Fundamentals of Graphene-Enabled Wireless On-Chip Networking

Sergi Abadal, Ignacio Llatser, Albert Mestres, Josep Solé-Pareta, Eduard Alarcón and Albert Cabellos-Aparicio

Abstract In the broad sense of the term, *nanonetworks* may refer not just to networks composed of nanosized devices, but also to communication networks enabled by nanotechnology. Nanoscale communication techniques can be suitable to interconnect elements far larger than a few square micrometers in applications subject to strong size constraints or bandwidth requirements. Here, the concept *Graphene-enabled Wireless Network-on-Chip* (GWNoC) is introduced as a clear example of this category. In GWNoC, graphene plasmonic antennas are used to wirelessly communicate the components of a multicore processor, which are located in the same chip. This shared medium approach is opposed to current chip communication trends and aims to reduce many of the issues that hamper the development of scalable multiprocessor architectures. In this chapter, we describe the scenario and the communication requirements that justify the employment of nanonetworking techniques, as well as the main challenges that still need to be overcome in this new research avenue.

1 Introduction

Parallelization has been the natural trend in microprocessor architecture design for the last decades and it is expected to continue in the near future. Parallelism can be found and exploited at different granularities, being the instruction-level parallelism the traditional approach which takes advantage of the potential overlap among simple instructions. However, fundamental limits at this level rapidly caused diminishing
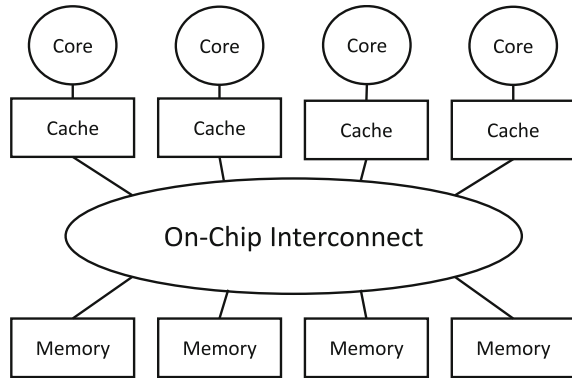
S. Abadal (✉) · I. Llatser · A. Mestres · J. Solé-Pareta · E. Alarcón · A. Cabellos-Aparicio
NaNoNetworking Center in Catalunya (N3Cat), Universitat Politècnica de Catalunya,
Barcelona, Spain
e-mail: abadal@ac.upc.edu

293

Fig. 1 Schematic diagram
of a shared-memory
multiprocessor

returns in its exploitation and finally caused power consumption in uniprocessors to
grow way faster than its performance [27].

Alternatively, single-chip multiprocessor architectures have emerged aiming to
keep pace with the performance trends predicted by Moore's Law, while maintain-
ing an acceptable energy footprint. Parallelism is exploited either by simultaneously
executing instances of independent applications or by dividing the application in a set
of tasks and processing them in a collaborative manner. To do so, multiprocessors
or multicore processors consist of the interconnection of a given number of inde-
pendent processor *cores* and a memory system within a single die. Figure 1 shows a
generic scheme of a shared-memory multicore processor, wherein the memory sys-
tem is generally hierarchical with some levels of the hierarchy being shared by all
the processors. The shared memory paradigm is widely used in current multicore
processors and will be the architecture assumed throughout the chapter.

The on-chip interconnect is a central element of a multicore processor since it
implements the communication between cores and memory and has a large impact
on performance. In shared memory schemes, communication between cores actually
occurs implicitly as a result of conventional memory access instructions. Coopera-
tion and coordination among threads is accomplished by reading and writing shared
variables [17]. The presence of caches within the memory system decreases the aver-
age latency of such memory accesses but, at the same time, it also introduces the
problem of cache coherence. Multiple copies of the same shared data may be dis-
tributed in a plurality of caches, so that different cores may be seeing different val-
ues if this data is modified. Cache coherence protocols are designed to enforce that
a read to a shared variable returns the last written value at the expense of generat-
ing extra communication. Other issues such as data consistency or synchronization
among threads are equally critical for the operation of a multicore processor, as well
as additional sources of traffic that the on-chip interconnect must deal with [17, 27].

Given the direct relation between memory architecture, communication and over-
all performance, the research focus in multiprocessors has gradually shifted from
how cores compute to how cores communicate. Buses were first widely considered
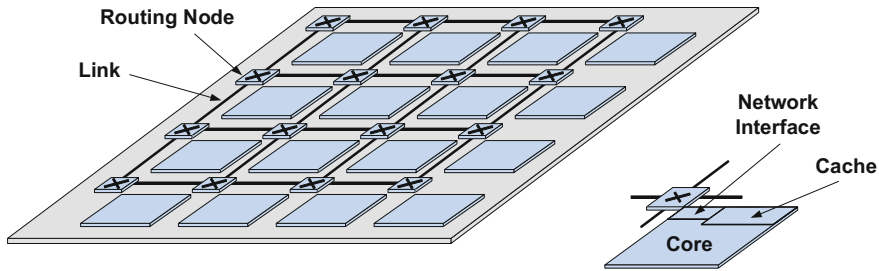
Fig. 2 Schematic diagram of a conventional network-on-chip (NoC)

for the implementation of the on-chip interconnect, but their use is restricted to small-scale architectures given their limited scalability beyond a few cores [9]. Instead, Network-on-Chip (NoC) has been widely adopted as the paradigm of choice for on-chip interconnection networks. NoC can be defined as the application of networking theory and methods to on-chip communication and it generally consists in the employment of point-to-point packet-switched schemes. Figure 2 represents a simple example of NoC, where a given number of on-chip Resistive-Capacitive (RC) wires interconnect the cores (and caches) by means of their respective network interfaces and passing through a network of routing nodes. The interconnection to main memory and the I/O system is omitted for simplicity. Such designs offer improvements in fault tolerance, in modularity and, most importantly, in the overall scalability of the interconnection network; still, it remains unclear whether NoCs based on RC wires will be able to meet the increasingly stringent requirements of next-generation multiprocessors. There are numerous reasons, the most important being the expected increase in delay and power consumption of the wires [55].

As we approach the manycore era, where chips will integrate thousands of cores, several challenges need to be addressed in order to prevent communication to become the performance bottleneck of multicore processors. On-chip interconnects must provide higher throughput levels while maintaining a low latency on a chip-wide basis, taking into consideration that the area and power of the solution must remain bounded (see Sect. 2 for more details). Given that on-chip wires will not be able to cope with such combination of demands, considerable research efforts are devoted to extending the original NoC paradigm with interconnect technologies yielding improved performance. Four emerging alternatives, namely, 3D stacking, Radio Frequency (RF) interconnects, wireless on-chip communication and nanophotonic communication, appear as serious contenders for this regard and are briefly presented next. These will provide both important improvements at the physical layer due to their higher bandwidth per area densities or energy efficiencies, as well as additional degrees of freedom for the design of scalable network architectures.

First, three-dimensional stacking consists in the superposition of different layers of active devices. These layers are separated by just a few tens of micrometers and are vertically interconnected by means through-silicon vias [12] or near-field coupling schemes [15]. The creation of 3D integrated circuits has proved to be a

promising paradigm, since it has shown to imply significant benefits such as higher packing density, improved noise immunity, and overall superior performance. From a NoC perspective, 3D stacking reduces the average propagation delay and energy per bit due to the short distance between layers and enables the use of topologies not considered in the 2D design space [21]. However, it is important to note that 3D stacking presents considerable challenges. For instance, the superposition of active layers produces an increase in the heat density that must be circumvented in order to avoid thermal effects. Also, refined techniques are needed for the manufacturing and integration of such tridimensional integrated circuits and networks, notably alignment methodologies for accurate placement of the vertical vias.

Second, the RF interconnection paradigm is presented as an alternative to traditional voltage and current signaling through metallic wires. Baseband signals are modulated using gigahertz carriers and then sent at the speed of light through transmission lines printed in the chip surface [55]. In long-range links, the improvement in terms of propagation time is very large with respect to the delay introduced by the modulation process and, therefore, the communication latency can be effectively reduced. RF interconnects also enable multiple access schemes in shared transmission lines, e.g. by means of frequency-multiplexing or code-multiplexing schemes. Each core is assigned a set of channels, enabling the possibility of interconnecting several cores using the same transmission line and therefore reducing the number of wires. Further, the bandwidth for each core could be dynamically assigned according to real-time demands. Due to these advantages, complementing a baseline NoC with an overlaid global RF interconnect has been proposed [16]. The main downturn is that the physical topology must be carefully designed as impedance mismatch reflections at the transmission line ends may generate interferences, limiting the number of practical network architectures and their scalability.

A possible solution to the RF-interconnect issues is to transmit the signals wirelessly instead of through transmission lines. The resulting Wireless Network-on-Chip (WNoC) approach not only inherits the advantages of RF-interconnects, but also adds natural adaptability and broadcast capabilities to the equation as no path infrastructure is needed. WNoC is feasible due to the availability of both on-chip antennas [46] and high-speed transceivers [23], and has recently given rise to a plethora of proposals (see [18] and references therein). However, as we will see in the following sections, the size of current and future metallic on-chip antennas largely limits the potential of this approach and motivates the employment of nanoscale communication techniques.

Last but not least, nanophotonics is enabling the creation of CMOS-compatible optical building blocks for, among others, on-chip communications [8]. Chip-scale transmissions at speeds of 50 Gbps have been accomplished thus far [45], whereas potential for energy figures several orders of magnitude lower than conventional interconnects is envisaged [13]. In light of the promise that this technology shows for low energy per bit communications, intense research efforts have been directed towards creating *photonic NoCs* by means of the integration of nanophotonic devices. Apart from yielding an outstanding potential for low power consumption, such networks also maintain the main advantages of RF interconnects as signals can be

wavelength-multiplexed. Such feature provides both potential for extremely high bandwidth per area, as well as a wide range of possibilities from the network architecture perspective. Extensive works in the design and development of photonic NoCs, including a wide variety of topologies and network architectures, serve as proof of this trend (see [3, 36, 54, 62] and references therein). It must be noted that these works, in most cases, aim to overcome the main limitations of the nanophotonic approach. Existing on-chip laser sources are excessively large in terms of area or involve costly integration processes; whereas the implementation of all-optical packet routing schemes remains as a grand challenge at the chip level.

Even though considerable advances have been accomplished in the field of on-chip networking, efficiently delivering multicast and broadcast traffic remains an open challenge at the time of writing this book chapter. The case is particularly concerning within manycore settings, where one-to-many communications will play a crucial role (see Sect. 2 for more details), and even considering the new interconnect technologies mentioned above. While one may think that the advent of WNoC would solve this issue given the inherent broadcast capabilities of this technology, the reality is that the size of metallic antennas prevents the integration of one antenna per core to fully take advantage of such competitive advantage.

In this chapter, we present the concept *Graphene-enabled Wireless Network-on-Chip* (GWNoC), which aims to address this grand challenge by providing each core with wireless communication capabilities and sharing the medium [2]. The approach is enabled by graphene antennas, whose plasmonic effects allow them to radiate electromagnetic waves in the terahertz band (0.1–10 THz) while occupying an area up to two orders of magnitude lower than metallic antennas for the same radiation frequency [40, 59]. This way, the stringent requirements of the scenario in terms of area and bandwidth, which are detailed in Sect. 2, can be met. Section 3 presents a description of GWNoC and its advantages over emerging alternatives, as well as an outline of the main communications and networking design considerations. Section 4 concludes the chapter.

2 Open Issues in Communication Within Manycore Chip Multiprocessors

Taking into consideration several physical constraints such as the thermal design power, technology improvements are foreseen to steadily provide a scaling of at least 1.4X, per technology generation, of the number of cores that can be integrated within a chip [30]. However, the entire system must scale before this trend translates into effective parallel performance improvement. This implies solving the open issues that are found when scaling aspects such as parallel programming models, the memory system or the on-chip interconnect fabric. In this chapter, we focus upon the on-chip interconnect while being aware of the memory system, since the performance of a multicore processor is largely dictated by how fast both memory accesses and the traffic generated by these accesses are served.

From an architectural perspective, a balance must be struck between the effectiveness of the memory system and the communication requirements cast upon the on-chip interconnect. However, this task becomes especially challenging as the number of cores per chip grows, since the communication demands of existing architectures exponentially increase when upscaled. In this regard, unconventional and less communication-intensive architectures need to be explored. From a communications perspective, the main objective is to match the performance of the on-chip interconnect with the potential communication demands placed by the architecture, while complying with some design constraints. For instance, it has been widely proved that chip communication mainly occurs among neighboring cores due to the spatial locality of code [27, 56] and initial NoC designs were better suited for this type of traffic [49]. However, the complexity of this matching process grows with the number of cores, as the architectural aspects that impact upon the characteristics of the on-chip traffic may change significantly. New interconnect solutions will be therefore required.

In the following, we detail the physical constraints and driving requirements that challenge the design of on-chip interconnects for manycore settings, putting special attention upon the case of multicast and broadcast on-chip communication.

Power Consumption

Thermal effects are a primary concern when designing a processor. In order to hold down the costs of thermal cooling, manufacturers generally impose constant power limits across generations. The energy efficiency of the on-chip interconnect will need to be improved, since the communication demands are foreseen to sharply increase with the number of cores. Projections derived from the International Technology Roadmap for Semiconductors (ITRS) calculate that transmitter energies of between 10 to 100 fJ/bit must be targeted in the near future [44].

Efficiencies around 200 fJ/bit have been demonstrated using conventional interconnects [52]. Even though these figures can be still improved, it remains unclear whether it is possible to meet future energy requirements without largely affecting other metrics such as the data rate. This situation has been the main driving force behind proposing nanophotonics for on-chip communication, as it promises to push figures down to around 1 fJ/bit [13].

Power consumption is also a concern at the network level as the core density and the complexity of NoCs increase. The average number of hops of widely-used mesh topologies increases with the number of cores, incurring in a proportional increase in power as routers consume a significant fraction of energy. One of the first implementations of NoCs for manycore chip multiprocessors is described in [29], where the NoC consumes approximately 40% of the total 100 W chip power. This suggests that alternative topologies (perhaps enabled by 3D stacking) may be needed in manycore settings to reduce the number of hops and, therefore, the average power consumption [21, 49]. However, the energy savings are generally traded off with area as these topologies require additional wires and more complex routers. In the case of photonic NoCs, all-optical alternatives at the network level are reduced and do not scale due to current laser and router complexity limitations [3]. Designs combining electrical

and optical planes offer higher degrees of freedom and have recently been considered instead [36, 54]. Tools are available for the evaluation of their energy efficiency [14, 58].

Area

In order to ensure a growing yield, manufacturers aim to keep the die size as small as possible. Processor dies are currently on the order of a few hundreds of square millimeters and grow slower than the area occupied by cores for each technology generation [30]. Therefore, the area overhead of the on-chip interconnect is a critical evaluation factor as chip real estate becomes an extremely scarce resource in many-core environments. For instance, the high bandwidth per area figures of nanophotonic interconnects is one of the reasons for considering them among the plethora of emerging contenders. Technological models are broadly used for the early-stage evaluation of the area in conventional and photonic NoCs [33, 58].

Closely related to the area constraint is the wiring complexity. NoCs based on wires or waveguides may need to include a large number of links to implement a complex topology suitable to the demands of manycore chip multiprocessors. Regardless of whether the area limitations are respected or not, finding an appropriate layout strategy may be unfeasible due to the increasing wire routing congestion. A potential solution would be to replace part of the wiring with wireless RF interconnects [18]. However, the size of on-chip antennas limits the usefulness of the WNoC approach and motivates the use of graphene-enabled wireless communications.

Performance

The multicore scenario imposes a set of general requirements to the on-chip interconnects. Cores generally send messages after a given computation and stop their execution until a response is received. A slow or lossy delivery of these messages must be avoided, as it will cause the cores to reduce their speed and therefore to reduce general performance. Hence, any on-chip interconnect must guarantee a given performance in terms of latency, throughput, and losses. The main challenge here is to provide solutions that will allow us to maintain these conditions as the number of cores grows.

Latency is arguably the most important constraint in on-chip networks despite the strong requirements in area, power, and bandwidth. The communication delay in operations that are in the critical path of the processor will directly impact upon its performance. Therefore, latency must be kept within certain bounds (ideally constant) when scaling the number of nodes. This is not possible in conventional mesh designs due to the increase of the average hop count. Again, alternative topologies or the use of RF/nanophotonic long-range links has been proposed to improve the overall latency [36, 38, 55, 62].

Secondly, multicore processors are extremely data intensive scenarios and, as a result, the bandwidth of the interconnect is also crucial. A rule of thumb is that its throughput should scale at least proportionally with the number of cores. Conventional NoCs meet such scalability demands, but optimizations are still needed as chip resources become more scarce. Overprovisioning is generally employed in order to

avoid the network to saturate in high contention phases, which are typical in parallel programs and generate large bursts of communication. Fine-grained reconfigurable links have been proposed in order to save area and power wasted in such process [28]. Nanophotonics are also taken into consideration as they yield much improved bandwidth per area figures.

Finally, all packets need to be delivered free of errors in order to guarantee a correct operation of the processor. At the link level, on-chip interconnects are designed to operate with a bit error rate (BER) around 10^{-15} [38] and generally apply Forward Error Correction (FEC) schemes to correct infrequent errors. At the network level, congestion may cause packets to be discarded due to network buffers being overrun, motivating the need for flow control mechanisms and retransmission policies.

Multicast/Broadcast

Area, power, and performance are general requirements that apply to all the traffic generated by the memory system, regardless of its characteristics. As the core density grows, the general tendency is to scale current multicore architectures and then to address the resulting increase in communication by means of the improved on-chip networks. However, it occurs that the traffic may not necessarily scale in the same direction than the interconnect performance does. One clear example is the communication between topologically distant cores: whereas the number of these transmissions increases with the core density, the performance of conventional NoCs worsens in this situation. Although a possible strategy is to design a memory system or a programming model aiming to reduce long-range communication, this implies facing additional challenges that are out of the scope of this chapter.

Within this context, a particularly concerning case is that of multicast and broadcast communication. From a computer architecture standpoint, broadcast communications have been traditionally regarded as expensive and its use is avoided whenever possible. However, operations such as thread communication or data synchronization generate a significant amount of multicast messages even in moderately size multiprocessors [20]. As the core density grows, one-to-many traffic will increase not only in number of messages but also in number of destinations of each message. Figure 3 exemplifies this trend by plotting both metrics as a function of the number of cores for a set of applications from the SPLASH-2 and PARSEC benchmarks [10, 64]. The results are obtained by means of full-system simulation using gem5 [11] and assuming two different types of coherence. The simulated architecture consists of N cores, each of which accounts for two private 32-kB 2-way associative L1 caches (one for instructions and another one for data), as well as a bank of shared 8-way associative L2 cache of size 512 kB. We modified the network interfaces in order to register the characteristics of the multicast traffic generated by the cache [6].

Whereas the importance of multicast and broadcast increases with the core density, the performance of NoCs is likely to decrease. Conventional designs are based upon point-to-point links and messages with M destinations are generally treated as M unicast messages. At low core counts, the impact of such type of traffic can be neglected. Nevertheless, the interconnect fabric will saturate as the number of multicast packets and the average number of destinations grow for high core counts.
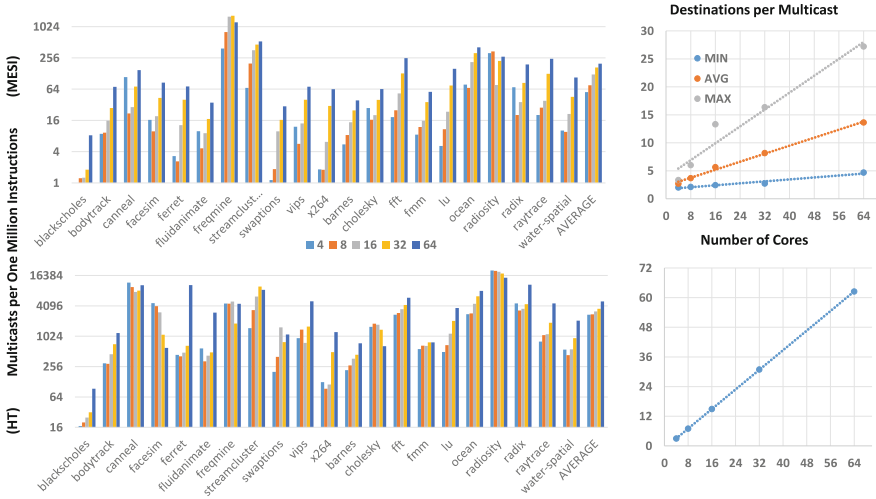
Fig. 3 Number of multicast messages per instruction (*left*) and average number of multicast destinations (*right*) as a function of the number of cores for different benchmark applications assuming MESI and HyperTransport coherence [6]

The size and wired nature of current NoCs render broadcast and multicast communications excessively costly and motivate the need for an alternative and efficient broadcast platform. The transition of broadcast from a constraint to an opportunity will not only provide means for the scaling of current architectures, but will also open a vast design space for the design and development of new architectures [20].

Although proposals to improve the performance of multicast and broadcast have been formulated for conventional NoCs [35, 50], RF NoCs [16] and photonic NoCs [36, 62], their scalability in terms of performance and cost remain largely unexplored. In light of the growing importance of one-to-many communications in the manycore scenario, a cost-effective solution is required. In the next section, we aim to address this issue by proposing the application of nanoscale communication techniques for chip-scale communications.

3   Graphene-Enabled Wireless Network-on-Chip

Among the plethora of emerging alternatives for on-chip communication, WNoC stands as a promising approach to complement existing wired interconnects. Wireless long-range links can be used to considerably decrease the multihop latency of a conventional NoC and even to provide one-hop communication for delay-critical traffic. Existing proposals adopt such approach by either placing antennas in a regular layout [38, 43] or following the principles of small-world networks [22], whereas multiple access is achieved by means of frequency or time channelization. Another advantage of WNoC is that implementing wireless links only requires, in physical

terms, the integration of an antenna and a transceiver at the nodes that we want to communicate. The network is not bound to any path infrastructure and, therefore, offers potential to adapt to varying delay and bandwidth requirements of the architecture. Such advantage is explored in [19], where a given set of time slots can be dynamically assigned depending upon link utilization.

While these designs have achieved significant delay and energy improvements with respect to conventional NoCs, their scalability is mainly compromised by the size of the on-chip antennas. Future on-chip metallic antennas are predicted to be hundreds of micrometers long, commensurate to the wavelength of terahertz electromagnetic waves [38]. This might render unfeasible the approach of integrating at least one antenna per core, as the cores 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 metallic antennas, as this would impose the use of frequencies from the near infrared to the optical ranges. Due to the low mobility of electrons in metals when nanometer scale structures are considered, and the challenges in implementing a transceiver which will be able to operate at this extremely high frequency, the feasibility of wireless communications at the core level would be compromised if this approach would be followed. Given these constraints, the current approach when integrating hundreds or thousands of cores is to use wireless links among sets of cores and then internally communicate these sets using on-chip wires [22, 43]. A packet may therefore propagate through the wired plane, then traverse a wireless link and finally return to the wired plane; whereas broadcast packets are distributed from the sender to the rest of sets and then internally within each set. In all cases, the performance improvements are ultimately limited by the performance of the wired network.

Instead, we propose to apply novel nanoscale communication techniques seeking to enable the integration of one or more antennas per core. This approach, to which we already referred to as GWNoC, consists in delivering core-level broadcast capabilities by means of the employment of graphene planar antennas. Antennas based upon a graphene patch just a few micrometers in size, i.e. two orders of magnitude below the dimensions of future metallic on-chip antennas, are expected to radiate in the terahertz (0.1–10 THz) band. These unique characteristics will both enable size compatibility with each processor core and offer enough bandwidth in massively parallel settings [31]. With a proper protocol stack, the latter will lead to low-latency and high-throughput schemes while complying with the severe area and power constraints of the manycore scenario.

Figure 4 shows the schematic representation of GWNoC within a manycore processor. We assume a hybrid approach, where the GWNoC is used to transport control flows and significant part of the broadcast-based data, and is deployed over a state-of-the-art NoC which serves heavy flows of data (not represented for simplicity). Each core is equipped with a network interface, a transceiver and at least one graphene antenna. Upon the release of a packet from a core, its network interface decides whether it must be transmitted through the wired or the wireless plane; in the second case, the transceiver modulates the information to be sent through the graphene antenna. At the receiver side, the graphene antenna picks up the wireless
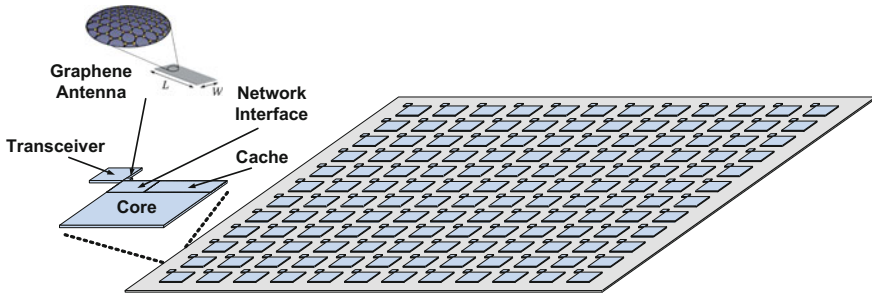
Fig. 4 Schematic Diagram of a 144-core graphene-enabled wireless network-on-chip

signal and passes it to the transceiver, which demodulates the data. The network inter-
face then checks the address of the packet and decides whether it must be delivered
to the core or discarded.

Since the information is radiated and can be received by any receiver within the
chip, GWNoC not only provides native broadcasting capabilities, but also makes data
transmission transparent with respect to the location of data. This heavily alleviates
the constraints of parallel architecture design, therefore reducing the complexity of
parallel programming and impacting upon the performance of virtually any future
application. Further, the integration of a wireless communication unit on a per-core
basis confers replicability and modularity to the on-chip design by means of the con-
cept of *wireless core*. A library of general-purpose or specific wireless cores could
be created, allowing the building of custom multicore processors by the integration
and configuration of a set of such pre-designed cores.

3.1   Modeling GWNoC Communications

Communications in the GWNoC scenario are unique since they are enabled by novel
antennas and occur within a unique environment and in the terahertz band. Under-
standing these aspects is an important step before the actual implementation of such
communications can be addressed. Conventional models, methods and tools cannot
be used and need to be profoundly revised to this end. In the following, we detail the
characteristics, requirements, and potential impact of each element involved in the
communication upon the system performance. Models and design methodologies
are briefly summarized whenever available.

3.1.1   Graphene Plasmonic Antennas

As conceptually represented in Fig. 5, a graphene antenna is composed of a finite-
size graphene layer, mounted over a metallic flat surface (the ground plane) with a
dielectric material layer in between, and an ohmic contact. These antennas are the

Fig. 5 Conceptual representation of a graphene plasmonic antenna
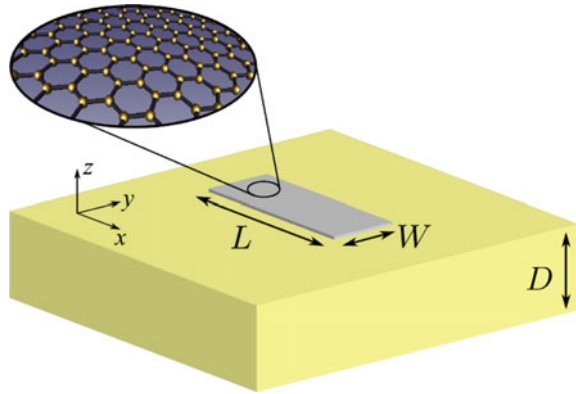
main enabler of the GWNoC approach due to their unique relation between size and radiation frequency. On the one hand, being up to two orders of magnitude smaller than metallic antennas for the same resonant frequency allows the integration of one or more antennas per processing core. On the other hand, the potential to radiate in the terahertz band provides a huge transmission bandwidth, allowing not only the transmission of information at extremely high speeds, but also the design of ultra-low-power and low-complexity schemes.

The reason behind such subwavelength behavior is that graphene antennas support the propagation of tightly confined Surface Plasmon Polariton (SPP) waves. Such phenomenon occurs at the interface between any metallic and dielectric material pair when an electromagnetic wave impacts upon the metal (graphene in our case). The wavelength of the SPPs within the metal determines the resonance condition and is given by λ/n_{eff}, where λ is the free-space wavelength and the effective mode index n_{eff} is:

$$n_{eff}(\omega) = \sqrt{1 - 4\frac{\mu_0}{\epsilon_0}\frac{1}{\sigma(\omega)^2}} \tag{1}$$

and yields, in the case of graphene, strong resonances at terahertz frequencies [5, 32, 59]. The effective mode index, among many other properties of the SPP waves, depends upon the frequency characteristics of the electrical conductivity of the metallic material $\sigma(\omega)$. Conductivity models of graphene are thus key to explore the radiation properties of graphene antennas. To model the conductivity of graphene, the main approach is to consider two approximations [5, 26]. Firstly, since we consider antennas with a size larger than 50 nm, it is possible to disregard the effects at the graphene edges. Secondly, we consider the interband contribution of the conductivity to be negligible in the frequency band of interest, which is a valid assumption when considering the terahertz band. With this, the conductivity is expressed as:

$$\sigma(\omega) = \frac{2e^2}{\pi\bar{h}} \frac{k_B T}{\bar{h}} \ln\left[2\cosh\left[\frac{E_F}{2k_B T}\right]\right] \frac{i}{\omega + i\tau^{-1}}, \qquad (2)$$

where e, $\bar{h}$ and k_B are constants. Variables T, τ and E_F correspond to the temperature, the relaxation time, and the chemical potential of the graphene layer. The relaxation time is the interval required for a material to restore a uniform charge density after a charge distortion is introduced, and it highly depends upon the quality of the graphene sheet and of the underlying substrate. The chemical potential or Fermi energy E_F refers to the level in the distribution of electron energies at which a quantum state is equally likely to be occupied or empty. The chemical potential can be modified by applying a voltage to the antenna (thereby allowing to dynamically tune its radiation properties) or by means of chemical doping.

Using Eq. 2, the frequency response of the conductivity is evaluated for a fixed chemical potential and relaxation time pair. Since graphene is a one-dimensional material, the antenna can be then modeled as a patch with an equivalent surface impedance of $Z = \frac{1}{\sigma}$. Such possibility is available in commercial electromagnetic field solving simulators and allows to obtain the frequency response of the antenna upon the presence of incident electromagnetic waves. By means of this methodology, important performance aspects of the antenna can be determined as functions of graphene technological parameters (i.e., chemical potential and relaxation time), as well as the antenna design parameters (e.g., size and shape), including but not limited to:

The antenna impedance and radiation efficiency: the frequency response of the impedance and of the radiation efficiency are crucial for the design of a transceiver that will drive the antenna. The frequency and power of the input signals, as well as the characteristic impedance of the source of those signals need to be determined taking into consideration the antenna impedance and radiation efficiency. Recent works report a radiation efficiency of up to 25% for graphene patch antennas and a very high impedance in the kΩ range [59].

The antenna bandwidth: is a crucial performance metric since a high data rate potentially leading to high throughput is required. The peculiarity of graphene antennas is that bandwidth depends not only upon the shape of the antenna but also upon the technological parameters of the material. In the former case, high bandwidths can be obtained with fractal or inherently broadband structures. In the latter case, recent results state that the relaxation time of the graphene sheet has a significant impact upon the resonance bandwidth [40].

The radiation pattern: which indicates the strength of the radiated signal as a function of the radiation direction. Recent works demonstrate that the radiation pattern of graphene patches is similar to that of their metallic counterparts [40, 59], suggesting a dependence on the antenna size and shape rather than on the radiative material. For antennas based on graphene patches, the radiation efficiency is extremely low in the plane of the antenna and substantially higher in the perpendicular direction.

Graphene antennas is a thriving albeit still wide open research area. At the time of writing this book chapter, several groups are currently conducting intense research

towards a further characterization of graphene patch antennas [39, 40, 59, 60]. For instance, the impact of the substrate material and thickness upon the radiation of the antenna must be taken into consideration [39]. Also, studying graphene antennas in transmission requires defining a feeding mechanism. This represents a challenge by itself, since the feeder must support the propagation of SPPs and must be matched to the antenna. The design of a matching mechanism requires, in turn, modeling the effects of the contacts between the feeder and the edge of the graphene patch.

3.1.2 Terahertz Within-Package Channel

A channel model that takes into consideration the peculiarities of the GWNoC scenario is fundamental in order to evaluate the available on-chip communication bandwidth. Mainly, the enclosed nature of chip processors causes the apparition of a large number of reflections that must be taken into consideration at the receiver. The physical landscape of a multiprocessor involves multiple dielectric/metallization layers and components printed on the chip surface, among other elements that need to be accurately described in order to model the channel [42]. Since such landscape is static, the model will be time-invariant.

In the general setting shown in Fig. 6, radiated signals reach the receiver via different paths [66]. First, surface waves propagate at the interface of the chip and the package medium. These waves show particularly low attenuation per unit of distance due to their cylindrical characteristics and are affected by the circuits printed on the chip surface. However, since graphene patch antennas show an extremely low radiation efficiency in the coplanar direction, the contribution of surface waves at the receiving end may be negligible. Second, part of the energy of patch antennas is radiated into the substrate. These waves are guided within the substrate and reach the receiver after repeated reflections upon the ground plane of the chip and the insulating layer. However, the substrate is generally lossy and introduces a very high attenuation per unit of distance. Given that surface and guided waves are highly attenuated by the antenna and the substrate, respectively, in most cases we can con-
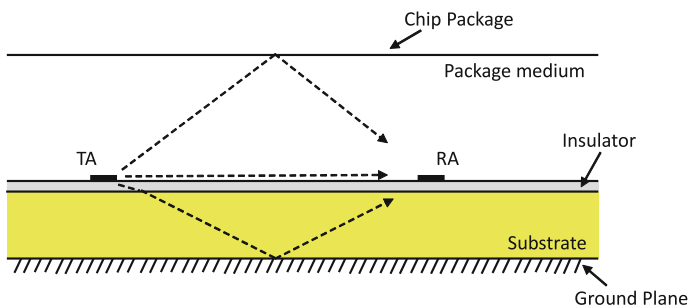

Fig. 6 Electromagnetic waves that may potentially reach the receiver

sider that communication occurs by means of a third mechanism: space waves that propagate through the medium and reflect upon the chip package and surface.

Modeling the channel implies evaluating all possible rays reaching the receiving antenna for each and every pair of antennas within the chip [42]. The result in the time domain will then be a sum of channel impulse responses: $h(t) = \sum_i \alpha_i e^{j\phi_i} g_i(\tau - \tau_i)$, where α_i, ϕ_i and τ_i are the amplitude, phase shift, and delay of the i-th ray. Note that this function is generally evaluated using Dirac deltas $\delta(t)$ instead of waveforms $g_i(t)$. In the GWNoC case, though, such ideal approach is not accurate as both propagation and reflections are frequency-dependent phenomena and antennas will radiate over a large bandwidth in the terahertz band.

Propagation: Communicating by means of terahertz waves has two main implications on propagation. First, we can assume that the far-field condition holds given both the short radiation wavelength and the fact that communication occurs by means of reflected waves. Second, the phenomenon of *molecular absorption* must be accounted for on top of typical spreading losses. Molecular absorption is the process by which part of the wave energy is converted into internal kinetic energy of the excited molecules in the medium. Molecules present in standard media have numerous resonances in the terahertz band, causing a frequency-selective attenuation of terahertz electromagnetic waves radiated by antennas [41]:

$$\alpha(f, d) = e^{k(f)d}$$

Note that the attenuation highly depends on the medium absorption coefficient k, which models the particular mixture of molecules in the medium; as well as on the transmission distance d that determines the number of molecules that the waves will find along their path. The inset of Fig. 7 exemplifies the latter dependence by representing the molecular absorption of the terahertz channel for transmission distances of 1 cm and 10 cm. Both the number of absorption peaks and their amplitude

Fig. 7 Available bandwidth in the frequency band from 0 to 50 THz due to molecular absorption, as a function of the transmission distance. The *inset* shows the molecular absorption in dB and available bandwidth for two particular distances: 1 cm (*blue*) and 10 cm (*red*)
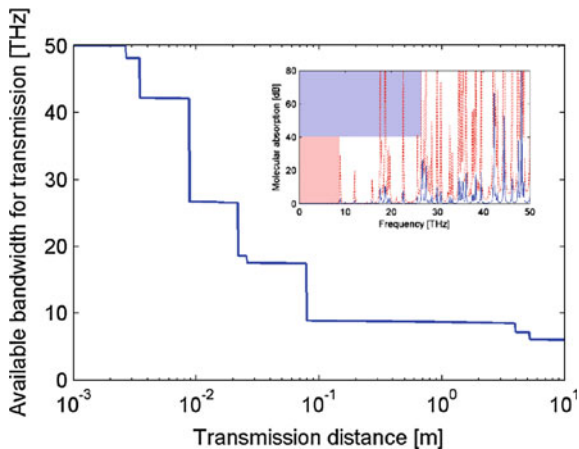
notably increase in the latter case, reducing the 10-dB bandwidth from 27 THz (top blue background) to 9 THz (bottom red background). Figure 7 shows the available 10-dB bandwidth for a range between one millimeter to ten meters [41]. In light of these results, it is concluded that molecular absorption has a limited impact on transmissions at the chip scale, fact that may lead to channel capacities over the terabit-per-second barrier [31].

Reflections: the characteristics of reflected waves depend both on the roughness of the surface and on the reflective material. The effects of the former can be neglected for conventional metallic materials in the frequency range of interest [37], whereas the latter is polarization-sensitive and given by the Fresnel coefficients of the different media [51]. The main issue here is that these coefficients are frequency-dependent and require knowing the frequency response of the materials present in on-chip environments; however, only a few materials have been characterized in the terahertz band [24, 51].

3.1.3 Transceiver

In order to enable on-chip wireless communication, it is necessary to develop a transceiver to modulate and demodulate the data and to drive the antenna. To this end, such transceiver needs to operate at the same frequency than the antenna itself. This represents a grand challenge since terahertz transceivers are still not available, even though advancements in CMOS [53] and alternative technologies based on InP [34] or graphene [25, 65] may enable their creation in the near future.

Since critical metrics such as the area and power consumed by the wireless communication unit mainly depend on the characteristics of the transceiver, accurate models are key to assess the feasibility of a GWNoC design. However, such models are not available since terahertz technologies are still in their infancy. Instead, behavioral area and energy models could be created from state-of-the-art transceiver implementations and then extrapolated to extract results in the terahertz region [4].

On the one hand, recent works point towards a promising decrease in the transceiver area when the frequency is upscaled (see Fig. 8, [4]). The reasons for the observed tendency may stem from the strong downsizing that is applied to the passive RF components of a transceiver when the operation frequency is increased. Rational fitting is chosen on the grounds that it delivers the most accurate result among the possible fittings and that it does not yield negative values for high frequencies, which would be unrealistic. On the other hand, since the transceiver energy is highly dependent on the transmission range, authors in [23] propose and discuss a figure of merit Φ that encompasses both their metrics as: $\Phi = \frac{E_{bit}}{\sqrt{d_{max}}}$. Figure 8 shows how this figure of merit scales as a function of the frequency for state-of-the-art implementations [4]. In this case, we also observe a decay of the energy per bit proportional to the radiation frequency.
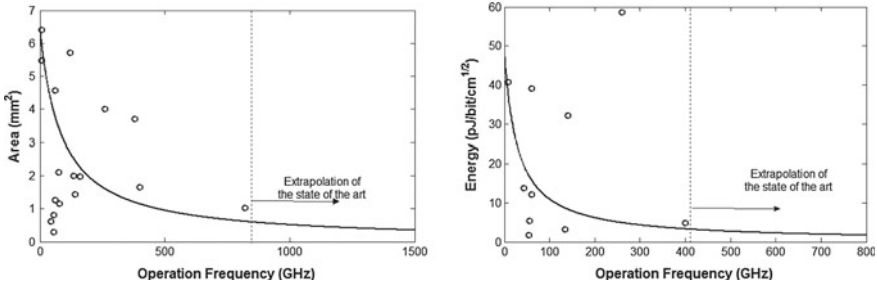
Fig. 8 Area and energy efficiency of state-of-the-art wireless transceivers as a function of their central frequency. See [4, 23, 47, 48] and references therein for more details

The results here presented confirm that terahertz circuits will likely be suitable for wireless on-chip communication purposes given the inverse relation between area, energy, and operation frequency.

3.1.4 Network Interface

As its name implies, the function of the network interface is to bridge the memory system and the network. In conventional NoCs, the network interface receives data from the memory system and creates a packet with it, to then split the packet into flow control units (flits). Finally, the network interface puts the flits in its output queue and sends them to the associated router whenever possible. At the other end, the network interface receives the flits, reconstructs the packet, and then checks that the destination address corresponds to its core address.

In our scenario, the on-chip interconnect accounts for both a wired and a wireless plane. A controller must be added to each network interface in order to determine through which plane a given piece of data should be sent. Before being sent to the transceiver, the data to be wirelessly transmitted must be packetized and serialized into a stream of bits. Finally, the inverse process is performed at the receiver: a stream of bits is received from the transceiver and interpreted. The network interface then checks the address in order to decide whether the packet must be either yielded to the core or discarded.

3.2 Design Decisions for GWNoC Protocols

The peculiarities of the GWNoC scenario require the design and development of a unique network architecture. Protocols for classical wireless networks cannot be applied to on-chip communications due to the blend of power, area limitations and stringent performance demands of multiprocessors. Luckily, some favorable condi-

tions may compensate for these challenging requirements and lead to opportunistic solutions. For instance, methods at compile-time could allow including traffic information in the code to be executed, so that the network can prepare for traffic bursts or high-contention phases. Next, we detail the challenges that must be addressed at each level of design, as well as possible approaches that could be adopted.

3.2.1 Modulations

The area and energy figures of a transceiver not only strongly depend on the implemented modulation, but are also generally traded off against performance. Therefore, modulations are an important design step in the GWNoC scenario, as a balance between area, energy, and performance is sought. Working at terahertz speeds may allow achieving these goals provided that additional challenges are addressed. Mainly, the solution must be feasible and adapt to the terahertz components that technology progress will made available in the years to come. Jitter should also be taken into consideration with special attention, as it may become an important performance bottleneck due to the extremely fine temporal resolution needed at the receiver. Such unique features strongly limit the boundaries of the practical design space.

Within this context, Impulse Radio Ultra-Wideband (IR-UWB) techniques stand out as promising candidates for the implementation of on-chip wireless communication. The IR-UWB consists of the transmission of very short baseband pulses, the length of which determines the bandwidth of such spread spectrum signal. Academic research efforts have gone beyond commercial implementations at the 3.1–10.6 GHz band and explored frequencies up to 110 GHz [57]. Following this trend, communication in the terahertz band can be accomplished by means of the transmission and reception of picosecond long pulses. Furthermore, IR-UWB yields potential for the devising of simple and low-power systems by means of non-coherent detection. This approach advocates for the detection of the energy of the signal rather than its phase, offering simplicity at the expense of a lower performance for fixed levels of noise. Non-coherent detection eliminates the need for channel estimation and makes the system more robust against timing issues induced by jitter [63]. From an implementation standpoint, the use of power hungry components such as a phase-locked loop can be avoided in asynchronous schemes. Also, it allows to perform initial signal processing tasks in the analog domain, leading to sub-Nyquist sampling rates [7]. This aspect is critical since Nyquist sampling rates imply a need for power demanding analog-to-digital converters able to operate in the terahertz band.

Energy detection is compatible with a limited number of modulations. Among them, On-Off Keying (OOK, modulating by means of the presence/absence of pulses) is particularly suitable to the GWNoC scenario due to its simplicity and relaxed timing constraints. The probability density of OOK zeroes and ones at the energy detector are evaluated using well-known central and non-central chi-square distributions, $\chi^2(k)$ and $\chi^2(k, \mu)$ [61]. The k represents the degrees of freedom or number of samples per symbol and is generally taken as $2 \cdot TW$, where TW is the time-bandwidth product of pulses at the receiver. The non-central distribution has

a non-centrality parameter μ equal to the signal to noise ratio $\gamma = hE_b/N_0$, where h accounts for the loss of energy due to jitter-induced effects. Assuming a threshold λ calculated following the *maximum a priori* criterion, the error probabilities are:

$$P(1|0) = \int_{\lambda}^{\infty} \chi^2(2TW) = \Gamma(TW, \lambda/2)/\Gamma(TW) \qquad (3)$$

$$P(0|1) = \int_{-\infty}^{\lambda} \chi^2(2TW, \gamma) = 1 - Q_{TW}(\sqrt{2\gamma}, \sqrt{\lambda})$$

where $\Gamma(\cdot, \cdot)$ corresponds to the incomplete Gamma function and $Q_u(\cdot, \cdot)$ is the generalized Marcum Q-function of order u. The error probability is then evaluated as: $P_e = P(0)P(1|0) + P(1)P(0|1)$. Upon the presence of jitter affecting the signal to noise ratio, the BER is calculated as the weighted average over the jitter probability density function: $BER = \int P_e(\varepsilon_i)f(\varepsilon_i)d\varepsilon_i$.

Next, we quantify the performance of OOK with energy detection and compare it to that of more complex options. Coherent schemes, i.e. matched filter and autocorrelator, are considered as receivers of Binary Pulse Position Modulation (BPPM), Binary Phase Shift Keying (BPSK), Transmitted Reference (TR) and differential (DIFF) schemes. On the one hand, a matched filter assumes perfect channel estimation and recovers the phase of the signal directly from the received pulse. Unlike the energy detector, this enables the demodulation of BPSK signals, the information of which is encoded within the pulse polarity. On the other hand, the autocorrelator relies upon a previously received pulse in order to estimate the channel and recover the information. In TR, a pair of pulses is sent for each symbol: the first one serves as a pilot and the second one modulates the information. In DIFF, the information is modulated differentially between each two consecutive pulses, allowing to save half of the energy per symbol.

The simulation framework assumes fixed signal-to-noise ratio (SNR) and data rate objectives, to then calculate the appropriate pulse characteristics for each scheme. Each value of jitter implies a different effective received power, which is used to evaluate the BER by using *(1)* the model explained above for the OOK case and *(2)* additional equations for the rest of cases (see [63] for more details). The BER is then averaged over all the probability density function of the jitter. Figure 9 shows the BER performance with respect to the jitter level in a system working at 100 Gbps and a nominal SNR of 18 dB. We observe that the combination of OOK and energy detector yields a performance comparable to that of coherent receivers for high levels of jitter, suggesting that the reduction in synchronization requirements of non-coherent detection makes up for its worse nominal performance.

3.2.2 Coding

As mentioned in Sect. 2, on-chip interconnects are designed to operate with a BER on the order of 10^{-15}. In light of the BER results shown above, such target will not

Fig. 9 Receiver
performance comparison as
a function of the jitter level
for a SNR of 18 dB and
working at 100 Gbps. The
performance of the ED-OOK
combination is compared to
that of *(1)* ideal coherent
detection (Matched) with
BPPM and BPSK, as well as
(2) autocorrelation (Auto)
with TR and DIFF
modulations
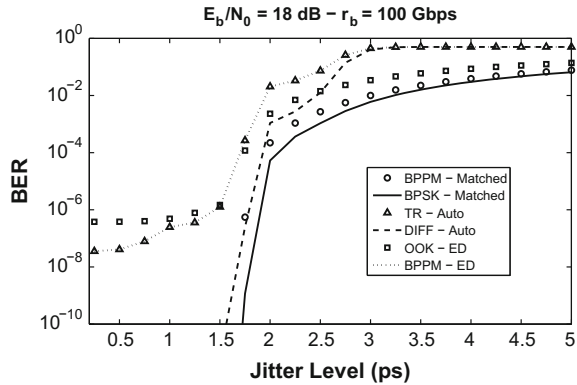
be likely achieved by means of increasing the signal to noise ratio. Since errors in GWNoC are due to thermal noise and will not occur in bursts, forward error correction could provide an effective way to reduce the error probability at the expense of reducing the effective data rate. Reed-Solomon (RS) with low-density parity check (LDPC) coding schemes have been proposed in the 802.15.3c standard [1], which works upon the physical layer of millimeter-wave radio for high-rate WPAN networks. These may be suitable in GWNoC environments as they are expected to provide a low-complexity implementation and high error-correcting capability. $RS(n, k)$ codes build codewords of length n, k of which are data; the remaining $2t$ bits are for parity check, allowing the correction of up to t erroneous symbols within the codeword. Assuming p to be the bit error probability considering a raw channel, the use of $RS(n, k)$ codes reduces the BER to:

$$BER = 1 - \frac{(1-p)^n}{k} - \frac{n}{k}p(1-p)^{n-1}$$

at the expense of reducing effective data rate a factor of k/n with respect to the raw data rate. In addition to codes for error correction, low-weight codes could be also employed. When combined with the OOK modulation, low-weight coding reduces the average power consumption for a given link budget.

3.2.3 Medium Access Control (MAC)

Coordinating the access to the shared medium is a huge challenge in GWNoC. All processors will be located within the same transmission range, implying the existence of one collision region accounting for hundreds or thousands of nodes. Furthermore, some applications generate large amounts of communication throughout the chip. The design of a MAC protocol that coordinates the expectedly high number of simultaneous transmissions is therefore key to guarantee that the performance requirements of GWNoC will be met. Above all, such protocol must be scalable in

terms of latency since it has a critical impact in the performance of the multiprocessor. Another important aspect is that GWNoC, unlike other wireless networks, must guarantee the delivery of broadcast packets to all nodes. Acknowledgement (ACK) packets must be conveyed to the transmitter in order to avoid losses due to collisions. However, this cannot be performed through the wireless plane as the "reply storm" would saturate the medium. Instead, the wired plane could be used.

The GWNoC scenario presents a set of peculiarities that may allow the design of opportunistic solutions. For instance, hidden or exposed terminal problems are avoided as all nodes are static and within the same transmission range. Also, most of the MAC protocols are designed assuming that no prior information on the traffic is available. However, this is not the case in our scenario. The size of the messages is known as it depends upon the multiprocessor architecture and the function of broadcast transmissions. In coherence protocols, coherence requests are short messages of around 8–16 bytes and responses may include cache lines up to 128 bytes. Further, estimations on the traffic to transmit can be generated by the compiler and provided at run-time so that the MAC protocol can adapt to the instantaneous traffic load. This feature may be employed to improve fairness or to avoid saturation in high-contention phases of some parallel programs.

Existing proposals rely on channelization approaches to control the medium access. Frequency-multiplexing schemes have been evaluated [38], but their scalability is compromised by the number of channels that will be required in manycore processors. Combinations of time-multiplexing and frequency-multiplexing schemes have been also proposed, seeking to increase the number of channels [19, 43] or the available bandwidth [22]. In this case, time-multiplexing schemes introduce a latency which is proportional to the number of cores and that may not be tolerated in manycore settings. Except for the work in [19], current proposals do not offer any reconfigurability option to adapt to the time-varying needs of the application. In this scenario, MAC protocols where nodes contend for the channel are generally a better choice. A carrier sensing approach (or energy sensing approach for impulse radio) may be adopted and adapted to provide means to take advantage of the information that the compiler could provide.

3.2.4 Network Layer

Since the main aim of GWNoC is to provide one-hop broadcast communication, routing or switching strategies are not required in the wireless plane. On the contrary, switching functionalities have to be added at the network interface as it has to deliberate through which plane a message is to be sent. The decision may be simply taken depending on whether the message is broadcast or not, or depending on upper layer policies such as congestion control. Another important aspect to carefully consider is multicast addressing. Since all messages are broadcast, the network interface must be provided with means to decide whether to keep or discard a message based on the address of the packet.

4 Conclusions and Future Work

In the manycore era, the exponential increase in *one-to-many* communication requirements intensifies the need for a scalable broadcast on-chip platform. Although the concept of wireless on-chip networks has been proposed and may be suitable to this end, size constraints hinder the use of metallic antennas and require the use of nanoscale techniques. In this chapter, we presented the concept of GWNoC, wherein graphene antennas, by virtue of their downscaled size which allow per-core wireless capabilities, deliver broadcast at the core level. We analyzed the GWNoC approach from both the communication performance and protocol design perspectives, providing models and guidelines for their evaluation.

On the one hand, we analyzed the unique properties of a graphene-enabled wireless link in Sect. 3.1. We introduced a methodology based on conductivity models for the simulation of graphene antennas as a function of different technological parameters. However, further research is required in order to fully understand these antennas and develop models that will capture all the phenomena that affect radiation. In the case of the propagation channel, we presented a general model and detailed the three propagation mechanisms present in the GWNoC scenario. We conclude that communication will mainly occur by means of the space waves that propagate through air and reflect upon the chip package. Terahertz wave propagation could be challenged by molecular absorption, but we have shown that its impact becomes negligible at the chip scale; whereas reflections are frequency-dependent and will require further work in material characterization at terahertz frequencies. Finally, in the case of the transceiver, we extrapolated performance trends from the state of the art to show that area and power objectives could be met by operating in the terahertz band.

On the other hand, we discussed the main protocol design aspects in Sect. 3.2. We first qualitatively analyzed the physical layer. We proposed to employ IR-UWB modulations using subpicosecond long pulses leading to terahertz-wide signals. Seeking simplicity and energy efficiency, we both discussed the use of OOK in combination with non-coherent detection and reviewed a model for its performance evaluation. Results show a good compromise between performance and robustness in front of timing effects, but also suggest the use of coding schemes to reduce the BER to acceptable levels for on-chip communication. At the MAC layer, we analyzed the main peculiarities of the scenario and concluded that frequency- or time-multiplexing options are not suitable due to their poor scalability. Instead, protocols where nodes contend for the channel could be used due to their potential adaptability to the time-varying communication requirements of manycore processors. Furthermore, information on the traffic to be served may be available at run-time and could be used to improve the network performance. In this regard, a detailed characterization of the traffic generated by cores when running a set of benchmark applications will be a helpful tool for the design of opportunistic MAC solutions.

References

1. 802.15.3c-Part 15.3 (2009) Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs)—Amendment 2: Millimeter-wave-based alternative physical layer extension
2. Abadal S, Alarcón E, Lemme MC, Nemirovsky M, Cabellos-Aparicio A (2013) Graphene-enabled wireless communication for massive multicore architectures. IEEE Commun Mag 51(11):137–143
3. Abadal S, Cabellos-Aparicio A, Lázaro JA, Nemirovsky M, Alarcón E, Solé-Pareta J (2013) Area and laser power scalability analysis in photonic networks-on-chip. In: Proceedings of the ONDM '13
4. Abadal S, Iannazzo M, Nemirovsky M, Cabellos-Aparicio A, Alarcon E (2015) On the area and energy scalability of wireless network-on-chip: a model-based benchmarked design space exploration. IEEE/ACM Trans Netw 23(5):1501–1513
5. Abadal S, Llatser I, Mestres A, Lee H, Alarcón E, Cabellos-Aparicio A (2015) Time-domain analysis of graphene-based miniaturized antennas for ultra-short-range impulse radio communications. IEEE Trans Commun 63(4):1470–1482
6. Abadal S, Martínez R, Solé-Pareta J, Alarcón E, Cabellos-Aparicio A (2016) Characterization and modeling of multicast communication in cache-coherent manycore processors. Comput Electr Eng (51):168–183
7. Arslan H, Chen Z, Benedetto MD (2006) Ultra wideband wireless communication
8. Beausoleil RG, Kuekes PJ, Snider GS, Wang, SY, Williams RS (2008) Nanoelectronic and Nanophotonic Interconnect. Proc IEEE 96(2):230–247
9. Benini L, De Micheli G (2002) Networks on chips: a new SoC paradigm. Computer 35(1): 70–78
10. Bienia C, Kumar S, Singh JP, Li K (2008) The parsec benchmark suite: characterization and architectural implications. In: Proceedings of the PACT '08, pp 72–81. ACM
11. Binkert N, Sardashti S, Sen R, Sewell K, Shoaib M et al (2011) The gem5 simulator. ACM SIGARCH Comput Arch News 39(2):1
12. Burns J, McIlrath L, Keast C, Lewis C, Loomis A, Warner K, Wyatt P (2001) Three-dimensional integrated circuits for low-power, high-bandwidth systems on a chip. In: IEEE ISSCC Dig Tech Papers:268–269
13. Cai W, White J, Brongersma M (2009) Compact, high-speed and power-efficient electrooptic plasmonic modulators. Nano Lett 9(12):4403–4411
14. Chan J, Hendry G, Biberman A, Bergman K, Carloni LP (2010) PhoenixSim: a simulator for physical-layer analysis of chip-scale photonic interconnection networks. In: Proceedings of the DATE '10, pp 691–696
15. Chang MCF, Verbauwhede I, Chien C, Xu Z, Kim J, Ko J, Gu Q, Lai BC (2005) Advanced RF/baseband interconnect schemes for inter- and intra-ULSI communications. IEEE Trans Electron Devices 52(7):1271–1285
16. Chang MF, Cong J, Kaplan A, Naik M, Reinman G, Socher E, Tam SW (2008) CMP Network-on-chip overlaid with multi-band RF-interconnect. In: Proceedings of the HPCA '08, pp 191–202
17. David Culler AG (1999) Parallel computer architecture: a hardware/software approach
18. Deb S, Ganguly A, Pande PP, Belzer B, Heo D (2012) Wireless NoC as interconnection backbone for multicore chips: promises and challenges. IEEE J Emerg Sel Topics Circuits Syst (JETCAS) 2(2):228–239
19. DiTomaso D, Kodi A, Matolak D (2013) Energy-efficient adaptive wireless NoCs architecture. In: Proceedings of the NoCS '13, pp 1–8
20. Enright Jerger N, Peh LS, Lipasti M (2008) Virtual circuit tree multicasting: a case for on-chip hardware multicast support. In: Proceedings of the ISCA-35, pp 229–240
21. Feero BS, Pande PP (2009) Networks-on-Chip in a three-dimensional environment: a performance evaluation. IEEE Trans Comput 58(1):32–45

22. Ganguly A, Chang K, Deb S, Pande PP, Belzer B, Teuscher C (2010) Scalable hybrid wireless network-on-chip architectures for multi-core systems. IEEE Trans Comput 60(10):1485–1502
23. Gorisse J, Morche D, Jantunen J (2012) Wireless transceivers for gigabit-per-second communications. In: Proceedings of the NEWCAS '12, pp 545–548
24. Grischkowsky D, Keiding S, van Exter M, Fattinger C (1990) Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors. J Opt Soc Am 7(10): 2006–2015
25. Han SJ, Garcia AV, Oida S, Jenkins KA, Haensch W (2014) Graphene radio frequency receiver integrated circuit. Nat Commun 5
26. Hanson GW (2008) Dyadic Green's Functions for an Anisotropic, Non-Local Model of Biased Graphene. IEEE Transactions on Antennas and Propagation 56(3):747–757
27. Hennessy J, Patterson D (2012) Computer architecture: a quantitative approach
28. Hesse R, Nicholls J, Jerger NE (2012) Fine-grained bandwidth adaptivity in networks-on-chip using bidirectional channels. In: Proceedings of the NoCS '12, pp 132–141. IEEE
29. Hoskote Y, Vangal S, Singh A, Borkar N, Borkar S (2007) A 5-GHz mesh interconnect for a teraflops processor. IEEE Micro 27(5):51–61
30. Huang W, Rajamani K, Stan M, Skadron K (2011) Scaling with design constraints: predicting the future of big chips. IEEE Micro:16–29
31. Jornet JM, Akyildiz IF (2011) Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band. IEEE Trans Wirel Commun 10(10):3211–3221
32. Jornet JM, Akyildiz IF (2013) Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks. IEEE J Sel Areas Commun 31(12):685–694
33. Kahng A, Li B, Peh L, Samadi K (2009) Orion 2.0: A fast and accurate noc power and area model for early-stage design space exploration. In: Proceedings of the DATE '09
34. Kim M, Rieh JS, Jeon S (2012) Recent progress in terahertz monolithic integrated circuits. In: Proceedings of the ISCAS '12, pp 746–749
35. Krishna T, Peh LS (2011) Towards the ideal on-chip fabric for 1-to-many and many-to-1 communication. In: Proceedings of the MICRO-44, pp 71–82
36. Kurian G, Miller J, Psota J, Eastep J et al (2010) ATAC: A 1000-Core Cache-Coherent Processor with On-Chip Optical Network. In: Proceedings of the PACT '10
37. Kürner T, Priebe S (2013) Towards THz communications—status in research, standardization and regulation. J Infrared, Millimeter Terahertz Waves 35(1):53–62
38. Lee SB, Tam SW, Pefkianakis I, Lu S et al (2009) A scalable micro wireless interconnect structure for CMPs. In: Proceedings of the Mobicom '09, p 217
39. Llatser I, Kremers C, Cabellos-Aparicio A, Jornet JM, Alarcón E, Chigrin DN (2012) Graphene-based nano-patch antenna for terahertz radiation. Photonics Nanostruct: Fund Appl 10(4):353–358
40. Llatser I, Kremers C, Chigrin D, Jornet JM, Lemme MC, Cabellos-Aparicio A, Alarcón E (2012) Radiation characteristics of tunable graphennas in the terahertz band. Radioeng J 21(4)
41. Llatser I, Mestres A, Abadal S, Alarcón E, Lee H, Cabellos-Aparicio A (2015) Time and frequency domain analysis of molecular absorption in short-range terahertz communications. IEEE Antennas Wirel Propag Lett 14:350–353
42. Matolak D, Kaya S, Kodi A (2013) Channel modeling for wireless networks-on-chips. IEEE Commun Mag 51(6):180–186
43. Matolak D, Kodi A, Kaya S, DiTomaso D, Laha S, Rayess W (2012) Wireless networks-on-chips: architecture, wireless channel, and devices. IEEE Wireless Commun 19(5):58–65
44. Miller DAB (2009) Device requirements for optical interconnects to silicon chips. Proc IEEE 97(7):1166–1185
45. Novack A, Liu Y, Ding R, Gould M, Baehr-jones T, Li Q, Yang Y, Zhang Y, Padmaraju K, Bergmen K, Lim AEJ, Lo GQ, Hochberg M (2013) A 30 GHz silicon photonic platform. In: Proceedings of the SPIE—Integrated optics: physics and simulations, vol 8781
46. O KK, Kim K, Floyd B, Mehta J, Yoon H, Hung CM, Bravo D, Dickson T, Guo X, Li R, Trichy N, Caserta J, Yang D, Bohorquez J, Seok E, Gao L, Sugavanam A, Lin JJ, Chen J, Brewer, JE (2005) On-chip antennas in silicon ICs and their application. IEEE Trans Electron Devices 52(7):1312–1323

47. Öjefors E, Grzyb J, Heinemann B, Tillack B, Pfeiffer UR (2011) A 820 GHz SiGe chipset for terahertz active imaging applications. In: Proceedings of the ISSCC '11, pp 224–225
48. Park JD, Kang S, Thyagarajan S, Alon E, Niknejad A (2012) A 260 GHz fully integrated CMOS transceiver for wireless chip-to-chip communication. In: Proceedings of the VLSIC '12, pp 48–49
49. Pande P, Grecu P, Jones C, Ivanov M, Saleh A (2005) Performance evaluation and design trade-offs for network-on-chip interconnect architectures. IEEE Trans Comput 54(8):1025–1040
50. Rodrigo S, Flich J, Duato J, Hummel M (2008) Efficient unicast and multicast support for CMPs. In: Proceedings of the MICRO-41 pp 364–375
51. Ronne C, Thrane L, AŁstrand PO, Wallqvist A et al (1997) Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation. J Chem Phys 107(14):5319
52. Schinkel D, Mensink E (2009) Low-power, high-speed transceivers for network-on-chip communication. IEEE Trans VLSI Syst 17(1):12–21
53. Seok E, Shim D, Mao C, Han R, Sankaran S, Cao C, Knap W (2010) Progress and challenges towards terahertz CMOS integrated circuits. IEEE J Solid-State Circuits 45(8):1554–1564
54. Shacham A, Bergman K, Carloni LP (2008) Photonic networks-on-chip for future generations of chip multiprocessors. IEEE Trans Comput 57(9):1246–1260
55. Socher E, Chang MCF (2007) Can RF Help CMOS processors? IEEE Commun Mag 45(8):104–111
56. Soteriou V, Wang H, Peh LS (2006) A statistical traffic model for on-chip interconnection networks. In: Proceedings of the MASCOTS '06
57. Stallo C, Mukherjee S (2010) IR-UWB for high bit rate communications beyond 60 GHz. In: Proceedings of the PIMRC '10
58. Sun C, Chen C, Kurian G (2012) DSENT—a tool connecting emerging photonics with electronics for opto-electronic networks-on-chip modeling. In: Proceedings of the NoCS '12, pp 201–210
59. Tamagnone M, Gomez-Díaz JS, Mosig JR, Perruisseau-Carrier J (2012) Analysis and design of terahertz antennas based on plasmonic resonant graphene sheets. J Appl Phys 112:114, 915
60. Tamagnone M, Gomez-Díaz JS, Mosig JR, Perruisseau-Carrier J (2012)Reconfigurable terahertz plasmonic antenna concept using a graphene stack. Appl Phys Lett 101(21), 214, 102
61. Urkowitz H (1967) Energy detection of unknown deterministic signals. Proc IEEE 55(4)
62. Vantrease D, Schreiber R, Monchiero M, McLaren M, Jouppi N, Fiorentino M, Davis A, Binkert N, Beausoleil R, Ahn J (2008) Corona: system implications of emerging nanophotonic technology. ACM SIGARCH Comput Architect News 36(3):153–164
63. Witrisal K, Leus G, Janssen GJM, Pausini M, Troesch F, Zasowski T, Romme J (2009) Noncoherent ultra-wideband systems. IEEE Signal Process Mag 26(4):48–66
64. Woo S, Ohara M, Torrie E, Singh J (1995) The SPLASH-2 programs: characterization and methodological considerations. In: Proceedings of the ISCA-22, vol 23, issue no 2, pp 24–36
65. Wu Y, Farmer DB, Xia F, Avouris P (2013) Graphene electronics: materials, devices, and circuits. Proc IEEE 101(7):1620–1637
66. Zhang YP, Chen ZM, Sun M (2007) Propagation mechanisms of radio waves over intra-chip channels with integrated antennas: frequency-domain measurements and time-domain analysis. IEEE Trans Antennas Propag 55(10):2900–2906