# Scalable NoC Architectures: Efficient and Low Energy Consumption Chip Communication

Sergi Abadal, Albert Cabellos-Aparicio, José A. Lázaro*, Eduard Alarcón, Josep Solé-Pareta

NaNoNetworking Center in Catalunya (N3Cat)
*Optical Communications Group (GCO)
Universitat Politècnica de Catalunya (UPC)
C/ Jordi Girona 1-3, 08034 Barcelona, Spain
{abadal,acabello,pareta}@ac.upc.edu, jose.lazaro@ tsc.upc.edu and eduard.alarcon@upc.edu

*Abstract*— **Energy has been the primary reason for shifting from traditional single-core processors to current multicore processors. Such multicore designs require an interconnection network to communicate cores among themselves and with memory. As the number of cores per chip increases, the energy consumption of these Networks-on-Chip (NoC) has become comparable to that of the cores computation, to the point of being expected to be the performance bottleneck of near future processors. Novel technologies, such as nanophotonics, have recently appeared as a response to the urgent need for efficient and scalable chip communication. In this paper, different alternatives for the reduction of the NoC power demands are analyzed from the interconnect level up to the microarchitectural level. Additionally, a graphene-enabled wireless/optical-wired communication architecture, consisting on both a photonic NoC and a wireless NoC, is proposed. The former for transferring heavy flows of data, the latter for supporting the control plane of the whole network, and carrying light data flows.**

*Keywords— Network-on-Chip; Efficient Chip Communication; Nanophotonics; Silicon-on-Insulator; Graphene; Plasmonics;*

## I. INTRODUCTION

The energy consumption of a traditional CMOS-based processor scales linearly with the operation frequency. Therefore, achieving better performance by increasing the operation frequency would result in unaffordable energy figures. Instead, multiprocessor designs have recently appeared, wherein several independent processors or cores are interconnected by means of a Network-on-Chip (NoC).

Thus far, NoC consists of a network of electrical interconnects and routers that convey the information from the transmitting core to the receiving core. However, it remains unclear whether such copper-based designs will scale beyond several tens of cores, since the energy requirements of these on-chip networks become more stringent as technology advancements allow the integration of more cores per chip. Fig. 1 quantifies this trend by showing a near-future projection of the chip communication requirements in terms of energy per bit [1].

Indeed, a higher number of cores entail a significant surge in the energy consumed in the communication for the same interconnect technology. For instance, in [2], the NoC consumes approximately a 40% of the total chip power. If this trend continues, the energy spent in communication could even surpass the energy consumed by the computation itself. In other words, the performance bottleneck of modern processors could soon shift from the computation to the communication if the energy per bit increasing requirements are not satisfied.

In light of this, the research community has been devoting considerable efforts in finding feasible alternatives to traditional on-chip networks, in order to achieve the necessary energy efficiency in chip communication. On the one hand, while the state-of-the-art in copper-based interconnects has demonstrated efficiencies below 100 fJ/b [3], this comes at the expense of lower data rates or larger transceiver sizes and, as seen in Fig. 1, might not suffice. On the other hand, novel interconnect technologies such as nanophotonics are being investigated seeking extremely low energy per bit, i.e. well below the 10 fJ/b barrier, as well as good bandwidth density figures [1, 3, 4].

In this paper, we aim to go one step further by exploring different ways to overcome this issue not just from the interconnect perspective, but also from different levels of design (Section 2). We also present a preliminary approach aiming to substantially reduce the energy consumption of modern multiprocessors to, this way, ensure their scalability in terms of number of cores per chip (Section 3). Finally, the paper is concluded in Section 4.

## II. REDUCING THE ENERGY FOOTPRINT OF A NoC

The field of on-chip networking is an extremely transversal field that covers several areas such as electronics for the electronic implementation, communications for the design of the inter-core networks, up to computer architecture for the design of the resulting multiprocessors. Consequently, the task of reducing the energy footprint of a NoC can be seen as the result of combining efforts at different levels of design, namely:

### A. Interconnect Level

The energy efficiency yielded by the on-chip interconnects has a major impact in the energy consumed in communication regardless of the on-chip network architecture. Nanophotonics and plasmonics are regarded as serious contenders in this respect since they open the door for CMOS-compatible interconnects with huge bandwidth per energy figures, which
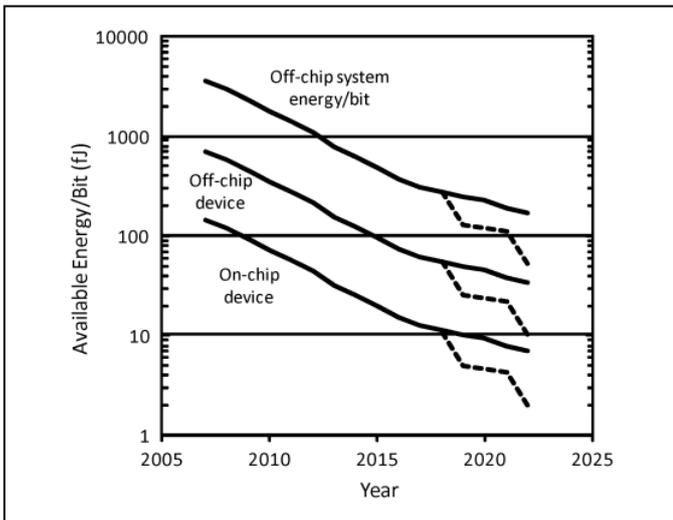
*Figure 1. Projected evolution of the chip communication requirements in terms of energy per bit (extracted from [1])*

are projected to be orders of magnitude higher that of traditional electrical interconnects [3]. Such performance and efficiency are possible by virtue of the recent advancements in the field of plasmonics [5], as well as two novel photonics technologies, namely Silicon-on-Insulator [4] and Graphene Nanophotonics [6]. Indeed, the aforementioned technologies are providing ways to design and develop small footprint devices for the integrated on-chip transmission [5, 7, 8], switching [9] and reception [10, 11] of optical signals. Potential has been demonstrated for the operation of such devices at energies as low as 1 fJ/b [7] with extremely high bandwidth leading to data rates of several tens of gigabits per second.

### B. Network Level

The design of the on-chip network needs to take into account the features of the underlying interconnect technology in order to provide efficient and fair end-to-end communication among the cores of the multiprocessor, including the design of appropriate physical and logical topologies and protocols. One example of efficiency improvement at the network level can be found in traditional NoCs. The path between the transmitter and the receiver in these networks is decided by a set of on-chip routers. This process consumes time and energy proportionally to the number of routers or 'hops' in the information path. By reducing the distance between cores in terms of number of hops, the energy devoted in the communication is greatly reduced. In this respect, advancements in the manufacturing of 3D active devices are allowing the design of network topologies which reduce both the physical and logical (hop) distance among cores and which would be not possible in a planar environment [12].

### C. Microarchitecture Level

The architecture of a multiprocessor, which includes the definition of operations that are needed for the correct functioning of these processors, e.g. memory coherency and consistency, is generally adapted to the characteristics of the underlying on-chip network and interconnect technology. However, changes in these two aspects could allow the design of novel multiprocessor architectures that would need a significantly lower amount of communication. In other words, higher overall efficiency could be accomplished by devising architectures needing less communication, on top of interconnection networks that might not be strictly the most efficient in terms of energy per bit.

### III. A GRAPHENE-ENABLED HYBRID ON-CHIP NETWORK ARCHITECTURE

The objective of this section is to present our tentative approach in the area of efficient and low energy chip communication, to later discuss the improvements that this architecture could yield at the aforementioned levels of design. Fig. 2 shows a schematic representation of such approach. The proposed design is hybrid since it combines two different interconnect technologies, both enabled by graphene, as follows:

- **Nanophotonic plane:** a photonic NoC will be implemented by means of nanophotonic components for the efficient transmission of heavy flows of data. The employment of graphene-based components could ensure, on the one hand, broadband operation with intrinsic bandwidths of up to 500 GHz [11], adequate for the high data rates needed in this scenario; on the other hand, size compatibility with future core sizes by virtue of the high absorption of graphene leading to reduced footprint devices. For example, preliminary results have demonstrated modulators of only 25 $\mu m^2$ [7].

- **Wireless plane:** in this case, wireless communication can be implemented by means of graphene-based nano-antennas [13, 14], for the control of the photonic plane and the transmission of selected flows of data. The employment of graphene-based nano-antennas is justified by the size and potential bandwidth they offer. Indeed, due to the plasmonic effects present at the surface of a graphene patch, graphene nano-antennas are able to radiate at the same frequency band that of a metallic antenna two orders of magnitude larger. Preliminary results show that a few micrometers long and wide graphene nano-antenna, which would be size compatible with future core sizes, would be able to radiate in the Terahertz band. The interested reader will find more details in [13, 14].
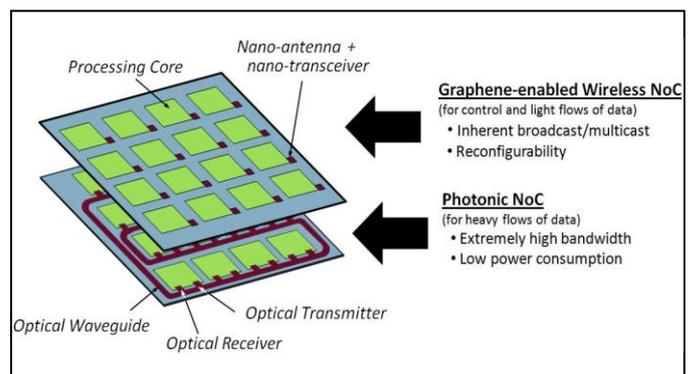


*Figure 2. Example of a graphene-enabled hybrid wireless-optical NoC*

The energy efficiency of multiprocessors relying on such graphene-enabled hybrid architecture could be greatly enhanced at all levels of design due to the following reasons:

- *Interconnect Level:* as pointed out above, the potential energy per bit figures of photonic interconnects substantially outperforms that of its electrical counterparts, while offering extremely high data rates which are might be necessary as the number of cores per chip increases.

- *Network Level:* wireless communication, with its inherent flexibility and broadcast capabilities, enables the design of communication schemes which are extremely costly to implement with other interconnect technologies, such as multicast communication. Hence, by transmitting selected flows of data through the wireless plane, a significant amount of power can be saved. Also, it is worth noting that wireless communication can be used to control the photonic NoC, instead of devoting resources to implement contention-free and often complex photonic network architectures that do not scale well with increasing number of nodes [15, 16]. For instance, the wireless control plane could be used to drive the switches of a photonic NoC like the one presented in [9].

- *Microarchitecture Level:* the special characteristics of on-chip wireless communication create a wide range of opportunities at the microarchitectural level. This way, novel multiprocessor architectures could be designed requiring less communication among cores. For instance, new ways of defining cache coherence might be created based on the potentially reduced cost of multicast communications.

## IV. CONCLUSIONS

Chip communication is a key component for the correct operation of multicore processors. As the number of cores per chip keeps increasing, there is a heightened need for efficient and low energy chip communication. In this paper, we analyzed this research challenge from different design perspectives. We then proposed a tentative graphene-based hybrid architecture and qualitatively discussed the potential improvements of such architecture from the aforementioned design perspectives.

REFERENCES

[1] D. A. B. Miller, "Device requirements for optical interconnects to silicon chips," Proc. IEEE 97, 1166-1185, 2009.

[2] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, et al, "An 80-tile 1.28TFLOPS network-on-chip in 65 nm CMOS," in IEEE ISSCC Dig. Tech. Papers, Feb. 2007, pp. 98–99.

[3] J. Postman and P. Chiang, "A survey addressing on-chip interconnect: energy and reliability considerations," ISRN Electronics, vol. 2012, pp. 1-9, 2012.

[4] Y. Vlasov, "Silicon CMOS-integrated nano-photonics for computer and data communications beyond 100G," *Communications Magazine, IEEE*, no. February, pp. 67-72, 2012.

[5] H. M. G. Wessel, D. Dai, M. Tiwari, J. K. Valamehr, L. Theogarajan, J. Dionne, *et al,* "Opportunities and challenges of using plasmonic components in nanophotonic architectures", IEEE Journal in Emerging and Selected Topics in Circuits and Systems, vol. 2, no. 2, 2012, pp. 154-168.

[6] F. Xia, T. Mueller, Y.-M. Lin, P. Avouris, "Graphene nanophotonics," 23rd Annual Meeting of the IEEE Photonics Society, 2010, pp.76-77, Nov. 2010

[7] W. Cai, J. S. White, Mark L. Brongersma, "Compact, high-speed and power-efficient electrooptic plasmonic modulators," Nano Letters, vol. 9, no. 12, 2009, pp. 4403-4411.

[8] M. Liu, X. Zhang, "Graphene-based optical modulators", Optical Fiber Communication Conference (OFC), 2012.

[9] 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, Sep. 2008.

[10] S. Assefa, F. Xia, W. M. J. Green, C. L. Schow, A. V. Rylyakov, Y. Vlasov, "CMOS-Integrated optical receivers for on-chip interconnects," IEEE Journal of Selected Topics in Quantum Electronics, vol. 16, no. 5, pp. 1376-1385, Sep. 2010.

[11] F. Xia, T. Mueller, Y.-M. Lin, Alberto Valdés-García, P. Avouris, "Ultrafast graphene photodetector", Nature Nanotechnology, vol. 4, pp. 839-843, Dec 2009.

[12] B. S. Feero and P. P. Pande, "Networks-on-Chip in a three-dimensional environment: a performance evaluation," IEEE Transactions on Computers, vol. 58, no. 1, pp. 32-45, Jan. 2009.

[13] J. M. Jornet and I. F. Akyildiz, "Graphene-based nano-antennas for electromagnetic nanocommunications in the terahertz band," in Proc. of 4th European Conference on Antennas and Propagation, 2010.

[14] I. Llatser, C. Kremers, D. N. Chigrin, J. M. Jornet, M. C. Lemme, A. Cabellos-Aparicio, E. Alarcón, "Characterization of graphene-based nano-antennas in the terahertz band," in Proc. of 6th European Conference on Antennas and Propagation, 2012.

[15] G. Kurian, J. E. Miller, J. Psota, J. Eastep, J. Liu, J. Michel, *et al.* "ATAC: A 1000-Core Cache-Coherent Processor with On-Chip Optical Network," in Proceedings of the 19th international conference on Parallel architectures and compilation techniques. ACM, 2010, pp. 477–488.

[16] S. Koohi, M. Abdollahi, and S. Hessabi, "All-optical wavelength-routed NoC based on a novel hierarchical topology," in Fifth IEEE/ACM International Symposium on Networks on Chip (NoCS), Pitts, 2011, pp. 97–104.