TABLE I: Trade-off between the studied transmission approaches for an optical link populated with 40 wavelengths.

As shown in Table I, choosing between both approaches presents a trade-off in the network features. For an optical link composed of 40 wavelengths, the coarse-grain cycle approach can provide up to 40 parallel channels, but each one of these channels only can offer the bandwidth of a single wavelength (40 Gbps). In contrast, the packed approach offers a limited maximum number (i.e. 5) of channels but each one aggregates the bandwidth of 8 wavelengths ($40 \times 8 = 320$ Gbps).

Note that between both approaches there are several possible *hybrid* approaches. For instance, the amount of channels can be halved with respect to the coarse-grain cycles approach (second and thrid line of the table). Then, instead of transmitting 1 byte in 1 network cycle using 1 wavelength, the byte can be divided in 2 nibbles that are transmitted by 2 wavelengths (doubling the network frequency).

F. Topologies and Routing

The network topologies implemented in INSEE need to be tailored to work under photonics technology. In particular, the routing algorithm must be adapted to circuit switching. We implemented minimal routing to establish the path between each source-destination pair. That is, only the paths that make the minimum number of hops are considered. To explore the benefits of photonics, a 3D torus topology has been modeled. We use deterministic minimal routing for the torus, in particular XYZ. Other topologies are left as for future work.

V. Performance Evaluation

The main aim of this section is to illustrate how a photonic network –consisting of the mechanisms and methods discussed in Section IV– performs in terms of network bandwidth and network latency with respect to an electrical network. In other

Technique	# Channels	Phit Size (bytes)	Channel BW (Gbps)
Electrical	-	4	10
Packed-wavelength	5	1	320
Hybrid	10	1	160
Hybrid	10	2	160
Hybrid	10	4	160
Hybrid	20	1	80
Hybrid	20	2	80
Coarse-Grain	40	1	40

TABLE II: Studied network configurations considering a link populated with 40 wavelengths.

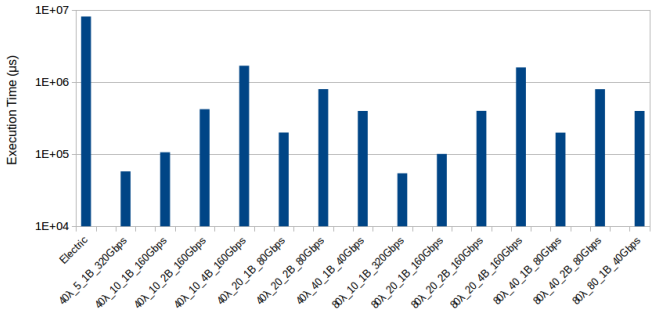


Fig. 2: Execution time (in us) for *Gadget*.

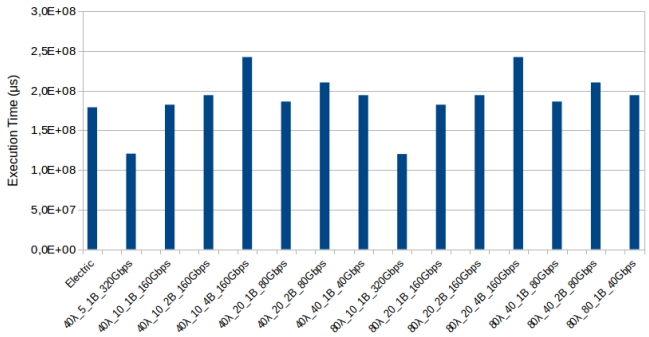


Fig. 3: Execution time (in us) for *Lammps*.

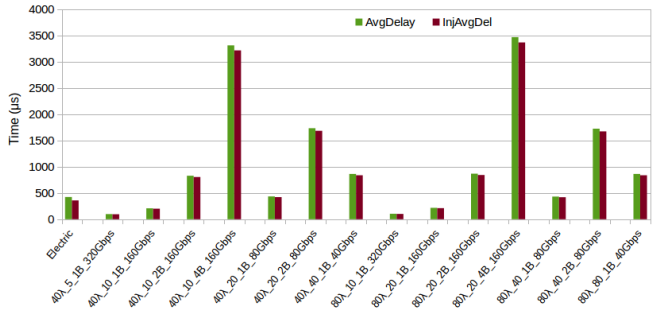


Fig. 4: Average network delay (in us) for *Gadget*.

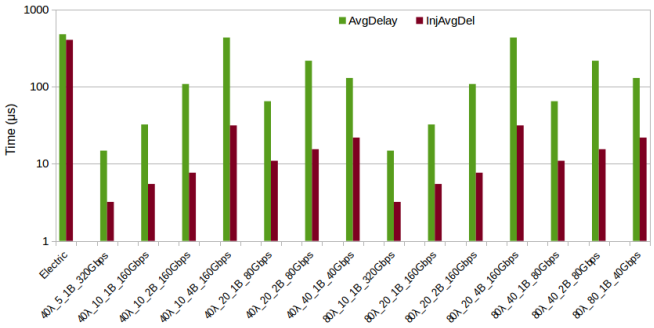


Fig. 5: Average network delay (in us) for *Lammps*.

words, we compare the results of the proposed extensions with the results provided by the baseline INSEE network.

Remember that multiple wavelengths share a link, where each of them can be part of an individual channel. Next, we discuss the design options that have been selected to carry out the experiments presented in this paper.

A. System Details

This section specifies the network bandwidth, the number of wavelengths per link and the network topology that have been considered to obtain the results.

The experiments consider a 10 Gigabit Ethernet electrical network. In the case of the photonics network 40 Gbps is considererd as the bandwidth provided by each individual wavelength. The electrical bandwidth was chosen because an important set (by 35.6%) of the supercomputers ranked in Top500 [1] implement this network technology, while the optical was selected according to the actual VCSEL technology commented in Section II.

On the other hand, photonic networks are limited by the optic communication band [21] and, as explained in Section IV-C, the amount of wavelength depends on the spacing between them. Nowadays, 100 GHz channel spacing is typically used, which gives 40 wavelengths per single optical fiber or link [22], but this spacing can be reduced in order to populate more wavelengths per single fiber or optical link. For instance, 50 GHz spacing allows to populate the link with 80 wavelengths, or recently [23], 160 wavelengths are allowed with a 25 GHz spacing.

Table II summarizes the main design choices of the studied configurations for a photonic link populated with 40 optical wavelengths. In addition to the two main transmission approaches, labelled as *packing wavelength* and *coarse grain*, hybrid schemes combining both approaches have been studied. All the configurations present an aggregate bandwidth of 1.6 Tbps with 40 wavelengths per link. This aggregate bandwidth is equally split among the channels of each configuration. In this work, we also study future optical links with 80 wavelengths, which gives 3.2 Tbps as aggregate bandwidth per single fiber. To carry out this study, we devised new configurations obtained by doubling the amount of channels of the configurations depicted in Table II.

To launch the experiments we consider two excerpts of traces collected from the execution of two ExaNest workloads, `Gadget` and `Lammps`. These traces have different number of nodes; thus, for `Gadget`, which has 72 nodes, we have used a 3D torus with 4x6x3 nodes and for `Lammps`, which has 192 nodes, the 3D torus is configured with 4x8x6 nodes.

B. Experimental Results

Figure 2 and Figure 3 present the execution time obtained with the electrical network and the studied photonic network configurations for both excerpts. Each configuration has associated four arguments, referring to (from left to right)

the amount of wavelengths per link, the number of possible channels, the size of the phit and the bandwidth per channel.

First, it can be observed that the configurations with 40 wavelengths per link provide the same results as those with 80 wavelengths regardless of the workload. This means that an aggregate link bandwidth of 1.6 Tbps, which is achievable with current photonics technology suffices to achieve the best performance. Thus, from now on, we focus on configurations with 40 wavelengths per link. Among these, the configuration that provides the best results in both studied workloads is the *packed-wavelength* one (see Table II). This is because it features the highest bandwidth per channel (320 Gbps), which means that this parameter is the one that most impacts on performance.

On the other hand, for a given bandwidth per channel (e.g. 160 Gbps), the best strategy is to reduce the phit size. For instance, in both excerpts, the best configuration for 80 Gbps per channel is the one with 1-byte phit size (i.e. 40λ_20_1B_80Gbps). This is because large phit sizes waste bandwidth when they are underused. This effect is so significant that sometimes it is better to reduce the phit size even if the bandwidth per channel is reduced (e.g. comparing 40λ_10_2B_160Gbps with 40λ_20_1B_80Gbps in `Gadget`).

Comparing both figures, it can be seen that while in `Gadget` the photonic network widely improves the performance against the electric network (ranging the benefits from 1 to 2 orders of magnitude), in `Lammps` the execution time is similar or better in the electrical network except in the optimal configurations with 320 Gbps per channel. The reason behind these results is that the former excerpt belongs to a phase of the execution dominated by network transactions with scarce computation time while the latter has a large amount of computation time interleaved with network usage. Thus, the benefits of photonics are only significant with the highest bandwidths.

Figure 4 presents the average transmission delay, which include the injection delay and the transit delay of packets, for *Gadget*. As can be observed, almost all the delay is caused by injection. Although in this figure the electric network reaches a low delay, it must be taken into account that in the electric network the delay is computed per each 64-byte packet while in the photonic configurations the delay is calculated for the transmission of a whole message, thus they cannot be directly compared. Nevertheless, the results confirm a strong correlation between network delay and execution time for the optical configurations.

In `Lammps`, as shown in Figure 5, the distribution of delays is completely different between electrical and photonic networks. While in the former most of the delay is caused by injection, in the latter the average delay is evenly distributed between injection and transit delays. This is because the size of the messages in `Lammps` is longer on average. Thus, a low photonic bandwidth per channel has a high impact on the total delay, affecting execution time accordingly.

VI. Related Work

The interest of the academia and industry communities in the development of Exascale systems, and in improving on-chip architectures, has fostered the research on photonics technology. In order to study these systems, novel simulation and estimation tools are required.

Current network simulators [24], [25], [26] focus on packet-switching electrical networks. These tools can be easily adapted to model packet-switching hybrid electro-optical networks. However, in order to adapt them to model the circuit switching capabilities of all-optical networks, a significant amount of programming effort is required. Due to this fact, some tools have been recently proposed designed from the ground up to support all-optical networks. In this regard, a well known simulator is PhoenixSim [27], [28]. This framework models multiprocessor systems that use electrical networks, optical networks, and hybrid networks. PhoenixSim is based on the OMNeT++ simulation environment [29] and allows the analysis of interconnection networks from both the physical level (e.g. optical insertion loss, crosstalk, energy dissipation) and the system level (e.g. latency, performance, execution time).

The Design Space Exploration of Networks Tool (DSENT) [30] improves the PhoenixSim model of electro-optical interface circuitry such as modulators, receivers, and thermal tuning, capturing trade-offs among photonic devices and modulator/receiver specifications that can be exploited to reach optimal configurations in terms of area and power. DSENT is designed to enable fast area and power evaluation of multiple optical network configurations and, when coupled with an architectural simulator, to obtain power and area estimations for the simulated network. However, DSENT does not model photonic switches so it cannot be used to simulate circuit-switched networks such as the evaluated in this work. In addition, DSENT does not support traffic patterns and workload traces, so it cannot provide the details of a system-level simulation.

LioeSim [31] is a electrical and optical network simulator that uses Orion [32] for the models of electrical routers and links. Unlike DSENT, it models photonic switches and allows analyzing both physical level (optical insertion loss, crosstalk, optical power budget, energy dissipation) and system level (latency, energy delay product) performance metrics of interconnection networks. Unfortunately, LioeSim is focused on on-chip networks. In contrast, in this work we simulate off-chip all-optical networks for intra-rack and inter-rack communications.

Finally, there is also a need for aiding designers in layout tasks such as visually placing photonic devices, connecting waveguides, etc. To this end, in [33], Hendry et al. introduce the Visual Automated Nanophotonic Design And Layout (VANDAL), which also can be interfaced with industry-standard software tools for chip fabrication processes.

VII. Conclusions

The data movement is one, if not the most, critical challenge to reach Exascale computation in future supercomputers. Exascale networks will count with thousands of nodes, and photonics has emerged as a promising technology to face this data challenge. To this end, photonic networks simulation tools are required to guide designers in decision taking.

This work has discussed the process required to build a photonic computer network simulator on top of the INSEE electrical network simulation framework. We have identified the key components that are involved, described the major working differences between such components depending on the underlying technology (electrical and optical), and studied optical-only features of some components (e.g. WDM). This paper has discussed major simulator extensions in the router and the link architecture, WDM implementation over circuit switching, and specific routing methods.

For illustrative purposes some experiments have been conducted aimed at comparing photonic networks with electrical networks in a 3D torus topology. Multiple configurations have been evaluated varying four main parameters: the amount of wavelengths, the number of possible channels, the size of phit, and the bandwidth per channel. Experimental results, obtained with excerpts of real applications used in the ExaNeSt project, show that depending on the optical network configuration the execution time of the application can widely differ even with two optical network technologies (e.g. 1.6 and 3.2 Tbps aggregate link bandwidth, that is, 40 and 80 wavelengths respectively). In general, the future 3.2 Tbps aggregate link bandwidth will not provide additional performance benefits for the studied workloads, but 1.6 Tbps and 320 Gbps per channel is enough to obtain the best results across the studied configurations. Moreover, we found that the parameter that most impacts on performance is the bandwidth per channel, achieving the best results with 320 Gbps channels. Finally, for lower bandwidths per channel (e.g. 160 and 80 Gbps), reducing the phit size provides the best trade-off.

Acknowledgments

This work was supported by the ExaNest project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 671553, and by the Spanish Ministerio de Economía y Competitividad (MINECO) and Plan E funds under Grant TIN2015-66972-C5-1-R.

References

[1] (2015, Apr) Top500 website. [Online]. Available: http://www.top500.org/

[2] A. K. Kodi, B. Neel, and W. C. Brantley, "Photonic interconnects for exascale and datacenter architectures," *IEEE Micro*, vol. 34, no. 5, pp. 18–30, 2014.

[3] S. Rumley, D. Nikolova, R. Hendry, Q. Li, D. Calhoun, and K. Bergman, "Silicon photonics for exascale systems," *Journal of Lightwave Technology*, vol. 33, no. 3, pp. 547–562, 2015.

[4] 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, pp. 1246–1260, 2008.

[5] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. W. Holzwarth, M. A. Popovic, H. Li, H. I. Smith, J. L. Hoyt *et al.*, "Building many-core processor-to-dram networks with monolithic cmos silicon photonics," *IEEE Micro*, vol. 29, no. 4, 2009.

[6] S. Werner, J. Navaridas, and M. Lujan, "Designing low-power, low-latency networks-on-chip by optimally combining electrical and optical links," in *The 23rd IEEE Symposium on High Performance Computer Architecture*. IEEE, 2016.

[7] J. Puche, S. Lechago, S. Petit, M. E. Gómez, and J. Sahuquillo, "Accurately modeling a photonic noc in a detailed cmp simulation framework," in *High Performance Computing & Simulation (HPCS), 2016 International Conference on*. IEEE, 2016, pp. 387–394.

[8] G. Chen, H. Chen, M. Haurylau, N. A. Nelson, D. H. Albonesi, P. M. Fauchet, and E. G. Friedman, "On-chip copper-based vs. optical interconnects: delay uncertainty, latency, power, and bandwidth density comparative predictions," in *Interconnect Technology Conference, 2006 International*. IEEE, 2006, pp. 39–41.

[9] (2017, Apr) ExaNeSt website. [Online]. Available: http://exanest.eu/

[10] M. Katevenis, N. Chrysos, M. Marazakis, I. Mavroidis, F. Chaix, N. Kallimanis, J. Navaridas, J. Goodacre, P. Vicini, A. Biagioni *et al.*, "The exanest project: Interconnects, storage, and packaging for exascale systems," in *Digital System Design (DSD), 2016 Euromicro Conference on*. IEEE, 2016, pp. 60–67.

[11] F. J. Ridruejo Perez and J. Miguel-Alonso, "Insee: An interconnection network simulation and evaluation environment," in *Proceedings of the 11th International Euro-Par Conference on Parallel Processing*, ser. Euro-Par'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 1014–1023.

[12] G. . H. Duan, C. Jany, A. L. Liepvre, J. G. Provost, D. Make, F. Lelarge, M. Lamponi, F. Poingt, J. M. Fedeli, S. Messaoudene, D. Bordel, S. Brision, S. Keyvaninia, G. Roelkens, D. V. Thourhout, D. J. Thomson, F. Y. Gardes, and G. T. Reed, "10 gb/s integrated tunable hybrid iii-v/si laser and silicon mach-zehnder modulator," in *2012 38th European Conference and Exhibition on Optical Communications*, Sept 2012, pp. 1–3.

[13] G. H. Duan, C. Jany, A. L. Liepvre, M. Lamponi, A. Accard, F. Poingt, D. Make, F. Lelarge, S. Messaoudene, D. Bordel, J. M. Fedeli, S. Keyvaninia, G. Roelkens, D. V. Thourhout, D. J. Thomson, F. Y. Gardes, and G. T. Reed, "Integrated hybrid iiiv/si laser and transmitter," in *2012 International Conference on Indium Phosphide and Related Materials*, Aug 2012, pp. 16–19.

[14] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE Journal of Quantum Electronics*, vol. 23, no. 1, pp. 123–129, Jan 1987.

[15] A. Liu, L. Liao, D. Rubin, H. Nguyen, B. Ciftcioglu, Y. Chetrit, N. Izhaky, and M. Paniccia, "High-speed optical modulation based on carrier depletion in a silicon waveguide," *Opt. Express*, vol. 15, no. 2, pp. 660–668, Jan 2007. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-15-2-660

[16] D. J. Thomson, F. Y. Gardes, Y. Hu, G. Mashanovich, M. Fournier, P. Grosse, J.-M. Fedeli, and G. T. Reed, "High contrast 40gbit/s optical modulation in silicon," *Opt. Express*, vol. 19, no. 12, pp. 11 507–11 516, Jun 2011. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-19-12-11507

[17] K. Bergman, L. P. Carloni, A. Biberman, J. Chan, and G. Hendry, *Photonic network-on-chip design*. Springer.

[18] P. Dong, L. Chen, C. Xie, L. L. Buhl, and Y.-K. Chen, "50-gb/s silicon quadrature phase-shift keying modulator," *Opt. Express*, vol. 20, no. 19, pp. 21 181–21 186, Sep 2012. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-20-19-21181

[19] P. Dong, X. Liu, C. Sethumadhavan, L. L. Buhl, R. Aroca, Y. Baeyens, and Y.-K. Chen, "224-gb/s pdm-16-qam modulator and receiver based on silicon photonic integrated circuits," in *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2013*. Optical Society of America, 2013, p. PDP5C.6. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=OFC-2013-PDP5C.6

[20] J. Navaridas, J. Miguel-Alonso, J. A. Pascual, and F. J. Ridruejo, "Simulating and evaluating interconnection networks with {INSEE}," *Simulation Modelling Practice and Theory*, vol. 19, no. 1, pp. 494 – 515, 2011, modeling and Performance Analysis of Networking and Collaborative Systems. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1569190X1000184X

[21] V. Alwayn, *Optical network design and implementation*. Cisco Press, 2004.

[22] R.-J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," *Proceedings of the IEEE*, vol. 100, no. 5, pp. 1035–1055, 2012.

[23] E. Temprana, E. Myslivets, B.-P. Kuo, L. Liu, V. Ataie, N. Alic, and S. Radic, "Overcoming kerr-induced capacity limit in optical fiber transmission," *Science*, vol. 348, no. 6242, pp. 1445–1448, 2015.

[24] Y. Ben-Itzhak, E. Zahavi, I. Cidon, and A. Kolodny, "Hnocs: Modular open-source simulator for heterogeneous nocs," in *Embedded Computer Systems (SAMOS), 2012 International Conference on*. IEEE, 2012, pp. 51–57.

[25] H. Hossain, M. Ahmed, A. Al-Nayeem, T. Z. Islam, and M. M. Akbar, "Gpnocsim-a general purpose simulator for network-on-chip," in *Information and Communication Technology, 2007. ICICT'07. International Conference on*. IEEE, 2007, pp. 254–257.

[26] L. Jain, B. Al-Hashimi, M. Gaur, V. Laxmi, and A. Narayanan, "Nirgam: a simulator for noc interconnect routing and application modeling," in *Design, Automation and Test in Europe Conference*, 2007, pp. 16–20.

[27] J. Chan, G. Hendry, A. Biberman, K. Bergman, and L. P. Carloni, "Phoenixsim: A simulator for physical-layer analysis of chip-scale photonic interconnection networks," in *Proceedings of the Conference on Design, Automation and Test in Europe*, ser. DATE '10. 3001 Leuven, Belgium, Belgium: European Design and Automation Association, 2010, pp. 691–696. [Online]. Available: http://dl.acm.org/citation.cfm?id=1870926.1871093

[28] S. Rumley, M. Bahadori, K. Wen, D. Nikolova, and K. Bergman, "Phoenixsim: Crosslayer design and modeling of silicon photonic interconnects," in *Proceedings of the 1st International Workshop on Advanced Interconnect Solutions and Technologies for Emerging Computing Systems*. ACM, 2016, p. 7.

[29] A. Varga and R. Hornig, "An overview of the omnet++ simulation environment," in *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops*, ser. Simutools '08. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 60:1–60:10. [Online]. Available: http://dl.acm.org/citation.cfm?id=1416222.1416290

[30] C. Sun, C.-H. O. Chen, G. Kurian, L. Wei, J. Miller, A. Agarwal, L.-S. Peh, and V. Stojanovic, "Dsent-a tool connecting emerging photonics with electronics for opto-electronic networks-on-chip modeling," in *Networks on Chip (NoCS), 2012 Sixth IEEE/ACM International Symposium on*. IEEE, 2012, pp. 201–210.

[31] X. Ma, J. Yu, X. Hua, C. Wei, Y. Huang, L. Yang, D. Li, Q. Hao, P. Liu, X. Jiang, and J. Yang, "Lioesim: A network simulator for hybrid opto-electronic networks-on-chip analysis," *Journal of Lightwave Technology*, vol. 32, no. 22, pp. 4301–4310, Nov 2014.

[32] A. B. Kahng, B. Li, L.-S. Peh, and K. Samadi, "Orion 2.0: a fast and accurate noc power and area model for early-stage design space exploration," in *Proceedings of the conference on Design, Automation and Test in Europe*. European Design and Automation Association, 2009, pp. 423–428.

[33] J. Chan, G. Hendry, K. Bergman, and L. P. Carloni, "Physical-layer modeling and system-level design of chip-scale photonic interconnection networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 10, pp. 1507–1520, Oct 2011.