Dynamic Management Policies for Exploiting Hybrid Photonic-Electronic NoCs

Antonio García-Guirado
University of Murcia
toni@ditec.um.es

Ricardo Fernández-Pascual
University of Murcia
rfernandez@ditec.um.es

José M. García
University of Murcia
jmgarcia@ditec.um.es

Sandro Bartolini
University of Siena
bartolini@dii.unisi.it

ABSTRACT

Nanophotonics promises to solve the scalability problems of current electrical interconnects thanks to its low sensitivity to distance in terms of latency and energy consumption. Before this technology reaches maturity, hybrid photonic-electronic networks will be a viable alternative. Ideally, an ordinary electrical mesh and a ring-based photonic network should cooperate to minimize overall latency and energy consumption, but currently we lack mechanisms to do this efficiently. In this paper we present novel fine-grain policies to manage the photonic resources in a tiled-CMP scenario. Our policies are dynamic and base their decisions on parameters such as message size, ring availability and distance between endpoints, at the message level. The resulting network behavior is also fairer to all cores, reducing processor idle time thanks to faster thread synchronization. Of these policies, only the most elaborate ones reduce the overall network latency by 50%, execution time by 36% and network energy consumption by 52% on average, in a 16-core CMP for the PARSEC benchmark suite, when compared to the same CMP without the photonic ring.

1. INTRODUCTION

To be able to keep pace with Moore's Law, the latest generations of most microprocessors have adopted an on-chip multicore architecture where the last-level cache (LLC) is typically distributed across tiles [13]. This configuration enables scalability, but while each core can directly access the portion of cache in its own tile, it needs to use an interconnection network to access the cache resources in other tiles. The tiled-CMP design paradigm is devised to run complex and heterogeneous multi-threaded applications, which require efficient communication and synchronization between threads within the chip. This leads to the need for efficient on-chip interconnection mechanisms such as high-performance and structured network-on-chips (NoCs) for interconnecting the tiles.

The execution time of applications is becoming more and more affected by network traffic and particularly by the average distance traversed to retrieve data from the correct LLC tile in the chip (number of network hops). As the core count increases, the number of retransmissions that messages suffer in the electrical network also increases, compromising the scalability of future chip multiprocessors. In addition, data transmission through the on-chip interconnect is starting to account for most of the energy consumption of a chip, a scenario that is expected to get worse [16, 5]. This problem must be addressed to continue increasing the performance of future chips within a reasonable power budget.

Advances in silicon photonics have enabled the integration of optical interconnects inside silicon chips [9, 11]. This disruptive technology provides low-energy fast data transmission across the whole chip regardless of the distance, and can be a solution to the scalability problems of NoCs. For instance, transmitting information between two opposite corners of a 8×8 electronic mesh at 4 GHz requires traversing fifteen routers and fourteen inter-tile links, taking up tens of processor cycles (also at 4 GHz). Traversing the same distance in a photonic waveguide[1] can take as little as two processor cycles, needs no retransmissions, and uses significantly less energy in the process.

Since silicon-photonic integration became feasible for CMP interconnects, a myriad of photonic networks have been proposed. Many topologies have been studied, from simple photonic rings [24, 20, 25] that operate like a crossbar, to complex articulated topologies [23, 22] that require or combine different transmission technologies, and to logical all-to-all interconnect designs [18]. Some complex photonic interconnects use supporting circuit-establishing electrical networks [23, 22] which limit severely the latency and energy advantages of nanophotonics in scenarios like hardware-cache-coherent CMPs. We focus on a simple photonic structure (ring) instead and show that, if properly managed, it can potentially deliver large latency and energy improvements without needing big investments in complex topologies and/or organization.

Investigation on the maximum benefits achievable from simple optical topologies (e.g., ring-based [21, 24, 20, 25]) is strategic because, for relatively short-term commercial solutions, the use of a simple 3D-stacked photonic network can be an interesting design choice, especially for pursuing low-power solutions.

Hybrid photonic-electronic NoCs attempt to make the most of both transmission technologies [21, 10, 15, 1]. While in the classic electronic heterogeneous network scenario [6, 7] the low-latency network was very power-hungry and was used selectively for accelerating certain messages, in this new scenario the low-latency raw photonic technology enables the faster and less power consuming network of the system in terms of dynamic power. But, when the load increases, it can suffer from long message queuing due to serialization and contention, reducing performance. Furthermore, increasing photonic resources to limit serialization effects introduces increased waveguide crossings, thus higher insertion loss and laser power, which in turn further increases static power consumption (e.g., laser, thermal tuning). Moreover, for very short distances (e.g., 1-hop) also the latency advantage of photonics over electronics can be quite limited due to conversion overhead.

For these reasons, hybrid photonic-electronic NoCs, especially those based on simple physical topologies which will soon be implementable, need to be carefully managed to take advantage of the latency and energy advantages of photonic technology. Currently, there is a lack of adequate policies to carry out this management. Our purpose is to develop mechanisms to make the best use of a photonic network that works in cooperation with an electrical mesh

[1] about 15 ps/mm (group velocity of light into silicon)

in a modern CMP. To our knowledge, these are the first proposed ad-hoc management strategies that use real-time information for hybrid photonic-electronic NoCs at the message level.

The main contributions of this work can be summarized as follows: (1) we propose novel policies to efficiently use photonic resources and test them by enabling a profitable usage of a simple photonic ring in collaboration with an electrical network and (2) we evaluate these policies and analyze their different characteristics in terms of performance and energy consumption, showing that our more advanced policies significantly outperform naïve ones both in execution time and power consumption and demonstrating the importance of appropriate dynamic electro/photonic message management policies. Only the best proposed policy is able to reduce the average execution time of the PARSEC benchmarks by 36% with respect to an electrical mesh while network energy consumption is also reduced by 52% on average.

Methodologically, considering a simple one-waveguide network (a likely representative of near-future CMPs) allows dissecting and tuning policies in isolation from other effects that might occur in more complex network organizations. Nevertheless, the proposed mechanisms are applicable, with appropriate tuning, to different network topologies, sizes and arbitration mechanisms.

The rest of this paper is organized as follows: Section 2 discusses the ring-based photonic architecture used as a case study in this paper. Section 3 describes the management policies designed and their strong and weak points. Sections 4 and 5 evaluate the proposed policies from performance and energy-consumption points of view and, finally, Section 6 concludes.

2. BACKGROUND

The basic elements necessary to build a photonic network are waveguides, light sources, modulators and detectors. *Silicon waveguides* are on-chip channels that carry light modulated to convey information, enabling communication without retransmissions across the chip at higher bitrates and with lower losses than electrical wires, resulting in lower delays. As a *light source*, either on-chip or off-chip lasers are used to inject light into the waveguide. WDM (Wavelenght Division Multiplexing) is employed to transmit information in several channels simultaneously, with one bit of width for each channel, corresponding to a particular wavelength or *carrier*. *Microring resonators* are made of looped optical waveguides [4] that can be used as modulators and filters for sending and receiving information. For instance, when being used for sending information, a microring coupling the light corresponding to its resonant wavelength injects a '0' value. Alternatively, letting pass the light injects a '1'. When being used for receiving, a microring using a germanium doped photodiode detects the injected stream of '0's and '1's.

2.1 Ring-based photonic networks

Due to their simplicity and flexibility, optical topologies based on open or closed waveguides implementing logical *rings* [21, 24, 20, 25] are likely to be implemented in commercial machines before more complex photonic architectures based on photonic switches, *passive* [26, 19] or *active* [23, 22] (i.e., dynamically reconfigurable), as these employ a significant number of optical switches and incur many waveguide crossings that can introduce significant optical attenuations. Active switches need a photonic-circuit establishment mechanism (using a supporting electrical network) to configure the microring resonators before transmission due to the incapacity to implement routing within the optical domain, which has a very unattractive overhead for short messages typical of cache coherence based systems.

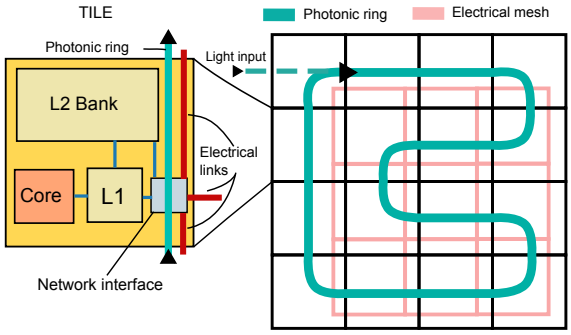

Figure 1: Hybrid photonic-electronic NoC on a 16-core tiled CMP. Every tile can read from or write in the photonic ring.

Figure 1 exemplifies a ring-shaped waveguide in a hybrid photonic-electrical NoC for a 4×4 CMP. Even simple topologies expose various design choices to be made [2]. For example, each origin-destination pair can be assigned to a different set of wavelengths, preventing the need for arbitration. However, this would limit the bandwidth for single transmissions noticeably. Other designs provide flexibility but require some arbitration. A Single Writer Multiple Reader (SWMR) [21, 14] configuration allocates a different set of wavelengths for each writer, which uses a destination selection mechanism to make the desired receiver turn on its photodetectors to read the data. In Multiple Writer Single Reader (MWSR) [24], each receiver reads different wavelengths and arbitration is needed on the writer's side to avoid collisions between writers. Multiple Writer Multiple Reader (MWMR) [20] is the most flexible, allowing a single transmission to use all the data wavelengths of the ring, but requires both arbitration in the writer's side and destination selection as well as more ring modulators and photodetectors for every destination to read and write every wavelength, which increases area and power consumption.

3. DYNAMIC MANAGEMENT POLICIES FOR HYBRID NOCS

This section presents a set of novel policies to manage hybrid networks comprising a ring-based photonic sub-network and an electrical sub-network such as a mesh. They decide which sub-network to use for each message using real time information based on the following parameters: message size, photonic ring availability and distance between endpoints. Table 1 provides a summary.

An adequate interface between the electronic and the phonotic subnetworks is required to apply our management policies. This interface adds little complexity over the one used in previous works (e.g., Firefly [21]). In each tile, one *pre-photonic buffer* interfaces each traffic source (e.g., L1 or L2 caches) with its associated external port to the electrical mesh router. Upon injection, those network messages that are candidate for photonic transmission are first stored in the corresponding pre-photonic buffer, and they are eventually either sent through the photonic ring or transferred to the corresponding external port of the router for electrical transmission.

Only one pre-photonic buffer per node can be *active* at each time, meaning that the message at its head is the one being considered for photonic transmission. As soon as the arbitration token is acquired, the active message is sent through the photonic ring and then the token is reinjected. In addition for some policies, when a message enters a pre-photonic buffer, a countdown timer is set to the appropriate wait time (i.e., AVAIL's 2-processor-cycle token wait). If it reaches zero, photonic transmission is ruled out and the message enters the electrical router.

Table 1: Summary of policies

Policy	Messages on photonic ring	Messages on mesh
SIZE	control messages (8 bytes)	data messages (72 bytes)
AVAIL-x	token acquired within x cycles	other messages (x-cycle delay)
DDA-x	token acquired within $(l_m - l_p) \times 0.x$ cycles	other messages ($(l_m - l_p) \times 0.x$ delay)
CDDA-x	control if token acquired within $(l_m - l_p) \times 0.x$ cycles data if token acquired within 2 cycles	other messages ($(l_m - l_p) \times 0.x$-cycle delay for control, 2-cycle delay for data)
MTDDA-x-y	control if token acquired within $(l_m - l_p) \times 0.x$ cycles, data if token acquired within $(l_m - l_p) \times 0.y$ cycles	other messages ($(l_m - l_p) \times 0.x$-cycle delay for control, $(l_m - l_p) \times 0.y$-cycle delay for data)

l_m = idle mesh latency, l_p = idle photonic ring latency.

3.1 SIZE: Message size

Our first criterion to decide which subnetwork to use is extremely simple, taking into consideration just the size of the message. There are typically two different kinds of messages in a cache-coherent CMP: control messages and data messages, with typical respective sizes of 8 and 72 bytes (the 64-byte difference is accounted by the data block) [17]. Transmitting a data message makes the photonic ring unavailable for much longer than a control message, potentially increasing the latency of other messages.

For instance, our chosen near-future network (see Section 4) comprises a high-performance waveguide with 64 data wavelengths that enables the transmission of up to 8 bytes per ring-cycle. Therefore, sending control and data messages takes 1 and 9 ring-cycles, respectively. In 9 ring-cycles, nine short messages (one per ring-cycle) could be sent, greatly benefiting from the low latency of the ring. That is, sending long data messages increases the chances of creating long message queuing times due to serialization, and decreases the opportunities to accelerate many other messages.

In addition, short messages account for just a small percentage of the overall traffic in bytes due to their small size, although they account for most of the messages injected. Thus, their acceleration provides a large performance gain with small bandwidth usage.

All of this makes it interesting to use a simple policy, which we call *SIZE*, that only sends through the photonic ring those messages of small size (control). Notice that the opposite policy (sending data messages on the photonic ring) would be prone to serialization, resulting in high queuing times.

This policy has some shortcomings as it cannot adapt to the burst nature of traffic in parallel applications, underutilizing photonic resources under low traffic loads (especially in large photonic NoCs) and serializing and delaying many messages under high traffic loads (especially if there are limited photonic resources).

3.2 AVAIL: Ring availability

This policy sends a message (control or data) through the photonic ring only if the ring is readily available when the transmission is attempted or before a parametrized wait time passes. Hence, the electrical mesh is used for those messages that find the photonic ring busy. This adjusts dynamically the traffic injected in each subnetwork, preventing the shortcomings of *SIZE*.

Since we have to acquire the token before transmitting, we must wait for the token round-trip time (2 processor-cycles in the NoC evaluated in Section 4) before knowing whether the ring is busy or not. If the token is acquired, then the message is sent through the ring. If it is not, the ring is busy and the message is sent through the mesh (after having waited unfruitfully during 2 processor-cycles).

With this policy we make sure that messages are never serialized waiting for the photonic ring under high traffic load scenarios, and that every message has a chance of using the photonic ring. Also, for low traffic loads, we prevent messages from being sent through the mesh while the photonic ring is free, hence increasing ring utilization and reducing overall energy consumption.

Table 2: Hops and link traversals in a 4×4 mesh for messages to access the LLC that leave the tiles.

Hops	Msgs.	Link traversals	Aggr. msgs.	Aggr. link traversals
1	20.0%	7.5%	20.0%	7.5%
2	28.3%	21.3%	48.3%	28.8%
3	26.7%	30.0%	75.0%	58.8%
4	16.7%	25.0%	91.7%	83.8%
5	6.7%	12.5%	98.3%	96.3%
6	1.7%	3.8%	100.0%	100.0%

3.3 DDA: Distance Dependent Availability

The energy consumption and latency of a message transmission through an electrical mesh varies depending on the distance between the origin and the destination. In a 4×4-CMP, sending a message between opposite corners of the chip requires six intermediate routing operations and six retransmissions through inter-tile electrical links. Transmitting the same message between adjacent tiles requires just one routing operation and one transmission on the link connecting the tiles. Therefore, communication between distant destinations requires much more energy and takes longer.

This is also why cores in the corners and borders of the chip suffer from longer average latencies than those in the center [8], making threads running on them execute more slowly than those running on central cores and harming parallel application performance, as slower cores limit the overall execution speed.

Table 2 shows the percentage of messages and link traversals caused by communications between cores at different distances (in number of hops) for a uniform random distribution of accesses to a NUCA last-level cache in a 16-core CMP, which matches our observations on the PARSEC benchmarks. Transmissions between neighbor nodes account for 20% of messages but only generate 7.5% of link traversals, while 5-hop transmissions account for only 6.7% of messages but generate a significant 12.5% of link traversals. Moreover, the half of messages between closest nodes generates just 30% of link traversals, while the other half (between furthest nodes) generates 70% of link traversals. A smart use of the photonic ring, which is almost insensitive to distance, consists of reserving it for messages that incur the most energy consumption and latency if transmitted through the mesh (i.e., between corners) and using the mesh for many short-distance transmissions. This would also make cores far from the center of the chip benefit greatly, making them catch up with center cores, boosting performance even further.

To leverage this, we have developed an heuristic policy called Distance Dependent Availability (DDA), which combines the benefits of preventing long waits for the photonic ring and favoring its

use for distant endpoint communications. We achieve this mix of goals by allowing a different token-wait time for each particular message that is proportional to the theoretical advantages of using the photonic ring over using the mesh. To calculate this advantage of the ring, we use the best case theoretical latency of transmitting each message in the ring (l_p) and in the mesh (l_m) in the absence of other transmissions. In our setup $l_p = 2$ or $l_p = 5$ processor cycles in case of a control (8-byte) or data (72-byte) message, respectively and, correspondingly, $l_m = 5/hop$ processor cycles for control messages and $l_m = 5/hop + 8$ for a data message. Every message is, at first, considered for sending through the photonic ring. If the ring is found idle, the message is sent. Otherwise, the message waits for $(l_m - l_p) \times th$ cycles, where th is a configurable threshold with values between 0 (no wait) and 1 (wait as long as there is any potential benefit in using the ring), before sending the message through the mesh. Small values of th avoid serialization of messages, while large values increase ring utilization. In any case, messages involving distant endpoints wait longer, hence acquiring the token and using the photonic ring more often. In the evaluation section we consider several values for th to explore possible trade-offs.

In order to account for the different message sizes of the messages and capture the potential benefits, we give differentiated treatment to them in the photonic ring in two additional heuristic policies explained below that add message size to the variables considered for photonic ring management.

Control DDA (CDDA) consists of using DDA for control messages and AVAIL for data messages. This policy tries to obtain an average low message latency by transmitting many short messages, while retaining a high utilization of the network by sending data messages when the ring is otherwise idle.

Multi-Threshold DDA (MTDDA) uses DDA for both control and data messages, but different thresholds are used for each message type. A longer threshold is used for control messages to prioritize their transmission in the ring. In this case we give more importance to the ring utilization than in CDDA, since data messages are now more likely to be transmitted.

4. EVALUATION METHODOLOGY

We have tested the policies using detailed full-system simulation on a photonic ring based on FlexiShare [20]. FlexiShare [20] is a MWMR photonic ring proposed for a 64-core CMP. We scale down FlexiShare to just one waveguide, resulting in an affordable design suitable for near-future commercial CMPs that requires just around 2000 microring resonators. Figure 1 shows our base hybrid photonic-electrical NoC in a 4×4 CMP. We consider a realistic five ring-cycle full-ring traversal time at 10 GHz for light pulses (i.e., two processor cycles at 4 GHz). All of the data wavelengths of the ring can be simultaneously used by one emitter to communicate with one receiver, and we limit the number of concurrent transmissions in the ring to just one. In all, this FlexiShare-like configuration requires 65 wavelengths for its operation. During destination selection, four wavelengths identify the receiver and one wavelength indicates the size of the message to transmit. Sixty-four wavelengths are used for data transmission. One extra wavelength is needed to circulate the single-bit token in which the selected arbitration mechanism is based.

Transmitting on the 10 GHz MWMR photonic ring when it is idle requires to acquire the free token (up to five ring cycles), activate the destination's receivers (three ring cycles) and then reach the destination with the first photonic pulse (up to five ring cycles). In the case of control messages this is enough to transmit the 8-byte message. In the case of data messages a second photonic pulse carries the eight-byte requested word (the first pulse contains an 8-byte header). The rest of the data block is transmitted in consecutive photonic pulses. In the most favorable case (no wait for the token and a transmission to the closest node), an idle photonic ring provides a 2-processor-cycle transmission latency (rounded up) for control messages or for the requested word of data messages. In the most unfavorable case (round-trip time wait for the token and transmission to the farthest node) this latency increases to 6 processor-cycles. In comparison, the idle 4×4 electrical mesh requires up to 31 processor-cycles between the most distant destinations.

Table 3: Simulated machine

Processors
16 Alpha cores @ 4 GHz, 2-ways, in-order
L1 Cache
Split I&D. Size: 16 KB, 4-ways, 64 bytes/block
Access latency: 1 cycle
MOESI coherence protocol (directory cache in L2 cache)
L2 Cache
Size: 1 MB per bank. 16 MB total (NUCA), 8-ways, 64 bytes/block
Access latency: 15 cycles
RAM
4 GB DDR2 DRAM
16 3D-stacked memory controllers
Interconnection - Electronic
4 GHz, 2D mesh: 4×4
16-byte bi-directional links
Latency: 1 processor-cycle/link, 4-processor-cycle pipelined routers
Flit Size: 16 bytes
Control/Data packet size: 8/72 bytes (1/5 flits)
Dynamic energy (1-hop switch+link): 282 pJ/flit
Static power (switch+link): 52.7 mW
Interconnection - Photonic
10 GHz MWMR Photonic Ring, 3D-stacked, 65 wavelengths
Latency: 2 processor-cycle round-trip time
2 processor-cycle minimum transmission time on idle ring (no token wait, closest node)
6 processor-cycle maximum transmission time on idle ring (roun-trip time token wait, furthest node)
Flit Size: 8 bytes.
Control/Data packet size: 8/72 bytes (1/9 flits)
Dynamic energy: 0.41 pJ/bit
Static power (laser+microrings): 318 mW

As for flow control, in case of buffer overflow a NACK signal is sent by the receiver through the data wavelength in the complementary ring portion to the transmission (hence closing the circle without disturbing any other photonic transmission) and it is read by the transmitter that backs down, releases the token and repeats the same transmission procedure again after some time. This flow control mechanism is rarely needed and has little impact in the overall performance. More complex mechanisms can be used.

The GEM5 simulator [3] was used to perform the evaluation. The common system characteristics can be found in Table 3. The L2 cache uses a NUCA design and a directory-based MOESI protocol enforces coherence between the private L1 caches. Time and power parameters of the electronic NoC are derived from Orion 2.0 [12] using a 32 nm silicon process. We consider state-of-the-art optical devices [27] and their behavior in our reference architecture.

We have used the PARSEC benchmark suite with the medium-sized working sets to perform this study. We evaluate the SIZE policy, the AVAIL policy with three different wait times (2, 6 and 10 processor cycles), the DDA and CDDA policies with three different thresholds (25%, 50% and 75%) and the MTDDA policy with two different configurations (60%-40% and 75%-25% thresholds

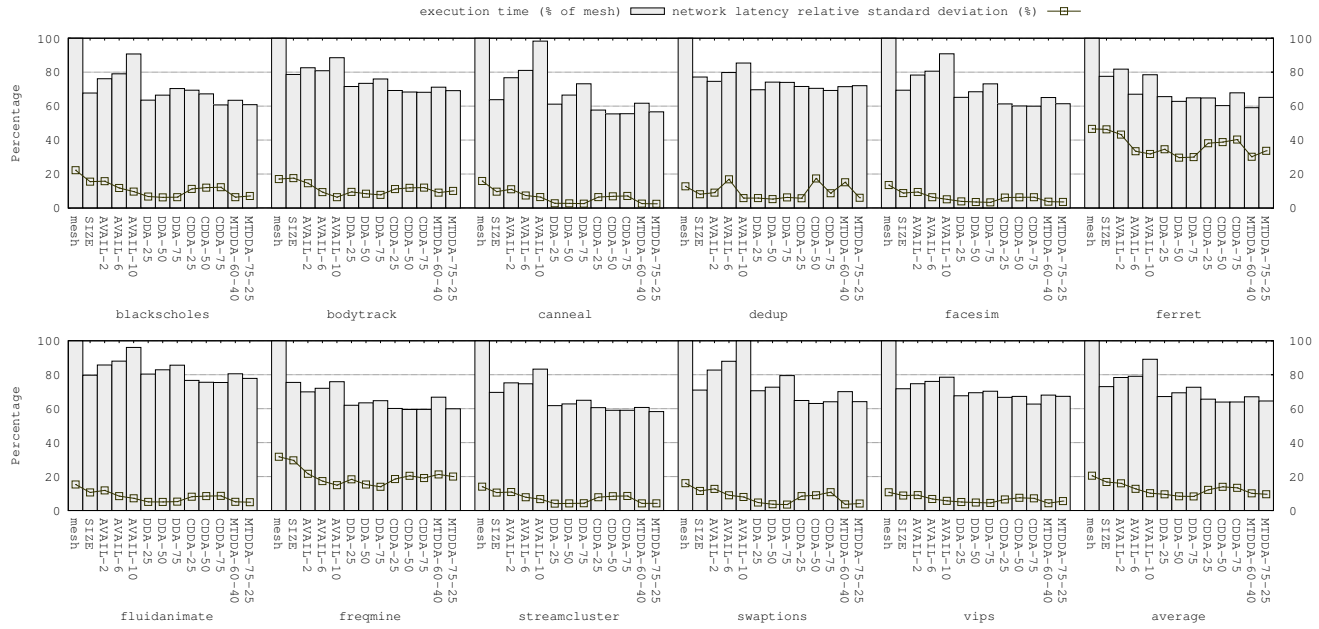

Figure 2: Execution time. Normalized to electronic mesh.

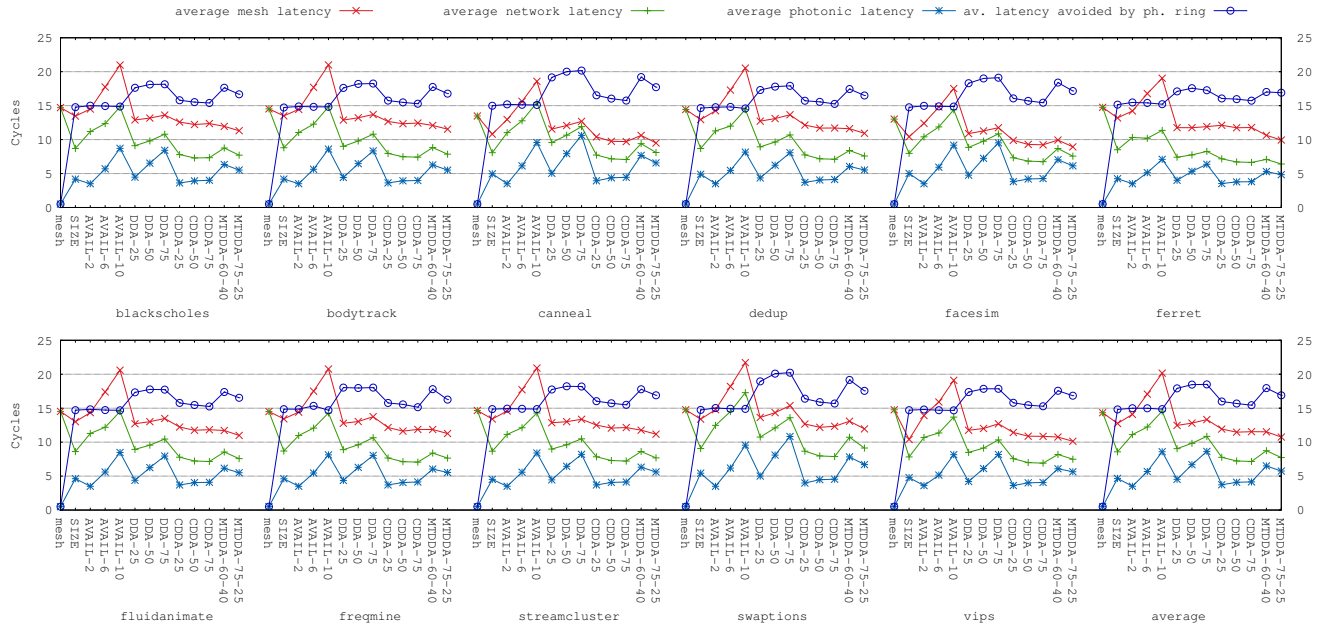

Figure 3: Network latency. Normalized to electronic mesh.

for control and data messages). Further details on these policies can be found in Table 1 and in Section 3.

5. EVALUATION RESULTS

Figure 2 shows, for each policy, the execution time and the relative standard deviation of the network latency suffered by the cores. Figure 3 shows the average latencies for message transmissions in the electrical mesh, in the photonic ring, and overall. Although the frequencies of these networks are different (4 GHz for the electrical and 10 GHz for the photonic), all the data are plotted at 4 GHz. We also show the theoretical latency gain for the messages that are finally sent through the photonic network. Figure 5 shows the energy consumption of the electrical mesh and the photonic ring, along with the photonic network usage and the fraction of messages sent through it. A higher percentage of messages does not imply a higher utilization of the ring, since two different message sizes exist.

In general, all of the photonic management policies improve performance compared to the baseline electrical mesh. Also, the trends shown by each policy remain stable across all benchmarks.

The SIZE policy reduces execution time by 27% with respect to the baseline mesh thanks to a reduction in the latency of control

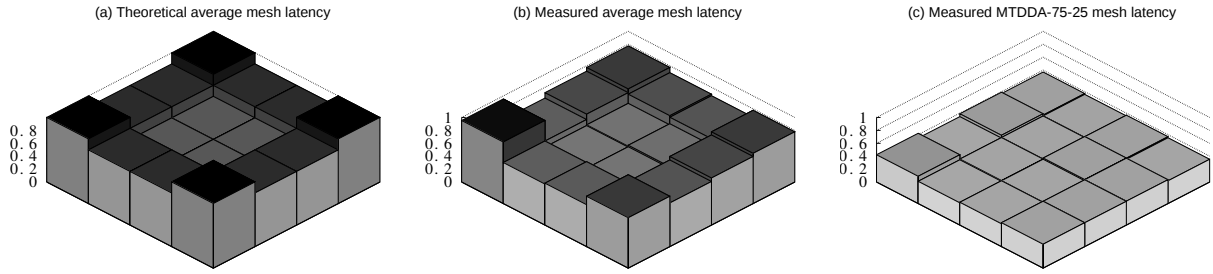

Figure 4: Total network latency in the critical path of L1 cache misses suffered by each core. Average results for electrical mesh and MTDDA-75-25. The data is normalized to the core with the highest network latency in the mesh.

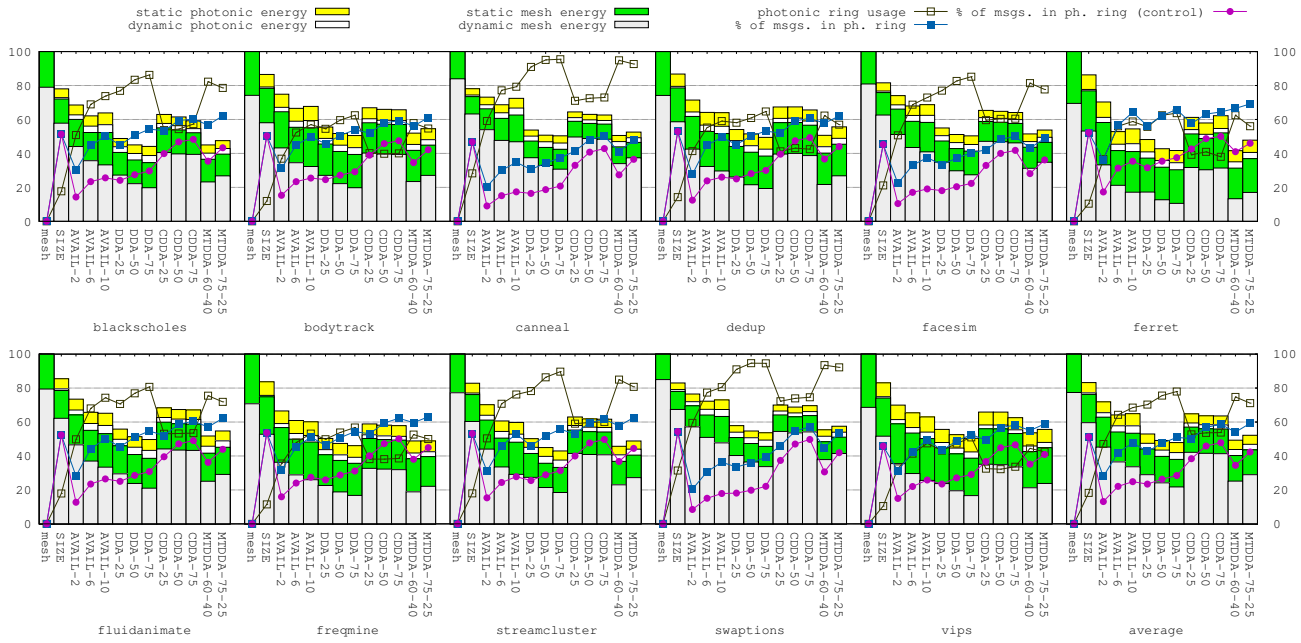

Figure 5: Network energy consumption (normalized to electronic mesh) and photonic ring usage.

messages. The latency gets reduced by 40% on average. In general, the wait time for acquiring the ring is low (1.2 cycles) thanks to the small size of the messages transmitted, avoiding serialization. However, the ring is underutilized because many times there are no control messages to transmit. Nevertheless, a 17% reduction of the energy consumption of the network is achieved with respect to the base architecture.

The execution time of AVAIL is noticeably higher than SIZE's, as AVAIL-2 reduces execution time by just 21% compared to the baseline. The latency reduction is also lower. The reason is the larger average size of messages transmitted through the ring (control and data in AVAIL, only control in SIZE) that causes fewer messages to be accelerated compared to SIZE (each data message sent prevents sending up to 9 control messages). As the number of wait cycles increases —AVAIL-6 and AVAIL-10— the average network latency increases too, resulting in average reductions of just 20% and 11% in time over the baseline. However, the sending of data increases the usage of the ring (48% in AVAIL-2), increasing the energy reduction (28% in AVAIL-2). In addition, increasing the wait cycles also increases the usage of the ring, which results in a noticeable reduction of the energy consumption of the network. We can say that AVAIL prioritizes energy reduction over execution time when compared to SIZE. Also, the wait threshold provides a way to tune between lower execution time and lower energy consumption.

Distance based policies make more efficient use of photonics and improve the performance results of SIZE and AVAIL. DDA-25 reduces execution time by 33%. As the policy threshold goes up, the execution time average benefit is reduced. This performance drop is caused by the increase in the average wait times to transmit when using higher thresholds. This policy provides lower network latencies on average because the photonic ring is preferentially used to send long-distance messages, while the is now mainly used for short-distance messages. DDA achieves the highest reductions in energy because its photonic ring usage is the highest and the ring is used for messages between distant endpoints, reducing the need for retransmissions in the mesh. Similarly to AVAIL, DDA's thresholds are useful to trade-off between speed and energy consumption.

CDDA provides the highest reduction in execution time. CDDA-25 reduces execution time by 33%, and this value increases to 36% for CDDA-50 and CDDA-75. In CDDA, the distance-dependent thresholds keep high the mesh latency avoided by the ring, although lower than in DDA because now most control messages use the photonic ring, including messages between close endpoints.

The latency reduction is the highest of any policy thanks to the higher number of messages accelerated. The energy consumption of CDDA is higher than DDA's.

Finally, MTDDA performs close to CDDA in execution speed and network latency, with 33% and 35% lower execution times than the baseline for MTDDA-60-40 and MTDDA-75-25, respectively. These policies allow some wait for data messages, based on distance, in order to achieve a balance between execution time speedup and photonic ring utilization. By using different thresholds for control and data we can still prioritize the sending of short messages in order to reduce execution time while retaining the ability to achieve a high utilization of the photonic ring with data messages. This results in a good trade-off between DDA and CDDA. We believe that MTDDA is the most versatile policy, providing good results in both execution time (almost matching CDDA) and energy consumption (close to DDA).

As mentioned in section 3.3, cores in the center of the chip suffer less latency when acessing the LLC. Figure 4 shows the overall absolute latency suffered by each core for the electrical mesh and for MTDDA-75-25 in each benchmark. The pattern observed in the mesh (subfigure b) matches closely the theoretical analysis (subfigure a) which assumes a completely uniform communication pattern. Subfigure c gives a clear view of how those cores that suffer longer latencies in the mesh are the most benefited by MTDDA-75-25 (corners and borders of the chip). The threads running on these cores get a higher speed-up and their performance can now match that of the threads running in the central cores. Figure 2 also shows that our policies reduce the standard deviation of the network latency suffered by cores. This reduction is especially noticeable for DDA and MTDDA where it reaches 60%. This means that a much more homogeneous network latency is perceived by all cores, which has the important benefit of reducing the wait times for thread synchronization. This provides a noticeable portion of the acceleration of applications seen in Figure 2 by reducing wait times between threads.

6. CONCLUSIONS

In this paper we have shown the importance of using adequate management policies to enable efficient use of photonic resources in an hybrid photonic-electronic network. We have proposed the first dynamic fine-grain policies to enable such management, based on distance between endpoints, ring availability and message size at a message granularity. We have tested these policies both on an near-future affordable photonic ring, obtaining significant performance improvements and energy consumption reductions. By using photonics for the messages most likely to benefit from it (distant, short, and keeping low wait times), we reduce the number of electric mesh retransmissions (that cause large energy consumption and latency) while preventing serialization in the ring. In addition, these policies level out the network latency suffered by all cores in the chip compared to an electrical mesh, resulting in an additional performance boost. A proposed performance oriented policy (CDDA-75) reduces execution time by 36%, and a proposed energy oriented policy (DDA-75) reduces network energy consumption by 52% for the PARSEC benchmark suite in a 16-core CMP. In addition, we propose a balanced policy (MTDDA-75-25) which reduces execution time by 35% and reduces network energy consumption by 48%.

7. ACKNOWLEDGMENTS

Part of this work was performed by Antonio García Guirado during a research stay at the University of Siena (February–August 2012) partially supported by a HiPEAC collaboration grant. This work was also supported by the Spanish MEC and European Commission FEDER funds under grants "Consolider Ingenio-2010 CSD2006-00046" and "TIN2009-14475-C04-02" and by IT FIRB Photonica (RBFR08LE6V) and IT PRIN 2008 (200855LRP2) projects. Antonio García-Guirado is also supported by a research grant from the Spanish MEC under the FPU National Plan (AP2008-04387).

References

[1] S. Bahirat and S. Pasricha. A particle swarm optimization approach for synthesizing application-specific hybrid photonic networks-on-chip. In *13th International Symposium on Quality Electronic Design (ISQED)*, pages 78–83. IEEE, 2012.

[2] C. Batten, A. Joshi, V. Stojanovic, and K. Asanovic. Designing chip-level nanophotonic interconnection networks. *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, 2(2):137–153, 2012.

[3] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, and D. A. Wood. The gem5 simulator. *SIGARCH Computer Architecture News*, 39(2), Aug. 2011.

[4] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets. Silicon Microring Resonators. *Laser & Photonics Reviews*, 6(1):47–73, 2012.

[5] S. Borkar and A. A. Chien. The future of microprocessors. *Communications of the ACM*, 54(5):67–77, May 2011.

[6] L. Cheng, N. Muralimanohar, K. Ramani, R. Balasubramonian, and J. B. Carter. Interconnect-aware coherence protocols for chip multiprocessors. In *Proceedings of the 33rd annual international symposium on Computer Architecture (ISCA)*, pages 339–351, 2006.

[7] A. Flores, J. L. Aragón, and M. E. Acacio. Heterogeneous Interconnects for Energy-Efficient Message Management in CMPs. *IEEE Transactions on Computers*, 59(1):16–28, 2010.

[8] A. García-Guirado, R. Fernández-Pascual, A. Ros, and J. M. García. DAPSCO: Distance-aware partially shared cache organization. *ACM Trans. Archit. Code Optim.*, 8(4):25:1–25:19, Jan. 2012.

[9] C. Gunn. CMOS Photonics for High-Speed Interconnects. *IEEE Micro*, 26(2):58–66, Mar. 2006.

[10] G. Hendry, S. Kamil, A. Biberman, J. Chan, B. G. Lee, M. Mohiyuddin, A. Jain, K. Bergman, L. P. Carloni, J. Kubiatowicz, L. Oliker, and J. Shalf. Analysis of photonic networks for a chip multiprocessor using scientific applications. In *Proceedings of the 2009 3rd ACM/IEEE International Symposium on Networks-on-Chip (NOCS)*, pages 104–113, Washington, DC, USA, 2009. IEEE Computer Society.

[11] B. Jalali and S. Fathpour. Silicon photonics. *Lightwave Technology, Journal of*, 24(12):4600–4615, dec. 2006.

[12] A. B. Kahng, B. Li, L.-S. Peh, and K. Samadi. ORION 2.0: a Fast and Accurate NoC Power and Area Model for Early-Stage Design Space Exploration. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE)*, pages 423–428, 2009.

[13] C. Kim, D. Burger, and S. W. Keckler. An Adaptive, Non-Uniform Cache Structure for Wire-Delay Dominated On-Chip Caches. In *Proceedings of the 10th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 211–222, 2002.

[14] N. Kirman, M. Kirman, R. K. Dokania, J. F. Martinez, A. B. Apsel, M. A. Watkins, and D. H. Albonesi. Leveraging op-

tical technology in future bus-based chip multiprocessors. In *Proc. of the 39th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pages 492–503, 2006.

[15] Z. Li, D. Fay, A. R. Mickelson, L. Shang, M. Vachharajani, D. Filipovic, W. Park, and Y. Sun. Spectrum: a hybrid nanophotonic-electric on-chip network. In *46th Annual Design Automation Conference (DAC)*, pages 575–580, 2009.

[16] N. Magen, A. Kolodny, U. Weiser, and N. Shamir. Interconnect-power dissipation in a microprocessor. In *Int'l workshop on System Level Interconnect Prediction (SLIP)*, pages 7–13, Feb. 2004.

[17] M. M. K. Martin, M. D. Hill, and D. J. Sorin. Why on-chip cache coherence is here to stay. *Communications of the ACM*, 55:78–89, 2012.

[18] C. Nitta, M. Farrens, and V. Akella. DCAF - A Directly Connected Arbitration-Free Photonic Crossbar For Energy-Efficient High Performance Computing. In *26th International Parallel & Distributed Processing Symposium (IPDPS)*, pages 1–12, 2012.

[19] I. O'Connor, D. Van Thourhout, and A. Scandurra. Wavelength division multiplexed photonic layer on cmos. In *Proceedings of the 2012 Interconnection Network Architecture: On-Chip, Multi-Chip Workshop*, pages 33–36, 2012.

[20] Y. Pan, J. Kim, and G. Memik. Flexishare: Channel sharing for an energy-efficient nanophotonic crossbar. In *Proc. of the 16th Int.l Symposium on High-Performance Computer Architecture (HPCA)*, pages 1–12. IEEE CS, 2010.

[21] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary. Firefly: Illuminating future network-on-chip with nanophotonics. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*, pages 429–440, 2009.

[22] M. Petracca, B. Lee, K. Bergman, and L. Carloni. Design exploration of optical interconnection networks for chip multiprocessors. In *16th IEEE Symposium on High Performance Interconnects*, pages 31–40, 2008.

[23] A. Shacham, K. Bergman, and L. P. Carloni. Photonic networks-on-chip for future generations of chip multiprocessors. *IEEE Trans. Comput.*, 57(9):1246–1260, Sept. 2008.

[24] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn. Corona: System implications of emerging nanophotonic technology. In *Proceedings of the 35th Annual International Symposium on Computer Architecture (ISCA)*, pages 153–164, 2008.

[25] Y. Xu, Y. Du, Y. Zhang, and J. Yang. A composite and scalable cache coherence protocol for large scale CMPs. In *Proceedings of the International Conference on Supercomputing (ICS)*, pages 285–294, 2011.

[26] L. Zhang, M. Yang, Y. Jiang, E. Regentova, and E. Lu. Generalized wavelength routed optical micronetwork in network-on-chip. In *Proc. of 18th IASTED Int.l Conference on Parallel and Distributed Computing and Systems*, pages 698–703, 2006.

[27] X. Zheng, D. Patil, J. Lexau, F. Liu, G. Li, H. Thacker, Y. Luo, I. Shubin, J. Li, J. Yao, P. Dong, D. Feng, M. Asghari, T. Pinguet, A. Mekis, P. Amberg, M. Dayringer, J. Gainsley, H. F. Moghadam, E. Alon, K. Raj, R. Ho, J. E. Cunningham, and A. V. Krishnamoorthy. Ultra-efficient 10gb/s hybrid integrated silicon photonic transmitter and receiver. *Optics Express*, 19(6):5172–5186, 2011.