

Evaluating the Feasibility of Wireless Networks-on-Chip Enabled by Graphene

Sergi Abadal, Albert Mestres, Mario Iannazzo, Josep Solé-Pareta, Eduard Alarcón,
Albert Cabellos-Aparicio
NaNoNetworking Center in Catalunya
Campus Nord - UPC, Jordi Girona 1-3, 08034 Barcelona, Spain
abadal@ac.upc.edu

ABSTRACT

Network-on-Chip (NoC) is currently the paradigm of choice for covering the on-chip communication needs of multicore processors. As we reach the manycore era, though, electrical interconnects present performance and power issues that are exacerbated in the presence of multicast communications due to the point-to-point nature of NoCs. This dramatically limits the available design space in terms of manycore architecture, sparking the need for new solutions. In this direction, the use of wireless interconnects has been recently proposed as a complement of a wired plane. In this paper, the concept of Graphene-enabled Wireless Network-on-Chip (GWNoC) is introduced, which extends the native broadcast capabilities of existing wireless NoCs by enabling the per-core integration of antennas that radiate in the terahertz band (0.1 - 10 THz). Preliminary results on the feasibility of GWNoC are presented, covering implementation, on-chip networking and multiprocessor architecture aspects.

Categories and Subject Descriptors

C.2.1 [Network Architecture and Design]: Wireless Communication; C.1.2 [Multiple Data Stream Architectures (Multiprocessors)]: Interconnection architectures

Keywords

Network-on-Chip, Manycore, Wireless, Graphene Antennas, Terahertz, Feasibility, Scalability

1. INTRODUCTION

Following the wide adoption of multicore processors and recent dawn of the manycore era, communication has been gradually replacing computation as the main determinant of the performance of computing systems. Within this context, Network-on-Chip (NoC) has become the paradigm of choice for covering the communication needs arising from, among others, coherency, consistency, or synchronization in shared-memory multiprocessors [1]. A NoC generally consists in

a fabric of wireline routed interconnections and was proposed as a solution to the very limited scalability of buses. However, as number of cores per chip increases, traditional NoCs suffer from fundamental issues that will render them impractical in future multiprocessors [2]. Given the strong correlation between the multiprocessor architecture and the employed NoC, this fact is expected to have a large influence upon the design of future manycore processors.

Figure 1 exemplifies the main challenge that conventional NoCs will need to face as we reach the manycore era: latency. There is a strong contradiction between how the NoC performance scales and *how it should scale* with the system size. One of the objectives of integrating more cores in the same chip is to obtain an increase of the execution speed, which implies a significant growth in terms of on-chip communication requirements. While this increasing in traffic would ideally be served with negligible performance losses as the architecture scales [3], the reality is that the performance of conventional NoCs tends to drop significantly.

While recent works seek to improve the scalability of the underlying wires and routers in order to reverse this trend [4, 5], solutions are also required at upper levels of design. For instance, a major cause of the performance contradiction depicted in Figure 1 is the poor performance of NoCs in the presence of multicast communication. Such traffic is typically generated by synchronization or coherency methods and, due to the point-to-point nature of NoCs, requires flits to be forked (replicated) either at the source or within the network [6]. This *flit forking* approach degrades the network performance proportionally not only to the multicast rate [7], but also to the number of cores due to the expected increase in terms of number of destinations per message. Given the high cost of multicast in multicore processors, these are architected seeking to minimize such type of traffic at the expense of lower overall performance and higher system complexity [8]. Such strategy, though, may be rendered impractical in manycore settings as the performance and complexity penalties become more strict.

The introduction of a dedicated platform for on-chip multicast communication would therefore have a twofold impact. First, it would improve the performance of existing architectures by relieving the main NoC of inefficiently serving multicast traffic; and second, it would enhance the support of multicast-based architectural methods [7] in the pathway to simplify the design of manycore processor architectures. The implementation of such multicast plane will not be feasible with NoCs based on electrical interconnects, at least in its conventional form, due to evident performance and effi-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NoCArc '14 December 13 - 14 2014, Cambridge, United Kingdom

Copyright 2014 ACM 978-1-4503-3064-0/14/12\$15.00

<http://dx.doi.org/10.1145/2685342.2685345>

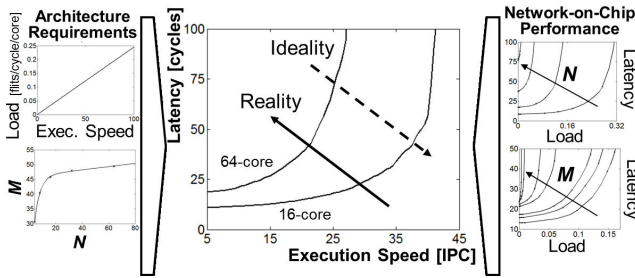


Figure 1: Left plots, derived from results in Sec. 3.3 (full-system simulation): as more cores (N) are integrated seeking higher execution speeds, load and percentage of multicasts (M) grow. Right plots, derived from results in Sec. 3.2 (NoC simulation): the performance of a 2D mesh scales poorly with N and with M . Center plot, combination of left and right plots: NoC latency should improve to enable speedups, yet in reality it decreases.

ciency issues. Instead, the use of emerging technologies that reach beyond the limits of such interconnects provide new opportunities towards cost-effective multicast at the chip scale. Some examples include nanophotonic interconnects [9, 2] and wireless on-chip networks [10]. It is important to note that scalability challenges may still arise due to, for instance, laser power issues in the former case or the size of the on-chip antennas in the latter case [11].

In this position paper, we present the Graphene-enabled Wireless Network-on-Chip paradigm (GWNoC, [12]) as a strong candidate for the implementation of a dedicated multicast plane. As outlined in Section 2, the choice is motivated by three main reasons. First, GWNoC has inherent broadcast capabilities due to the shared medium nature of wireless communication. Second, it offers such capabilities on a per-core basis by virtue of the reduced size of graphene-based micro-scale antennas [13, 14, 15]. Third, it delivers potential to support multi-hundred gigabit-per-second rates as graphene-based antennas radiate in the terahertz band (0.1-10 THz).

To evaluate the feasibility of GWNoC, we present in Section 3 the results of an on-going study that (a) compares the area, power, and performance scaling trends of different NoCs, (b) investigates the scalability of multicast traffic in different architectures, and that (c) will assess the impact of improving the multicast support upon the performance of current and future architectures. This paper combines our previous work in [11, 16] with further results from new network performance and traffic scalability analyses, and puts them in the context of this vertical feasibility study. The final aim and expected contribution of this study is to prove that, by means of its unique area and broadcast capabilities, GWNoC will have a profound impact upon the design of manycore architectures. Section 4 concludes the paper.

2. GRAPHENE-ENABLED WIRELESS NETWORK-ON-CHIP (GWNOC)

The wireless NoC paradigm has been recently proposed in a variety of forms to complement wireline NoCs [10]. The most extended approach consists in the CMOS-compatible

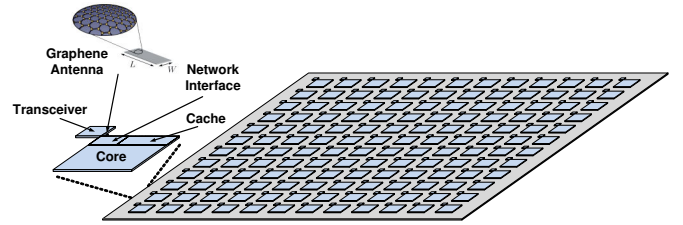


Figure 2: Schematic Diagram of a 144-core Graphene-enabled Wireless Network-on-Chip.

integration of antennas in particular chip locations, seeking to improve power and latency by reducing the number of hops between any pair of cores [17, 18]. Signals are radiated by the transmitting antenna following a radiation pattern, propagate through the medium being reflected at the chip package until reaching the receiving antennas. Given its shared medium nature and the lack of need for wiring between transmitters and receivers, the adoption of wireless NoC also adds reconfigurability and native multicast capabilities to the network, opening the door to a large number of possibilities at run-time.

The only downturn of the current wireless NoC designs resides in the size of the antenna, which forces proposals to either focus on moderately-sized processors or to use the wireless plane only to communicate clusters of cores [19, 17, 18]. Indeed, the size of future metallic on-chip antennas, i.e., hundreds of micrometers [19], renders unfeasible the approach of integrating one antenna per core, as the core sizes continue to shrink with each CMOS technology generation and reach sizes of a few hundreds of micrometers. Such issue cannot be solved by further reducing the size of a metallic antenna as (1) the performance of metallic antennas of a few micrometers is poor due to their low conductivity, and (2) this would impose the use of frequencies of a few hundreds of THz, which is not suitable for RF wireless communications due to attenuation and transceiver design issues.

The integration of at least one antenna per core is essential to fully exploit the potential benefits of wireless NoC, yet it is not possible with conventional technologies. This could be enabled by graphene instead, as its unique properties allow the creation of antennas with lateral dimensions of just a few micrometers that resonate in the terahertz band [13]. The main reason behind the exhibited subwavelength behavior, i.e. the size of the antenna is lower than the wavelength at which it resonates, is the presence of *surface plasmon polariton* (SPP) waves on the surface of the radiating structure [20]. These waves result from the coupling between an incident EM wave and surface electric charges at the interface between a dielectric and a metal, and their properties are determined by the characteristics of the metal (graphene in this case). For instance, the resonant behavior in the terahertz band is given by the particular dispersion of SPP waves in graphene [21]. A plethora of on-going works are focused on assessing the properties of these novel antennas and have predicted a similar performance than that of metallic antennas [13, 22, 15, 14, 23].

The integration and use of graphene-based antennas within a chip multiprocessor gives birth to the concept of GWNoC, as shown in Figure 2. Each computing core can be considered a *wireless core*, as it contains a graphene antenna and a

transceiver that prepares the information for outgoing transmissions and demodulates incoming transmissions. Within each network interface, a controller will be included capable of deciding whether a transmission should go through the wireless plane or not. GWNoC not only maintains the advantages of a wireless NoC, but also provides higher flexibility and broadcast support at the core level [12]. The terahertz band, on its turn, offers very high bandwidths that could be used to seek ultra-high data rates and low power schemes [24].

3. A MULTIFACETED FEASIBILITY STUDY

As any on-chip network, GWNoC must satisfy the traffic requirements cast by the architecture with a given performance, while being subject to a set of implementation constraints. The main aim of our feasibility study is to position GWNoC within the space formed by this combination of requirements, performance and implementation constraints in order to confirm its suitability in the manycore scenario. To this end, we inspect how different aspects of the system scale with respect to the number of cores N and the communication capacity of each core C .

3.1 Implementation: Area and Power

Chip area and energy are scarce resources and represent the main constraint in the manycore scenario. Therefore, we first analyze and compare how these metrics scale considering electrical, optical, and wireless interconnects. In the first case, ORION is used to model the area and power of a 2D mesh of on-chip links and routers [25]. In the second case, three shared-waveguide topologies based on ring resonators are evaluated using area and insertion loss figures from the literature. In the third case, the analysis of the state of the art in wireless transceivers (including analog and digital stages but neglecting the MAC overhead) from 8 to 820 GHz revealed that area and power are inversely proportional to the radiation frequency [11]. Provided that a given on-chip network may scale remarkably well in terms of area and perform poorly in terms of energy, or *vice versa*, we jointly evaluate both metrics by using the following figure of merit:

$$FoM = \frac{1}{A \cdot E_{bit}} \text{ [bits/J/mm}^2\text{]}, \quad (1)$$

where A is the total area and E_{bit} is the energy per bit. Such performance metric can be understood as the average number of bits that can be effectively transmitted for each consumed joule of energy and square millimeter of chip real estate. The interested reader can refer to [11] for more details on both the analyzed architectures and the evaluation methods.

Top plot in Figure 3 shows how the figure of merit scales as a function of the number of cores. Nanophotonic options scale worse than the rest of technologies due to laser power issues and the increase of area required to scale the evaluated network architectures. In contrast, electrical and wireless NoCs show a similar trend, with the former yielding the best absolute figures. In the wireless case, three different operation frequencies were chosen pointing towards the terahertz band. It is observed that the use of high frequencies is beneficial since it implies lower area and energy.

Bottom plot in Figure 3 shows how the figure of merit scales as a function of the communication capacity of each

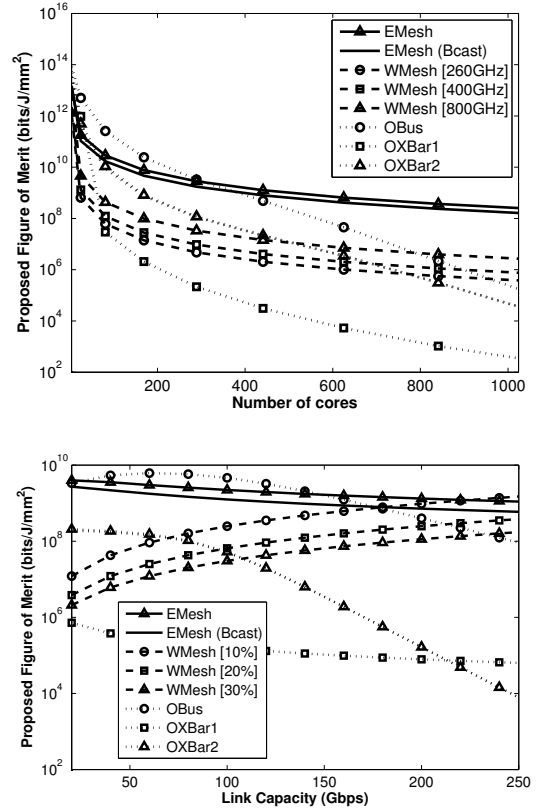


Figure 3: Proposed figure of merit, Eq. (1), as a function of both the number of cores assuming a core capacity of 80 Gbps (top), and of the core capacity assuming 256 cores (bottom). Higher is better.

core. In this scenario, a wireless NoC improves its performance as the communication requirements are pushed up. This is because higher throughput requirements imply the use of higher frequency technologies which, as mentioned above, are expected entail lower area and energy. We assumed three different wireless cases, corresponding to considering, for simplicity, that the data rate is a 10%, 20%, or 30% of the carrier frequency [11]. In the rest of options, performance decreases with the core capacity due to a sustained growth in the number of components required to scale the network architecture, as well as in the power associated to this additional circuitry.

Combined, the results shown above reveal that the wireless option may end up outperforming the rest of analyzed approaches in terms of area and energy efficiency when considering a scenario with both a very large number of cores and very high capacity requirements. These conditions will be only met with GWNoC, as it enables the integration of one antenna per core in manycore settings and offers enough bandwidth to reach such high data rates. This will be possible if the device and transceiver design challenges that are associated to the terahertz band are overcome [24, 11].

3.2 Network: Latency and Throughput

The latency and throughput that the NoC offers to the multiprocessor are critical since they have a large impact

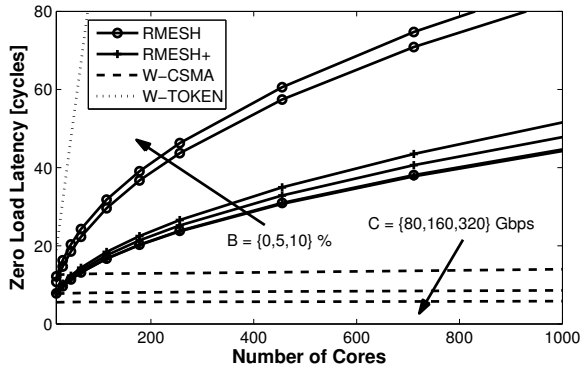


Figure 4: Zero-load latency as a function of the system size N , wireless capacity C and broadcast percentage B .

upon its execution speed. The latency quantifies the time required for a message to reach the intended destinations, whereas the throughput is an indicator of the number of messages that can travel through the NoC over time. Both are interrelated and depend on the injection load. For simplicity, we use the zero-load latency and the throughput for a given latency limit (150 cycles) as metrics of performance. The former assumes no contention and models the latency for low loads reasonably well, while the latter measures performance for higher loads. We set a common latency limit in the throughput metric seeking to obtain a fair comparison between NoCs (different designs may yield different saturation loads). The throughput is measured from the transmitter perspective, this is, a multicast packet will be taken into account one regardless of the number of destinations.

In our exploration, we use PhoenixSim [26] to simulate three representative NoC scenarios:

RMESH(+) A 2D mesh based on electrical interconnects, assuming an aggressive design where switch and link traversals take one clock cycle each [6]. Multicast support is tree-based: flits are replicated at intermediate routers and form a fixed virtual tree. A delay of one cycle per port and per router is assumed in RMESH and RMESH+, respectively.

W-CSMA A single channel shared among all cores and arbitrated through a non-persistent MAC protocol based on collision detection. Since correct transmissions will likely be more frequent than collisions, the protocol adopts a negative acknowledgement (NACK) strategy to reduce the control overhead. In the event of a collision, a burst of NACKs is sent through the same channel than data: the source will then assume that its transmission resulted in a collision and will schedule a retry. Otherwise, the source considers that the transmission is successful after a round-trip delay.

W-TOKEN A single channel shared among all cores and arbitrated through a token ring scheme, where only the core that possesses the token is able to transmit [10]. We assume that the token passing is performed through a dedicated channel and that it takes one clock cycle between two consecutive cores.

Two-cycle overheads that model processor-to-router communication delay in RMESH and modulation/demodulation

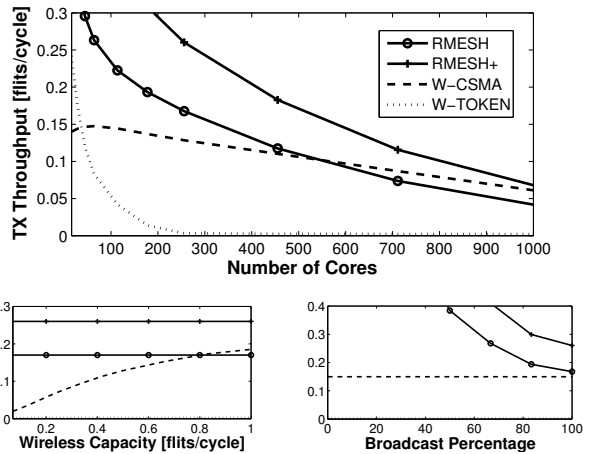


Figure 5: Network throughput assuming a latency of 150 cycles as a function of the system size (top), the capacity of the wireless plane (left) and the percentage of broadcast traffic (right). Default values are $C = 320 Gbps$ (half flit per cycle), $N = 256$ and $B = 100\%$.

delay in W-CSMA and W-TOKEN are included.

Each network architecture is stressed with traffic generated by means of a memoryless Poisson process and evenly distributed over all cores. To evaluate the broadcast percentage, a certain percentage of the generated packets B are tagged as broadcast. Note, however, that such distinction only affects the performance of RMESH since the rest of options treat all messages as broadcast. Clock frequency is set to 5 GHz and packet length is fixed to 128 bits including headers. We avoided using traffic traces in order to enable not only the extension of the analysis to thousands of cores, but also the use of the broadcast percentage as a parameter. More realistic traffic will be employed in future work.

Figure 4 shows the zero-load latency of the three network architectures considering different broadcast percentages B (which affect RMESH) and wireless capacities C (which affect W-CSMA and W-TOKEN). It is observed that the worst results are that of W-TOKEN, where the latency is dominated by the token passing and scales as $O(N/2)$. Similarly, the latency of the RMESH scales as $O(\sqrt{N})$ which corresponds to the average hop distance in a mesh topology. Note that the presence of broadcast traffic also has a negative impact upon the latency due to the serialization delay incurred in the flit replication process. Finally, W-CSMA shows a zero-load latency of a few cycles, invariant with respect to the number of cores since the die size and the wireless data rate remain constant.

Figure 5 shows the throughput of the three network architectures when the latency is 150 cycles. Top plot shows diminishing performance in all architectures as the system size increases, yet it seems that W-CSMA is able to restrain the performance drop. Left plot illustrates how the throughput can be improved in W-CSMA by increasing the wireless channel capacity, to the point of outperforming RMESH, but not RMESH+. As shown in the right plot, this is would be only possible when the percentage of broadcast reaches very high levels. Finally, note that W-TOKEN is dominated by the token passing delay and is not able to compete with

the other options beyond a few tens of cores.

Perhaps due to its on-demand nature, the results above reveal that W-CSMA may scale well enough to outperform conventional NoCs in the presence of multicast traffic as the system scales. This is, however, provided that either the capacity of the wireless channel is large enough, or the efficiency of the employed MAC mechanism is significantly improved. The simple protocol evaluated here yields a throughput below 15% of the wireless capacity; the challenge here is to devise a MAC protocol that systematically exploits the considerable cross-layer design opportunities of the scenario, e.g. prediction, to reach unprecedented throughput levels and be able to outperform advanced multicast schemes.

3.3 Architecture: Multicast Traffic

In order to accurately evaluate the performance of any NoC, it is of high value to have a deep understanding on the traffic requirements of commonly used applications running over the target architectures. At present, these can be obtained with full-system cycle-accurate simulators. However, given the lack of a well-established platform for the simulation of manycore processors, it is complex to capture the traffic requirements in systems with hundreds of cores and above. In light of this, we use GEM5 [27] to simulate moderately-sized processors and then extract scaling trends for different traffic aspects. GEM5 currently allows to simulate up to 64 cores and admits a variety of architectures. We started by running SPLASH-2 and PARSEC benchmarks over a tiled mesh, where each tile includes one processing core, a network interface, 32-KB 2-way L1 data and instruction caches, one bank of 512-KB 8-way L2 cache (64-B cache line size, strict inclusion). For the sake of brevity, we only consider a multicast-intensive coherence scheme: HyperTransport (HT). Note, though, that the methodology can be made extensive to alternative coherency protocols, benchmarks and architectures.

We made slight modifications to GEM5 in order capture statistics on multicast traffic. The left plot in Figure 6 reveals a sustained increase in terms of the multicast intensity of HT in bits per instruction. This is a NoC-agnostic measure of the multicast intensity as it solely depends upon the interaction of the multiprocessor architecture (it defines the methods that generate these messages) with the application (it determines the sharing structures and memory intensity). From these multicast intensity values, it is possible to infer the throughput that a dedicated multicast plane must support by assuming a target execution speed.

Right plots in Figure 6 represent the percentage of injected (top) and ejected (bottom) traffic that is multicast in conventional NoCs. It is observed that, in HT, the ratio of injected multicast traffic decreases with the number of cores mainly due to the increase of unicast acknowledgments. However, the ratio of ejected traffic consistently grows with the system size since flits need to be replicated an increasing number of times. This causes 1.5% of the transactions to be accountable for almost half of the served traffic in a 64-core system, effect that would be avoided by eliminating the need for explicit flit replication. Shared-medium alternatives like GWNOC offer such possibility and would therefore imply huge savings to the wired plane in multicast-intensive coherence if used as a dedicated multicast plane.

The employed methodology also enables the analysis of the spatial and temporal distributions of the multicast traf-

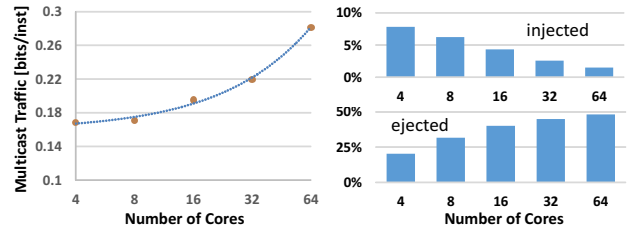


Figure 6: Left plot: multicast requirements of HT in bits per instruction averaged over all PARSEC and SPLASH-2 applications. Right plots: percentage of injected and ejected flits that are due to multicast transactions.

fic in an application-dependent manner [16]. The former is useful to inspect whether hotspots are to be expected, whereas the latter evaluates the burstiness of the traffic. A traffic model containing such information will be of special importance both for the design of specific MAC protocol for GWNOC and for the evaluation of NoC proposals.

3.4 Future Work

The feasibility study here presented is an on-going effort and we expect to continue it in future work by assessing, first, the impact of having an effective broadcast platform upon the performance of a set of selected architectures. To this end, we will add a broadcast plane with a given latency-throughput characteristic upon a conventional NoC. We will then evaluate, through full-system simulation, the speedup resulting from this addition as a function of the multiprocessor architecture and the assumed broadcast performance. Note that these two variables are directly related to the number of cores and core capacity, respectively, which have been consistently used throughout the feasibility study. To the best of the authors' knowledge, such analysis has been thus far performed only for a fixed number of cores: in [6], an average speedup of more than 10% was obtained by assuming an ideal multicast scheme in a 64-core system with HT coherence. The real challenge here, though, is to co-design novel architectural methods that fully take advantage of the benefits of GWNOC in the pathway to scalable architectures.

4. CONCLUSIONS

Owing to unique plasmonic properties, micrometer antennas based on graphene radiate in the terahertz band. These features allow envisaging the concept of GWNOC which, by integrating one antenna per core, would provide efficient chip-scale broadcast capabilities that could be extremely beneficial for the design of manycore architectures. If current technological trends continue in the future, GWNOC will be a viable approach in manycore processors as its area and power have been demonstrated to scale well with the number of cores and bandwidth requirements. It has been also shown that GWNOC offers a huge potential for low-latency broadcast communication with good, albeit still improvable, scalability in terms of throughput. Since the broadcast traffic requirements are expected to increase with the number of cores in a wide variety of architectures, we believe that the introduction of GWNOC is expected to have a profound impact in future multiprocessors.

Acknowledgments

This work has been partially funded by SAMSUNG, INTEL, the Catalan Government (Ref. 2014SGR-1427) and through a FI-AGAUR grant. The authors would like to thank Mario Nemirovsky, Max Lemme, Raúl Martínez and Ignacio Llatser for their invaluable research discussions.

5. REFERENCES

- [1] T. Bjerregaard and S. Mahadevan, "A survey of research and practices of Network-on-chip," *ACM Computing Surveys*, vol. 38, no. 1, pp. 1–51, Jun. 2006.
- [2] D. A. B. Miller, "Device Requirements for Optical Interconnects to Silicon Chips," *Proceedings of the IEEE*, vol. 97, no. 7, pp. 1166–85, 2009.
- [3] D. Culler, J. P. Singh, and A. Gupta, *Parallel computer architecture: a hardware/software approach*. Morgan Kaufman, 1999.
- [4] T. Krishna, C. Chen, W. Kwon, and L. Peh, "Smart: Single-Cycle Multihop Traversals over a Shared Network on Chip," *IEEE Micro*, vol. 34, no. 3, pp. 43–56, 2014.
- [5] G. Nychis, C. Fallin, and T. Moscibroda, "On-chip networks from a networking perspective: congestion and scalability in many-core interconnects," in *Proceedings of the SIGCOMM*, 2012, pp. 407–18.
- [6] T. Krishna, L. Peh, B. Beckmann, and S. K. Reinhardt, "Towards the ideal on-chip fabric for 1-to-many and many-to-1 communication," in *Proceedings of the MICRO-44*, vol. 2, 2011, pp. 71–82.
- [7] N. E. Jerger, L.-S. Peh, and M. Lipasti, "Virtual Circuit Tree Multicasting: A Case for On-Chip Hardware Multicast Support," in *Proceedings of the ISCA-35*. Ieee, Jun. 2008, pp. 229–240.
- [8] A. Ros, M. E. Acacio, and J. M. García, "A Direct Coherence Protocol for Many-Core Chip Multiprocessors," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 12, pp. 1779–92, 2010.
- [9] R. G. Beausoleil, P. J. Kuekes, G. S. Snider, S.-y. Wang, and R. S. Williams, "Nanoelectronic and Nanophotonic Interconnect," *Proceedings of the IEEE*, vol. 96, no. 2, pp. 230–247, Feb. 2008.
- [10] S. Deb, A. Ganguly, P. P. Pande, B. Belzer, and D. Heo, "Wireless NoC as Interconnection Backbone for Multicore Chips: Promises and Challenges," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 2, pp. 228–239, 2012.
- [11] S. Abadal, M. Iannazzo, M. Nemirovsky, A. Cabellos-aparicio, and E. Alarcón, "On the Area and Energy Scalability of Wireless Network-on-Chip: A Model-based Benchmarked Design Space Exploration," *IEEE /ACM Transactions on Networking*, vol. PP, no. 99, p. 1, 2014.
- [12] S. Abadal, E. Alarcón, M. C. Lemme, M. Nemirovsky, and A. Cabellos-Aparicio, "Graphene-enabled Wireless Communication for Massive Multicore Architectures," *IEEE Communications Magazine*, vol. 51, no. 11, pp. 137–143, 2013.
- [13] I. Llatser, C. Kremers, A. Cabellos-Aparicio, J. M. Jornet, E. Alarcón, and D. N. Chigrin, "Graphene-based nano-patch antenna for terahertz radiation," *Photonics and Nanostructures - Fundamentals and Applications*, vol. 10, no. 4, pp. 353–358, 2012.
- [14] M. Tamagnone, J. S. Goñamez-Diñáñez, J. R. Mosig, and J. Perruisseau-Carrier, "Analysis and design of terahertz antennas based on plasmonic resonant graphene sheets," *Journal of Applied Physics*, vol. 112, p. 114915, 2012.
- [15] J. M. Jornet and I. F. Akyildiz, "Graphene-based Plasmonic Nano-Antenna for Terahertz Band Communication in Nanonetworks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 685–694, Dec. 2013.
- [16] S. Abadal, R. Martínez, E. Alarcón, and A. Cabellos-Aparicio, "Scalability-Oriented Multicast Traffic Characterization," in *Proceedings of NoCS '14*, 2014, pp. 180–181.
- [17] A. Ganguly, K. Chang, S. Deb, P. P. Pande, B. Belzer, and C. Teuscher, "Scalable Hybrid Wireless Network-on-Chip Architectures for Multi-Core Systems," *IEEE Transactions on Computers*, vol. 60, no. 10, pp. 1485–1502, 2010.
- [18] D. Matolak, A. Kodi, S. Kaya, D. DiTomaso, S. Laha, and W. Rayess, "Wireless networks-on-chips: architecture, wireless channel, and devices," *IEEE Wireless Communications*, vol. 19, no. 5, 2012.
- [19] S.-B. Lee, S.-W. Tam, I. Pefkianakis, S. Lu, M.-C. F. Chang, C. Guo, G. Reinman, C. Peng, M. Naik, L. Zhang, and J. Cong, "A scalable micro wireless interconnect structure for CMPs," in *Proceedings of the MOBICOM '09*, 2009, p. 217.
- [20] A. Vakil and N. Engheta, "Transformation optics using graphene," *Science*, vol. 332, no. 6035, pp. 1291–4, 2011.
- [21] M. Jablan, H. Buljan, and M. Soljačić, "Plasmonics in graphene at infrared frequencies," *Physical review B*, vol. 80, no. 24, p. 245435, 2009.
- [22] I. Llatser, C. Kremers, D. Chigrin, J. M. Jornet, M. C. Lemme, A. Cabellos-Aparicio, and E. Alarcón, "Radiation Characteristics of Tunable Graphennas in the Terahertz Band," *Radioengineering Journal*, vol. 21, no. 4, 2012.
- [23] Y. Huang, N. Khiabani, Y. Shen, and D. Li, "Terahertz photoconductive antenna efficiency," in *Proceedings of the iWAT '11*, 2011, pp. 152–56.
- [24] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, 2014.
- [25] A. Kahng, B. Li, L. 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 DATE '09*, 2009, pp. 423–8.
- [26] 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 DATE '10*, 2010, pp. 691–696.
- [27] N. Binkert, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, D. a. Wood, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, and T. Krishna, "The gem5 simulator," *ACM SIGARCH Computer Architecture News*, vol. 39, no. 2, 2011.